Characterisation of the Rex1 exonuclease domain and RYS domain

Peter W Daniels

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Science

School of Biosciences

September 2023

Characterisation of the Rex1 exonuclease domain and RYS domain

Peter W Daniels

September 2023

Abstract

Gene expression requires the accurate production of RNA from DNA. Raw transcription products are usually non-functional, requiring processing by nucleases to reach a functional, mature state. If a transcript is not processed correctly it will be targeted by nucleases for degradation; in some cases the same nuclease required for maturation can mediate degradation. By understanding the mechanisms nucleases use to bind their substrates, it is possible to understand nucleases features that direct a transcript towards maturation or degradation. This study aims to dissect the substrate recognition mechanisms of Rex1, a 3´ exonuclease belonging to the DEDD nuclease superfamily found in the model eukaryote *Saccharomyces cerevisiae*. Rex1 is unique in the yeast proteome for its ability to precisely trim short stem-loop adjacent 3´ overhangs found at the end of small RNA precursors including tRNA and 5S rRNA. The equivalent processing event is performed in *Escherichia coli* by a related DEDD exonuclease RNase T, however the sequence features responsible for substrate binding in RNase T are not found in Rex1. This study uses phylogenetic analysis, sequence conservation, and structure prediction to identify candidate substrate-binding features within the Rex1 exonuclease domain, and defines a novel Alphafold2-predicted 'RYS' domain that is conserved throughout eukaryotes. An *in vitro* biochemical analysis of recombinant Rex1 is performed by observing the trimming of single-stranded oligo substrates. These trimming assays are used alongside *in vivo* complementation and RNA crosslinking to examine the contribution of Rex1 prospective substrate-binding features to function. This study demonstrates that the RYS domain is central to Rex1 substrate binding, which may be conserved in the structural homologues of Rex1 that are found in eukaryotes spanning plants to humans.

Acknowledgements

Firstly, I would like to thank my primary supervisor, Phil Mitchell. I couldn't have asked for a better supervisor; his enthusiastic instruction and guidance have made my PhD an overwhelmingly positive experience.

I would like to thank my second supervisor John Rafferty, not only for his assistance in crystallography, but also for the advice and support he has given me throughout my studies.

I would like to thank my advisors, Pat Baker and Stuart Wilson, and additionally Dan Bose and Emma Thomson for their advice and support throughout my PhD.

I would like to thank my predecessor Taib Hama Soor, whose PhD work laid the foundations for this project. I am also grateful to Sophie Kelly, who cloned plasmids that have been tremendously useful in this work.

I am very grateful for the contributions that Sveta Sedelnikova of the University of Sheffield Protein Purification Facility; her technical expertise in protein purification was essential to the crystallography component of this project. I am also grateful to Diamond Lightsource for their diffraction screening services.

I am very grateful to my fellow Mitchell Lab members who have supported me over the years: Quentin Levicky, Hunouf Mohammad, Selva Turkolmez, Kat Jameson, Sophie Kelly, Iwan Tebbs, Gareth Owen, Ffion Jones, Heppy Robb, Lizzy O'Leary, and Anamaria Buzoiano. You've made E19 a joyful place to be.

I feel very fortunate to have shared office space with a supportive group of colleagues, who have created a wonderful working environment throughout my PhD: from the Bose laboratory Petra Celadova, Nicola Carruthers, Katie Gelder, Laura Harrison, Archna Shah, Grace Gilbert, and Bia Adawiyah; from the Thomson laboratory Moni Feigenbutz, Holly Sutherland, Vincent Chan, Sara Ilic, Charlotte Vandermeulen, and Jo Cunningham. I'd particularly like to thank writing club, who have been an amazing support network during the writing of this thesis.

I am deeply grateful to my parents and brother, who have been an unwavering source of support and encouragement throughout my studies.

My final thanks I give to my partner Sarah, you've brightened every day of my PhD.

Contents

List of Figures

List of Tables

Nomenclature

A. thaliana	*Arabidopsis thaliana*
B. cereus	*Bacillus cereus*
C. elegans	*Caenorhabditis elegans*
D. melanogaster	*Drosophila melanogaster*
E. coli	*Escherichia coli*
H. sapiens	*Homo sapiens*
M. musculus	*Mus musculus*
P. aeruginosa	*Pseudomonas aeruginosa*
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
Z. mays	*Zea mays*
G. max	*Glycine max*
S. mutans	*Streptococcus mutans*
aa	Amino acid
Alphafold DB	Alphafold database
AMP	Adenosine mono-phosphate
AP	Alkaline phosphatase
ATP	Adenosine tri-phosphate
BSA	Bovine serum albumin
CFU	Colony forming units
CUT	Cryptic unstable transcript
DEDD	Asp-Glu-Asp-Asp

DEDDh	Asp-Glu-Asp-Asp-His
DEDDy	Asp-Glu-Asp-Asp-Tyr
DELTA-BLAST	Domain enhanced lookup time accelerated-Basic Local Alignment Search Tool
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dnaQ	DNA polymerase III
dNTP	Deoxynucleotide tri-phosphate
DrRP	DNA-directed RNA polymerase
dsRNA	Double-stranded ribonucleic acid
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
FATCAT	Flexible structure alignment by chaining aligned fragment pairs allowing twists
FOA	5-Fluoroorotic acid
GST-tagged	Glutathione-S-transferase
GTC	Guanidinium thiocyanate
GTEx	Genotype tissue expression
HF	High-fidelity
HRDC	Helicase and RNase D C-terminal
HTP	His-6xTEV-Protein A
IgG	Immunoglobulin G
IPTG	Isopropyl beta-D-1-thiogalactopyranoside
iTOL	Interactive tree of life
kb	Kilobase
Lhp1	Lupus antigen homology protein
LSm	Like Sm

mad	Mean absolute deviation
mRNA	Messenger ribonucleic acid
NBS	Nucleotide binding site
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NLS	Nuclear localisation sequence
NNS	Nrd1-Nab3-Sen1
NTA	Nitrilotriacetic acid
OD	Optical density
Oligo	Oligonucleotide
ORF	Open reading frame
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethlyene glycol
PLB	Protein loading buffer
pLDDT	Predicted local distance difference test
PNK	Polynucleotide kinase
polA	DNA polymerase I
polyA	Poly adenosine
polyU	Poly uracil
PPM	Pentophosphomutase
PPM	Phosphopentomutase
PROMPT	Promoter upstream transcript
rDNA	Ribosomal deoxynucleic acid
Rex1	RNA exonuclease 1
RISC	Ribonucleic acid-induced silencing complex
RMSD	Root mean square of deviation
RNA	Ribonucleic acid

RNAi	RNA interference
RNAP	RNA Polymerase
RNase	Ribonuclease
RNP	Ribonucleoprotein
RRM	RNA recognition motif
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription PCR
RTD	Rapid tRNA decay
RYS	Rex1-Yfe1-SDN5
S	Svedburg unit
SAFA	Semi-automated footprint analysis
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
siRNA	Small interfering ribonucleic acid
snoRNA	Small nucleolar ribonucleic acid
snRNA	Small nuclear ribonucleic acid
SRP	Signal recognition particle
ssDNA	Single-stranded deoxyribonucleic acid
SSPE	Sodium chloride-sodium phosphate-EDTA
ssRNA	Single-stranded ribonucleic acid
TBE	Tris-borate-EDTA
TBST	Tris-buffered saline-TWEEN-20
TCA	Trichloroacetic acid
TEMED	N,N,N´,N´-Tetramethylethane-1,2-diamine
TGS	Tris-glycine-sodium dodecyl sulphate
TPM	Transcripts per million
TRAMP	Trf4/Air2/Mtr4

tRNA	Transfer ribonucleic acid
TSEN	tRNA splicing endonuclease
UTR	Untranslated Region
UV	Ultra-Violet
WT	Wild type

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously presented for an award at this, or any other, university. The following publication has arisen from work presented in this thesis:

Daniels, P*., Hama Soor, T*., Levicky, Q., Hettema, E. H. and Mitchell, P. (2022), 'Contribution of domain structure to the function of the yeast DEDD family exoribonuclease and RNase T functional homologue, Rex1', *RNA* **28**(4), 493-507.

* These authors share join first authorship, and contributed equally to this work.

Chapter 1

Introduction

1.1 Nucleases define gene expression.

Nucleases are critical to gene expression, with this criticism taking the form of phosphodiester bond breakage. Evolution takes a broad approach to this end, with nuclease families harnessing a range of catalytic approaches with little correlation to biological function (for an extensive review of the full diversity of nucleases, see Yang, 2011).

Nuclease activity is required at every stage of RNA synthesis, as transcription by DNA-directed RNA polymerase (DrRP; commonly abbreviated to RNA polymerase or RNAP) innately polymerises more RNA than is necessary. This is a consequence of the mechanistic imprecision at each stage of transcription from initiation to termination.

Stochastic initiation throughout euchromatic regions of the genome can result in RNA production independent of promoter motifs, described as 'pervasive transcription' (Wyers *et al*., 2005; Thompson and Parker, 2007; Lee *et al*., 2008; for recent review see Villa *et al*., 2023). These pervasive transcripts are thought to be degraded to avoid titrating the components of RNA metabolism away from canonical gene expression. At canonical promoter sites for RNAPII, initiation seems to be bi-directional with antisense transcripts arising that are rapidly degraded (Seila *et al*., 2008; Neil *et al*., 2009; Xu *et al*., 2009). Further bidirectional transcription is observed roughly 1.5kb upstream of such promoters in mammalian cells (PROMoter upstream Transcripts, or 'PROMPTs'; Preker *et al*., 2008).

Transcription termination is a multi-faceted source of phosphodiester bonds, requiring numerous cleavage events (for recent review, see Xie *et al*., 2023). Premature transcription termination can give rise to truncated transcripts incapable of function, again risking titration of gene expression apparatus and requiring degradation. At the other extreme, a failure to terminate transcription may lead to unnecessary extension of a transcript that will be either trimmed off or serve to target the transcript for degradation, for example cytoplasmic nonsense-mediated decay in the case of mRNA (Losson and Lacroute, 1979; Maquat *et al*., 1981; reviewed in Kishor *et al*., 2019). Even for correctly executed transcription termination, the 'torpedo' models of RNAPI and RNAPII require nuclease activity first to define the 3´ end of the transcript (Kufel *et al*., 1999), then to simultaneously degrade the separate downstream sequence in a 5´-3´ direction and dislodge the polymerase (Connelly and Manley, 1988; Kim *et al*., 2004; West *et al*., 2004, 2008; El Hage *et al*., 2008; Kawauchi *et al*., 2008; Han *et al*., 2023). By contrast the termination mechanisms of RNAPIII can be induced by the transcription of four to six U nucleotides (Bogenhagen and Brown, 1981; Allison and Hall, 1985; Braglia *et al*., 2005). This mechanism is en-

hanced by adjacent stem loops (Nielsen *et al*., 2013; Xie *et al*., 2022), but these U nucleotides require removal nucleases to define the mature 3´ end (with the exception of the *SCR1* RNA, which is produced with a mature 3´ terminus; Felici *et al*., 1989).

The sum of a cell's ribonuclease activity is mediated by a large number of often overlapping specificities. Each nuclease demonstrates substrate recognition mechanisms that require the presence or absence of sequence motifs, secondary structures, or bound proteins. By acting in concert a cell's nucleases mediate a delicate balance between stabilisation and destabilisation. Certain transcripts require specific nucleases for maturation, and can be left with sequence features rendering them vulnerable to degradation in their absence; nucleases can protect or sensitise an RNA to other nucleases. In order to understand how a cell's nucleases jointly define the sequence and lifespan of RNA, we need an understanding of nuclease substrate recognition mechanisms.

1.2 Nucleases can be classified by catalytic mechanism.

Nucleases can be broadly distinguished by their catalytic mechanisms: endoribonucleases can cleave phosphodiester bonds found throughout the length of an RNA strand, while exoribonucleases cleave the terminal phosphodiester bonds of an RNA as outlined in Figure 1.1A. Exonuclease activity can be differentiated further by polarity; a given substrate recognition mechanism is only able to target either the 5´ or 3´ terminus of a nucleic acid, but certain conditions may enable an exonuclease to act as an endonuclease (for example Mre11; Cannavo and Cejka, 2014) and *vice versa* (for example T4 RNase H, Bhagwat *et al*., 1997; and Ape1, Chou and Cheng, 2002). Far fewer 5´ exoribonucleases than 3´ exoribonucleases are found in any given genome (Yang, 2011; Bechhofer and Deutscher, 2019), with only a single 5´ exonuclease identified in *E. coli* (Ghodge and Raushel, 2015; Jain, 2020). Another property of exonucleases is processivity, the number of individual nucleotides removed per binding event. Processive nucleases may be able to degrade a transcript fully with a single binding event, whereas distributive exonucleases will tend to remove one or only a few nucleotides before dissociating, as demonstrated in Figure 1.1B. Nucleases can be further classified by the nucleophile used: hydrolytic nucleases use H_2O, whereas phosphorolytic nucleases use inorganic phosphate (for review see Yang, 2011).

The classification of nucleases into superfamilies is challenging, with one heroic effort manually describing over 30 superfamilies, spanning the breadth of known nucleases as of 2011 (Yang, 2011). A more systematic classification can be seen by searching the InterPro database, which as of September 2023 describes 68 superfamilies that encode nuclease activity (Jones *et al*., 2014; Paysan-Lafosse *et al*., 2023. for full list see Appendix B).

The focus of this thesis centres on the DEDD superfamily of 3´ exonucleases, so named for four conserved acidic residues (Asp-Glu-Asp-Asp). The DEDD residues mediate the binding of two divalent cations required for phosphodiester backbone hydrolysis, as shown in Figure 1.2 (Beese and Steitz, 1991; Steitz and Steitz, 1993). This hydrolysis reaction also uses a conserved histidine or tyrosine residue as a general base for activation of the water nucleophile, which gives the further designation of DEDDh and DEDDy . This active site catalyses the breakage of phosphodiester bonds using a '2-ion' mechanism (Beese and Steitz, 1991; Steitz and Steitz, 1993), summarised in Figure 1.2 B. Two catalytic Mg^{2+} ions are coordinated by the DEDD residue acidic patch, the ions interacting with both the 3´ end of a nucleic

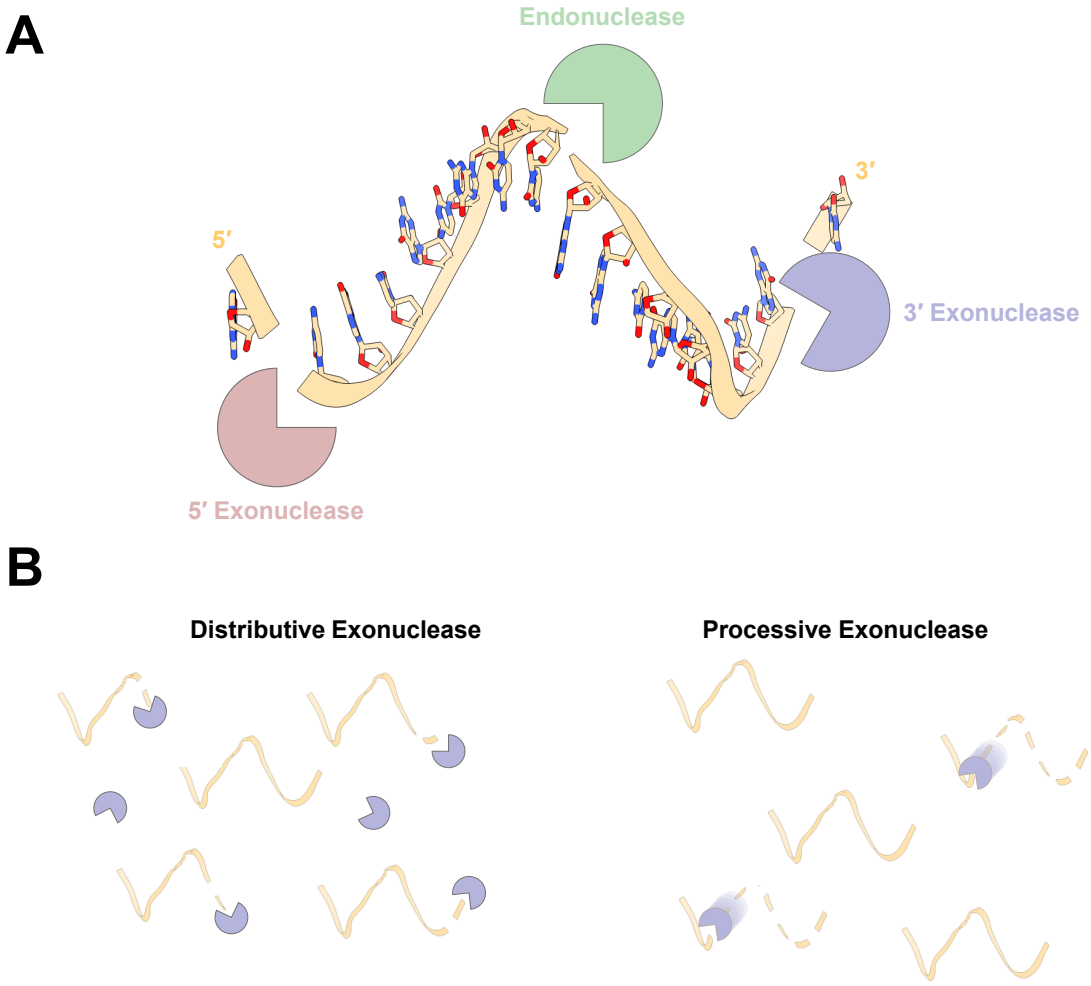


Figure 1.1: The properties of nuclease activity
A: Demonstration of different modes of nuclease activity, depicted against a short CG-rich RNA. RNA represented as a hybrid between stick models for ribose and base groups, with a cartoon ribbon for the phosphodiester backbone. Oxygen atoms are shown in red, nitrogen atoms are shown in blue, all other atoms are shown in beige. **B:** Visualisation of distributive and processive exonuclease activity. RNA is represented as beige ribbons. *RNA generated using UCSF ChimeraX 1.2.5 to represent PDB entry 1RC7, RNAse III complexed with nicked dsRNA* (Blaszczyk *et al.*, 2004).

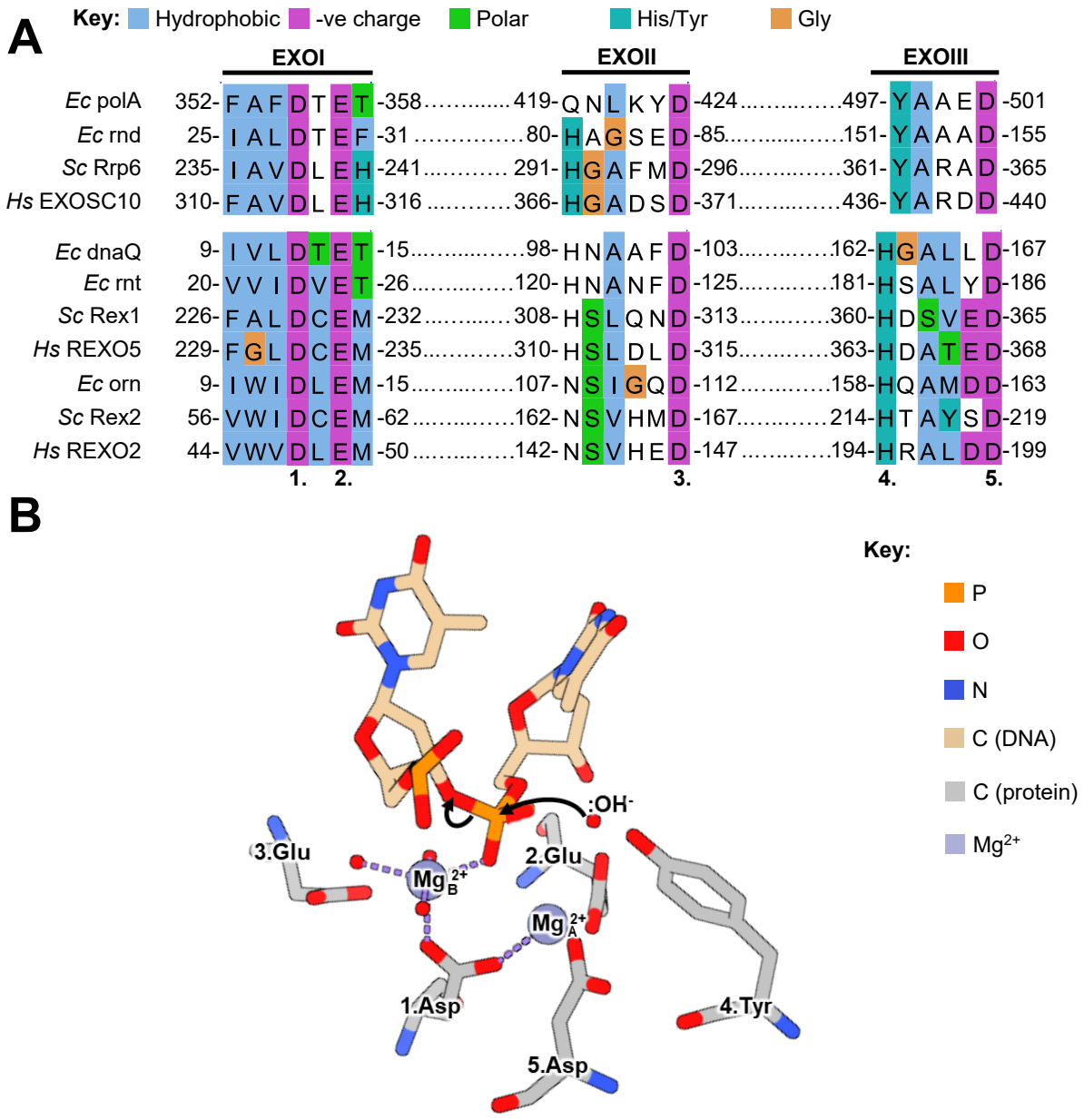


Figure 1.2: DEDD exonucleases possess conserved catalytic residues spread across three EXO motifs **A:** Multi-alignment of DEDD EXO motifs from *E. coli*, *S. cerevisiae*, and human, aligned using Clustal-Omega with default settings (Sievers *et al.*, 2011; Madeira *et al.*, 2022) and displayed in Jalview v2.11.2.7 (Waterhouse *et al.*, 2009). DEDDh (top) and DEDDy (bottom) alignments performed separately. Numbers '1.' - '5.' correspond to positions in catalytic scheme in part B. Amino acids are coloured based on the Clustal X colour scheme (shown in full in Figure A.1): all glycines are shown in orange, all other colours are shown when >60% of aligned residues share a residue group, with polar and His/Tyr residues also shown if the 60% threshold is met for aligned hydrophobic residues. **B:** The '2-ion' model of nucleic acid hydrolysis, demonstrated in an experimentally derived DEDDy active site. Annotations based on Steitz and Steitz, 1993, structural model of *E. coli* DNA polymerase I Klenow fragment (PDB: 1KRP; Hamdan *et al.*, 2002) displayed using ChimeraX v1.2 (Pettersen *et al.*, 2021). Majority of protein chain not shown, with main chain N- and C- termini shown instead as blue and red respectively. 'Ec' = *E. coli*, 'Sc' = *S. cerevisiae*, 'Hs' = *H. sapiens*.

acid and a OH^- nucleophile to stabilise a pentavalent phosphate intermediate.

The DEDD motif was first identified in the proofreading modules of *E. coli* DNA polymerases I (polA) and III (dnaQ), featuring a DEDDy and DEDDh motif respectively (Bernad *et al*., 1989). The DEDD superfamily was later coined in a bioinformatic analysis that aimed to classify the known exonucleases into five superfamilies (Zuo and Deutscher, 2001). The DEDD superfamily has since come to include the domains in enzymes involved with RNA maturation and turnover, but also domains in enzymes with roles as varied as DNA repair (TREX1, Lindahl *et al*., 1969; WRN, Huang *et al*., 1998), replication proofreading for the ssRNA genome of nidoviruses (for example SARS-CoV nsp14; Minskaia *et al*., 2006), and the acquisition of DNA fragments in certain bacterial CRISPR immune systems (recently identified in the *Megasphaera* trimmer integrase; Wang *et al*., 2023).

The DEDD family describes a conserved globular domain. Some DEDD exonucleases manage to achieve their specificity through internal features of the DEDD domain, however many achieve their specificity through the contribution of additional substrate-binding domains found within the protein, and/or contributed through the binding of partner proteins and complexes.

1.3 The mechanisms and domain arrangements of DEDD exonucleases are conserved to varying extents between *E. coli* and eukaryotes.

The bulk of our early understanding of RNA metabolism has been gained from fast growing and genetically pliable models such as *E. coli* and *S. cerevisiae* (for recent reviews, see Bechhofer and Deutscher, 2019 for prokaryotic nuclease activity; Phizicky and Hopper, 2023 for nuclease activity in eukaryotic tRNA metabolism; and Schneider and Bohnsack, 2023 for nuclease activity in eukaryotic rRNA metabolism). By applying the foundational knowledge of RNA metabolism gained in these organisms, it is possible to extrapolate similar mechanisms in humans.

Three of the eight 3´ exonucleases identified in *E. coli* belong to the DEDD superfamily (Zuo and Deutscher, 2001), which serve well to demonstrate the diversity of substrate preferences for 3´ exonucleases. Oligoribonuclease is uniquely adapted to break down 2-5 nucleotide end products from the degradation of larger RNA substrates (Stevens and Niyogi, 1967; Niyogi and Datta, 1975; Ghosh and Deutscher, 1999); RNase D is thought to trim the unstructured distal 3´ sequence of tRNA precursors (Ghosh and Deutscher, 1978; Zhang and Deutscher, 1988; Reuven and Deutscher, 1993); and RNase T is unique in the *E. coli* proteome for its ability to precisely trim short overhangs immediately adjacent to stem loop structures, such as those found in the final stages of tRNA maturation (Deutscher *et al*., 1984; Zuo and Deutscher, 2002b).

The *E. coli* DEDD exonucleases each have known functional homologues in *S. cerevisiae* and humans: Rex2/REXO2 for oligoribonuclease (Hanekamp and Thorsness, 1999; van Hoof *et al*., 2000a; Nguyen *et al*., 2000), Rrp6/EXOSC10 for RNase D (Ge *et al*., 1992; Briggs *et al*., 1998), and Rex1 (also known as RNA82 and RNH70) /REXO5 (also known as Nef-sp) for RNase T (Piper *et al*., 1983; Frank *et al*., 1999; van Hoof *et al*., 2000a; Gerstberger *et al*., 2017; Silva *et al*., 2017). The eukaryotic homologues of oligoribonuclease and RNase D, REXO2 and Rrp6 have received thorough mechanistic characterisation, while the absence of structural models for Rex1 and REXO5 have prevented the same level of understanding for functional homologues of RNase T, despite rigorous characterisation of RNase T itself.

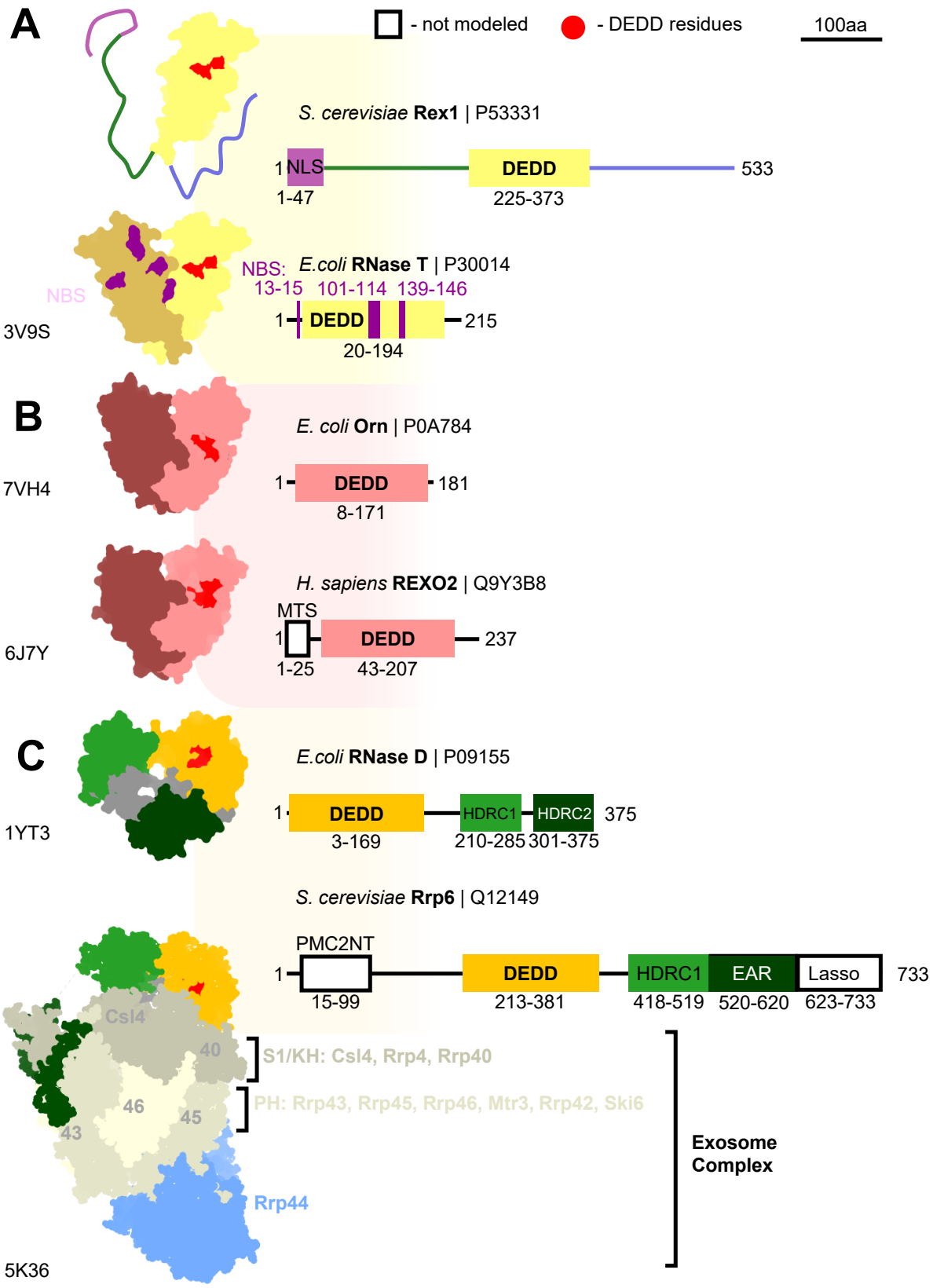

A
- not modeled
- DEDD residues
100aa
S. cerevisiae Rex1 | P53331
1
NLS
1-47
DEDD
225-373
533
NBS
3V9S
NBS:
E.coli RNase T | P30014
13-15
101-114
139-146
1
DEDD
20-194
215
B
E. coli Orn | P0A784
7VH4
1
DEDD
8-171
181
H. sapiens REXO2 | Q9Y3B8
MTS
6J7Y
1
DEDD
1-25
43-207
237
C
E.coli RNase D | P09155
1YT3
1
DEDD
3-169
HDRC1
210-285
HDRC2
301-375
375
S. cerevisiae Rrp6 | Q12149
PMC2NT
1
15-99
DEDD
213-381
HDRC1
418-519
EAR
520-620
Lasso
623-733
733
Csl4
40
46
45
43
Rrp44
S1/KH: Csl4, Rrp4, Rrp40
PH: Rrp43, Rrp45, Rrp46, Mtr3, Rrp42, Ski6
Exosome Complex
5K36

Figure 1.3: Conservation of domain structure for DEDD exonucleases
Comparison of domain arrangements for six DEDD exonucleases. Domain cartoons based on PDB entries indicated to the left with the DEDD active site highlighted in red. Domains not shown in cartoons are shown as white boxes with black outlines. Domain cartoons displayed using ChimeraX v1.2 (Pettersen *et al*., 2021). Linear domain models represented in scale with each other, based on UniprotKB entries of accession numbers indicated next to gene names. **A:** Comparison of Rex1 and RNase T. RNase T based on PDB entry 3V9S (Hsiao *et al*., 2012), represented as a homodimer with purple spots indicating the position of the nucleotide binding site (NBS; Zuo and Deutscher, 2002a). Rex1 represented by the isolated DEDD domain of RNase T from the same PDB entry. **B**: Comparison of *E. coli* oligoribonuclease (PDB 7VH4; Badhwar *et al*., 2022) with human REXO2 (6J7Y; Chu *et al*., 2019). **C**: Comparison of *E. coli* RNase D (PDB 1YT3; Zuo *et al*., 2005 with *S. cerevisiae* Rrp6 in complex with the exosome complex (5K36; Zinder *et al*., 2016). Exosome subunits Rrp40, Rrp43, Rrp45, and Rrp46 are abbreviated.

1.3.1 Oligoribonuclease and Rex2

Oligoribonuclease has received extensive structural study in *E. coli* (Badhwar *et al*., 2022; additional unpublished PDB entries 1YTA and 2IGI) and humans (Kim *et al*., 2019; Chu *et al*., 2019; Szewczyk *et al*., 2020), revealing a functional homodimer that is conserved between prokaryotes and eukaryotes. The subunits of this dimer consist of the DEDD exonuclease domain with little extra sequence besides a short N-terminal mitochondrial localisation sequence in eukaryotes. While oligoribonuclease and its homologues are best known for their unique ability to trim nucleotide end products that are 2-5 nucleotides in length, yeast Rex2 is also known to provide redundancy in the processing of stable RNAs including 5.8S rRNA and the U4 and U5 spliceosomal snRNAs, indicating a degree of activity on large, structured substrates (van Hoof *et al*., 2000a).

1.3.2 RNase D, Rrp6, and the exosome complex

In contrast, RNase D and Rrp6/EXOSC10 demonstrate significant divergence in structural organisation. Both RNase D and Rrp6/EXOSC10 feature a central DEDD exonuclease domain followed by a HRDC (Helicase and RNase D C-terminal; Morozov *et al*., 1997; Phillips and Butler, 2003) domain, but while RNase D functions as a monomer, Rrp6/EXOSC10 possesses an additional N-terminal PMC2NT domain that results in stoichiometric binding to Rrp47 (Stead *et al*., 2007; C1D in humans); this binding mutually increases stability of Rrp6 and Rrp47 (Feigenbutz et al., 2013a,b) and enables association of the Rrp6/Rrp47 heterodimer with the exosome complex (Peng *et al*., 2003).

The exosome complex is central to eukaryotic gene expression, coordinating RNA maturation, quality control, turnover, and transcription termination (for review see Lemay and Bachand 2015; Zinder and Lima 2017; Bresson and Tollervey 2018). RNase D serves as a comparatively peripheral source of 3´ exonuclease activity in *E. coli,* acting as a backup pathway for the exonucleolytic trimming of tRNAs and other small ncRNAs (Zhang and Deutscher, 1988; Reuven and Deutscher, 1993; Li *et al*., 1998).

The exosome is a sophisticated arrangement of nucleases and pseudonucleases (Liu *et al*., 2006; Dziembowski *et al*., 2006). The exosome core features a hexameric barrel resembling RNase PH, but with a unique protein for each subunit and a loss of phosphorolytic exonuclease activity. The barrel is capped with a trimer featuring folds known to be active in RNA binding (Liu *et al*., 2006). In yeast, these 9 subunits form a stable interaction with Rrp44 via the PIN endonuclease domain fused to a C-

terminal RNase II-like 3´ exonuclease (Schneider *et al*., 2007; Lebreton *et al*., 2008; Schaeffer *et al*., 2009; Schneider *et al*., 2009). Association of Rrp6 is observed in the nucleus and not observed in the cytoplasm (Allmang *et al*., 1999a). The human exosome complex features a greater degree of specialisation between compartments: the Rrp6 homologue EXOSC10 is the sole nuclease bound to the 9 subunit core in the nucleolus, both EXOSC10 and the Rrp44 homologue DIS3 are bound in the nucleoplasm, and a separate Rrp44 homologue DIS3-like (DIS3L) is the only nuclease bound to the core in the cytoplasm (Staals *et al*., 2010; Tomecki *et al*., 2010).

The activity of the yeast exosome complex is stimulated by additional cofactors. Mpp6 is a small RNA-binding protein that assists with the turnover of structured substrates through binding at the top of the barrel to the Rrp6/Rrp47 heterodimer (Schilders *et al*., 2005; Milligan *et al*., 2008; Feigenbutz *et al*., 2013a; Garland *et al*., 2013; Wasmuth *et al*., 2017). The TRAMP (Trf4/Air2/Mtr4 Polyadenylation) adaptor complex is a multi-protein complex featuring an untemplated polyA polymerase (either Trf4 or Trf5), a zinc knuckle RNA-binding protein (Air1 or Air2), and an ATP-dependent helicase (Mtr4; LaCava *et al*., 2005; Vanacova *et al*., 2005; Wyers *et al*., 2005). The polyadenylation of substrates and binding to the TRAMP complex seems to enhance the activity of Rrp6 several-fold (Callahan and Butler, 2010), and is capable of targeting substrates for exosome-mediated maturation or turnover.

1.3.3 RNase T and Rex1

RNase T has received rigorous biochemical (Deutscher and Marlor, 1985; Viswanathan *et al*., 1998; Zuo and Deutscher, 1999, 2002a,b,c) and structural (Zuo *et al*., 2007; Hsiao *et al*., 2011, 2012, 2014) characterisation, revealing a homodimeric structure wherein both subunits contribute to substrate specificity. In addition to processing the short stem-loop adjacent overhangs of stable RNAs (Deutscher *et al*., 1984), RNase T has been shown to play a role in the processing of DNA damage-associated ssDNA overhangs (Viswanathan *et al*., 1999; Hsiao *et al*., 2014). The RNase T open reading frame includes minimal sequence aside from the DEDD exonuclease domain, with a positively charged Nucleotide Binding Site (NBS) found towards the C-terminal end (Zuo and Deutscher, 2002c), which is responsible for accommodating the non-scissile strand of the RNA duplex in the active site (Zuo *et al*., 2007; Hsiao *et al*., 2011, 2012). An additional layer of substrate specificity is imposed on top of the binding of dsRNA, the residues E73 and 5 phenylalanines are able to recognise a 3´ terminal CC sequence, and inhibit nuclease activity through mis-positioning of the DEDDh residues to mediate a so-called 'C-effect' that serves to prevent over-trimming of tRNA (Zuo and Deutscher, 2002b; Hsiao *et al*., 2011).

This domain arrangement is substantially different in the yeast functional homologue of RNase T, Rex1. Rex1 is known to behave as a monomer (Frank *et al*., 1999; Hama Soor, 2017) and possesses extra N- and C-terminal sequence flanking the DEDD exonuclease domain including a 50aa Nuclear Localisation Sequence (NLS) at the N-terminus (Frank *et al*., 1999; Hama Soor, 2017; Daniels *et al*., 2022).

Although DEDD exonucleases demonstrate unique substrate recognition mechanisms, a degree of redundancy is shared in the processing of RNAs (for review, see Bechhofer and Deutscher, 2019; Phizicky and Hopper, 2023; Schneider and Bohnsack, 2023). Through considering common features between substrates that are shared between and unique to certain exonucleases, it is possible to identify the substrate features that are required for an exonuclease to recognise its substrate.

1.4 Rex1 uses RNase T-like activity for the accurate production of stable RNAs in yeast, primarily RNAPIII transcripts.

Rex1 and RNase T process similar sets of substrates in their respective organisms, as outlined in Figure 1.4. In *S. cerevisiae* and *E. coli,* Rex1 and RNase T are the sole nucleases with the ability to trim the 3 nucleotide 3´ overhang adjacent to the terminal stemloop of 5S rRNA, but other substrates demonstrate redundant processing by alternative nucleases as outline in Figure 1.4. The best characterised substrates of Rex1 are transcribed by RNAPIII, although Rex1 also plays a role in the maturation of 25S rRNA from RNAPI transcripts.

The most abundant RNAPIII transcripts are 5S rRNA and tRNAs, the genes coding for which feature Type I and Type II RNAPIII promoters respectively. These promoters require specific sequence elements embedded downstream of the initiation site in the transcribed sequence itself, termed A-boxes and B-boxes. Other transcripts such as U6 snRNA (Das *et al*., 1988), the RNase P RNA component Rpr1 (Lee *et al*., 1991), and the SRP importin RNA component scR1 (Felici *et al*., 1989) use Type III RNAPIII promoters that do not feature promoter sequences downstream of the initiation site, with the A-boxes and B-boxes instead occurring upstream (for review see Acker *et al*., 2013).

The termination of RNAPIII is induced by the transcription of a poly dA template to create a nascent run of rUs, resulting from the innate instability of dA:rU base pairing (Martin and Tinoco, 1980). Human RNAPIII efficiently terminates upon transcription of four Us, but *S. cerevisiae* seems to require five to six (Braglia *et al*., 2005; Iben and Maraia, 2012). Transcriptome sequencing studies have revealed that read-through of canonical RNAPIII termination sites is common (Turowski *et al*., 2016), in which case a backup termination site may be used, for example as mediated by the NNS (Nrd1-Nab3-Sen1) complex in yeast (Xie *et al*., 2022). RNA stemloop formation upstream of a polyU can increase termination efficiency and promote termination at weaker polyU sites (Nielsen *et al*., 2013; Xie *et al*., 2022). Upon release from the DNA template, the nascent RNA is removed from the RNAPIII using internal proteins of the complex including Rpc11. Rpc11 stimulates the intrinsic nuclease activity of the RNAPIII active site in an analogous manner to TFIIS in RNAPII transcription (Chédin *et al*., 1998, for review of internal RNAPIII termination mechanisms see Arimbasseri *et al*., 2013). The 3´ end of the released nascent transcript is bound and protected by the Lupus antigen Homology Protein (Lhp1) and LSm complexes, which both serve to enhance termination, protect the 3´ ends of nascent RNAs from exonuclease activity, and promote correct folding (Mayes *et al*., 1999; Kufel *et al*., 2002). The mechanistic preference of canonical RNAPIII termination for a stemloop followed by a polyU site (Xie *et al*., 2022) seems to produce 3´ ends that are natural substrates for Rex1 by necessity.

1.4.1 RNase T and Rex1 target the 3´ ends of tRNA.

In eukaryotes, tRNAs are transcribed by RNAPIII at hundreds of loci: 275 tRNA genes are found in *S. cerevisiae*, and 466 are found in humans (Goodenbour and Pan, 2006). Mature tRNAs contain between 75-95 nucleotides, and feature variation far beyond that seen in the acceptor stem anticodons that are used to decode mRNA (Phizicky and Hopper, 2023). In yeast, 61 of the 275 tRNA genes contain introns that require excision by tRNA splicing endonuclease and tRNA ligase (for recent review see Hayne *et al*., 2023). These introns serve to guide modification of nucleotides, aid recognition of the type II RNAPIII promoter, and potentially serve as insulating elements for the spread of chromatin (Donze *et al*., 1999;

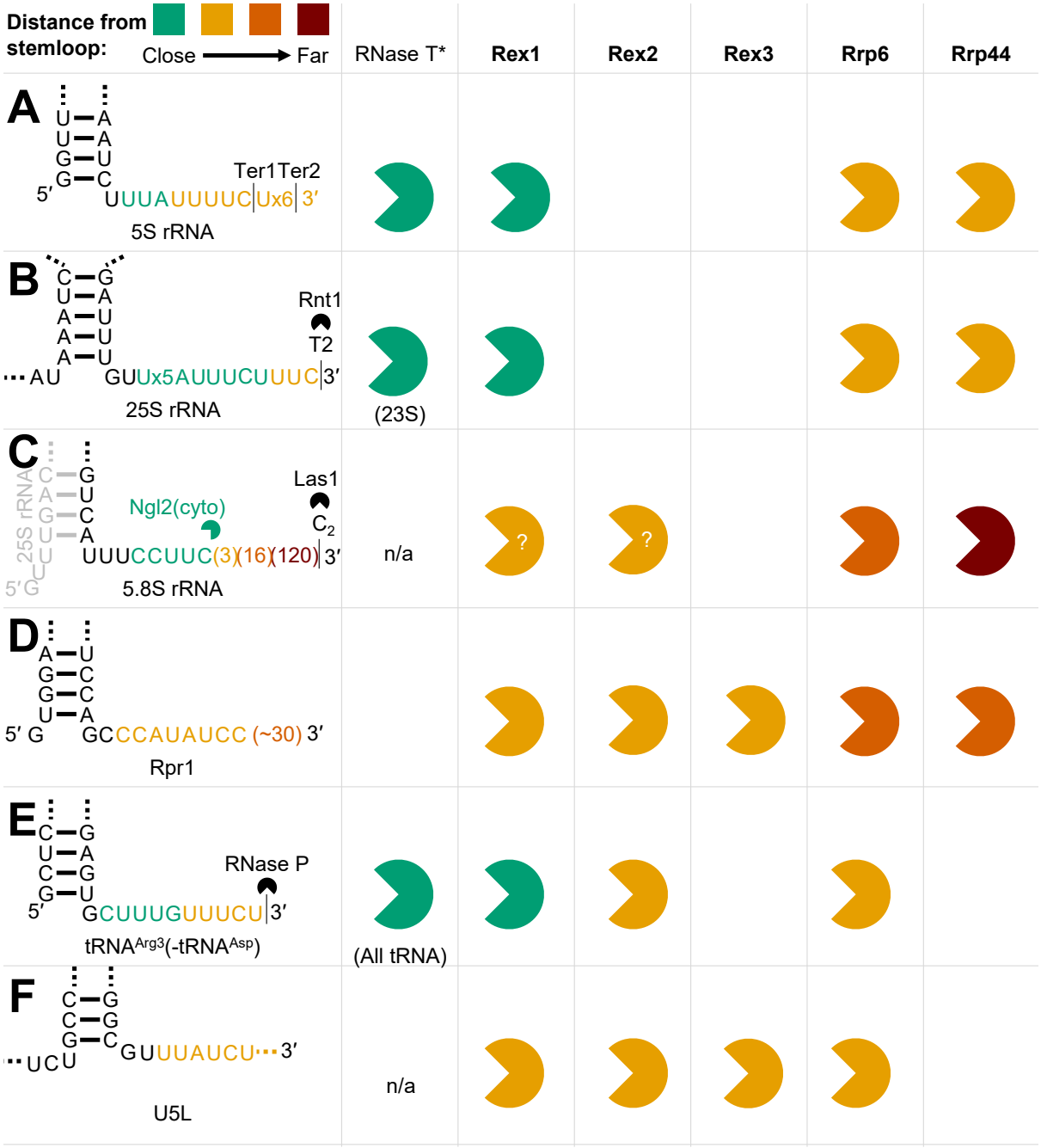


Figure 1.4: Rex1 shares substrates with other yeast exonucleases, and with *E. coli* RNase T
Summary of overlapping substrate processing roles for RNase T, Rex1-3, and the exosomal nucleases Rrp6 and Rrp44. For each substrate, a heavily simplified representation is shown of the final stemloop of the mature species. The nucleotides of the mature sequence are shown in black, with trimmed precursor sequences shown in colours that correspond to the nucleases that process them, assigned according to the scale (top left). RNase T-like activity is represented in green, with broader exonuclease activity is shown in yellow-red. For simplicity no distinction is made between Rrp44 exonuclease or endonuclease activity, and the multiple threading routes of the exosome complex are not shown. Each substrate processing event is based on published models indicated as follows: **A**, Based on Piper *et al.*, 1983, the two different termination sites are indicated; **B**, Based on Abou Elela *et al.*, 1996; Kufel *et al.*, 1999; **C**, Based on Thomson and Tollervey, 2010; Faber *et al.*, 2002; Leshin *et al.*, 2011, green nucleotides show the cytoplasmic action of Ngl2; **D**, Based on Lan *et al.*, 2018; **E**, Based on Engelke *et al.*, 1985; **F**, Based on Nguyen *et al.*, 2015. *RNase T is included in cases where the equivalent processing event takes place in *E. coli*, in which case the corresponding species is stated if named differently. RNA species not included in *E. coli* are marked with 'n/a'.

Hayashi *et al*., 2019). Prokaryotic tRNA genes by comparison tend to be self-splicing if present (Xu *et al*., 1990; Reinhold-Hurek and Shub, 1992; Ferat and Michel, 1993), although no tRNA introns are found in *E. coli*. In both *S. cerevisiae* and *E. coli*, tRNAs are produced as precursor transcripts featuring 5´ and 3´ extensions that require removal by nucleases.

1.4.1.1 The maturation of tRNA in *E. coli* and *S. cerevisiae*

In both *E. coli* and *S. cerevisiae*, the mature 5´ end of tRNAs is generated by the endonuclease ribozyme RNase P (Guerrier-Takada *et al*., 1983; Lee *et al*., 1991). There are however differences in the 3´ maturation of tRNA between *E. coli* and eukaryotes: the initial 3´ CCA of all *E. coli* tRNAs is templated by the genome, meaning RNase T is required to mediate trimming up until the CCA (Reuven and Deutscher, 1993; Zuo and Deutscher, 2002b); in eukaryotes the 3´ CCA is added in a non-templated manner after removal of the 3´ precursor sequence, requiring trimming until the discriminator base prior to untemplated CCA addition (Marck and Grosjean, 2002).

A further difference is that a subset of eukaryotic tRNAs possess introns (Hopper *et al*., 1978) that are spliced out using the tRNA splicing endonuclease complex and tRNA ligase; tRNA splicing takes place independently of trimming the 5´ and 3´ termini, and in *S. cerevisiae* requires export of the tRNA from the nucleus to access the tRNA splicing endonuclease complex (TSEN) and tRNA ligase that are anchored to the surface of the mitochondria (Yoshihisa *et al*., 2003).

E. coli and *S. cerevisiae* differ in the ordering of 5´ and 3´ trimming events. In *E. coli,* 5´ cleavage by RNase P only precedes 3´ trimming in cases where the 5´ extension of a pre-tRNA base pairs to the 3´ extension (Li and Deutscher, 2002); 3´ processing only precedes RNase P cleavage in cases where the 3´ extensions are particularly long (Li and Deutscher, 1994), with the order of activity otherwise proceeding stochastically (Li and Deutscher, 1996, 2002). In *S. cerevisiae*, RNase P cleavage of the 5´ precursor extension is required before 3´ trimming by default (Lee *et al*., 1991; O'Connor and Peebles, 1991; Lygerou *et al*., 1994; Engelke *et al*., 1985), with the reverse occurring in the case of $tRNA^{Trp}$ (Kufel and Tollervey, 2003) and for other tRNAs in the absence of Lhp1 (Yoo and Wolin, 1997).

While the bulk of tRNA 3´ trimming in *E. coli* is exonucleolytic, in eukaryotes it seems that 3´ trimming is mediated by both the endonuclease tRNase Z (a homologue of RNase BN) and Rex1 with a minor contribution from Rrp6 and Rex2 (Skowronek *et al*., 2014). Distribution of substrates between endonucleolytic and exonucleolytic maturation seems to be based in part on the length and composition of the trailer sequence (Skowronek *et al*., 2014).

Short trailers (1-10 nucleotides) can be processed by both pathways, but the binding of Lhp1 to a 3´ terminal trailer of at least 3 Us protects the tRNA end from exonucleases, likely acting as a chaperone to facilitate folding, and actively stimulating the endonucleolytic pathway (Yoo and Wolin, 1997; Copela *et al*., 2008). Longer trailers (>25 nucleotides) featuring secondary structures are proposed to be poor substrates for exonucleolytic trimming, instead requiring the activity of tRNase Z (Skowronek *et al*., 2014).

Conversely, the Arg3-Asp tRNA gene is a case in yeast where only Rex1 can produce a mature 3´ end; two tRNAs are expressed as exons in the dicistronic $tRNA^{Arg3}$-$tRNA^{Asp}$ transcript, with the initial 3´ end of the 5´ $tRNA^{Arg3}$ precursor generated by RNAse P endonuclease activity instead of terminating RNAPIII (Engelke *et al*., 1985; Piper and Stråby, 1989; van Hoof *et al*., 2000a); this resembles the endonucleolytic release of tRNA from polycistronic precursors in *E. coli* (Li and Deutscher, 2002;

Mohanty and Kushner, 2007; Agrawal *et al*., 2014), albeit with the additional involvement of splicing. Maturation of the $tRNA^{Arg3}$-$tRNA^{Asp}$ transcript is further complicated by the fact that the 5´ precursor extension base pairs to 5 of the 10 nucleotide 3´ extension, forming an extended acceptor stem precursor (Engelke *et al*., 1985). Until the 5´ precursor is removed, the 5 base paired 3´ extension nucleotides are protected from degradation, and the remaining 3´ overhang of this precursor duplex require RNase T-like activity for their removal. A schematic of this maturation pathway is shown in Figure 1.5.

1.4.1.2 Quality control and turnover of tRNA

Exonuclease activity also plays an essential role in the quality control and turnover of tRNAs (for review, see Hopper, 2013). RNase T was initially isolated based on its unique ability to turn over the CCA terminal adenine at the end of mature tRNAs (Deutscher *et al*., 1984). This trimming role underpins the futile cycle of CCA removal by RNase T and addition by tRNA terminal nucleotide transferase at the end of mature tRNAs in bacteria, which serves to couple quality control of a tRNA to its aminoacylation status (Dupasquier *et al*., 2008). RNase T was later observed to have a role in the exonucleolytic maturation of the 3´ precursor extension of *E. coli* tRNAs with joint contribution from RNases PH, and with minor contributions from RNases D, II, and BN (Reuven and Deutscher, 1993; Li and Deutscher, 1996).

These roles for RNase T resemble the roles proposed for Rex1 in *S. cerevisiae* tRNA metabolism. It has been suggested that Rex1 may be the exonuclease responsible for trimming the 3´ terminal adenine for an equivalent futile CCA cycle in yeast (Copela *et al*., 2008), and that the RNase A-like endonuclease angiogenin may fill this role during oxidative stress in humans (Czech *et al*., 2013; for review of bacterial and eukaryotic CCA cycling see Wellner *et al*., 2018).

In *S. cerevisiae*, turnover takes place through three pathways: 3´ degradation by Rrp44 as part of the exosome complex, promoted by tRNA polyadenylation by the TRAMP complex (Kadaba *et al*., 2004, 2006; Schneider *et al*., 2007); 5´ degradation by Rat1 and Xrn1, 'Rapid tRNA Decay' (RTD), directed by the instability of tRNA folds that lack the correct modifications (Alexandrov *et al*., 2006; Whipple *et al*., 2011); and jointly-induced 5´ and 3´ degradation directed against tRNAs that have been marked with a 3´ CCACC(A) sequence, although this addition of a second CCA motif is only seen in tRNAs featuring a 5´GG (Wilusz *et al*., 2011).

The levels of attrition directed against tRNAs in *S. cerevisiae* is high, with 50% of transcribed tRNAs in yeast undergoing degradation by the exosome before reaching maturation (Gudipati *et al*., 2012). This TRAMP-mediated pathway seems to be inhibited by timely Lhp1 binding or Rex1 processing (Copela *et al*., 2008; Ozanick *et al*., 2009), meaning the down-regulation of Rex1 activity by temperature or nutrient stress results in an increased direction of tRNA towards Lhp1-mediated end protection or TRAMP/exosome-mediated turnover (Foretek *et al*., 2016). The race between maturation and turnover is emblematic of RNA quality control and represents a common theme throughout the production of stable RNAs (Porrua and Libri, 2013).

1.4.2 Rex1 and RNase T are required for the full maturation of 5S rRNA.

The transcription of 5S rRNA differs from the transcription of other rRNAs (for review see Ciganda and Williams, 2011): 5S is the only rRNA not to be transcribed by RNAPI as part of the rDNA repeat, instead being transcribed by RNAPIII from a type I promoter; 5S genes are found in-between

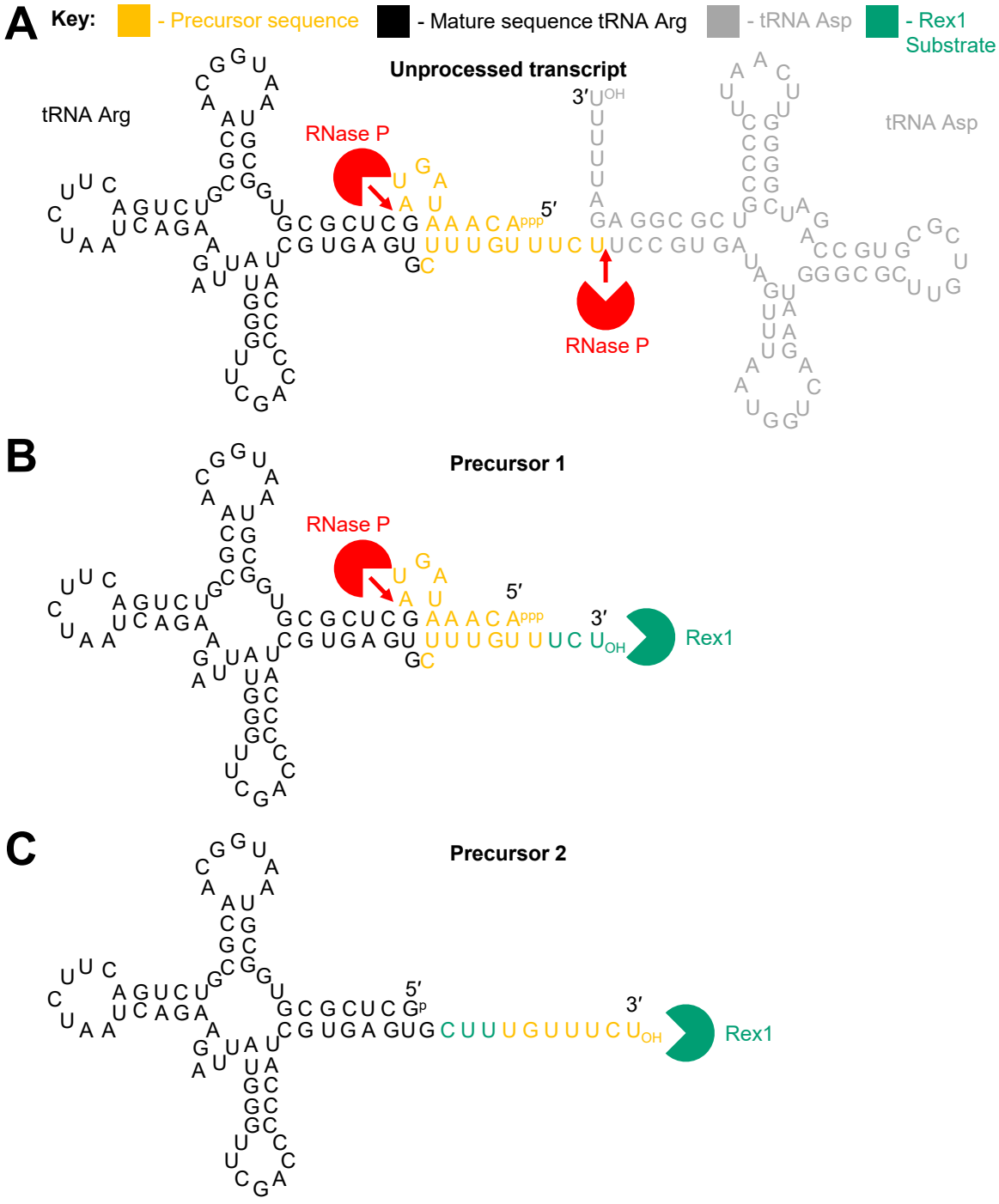

A
Key:
- Precursor sequence
- Mature sequence tRNA Arg
- tRNA Asp
- Rex1 Substrate
Unprocessed transcript
tRNA Arg
tRNA Asp
RNase P
RNase P
5′
3′
B
Precursor 1
RNase P
5′
3′
Rex1
C
Precursor 2
5′
3′
Rex1

Figure 1.5: Maturation of *S. cerevisiae* $tRNA^{Arg3}$-$tRNA^{Asp}$ requires multiple nucleolytic processing events
Schematic of cleavage events required for the maturation of $tRNA^{Arg3}$-$tRNA^{Asp}$ in *S. cerevisiae* based on secondary structure map shown in Engelke *et al*., 1985. Two RNase P cleavage sites result in multiple precursors of $tRNA^{Arg3}$ that serve as substrates for Rex1; nucleotides that require Rex1 for removal are indicated in green. The order of RNase P cleavage events is not indicated, and the discistronic precursor that would be produced from cleavage at the 5´ site is not shown. **A**: The raw RNAPIII transcription product of discistronic $tRNA^{Arg3}$-$tRNA^{Asp}$. Cleavage at the 3´ RNase P site gives 'Precursor 1'. **B**: Precursor 1, which features an extended acceptor stem. Base paired nucleotides are inaccessible to maturation by 3´ exonucleases, with the 3 nucleotides accessible to RNase T-like trimming by Rex1. Cleavage of this precursor at the 5´ RNase P cleavage site gives 'Precursor 2'. **C**: Precursor 2, which may lack 1-3 3´ nucleotides if processed by Rex1 prior to cleavage at the 5´ RNase P site. Nucleotides indicated in orange may be trimmed by 3´ exonuclease of other DEDD exonucleases such as Rrp6.

rDNA repeats in the yeast nucleolus (Philippsen *et al*., 1978) or nuclear tandem repeats in most other eukaryotes (Douet and Tourmente, 2007); and 5S rRNA doesn't seem to require any nucleotide modification in most eukaryotes, although a pseudouridylation at position 50 has been observed in *S. cerevisiae* (Miyazaki, 1974). Transcription at a separate locus can lead to a stoichiometric discrepancy with the jointly-transcribed 18S-5.8S-25S species, which is leveraged to couple the dysregulation of ribosome biogenesis to induction of apoptosis by p53 in higher eukaryotes (for recent review see Lindström *et al*., 2022).

1.4.2.1 The maturation of 5S rRNA in *E. coli* and *S. cerevisiae*

The eukaryotic 5S rRNA precursor is transcribed without additional 5´ sequence (Dieci *et al*., 2013), explaining reports in *S. cerevisiae,* mice, and humans of 5´ tri- and di- phosphorylated 5S species persisting from the initiating ribonucleotide tri-/di- phosphates (Soave *et al*., 1973; Mori and Ichiyanagi, 2021). *E. coli* by comparison requires the removal of 5´ extensions by the activity of one or more nucleases (Feunteun *et al*., 1972) including the recently identified 5´ exonuclease RNase AM (Jain, 2020), leaving a 5´ monophosphate (Soave *et al*., 1973).

Eukaryotic and prokaryotic 5S demonstrate more similarity in the processing of 3´ precursors of their extensions, both requiring the action of 3´ exonucleases, Rex1 and RNase T respectively for the removal of the final 3nts (Piper *et al*., 1983; Li and Deutscher, 1995). The upstream processing of these 3´ precursor sequences have different origins: in eukaryotes, the distal sequence of the raw 5S transcription product is removed by exonucleases, most likely the exosome (Wlotzka *et al*., 2011), with the final 3 nucleotides removed by Rex1 (Piper *et al*., 1983; van Hoof *et al*., 2000a); in prokaryotes, the 3´ end of the 5S transcript is defined by the endonuclease RNase E, with distal nucleotides potentially trimmed with varying contributions from RNases II, D, PH, and BN, with only RNase T able to remove the final 2nts (Li and Deutscher, 1995). This extension of 5S by 2-3 nucleotides is characteristic for loss of the exonuclease activity of Rex1 and RNase T, indeed this defect was how the *rna82-1* yeast mutant was initially identified (Piper *et al*., 1983), which would later be revealed by sequencing to be a premature termination mutant of Rex1 (van Hoof *et al*., 2000a).

The extended 5S rRNA is still readily incorporated into the ribosomes of both *E. coli* and *S. cerevisiae* with seemingly no detriment to ribosome function, indeed *E. coli* 5S precursors seem to be more effectively targeted by RNase T post-integration into the large subunit, likely stemming from the need for

trimming of the unpaired 5´ overhang by RNase AM before the full 3´ precursor overhang can be accessed (Jain, 2020).

1.4.2.2 Quality control and turnover of 5S rRNA

In eukaryotes, the maturation of 5S seems to be limited to 3´ trimming. In light of this, there are limited scenarios for which 5S would be targeted for degradation. Premature termination of RNAPIII can give rise to truncated 5S products, which seem to be degraded by the exosome (Wlotzka *et al*., 2011; Schneider, Kudla, Wlotzka, Tuck and Tollervey, 2012; Han and van Hoof, 2016; Delan-Forino *et al*., 2017). Over trimming of 5S by Rex1 and the exosome may give rise to truncation products, with a discrete truncation product seen in mutants of *TRF4, RRP44* (specifically the exonuclease module; Schneider, Kudla, Wlotzka, Tuck and Tollervey, 2012), *RRP6* (Kadaba *et al*., 2006; Schneider, Kudla, Wlotzka, Tuck and Tollervey, 2012; Han and van Hoof, 2016; Delan-Forino *et al*., 2017) and *RRP47* (Garland *et al*., 2013; Feigenbutz *et al*., 2013a), which is abolished with the co-deletion of *REX1* (Copela *et al*., 2008; Garland *et al*., 2013). As the activity of Rex1 should be stopped by the terminal stemloop found in correctly-folded 5S rRNA, one explanation for this observation could be that Rex1 activity is able to truncate and target for degradation 5S precursors that have failed to fold correctly. Conversely, it may be that rapid Rex1 activity directed against 5S that isn't protected by Lhp1 may lead to the erroneous degradation of 5S that hasn't been allowed sufficient time to fold correctly, as proposed for Rex1 activity directed against scR1 (Copela *et al*., 2008).

1.4.3 Rex1 activity is detrimental to accurate maturation of the signal recognition particle RNA component scR1.

RNAPIII transcription termination by necessity produces transcripts ending with at least 4 U nucleotides. In many cases this polyU tail is removed to form a mature product, however scR1 represents an exception; the 6xU tail of the transcript forms a 3´ terminal stemloop in the mature product (Felici *et al*., 1989). For scR1, a joint protective effect against Rex1 seems to be mediated by Lhp1 and polyadenylation by the TRAMP complex (Leung *et al*., 2014), with a TRAMP-destablised truncation product observed in *tfr4*Δ yeast that isn't seen in *tfr4*Δ*/rex1*Δ mutants (Copela *et al*., 2008). A sequencing study of scR1 revealed a dependence on the TRAMP complex and exonuclease activity of the exosome for the correct maturation of the scR1 3´ terminus, and suggests that Rex1 may not be the only exonuclease responsible for over-trimming of the terminal 6xU in the absence of Lhp1 (Leung *et al*., 2014).

1.4.4 Maturation of the RNase P RNA component requires the activity of Rex1, Rex2, or Rex3.

The maturation of *RPR1*, the RNA component of RNase P in *S. cerevisiae,* is unusual amongst eukaryotes: most eukaryotes express *RPR1* as a mature transcript requiring no further end processing, however maturation of an extended precursor has been shown in *S. cerevisiae* to accumulate with the joint deletion of *REX1*, *REX2*, and *REX3* (van Hoof *et al*., 2000a). This suggests that removal of the 30nt of sequence between the mature 3´ end and the canonical termination site (Lee *et al*., 1991) can be mediated by any of Rex1, 2 or 3, with inefficient processing mediated by other nucleases, potentially the

exosome complex (Delan-Forino *et al*., 2017), giving rise to a heterogeneous 3´ extended end product in addition to the mature species (van Hoof *et al*., 2000a). Interestingly, the initial characterisation of the *RPR1* gene revealed a faint 28-30nt extended precursor (Lee *et al*., 1991) in the *rna82-1* Rex1 loss of function strain (Piper *et al*., 1983).

1.4.5 Non-RNAPIII Rex1 targets

1.4.5.1 RNAPI: Rex1 is involved in 3´ maturation of 25S rRNA and 5.8S rRNA, but not 18S rRNA.

Rex1 has also been observed to process transcripts from RNAPI. 18S rRNA is the only rRNA that seems to lack involvement of Rex1 in the generation of its 3´ end, as the mature 3´ end is produced in a single step through endonucleolytic cleavage by Nob1 (Fatica *et al*., 2003, 2004).

The *rna82-1* mutant corresponding to a premature termination mutant of Rex1 was observed to accumulate extended 25S rRNA precursors (Kempers-Veenstra *et al*., 1986). This observation was later contextualised as acting downstream of the rnt1 endonucleolytic cleavage site T2, requiring the trimming of 14nt (Abou Elela *et al*., 1996; Kufel *et al*., 1999). The final maturation step of 25S rRNA in yeast is proposed to be the removal of a non-base-paired 5´ overhang (Geerlings *et al*., 2000), recently discovered through serendipity to be mediated by the cytoplasmic 5´ exonuclease Dxo1 (Hurtig and van Hoof, 2022). This mature 5´ end is found upstream of a stemloop formed with the 3´ end of 5.8S rRNA, forming an unpaired 5´-UUU and UUG-3´ respectively (Leshin *et al*., 2011; see Figure 1.4C). As the final maturation step of 5.8S rRNA is also performed by the cytoplasmic 3´ exonuclease Ngl2 (Thomson and Tollervey, 2010), it is interesting to note how both exonucleases are unable to trim any closer to this stemloop than 3nt.

Rex1 was initially thought to share a redundant processing role with Rex2 in the 3´ trimming of 5.8S rRNA with a minor contribution from Rex3 (van Hoof *et al*., 2000a), although recent studies suggest this involvement may be limited, with other nucleases responsible for the bulk of 3´ 5.8S rRNA processing (Thomson and Tollervey, 2010; Hurtig and van Hoof, 2022; reviewed in Schneider and Bohnsack, 2023): the endonuclease Las1 generates the initial 5.8S 3´ end through cleavage of the rDNA transcription product at site C_2 (Castle *et al*., 2010; Schillewaert *et al*., 2012; Castle *et al*., 2013), the exosome trims the bulk of the precursor sequence through the successive activities of Rrp44 (7S to 5.8S+30; Mitchell *et al*., 1997) and Rrp6 (5.8S+30 to 6S; Briggs *et al*., 1998), at this point Rex1/2 has been suggested, but not proven, to trim nucleotides between the 6S species and the 5.8S+6 species (van Hoof *et al*., 2000a; Faber *et al*., 2002; Thomson and Tollervey, 2010), with the final 6 nucleotides trimmed by the cytoplasmic exonuclease Ngl2 (Faber *et al*., 2002; Thomson and Tollervey, 2010). A simplified schematic of this processing is shown in Figure 1.4C.

1.4.5.2 RNAPII: Rex1-3 are required for U5L maturation, and the loss of *REX1* exacerbates snoRNA processing defects in *rrp6*Δ yeast.

In addition to transcribing mRNA, *S. cerevisiae* RNAPII also transcribes a wide range of ncRNAs including all snoRNA (with the exception of snR52) and all spliceosomal RNAs (with the exception of U6; Brow and Guthrie, 1990). The activity of snoRNA-containing RNPs includes accurate modification of ribosomes, primarily 2´-*O*- methylation (C/D box snoRNA) or pseudouridylation (H/ACA box

snoRNA). In yeast, the 3´ ends of snoRNAs arise from NNS-dependent termination at one or more site if the gene is not contained within an intron, followed by TRAMP-guided trimming by Rrp6 and the exosome (Allmang *et al*., 1999b; van Hoof *et al*., 2000b). Although the deletion of *REX1* individually doesn't lead to snoRNA processing defects (Garland *et al*., 2013; Feigenbutz *et al*., 2013a), combining this mutant with an inducible *RRP6* loss of function mutant led to accumulation of extended snR13, snR38, snR50, and U14 with a depletion of mature species (Garland *et al*., 2013). It is suggested that widespread use of polyadenylation for backup snoRNA processing underpins the synthetic lethality observed with the loss of both proteins, resulting in titration of RNA processing factors away from essential mechanisms of RNA metabolism (Garland *et al*., 2013; Feigenbutz *et al*., 2013a).

Of the spliceosomal RNAs, Rex1 has been implicated jointly with Rex2 and Rex3 in the formation of U5L snRNA, an isoform of U5 snRNA (van Hoof *et al*., 2000a). The mature U5L snRNA possesses a 3´ terminal stemloop, whereas the other isoform U5S features a short single-stranded 3´ extension (Patterson and Guthrie, 1987; Chanfreau *et al*., 1997). Joint deletions of *REX1*, *REX2*, and *REX3* demonstrate a dramatic depletion of U5L with a mild extension phenotype visible for U5S, leading to the suggestion that Rex1-3 are primarily responsible for maturation of U5 (van Hoof *et al*., 2000a), with less efficient processing mediated by the exosome in their absence (Allmang et al., 1999b), and with both exonucleolytic pathways acting downstream of endonucleolytic cleavage by Rnt1 (Chanfreau *et al*., 1997). It is unclear to what extent Rex1-3 impact the exonucleolytic processing of the other spliceosomal RNAs, although there is clear involvement of the exosome (for review, see Peart *et al*., 2013).

1.4.6 Do Rex1 substrates have common features?

By considering which substrates are uniquely processed by Rex1, and which substrates are shared between Rex1 and other exonucleases, a trend seems to emerge. Where Rex1 is uniquely able to process substrates such as 5S rRNA and dicistronic $tRNA^{Arg3}$, these are cases where the 5´ end of the substrate RNAs are found base-paired in close proximity to the 3´ end. For substrates where Rex1 is one of several nucleases able to perform maturation, for example in the U5L snRNA and 5.8S rRNA, there is a terminal stemloop, but with extensive 5´ sequence. One simple model that may explain the differing efficiencies of Rex1 on different substrates could be that a base-paired 5´ terminal nucleotide is required for efficient processing, with stemloops featuring both 5´ and 3´ overhangs presenting similar, but sub-optimal substrates that are equally well processed by other exonucleases.

At the boundary between these two classifications is *RPR1*, the RNA component of RNAse P. In a high resolution electron cryo-microscopy map of the mature RNase P complex (Lan *et al*., 2018), a single unpaired G is seen at the 5´ end that is unable to hydrogen bond with the 3´ terminal GC overhang. If it is the case that a base paired 5´ nucleotide is a hallmark of efficient Rex1 processing, then it may be the case that a single unpaired 5´ nucleotide is sufficient to render Rex1's efficiency on the substrate as comparable to other exonucleases such as Rex2 or Rex3. This model is summarised in Figure 1.6.

This model for substrate specificity has been proposed for RNase T (Li *et al*., 1998, 1999), where it is suggested that a base-paired 5´ terminal nucleotide is required for efficient trimming by RNase T. An informative comparison can be drawn between the processing of 25S rRNA by Rex1 in *S. cerevisiae* and the equivalent processing of 23S rRNA by RNase T in *E. coli*: the 3´ terminus of 25S in *S. cerevisiae* is found adjacent to a stemloop that ends at the 5´ end with unpaired nucleotides, leaving a 3´ GU overhang

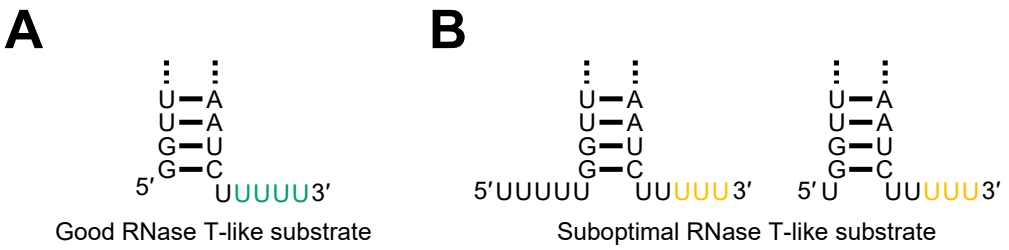


Figure 1.6: Substrate binding patterns suggest stemloop context influences optimal processing by Rex1 or RNase T
Model of Rex1 and RNase T substrate preferences based on the common features of Rex1/RNase T-unique substrates. Arbitrary nucleotides shown, with optimally processed nucleotides shown in green (**A**), with suboptimally processed nucleotides shown in yellow (**B**).

that can't be efficiently processed by Rex1 (Leshin *et al*., 2011); the 3´ terminus of 23S rRNA in *E. coli* is adjacent to a stemloop that terminates at the 5´ end with a base-paired nucleotide, meaning the 3´ overhang is efficiently trimmed by RNase T (Li *et al*., 1999), although this 5´ end doesn't represent the mature 5´ end that is generated by the subsequent action of RNase AM (Jain, 2020).

If the activities of Rex1 and RNase T are so similar, why have the two proteins adopted such different folds? One difference may be in the number of 5´ phosphates that are found in the 5S rRNA of their respective organisms: *S. cerevisiae* 5S rRNA often contain 5´ tri-phosphates in their mature products, whereas *E. coli* 5S tend to be produced with 5´ monophosphates (Soave *et al*., 1973; Guerrier-Takada *et al*., 1983; Mori and Ichiyanagi, 2021). With the hallmarks of RNase T-like activity in mind, it is possible to consider whether this activity seen in human RNA metabolism.

1.4.7 Human RNase T-like activity may be distributed between multiple DEDDh enzymes.

The most likely sequence homologue of Rex1 in humans is REXO5 (Silva *et al*., 2017; Gerstberger *et al*., 2017), which possesses predicted structural features of Rex1 that may be universally conserved throughout eukaryotes, but a close sequence homologue of RNase T itself is represented in mammals by the TREX1 family (Yuan *et al*., 2015).

1.4.7.1 REXO5

REXO5 has received characterisation in mouse (Silva *et al*., 2017) and *D. melanogaster* (Gerstberger *et al*., 2017). The initial study in mouse examined both murine and human REXO5, and described how expression measured by reverse transcription PCR (RT-PCR) of *REXO5* RNA in mice seems to be limited to the testes, and that REXO5 seems not to be expressed in HeLa or HEK293 cells. This resembles what can be seen for human REXO5 in the Genotype Tissue expression (GTEx) database, which suggests that REXO5 levels in the testes reach wildly variable levels with a median of roughly 400 transcripts per million (TPM), with other tissues showing baseline expression of between 0.5-10 TPM as shown in Figure 1.7A. Although not detected by Silva *et al*., the cell line gene expression database suggests that REXO5 may be expressed at low levels in most cell lines with the exception of Jurkat cells as shown in Figure 1.7B.

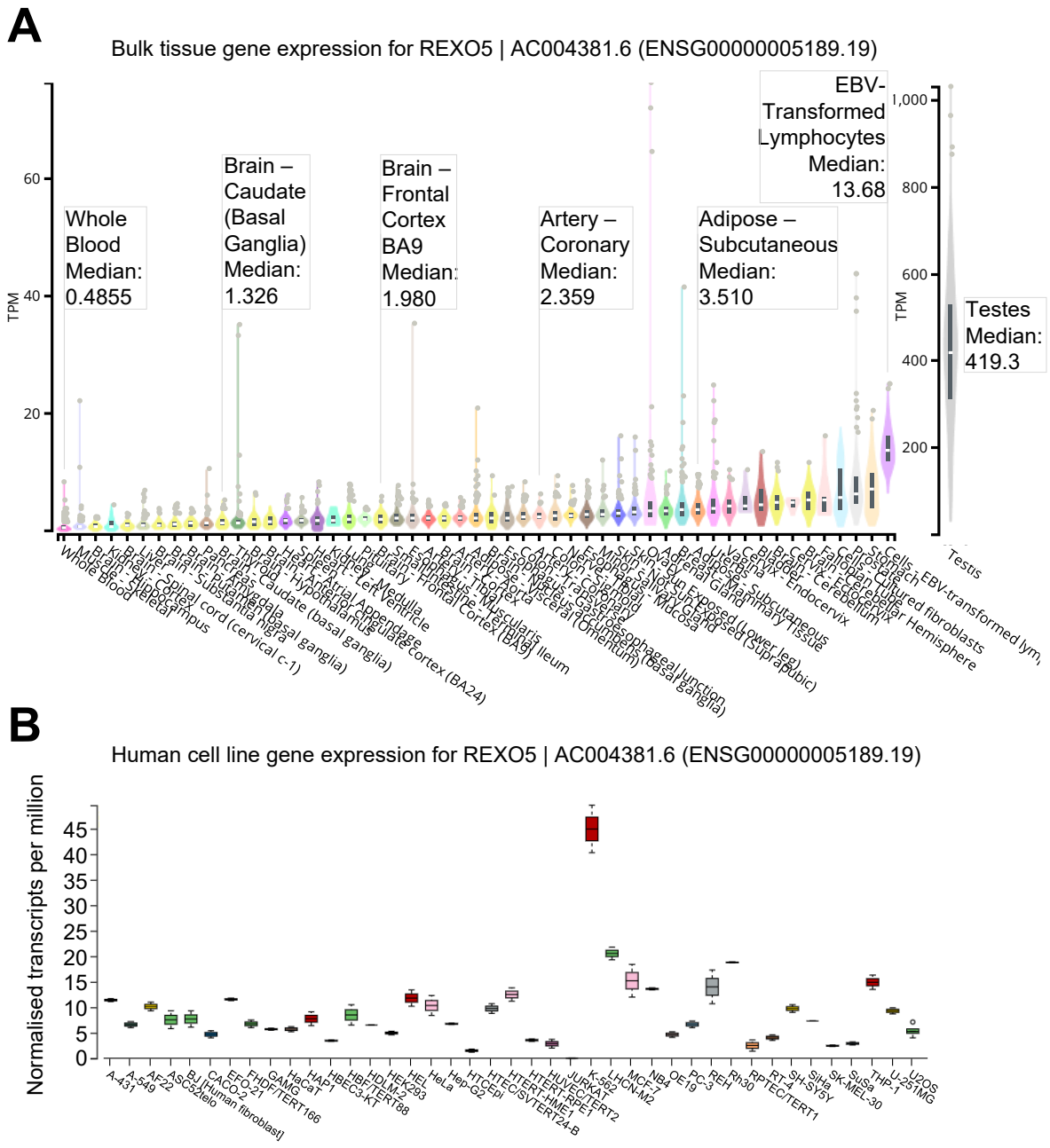

A
Bulk tissue gene expression for REXO5 | AC004381.6 (ENSG00000005189.19)
TPM
Whole Blood Median: 0.4855
Brain – Caudate (Basal Ganglia) Median: 1.326
Brain – Frontal Cortex BA9 Median: 1.980
Artery – Coronary Median: 2.359
Adipose – Subcutaneous Median: 3.510
EBV-Transformed Lymphocytes Median: 13.68
Testes Median: 419.3
B
Human cell line gene expression for REXO5 | AC004381.6 (ENSG00000005189.19)
Normalised transcripts per million

Figure 1.7: *REXO5* expression in human tissues is at its highest in the testes
The expression of human *REXO5* mRNA (ENSG00000005189.19) based on online expression data repositories. **A:** The expression of *REXO5* mRNA in different tissues of the human body. Tissues, shown on the X-axis, are arranged in order of increasing expression. Boxes represent lower and upper quartiles of data points, median indicated with a white line. The following modifications are made for clarity: the median expression several tissues at regular percentiles of the distribution are stated in boxes; the expression of *REXO5* in the testes is shown on a separate graph to increase Y-axis resolution for the other tissues. **B:** The expression of *REXO5* mRNA in a range of common human cell lines. Boxes represent upper and lower quartiles, with the median indicated by a horizontal black line. Two replicates are shown for each cell line with the exception of HDLM-2 and SiHa for which only one replicate is shown, and U2OS for which 4 replicates are shown. The following modifications were made for clarity: The axis labels were rewritten in larger fonts; a notched line was added to link boxes to cell line names. **Licenses:** ***A****: Graph .svg image reproduced in accordance with the GTEx Portal Data License. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 31/8/2023 and dbGaP Accession phs000424.v8.p2, at the following link: https://www.gtexportal.org/home/gene/ENSG00000005189 on 31/8/2023.* ***B****: Screenshot taken 31/8/2023 from the Human Protein Atlas v23.0, subsection 'Cell line dataset' in accordance with 'Creative Commons Attribution-ShareAlike 3.0 International' License, graph available at v23.proteinatlas.org/ENSG00000005189-REXO5/summary/rna* (Uhlén *et al*., 2015).

Silva *et al*. demonstrate properties of human REXO5 including localisation, *in vitro* activity, and examined the phenotype of *REXO5* loss using single-cell sequencing of RNA for a mouse CRISPR-Cas9 knockout model. In a *REXO5* plasmid-transfected HeLa cell model, human REXO5 was observed to primarily accumulate in the nucleolus in an N-terminal NLS-dependent manner. Recombinant REXO5 demonstrated distributive 3´ exonuclease activity against RNA and DNA substrates, with a preference for RNA. REXO5 was also shown to be active against an RNA hairpin, although the substrate may have been a sub-optimal substrate for RNAse T-like activity due to the presence of an unpaired 5´ overhang. This study concluded with analysis of a mouse *REXO5* knockout, which revealed no defects in fertility or growth. A single-cell sequencing approach was used to identify differences in gene expression, however defects in ribosomal RNA processing were not examined as ribosomal RNA was depleted before library preparation.

An analysis of REXO5 in *D. melanogaster* was published shortly after, describing a *REXO5* knockout strain with embryonic lethality and defects in snoRNA and rRNA maturation (Gerstberger *et al*., 2017). Small RNA sequencing and Northern blotting revealed an accumulation of 3´ extensions for 27S rRNA, 5S rRNA, and the majority of snoRNAs. The depletion of mature snoRNA was proposed to cause ribosomal assembly defects, leading to nucleolar stress and genomic instability. The authors noted that unlike in *S. cerevisiae*, Rrp6 homologues in higher eukaryotes are not involved with the 3´ maturation of snoRNAs, with PARN serving to mature a subset of snoRNAs in a polyA-dependent manner as a potentially inefficient alternative in the absence of REXO5.

Considered through the lens of RNase T activity, snoRNAs don't seem to represent optimal substrates: both C/D and H/ACA box snoRNAs feature a 3´ overhang adjacent to a stemloop preceded by a 2nt unpaired 5´ extension, or protracted extra sequence for C/D and H/ACA box snoRNAs respectively (for review, see Reichow *et al*., 2007). Although 3´ extended snoRNAs accumulate in *REXO5* knockout

flies, a large fraction of mature length snoRNAs still accumulate (Gerstberger *et al*., 2017). This may mirror Rex1's role in *S. cerevisiae* RNA metabolism where its broad exonuclease activity enables processing at comparable levels to other exonucleases that can equally process 3´ overhangs of stemloop with unpaired 5´ extensions, as opposed to relying on RNase T-like activity (as discussed in section 1.4.5.2).

A clear indicator of RNase T-like activity is the accumulation of extended 5S rRNA, and the study of *D. melanogaster* REXO5 clearly demonstrates loss of the mature 5S species. This may suggest that REXO5 is the only exonuclease in *D. melanogaster* with RNase T-like activity. The lack of REXO5 activity seems to have a severe phenotype in *D. melanogaster*, but has seemingly little impact in mice. One simple explanation for this could be that there is an additional source of RNase T-like activity in the mouse proteome, allowing redundancy in the processing of RNase T-like substrates. TREX1 represents a good candidate as it bears close structural and sequence homology to RNase T (Yuan *et al*., 2015) and is found in mammals, but a homologue is yet to be identified in *D. melanogaster*.

1.4.7.2 TREX1

TREX1 is a homodimeric DEDDh exonuclease (Mazur and Perrino, 1999, 2001). Since its initial isolation (Lindahl *et al*., 1969), TREX1 has received intense characterisation in its capacity as a DNA repair enzyme (for review see Lindahl, 2013). TREX1 also mediates the turnover of cytoplasmic DNA that otherwise leads to cGAS-mediated inflammation in the pathogenesis of autoimmune disease (for review, see Wang *et al*., 2022).

A recent biochemical study however may overturn the long-held view that TREX1 is only active on DNA (Yuan *et al*., 2015). Previous *in vitro* studies of TREX1 activity made the observation that TREX1 was unable to process ssRNA to any extent (Mazur and Perrino, 1999), however Yuan and colleagues made the incisive observation that the initial biochemical studies of TREX1 used an isoform with 10 fewer amino acids than the most common isoform, and that the addition of these 10 amino acids to untagged TREX1 is sufficient to confer RNA exonuclease activity. The RNA exonuclease activity of TREX1 was measured as 10-fold less than is seen against DNA (Yuan *et al*., 2015), but this is also known to be the case for RNase T (Zuo and Deutscher, 1999). A degree of sequence-dependence was observed for ssRNA trimming: TREX1 processed polyA efficiently, but seemed to process polyU less efficiently. This is surprising, given the high U content of 5S and tRNA trailer sequences. However, Yuan and colleagues demonstrated the *in vitro* ability of TREX1 to trim the terminal adenine from the CCA-3´ of a tRNA, demonstrating RNase T-like activity. An additional case of *in vitro* RNA trimming has been recently reported for TREX1 against an siRNA-RISC RNP (Sim *et al*., 2022).

1.4.7.3 Eri1

Eri1 was initially characterised under the name 3´hExo, a DEDDh exonuclease that was observed to trim the 3´ overhangs of the stemloop structures found at the end of histone mRNAs (Dominski *et al*., 2003). In the years since, Eri1 has been implicated in the trimming of similar stemloop overhangs in the turnover of siRNA (Kennedy *et al*., 2004), and in the final, cytoplasmic step of 5.8S rRNA maturation (Ansel *et al*., 2008; Gabel and Ruvkun, 2008). The mechanistic basis for Eri1's ability to trim in close proximity to stem loops seems to be mediated by a C-terminal SAP domain (Yang *et al*., 2006), a domain that is otherwise thought to mediate DNA binding (Aravind and Koonin, 2000). Eri1 also demonstrates a

preference for a consensus ACCCA motif in the binding of the 3´ overhang, and has a reduced affinity to histone mRNAs where this sequence has been artificially replaced with a AUUUU sequence (Yang *et al*., 2006). Of Eri1's substrates, the substrate that most closely resembles an RNase T-like base paired 5´ terminal stemloop with 3´ overhang is siRNA, a 22 nucleotide dsRNA duplex flanked with 3 nucleotide 3´ overhangs, although there are hints that this substrate may not be processed in an RNase T-like manner; as opposed to trimming a 3´ overhang up until the base of the stemloop for RNA maturation as seen for RNase T, Eri1 seems to be able to process into the stemloop itself to mediate a broader function of degrading these RNAs (Hoefig *et al*., 2012). While Eri1 does serve a role in the cytoplasmic maturation of 5.8S rRNA (Ansel *et al*., 2008; Gabel and Ruvkun, 2008), the mature end it produces is 3 nucleotides away from the terminal stemloop, and this stemloop possesses a 2 nucleotide 5´ extension (Petrov *et al*., 2014). The 3´ maturation of 5.8S thus resembles a different niche of exonuclease activity, rather than RNase T-like activity; it remains to be seen whether Eri1 is capable of mediating RNase T-like processing.

1.4.7.4 RNase T-like activity in humans: specialisation or redundancy?

REXO5 and TREX1 seem to represent the best candidates for RNase T-like activity in humans, and it will be intriguing to examine the extent of redundant processing shared between them. There may also be uncharacterised human RNases that also harbour this activity; Eri1 has two close sequence homologues, Eri2 and Eri3, that are yet to receive any characterisation. Aside from the potential for redundancy, additional barriers exist to the study of these exonucleases. The extensively studied DNA exonuclease behaviour of TREX1 may render interpretation of cell-level defects in nucleotide metabolism difficult, but the dependence on the 10 N-terminal amino acids for RNA exonuclease activity may present an opportunity for selective depletion by RNAi, or a separation of function mutation to remove the relevant exon. In the study of REXO5, it will be important to assess the extent of expression at the protein level by proteomics or Western blotting in mammalian tissues and cell lines, as the focus on mRNA-level expression currently suggests that mammalian *REXO5* is only expressed in the testes (Silva *et al*., 2017). When attempting to study sources of RNase T-like activity in human cells, it will be important to consider the possibility that there may be several nucleases with mechanistic means to the same 3´ ends.

1.5 Aims and objectives

The RNase T-like activity of Rex1 is required for the full maturation of many stable RNAs in *S. cerevisiae*. This project aimed to dissect the substrate recognition mechanisms that allow Rex1 to perform this activity, and identify analogous mechanisms that may be conserved in humans. A bioinformatic sequence analysis of the Rex1 DEDDh exonuclease domain was used to identify Rex1 sequence homologues, and identify conservation of substrate binding mechanisms. A later structural model of Rex1 released in the Alphafold Database (Jumper *et al*., 2021; Varadi *et al*., 2022) was analysed, revealing a novel and widely conserved 'RYS' domain composed of the uncharacterised N- and C-terminal sequence that flanks the exonuclease domain. Structural models of Rex1 in complex with an RNA substrate were generated to give a list of prospective substrate binding features in both the exonuclease domain and RYS domain of Rex1. This project next aimed to conduct a biochemical analysis of wild-

type Rex1, examining the buffer requirements and reaction parameters for the trimming of a simple ssDNA substrate. Rex1 mutants for analysis by X-ray crystallography were purified with the aim of producing an experimentally-derived Rex1 structure. In the absence of an experimental Rex1 structure, the substrate binding features of Rex1 predicted by the bioinformatic analysis were examined. An *in vivo* RNA-crosslinking approach revealed RNA binding to both the N- and C-terminal regions of the RYS domain. Rex1 mutants lacking features of the RYS domain and residues of the exonuclease domain were cloned and examined for *in vivo* and *in vitro* Rex1 activity, revealing specific defects seen in the absence of specific Rex1 features. This study concluded with the generation of a yeast expression vector designed for the expression of human DEDDh exonucleases, and observed a lack of complementation by *REXO5* of yeast mutants lacking or deficient in Rex1 activity.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 General buffers and reagents

Table 2.1: List of buffers and solutions for SDS-PAGE and Western blotting

Name:	Use:	Composition:
5x Protein loading buffer	Loading of proteins onto SDS-PAGE gels, denaturing lysis of *E. coli* and yeast	25% β-mercaptoethanol, 25% glycerol, 400mM Tris-HCl pH 6.8, 0.25% bromophenol blue, 5% SDS
1x Protein loading buffer		5% β-mercaptoethanol, 5% glycerol, 80mM Tris-HCl pH 6.8, 0.05% bromophenol blue, 1% SDS
TGS running buffer	Running buffer for electrophoresis of proteins	25mM Tris, 198mM glycine, 0.1% SDS
Stacking protein acrylamide gel solution	Electrophoresis of proteins	4.67% 38:1 Protogel acrylamide, 117mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulphate
12% resolving protein acrylamide gel solution	Electrophoresis of proteins	12% 38:1 Protogel acrylamide, 42mM Tris-HCl pH 8.7, 0.1% SDS, 0.1% ammonium persulphate, 0.001% TEMED
Western transfer buffer	Western blot transfer of protein resolved by PAGE onto nitrocellulose membrane	12.5mM Tris, 96mM glycine, 20% methanol, 0.1% SDS
TBST	Washing and probing of membranes during Western blotting	10mM Tris-HCl pH 6.8, 150mM NaCl, 0.1% Tween-20
ECL solution I	Visualisation of antibodies bound during western blotting by chemiluminescence	2.5mM luminol, 400µM p-coumaric acid, 100mM Tris-HCl pH 8.7
ECL solution II		5.4mM H_2O_2, 100mM Tris-HCl pH 8.7

Table 2.2: List of buffers and solutions for nucleic acid electrophoresis and Northern blotting

Name:	Use:	Composition:
6x DNA loading buffer	Loading of nucleic acids into agarose gels	30% glycerol, 0.25% bromophenol blue
1x DNA loading buffer		5% glycerol, 0.042% bromophenol blue
2x Formamide RNA loading buffer	Loading of nucleic acids onto Urea-TBE gels, stopping solution for exonuclease assays	95% formamide 20mM EDTA, 0.05% bromophenol blue 0.05% xylene cyanol
1x Formamide RNA loading buffer		47.5% formamide 10mM EDTA, 0.025% bromophenol blue 0.025% xylene cyanol
10x TBE	Running buffer and gel solution for electrophoresis of nucleic acids	900mM Tris, 900mM boric acid, 20mM EDTA
0.5x TBE		45mM Tris, 45mM boric acid, 1mM EDTA
16% DNA acrylamide gel solution	Electrophoresis of short DNA oligonucleotides	16% 19:1 acrylamide, 0.5x TBE, 50% urea, 0.07% ammonium persulphate, 0.007% TEMED
8% RNA acrylamide gel solution	Electrophoresis of RNA	8% Accugel 19:1 acrylamide, 45mM Tris, 45mM boric acid, 1mM EDTA, 50% urea, 0.07% ammonium persulphate, 0.007% TEMED
20x SSPE	Transfer, washing, probing, and stripping of Northern blot membranes	3M NaCl, 180mM NaH_2PO_4, 20mM EDTA, pH 7.4 with NaOH
6x SSPE		900mM NaCl, 54mM NaH_2PO_4, 6mM EDTA, pH 7.4 with NaOH
2x SSPE		300mM NaCl, 18mM NaH_2PO_4, 2mM EDTA, pH 7.4 with NaOH
0.1x SSPE		15mM NaCl, 0.9mM NaH_2PO_4, 0.1mM EDTA, pH 7.4 with NaOH
50x Denhardt's solution		2% ficoll, 2% polyvinylpyrrolidone, 2% BSA
Northern oligohybridisation buffer		300mM NaCl, 18mM NaH_2PO_4, 2mM EDTA, pH 7.4 with NaOH, 0.2%ficol, 0.2 polyvinylpyrrolidone, 0.2% BSA
Northern stripping buffer		15mM NaCl, 0.9mM NaH_2PO_4, 0.1mM EDTA, pH 7.4 with NaOH, 0.1% SDS

Table 2.4: List of antibodies used in this study

Primary antibodies				
Name:	**Dilution**	**Incubation time (h)**	**Supplier:**	**Secondary antibody:**
Peroxidase anti-peroxidase (PαP)	1/10,000	2 hour	Sigma-Aldrich (P1291)	None
Rabbit anti-Glutathione-*S*-Transferase (α-GST)	1/10,000	2 hour	Merck (G7781)	GαRPO
anti-phosphoglycerate kinase (α-PGK1)	1/10,000	2 hour	Life technologies (clone 22C5D8)	GαMPO
Secondary antibodies				
Name:	**Dilution**	**Incubation time**	**Supplier:**	
Goat anti-rabbit peroxidase (GαRPO)	1/10,000	1 hour	Sigma-Aldrich (A4914)	
Goat anti-mouse peroxidase (GαMPO)	1/10,000	1 hour	Bio-Rad Laboratories (1706516)	

Table 2.3: List of buffers for extraction and purification of proteins and nucleic acids

Name:	**Use:**	**Composition:**
Plasmid Solution I	Alkaline lysis for plasmid miniprep	50mM glucose, 25mM Tris pH 8, 10mM EDTA pH8
Plasmid Solution II	Alkaline lysis for plasmid miniprep	0.2M NaOH, 1% SDS
Plasmid Solution III	Alkaline lysis for plasmid miniprep	3M KAc
DNA Phenol Chloroform	Alkaline lysis for plasmid miniprep	25:24:1 phenol (pH 6.5):chloroform:isoamyl alcohol
5x Protein Loading Buffer	Loading of proteins onto SDS-PAGE gels, denaturing lysis of *E. coli* and yeast	25% β-mercaptoethanol, 25% glycerol, 400mM Tris-HCl pH 6.8, 0.25% bromophenol blue, 5% SDS
1x Protein Loading Buffer		5% β-mercaptoethanol, 5% glycerol, 80mM Tris-HCl pH 6.8, 0.05% bromophenol blue, 1% SDS
E. coli native lysis buffer	Native extraction of soluble protein from *E. coli*	10mM Tris-HCl, 150mM NaCl, 0.1% Tween-20, 1mM PMSF, 5ng/µl Leupeptin, 7ng/µl Pepstatin A, 1mM DTT, 100µg/µl lysozyme
Yeast denaturing lysis buffer	Denaturing extraction of soluble protein from yeast	0.2M NaOH, 0.2% β-mercaptoethanol
Yeast native lysis buffer	Native extraction of soluble protein from yeast	50mM HEPES pH 7.6, 150mM NaCl, 1mM EDTA, 100µM PMSF, 1x yeast protease inhibitor cocktail; Melford, Ipswich, UK)
Yeast protein wash buffer	Wash steps for immunoprecipitation of yeast protein	50mM HEPES pH 7.6, 150mM NaCl
Yeast stringent wash buffer	Stringent wash steps for immunoprecipitation of yeast protein	25mM HEPES pH 7.6, 75mM NaCl, 2M $MgCl_2$
PNK buffer	^{32}P-labeling of RNA-protein crosslinked complexes	50mM HEPES pH 7.6, 10mM $MgCl_2$, 1mM DTT
GTC mix	RNA extraction from yeast	47% GTC, 50mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0, 2% sarkosyl, 1% β-mercaptoethanol
NaAc mix	RNA extraction from yeast	100mM NaAc pH 5.0, 10mM Tris HCl pH 8.0, 1mM EDTA, DEPC-H_2O
RNA phenol chloroform	RNA extraction from yeast	25:24:1 pH 4 phenol: chloroform: isoamyl achohol

2.1.2 List of *E. coli* strains, media, and transformation buffers used in this study

Table 2.5: List of *E. coli* strains

Strain	Use:	Genotype:	Source:
DH5α	Propagation of plasmid stocks and cloning of ligation products	F–, *endA1, glnV44, thi-1, recA1, relA1, gyrA96, deoR, nupG, Φ80dlacZΔM15, Δ(lacZYA-argF)U169, hsdR17(rK-mK+)*, λ–	Invitrogen, Paisley, UK
BL21 (DE3) pLysS	Production of recombinant protein through autoinduction	F–, *dcm, ompT, lon, hsdS(rB-, mB-), gal,* λ(DE3), pLysS	Stratagene, California, U.S.A

Table 2.6: List of *E. coli* media compositions and antibiotics

Media	Composition:
LB (Luria Bertani)	1% Bactotryptone, 0.5% Yeast extract, 1% NaCl, 2% Bactoagar for plate media
2xYT	2% Bactotryptone, 1% Yeast extract, 1% NaCl
Auto-induction (Studier, 2005)	1.2% Bactotryptone, 2.4% Yeast extract, 0.6% Glycerol, 0.23% KH_2PO_4, 1.25% K_2HPO_4, 0.5% Glucose, 2% Lactose
1000x Ampicillin	80mg/ml, 50% ethanol
1x Ampicillin	80µg/ml, 0.05% ethanol
1000x Chloramphenicol	25mg/ml, 100% ethanol
1x Chloramphenicol	25µg/ml, 0.1% ethanol

Table 2.7: Competent *E. coli* buffers

Transformation Buffer	Composition:
TFI	30mM Kac, 100mM $RbCl_2$, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% Glycerol, pH 5.8 with acetic acid, sterile filter
TFII	10mM MOPS, 10mM $RbCl_2$, 75mM $CaCl_2$, 15% Glycerol, pH 6.5 with KOH, sterile filter

2.1.3 List of *S. cerevisiae* strains, media, and transformation buffers used in this study

All strains are listed with ‘P’ numbers as part of the Mitchell laboratory nomenclature system.

Table 2.8: List of *S. cerevisiae* strains

Strain	Use:	Genotype:	Source:
BY4741 (P364)	Wild-type laboratory strain	Mata, *his3*Δ1, *leu2*Δ0, *met15*Δ0, *ura3*Δ0	Euroscarf, The University of Frankfurt, Germany
P356	*RRP47* deletion strain	Mata, *his3*Δ1, *leu2*Δ0, *met15*Δ0, *ura3*Δ0, *rrp47*Δ::*KANMX4*	
P550	*REX1* deletion strain	Matα *his3*Δ1 *leu2*Δ1 *lys2*Δ0 *ura3*Δ0 *rex1*Δ::*KANMX4*	
P1604	*rex1*Δ*/rrp47*Δ plasmid shuffle strain	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + p659(*zz-REX1, URA3*)	Taib Hama Soor, (Hama Soor, 2017; Daniels et al., 2022)
P2100	HTP-tagged *rex1* strain for measuring endogenous *rex1* expression	Mata *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 *rex1-HTP*::*URA3*	
P2240, P2241, P2242	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*K340A)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*K340A)	This study
P2243, P2244, P2245	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*Y272A)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*Y272A)	
P2246, P2247, P2248	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*N312A)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*N312A)	
P2249, P2250, P2251	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*H308A)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*H308A)	
P2252, P2253, P2254	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*S342A)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*S342A)	
P2255, P2256, P2257	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*ΔLoop1)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*ΔLoop1)	
P2258, P2259, P2260	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*ΔLoop2)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*ΔLoop2)	
P2261, P2262, P2263	5-FOA isolates 1-3 of P1604 + pRS313(zz-*rex1*ΔLoop3)	*Mata his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0 rex1*Δ*::KANMX4 rrp47*Δ*::HphMX4* + pRS313(zz-*rex1*ΔLoop3)	

Table 2.9: List of *S. cerevisiae* media compositions

Media	Stock composition:
YPD	2% Bactopeptone, 1% Yeast extract, 2% Glucose, 2% Bactoagar for plate media
Synthetic complete media (-His, -Leu; SD, SGal, SRaff)	0.17% yeast nitrogen base, 0.5% ammonium sulphate, amino acids as described in 'List of amino acid stocks and supplements for *S. cerevisiae* media' , excluding histidine for '-His', or leucine for '-Leu', 2% carbon source: Glucose, 'SD' ; galactose, 'SGal' ;or raffinose, 'SRaf'

Table 2.10: List of amino acid stocks and supplements for *S. cerevisiae* media

Amino acids/ supplements	Stock composition:	Final Concentration (mg/L, % w/v)
'-4 mix' dry powder mix	2g adenine hemisulphate, 2g arginine monohydrochloride, 3g lysine monohydrochloride, 2g methionine, 5g phenylalanine, 20g threonine, 3g tyrosine	370mg/L (0.002% arginine, 0.02% threonine, 0.003% tyrosine, 0.005% phenylalanine, 0.002% methionine, 0.003% lysine, 0.002% adenine, 0.002% arginine
Leucine	0.6% leucine (100x)	0.006%
Histidine	0.2% histidine monohydrochloride monohydrate (100x)	0.002%
Tryptophan	0.2% tryptophan (100x)	0.002%
Uracil	n/a	20mg/L or 5mg/L for 5′FOA plates
5-FOA	10% in DMSO	0.1%

Table 2.11: *S. cerevisiae* transformation buffers

Transformation Buffer	Composition:
TE	10mM Tris-HCl pH 7.5, 10mM EDTA pH 8.0, filter sterilised
LiT	10mM Tris-HCl pH 7.5, 10mM EDTA pH 8.0, 100mM LiAc pH 7.5, filter sterilised

2.1.4 List of plasmids used and generated in this study

All plasmids originating from the Mitchell laboratory are listed with 'p' numbers without additional letters (such as pRS) as part of the lab nomenclature system.

Table 2.12: List of plasmid backbones used in this study

Plasmid	Contents	Use in this study	Source:
pGEX-6-P1	*amp*, *GST::ORF*	Vector for *E. coli* expression of GST-tagged protein	(Smith and Johnson, 1988)
pRSETb	*amp*, *6xHis::ORF*	Vector for *E. coli* expression of 6xHis-tagged and untagged protein	(Schoepfer, 1993)
pBlueScript II KS	*amp*	Compact vector for site-directed mutagenesis	(Alting-Mees and Short, 1989)
pRS425	*amp*, *LEU2*, *2μ*, *zz::ORF*	High-copy number yeast expression vector for zz-tagged protein	(Christianson et al., 1992)
pRS415	*amp*, *LEU2*, *CENP*, *zz::ORF*	Low-copy number yeast expression vector for zz-tagged protein	(Sikorski and Hieter, 1989)
pRS313	*amp*, *HIS3*, *CENP*, *zz::ORF*		
pRS314	*amp*, *TRP1*, *CENP*, *zz::ORF*		
p956	*amp*, *TRP1*, *CENP*, *GAL1 – zz::ORF-Adh13'UTR*	Low-copy number yeast expression vector derived from pRS314. For human DEDDh ORFs, featuring *GAL* promoter and the 3′ UTR of *Adh1*	This study
p968	*amp*, *LEU2*, *CENP*, *GAL1 – zz::ORF-Adh13'UTR*	Low-copy number yeast expression vector derived from pRS415. For human DEDDh ORFs, featuring *GAL* promoter and the 3′ UTR of *Adh1*	
p969	*amp*, *HIS3*, *CENP*, *GAL1 – zz::ORF-Adh13'UTR*	Low-copy number yeast expression vector derived from pRS313. For human DEDDh ORFs, featuring *GAL* promoter and the 3′ UTR of *Adh1*	

Table 2.13: List of pre-existing plasmids used in this study

Plasmid	Contents	Use in this study	Source:
p674	pRS313[*amp*, *CENP*, HIS3, *rrp4pro - zz-REX1*]	Yeast expression of *REX1* (HIS marker)	Taib Hama Soor (Hama Soor, 2017; Daniels et al., 2022)
p675	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-REX1*]	Yeast expression of *REX1* (LEU marker)	
p679	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*D229A]	Plasmid shuffle assay of *rex1* truncation mutants based on Hama Soor, 2017 and Daniels et al., 2022	
p680	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*Δ509-553]		
p701	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*Δ1-205]		
p705	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*Δ509-553]		
p706	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*Δ84-205]		
p713	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*Δ1-84]		
p716	pRS415[*amp*, *CENP*, *LEU2*, *rrp4pro - zz-rex1*Δ428-470]		
p719	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*Δ1-205]		
p720	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*Δ1-84]		
p721	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*Δ428-470]		
p748	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*Δ84-205]		
p752	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-REX1*]		
p754	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro - zz-rex1*D229A]		
p869	pRS414[*amp*, *CENP*, *TRP1 GAL – HA-MTR4*]	Cloning source of *GAL* promoter sequence	Phil Mitchell
p872	pGEX-6P-1[*amp*, *tac – GST-REX1*]	*E. coli* expression of *GST-REX1* for *in vitro* assays	William Royle
p880	pGEX-6P-1[*amp*, *tac – GST-rex1*D229A]	*E. coli* expression of *GST-rex1*D229A for *in vitro* assays	Sopida Wongwas
p966	pRS313[*amp*, *CENP*, *HIS3*, *rrp4pro – zz-rex1*Δ459-472]	Yeast expression of *rex1* truncation mutants based on Alphafold2 model (Jumper et al., 2021, Varadi et al., 2022)	Sophie Kelly
p967	pRS313[*amp*, *CENP*, *HIS3*, *rrp4pro – zz-rex1*Δ145-154]		
p972	pRS313[*amp*, *CENP*, *HIS3*, *rrp4pro – zz-rex1*Δ506-544]		

Table 2.14: List of plasmids constructed during this study: p907-921

Plasmid	Contents	Use in this study	Plasmid Construction:
p907	pBlueScript II KS[*amp*, *rex1*K340AΔ543-553]	SDM product cloning intermediate of *rex1*K340A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p933 using o1255/o1256 (HindIII)
p908	pBlueScript II KS[*amp*, *rex1*Δ543-553]	Cloning intermediate of *rex1*, truncated at *rex1*'s internal XmaI site. For use as an extended PCR SDM template	EcoRI-XmaI fragment of p675 ligated into pBlueScript II KS
p909	pBlueScript II KS[*amp*, *rex1*Y272AΔ543-553]	SDM product cloning intermediate of *rex1*Y272A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1271/o1272 (NheI)
p910	pBlueScript II KS[*amp*, *rex1*H220-SR(BglII)-G221Δ543-553]	SDM product cloning intermediate of *rex1*H220-SR(BglII)-G221, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1243/o1244 (BglII)
p911	pBlueScript II KS[*amp*, *rex1*R366-SR(BglII)-A367Δ543-553]	SDM product cloning intermediate of *rex1*R366-SR(BglII)-A367, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1245/o1246 (BglII)
p912	pBlueScript II KS[*amp*, *rex1*N511-SR(BglII)-A512Δ543-553]	SDM product cloning intermediate of *rex1*N511-SR(BglII)-A512, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1247/o1248 (BglII)
p913	pBlueScript II KS[*amp*, *rex1*N312AΔ543-553]	SDM product cloning intermediate of *rex1*N312A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1265/o1266 (PstI)
p914	pBlueScript II KS[*amp*, *rex1*H308AΔ543-553]	SDM product cloning intermediate of *rex1*H308A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1267/o1268 (NheI)
p915	pBlueScript II KS[*amp*, *rex1*S342AΔ543-553]	SDM product cloning intermediate of *rex1*S342A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1269/o1270 (AflII)
p916	pBlueScript II KS[*amp*, *rex1*H220-SR-PsP(concatenated)-SR-G221Δ543-553]	Linker insertion cloning intermediate of *rex1*H220-SR-PsP(concatenated)-SR-G221, truncated at *rex1*'s internal XmaI site	Insertion of linker o1259/o1260 into BglII site of p910
p917	pBlueScript II KS[*amp*, *rex1*R366-SR-PsP(concatenated)-SR-A367Δ543-553]	Linker insertion cloning intermediate of *rex1*R366-SR-PsP(concatenated)-SR-A367, truncated at *rex1*'s internal XmaI site	Insertion of linker o1259/o1260 into BglII site of p911
p918	pBlueScript II KS[*amp*, *rex1*N511-SR-PsP-SR-A512Δ543-553]	Linker insertion cloning intermediate of *rex1*N511-SR-PsP-SR-A512, truncated at *rex1*'s internal XmaI site	Insertion of linker o1259/o1260 into BglII site of p912
p919	pRS313[*amp*, *CENP*, HIS3, *rrp4pro - zz-rex1K340A*]	Yeast expression of *zz-rex1K340A* (HIS marker)	EcoRI-XmaI fragment of p907 ligated into pRS313
p920	pRS313[*amp*, *CENP*, HIS3, *rrp4pro - zz-rex1Y272A*]	Yeast expression of *zz-rex1Y272A* (HIS marker)	EcoRI-XmaI fragment of p909 ligated into pRS313
p921	pRS313[*amp*, *CENP*, HIS3, *rrp4pro - zz-rex1N312A*]	Yeast expression of *zz-rex1N312A* (HIS marker)	EcoRI-XmaI fragment of p913 ligated into pRS313

Table 2.15: List of plasmids constructed during this study: p922-935

Plasmid	Contents	Use in this study	Plasmid Construction:
p922	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - *zz-rex1H308A*]	Yeast expression of *zz-rex1H308A* (HIS marker)	EcoRI-XmaI fragment of p914 ligated into pRS313
p923	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - *zz-rex1S342A*]	Yeast expression of *zz-rex1S342A* (HIS marker)	EcoRI-XmaI fragment of p915 ligated into pRS313
p924	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - zz-*rex1*H220-SR-PsP(concatenated)-SR-G221]	Sub-cloning intermediate of *rex1*H220-SR-PsP(concatenated)-SR-G221	EcoRI-XmaI fragment of p916 ligated into pRS313
p925	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* – zz-*rex1*R366-SR-PsP(concatenated)-SR-A367]	Sub-cloning intermediate of *rex1*R366-SR-PsP(concatenated)-SR-A367	EcoRI-XmaI fragment of p917 ligated into pRS313
p926	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* – zz-*rex1*N511-SR-PsP-SR-A512]	Yeast expression of *zz-rex1*N511-SR-PsP-SR-A512 (HIS marker) for testing viability of SR-PsP-SR insertion *rex1* mutant	EcoRI-XmaI fragment of p918 ligated into pRS313
p927	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - zz-*rex1*H220-SR-LEVLLQDFQR-SR-G221]	Sub-cloning intermediate of zz-*rex1*H220-SR-LEVLLQDFQR-SR-G221	Removal of concatemer from p924 using EcoNI
p928	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* – zz-*rex1*R366-SR-PsP-SR-A367]	Yeast expression of *zz-rrex1*R366-SR-PsP-SR-A367 (HIS marker) for testing viability of SR-PsP-SR insertion *rex1* mutant	Removal of concatemer from p925 using EcoNI
p929	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - zz-*rex1*H220-SR-PsP-SR-G221]	Yeast expression of zz-*rex1*H220-SR-PsP-SR-G221 (HIS marker) for testing viability of SR-PsP-SR insertion *rex1* mutant	Extended PCR SDM of p927 using o1273/o1274 (ApaI)
p930	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - zz-*rex1*H220-SR-PsP-SR-G221,D229A]	Yeast expression of zz-*rex1*H220-SR-PsP-SR-G221,D229A for Rex1-RNA crosslinking mapping	Extended PCR SDM of p929 using o1275/o1276 (SphI)
p931	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* – zz-*rex1*D229A,R366-SR-PsP-SR-A367]	Yeast expression of zz-*rex1*D229A,R366-SR-PsP-SR-A367 for Rex1-RNA crosslinking mapping	Extended PCR SDM of p928 using o1275/o1276 (SphI)
p932	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* – zz-*rex1*D229A,N511-SR-PsP-SR-A512]	Yeast expression of zz-*rex1*D229A,N511-SR-PsP-SR-A512 for Rex1-RNA crosslinking mapping	Extended PCR SDM of p926 using o1275/o1276 (SphI)
p933	pBlueScript II KS[*amp*, zz-*rex*1Δ543-553, -HindIII site, +NheI site]	Klenow reaction cloning intermediate for extended PCR SDM to give p907	Klenow fill-in treatment of p908 HindIII site
p934	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro* - zz-*rex1*H220-SR-PsP-SR-G221(D229A)]	High copy number yeast expression of zz-*rex1*H220-SR-PsP-SR-G221,D229A for Rex1-RNA crosslinking mapping	NotI-XmaI fragment of p930 ligated into pRS425
p935	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro* – zz-*rex1*(D229A)R366-SR-PsP-SR-A367]	High copy number yeast expression of zz-*rex1*D229A,R366-SR-PsP-SR-A367 for Rex1-RNA crosslinking mapping	NotI-XmaI fragment of p931 ligated into pRS425

Table 2.16: List of plasmids constructed during this study: p936-954

Plasmid	Contents	Use in this study	Plasmid Construction:
p936	pRS425[*amp*, *2μ*, *LEU2*, *rrp4pro* – zz-*rex1*(D229A)N511-SR-PsP-SR-A512]	High copy number yeast expression of zz-*rex1*D229A,N511-SR-PsP-SR-A512 for Rex1-RNA crosslinking mapping	NotI-XmaI fragment of p932 ligated into pRS425
p941	pBlueScript II KS[*amp*, *rex1*H360AΔ543-553]	SDM product cloning intermediate of *rex1*H360A, truncated at *rex1*'s internal XmaI site	Extended PCR SDM of p908 using o1289/o1290 (PflFI)
p943	pBlueScript II KS[*amp*, *hREXO5*]	cDNA PCR product cloning intermediate of human *REXO5*	PCR of HCT116 cDNA (courtesy of Ang Li from the Wilson Laboratory) using o1283(SalI)/o1284(ApaI) into pBlueScript II KS
p944	pRS313[*amp*, *CENP*, HIS3, *rrp4pro* - *rex1*H360A]	Yeast expression of zz-*rex1*H360A (HIS marker)	EcoRI-XmaI fragment of p941 ligated into pRS313
p945	pGEX-6P-1[*amp*, *tac* – *GST-rex1*H360A]	*E. coli* expression of *GST-rex1*H360A for *in vitro* assays	PCR on p944 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p946	pGEX-6P-1[*amp*, *tac* – *GST-rex1*K340A]	*E. coli* expression of *GST-rex1*K340A for *in vitro* assays	PCR on p919 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p947	pGEX-6P-1[*amp*, *tac* – *GST-rex1*Y272A]	*E. coli* expression of *GST-rex1*Y272A for *in vitro* assays	PCR on p920 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p948	pGEX-6P-1[*amp*, *tac* – *GST-rex1*N312A]	*E. coli* expression of *GST-rex1*N312A for *in vitro* assays	PCR on p921 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p949	pGEX-6P-1[*amp*, *tac* – *GST-rex1*H308A]	*E. coli* expression of *GST-rex1*H308A for *in vitro* assays	PCR on p922 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p950	pGEX-6P-1[*amp*, *tac* – *GST-rex1*S342A]	*E. coli* expression of *GST-rex1*S342A for *in vitro* assays	PCR on p923 using o1133(BamHI)/o1215(XhoI), ligated into pGEX-6P-1
p953	pRS414[*amp*, *CENP*, *TRP1, GAL*(-AgeI site) – *HA-MTR4*]	Klenow reaction cloning intermediate for generating *GAL1* promoter that lacks an AgeI site	Klenow treatment of p869 AgeI site
p954	pRS314[*amp*, *CENP*, *TRP1*, *GAL*(-AgeI site)]	Cloning intermediate for construction of human DEDDh ORF yeast expression vector, addition of *GAL1* promoter(lacking AgeI)	PCR on p953 using o1291(NotI)/o1292(SpeI), ligated into pRS313

Table 2.17: List of plasmids constructed during this study: p955-979

Plasmid	Contents	Use in this study	Plasmid Construction:
p955	pRS314[*amp*, *CENP*, *TRP1*, *GAL*(-AgeI site) – *zz-ORF*]	Cloning intermediate for construction of human DEDDh ORF expression vector, addition of *zz-* tag	PCR on p674 using o1293(SpeI)/o1294(BamHI), ligated into p954
p956	pRS314[*amp*, *CENP*, *TRP1*, *GAL*(-AgeI site) – *zz-ORF-Adh1 3′ UTR*]	Human DEDDh ORF yeast expression vector (TRP marker)	Colony PCR of P550 yeast using o1295(PstI)/o1296(ClaI), ligated into p954
p961	pRS314[*amp*, *CENP*, *TRP1*, *GAL*(-AgeI site) – *zz-REXO5-Adh1 3′ UTR*]	Gal-inducible yeast expression of *zz-REXO5* (TRP marker)	PCR on p943 using o1299(AgeI)/o1300(NsiI), ligated into p956(XmaI/PstI)
p962	pRSETb[*amp*, *T7* – *rex1*H360AΔ1-51]	*E. coli* expression of *rex1*H360AΔ1-51 for crystallisation screening	PCR on p945 using o1306(NdeI)/o1215(XhoI), ligated into pRSETb
p963	pRSETb[*amp*, *T7* – *6xHis-rex1*H360AΔ1-51]	*E. coli* expression of *6xHis-rex1*H360AΔ1-51 for crystallisation screening	PCR on p945 using o1336(NdeI)/o1215(XhoI), ligated into pRSETb
p968	pRS415[*amp*, *CENP*, *LEU2*, *GAL*(-AgeI site) – *zz-ORF-Adh1 3′ UTR*]	Human DEDDh ORF yeast expression vector (LEU marker)	NotI/SalI fragment of p956 ligated into pRS415
p969	pRS313[*amp*, *CENP*, *HIS3*, *GAL*(-AgeI site) – *zz-ORF-Adh1 3′ UTR*]	Human DEDDh ORF yeast expression vector (HIS marker)	NotI/SalI fragment of p956 ligated into pRS313
p970	pRS415[*amp*, *CENP*, *LEU2*, *GAL*(-AgeI site) – *zz-REXO5-Adh1 3′ UTR*]	Gal-inducible yeast expression of *zz-REXO5* (LEU marker)	NotI/SalI fragment of p968 ligated into pRS415
p971	pRS313[*amp*, *CENP*, *HIS3*, *GAL*(-AgeI site) – *zz-REXO5-Adh1 3′ UTR*]	Gal-inducible yeast expression of *zz-REXO5* (HIS marker)	NotI/SalI fragment of p968 ligated into pRS313
p976	pGEX-6P-1[*amp*, (-BamHI), *GST-ORF*]	T4 cloning intermediate for subcloning of *rex1* truncation mutants (p966, p967, p972) based on Alphafold2 model (Jumper et al., 2021, Varadi et al., 2022) into pGEX-6P-1	T4 fill-in treatment of pGEX-6P-1 BamHI site
p977	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ145-154]	*E. coli* expression of *GST-rex1*Δ145-154 for *in vitro* assays	PCR on p967 using o1133(BamHI)/o1215(XhoI), ligated into p976
p978	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ459-472]	*E. coli* expression of *GST-rex1*Δ459-472 for *in vitro* assays	PCR on p966 using o1133(BamHI)/o1215(XhoI), ligated into p976
p979	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ506-544]	*E. coli* expression of *GST-rex1*Δ506-544 for *in vitro* assays	PCR on p972 using o1133(BamHI)/o1215(XhoI), ligated into p976

Table 2.18: List of plasmids constructed during this study: p980-984

Plasmid	Contents	Use in this study	Plasmid Construction:
p980	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*N312A,K340A,S342A]	*E. coli* expression of*GST-rex1*N312A,K340A,S342A for *in vitro* assays	Extended PCR SDM of p948 using o1330/o1331 (NaeI)
p981	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ145-154,Δ459-472]	*E. coli* expression of *GST-rex1*Δ145-154,Δ459-472 for *in vitro* assays	Fusion PCR of o1345/o1347 on p967 and o1346/o1215 on p966, amplified with o1345(EcoRI)/p1215(XhoI) into pGEX-6P-1
p982	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ145-154,Δ506-544]	*E. coli* expression of *GST-rex1*Δ145-154,Δ506-544 for *in vitro* assays	Fusion PCR of o1345/o1349 on p967 and o1346/o1215 on p972, amplified with o1345(EcoRI)/p1215(XhoI) into pGEX-6P-1
p983	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ459-472,Δ506-544]	*E. coli* expression of *GST-rex1*Δ459-472,Δ506-544 for *in vitro* assays	Fusion PCR of o1345/o1349 on p967 and o1348/o1063 on p972, amplified with o1345(EcoRI)/p1215(XhoI) into pGEX-6P-1
p984	pGEX-6P-1[*amp*, (-BamHI), *GST-rex1*Δ145-154,Δ459-472,Δ506-544]	*E. coli* expression of *GST-rex1*Δ145-154,Δ459-472,Δ506-544 for *in vitro* assays	Fusion PCR of o1345/o1347 on p967 and o1346/o1349 on p966, and o1348/o1063 on p972 amplified with o1345(EcoRI)/p1215(XhoI) into pGEX-6P-1

2.1.5 List of oligonucleotides

All oligos are listed with ‘o’ numbers as part of the Mitchell laboratory nomenclature system.

Table 2.19: List of oligonucleotides used for sequencing

Oligo:	Name:	Sequence (5′ – 3′)	Used to sequence:
o235	pGEX-2T seq+	GCAAGCCACGTTTGGT	p945, p946, p947, p948, p949 p950,
o774	zz tag primer	CGAAAGTAGACAACAAATTC	p929, p931, p932, p956
o869	REX1_5'R1	CGCGAATTCAATGCAAGTAGAAGGGCC	p927
o884	REX1_+_(601)	CTGGATTCTGGAGACAC	p907, p909, p916, p928, p929, p931, p932,
o986	pGEX6P_rev	CTGCATGTGTCAGAGG	p945, p946, p947, p948, p949, p950
‘In-house’ oligos from sequencing company; Source Bioscience, Cambridge, UK	T3	ATT AAC CCT CAC TAA AG	p941, p954, p955, p956,
	T7F	TAA TAC GAC TCA CTA TAG GG	p941, p953, p954, p955, p956, p962, p963
	T7R	GCT AGT TAT TGC TCA GCG G	p962, p963
	PGEX5	GGGCTGGCAAGCCAC GTTTGGTG	p976, p977, p981, p982, p983, p984
	PGEX3	CCGGGAGCTGCATGT GTCAGAGG	p976, p977, p981, p982, p983, p984

Table 2.20: List of oligonucleotides used for exonuclease assays

Oligo:	Length (nt)	Sequence (5′ – 3′)
o1103	18	TTT TTT TTT TTT TTT TTT
o1165	27	CAC GGA TCC GAT GAA GTG GTT GTT GTT
o1337	18	TTT TTT TTT TTT CCT TTT
o1338	18	TTT TTT CCT TTT TTT TTT

Table 2.21: List of oligonucleotides used for Northern blotting

Oligo:	Sequence (5′ – 3′)	RNA species
o238	TCA CTC AGA CAT CCT AGG	U14
o243	GAG AGG TTA CCT ATT ATT A	snR38
o339	AGA AAC AAA GCA CTC ACG AT	Arg tRNA (UCU)
o517	ATC TCT GTA TTG TTT CAA ATT GAC CAA	U6
o821	GGT CAG ATA AAA GTA AAA AAA GGT AGC	snR13
o925	CTA CTC GGT CAG GCT C	5S rRNA
o950	CCT TGC TTA AGC AAA TGC	Lys tRNA(UUU), intron
o951	AAG ATT TCG TAG TGA TAA	Tyr tRNA (GUA)

Table 2.22: List of oligonucleotides used for cloning: o1133-o1291

Oligo:	Name:	Sequence (5′ – 3′):	For cloning plasmid:
o1133	REX1 5' Bam	GGT GGA TCC ATG CAA GTA GAA GGG CC	p945, p946, p947, p948, p949, p950, p962, p963, p977, p978, p979 (PCR insert cloning)
o1215	rex1_end_XhoI	GCG CTC GAG TTA TTT TAC AGT AAA GGA TG	p945, p946, p947, p948, p949, p950, p977, p978, o979, p981, p982, p983, p984 (PCR insert cloning)
o1243	REX1_H220_BglII_G221_F	GAA AGA TCT GGT GGC TCC CAC ATC	p910 (Extended PCR SDM)
o1244	REX1_H220_BglII_G221_R	GAA AGA TCT ATG GGT AAA GTC AAC AGT	
o1245	REX1_R366_BglII_A367_F	GAA AGA TCT AGG GCT TGT CTT GAA TTG	p911 (Extended PCR SDM)
o1246	REX1_R366_BglII_A367_R	GAA AGA TCT TGC ATC TTC GAC AGA ATC	
o1247	REX1_N511_BglII_L512_F	GAA AGA TCT CTT TCA ACT GAG TTA GAG T	p912 (Extended PCR SDM)
o1248	REX1_N511_BglII_L512_R	GAA AGA TCT ATT GTT CCA TGG TCT TGT	
o1255	Rex1_K340A_HindIII_F	GAA GCG CCA AGC TTA AAA TAC TTG AGC GAA ACC TTT C	p907 (Extended PCR SDM)
o1256	Rex1_K340A_HindIII_R	GAA GCT TGG CGC GAA AGG ATC GCC AGC TTT ATG	
o1259	PreScission_BglII_F2	GAT CTC TGG AAG TCC TGT TCC AGG GGC CCA	p916, p917, p918 (anneal with o1260 for linker insertion)
o1260	PreScission_BglII_R2	GAT CTG GGC CCC TGG AAC AGG ACT TCC AGA	p916, p917, p918 (anneal with o1259 for linker insertion)
o1265	Rex1_N312A_PstI_F	GAA CTG CAG GCT GAT TTG AAA GTC ATG AAA TTG	p913 (Extended PCR SDM)
o1266	Rex1_N312A_PstI_R	GAA GCC TGC AGC GAA TGT CCT ATC AAA ATA TC	
o1267	Rex1_H308A_NheI_F	GAA GCT AGC CTA CAG AAT GAT TTG AAA GTC	p914 (Extended PCR SDM)
o1268	Rex1_H308A_NheI_R	GAA GCT AGC TCC TAT CAA AAT ATC TGA ACG	
o1269	Rex1_S342A_AflII_F	GAA CTT AAG AGC TGG CTT GAA AGG ATC GCC A	p915 (Extended PCR SDM)
o1270	Rex1_S342A_AflII_R	GAA GCT CTT AAG TAC TTG AGC GAA ACC TTT CTG	
o1271	rex1_Y272A_NheI_F2	GAA GCT AGC GGT ATA ACT GAA GAG AAA CT	p908 (Extended PCR SDM)
o1272	rex1_Y272A_NheI_R2	GAA GCT AGC ACG TGT CAA ATA GTC CAC T	
o1273	Rex1_G221_p927-1_fixer_F	GAA GGG CCC AGA TCT GGT GGC TCC CAC	p929 (Extended PCR SDM)
o1274	Rex1_G221_p927-1_fixer_R	GAA GGG CCC CTG GAA CAG GAC TTC CAG AGA TCT	
o1275	Rex1_D229A_SphI_F	GAA GCA TGC GAA ATG TGT CTT TCC GAA C	p930, p931, p932 (Extended PCR SDM)
o1276	Rex1_D229A_SphI_R	GAA GCA TGC TAG TGC AAA GAT GTG GGA G	
o1289	Rex1_H360A_PflFI_F	CGC GCG GAC TCT GTC GAA GAT GCA AGG	p941 (PCR insert cloning)
o1290	Rex1_H360A_PflFI_R	CGC GAC AGA GTC CGC TTC TCC GTT TTG AAT GCT	
o1291	NotI_GalI_F	GGA GCG GCC GCG TTT TTT CTC CTT GAC GT	p954 (PCR insert cloning)

Table 2.23: List of oligonucleotides used for cloning: o1292-o1349

Oligo:	Name:	Sequence (5′ – 3′):	For cloning plasmid:
o1292	GalI_5'UTR_SpeI_R	GGA ACT AGT AAT TTT CAA AAA TTC TTA CTT TTT	p954 (PCR insert cloning)
o1293	SpeI_zz_F	GGA ACT AGT CAT GGC AGG CCT TGC G	p955 (PCR insert cloning)
o1294	zz_BamHI_R	GGA GGA TCC CGC GTC TAC TTT CGG C	
o1295	PstI_Adh1_3'UTR_F	GGA CTG CAG CAC TTC TAA ATA AGC GAA TTT C	p956 (PCR insert cloning)
o1296	Adh1_3'UTR_ClaI_R	GGA ATC GAT TGC CGG TAG AGG TGT G	
o1299	AgeI_REXO5_F	GGA ACC GGT ATG GAG CCA GAG AGG GAA	p961 (PCR insert cloning)
o1300	REXO5_NsiI_R	GGA ATG CAT CAC GAA CAC AGG CCT GG	
o1330	Rex1_K340A_S342A_NaeI_F	GAA GCG CCG GC TTTAAAATACTTGAGCGAAACC	p980 (Extended PCR SDM)
o1331	Rex1_K340A_S342A_NaeI_R	GAA GC CGG CGC GAAAGGATCGCCAGCTTT	
o1336	NdeI_6xHis_Rex1_M52_F2	GCG CATATG CAC CAT CAT CAC CAT CAC ATG ACA TGC ACA TTG CTA	p962, p963 (PCR insert cloning)
o1345	Rex1_EcoRI_F2	GCG GAA TTC ATG CAA GTA GAA GGG CC	p981, p982, p983, p984 (PCR insert cloning)
o1346	Rex1_P155_F	CCC TAC AAT TCA TTT ATT AAT	p981, p982, p984 (Overlapping PCR)
o1347	Rex1_N458_R	GCG TGA CCT CTC TAA G	p981, p984 (Overlapping PCR)
o1348	Rex1_A473_F	GCG TCC ATG GTT CTT C	p983, p984 (Overlapping PCR)
o1349	Rex1_L498_R	GTC GCC TGT ACC CGA	p984 (Overlapping PCR)

2.2 Methods

2.2.1 Growth and lysis of *E. coli*

2.2.1.1 Generation of competent *E. coli* by the $RbCl_2$ method

500ml LB flask cultures were inoculated with a freshly saturated overnight culture, to a sufficient dilution that 3-4 doublings can take place to reach an OD_{550} of 0.48. Cultures were incubated at 37ºC until an OD_{550} of 0.48 . TfI (30mM KAc, 100mM $RbCl_2$, 10mM $CaCl_2$, 50mM $MnCl_2$, 15% glycerol, pH 5.8 with acetic acid) and TfII (10mM MOPS, 75mM $CaCl_2$, 10mM $RbCl_2$, 15% glycerol, pH 6.5 with KOH) buffers were prepared and filter sterilised using a 0.22μm pore syringe filter. Cells were harvested at an OD_{550} of 0.48 in a pre-chilled 4ºC centrifuge at 4000xg for 5 minutes. Cells were resuspended in 40ml of ice-cold TfI buffer, incubated on ice for 10 minutes, and pelleted in a pre-chilled 4ºC centrifuge at 4000xg for 5 minutes. Pellets were resuspended in 5ml ice-cold TfII buffer. Once resuspended, cells were incubated on ice for at least 30 minutes, typically 30 to 60 minutes. Cells were dispensed into ice-cooled 1.5ml microfuge tubes and snap-frozen in a liquid nitrogen vessel for storage at -80ºC. Competency was measured using 5μl of pUC19 at concentrations of 1ng/μl and 0.1ng/μl to give transformation efficiency in colony forming units per μg of plasmid (CFU/μg).

2.2.1.2 Plasmid transformation of *E. coli*

50µl Aliquots of competent *E. coli* were thawed on ice. Once thawed, no more than 10µl DNA was added to the cells, mixed by flicking the tube. Cells were incubated with DNA on ice for 30 minutes. Cells were heat-shocked at 42ºC in a water bath for 90 seconds, then incubated on ice for 2 minutes. Cells were incubated at 37ºC for 30-60 minutes with 1ml 2xYT. Cells were flash spun for 10s in a microcentrifuge. Cells were resuspended in 100µl of media and aspirated onto selective plates, and spread using 4mm glass roller beads until fully absorbed.

2.2.1.3 Growth of *E. coli* for plasmid DNA extraction

Freshly grown colonies DH5α *E. coli* transformed with plasmid were obtained by transformation, or streaking of a glycerol stock were inoculated into 2-5ml of 2xYT media with antibiotic, and incubated at 37ºC in a shaking incubator overnight to reach saturation. Cells were pelleted 1.2ml at a time into a 1.5ml microfuge tube, and plasmid DNA was extracted using alkaline lysis, a spin column kit, or frozen at -80ºC if needed.

2.2.1.4 Isolation of plasmid DNA from *E. coli* by alkaline lysis

Based on Bimboim and Doly, 1979. Before beginning, Plasmid solutions I (50mM glucose, 25mM Tris pH 8, 10mM EDTA pH 8), II (0.2M NaOH, 1% SDS), and III (3M KAc) were prepared. Cell pellets of saturated DH5α *E. coli* were resuspended in 100µl of Plasmid Solution I. 200µl of Plasmid Solution II was added to each suspension and mixed by inversion 6 times. Mixtures were incubated on ice for 5 minutes. 150µl of Plasmid Solution III was added to each lysate and mixed by inversion 6 times, then incubated on ice for 5 minutes. Precipitated cell debris was pelleted by centrifugation at 13,000xg for 10 minutes. Supernatants were retained in fresh microfuge tubes and placed on ice. 1 volume (typically 450µl-500µl) of ice-cold 25:24:1 phenol (pH 6.5):chloroform:isoamyl alcohol was added to each lysate. Mixtures were agitated on a vortex mixer briefly, then separated by centrifugation at 13,000xg for 5 minutes. Aqueous phases were retained in fresh microfuge tubes, and 2 volumes of ice-cold 100% ethanol was added to each sample for precipitation on ice for 30 minutes, or for 1-2 hours at -80ºC. Precipitant was pelleted 13,000xg for 20 minutes and washed with 200µl of ice-cold 70% ethanol. Pellets were air-dried for at least 10 minutes. Pellets were resuspended in 38µl H_20 per 5ml of culture used. For every 19µl of solution, 1µl of 1mg/ml RNase A was added. 1µl of each plasmid preparation was examined by agarose gel electrophoresis.

2.2.1.5 Isolation of plasmid DNA from *E. coli* using a spin column kit

Column minipreps were performed using the E.Z.N.A Plasmid DNA Mini Kit 1 (Omega, USA; D6942-01), according to the manufacturer's instructions. The elution step was performed in two 20µl 5 minute steps.

2.2.1.6 Growth of *E. coli* in autoinduction media for expression of recombinant protein

Auto-induction of *E. coli* was performed based on Studier, 2005. BL21 (DE3) + pLysS *E. coli* were transformed with a protein expression plasmid. 2-4 colonies were inoculated into 2-5ml 2xYT starter

culture (or several, sufficient for 1% inoculation of the induction culture), containing the corresponding antibiotic for the expression plasmid, and chloramphenicol for the pLysS lysozyme plasmid. The time was recorded, then 1% volume started cultures were added to the induction cultures. Cultures were incubated at 37ºC in a shaking incubator for 3 hours, the absorbance of the culture at 600nm was recorded after 3 hours passed (OD_{600nm}). 2ml cultures were retained at -80ºC serving as a pre-induction sample for SDS-PAGE analysis. At an OD_{600nm} of 0.8 or after 5 hours of incubation had passed, the shaking incubator was shifted to 18ºC for 24 hours. After 24 hours, culture OD_{600nm} was measured and cells were pelleted at 4ºC and 4,800xg, with culture flasks left shaking at 18ºC during centrifuge spins. 1ml culture pellets were retained as a post-induction sample for SDS-PAGE analysis. Pellets were dried, weighed, then stored at -80ºC.

2.2.1.7 Denaturing extraction of whole-cell protein from *E. coli* by alkaline lysis

Denaturing extractions of protein from *E. coli* were performed to measure induction efficiencies using pellets taken pre- and post-induction. To each pellet, an equivalent volume of 0.5mm glass disruption beads (RPI, USA; 9831) was added. Samples were lysed in 1xPLB (5% β-mercaptoethanol, 5% glycerol, 80mM Tris-HCl pH 6.8, 0.05% bromophenol blue, 1% SDS)/40% urea, normalised by OD_{600nm} to give an equivalent cell concentration to the smallest pellet lysed in 10µl. Once resuspended, samples were boiled at 95ºC in heatblock and vortex-mixed every 2 minutes. Samples were centrifuged at 13,000xg for 1 minute. Samples were used immediately for SDS-PAGE analysis, or stored at -80ºC.

2.2.1.8 Native extraction of soluble protein from *E. coli*

1.5ml microfuge tube-scale native extractions were performed in the same manner as denaturing extraction with the following modification: TBST (10mM Tris-HCl pH 6.8, 150mM NaCl, 0.1% Tween-20) was used as lysis buffer, and the sample was not boiled.

Preparative large-scale protein purification based on Stead *et al*., 2007. 1ml of lysis buffer (10mM Tris-HCl pH 6.8, 150mM NaCl, 0.1% Tween-20, 1mM PMSF, 5ng/µl Leupeptin, 7ng/µl Pepstatin A, 1mM DTT, 100µg/µl lysozyme) was added per gram of cell pellet on ice, cells were resuspended and left on ice for 30 minutes. Cells were disrupted by sonication using a 20KHz sonicator 2mm probe (Fisher; FB-505) at 30% amplitude, with 30s on/30s off pulses for a total of 4 minutes, or until the lysate was no longer viscous. Unlysed cells and debris were pelleted at 4ºC and 4,800xg for 10 minutes. Crude extracts were clarified in in 1.5ml microfuge tubes at 4ºC and 20,000xg for 30 minutes. Supernatants were used for purification, or snap-frozen in liquid nitrogen. Before downstream processing, 50µl of each sample was retained and snap-frozen for analysis by SDS-PAGE.

2.2.2 Growth and lysis of *S. cerevisiae*

2.2.2.1 Colony Method for Plasmid transformation of *S. cerevisiae*

Protocol based on Gietz *et al*., 1992. A fresh streak plate was made and allowed to grow until small colonies were seen. A 1µl inoculation loop's worth of small colonies for each strain was mixed into 1ml of sterilised TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8). Cells were pelleted at 13,000xg for 1 minute, and supernatants were discarded. The cell pellet was resuspended in 1ml LiT buffer (10mM Tris-HCl pH 7.5, 1mM EDTA pH 8.0, 100mM LiAc pH 7.5), and pelleted at 13,000xg for 1 minute. For

each transformation, 50µl of cells were added to 1µl of plasmid DNA (typically >150ng/µl) and 5µl of 10µg/µl sonicated herring sperm carrier DNA and incubated at room temperature for 30 minutes. Sterile filtered PEG4000 solution was added to a concentration of roughly 30% followed by vortex mixing and a further 30 minute incubation at room temperature. DMSO was added to each transformant to a final concentration of 10%. Cells were heatshocked at 42ºC for 15 minutes, then pelleted at 13,000xg for 1 minute and washed with TE buffer. Cells were pelleted at 13,000xg for 1 minute and resuspended in 100µl of TE buffer Cells were spread using autoclaved 4mm glass roller beads until absorbed and plates were incubated at 30ºC for 2-3 days.

2.2.2.2 Denaturing extraction of whole-cell protein from *S. cerevisiae* by alkaline lysis

Protocol based on Motley *et al*., 2012. Fresh colonies were inoculated into 2-5ml starter cultures that were incubated at 30ºC for 2-3 days until visibly saturated. 50-500µl of starter cultures were used to inoculate 50ml cultures, which were incubated overnight at 30ºC. Cultures that has reached an OD_{600nm} of 0.5-1 were pelleted at 4,800xg in a pre-chilled centrifuge for 10 minutes, to give 10 OD_{600nm}-worth of cells in triplicate for each culture. Pellets were stored at -80ºC or processed further. Pellets were resuspended in 500µl of freshly prepared ice-cold 0.2M NaOH/0.2% β-mercaptoethanol, then transferred to 1.5ml microfuge tubes and incubated on ice for 10 minutes. Trichloroacetic acid (TCA) was added to give a final concentration of 5%. Samples were precipitated on ice for 5 minutes, then pelleted at 13,000xg for 5 minutes. Protein pellets were resuspended in 10µl of 1.5M Tris pH 8.7. To the resuspended pellets, 100µl of 1x PLB/40% urea were added, then boiled at 95ºC for 10 minutes. If the bromophenol blue indicator turned green or yellow at this stage, concentrated base such as 10M NaOH was added drop-wise until colour returned to blue. Samples were clarified by centrifugation at 13,000xg for 1 minute. Samples were used immediately for SDS-PAGE analysis, or stored at -80ºC.

2.2.2.3 *In vivo* crosslinking of *S. cerevisiae*

This protocol is described in Daniels *et al*., 2022. 1% freshly saturated starter cultures were added to 1L selective SD media in 3L flasks, and incubated overnight at 30ºC. Upon reaching an OD_{600nm} of 1.0, each 1L culture was pelleted into 4x 50ml centrifuge tubes at 4,800xg for 5 minutes at 4ºC. 100ml of culture supernatant was retained and used for downstream washes of cells. Each pellet was resuspended thoroughly in 6ml spare media, and pipetted onto a 10cm Petri dish to give a minimal surface depth of roughly 1mm. Four Petri dishes were cooled on a 20cm x 20cm glass plate capping a vessel of ice water. Lidless Petri dishes were crosslinked using a CL1000 Crosslinker (UVP Plastics) with 12000J of UV 30s on/ 30s off over 10 minutes. Cells were pooled after crosslinking into a single tube. Each plate was washed with 10ml of spare media and collected into a second tube. Both suspended cells and wash supernatants were pelleted, pooled into a single tube, and stored at -80ºC.

2.2.2.4 Purification, protease cleavage, RNase A treatment, and radio-labelling of crosslinked Rex1 from *S. cerevisiae*

This protocol is described in Daniels *et al*., 2022. Pellets were thawed on ice and lysed with 1 volume 0.5mm glass disruption beads (RPI, USA; 9831) and 1ml of lysis buffer (50mM HEPES pH 7.6, 150mM NaCl, 1mM EDTA, 100µM PMSF, 1x yeast protease inhibitor cocktail). Pellets were vortexed for 30s

/30s on ice for 5 minutes total vortexing. Unlysed cells were pelleted at 4,800xg at at 4ºC for 10 minutes. Beads werere-extracted with a further 1ml of lysis buffer. IgG beads (20-50µl per pulldown; GE Healthcare, Sweden; 17-0969-01, lot 10253617) were washed three times with wash buffer (50mM HEPES pH 7.6, 150mM NaCl). 1ml of clarified lysate was added to 20-50µl of IgG sepharose beads in a 1.5ml microfuge tube. Beads were incubated on a rotating wheel at 4ºC for 1 hour. Beads were washed three times with wash buffer, three times with stringent wash buffer (wash buffer mixed 1:1 with 4M $MgCl_2$), then 3x with 1x PNK buffer (50mM HEPES pH 7.6, 10mM $MgCl_2$, 1mM DTT). To washed beads, 50µl of 1xPNK buffer and 1µl of 1mg/ml RNase A was added and incubated at 37ºC for 30 minutes. Beads were washed three times with 1x PNK buffer before labelling in 50µl of PNK buffer, 5U polynucleotide kinase (NEB; M0201S), and 3 pmol $\gamma[^{32}P]$-ATP (PerkinElmer) at 37ºC for 30 minutes. Beads were washed 3x with 50mM HEPES pH 7.6 before digestion with PreScission protease in the same buffer overnight at 4ºC. The supernatant was retained, along with two 70µl bead washes pooled to give a 210µl 'PsP' fraction. The beads were then washed three times with 70µl 0.5M acetic acid to give a 210µl 'Ac' fraction. To each fraction, 4 volumes of butanol were added for overnight precipitation at -80ºC. Protein pellets were dried, and resuspended in 40% urea/1xPLB for SDS-PAGE analysis.

2.2.2.5 Growth of *S. cerevisiae* for plasmid shuffle assay

Sterile 96-well plates were used to prepare a 5-well 10-fold serial dilution with 200µl of diluted culture in each well to give 6 wells per transformant. To the first well of each row, 220µl of saturated starter culture was added, normalised for cell number using the OD_{600nm}. Fresh sterile tips were used for each serial dilution step. The pinning stamp was washed and sterilised in Petri dishes by gently shaking: for 90s in 10% bleach, 30s each in three dishes of ddH_20, then 30s in 100% ethanol. Ethanol was removed by flaming over a Bunsen burner. Once dry, the pinning stamp was stamped onto three SD + 0.1%(w/v) 5-FOA plates, and three SD-Leu or SD-His plates. Plates were incubated at 30ºC, and imaged after 4-5 days. To account for temperature-sensitive mutants, sets of plates were also incubated at 25ºC.

2.2.2.6 *S. cerevisiae* growth curve assay

Growth rates of FOA isolates were measured by tracking the OD_{600nm} of 50ml SD flask cultures. Cultures were inoculated with 1% of saturated starter cultures and incubated overnight. Culture optical densities were tracked, and cultures were diluted 1/10 with fresh media if higher than an OD_{600nm} of 1.0, and remeasured. All strains were processed in parallel, with a full replicate prepared the following day using the same precultures. Over both days, each strain was tracked over 10 hours in logarithmic phase (<0.1 OD_{600nm}) at least once. Growth curves were normalised against the first OD_{600nm} and Log_2 - transformed using Microsoft Excel, and plotted using GraphPad Prism 10.0.2.

2.2.2.7 Guanidinium/phenol RNA extraction from *S. cerevisiae*

0.1-1% of starter culture were used to inoculate 50ml flask cultures that were incubated at 30ºC overnight. Cells were harvested at 0.5 OD_{600nm} at 4000xg and 4ºC for 5 minutes. Pellets were stored at -80ºC or proceeded to lysis. Pellets were thawed on ice and resuspended in 500µl GTC mix. 500µl of pH 4 phenol and 1ml of DEPC-treated glass disruption beads were added and vortex-mixed for 5 minutes. 3.5ml GTC mix and 3.5ml pH 4 phenol were added and mixed before incubation at 65ºC in a water bath for

10 minutes. 4ml of chloroform and 2ml of NaAc buffer were added, mixed, then centrifuged at 4,000xg for 5 minutes at 4ºC. Aqueous phases were retained and mixed with an equal volume of 25:24:1 pH4 phenol: chloroform: isoamyl alcohol before centrifugation at 4,000xg and 4ºC for 5 minutes. Aqueous phases were retained and mixed with 2 volumes of 100% ethanol for precipitation at -80ºC for 1-2 hours. Pellets were washed twice with 70% ethanol, and air dried for 10 minutes. Pellets were resuspended in DEPC-treated ddH_20, with 20µl/OD_{600nm} giving roughly 2µg/µl as measured by OD_{260nm}. RNA was stored at -80ºC.

2.2.2.8 *S. cerevisiae* Colony PCR - NaOH method

Fresh streak plates were made and allowed to grow until small colonies were seen. 1µl inoculation loops were used to pick a small individual colonies that were each resuspended in 50µl of fresh 20mM NaOH. Cell suspensions were heated at 95ºC for 10 minutes then pelleted at 13,000xg for 10 minutes. 2µl of each supernatant was used for PCR.

2.2.3 Gel electrophoresis and blotting

2.2.3.1 Agarose gel analysis of plasmid and PCR product DNA

Agarose gel electrophoresis was performed using 1% agarose dissolved in 0.5x TBE (45mM Tris, 45mM boric acid, 1mM EDTA), with 4% ethidium bromide added after microwaving. Gels were run at 120V until the bromophenol blue reached 2/3 the length of the gel. For each gel, 0.5µg of GeneRuler 1 kb DNA Ladder was included (Thermo Scientific; SM0311)

2.2.3.2 Poly-acrylamide gel analysis of protein by SDS-PAGE

Protocol based on Shapiro *et al*., 1967. Gels were cast using 10cm x 10cm glass plates featuring 1.5mm integrated glass spacers, and sealed at the base with 1% agarose in ddH_20. 10ml of 12% resolving solution per gel (12% 38:1 Protogel acrylamide, 42mM Tris-HCl pH 8.7, 0.1% SDS, 0.1% ammonium persulphate, 0.001% TEMED) and 3ml of stacking solution per gel (4.67% 38:1 Protogel acrylamide, 117mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulphate) were prepared fresh. TEMED was added to resolving gel solutions for casting, to which a layer of isopropanol was added. Isopropanol was washed off, stacking gels were polymerised with TEMED, cast, and the 1.5mm combs were inserted. 12% gels were used for all data presented in this work. 2 gels were assembled into each gel running tank (ATTO, Japan) containing TGS running buffer (25mM Tris, 198mM glycine, 0.1% SDS), samples were boiled until condensation was seen and spun down before loading. Protein ladders were loaded according to downstream analysis: for Coomassie staining, 2.5µl of Precision Plus Protein Unstained Standards were used (Biorad; 1610363); for Western analysis, 2.5µl of Precision Plus Protein All Blue Standards were used (Biorad; 1610373). Gels were initially run at 75V until the bromophenol blue reached the stacking interfaces, after which voltage was gradually increased to 130V until the bromophenol blue reached the base of the gels. At this stage gels were either stained with Coomassie stain (Insta Blue colloidal stain, APExBIO; B8226), or progressed to membrane transfer and Western blotting.

2.2.3.3 Membrane transfer and Western blotting of proteins resolved by SDS-PAGE

Protocol based on Burnette, 1981. Transfer stack components were soaked in transfer buffer (12.5mM Tris, 96mM glycine, 20% methanol, 0.1% SDS) and added to the positive plate of a Pierce fast semi-dry blotter in order: 3x 10cmx10cm squares Whatmann filter paper, 1x 10cmx10cm Protran 0.45μm nitrocellulose membrane (Amersham; 10600003), the acrylamide gel, three 10cmx10cm squares of Whatmann filter paper. Protein was transferred at 10V for 2 hours. Membranes were washed with TBST (10mM Tris pH 7.5, 150mM NaCl, 0.1% Tween-20), and incubated with 10-fold diluted Ponceau S solution (Sigma; P7170). If visible, stained protein was imaged and used to guide cutting of membranes. All TBST washes were preformed by manual shaking for 30s. Membranes were washed three times with TBST, then blocked with 5% powdered milk in TBST at room temperature for 1 hour. Membranes were washed three times with TBST, and antibodies were added to a 1/10,000 dilution in TBST for 1-2 hours. Unbound antibody was removed by three TBST washes. If required, membranes were incubated with secondary antibody diluted in TBST with three further washes. Freshly mixed ECL solution was made by 1:1 mixing of ECL Solution I (2.5mM luminol, 400μM p-coumaric acid, 100mM Tris-HCl pH 8.7) and 5ml ECL Solution II (5.4mM H_2O_2, 100mM Tris-HCl pH 8.7) and incubated with membranes for 1 minute before imaging by chemiluminescence using a SynGene (Cambridge, UK) G:Box Geldoc.

2.2.3.4 Poly-acrylamide gel electrophoresis (PAGE) analysis of whole-cell RNA extracts

60ml RNA acrylamide gel solutions (8% Accugel 19:1 acrylamide, 45mM Tris, 45mM boric acid, 1mM EDTA, 50% urea, 0.07% ammonium persulphate, 0.007% TEMED) were prepared. 20cmx20cm glass plates with 1.5mm plastic spacers were clamped and cast horizontally. Once gels were set, combs were removed under water or buffer and wells were washed with water. Gels were assembled into PAGE units (custom order; Engineering and Design Plastics Ltd., Cambridge, UK) filled with 0.5x TBE (45mM Tris, 45mM boric acid, 1mM EDTA) running buffer. Gels were pre-warmed 60V for at least 30 minute prior to loading. 5μg of each RNA sample was prepared in 2μl of DEPC-ddH_20, diluted 1:1 with 2xFormamide loading buffer (5% formamide 20mM EDTA, 0.05% bromophenol blue 0.05% xylene cyanol), and boiled at 65ºC for five minutes. Before samples were loaded, each well was washed through pipetting with buffer. Gels were run overnight at 60V, and voltage was increased to 160V the following day, until the xylene cyanol reaches the bottom. For 'Twice run' gels, at this point an empty lane was loaded with xylene cyanol, and allowed to run the same distance again. Gels were stained in 0.5x TBE with 2% ethidium bromide for 15 minutes on an orbital shaker and imaged using a UV transluminator. Gels then progressed to transfer for Northern blotting.

2.2.3.5 Membrane transfer and Northern blotting of RNA resolved by PAGE

Protocol based on Alwine *et al*., 1977. Transfer stack components were washed in 0.5x TBE (45mM Tris, 45mM boric acid, 1mM EDTA) and added to the positive plate of a Pierce fast semi-dry blotter, in order: 3x 20cmx20cm squares Whatmann filter paper, 1x 20cmx20cm positively charged 0.45μm nylon transfer membrane (Amersham; 10600003), the acrylamide gel, three 20cm x 20cm squares of Whatmann filter paper. Nucleic acids were transferred at 10V for 2 hours. Membranes were dried and crosslinked with 1200J UV-C using a CL1000 Crosslinker (UVP Plastics). Membranes were blocked in oligo hybridisation buffer (0.5% SDS, 2xSSPE; 300mM NaCl, 18mM NaH_2PO_4, 2mM EDTA, pH 7.4

with NaOH; 5x Denhardt's; 0.2% ficoll, 0.2% polyvinylpyrrolidone, 0.2% BSA) for 1 hour at 37ºC. 20µl labelling reactions for oligonucleotide probes were prepared in DEPC-treated ddH_20 (oligonucleotide probe 5pmol, 2µl 20x PNK buffer, 5U polynucleotide kinase; New England Biolabs; M0201S, 3pmol γ-[^{32}P]-ATP ; PerkinElmer) and incubated at 37ºC for 30 minutes. Labelling reactions were diluted to 1ml with oligo hybridisation buffer and inactivated at 65ºC for 5 minutes, before direct addition to hybridisation solution by syringing through a 0.22µm filter. Blots were hybridised at 37ºC for 4-16 hours. Membranes were washed with 6x SSPE buffer (900mM NaCl, 54mM NaH_2PO_4, 6mM EDTA, pH 7.4 with NaOH) three times. An additional final wash was performed in 6x SSPE buffer at 37ºC for 15 minutes. Membranes were dried, coated in Saran wrap, and used for radiography. Radiography was performed using a 20cm x 25cm phosphor Imaging Screen K (Bio-Rad; 1707843), placed white-side down over Saran-wrapped membrane inside a BioMax MS Intensifying Screen (Kodak) for 4 hours to several days. Phosphor imaging screens wereimaged using a Typhoon FLA-7000 scanner (GE healthcare, Sweden). Northern blots were stripped with boiling stripping buffer (0.1x SSPE and 0.1% SDS) on an orbital shaker for 15 minutes before re-probing.

2.2.4 Cloning of plasmids and recombinant DNA techniques

2.2.4.1 Restriction Digest

All restriction enzymes were ordered from New England Biolabs, using HF variants where possible. Diagnostic digests were made as 20µl reactions, preparative digests were made as 50-100µl reactions. All digests were conducted in PCR tubes using a PCR machine with a heated lid. Glycerol content was never permitted to exceed 5% to avoid Star activity (Barany, 1988 and references therein).

2.2.4.2 DNA gel purification using using a spin column kit

DNA bands were identified using a low-power UV hand lamp, and excised with a razor blade. Gel slices were processed using the E.Z.N.A Gel Extraction Kit (Omega; D2500-01), in accordance with the manufacturer's instructions. The elution step was performed twice with 20ul of ddH_20, incubated 5 minutes before spinning. 20µl elutions were retained separately. Gel purification was performed after preparative digestion for cloning, to give purified DNA for ligation.

2.2.4.3 DNA buffer exchange using a spin column kit (PCR purification)

Buffer exchange for DNA was performed using E.Z.N.A Cycle Pure Kit (Omega; D6492-01) according to manufacturer's instructions. Buffer exchange was performed on PCR reactions prior to digestion, overlapping PCR halves, or if incompatible reaction buffers meant restriction digest needed to be performed in two steps.

2.2.4.4 Klenow Reaction

Klenow reactions were used to fill in 5´ overhangs to produce blunt ends when a restriction site required removal. To 50µl restriction digestion mixtures, 5µl of 180µM dNTPs and 5U Klenow fragment were added. Reaction mixtures were incubated at 25ºC for 15 minutes. Klenow fragment was inactivated at 65ºC for 10 minutes. DNA was gel purified before ligation.

2.2.4.5 T4 Reaction

T4 reactions were used to remove 3´ overhangs to produce blunt ends when a restriction site required removal. 3 units of T4 DNA polymerase were added to a 50µl restriction digested mixture, and incubated at 37ºC for 2 minutes. dNTPs were added to a final concentration of 120µM, and polymerase reactions were incubated at 12ºC for 15 minutes. Reactions were stopped by addition of EDTA to a concentration of 10mM, followed by heat inactivation at 75ºC for 20 minutes. Products were purified by gel purification for use in monomolecular ligation.

2.2.4.6 Polynucleotide kinase treatment of oligonucleotides for linker insertion cloning

Oligonucleotides were treated with polynucleotide kinase to produce 5´ ends compatible with restriction digested and de-phosphorylated plasmid. Labelling reactions consisted of 25pmol/µl oligonucleotide, 5U polynucleotide kinase (New England Biolabs; M0201S), 1x DNA ligase buffer (New England Biolabs: 50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 1mM ATP, 10mM DTT). Oligonucleotides were labelled at 37ºC for 30 minutes. Polynucleotide kinase was inactivated at 65ºC for 20 minutes. For linker insertion cloning, each oligonucleotide was first labelled in separate 20µl reactions, then mixed 1:1 after heat inactivation. 100µl annealing reactions were prepared, containing the 40µl oligo mix, 300mM NaCl, 10mM Tris-HCl pH 7.5, and 2mM EDTA pH 8.0. Annealing reactions were heated and cooled in a PCR machine with a heated lid, from 95ºC to 25ºC in 5ºC/30s steps. Once annealed, linkers were ready for ligation with restriction digested and dephosphorylated plasmid.

2.2.4.7 Q5 PCR

Preparative PCRs were performed using Q5 polymerase (New England Biolabs; M0491S). With the exception of *REXO5*, all PCR amplifications used plasmid as template. Reactions mixtures consisted of 1x Q5 buffer, 1U Q5 polymerase, 200µM dNTPs, 2pmol of each primer, and 1ng plasmid. The following cycle conditions were used: 2 minutes initial denaturation, followed by thirty cycles of 30s 95ºC denaturation, 1 minute 50ºC annealing, and 1 minute/kb 72ºC extension.

2.2.4.8 Whole-plasmid Site-directed mutagenesis PCR

For most site-directed mutagenesis PCRs, diverging primers were designed from the mutation site. Induced mismatches typically spanned 1-4nt, which were positioned at the 5´ of the primer. Mutagenesis was designed to introduce a unique restriction site at the missense mutation, or as a silent mutation in nearby codons. This restriction site was used for digestion of the plasmid-sized PCR product, and for downstream screening of ligation products. Dpn1 was included with digestion reactions to remove plasmid template.

2.2.4.9 Overlapping PCR

Overlapping PCR was used to combine mutations in the same open reading frame. Primers were designed adjacent to mutation sites, to maximise the sequence overlaps between PCR products. Once overlapping PCR products had been separately amplified, DNA buffer exchange was used to isolate PCR products. Fusion reactions were prepared consisting of the same components of a Q5 PCR, but

lacking primers and including both overlapping PCR products. Fusions products that were sufficiently concentrated were used directly for restriction cloning, more dilute fusion products or reactions resulting in multiple bands were purified by gel purification, and used as templates in a standard PCR.

2.2.4.10 Ligation

Ligation reactions were prepared as 20µl reactions containing 1x T4 ligase buffer (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 1mM ATP, 10mM DTT), 200U T4 ligase (NEB; M0202S). Insert and vector DNA that had been digested with complementary restriction enzymes were gel purified, and included in a 3:1 molar ratio. Ligations were incubated overnight at 15ºC in a PCR machine, with an additional 200U of T4 ligase added the following morning. 10µl of ligation reactions were used to transform competent DH5α *E. coli*.

2.2.4.11 Phosphatase treatment of digested DNA

Phosphatase treatment was used to treat restriction digested vector DNA to prevent re-annealing and re-ligation in cases where inserts were being cloned into single restriction sites. Phosphatase reaction mixtures were prepared by addition of 1U of shrimp alkaline phosphatase (NEB; M0371S) and incubated at 37ºC for 30 minutes. Phosphatase was inactivated at 65ºC for 10 minutes.

2.2.5 Purification of recombinant protein from native *E. coli* extracts

1-step purification of GST-tagged proteins from clarified native *E. coli* extracts from 100ml auto-induction cultures were performed as follows: 1ml of clarified lysates were incubated with 20µl of glutathione sepharose beads ('4B'; Lots 1017095, 10324511; GE Healthcare, Sweden) at 4ºC for 2 hours. Beads were washed three times with TBST (10mM Tris-HCl pH 7.5, NaCl 150mM, 0.1% Tween-20). Beads were eluted in 5x bead volume of TBST containing 25mM glutathione (pH 7), split across three 10 minute elutions at 30ºC. Elutions were snap-frozen and stored at -80ºC

2-step purification of GST proteins from clarified native *E. coli* extracts from 500ml auto-induction cultures were performed as follows: 6ml of clarified lysates were concentrated through binding to 500µl of heparin sulphate beads ('6 Fast Flow'; Lot 10265932; GE Healthcare, Sweden), and eluted using heparin elution buffer (10mM Tris-HCl pH 7.5, 500mM NaCl, 0.1% Tween-20). These enriched lysates were purified using glutathione beads as described for the 1-step purification, with TBST substituted for heparin wash buffer for washes.

Crystallisation-scale purifications were prepared by Svetlana Sedelnikova of the University of Sheffield Protein Purification Facility as follows: roughly 40ml clarified lysates of 2L auto-induction cultures were fractionated using a 25ml Heparin-Sepharose column (GE Healthcare, Sweden) with an AKTA purifier running at 5ml/min with a gradient of 0mM-700mM NaCl-containing 50mM HEPES pH 7.4 buffer, giving 4ml fractions. Three fractions judged by SDS-PAGE analysis and Coomassie staining to contain the most protein at the correct size were combined. Fractions were either concentrated to 2ml using 2.8M ammonium sulphate precipitation, or used directly as 5ml use in gel filtration by 1.6/60 HiLoad Superdex200 column (GE Healthcare, Sweden), run at 1.5ml/min to give 3ml fractions. Fractions were analysed by SDS-PAGE analysis, and the purest fractions were combined, concentrated by precipitation with 2.8M ammonium sulphate or Vivaspin 300K column (Sartorius, Germany), and lastly buffer

exchanged into 50mM HEPES and 50mM NaCl using a Zeba column (Thermo Fisher, U.S.A).

2.2.6 Preparation of crystallisation screens

Purified protein was centrifuged to remove aggregates, then dispensed into 96-well sitting 3-drop vapour diffusion crystallisation trays using a Mosquito crystallisation robot (SPT Labtech, Melbourn), with 100nl protein +100nl of buffer per sitting drop. The following crystallisation screens were used, each from Molecular Dimensions (Rotherham, UK): JCSG, PACT, $AmSO_4$, and ProPlex. An additional two custom crystallisation screens were prepared using a Formulatrix (U.A.E) Formulator, illustrated in Figure 4.15.

2.2.7 *In vitro* exonuclease assays

2.2.7.1 Radiolabelling and gel purification of DNA oligonucleotides

Oligonucleotides were labelled in 20µl labelling reactions: (oligonucleotide probe 5pmol, 2µl 20x PNK buffer, 5U polynucleotide kinase; New England Biolabs; M0201S; 3pmol γ-[^{32}P]-ATP ; PerkinElmer) and incubated at 37ºC for 30 minutes. Labelling reactions were diluted to 1ml with oligo hybridisation buffer and inactivated at 65ºC for 5 minutes. DNA was precipitated overnight at -20ºC with 2 volumes 100% ethanol, 210mM NaAc, and 10µg glycogen. Pelleted substrates were resuspended in formamide loading buffer (95% formamide 20mM EDTA, 0.05% bromophenol blue 0.05% xylene cyanol) and resolved using 10cmx10cm polyacrylamide gel (16% 19:1 acrylamide, 0.5x TBE, 50% urea, 0.07% ammonium persulphate, 0.007% TEMED) using 0.5x TBE (45mM Tris, 45mM boric acid, 1mM EDTA) as running buffer. Bands were identified using UV shadowing and excised using a razor blade. Labelled oligonucleotides were extracted from gel slices by incubation with 100µl DEPC-H_20 for 1 hour, repeated four times. Pooled oligonucleotides were precipitated overnight as previously, then resuspended in DEPC-H_20. To account for variable recovery efficiency, labelled substrate was spiked 1/100 with 100nM unlabelled substrate.

2.2.7.2 Exonuclease time course assays

Assays typically consisted of 25.3nM protein thawed at 30ºC for 5 minutes in 28µl, with buffer contents varying from a standard 10mM Tris pH 6.8, 150mM NaCl, 0.1% Tween-20, 1.5mM $MgCl_2$ and 1.7mM glutathione, before the addition of substrate to a final concentration of 3.3nM. Assays were incubated in a PCR machine using a heated lid. 4µl time points were taken, and stopped with the addition of 1 volume formamide loading buffer (95% formamide 20mM EDTA, 0.05% xylene cyanol). Time point samples were resolved using 10cmx10cm polyacrylamide gel (16% 19:1 acrylamide, 0.5x TBE, 50% urea, 0.07% ammonium persulphate, 0.007% TEMED) using 0.5x TBE (45mM Tris, 45mM boric acid, 1mM EDTA) as running buffer. Radiolabelled substrates were transferred to positively charged 0.45µm nylon transfer membrane (Amersham; 10600003) for phosphoimaging as described in section 2.2.3.5, fluorescent-labelled substrates were imaged directly from gels without removing the glass plates.

2.2.8 Bioinformatic analysis

2.2.8.1 Image quantification using ImageJ

All quantification was performed using ImageJ v1.53k (Schneider, Rasband and Eliceiri, 2012). Bands were quantified using a rectangle of consistent dimensions with Analyse>Measure. Mean pixel intensities were normalised by subtracting the minimum intensity value to remove background, and account for variable background brightness. For DNA quantification, concentrations were calculated by measuring the intensities of the ladder bands, for which concentrations were known. For Western blot protein band quantification, only raw chemiluminescence images were used. The normalised average pixel intensity for each sample was divided by the normalised signal intensity for that sample's loading control. For quantifications performed in triplicate, averages of the loading-normalised average pixel intensity with standard deviations or mean absolute deviations were calculated using Microsoft Excel.

2.2.8.2 Image quantification using SAFA

All Rex1 trimming assays were quantified using SAFA (Semi-Automated Footprinting Analysis) v1.1b (Das *et al*., 2005). For each time course gel image, SAFA was first used to correct for uneven band migration and to assign band identities. SAFA next automatically fitted and integrated Lorentzians by least squares fitting giving a list peak area for each oligo length measured in a given gel lane. Matrices of oligo length against lane number were processed using Microsoft Excel to give an average oligo length for each time point. Background signal was filtered using the T_0 time point, measured as the fold difference between the highest non full-length oligo value and the total lane signal. Once filtered, each remaining value was normalised through division by the remaining total lane signal. These normalised values were multiplied by their oligo length, allowing the calculation of an average oligo length for each time point. Average oligo lengths were used to calculate trimming rates by dividing the nucleotides removed between two time points by the time difference. Time course graphs were plotted using GraphPad Prism 10.0.2.

2.2.8.3 Homology search using DELTA-BLAST

The full-length FASTA sequence of Rex1 was used as a search term with DELTA-BLAST (Domain enhanced lookup time accelerated-Basic Local Alignment Search Tool; Boratyn *et al*., 2012) to search the non-redundant protein database for the following organisms: *S.cerevisiae*, *S. pombe*, *C. elegans*, *D. melanogaster*, *M. musculus*, *H. sapiens*. A further DELTA-BLAST search was performed in the same manner for *H. sapiens* PARN.

2.2.8.4 Sequence alignment using Clustalo and Jalview

FASTA sequences identified by DELTA-BLAST for both Rex1 and PARN searches were combined with sequences for RNase T and oligoribonuclease from *E. coli* and *P. aeruginosa* as a primary input for alignment using EMBL-EBI's Clustal-omega tool (Sievers *et al*., 2011). All resulting alignments were visualised and trimmed using Jalview v2.11.2.7 (Waterhouse *et al*., 2009) with default settings. The initial clustalo alignment revealed subgroups of misaligned EXO motifs; this alignment was trimmed to include all DEDDh domain boundaries defined as 10aa preceding the EXOI motif and 10nt following

the EXOIII motif, and this trimmed alignment was used for a repeated clustalo alignment with default setting, but with 'DEALIGN INPUT SEQUENCES' set to 'yes'. The resulting alignment was used to produce consensus logos using Clustalo with settings: 'Annotations>Autocalculated Annotations' set to 'show consensus logo' and 'normalise consensus logo', but with 'show consensus histogram' disabled.

2.2.8.5 Phylogenetic tree building using Simplephylogeny and display using Interactive Tree of Life

EMBL-EBI's Simple Phylogeny tool (Madeira *et al*., 2022) was used with default settings to build a sequence distance tree of the Clustalo trimmed alignment of DEDDh domain sequences. The resulting text file was uploaded to the Interactive Tree of Life tool (iTOL; itol.embl.de; Letunic and Bork, 2019). The tree was displayed using 'unrooted' mode, but with *E. coli* RNase T selected as the root. Branches were coloured manually within the iTOL interface with clade boundaries illustrated manually using Microsoft PowerPoint.

2.2.8.6 Analysis of structure with UCSF ChimeraX

All structural analysis presented was produced using UCSF ChimeraX 1.2.5 (Pettersen *et al*., 2021). Models were exported without backgrounds using the command 'save browse format .tiff transparentBackground true'.

2.2.8.7 Structural threading

Structural threading of Rex1 with the RNA-bound Pan2 structure PDB 6R9M (Tang *et al*., 2019) was performed using the 1-1 threading option for the Phyre2 web tool (Kelley *et al*., 2015). All other structural threading was performed using the pairwise alignment web tool available on the protein databank (https://www.rcsb.org/alignment; accessed 10/09/2023), using the jFATCAT(rigid) algorithm (Li *et al*., 2020).

Chapter 3

Bioinformatic Analysis of Rex1 and the DEDDh Superfamily

3.1 Introduction

Members of the DEDD exonuclease family are readily identified by sequence homology, which allows the correlation of conserved sequence to identify prospective mechanisms. The highest level of homology is seen within three 'EXO' motifs that surround the metal ion-binding aspartates and glutamate (Zuo and Deutscher, 2001), shown in Figure 1.2. This homology extends to a broader conserved fold shared with the RNase H superfamily that consists of a twisted β-sheet with the second β-strand running antiparallel, interspaced with a minimum of 6 α-helices. This fold was first observed in a crystal structure of the proofreading module of *E. coli* DNA polymerase III (also known as DnaQ; McHenry, 1985), and has since been observed in a wide range of exonucleases that target DNA and/or RNA (reviewed in Yang, 2011). The main feature that distinguishes DEDD enzymes within the RNase H family is that the metal-chelating active site DEDD residues occur in conserved positions: the D and E in EXOI are found on β1, the D of EXOII is found on the α-helix between β4 and β5, with the h/y-D EXOIII motif found on a flexible loop and α-helix respectively at the C-terminus of β5 as outlined in Figure 3.1. By comparison, the metal-chelating residues in other RNase H homologues are found in variable positions throughout the fold.

Extensive structural biology studies of DEDD exonucleases have been undertaken to date featuring the active site in complex with nucleic acid substrates. As of September 2023, the Protein Data Bank contains 72 substrate-bound structures for the RNase T/DEDDh InterPro family (IPR013520), and 150 substrate-bound structures for the RNase D/DEDDy interpro family (IPR002562; Jones *et al*., 2014; Paysan-Lafosse *et al*., 2023). These studies reveal that the DEDD motif is accompanied by common structural features within the exonuclease domain that assist with orientation of the substrate in the active site, summarised in Table 3.1. For example, an aromatic residue is commonly observed to stack with the 3´ terminal base of a substrate, corresponding to Y272 in yeast Rex1, which is proposed to serve as a physical barrier that prevents the substrate from progressing further into the binding cleft (Hsiao *et al*., 2011; Chen *et al*., 2018; Tang *et al*., 2019; Schäfer *et al*., 2019). A further example is the 3´ terminal nucleotide 2´ OH interaction, thought to be responsible for a preference for RNA over DNA, which is mediated by a mainchain interaction with the peptide backbone C-terminally adjacent to

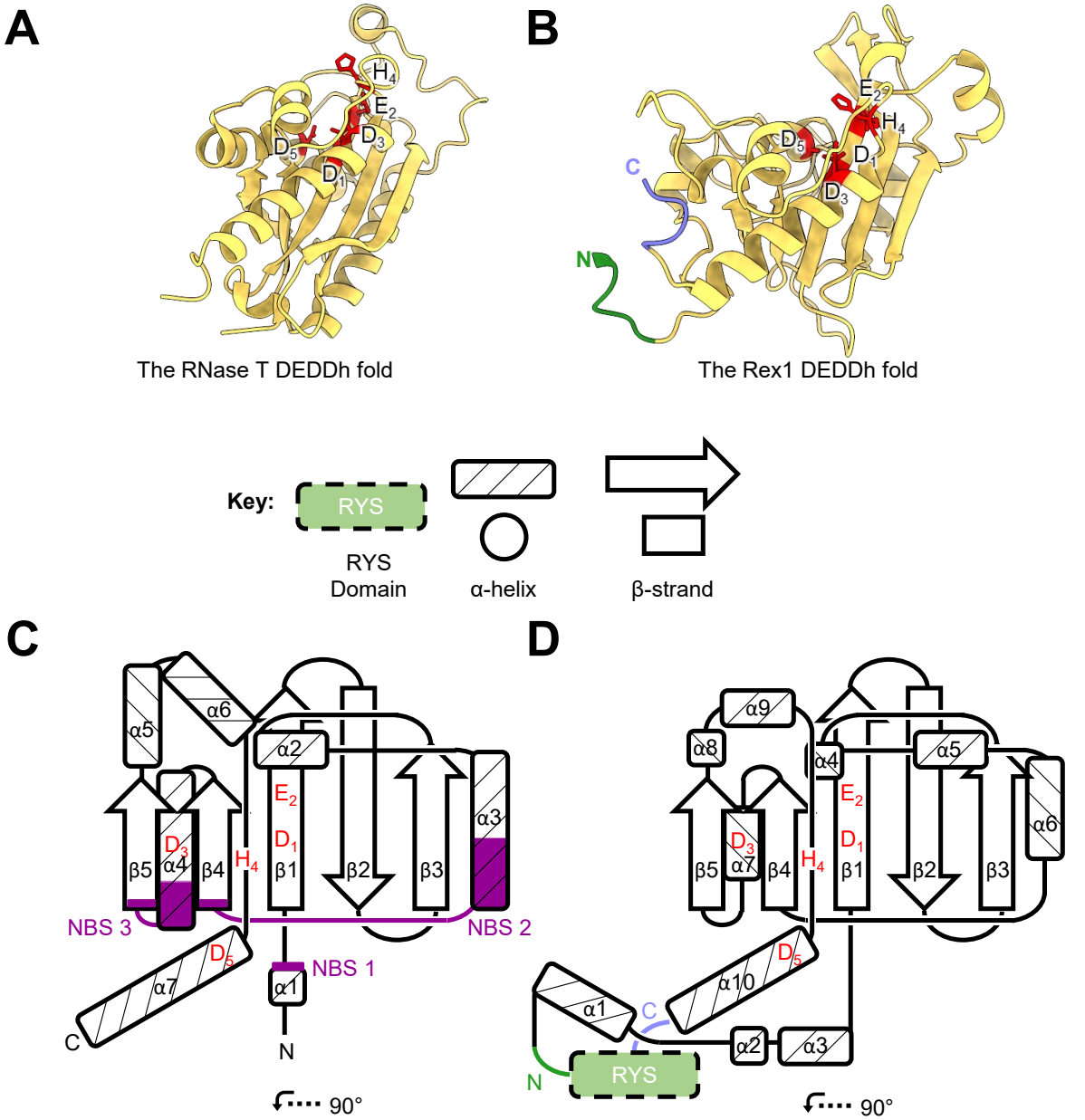

A
B
The RNase T DEDDh fold
The Rex1 DEDDh fold
Key:
RYS
RYS Domain
α-helix
β-strand
C
D
NBS 3
NBS 2
NBS 1
90°
E
F
RYS

Figure 3.1: The conserved topology of the DEDD folds in RNase T and Rex1
Structural comparison a crystal structure of RNase T (**A**; PDB entry 2IS3; Zuo *et al*., 2007) and the Alphafold DB entry for Rex1 (**B**; AF DB entry P53331; Jumper *et al*., 2021; Varadi *et al*., 2022). Red residues represent the DEDDh catalytic site, the green and purple sequence of Rex1 represent the N- and C- terminal flanking regions that compose the 'RYS' domain. **C,D**:Topological diagrams for the front views of RNase T and Rex1. DEDDh residues are indicated as red letters, numbered by the order they occur. The α-helices and β-strands are depicted roughly to scale, with positions represented as simplification of the exonuclease domain as a flat surface. Loop structure lengths not drawn to scale, but topology is preserved. **E,F**: Topological diagrams for the top views of RNase T and Rex1. DEDDh residues are indicated as red letters, numbered by the order they occur. The α-helices and β-strands are simplified as being upright, but the position of α-helices relative to the β-sheets are preserved. Loop structure lengths not drawn to scale, but orientation to either side of the β-sheets are preserved.

the EXOI motif and for some DEDDh exonucleases an additional side chain interaction (Horio *et al*., 2004).

The mechanisms of DEDD exonucleases can consist of features within and beyond the exonuclease domain. Accessory binding motifs and domains can contribute to a nuclease active site from the protein itself or a binding partner. In the case of *E. coli* RNase T, the non-scissile strand of the duplex nucleic acid substrate is coordinated in the active site by a positively charged Nucleotide Binding Sequence (NBS) found on the other subunit of the homodimer (Zuo and Deutscher, 2002c). This NBS confers RNase T its unique ability to trim 3´ nucleic acid ends found in close proximity to stemloop structures for example as found in 5S rRNA and tRNA, a mechanism that is functionally conserved in *S. cerevisiae* Rex1.

The RNase D homologue Rrp6 recognises its substrates through domain contributions within the protein and from bound proteins. Within the Rrp6 protein, a tandem pair of HRDC domains are proposed to funnel substrates towards the active site as seen for *E. coli* RNase D (Zuo *et al*., 2005), which can be observed in substrate bound structural models of Rrp6 in complex with the exosome complex (Makino *et al*., 2015). The Rrp6 binding partner protein Rrp47 not only provides mutual stability and an anchor to the exosome complex (Feigenbutz *et al*., 2013a,b), but also confers the substrate specificity necessary to process stable RNAs such as 5.8S, U6, and snR38 by both interacting with RNA and protein components of nascent RNPs (Costello *et al*., 2011).

Rex1 is unique in the yeast proteome for its ability to fully process 5S rRNA and certain dicistronic tRNAs (Piper *et al*., 1983; van Hoof *et al*., 2000a), and shares an ability to process essential snRNAs with Rrp6/Rrp47 that is proposed to form the basis of a synthetic lethality relationship (Garland *et al*., 2013; Feigenbutz *et al*., 2013a). Although Rex1 is able to process these substrates, it does not share the structural features used to this end in RNase T or Rrp6. Rex1 lacks the NBS at the equivalent position to RNase T, and has been demonstrated by analytical ultracentrifugation and size exclusion chromatography to act as a monomer when purified in a HTP (His-6xTEV-Protein A)-tagged form from yeast in prior work by Taib Hama Soor (Hama Soor, 2017; Daniels *et al*., 2022), in agreement with previous observations by Frank and colleagues (Frank *et al*., 1999). The N- and C- terminal sequence flanking the Rex1 exonuclease domain lack any homology to known RNA-binding folds such as HDRC domains, and its established monomeric behaviour rules out the association of a substrate-engaging protein binding partner as seen in Rrp6 (Costello *et al*., 2011).

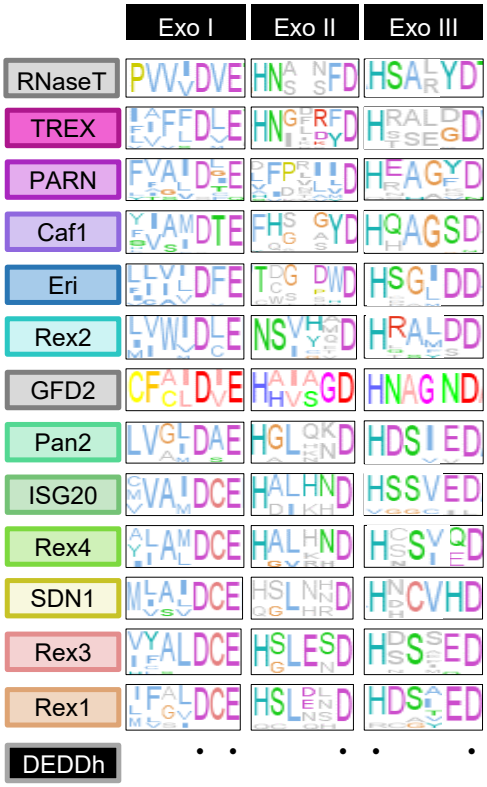


Figure 3.2: Sequence alignment of the DEDDh exonuclease domain EXO motifs.
Consensus logo representation of Clustal Omega (Sievers *et al.*, 2011) multi-alignment of DEDDh exonuclease domains, generated in Jalview v2.11.2.7 (Waterhouse *et al.*, 2009). Three EXO motifs are shown, each column represents roughly 150 aligned residues with each row corresponding to a clade shown in Figure 3.3. '•' indicates the positions of the DEDDh catalytic residues. Clades are shown in coloured boxes that correspond to positions in the phylogenetic tree Figure 3.3. Amino acids are coloured based on the Clustal X colour scheme: all glycines are shown in orange, all other colours are shown when >60% of aligned residues share a residue group, with polar and His/Tyr residues also shown if the 60% threshold is met for aligned hydrophobic residues. Within the GFD2 clade, hydrophobic residues are shown in peach, glycines are shown in magenta, polar residues are shown in green, acidic residues are shown in red, cysteines are shown in yellow, and phenylalanines are shown in orange. For full description of Clustal X colour scheme, see Table A.1.

The analysis presented in this chapter aimed to identify sequence features in *S. cerevisiae* Rex1 that confer its substrate specificity through examining sequence conservation between DEDDh enzymes, and the prediction of conserved structural features using protein structure prediction tools.

3.2 Results

3.2.1 A multi-species sequence alignment of the DEDDh exonuclease domain reveals conserved sequence features

The DEDD exonuclease family is widely conserved, and has received in-depth biochemical and structural characterisation for a subset of its member proteins. In order to extrapolate these findings onto Rex1, it is helpful to prepare a multi-alignment spanning several species in order to identify similarities between characterised sequence features and their Rex1 equivalents. Such an alignment would have

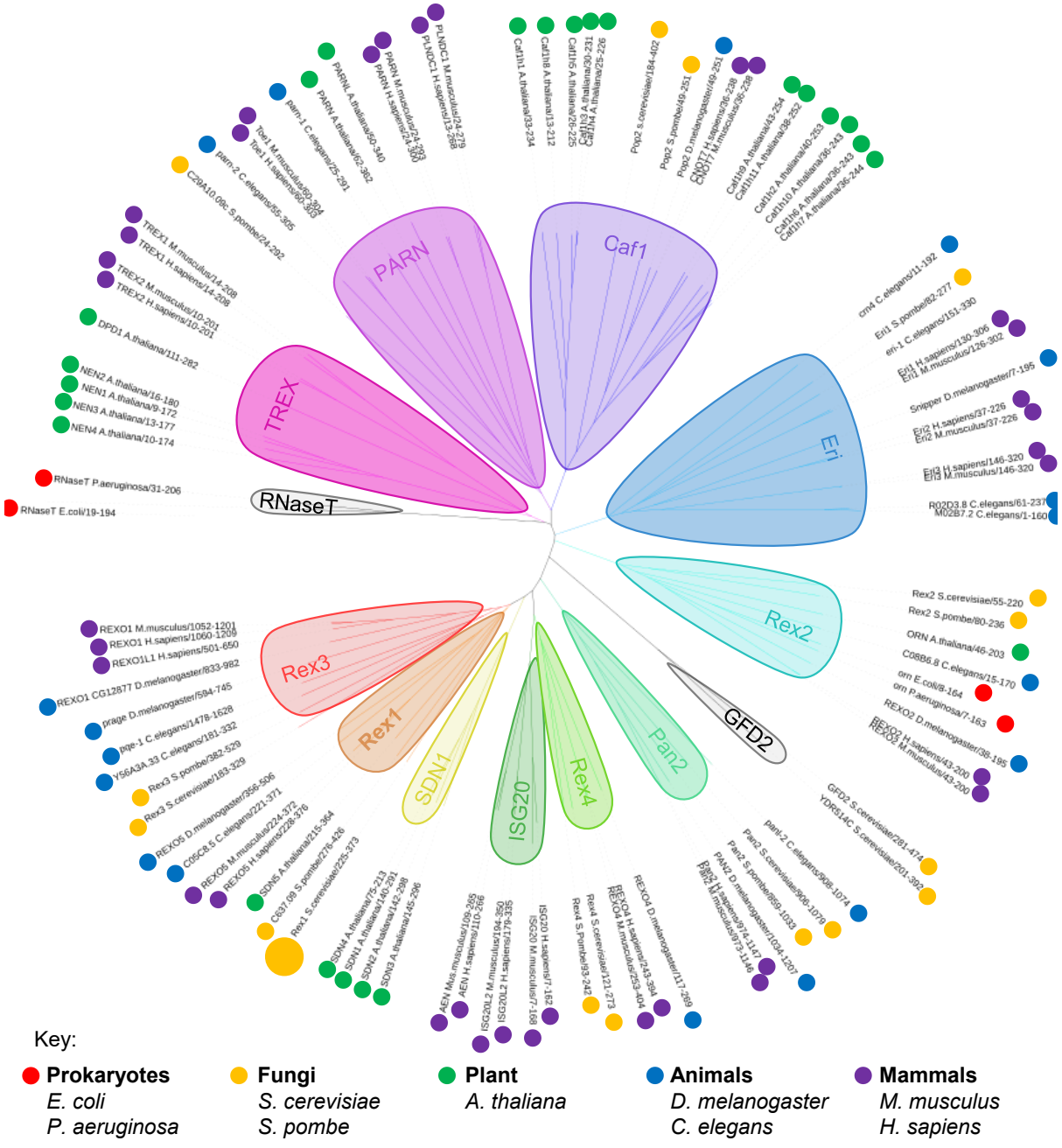


Figure 3.3: Phylogenetic analysis of the DEDDh exonuclease domain.
Phylogenetic tree built using the exonuclease domain alignment, rooted with *E. coli* RNase T. Rex1 is highlighted using a large yellow circle. Each coloured bubble corresponds to the same DEDDh subfamilies shown in A. Representative taxonomic groups of each individual protein's source is represented by a coloured dot, as described by the key. Tree graphic generated using Interactive Tree of Life (Letunic and Bork, 2019).

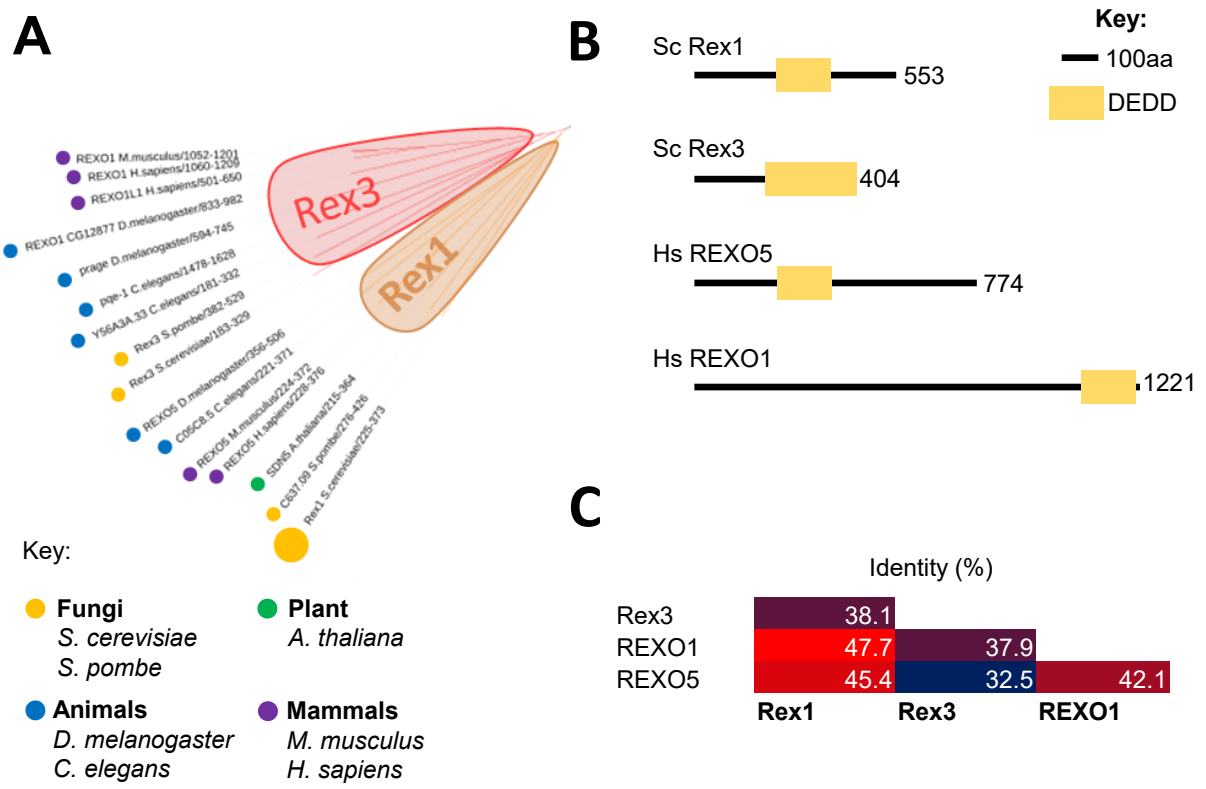


Figure 3.4: Unclear equivalence within the Rex1/Rex3 clade.
A: Magnification of the Rex1/Rex3 clade of the phylogenetic tree from Figure 3.3. **B:** Scale diagrams for the positioning of the exonuclease domain for Rex1/Rex3 clade members in *S. cerevisiae* and humans, based on the following UniprotKB entries: Rex1 - P53331, Rex3 - A7A111, REXO1 - Q8N1G1, REXO5 - Q96IC2 **C:** Four-way sequence identity comparison, Calculated using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). Cells are coloured from blue to red to indicate 30% - 50% sequence identity.

the additional benefit of revealing the phylogenetic relationships between DEDD exonucleases, which would be particularly useful in identifying sequence orthologues of Rex1 in humans.

At the outset of this project, a search of the Protein Data Bank was undertaken to identify experimentally-derived structural models of DEDDh exonucleases, specifically in complex with nucleotide substrates. Five structures were selected, each serving to represent structures of a specific protein: *A. thaliana* SDN1 (PDB entry 5Z9X; Chen *et al*., 2018), *S cerevisiae* Pan2 (PDB entry 6R9M; Tang *et al*., 2019), *H. sapiens* REXO2 (PDB entry 6STY; Szewczyk *et al*., 2020), *E. coli* RNase T (PDB entry 3NGZ; Hsiao *et al*., 2011), and *M. musculus* PARN (PDB entry 3D45; Wu *et al*., 2009); these proteins specifically would serve as sources for potentially conserved mechanistic features in Rex1. A DELTA-BLAST (Domain enhanced lookup time accelerated-Basic Local Alignment Search Tool; Boratyn *et al*., 2012) search of the non-redundant protein database was prepared using the Rex1 exonuclease domain amino acid sequence (225-373) as defined by the Pfam database (Mistry *et al*., 2021). This search was restrained to model organisms, particularly those with structural models available of DEDDh exonucleases in complex with polynucleotide substrates.

The results of the DELTA-BLAST search were aligned using EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute)'s Clustal-omega tool (Sievers *et al*., 2011). The alignment of members of the DEDD family is hampered by the non-uniform domain configuration of its members: some exonuclease domains are flanked with orthologue-specific N- and/or C-terminal regions. To account for this, we limited sequence alignment to the conserved boundaries of the exonuclease domain. To this end, the initial rough alignment of full-length proteins was manually cropped in Jalview v2.11.2.7 (Waterhouse *et al*., 2009) to leave only sequence within the 10 amino acids flanking the boundaries of the Rex1 exonuclease domain as defined by the Pfam database. Successful realignment of these cropped sequences was judged by the visible alignment of all three EXO motifs for each protein. This was the case for all clades in the final alignment with the exception of the PARN, where the clade-specific R3H subdomain insertion between EXOI and EXOII (Wu *et al*., 2005) lead to misalignment of the PARN EXOII motif with the EXOIII motif of other clades. A shortened 'consensus logo' view of the multi-alignment focussing on the three EXO motifs is shown in Figure 3.2A, with an unabridged view available in Appendix A.1. The multi-alignment was used with Clustal-omega's Simple Phylogeny tool to build a phylogenetic tree, which was visualised with Interactive Tree of Life (Letunic and Bork, 2019) to generate Figure 3.3. Using this alignment as a guide, the sequence features that had been observed to mediate substrate interactions in DEDDh exonuclease-substrate structural models were correlated to positions in Rex1; this analysis is compiled in Table 3.1.

All five DEDDh residues are conserved throughout the alignment, with the well-known exception of *S. cerevisiae* Pop2 that instead features an SEDQt motif (Daugeron *et al*., 2001). A handful of trends emerge: the EXOI motif almost always features four hydrophobic residues preceding the first DEDD aspartate, most clades feature a histidine residue at the beginning of the EXOII motif, and a cysteine is conserved in-between the catalytic aspartate-glutamate pair in EXOI for all Rex family clades. Using this alignment as a guide, it is possible to derive analogues for structurally-modelled substrate-interacting residues from the exonuclease domains of other DEDDh exonucleases. These substrate-binding predictions are listed in the fourth column of Table 3.1.

Members of the clades shown in Figure 3.3B readily cluster in agreement with their annotated names and resemble an existing analysis in the literature (Gerstberger *et al*., 2017), but differ in a handful of

Table 3.1: Curated list of DEDDh exonuclease residues from experimentally-determined structures showing nucleotide binding

Species	Protein	RNA-binding residues	Sequence aligned Rex1 residues	Structure(PDB ID)
A. thaliana	SDN1	M147	M232	5Z9X (Chen et al., 2018)
		R185	R271	
		H223	H308	
		S224	S309	
		R256	K340	
		S258	S342	
S. cerevisiae	Pan2	Y975	Y272	6R9M (Tang et al., 2019)
		N1019	N312	
		Y1046	K340	
		S1048	S342	
H. sapiens	REXO2	L53	S235 or L234	6STY (Szewczyk et al., 2020)
		W96	Y272	
		E146	N312	
		Y164	V324	
E. coli	RNaseT	F29	S235	3NGZ (Hsiao et al., 2011)
		E73	Y268	
		F77	Y272	
		F124	N312	
		F146	V324	
M. musculus	PARN	I34	S235	3D45 (Wu et al., 2009)
		R53	M232	
		H280	H308	
		H288	V316	

instances; likely a result of limiting alignment to within the DEDD domain. One notable difference is the presence of a *S. cerevisiae*-specific 'GFD2' clade, which contains a pair of closely related exonucleases: GFD2, and YDR514C. These exonucleases have no known function or obvious sequence homologues in other organisms. Another difference is that the *D. melanogaster* protein Snipper clusters within the Eri1 clade in agreement with the literature (Kupsco *et al*., 2006).

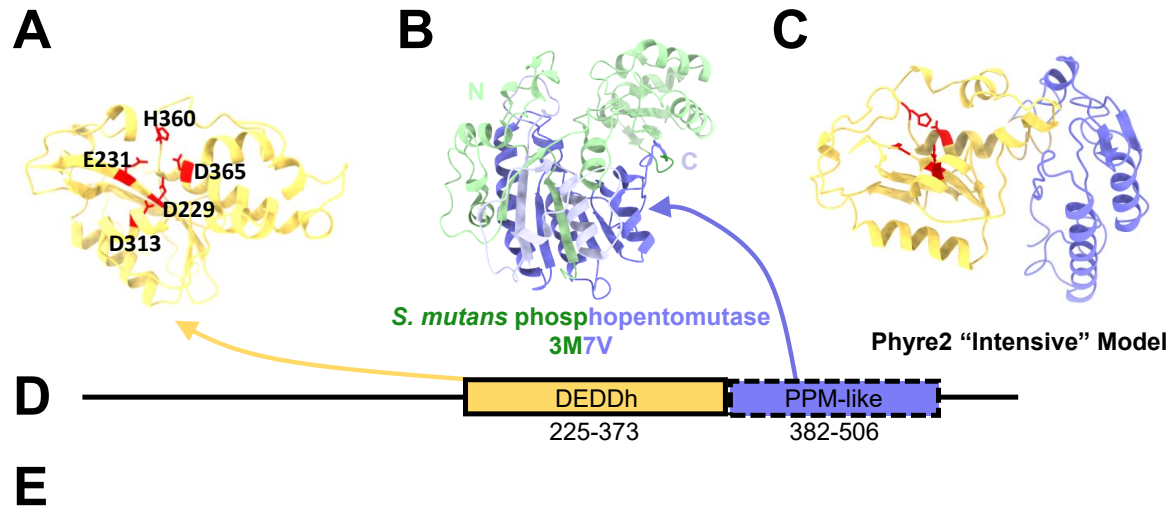
In the context of this project, the most important difference is seen when examining the clades containing the *S. cerevisiae* proteins Rex1 and Rex3. The phylogenetic analysis shown here suggests that Rex3 clusters with the human protein REXO1 and its homologues, whereas Rex1 seems to cluster more closely with homologues of REXO5. When examining the branches of the phylogenetic tree closely (magnified in Figure 3.4A), the branch points are far lower than those observed in other clades; this indicates a low similarity of amino acids within the exonuclease domain compared to other clades, and makes designation of clade boundaries more uncertain. This is reflected in the percentage sequence identities observed between yeast Rex1 and Rex3 with human REXO5 and REXO1 shown in Figure 3.4 C: Rex1 demonstrates higher identity with REXO5 and REXO1 than Rex3. The ambiguity in sequence homology between proteins in this clade is also observed in the analysis presented by Gerstberger *et al*., 2017, suggesting analysis of linear sequence alone may be insufficient for defining REXO5 and REXO1 as Rex1 or Rex3 orthologues

3.2.2 Structural modelling of the Rex1 exonuclease domain using Phyre2

A more compelling prediction of substrate recognition mechanisms is possible through comparison of computationally-predicted structural models of Rex1 with experimentally-derived models of homologous proteins bound to substrates. At the outset of this project, Phyre2 (Kelley *et al*., 2015) was a popular prediction tool for template-guided protein structure prediction. A Phyre2 search predicted two domains for Rex1: a model for the exonuclease domain based on a recent structure of *S. cerevisiae* Pan2 (PDB entry 6R5K; Schäfer *et al*., 2019) and a short C-terminal domain that resembled a fragment from a *S. mutans* phosphopentomutase structure (PDB entry 3M7V; unpublished), as shown in Figure 3.5. Phyre2 also offers an 'intensive' service that predicts how these domains may be arranged in an overall fold, using a combination of template-based modelling of predicted domains interspaced with *ab initio* modelling of non-templated sequence. The non-templated N- and C-terminal regions failed to produce a convincing model, but a model of how the two template-predicted domains may interact is shown in Figure 3.5C. With this model in mind, it was proposed that the C-terminal PPM-like motif may serve as an additional positively charged surface for substrate binding, owing to the high proportion of positively charged residues found in the motif. Later structural predictions generated by Alphafold2 however would reveal a more compelling domain model.

3.2.3 An Alphafold2 model of Rex1 reveals a novel 'RYS' domain resembling the alkaline-phosphatase family core fold

With the release of the Alphafold Database (Jumper *et al*., 2021; Varadi *et al*., 2022), a full-length structural model was proposed for Rex1 (entry P53331) shown in Figure 3.6. The associated quality metrics indicate confidence in the model with an average per-residue confidence score of 92% (pLDDT) for all regions except residues 1-52, which Mitchell group members Taib Hama-Soor and Quentin Levicky



E

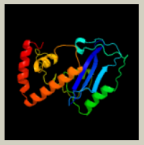
#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
2	c6r5kA_	Alignment		100.0	30	**PDB header:**rna binding protein **Chain:** A: **PDB Molecule:**pan2-pan3 deadenylation complex catalytic subunit pan2; **PDBTitle:** cryo-em structure of a poly(a) rnp bound to the pan2-pan3 deadenylase **PDB Entry:** PDBe RCSB PDBj Run Investigator
118	c2i09A_	Alignment	not modelled	87.8	11	**PDB header:**isomerase **Chain:** A: **PDB Molecule:**phosphopentomutase; **PDBTitle:** crystal structure of putative phosphopentomutase from streptococcus2 mutans **PDB Entry:** PDBe RCSB PDBj

Figure 3.5: A Phyre2 prediction of Rex1 structure predicts a short C-terminal fold.
A: Phyre2 generated model of the Rex1 exonuclease domain based on *S. cerevisiae* Pan2 (PDB entry 6R5K; Schäfer *et al*., 2019) displayed using ChimeraX. The DEDDh residue side chains are highlighted in red. **B:** A Phyre2-predicted region of Rex1 structural homology (residues 383-506) to *S. mutans* phosphoglycerate mutase (PPM; residues 224-342; PDB entry 3M7V; unpublished), highlighted in purple. The unaligned regions at the N- and C- termini of the PPM (Phosphopentomutase) structure are highlighted in light green and light purple respectively. **C:** A Phyre2 "intensive" model of how both predicted domains may interact. **D:** Diagram of the Phyre2-predicted domains represented on the primary sequence of Rex1. **E:** The Phyre2 web interface results page, limited to the two predictions shown (Kelley *et al*., 2015). Screenshot taken 19/5/2023.

have demonstrated act as a nuclear localisation sequence (Hama Soor, 2017; Daniels *et al*., 2022).

Unlike Phyre2, Alphafold2 predicted that the N- and C- terminal regions of Rex1 form a composite second domain featuring a central 8-strand β-sheet composed of N- and C-terminal sequence. As a similar domain configuration was observed in Alphafold2 models for the Rex1 homologues YFE1 from *S. pombe* and SDN5 from *A. thaliana*, we proposed the 'RYS' (Rex1-Yfe1-SDN5) domain as a prospective name. The exonuclease and RYS domains seem to form a cleft, with the two linker sequences joining the domains forming a double hinge at the base. Tantalisingly, this modelled domain features three loops capped with positively charged residues that reach across from the second domain towards the DEDDh active site, making natural candidates for RNA-binding features. These loops are highlighted in Figure 3.6D, and will be referred to with the following shorthand: Loop1 (145-154), Loop2 (459-472), and Loop3 (506-544).

Alphafold DB cannot list the structures that contributed to a given prediction in the manner of Phyre2. A DALI search of the model without the exonuclease domain suggests however that the Alphafold2 model resembles PDB entries for phosphopentomutases and phosphoglycerate mutases, specifically the 'Alkaline-phosphatase-like, core domain superfamily' fold (InterPro ID: IPR017850). The core of this domain resembles the structure for pentophosphomutase indicated by the Phyre2 prediction in Figure 3.5B, which also features a central domain flanked by a composite N- and C- terminal second domain. As an example, the core domain of PDB entry 3TX0 (Iverson *et al*., 2012) for *S. mutans* phosphopentomutase (residues 1-100//219-392) aligned well with the Alphafold2 model of the Rex1 RYS domain (residues 51-166//386-553) using the PDB's pairwise alignment tool (rcsb.org/docs/tools/pairwise-structure-alignment), with Figure 3.7A demonstrating an average root mean square deviation (RMSD) of 3.1Å for the 184 residues aligned.

While the three loop arrangement seems to be preserved in *S. pombe* (entry O94443) and *A. thaliana* (entry Q8L7M4), other clade members seem to possess different configurations- the Alphafold DB entries for REXO5 in human (entry Q96IC2) and mouse (entry D3YW29) seem to possess a tandem pair of Pfam-recognised canonical RNA-recognition motifs (RRMs) affixed to the C-terminal half of the pentophosphomutase-like 'core' fold at the equivalent insertion site to Loop 2, whereas no obvious '3-loop' or RRM-like features can be seen in the REXO5 Alphafold DB entries for *D. melanogaster* (entry Q9VRX7) or *C. elegans* (entry O16312). These three RYS domain configurations are summarised in Figure 3.8. All examined Rex1 homologues are predicted to possess a 7-strand β-sheet as seen in the alkaline phosphatase family, with the exception of a 8-strand β-sheet predicted for *S. cerevisiae* Rex1. The extra β-strand observed in *S. cerevisiae* Rex1 is found between β-strands 3 and 4, and has a comparatively less confident pLDDT score at an average of 73 over the 4 residues.

The biological relevance of the structural homology between these predicted structures and the alkaline phosphatase family remains unclear. A parsimonious explanation would be that this second domain provides a positively charged surface for the binding of substrates. The composite 'core' fold seen in pentophosphomutases is conserved throughout the broader alkaline phosphatase family (Jonas and Hollfelder, 2009; Galperin and Koonin, 2012), which demonstrates the mechanistic flexibility of adding, removing, and moving the phosphate groups of their substrates. The lack of conservation for known catalytic residues of the alkaline phosphatase family renders the preservation of this catalytic activity in Rex1 unlikely, but one possibility is that this domain may specifically recognise the base paired 5´-terminal nucleotide tri-phosphate observed in 5S rRNA (Soave *et al*., 1973; Mori and Ichiy-

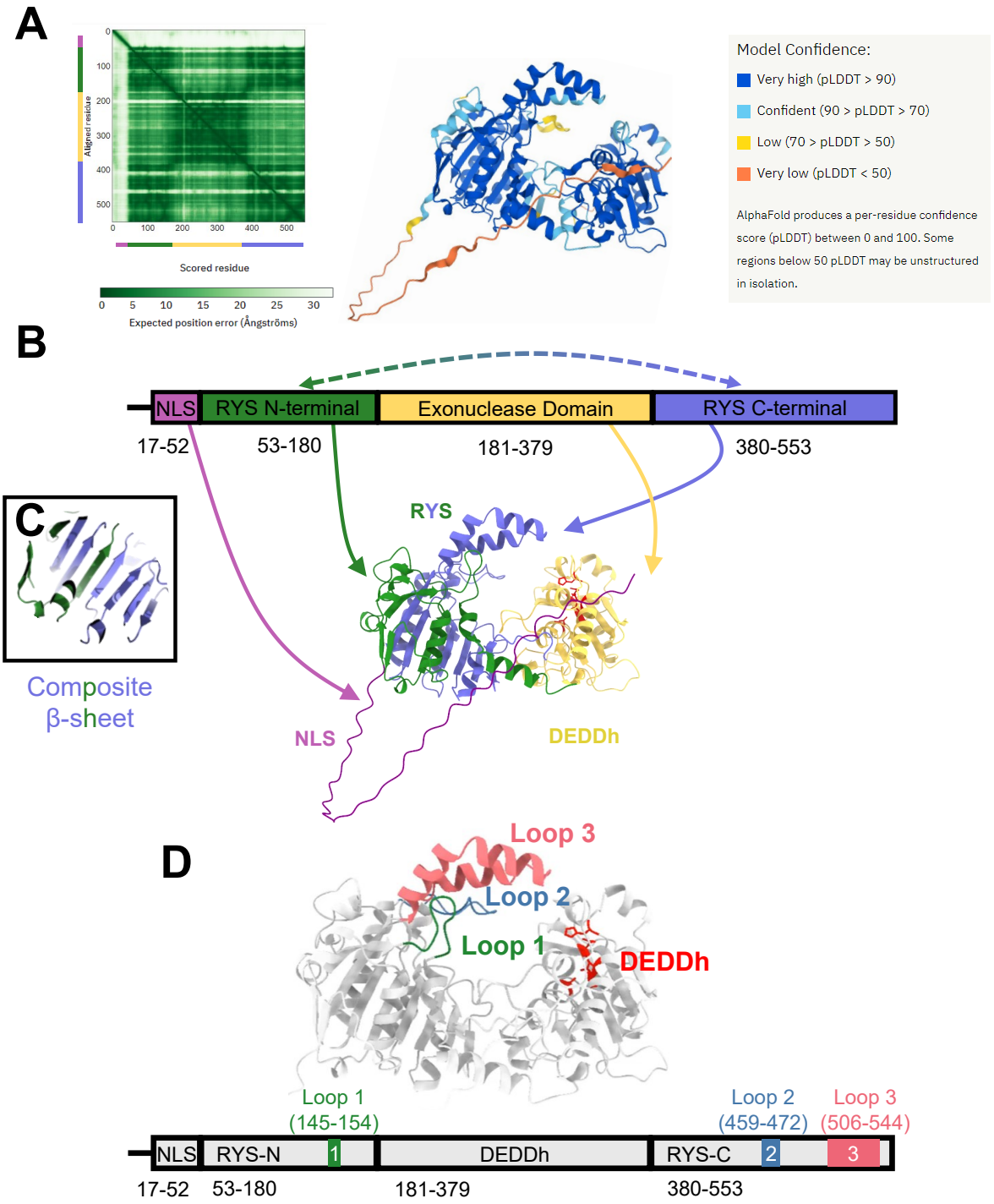


Figure 3.6: An Alphafold2 model for Rex1 structure

A: Summary of quality metrics for the EMBL-EBI Alphafold DB entry for Rex1 (https://www.alphafold.ebi.ac.uk/entry/P53331; Jumper *et al.*, 2021; Varadi *et al.*, 2022). Screenshots taken from the Alphafold DB entry webpage, with miniature domain diagrams added to the expected position error heatmap for clarity. **B:** A ChimeraX model of Alphafold2-predicted domain composition for Rex1. **C:** Cutaway of the composite RYS domain demonstrating the composite β-sheet. **D:** The position of three loop structures indicated on both the Alphafold model, and a domain diagram of Rex1.

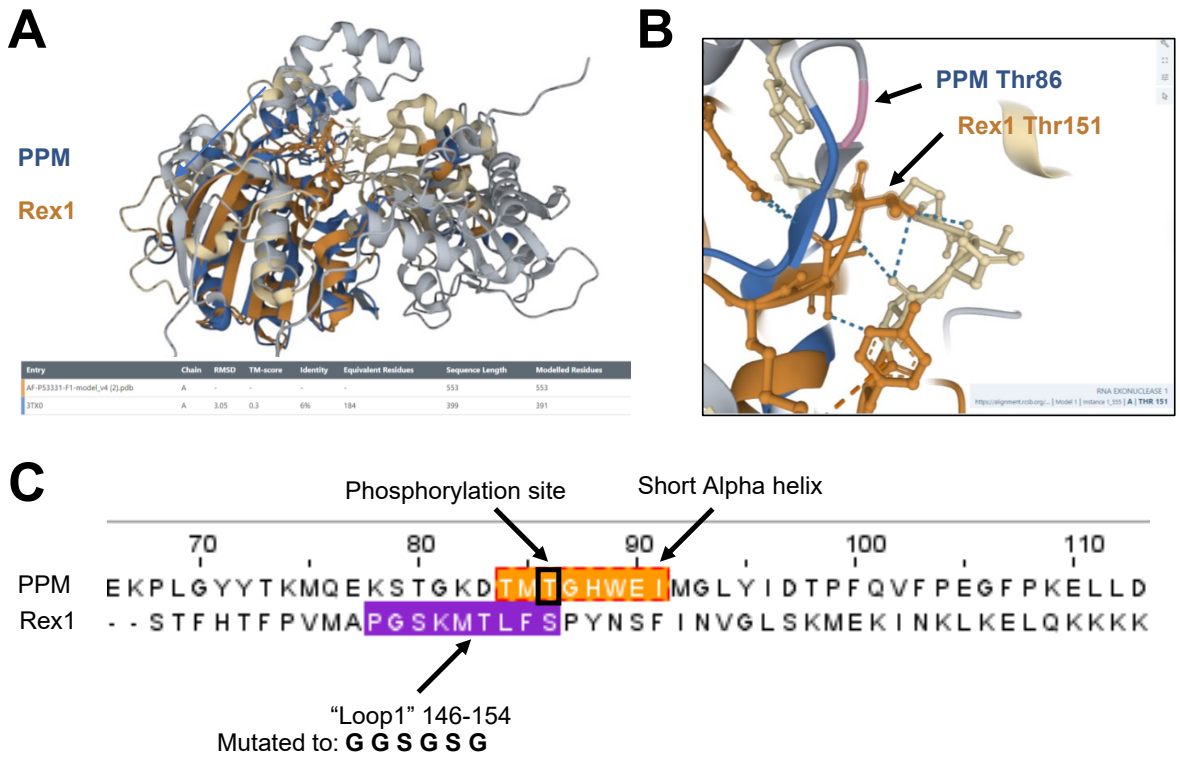


Figure 3.7: Pairwise structural alignment of Rex1 and pentophosphomutase

A: Pairwise alignment of the Alphafold DB entry for Rex1 (residues 51-166//386-553) and *B. cereus* pentophosphomutase (PDB 3TX0; residues 1-100//219-392) using the rigid FATCAT algorithm (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists; Li *et al*., 2020). Images are screenshots of the Protein Databank's pairwise structural alignment tool (https://www.rcsb.org/alignment; accessed 19/05/2023) **B:** Close-up of the catalytically phosphorylated Threonine 86, and a Threonine found in "Loop 1" of the Rex1 Alphafold2 model. **C:** Jalview display of the pairwise alignment between PPM and Rex1. The catalytic phosphorylation site of PPM is highlighted, along with the surrounding α-helix. "Loop 1" of the Rex1 Alphafold2 model is indicated, along with the 6-amino acid substitution that replaces it in the "Loop1Δ" rex1 mutant.

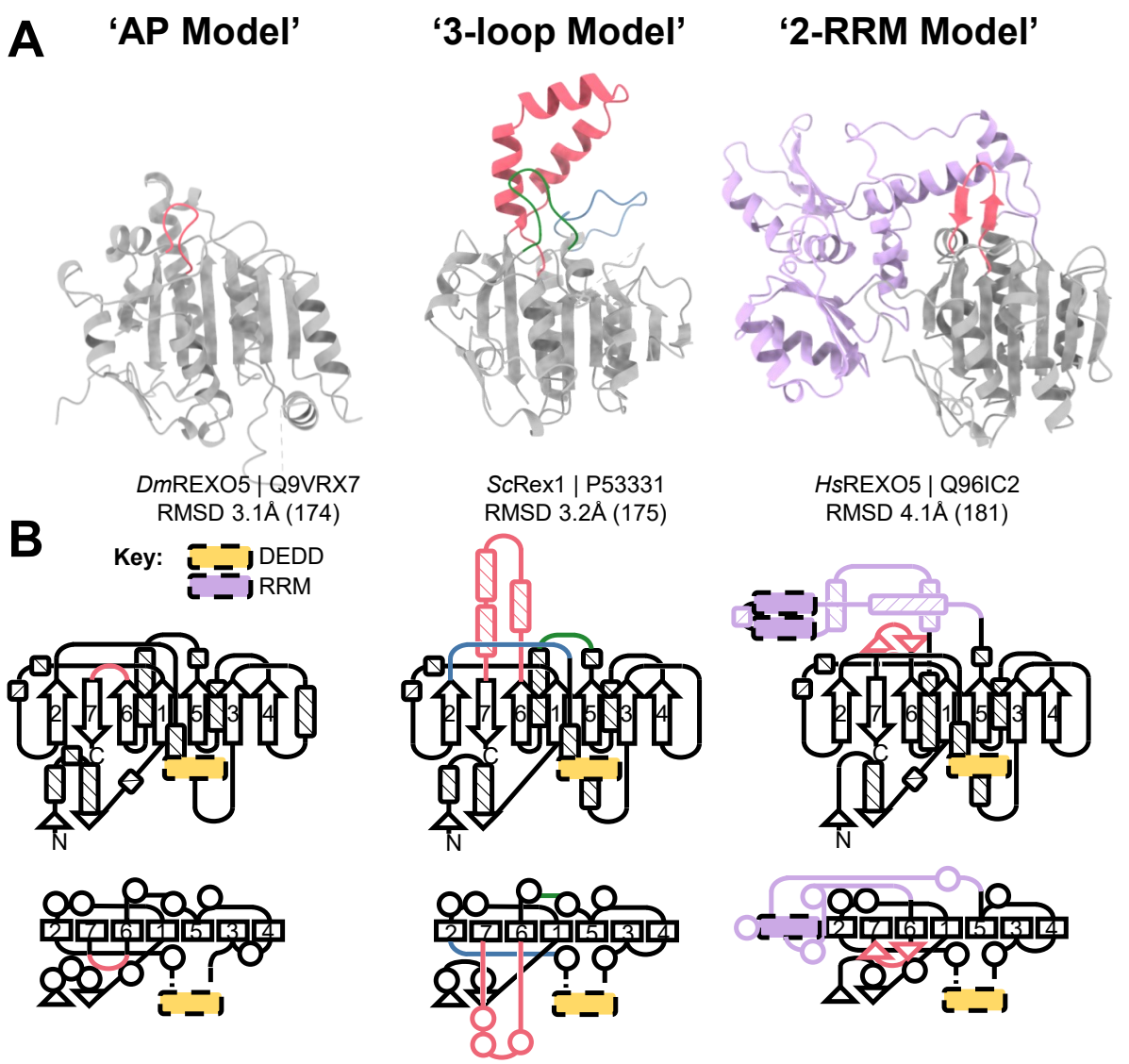

A
'AP Model'
'3-loop Model'
'2-RRM Model'
DmREXO5 | Q9VRX7
RMSD 3.1Å (174)
ScRex1 | P53331
RMSD 3.2Å (175)
HsREXO5 | Q96IC2
RMSD 4.1Å (181)
B
Key:
DEDD
RRM
C
N
C. Elegans REXO5
56-156
469-594 541-555
D. Melanogaster REXO5
144-269
545-681
'AP Model'
S. cerevisiae Rex1
53-180 145-154
380-553 459-472 506-544
S. pombe YFE1
89-209 187-196
443-631 514-542 576-622
A. Thaliana SDN5
41-137 118-125
381-567 455-486 520-558
'3 Loop Model'
M. musculus REXO5
44-138
408-489 490-564 585-655 656-759 730-740
H. sapiens REXO5
48-144
421-504 505-574 600-679 680-774 745-754
RRM
'RRM Model'

Figure 3.8: The RYS domain core is conserved throughout the Rex1 clade
A: The isolated RYS domains of the Rex1 clade each in turn 1-1 threaded to the PPM 'core domain' of PDB entry 3TX0 (Iverson *et al*., 2012), and grouped by common features. For each alignment to PPM, a Root Mean Square of Deviation (RMSD) is given, with the number of residues aligned given in brackets. **B:** Simplified structural models of the three RYS domain configurations, based on β-sheet model from ChimeraX display of 3TX0 (Iverson *et al*., 2012). β-strand are numbered from 1-7 from N-C termini, and change colour from red to blue after interruption by the exonuclease domain, and from blue to green if interrupted by RRM domains. **C:** Scale domain models of pentophosphomutase (PPM) 'core' and RYS domains. Intermediate break indicates the position of the cap domain in PPM, and a DEDDh exonuclease domain in all other cases. The positions of '3 Loop' features annotated by manual structure analysis. Tandem RRM domain boundaries defined in accordance with the Pfam database (Mistry *et al*., 2021), and are separated by a linker indicated by a purple stripe pattern.

anagi, 2021).

Another more remote possibility is that the mutase activity of this domain is preserved. The possibility that Rex1 is able to remove or confer unusual 3´ phosphate structures such as a 2´-3´ cyclic phosphate is an intriguing, if remote possibility that would suggest mechanisms by which Rex1 could access specific substrates with unusual end structures, or deposit unusual end structures to prevent degradation by exonucleases. In pentophosphomutase in *Streptococcus mutants*, a catalytic threonine 85 is phosphorylated and dephosphorylated during the enzyme's catalytic cycle (Iverson *et al*., 2012). The mechanism of the core domain is commonly observed to involve the coordination of metal ions such as Zn^{2+} and Mg^{2+} to stabilise this intermediate (Galperin *et al*., 1998), which might suggest additional cofactor requirements for Rex1 beyond the DEDD-coordinated Mg^{2+} pair typical of an exonuclease two-ion mechanism (Steitz and Steitz, 1993). While the general structural homology between the RYS domain and pentophosphomutase is high, the structural homology in the region of this catalytic threonine is low, correlating loosely with the position of 'Loop 1' as shown in Figure 3.7B and C. Loop 1 does however contain threonine and serine residues that could conceivably be phosphorylated in an equivalent mechanism. Activity that would involve a stably phosphorylated intermediate as seen in pentophosphomutase can be confidently ruled out, as mass spectrometry datasets indicate that no phosphorylation is observed in Rex1 outside of the NLS (BioGrid entry 33526; Stark *et al*., 2006); although this doesn't rule out the transient phosphorylation mechanism seem in other alkaline phosphatase-related enzymes (Iverson *et al*., 2012). It is possible that experimentally-determined Rex1 structures may reveal better analogues for conserved alkaline phosphatase-like catalytic features, however the analysis we present here suggests this fold most likely serves a non-catalytic role, perhaps involving substrate recognition.

In addition to predicting structural features, the Alphafold Database provides a new perspective for phylogenetic analysis of the Rex1/Rex3 clade. Rex3 and REXO1 share similar domain configurations: a C-terminal DEDD domain is preceded by an extended and uncharacterised N-terminal sequence. The contents of this extended N-terminal sequence suggest further similarity: the Alphafold Database predictions for each protein (Rex3: entry A7A111, REXO1: entry Q8N1G1 ; Jumper *et al*., 2021; Varadi *et al*., 2022) predict short structural motifs recognised by the DALI structure recognition server (Holm *et al*., 2023) as resembling the RNAPII cofactor TFIIS at positions 35-99 and 800-873 in Rex3 and REXO1 respectively. Taken together the analysis presented here provides a resolution to the uncertainty of sequence homology within the Rex1/Rex3 clade using predicted structural homology of Rex1 with REXO5, and of Rex3 with REXO1.

3.2.4 A threaded model of Rex1 in complex with substrate predicts potential RNA-binding residues

While the Alphafold2-predicted model of Rex1 is informative in isolation, a structural model displaying how sequence features of Rex1 engage with substrates is a fundamental aim of this project. A 2019 X-ray crystallographic study of *S. cerevisiae* Pan2 demonstrates DEDDh exonuclease domain interactions with substrates ranging in length from 5-7 nucleotides (Tang *et al*., 2019). Since Phyre2 had produced a threaded model of Rex1 (residues 181-381) onto the exonuclease domain of Pan2 (residues 432-638; PDB entry 6R9M; Tang *et al*. 2019) in the absence of a substrate, it was straightforward to thread this model onto a structure featuring a substrate with an RMSD between the Cα backbones of 0.75 Å, as shown in Figure 3.9A.

This model indicated several residues that may have a role in Rex1 substrate recognition: Y272, N312, K340, and S342, with a concurrently published structural study implicating these same residues in substrate binding (Schäfer *et al*., 2019). Strangely, K340 seemed to clip through the predicted trajectory of the substrate, however it is worth noting that Phyre2 orients side chains based on predicted energetic favourability and not based on information from structurally determined models. A further candidate was suggested upon generation of this model, H308, based on the positioning of the side chain from the base of the cleft towards the phosphodiester backbone. This residue is highly conserved at the beginning of the EXOII motif (as shown in Figure 3.2), with structural data suggesting a substrate binding role for the homologous residue in *A. thaliana* SDN1 (H223; Chen *et al*. 2018) and PARN (H280; Wu *et al*. 2009)

The positioning of this AAGGAA RNA substrate in the Alphafold2 model was observed by threading the exonuclease domains of the Phyre2-substrate model and the Alphafold2 model with an RMSD between the backbones of 2.0 Å, giving Figure 3.9B. This second substrate bound model differs in the relative orientation of the proposed RNA-interacting residue side chains. Unlike the Phyre2 model, the Lysine 340 chain no longer clips through the substrate, and is instead oriented towards the base of the cleft.

The positioning of the substrate for the model shown in Figure 3.9B presents a degree of ambiguity for the path of substrate entry, as the substrate seems to be oriented directly into the RYS domain. A potential hint for the path that substrates may take is given by the distribution of charge: the 'front' face is mostly positively charged, while the 'rear' face is negatively charged, as shown in Figure 3.9C.

It is however possible that there may be a degree of flexibility between the two domains, with the double hinge at the base serving as the pivot. In a structural study of *S. mutants* pentophosphomutase, Iverson and colleagues suggested that twisting between the protein's 'core' and 'cap' domains may have functional implications (Iverson *et al*., 2012). This echoed a mechanism proposed for the homologue in *B. cereus* where structural models of catalytic intermediates indicated that rotation of the cap domain may be coupled to the catalytic cycle (Lahiri *et al*., 2003; Zhang *et al*., 2005). In the context of Rex1, the reorientation of the exonuclease domain could potentially occlude the active site, or reorient the overall fold to accommodate larger substrates through coordinated movement of the 3 RYS domain flexible loops.

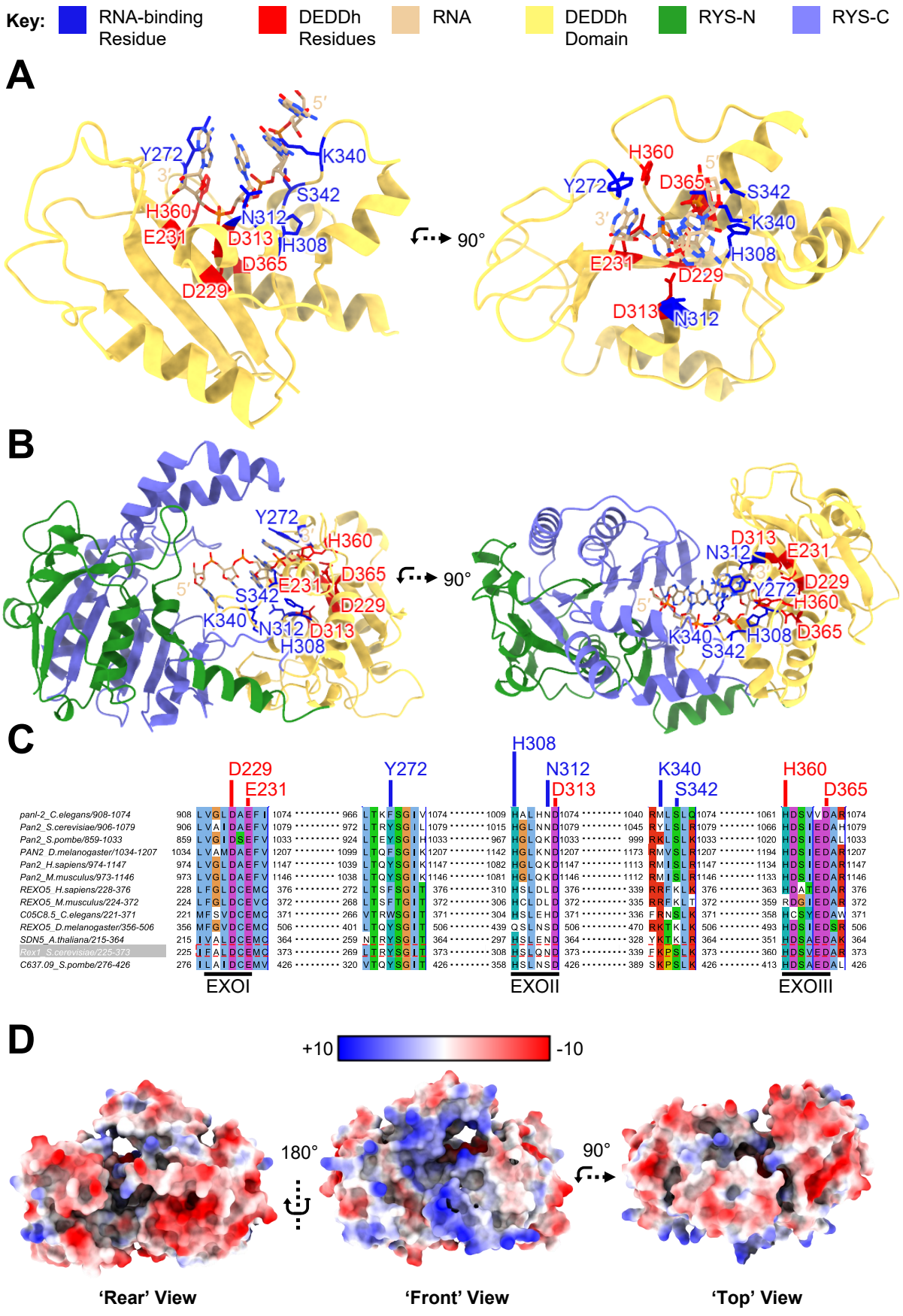

Key:
RNA-binding Residue
DEDDh Residues
RNA
DEDDh Domain
RYS-N
RYS-C
A
Y272
K340
S342
H360
N312
E231
D313
H308
D365
D229
90°
H360
Y272
D365
S342
K340
H308
E231
D229
D313
N312
B
Y272
H360
D365
E231
S342
D229
K340
N312
D313
H308
90°
D313
N312
E231
Y272
D229
H360
K340
H308
D365
S342
C
D229
E231
Y272
H308
N312
D313
K340
S342
H360
D365
panl-2_C.elegans/908-1074
Pan2_S.cerevisiae/906-1079
Pan2_S.pombe/859-1033
PAN2_D.melanogaster/1034-1207
Pan2_H.sapiens/974-1147
Pan2_M.musculus/973-1146
REXO5_H.sapiens/228-376
REXO5_M.musculus/224-372
C05C8.5_C.elegans/221-371
REXO5_D.melanogaster/356-506
SDN5_A.thaliana/215-364
Rex1_S.cerevisiae/225-373
C637.09_S.pombe/276-426
EXOI
EXOII
EXOIII
D
+10
-10
180°
90°
'Rear' View
'Front' View
'Top' View

Figure 3.9: Threading of substrate into Phyre2 and Alphafold-based Rex1 structural models. Comparison of 5 RNA-binding residue prediction positions on Phyre2 and Alphafold2-based Rex1 models (Kelley *et al*., 2015; Jumper *et al*., 2021; Varadi *et al*., 2022) displayed as cartoon models using ChimeraX v1.2 (Pettersen *et al*., 2021). Predicted RNA-binding residues are indicated in blue. Equivalent residues in sequence aligned structures are given in Table 3.1. **A:** ChimeraX display of Phyre2 one-one threaded model of the Rex1 exonuclease domain (181-381) onto a structure of *S. cerevisiae* Pan2 in complex with AAGGAA RNA (4nt shown; PDB 6R9M; Tang *et al*., 2019). Front view and top view shown **B:** Manual alignment in ChimeraX of the Rex1-Pan2 structure onto the Alphafold2 structure model for Rex1 allowing positioning of the AAGGAA RNA substrate (5nt shown). Front view and top view shown. **C:** Electrostatic potential map of the Rex1 Alphafold2 model. Surface potential calculated and displayed using 'Coulombic' command in ChimeraX, expressed as kcal/(mol·e) at 298 K.

3.3 Discussion

The phylogenetic analysis presented in Figure 3.3 is a useful resource for drawing parallels between the exonuclease domains of DEDDh enzymes, however a number of members from each organism seem to be missing, such as Rrp6 homologues and DNA polymerase proofreading modules. This may be a consequence of the evolutionary distance between bacteria, plants, and the rest of the included species, which are more closely grouped by comparison. This could also be a consequence of the sequence diversity of the DEDD superfamily, requiring a different approach to sequence alignment.

One approach could be to create a chain of overlapping alignments, by overlaying the results of DELTA-BLAST searches of the exonuclease domains of known absences. An ever-expanding alignment however, risks an increasingly ambiguous correlation of residues in-between the conserved EXOI-III motifs. If the goal of the phylogenetic analysis is just to observe the evolutionary distance between DEDDh enzymes, then a solution could be to exclude the residues found between the three EXO motifs, creating a compact and strictly defined benchmark of evolutionary distance.

A different approach may be required if, as attempted here, the aim of a phylogenetic analysis is to discover conserved structural features within the exonuclease domain. With the systematic generation of structural models using Alphafold2 by EMBL-EBI in the creation of the Alphafold Database (Jumper *et al*., 2021; Varadi *et al*., 2022), a more elegant alternative to the alignment of naked amino acid sequences may be the mass-threading of structural models. The online tools already exist that can be used to achieve this on a small scale, first through the identification of similar Alphafold structures using a search engine, the DALI and FoldSeek webservers both offer this service (Holm *et al*., 2023; van Kempen *et al*., 2023), which was followed in this analysis by alignment of full or partial structural models using the PDB's structural threading tool. While a multi-alignment as attempted in Figure 3.3 allows an estimate of equivalent residues between closely related proteins, a more confident correlation was drawn through the threading of the Rex1 fold onto another protein's structural model as in Figure 3.9. When attempting to draw parallels between proteins, the knowledge that a specific position on a protein's predicted structure is occupied by certain residues may provide more context than the positioning of an amino acid or motif in a protein's linear sequence.

Rex1 and its functional homologues are unique in their respective organisms for their ability to engage with RNase T-like substrates, consequently models built of the exonuclease domain based on related proteins that cannot process these substrates may obscure relevant mechanistic features. Pan2 possesses a substrate recognition mechanism whereby helically stacked polyA RNA is recognised through

shape (Tang *et al*., 2019), with the individual residues listed in Figure 3.9A interacting with the ribose-phosphate backbone instead of bases themselves. PolyA tails are very different substrates from the stem loop-adjacent short overhangs Rex1 is known to process, however these phosphate backbone interactions may be conserved in Rex1 as a generic mechanism for the binding and orientation of substrates - albeit without selectivity for helically stacked polyA RNA.

A key advantage of Alphafold2 over Phyre2 becomes apparent by comparing the respective models for Rex1. Phyre2 uses the assumption that a protein's domains form a series of discrete structurally conserved domains akin to beads on a string, and uses the 'intensive' setting to find an energetically favourable configuration of those domains to create an overall fold. This creates a blind spot for proteins with composite domains, as is seen for the alkaline phosphatase family - Alphafold2 largely avoids this bias, through its residue covariance approach that initially prioritises the proximity of residues over mainchain geometry. Another advantage is apparent when comparing the orientation of side chains between the two models - Alphafold2's residue covariance approach to side chain orientation resolved a predicted substrate amino acid side chain conflict that was present in the substrate-threaded Phyre2 model of the Rex1 active site.

The core of the Alphafold2-predicted RYS domain is conserved throughout the Alphafold2 models of the Rex1 clade. These models predict heterogeneity for the composition of this domain beyond the core fold, hinting at diverging mechanisms for substrate recognition. The lack of these additional features in *D. melanogaster*, an organism where *REXO5* gene deletion results in embryonic lethality (Gerstberger *et al*., 2017), may however indicate that the unadorned RYS domain core serves a role fundamental to function within the clade, while the additional features seem in the '3-loop' and '2-RRM' models to serve additional functions that are secondary to the primary function.

The intuitive explanation for the divergence of the domain configurations outline in Figure 3.8 would be that the '3-loop' and '2-RRM' members diverged from the more compact 'AP' model. The '3-loop' model however is found in both yeast and plants, with *Z. mays* (maize) and *G. max* (Soy bean) homologues also predicted by AlphafoldDB to possess this fold. While convergent evolution remains a possibility, the simpler explanation is that the '3-loop' model is the default configuration present in the last eukaryotic universal ancestor, with the 'AP' and '2-RRM' configurations arising between the common ancestor of fungi and Nephrozoa, the last common ancestor between insects and mammals (evolutionary relationships determined using https://www.onezoom.org/ ; Wong and Rosindell, 2022).

While Alphafold2 provides high confidence values for folding within each domain of the Rex1 model, it has lower confidence scores for the relative orientations of these two domains. Given the presence of a predicted flexible hinge at the base of the central cleft, it could be that two domains are able to open to accommodate larger substrates. Given Rex1's ability to remove nucleotides in close proximity to duplex structures, the Alphafold2 model as presented seems to possess a pore that is too narrow at 13.5Å to accommodate double-stranded RNA (roughly 20Å as measured from PDB 3VA3; Hsiao *et al*., 2012), suggesting such a domain rearrangement may be necessary for the engagement of substrates.

This bioinformatic analysis predicts several structural features of Rex1 that may contribute to its substrate recognition mechanisms. These predictions allow the design of experiments that may be able to test a structural model of Rex1 function before an experimentally-derived structural model is solved. While the Alphafold2 model for Rex1 is compelling, it fails to capture a number of details that could

be addressed by experimental data, including the binding of cofactors and conformational changes the Rex1 fold may undergo in response to the processing of different substrates.

Chapter 4

Biochemical and Crystallographic Analysis of Rex1

4.1 Introduction

With an integrated understanding of an exonuclease-substrate complex's structure and its biochemical properties, it is possible to derive the mechanisms through which substrates are recognised and processed. An extensive collection of DEDDh structures have been reported to date; 148 structures have been determined for the DEDDh exonuclease domain as defined by InterPro family IPR013520 as of September 2023 (Jones *et al*., 2014; Paysan-Lafosse *et al*., 2023). For exonucleases where structures have been solved, many cases exist where substrate binding can be observed (see Table 3.1 in Chapter 3). Among these proteins, Rex1's functional homologue in *E. coli* RNase T has received a seminal treatment of its substrate recognition mechanisms by the laboratory of Murray Deutscher (Deutscher *et al*., 1984; Deutscher and Marlor, 1985; Zuo and Deutscher, 2002a,b,c).

The initial biochemical assessment of RNase T revealed several buffer requirements that would later be observed throughout the DEDD superfamily (Deutscher *et al*., 1984; Deutscher and Marlor, 1985). Exonuclease activity was observed to be dependent on the binding of Mn^{2+} or Mg^{2+} divalent cations, with activity inhibited in the presence of other divalent cations including Ca^{2+}, with the highest level of activity observed at alkaline pHs of 8-9 (Deutscher *et al*., 1984). This is consistent with the 2-ion mechanism predicted for all DEDD enzymes (Steitz and Steitz, 1993), which states that hydrolysis is mediated through the binding of two Mg^{2+} or Mn^{2+} cofactors alongside a general base histidine or tyrosine that requires alkaline pH for de-protonation into a catalytically active state. RNase T was observed to be inhibited by increasing concentrations of KCl, and demonstrated instability at or above temperatures of 37ºC that could be protected against with high, albeit inhibitory KCl concentrations of 1M (Deutscher *et al*., 1984).

Later studies revealed a similar set of requirements in the trimming of DNA substrates by RNase T, which demonstrated tighter binding than RNA equivalents (Viswanathan *et al*., 1998; Zuo and Deutscher, 1999). Notably, the 3´ overhang of dsDNA bound more tightly than the 3´ overhang of dsRNA, but ssRNA bound more tightly than ssDNA. This reflects the physiological roles of RNase T, which is specialised to trim the 3´ ends of tRNA and 5S rRNA precursors (Zuo and Deutscher, 2002b), but can also trim 3´ DNA overhangs during DNA repair (Viswanathan *et al*., 1998, 1999; Hsiao *et al*., 2014).

The integration of biochemical data from the Deutscher laboratory (Zuo and Deutscher, 2002a,b,c) and extensive structural study of enzyme-substrate complexes from the laboratory of Hanna Yuan (Hsiao *et al*., 2011, 2012, 2014) have enabled a detailed mechanistic understanding of the unique substrate specificity of RNase T amongst the *E. coli* DEDD exonucleases. RNase T is known to cleave short 3´ overhangs adjacent to a stemloop structure, leaving either a blunt end or one-two 3´C nucleotides (Zuo and Deutscher, 2002b). This mechanism was first modelled using structural homology prediction (Zuo and Deutscher, 2002c), then confirmed and expanded upon with experimentally-derived structures of RNase T in complex with preferred and inhibitory sequences (Hsiao *et al*., 2011, 2012).

Another DEDD exonuclease that has benefited from integrated structural and biochemical study of substrate binding is Rrp6, which has received rigorous biochemical examination individually (Burkard and Butler, 2000; Axhemi *et al*., 2020), in complex with the RNA exosome (Liu *et al*., 2006), and with the additional presence of substrate-binding adaptor complexes (Das *et al*., 2021; Gerlach *et al*., 2022; Kögel *et al*., 2022). Notably, Rrp6 exonuclease activity was observed to be inhibited by CCA motifs in a similar manner to RNase T (Axhemi *et al*., 2020). A concurrently published pair of X-ray crystallography studies succeeded in observing Rrp6 as part of a substrate-bound exosome complex (Wasmuth *et al*., 2014; Makino *et al*., 2015), which in combination with transcriptome-wide crosslinking data (Schneider, Kudla, Wlotzka, Tuck and Tollervey, 2012; Delan-Forino *et al*., 2017) have given rise to a model for its contribution to substrate triage through the exosome's catalytic centres.

Rex1 has received *in vitro* biochemical assessment prior to this study, revealing exonuclease activity against tRNA precursor analogues that is dependent on the presence of Mg^{2+} (Ozanick *et al*., 2009), and an ability to trim the extended 5S rRNA purified from the mature ribosomes of *rex1*Δ yeast to their mature length (Hama Soor, 2017; Daniels *et al*., 2022). The work presented in this chapter aims to provide an expanded biochemical assessment of recombinant Rex1, and establishes substrate binding properties and preferences using simple ssDNA substrates. A long term aim of this project is to experimentally determine structural models of Rex1 in complex with substrate, and understand the mechanisms that allow engagement of Rex1 with its physiological substrates. By understanding the biochemical properties of Rex1 *in vitro*, we can build up a nuanced understanding of the roles Rex1 plays in the metabolism of nucleic acids.

4.2 Results

4.2.1 Generation of the *rex1* H360A mutant: an active site mutant that may leave metal ion binding intact

An aim of this project was to observe the positioning of a substrate in the Rex1 active site using structural biology methods. Structural biology experiments can require large quantities of purified protein, particularly in the case of crystallography, and the enzyme needs to be expressed in a form where it will not process its substrate if the enzyme-substrate complex is to be preserved. To achieve both of these aims, most structural studies of DEDDh exonuclease domains ablate one of the DEDD residues, however this comes at the cost of losing one or both catalytic metal ions. These metal ions not only coordinate hydrolysis, but also influence the positioning of the substrate; the loss of bound metal ions in a structure leads to loss of substrate binding information content.

To enable structure determination of Rex1 with native-like substrate contacts intact, H360 was ab-

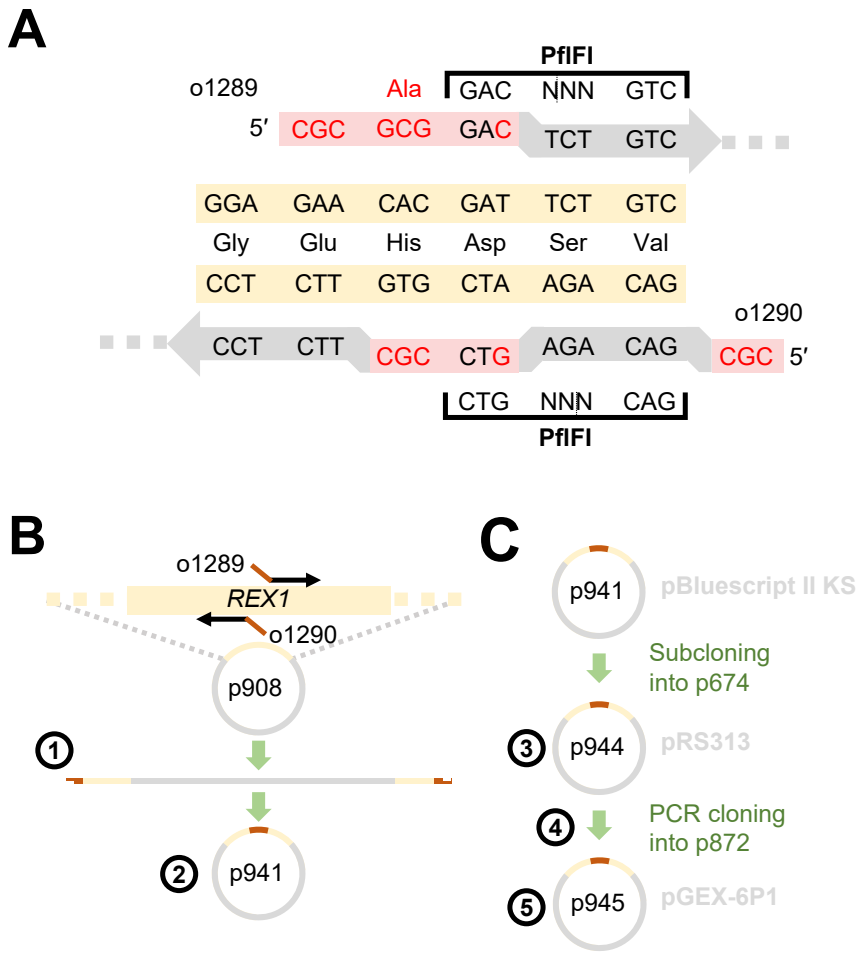


Figure 4.1: Cloning scheme for *rex1* H360A mutagenesis

A: Extended primer design for H360A mutagenesis. The resulting restriction site with cutting position is indicated. "N" describes a position that can be occupied by A, T, C, or G. Yellow double-stranded sequence represents *REX1* coding sequence spanning G358-V364. **B**: Flowchart explaining the use of mutagenesis primers. Primers o1289 and o1290 are shown amplifying the entire plasmid into a linear product. This linear product can then be ligated to produce a plasmid with the desired mutation. **C**: Summary of the steps taken to clone *rex1*H360A from a pBluescript II KS backbone into a yeast expression vector, pRS313, and an *E. coli* expression vector, pGEX-6P1. Circled numbers correspond to data shown in Figure 4.2.

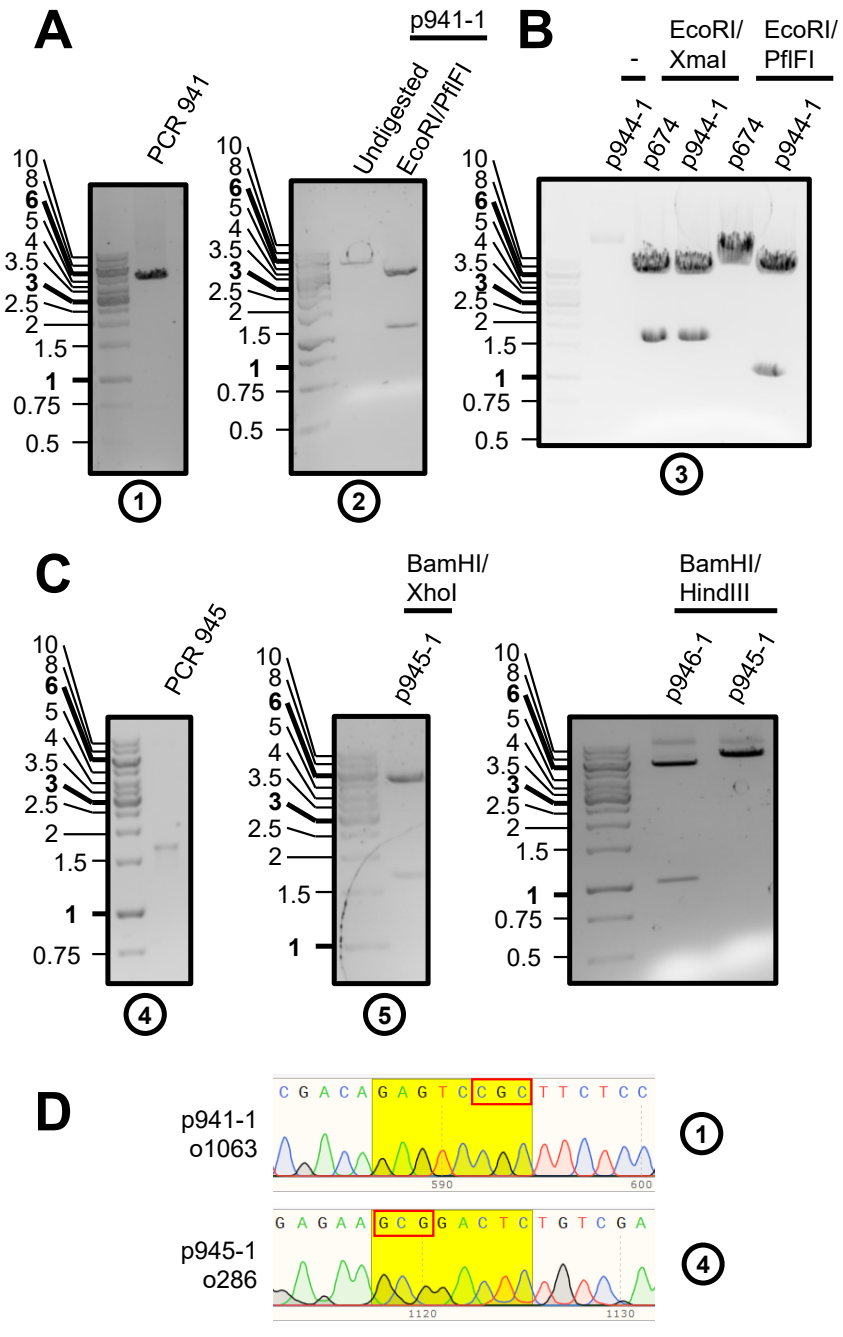

A
p941-1
PCR 941
Undigested
EcoRI/PflFI
B
-
EcoRI/
XmaI
EcoRI/
PflFI
p944-1
p674
p944-1
p674
p944-1
C
BamHI/
XhoI
BamHI/
HindIII
PCR 945
p945-1
p946-1
p945-1
D
p941-1
o1063
p945-1
o286

Figure 4.2: Confirming of *rex1* H360A plasmid construct cloning
Confirmatory molecular biology experiments to confirm successful generation of *rex1* H360A plasmid constructs. All gels shown are ethidium bromide stained 1% agarose 0.5x TBE gels. **A**: Extended PCR product "PCR 941" amplified from p908 using primers o1289/o1290. Successful introduction of a unique PflFI site is demonstrated by doubled restriction digest with EcoRI and PflFI. **B**: The EcoRI/XmaI fragment of p941-1 is subcloned into a wild type Rex1 yeast expression vector, p674. Successful replacement of the *REX1* insert is confirmed by double digest with EcoRI and PflFI. **C**: PCR product "PCR 945" is amplified from p944 using primers o1133/o1215 to generate a product with compatible restriction sites for an insert containing *E. coli* expression vector, p946. Successful regeneration of the BamHI/XhoI restriction sites is demonstrated. Successful exchange of the p946-1 insert for PCR 945 is confirmed by observing the loss of a HindIII restriction site. **D**: Presence of the full H360A-PflFI mutation described in Figure 4.1 is confirmed in the *rex1* ORF by Sanger sequencing, and highlighted in yellow, with the Histidine to Alanine GCG missense mutation indicated by a red box. The –ve strand sequence produced by the primer o1063 is shown for p941-1, and the +ve strand sequence produced by the primer o286 is shown for p945-1. Circled numbers correspond to positions in the cloning scheme outlined in Figure 4.1.

lated - the only residue of the 5 DEDDh residues involved in catalysis without contacting the substrate. This mutation has been previously observed to be sufficient to abolish *in vitro* processing of physiological RNA substrates in *E. coli* RNase T and *A. thaliana* SDN1 (Zuo and Deutscher, 2002a; Chen *et al*., 2018). To this end, we employed diverging primer mutagenesis on a plasmid-borne (pBluescript II KS; Alting-Mees and Short, 1989) copy of wild type *REX1*, as outlined in Figure 4.1A and B. A histidine to alanine codon substitution was introduced, alongside a neighbouring GAT to GAC silent aspartate mutation that introduced a unique PflFI restriction site to each end of the linear PCR product. This restriction site allows both re-circularisation of the plasmid, and served as a marker for screening for the presence of the mutation in down-stream cloning.

The pBluescript vector's short sequence suited it well to whole-plasmid PCR amplification (Alting-Mees and Short, 1989), however it was necessary to sub-clone the mutagenised *rex1* sequence into expression vectors, for *E. coli* pGEX-6P1 (Smith and Johnson, 1988), and for yeast pRS313 (Sikorski and Hieter, 1989) as outlined in Figure4.1C. The pGEX-6P-1 vector possesses a TRP-Lac promoter that allows induction of protein expression in *E. coli* by the addition of lactose, galactose, or IPTG (Isopropyl-β-D-1-thiogalactopyranoside) to the culture media, and allows expression of a protein fused at the N- terminus to a Glutathione-S-Transferase (GST) domain. This GST tag can be cleaved off at an 8aa 3C cleavage site (LEVLFQ/GP) using 3C-like proteases such as PreScission protease. The pRS313 vector features a CENP 'low copy number' replication origin and a HIS3 marker (Sikorski and Hieter, 1989), with the variant used here featuring the promoter sequence of yeast *RRP4* followed by an N-terminal zz-tag; the duplicated z domain from *Staphylococcus aureus* Protein A (Mitchell *et al*., 1996), and the 3′ UTR (untranslated region) of *REX1* (Hama Soor, 2017). This cloning scheme was successful, as evidenced in Figure 4.2 by Sanger sequencing for PCR steps, and diagnostic restriction digests for subcloning steps.

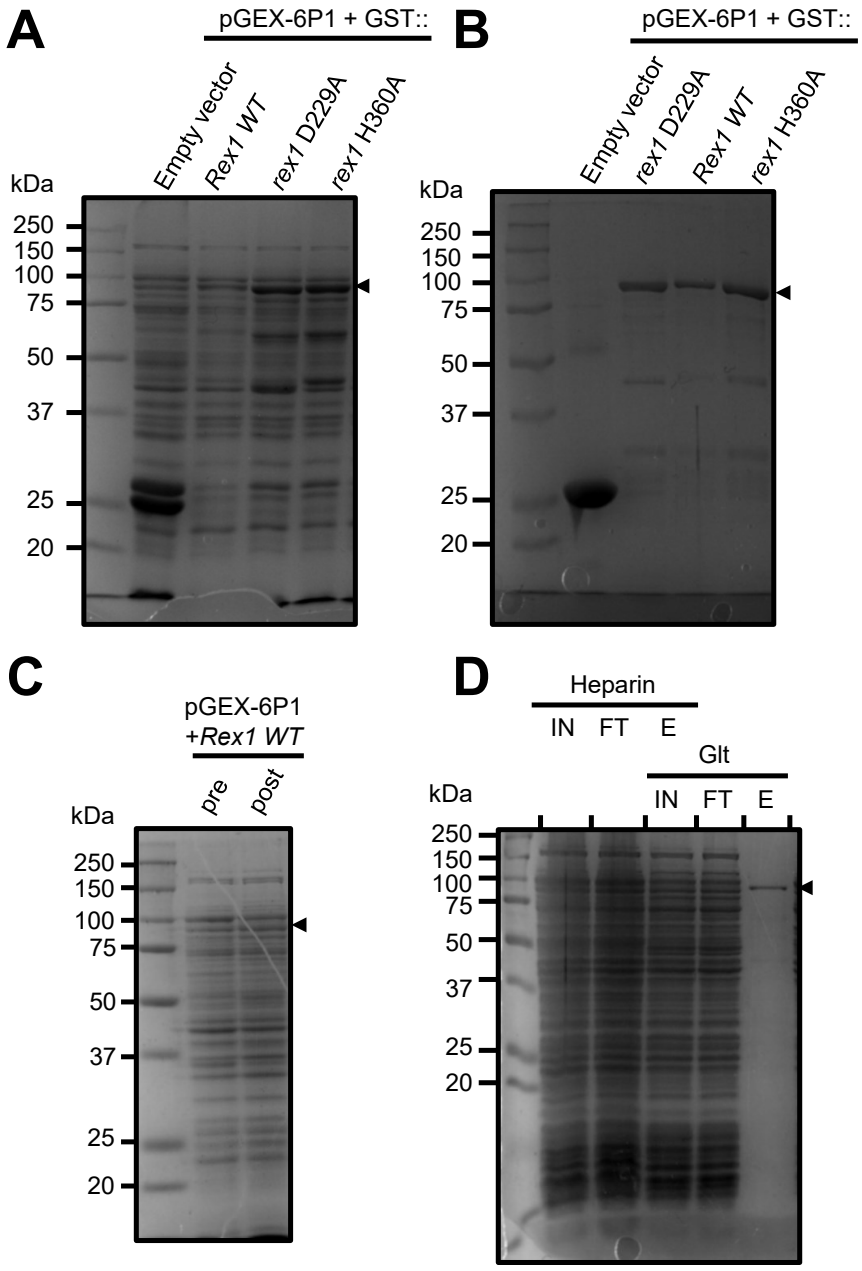


Figure 4.3: Autoinduction and purification of wild type, D229A and H360A Rex1

All panels show 12% denaturing acrylamide gels, imaged with Coomassie stain. The expected size of GST-Rex1 is 89kDa, indicated with a black triangle. **A**: Whole-cell lysates of BL21 *E. coli* autoinduction 100ml cultures. **B**: Eluates of small-scale GST pulldowns for Rex1 wild type, D229A, and H360A. **C**: Whole-cell lysate of BL21 *E. coli* autoinduction 500ml culture expressing GST-Rex1 wild type, equivalent cell mass taken before and after overnight incubation at 18°C. **D**: Two-step purification of GST-Rex1 from 500ml of *E. coli* autoinduction culture, with inputs ('IN'), flowthroughs ('FT'), and elutions ('E') shown. The following volumes were loaded: 2.5µl from 6ml of heparin bead input and flowthrough, 2.5µl from 1ml of heparin elution/ glutathione bead ("Glt") input and flowthrough, followed by 2.5µl from 100µl of glutathione elution.

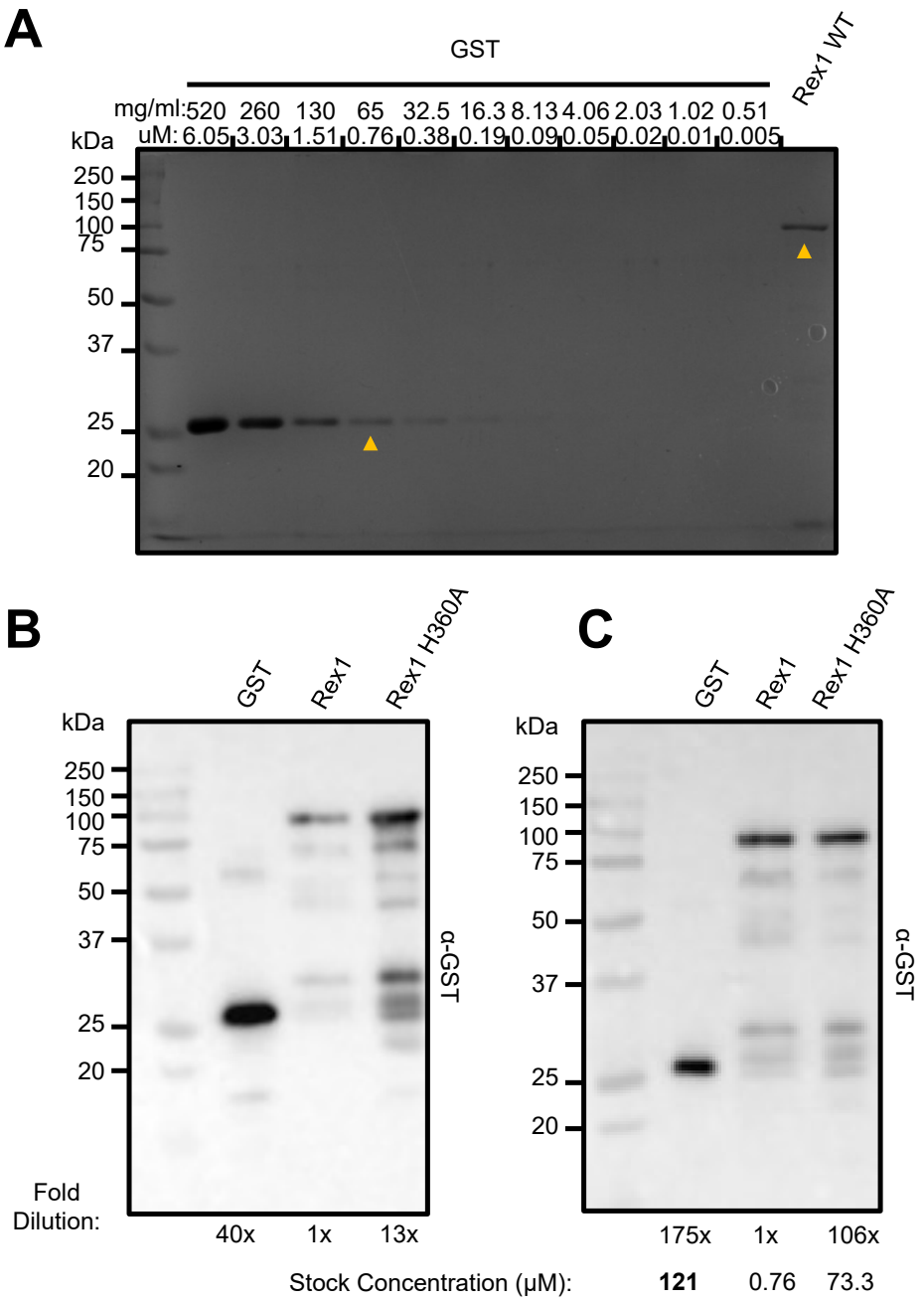


Figure 4.4: Analysis of GST-Rex1 purifications using Coomassie staining and Western blotting
A: Initial estimation of elution concentration for wild type and D229A Rex1 purifications by comparison with a 2-fold dilution series starting with 1/20 diluted GST purification. Concentration of GST in mg/ml calculated using absorbance at 280nm with an assumed extinction coefficient of 42860 (Expasy ProtParam of pGEX-6P1 protein product) with stock measured as 5.2mg/ml (121μM). Yellow arrows indicate the GST band that most closely resembled the Rex1 band. **B:** Western analysis of GST, GST-Rex1, and GST-Rex1 H360A. Demonstrates an initial dilution for each sample based on the band intensity of the Coomassie-stained gel A. **C:** Normalised protein concentrations prepared based on quantification of B multiplied by the fold dilutions. Concentrations for GST-Rex1 and GST-Rex1 H360A calculated based using the fold dilution and the stock GST concentration of 121μM are shown.

4.2.2 Recombinant expression and purification of wild-type, D229A and H360A GST-Rex1

We employed an autoinduction scheme for recombinant Rex1 expression (Studier, 2005). Conventional overexpression of a recombinant plasmid-borne gene in *E. coli* involves growing a flask culture to mid-logarithmic phase (0.5-1$OD_{600nm}ml^{-1}$), then inducing acute activation of the encoded TRP-Lac promoter through the addition of IPTG, a non-metabolisable analogue of allolactose (de Boer *et al.*, 1983). Autoinduction by comparison offers a less acute, more gradual expression induction process that requires no addition of the inducing compound during growth (Studier, 2005). Instead, multiple carbon sources are included in the culture media, which are metabolised in sequence throughout the culture's growth. The carbon sources used here are glucose, glycerol, and lactose. Upon the exhaustion of the glucose, the lac operon of the *E. coli* is derepressed allowing the allolactose product of metabolised lactose to activate the lac operon through binding the lac repressor protein lacl (Jobe and Bourgeois, 1972).

The biochemical analysis of recombinant Rex1 presented here requires the expression of three GST-tagged *Rex1* constructs: wild type, D229A, and H360A. The D229A mutant is a well characterised alanine substitution of the first aspartate in the DEDDh motif that was generated by previous PhD student Taib Hama Soor (Hama Soor, 2017; Daniels *et al.*, 2022), based on the D238A *rrp6* mutant (Phillips and Butler, 2003). The histidine-alanine substitution has been examined once before in SDN1 (H278A; Chen *et al.*, 2018), a related plant DEDDh exonuclease, and hasn't been examined directly in yeast Rex1. To ensure that the H360A mutant is catalytically inactive, an essential aim of this biochemical analysis is to compare the *in vitro* activity of GST-Rex1 H360A to GST-Rex1 wild type and GST-Rex1 D229A.

Autoinduction of 100ml *E. coli* cultures for the three GST-tagged *Rex1* constructs were performed alongside an empty vector GST control as shown in Figure 4.3A. In agreement with previous data in the Mitchell laboratory, the wild type *Rex1* construct was expressed far less than the active site mutants, which in turn had weaker expression than the robust expression of GST alone. The low expression of GST-Rex1 wild type may indicate a degree of toxicity to *E. coli* when overexpressed, leading to selection against retention of the expression plasmid. The resulting loss of the plasmid would lead to lower secretion of the plasmid-encoded ampicillin resistance protein, resulting in lower final cell density. This was anecdotally observed to be the case throughout this project- autoinduction cultures of wild type Rex1-expressing *E. coli* would typically reach $OD_{600nm}ml^{-1}$ endpoints of 1.5, whereas active site mutants would often reach endpoints of 3-5 OD_{600nm}.

Small-scale 1-step purification using glutathione beads were performed on native lysates for each of the cultures. The eluted products for each construct are shown in Figure 4.3B. Several bands that are smaller than the expected size of GST-Rex1 are visible, although their absence in the GST lane suggest these may be Rex1 proteolysis products. A further indicator is that the strength of these bands seems to correlate with the intensity of the full-length band; the lower bands are less intense in the comparably dilute wild type Rex1 preparation.

In order to purify larger quantities of GST-Rex1 wild type, a 500ml autoinduction culture was later prepared as shown in Figure 4.3C. The lysate of this culture was subjected to a 2-step purification scheme, drawing on later knowledge that Rex1 has an affinity for Heparin resin (see section 4.2.7), as demonstrated in Figure 4.3D.

The detergent TWEEN-20 was included at 0.1% concentration in the elution buffer of each of the four purified protein constructs with the aim of increasing the stability of the protein and reducing

adsorption to the interior surfaces of microfuge tubes. A consequence is that the detergent interferes with the estimation of protein concentration using the Bradford assay, requiring the use of an alternative measurement such as absorbance of UV light at 280nm (ABS_{280nm}). The reliable measurement of UV absorbance was only possible for the purified GST, which gave an estimated concentration of 121μM.

In order to extrapolate the measured GST concentration onto the other purified products, two SDS-PAGE (sodium-dodecyl-sulfate polyacrylamide gel electrophoresis)-based methods were used. The first method aimed for an initial estimation of GST-Rex1 wild type concentration by comparison of Coomassie staining to a GST dilution series as shown in Figure 4.4A. This allowed the initial estimation of GST-Rex1 wild type concentration as between 1.5 and 0.7μM.

The second method of concentration estimation used was the quantification of Western blot chemiluminescence, which has the advantage of calculating the concentration of the full length, non-cleaved Rex1. With GST-Rex1 wild type representing the lowest concentration of the constructs, the purified products of GST and GST Rex1 H360A were diluted 40-fold and 13-fold respectively, and were analysed alongside the GST-Rex1 wild type through Western blotting with an α-GST antibody in Figure 4.4B. This initial dilution did not result in equal chemiluminescence band intensities, however the full-length bands of the non-merged image could be quantified using ImageJ (Schneider, Rasband and Eliceiri, 2012) to calculate a more accurate dilution. The resulting dilutions of GST and GST-Rex1 H360A, 175-fold and 106-fold respectively, resulted in a visibly equal signal shown in Figure 4.4C. Taken with the measured GST concentration this indicated stock concentrations of 0.76μM for GST-Rex1 wild type, and 73.3μM for GST-Rex1 H360A. The concentration of GST-Rex1 D229A was not estimated in this manner, but based on the similar Coomassie stain intensity visible in Figure 4.3B, it can be assumed to be similar to GST-Rex1 H360A.

The Western blotting data shown in Figure 4.4 provides information beyond concentration estimation. There is a close correlation visible between Coomassie-stained banding patterns visible in Figure 4.3 and Figure 4.4 to the pattern visible by Western blots of Figure 4.4. This lends further credence to the identification of the reproducible smaller banding pattern as proteolysis products, with reactivity to the α-GST antibody indicating they still possess a GST-tag at their N-termini.

4.2.3 Measuring Rex1 activity by the degradation of DNA oligos

We next examined the ability of recombinant Rex1 to process short DNA oligos 5´-labelled with ^{32}P phosphate. This assay is routinely used in the literature, and a specific mixed sequence substrate had been used in the lab prior to this project ('o1165' Hama Soor, 2017; Daniels *et al.*, 2022). This substrate was incubated at 30ºC over a 1-hour time course with purified GST, GST-Rex1wild type, GST-Rex1 D229A, and GST-Rex1 H360A. Samples were quenched with the addition of 1:1 volume formamide loading buffer before resolution through denaturing 16% polyacrylamide gels, as shown in Figure 4.5A. Of these purified proteins, only GST-Rex1 wild type demonstrated exonuclease activity, with an absence of detectable activity in the GST and GST-Rex1 D229A indicating undetectable levels of co-purified exonuclease activity. The lack of exonuclease activity for Rex1 H360A indicates that mutation of the DEDDh histidine to alanine was sufficient to abolish exonuclease activity under these conditions, in agreement with examples from the literature (Zuo and Deutscher, 2002a; Chen *et al.*, 2018). An additional assay was performed using GST, GST-Rex1 H360A, and GST-Rex1 wild type that had been diluted to a standardised final reaction concentration of 25nM, as shown in Figure 4.5C, which revealed

4-fold less activity for 10-fold less wild type Rex1 concentration.

The exonuclease activity demonstrated by wild type GST-Rex1 in Figure 4.5 confirm the predicted characteristics of Rex1: no 5´ exonuclease activity is detectable, which would result in loss of the 5´-radiolabelled signal instead of the laddering effect seen. The single nucleotide spacing of this ladder suggests that GST-Rex1 is behaving as a distributive exonuclease as opposed to an endonuclease or processive exonuclease, which would manifest as conversion of full-length substrate to a defined end product as seen for example for Rrp44 as part of the exosome complex (Mitchell *et al*., 1997; Das *et al*., 2021),

The mixed-sequence substrate revealed an uneven processing rate for wild type Rex1, indicating a degree of sequence and/or structure specificity for Rex1's exonuclease activity. With this in mind, a 5´-radiolabelled 18-thymine substrate (oligo dT) was used, as shown in Figure 4.6. This substrate demonstrated a faster and more even degradation profile, to the point where an endpoint nucleotide length of 4 nucleotides could be observed. This is a common observation for exonucleases, with specialised oligoribonucleases breaking down these otherwise non-metabolisable end products (Niyogi and Datta, 1975; Ghosh and Deutscher, 1999).

To isolate more specifically the effect of sequence composition on the inhibition of Rex1 exonuclease activity, we prepared two further substrates featuring pairs of cytidines, shown in Figure 4.7. It was expected that CC pairs may inhibit activity in a similar manner to the RNase T 'C effect' (Zuo and Deutscher, 2002c), a mechanism intended to protect the functional CCA sequence at the 3´ end of tRNAs. This activity took an unexpected form for Rex1: a minor accumulation of certain length substrates was observed, but the lengths corresponded to a consistent CCTTT-3´ end sequence instead of the expected CC-3´ as seen for RNase T and Rrp6 (Axhemi *et al*., 2020; Zuo and Deutscher, 2002c). The inhibitory behaviour of substrate's sequence seemed to be able to mediate its effect 4 nucleotides downstream of the second C. With this 4 nucleotide footprint matching the size of the end product, this hints that when processing simple nucleotide substrates that Rex1's substrate recognition mechanisms may engage the last 4 nucleotides. This observation is consistent with the substrate-threaded structural models shown in Figure 3.9, which feature RNA-binding residues that only predict interactions with the last four nucleotides. This observation resembles a recent analysis of substrate selectivity in Rrp6 (Axhemi *et al*., 2020), which confirmed a previous structural model of Rrp6 in complex featuring an RNA-binding footprint in the active site of 4 nucleotides (Wasmuth *et al*., 2014).

In RNase T, the mechanism of the C effect is mediated by four conserved phenylalanines and a glutamate: F29, E73, F77, F124, and F146 (Hsiao *et al*., 2011). These residues interact with the two 3´ C nucleotides in a manner that positions the scissile phosphodiester bond away from the active site, and reorients the general base histidine into an inactive conformation (Hsiao *et al*., 2011). If this conformational change is conserved in Rex1, it may be the case that the responsible residues interact with C nucleotides found three nucleotides before the 3´ terminal nucleotide to mediate this effect on a DNA substrate.

5´-^{32}P-labelled substrates have a half-life of two weeks, and are undetectable after 6 months. We used 5´-Cy3 labelled oligo dT as a more stable alternative substrate, which could be processed by GST-Rex1 wild type to give the degradation profile shown in Figure 4.8C. A key difference between these substrates is readily apparent; while the ^{32}P-labelled substrate could be processed to a 4nt end product, the Cy3-labelled substrate could only be processed to a 9nt end product. The Cy3 dye molecule itself is

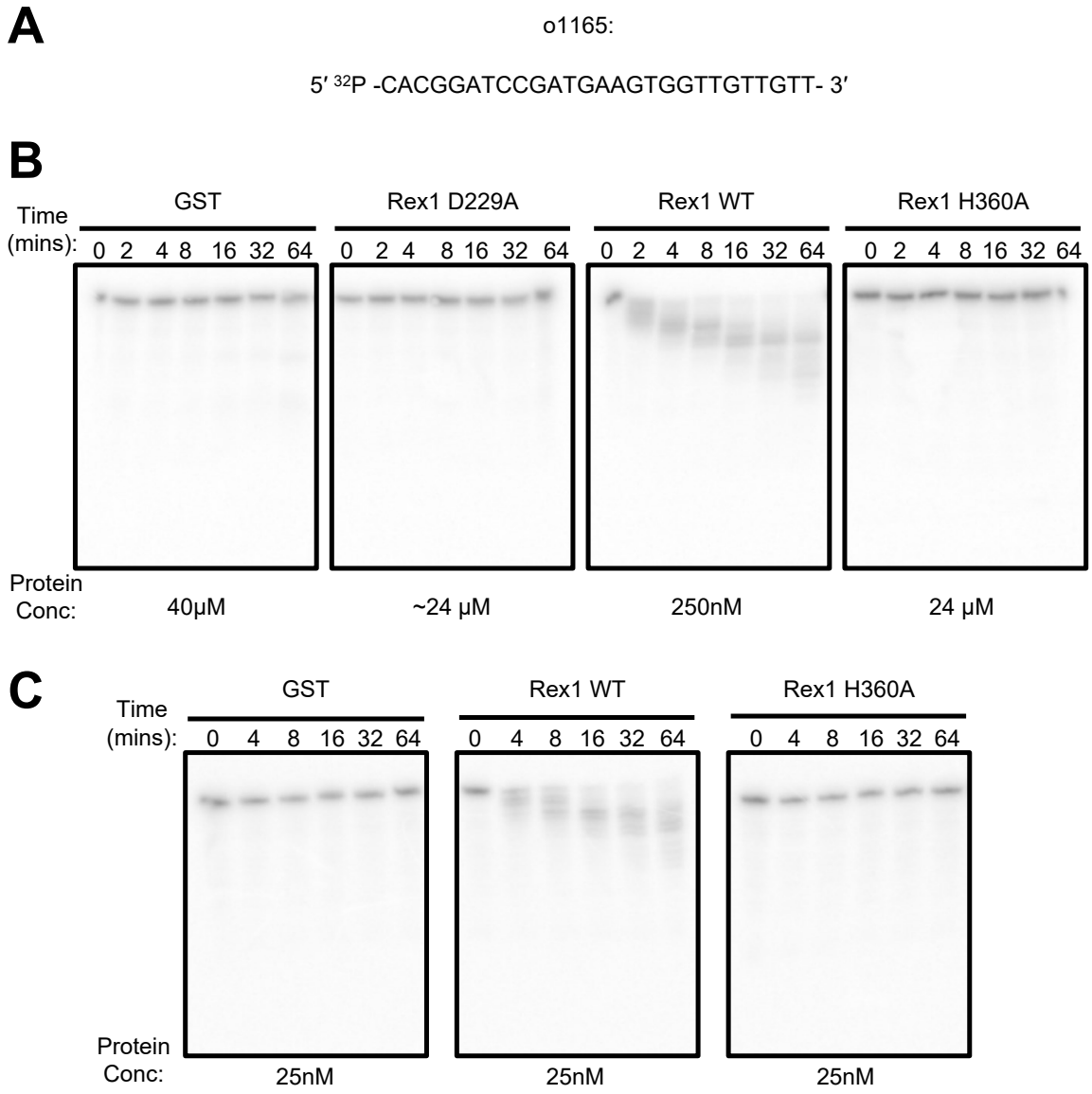


Figure 4.5: *In vitro* turnover of ^{32}P-labelled mixed-sequence DNA oligos by recombinant Rex1 **A:** Sequence composition of 5´- ^{32}P-labelled DNA oligo. **B:** PhosphoImaged blots of 1-hour degradation time course. Proteins were used at the concentrations they were purified at, and incubated with 3.3nM substrate. **C:** PhosphoImager scans showing 1-hour turnover time course using a standardised protein concentration with 3.3nM substrate.

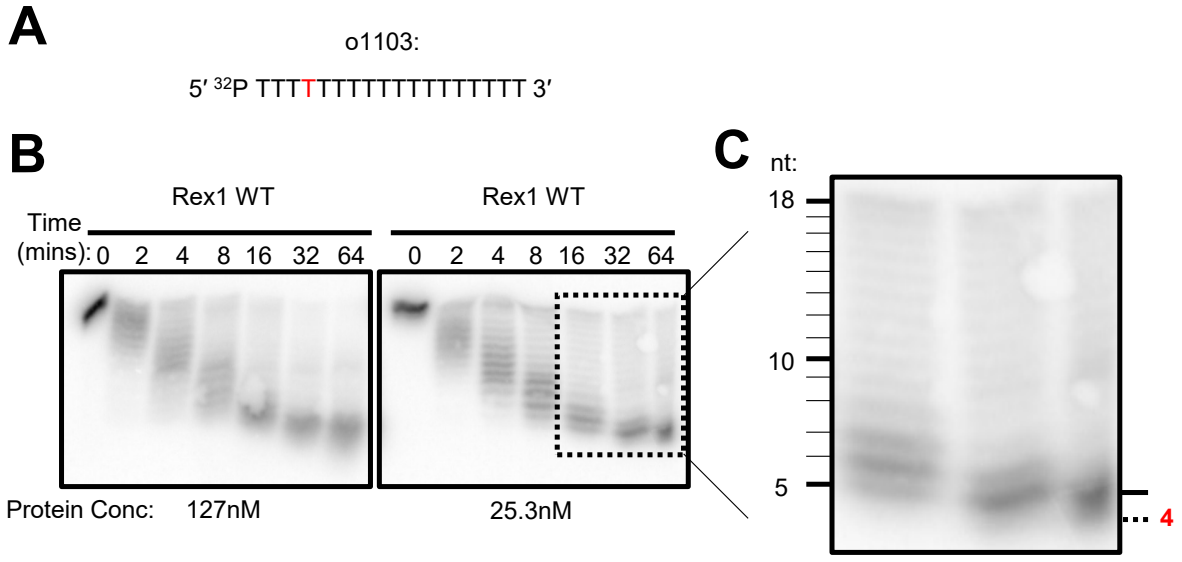


Figure 4.6: Degradation of ^{32}P-labelled oligo dT to a 4 nucleotide endpoint
A: Sequence composition of 5´-^{32}P-labelled oligo substrate, with measured degradation endpoint highlighted in red. **B:** PhosphoImaged blot of degradation time course for 127nM or 25.3nM Rex1 incubated with 3.3nM substrate at 30°C. Dotted box selection magnified in next panel. **C**: Magnified degradation pattern of o1103 to a measured 4 nucleotide endpoint.

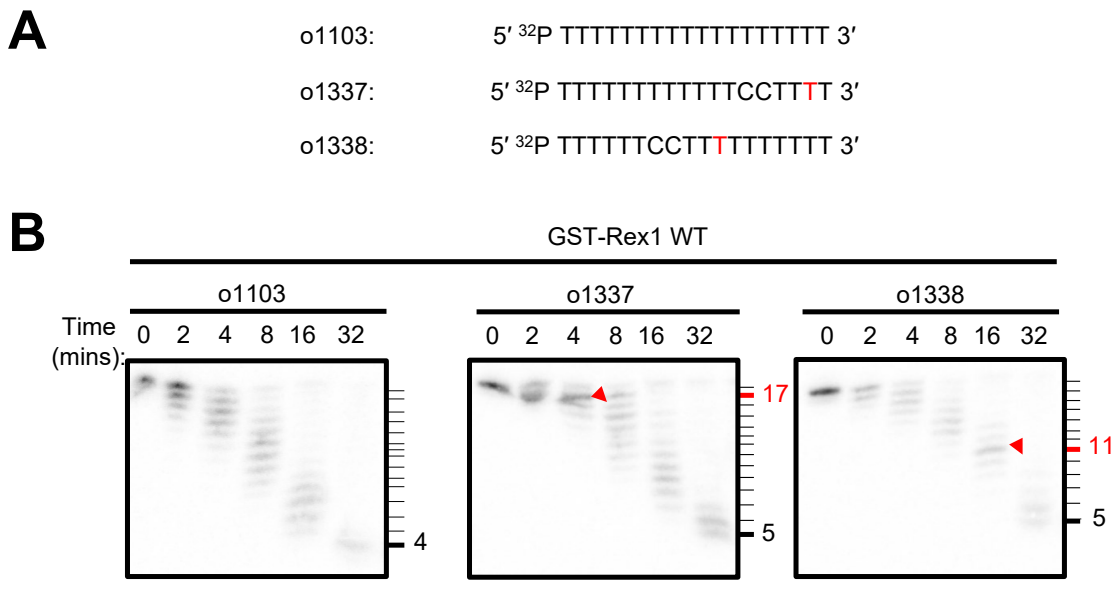


Figure 4.7: Impaired degradation of 5´ ^{32}P-labelled oligo dT by cytosine pairs.
A: Sequence composition of 5´ ^{32}P-labelled DNA oligo substrates. Red nucleotides indicate positions where degradation was stalled during the time course. **B:** Degradation time course of 25.3nM Rex1 incubated with 3.3nM substrate at 30°C. Positions with higher occupancy compared to the non-CC containing control are indicated with a red arrow, with the length of the substrate indicated in red on the ladder.

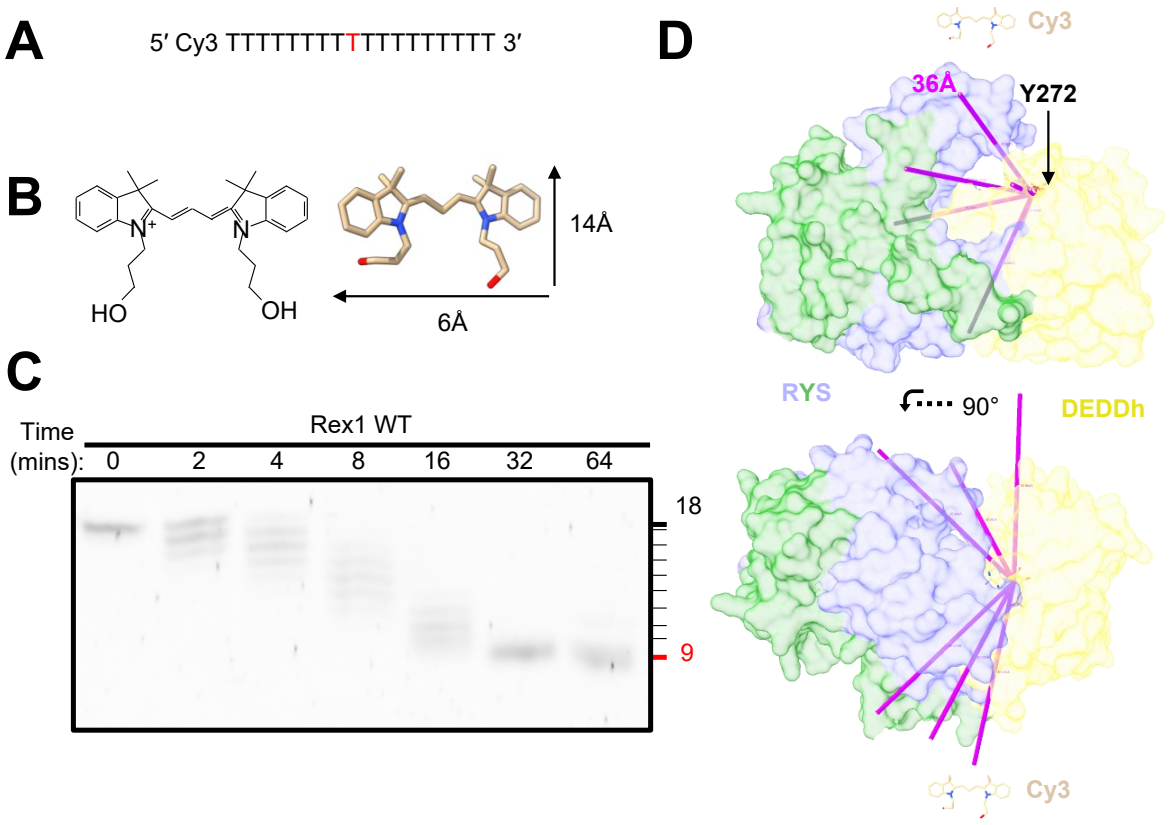


Figure 4.8: Impaired degradation of oligo dT by a 5´ Cy3 label.
A: Sequence composition of 5´ Cy3-labelled DNA oligo substrates. Red nucleotides indicate positions where degradation was stalled during the time course. **B:** Lewis diagram of Cy3 generated in PowerPoint demonstrating positive charge, alongside a ChimeraX model based on electron density from the crystal structure 5NS4. **C:** Degradation time course of 25nM Rex1 incubated with 3.3nM substrate at 30°C. The length of the substrate endpoint is indicated in red on the ladder. **D:** ChimeraX-generated model of a 9 nucleotide oligo length relative to the Rex1 Alphafold DB model. Rex1 shown as a surface model. 9 nucleotides were estimated as roughly 36Å based on crystal structure 6R9J (Tang *et al.*, 2019), which features yeast Pan2 in complex with polyA. Each 36Å ruler (purple) diverges from Y272- the residue predicted to base-stack with the 3´ base. Scale model of Cy3 is shown at the top right.

positively charged, and seems to be a common source of artefacts in the literature. In most cases, this arises from a characterised ability to base-stack with base-paired nucleotides (Norman *et al.*, 2000; Iqbal *et al.*, 2008). In the context of this substrate, base pairing between or within oligos is unlikely. A more likely explanation is that the charge of the Cy3 molecule may interact with features of the Rex1 binding cleft, potentially mispositioning the substrate for degradation. The modelled path length of a substrate through the two solution accessible clefts of the Alphafold2 model roughly match the length of a 9nt substrate, estimated here as 36Å as shown in Figure 4.8D.

The buffer conditions used for the exonuclease assays demonstrated measurable activity for GST-Rex1 wild type. We next aimed to establish the buffer and temperature parameters of the assay. Using the Cy3-labelled oligo dT substrate, we varied various reaction conditions as shown in Figures 4.9. Increasing NaCl concentration from 150mM reduced Rex1 activity as expected, with slightly reduced Rex1 activity observed at 300mM NaCl, and no Rex1 activity detectable at 600mM. For $MgCl_2$ concentration, highest activity was seen with 1.5mM, although it is possible that higher $MgCl_2$ concentrations

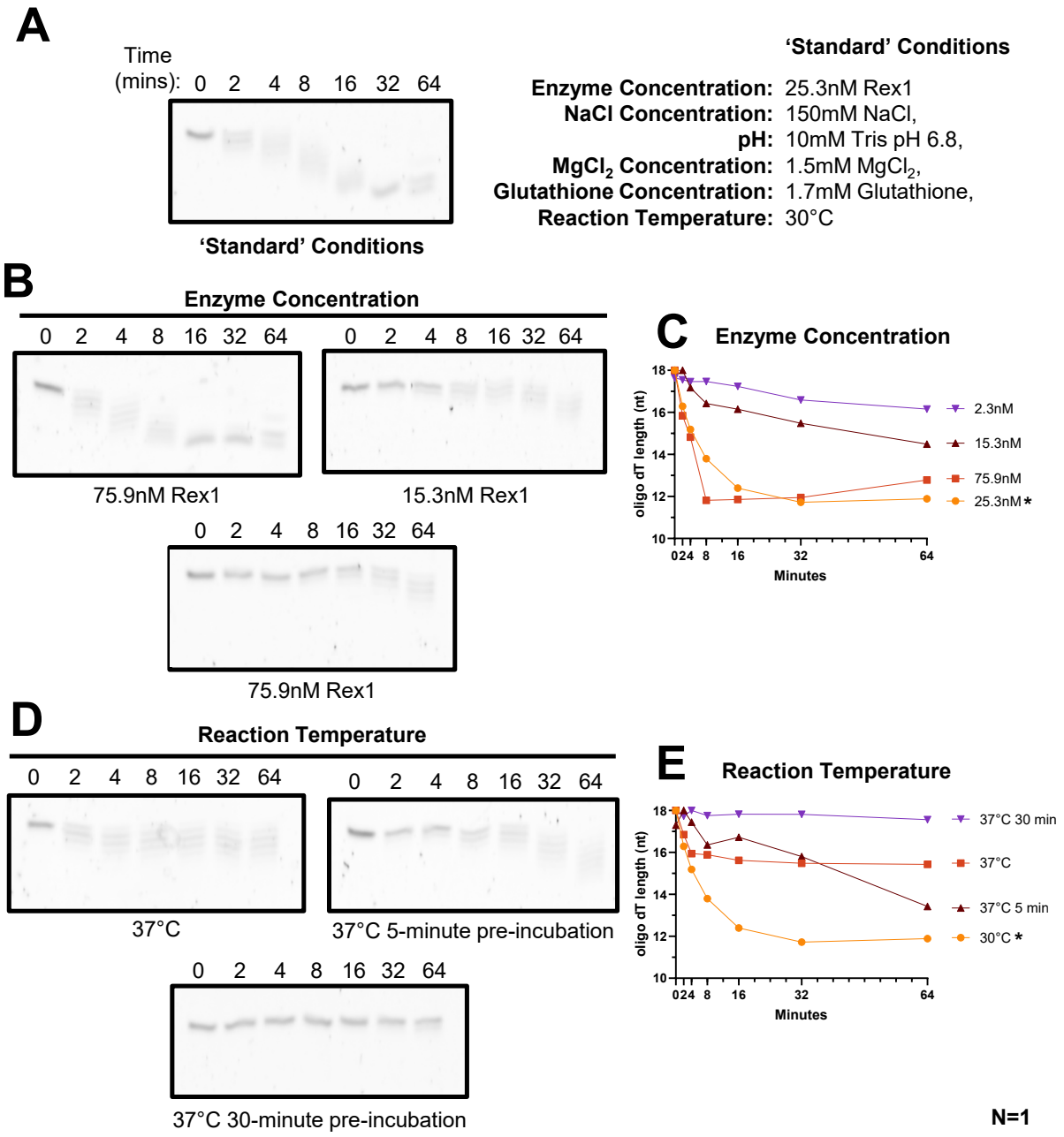


Figure 4.9: Preliminary optimisation of Rex1 exonuclease assay conditions using a 5´ Cy3-labelled oligo dT substrate

All assays use 3.3nM of Cy3-labelled substrate, and a single batch of purified Rex1. Each time course was performed once, using aliquots from a single preparation of GST-Rex1. Standard reaction conditions are described in A, with each reaction condition variable that differed from **A** indicated below each assay. All time points are measured in minutes. Band intensity for each nucleotide length was measured using SAFA software (Semi-automated footprint analysis; Das *et al.*, 2005), allowing calculation of average oligo length at each time point. Plots show oligo length over time. Each condition is quantified and plotted alongside the 'standard' condition assay in Figure 4.9, shown in each case as a yellow line and indicated with '*'. **B**,**C**: Reactions prepared using varying concentrations of GST-Rex1. **D**,**E**: Reactions were either incubated at a constant indicated temperature, or pre-incubated at 37ºC for an indicated time before substrate was added. **F-M:** Buffer components varied through the preparation of corresponding 10x reaction buffers.

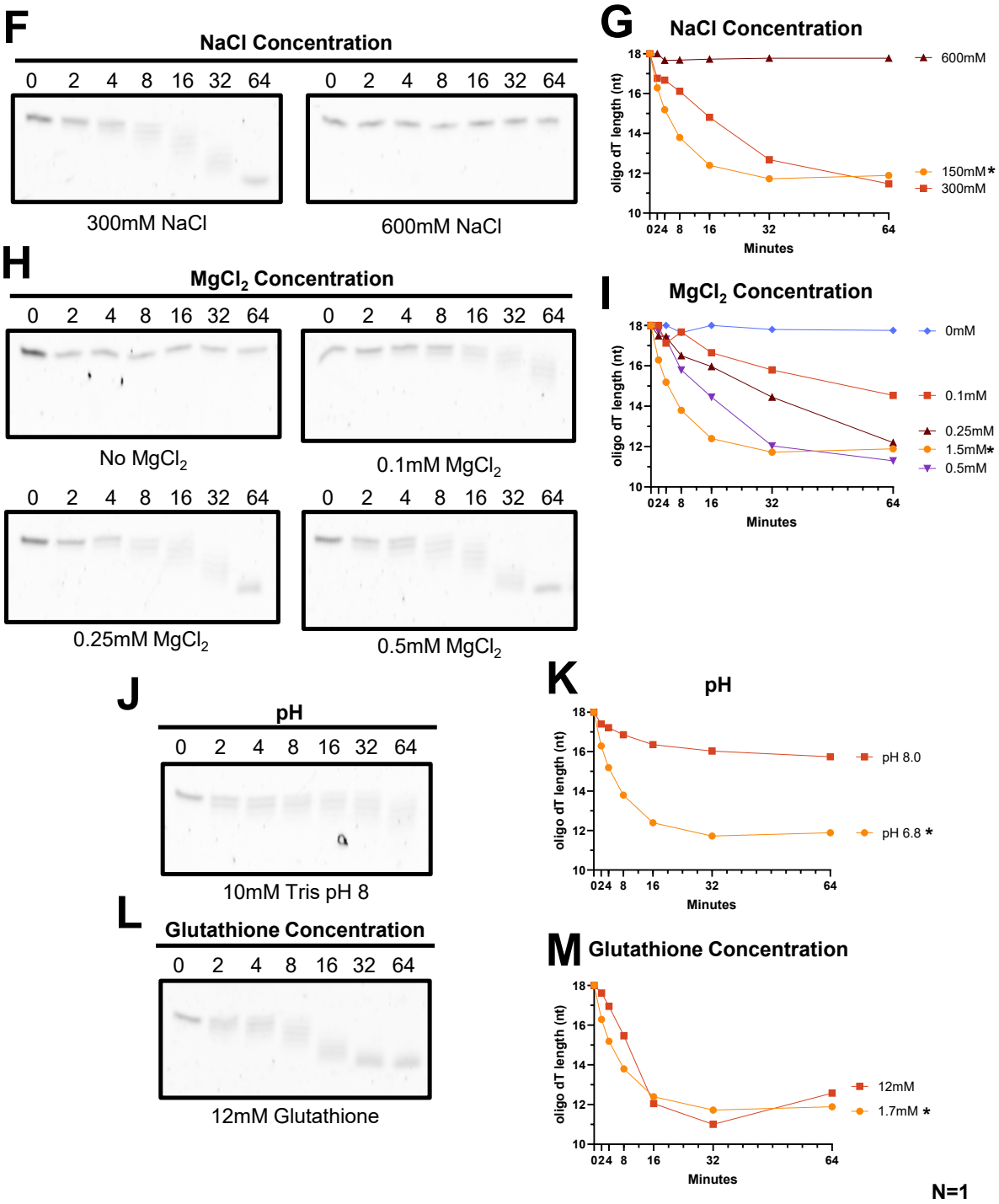


Figure 4.9 continued: Preliminary optimisation of Rex1 exonuclease assay conditions using a 5´ Cy3-labelled oligo dT substrate

may enhance activity further. No Rex1 activity is observed without supplemental $MgCl_2$; it is likely the case that no $MgCl_2$ cofactor was co-purified. A consequence of the purification method used is that glutathione accompanies the eluted protein. The concentrations of glutathione examined, 1.2mM and 12mM, did not cause a concentration-dependent inhibition of Rex1 activity, and may confer benefit to the assay by providing a reducing environment. It was expected that an alkaline pH would be required for Rex1 activity, as structural study of *E. coli* RNase T demonstrates a requirement of hydrolysis is de-protonation of the DEDDh histidine general base (Hsiao *et al.*, 2011). While Rex1 activity could be detected at pH 6.8, increasing pH to 8.0 seemed to result in a sharp reduction in activity by the 8 minute time point, indicating inactivation of the enzyme during the time course. If this result is confirmed with further replicates, it would suggest a very narrow range of pH in which recombinant Rex1 is active.

In addition to varying buffer components, we also observed Rex1 activity at 37ºC. An initial assay incubated at 37ºC demonstrated an activity that halted after 5 minutes. To expand on this observation, we also attempted 5 or 30 minute pre-incubations at 37ºC before shifting to 30ºC upon the addition of substrate, demonstrating that a reduced level of activity could be observed after 5 minutes at 37ºC, but a total loss of activity was observed after 30 minutes at 37ºC. In light of the temperature instability of this batch, it could be that buffer components may confer a protective effect. A common additive to this end is BSA, but higher salt concentrations have been shown to increase the thermal stability of RNase T (Deutscher *et al.*, 1984). Taken at face value, it seems that this preparation of recombinant Rex1 is destabilised at 37ºC. This assay requires further replicates to rule out instability arising from batch effects, but the instability of Rex1 at high temperatures agrees with unpublished observations in the Mitchell lab that endogenous Rex1 seems to be degraded *in vivo* for cultures incubated at 39ºC (Levicky and Mitchell, unpublished observations).

4.2.4 Generation of an NLS-truncated *rex1* H360A construct with and without a 6xHis-tag for crystallisation trials

Previous work in the Mitchell lab attempted to crystallise full-length recombinant Rex1 D229A that had been purified using a cleaved GST tag, however crystals were not forthcoming. In order to improve odds of successful crystallisation, we aimed to use the Alphafold2 model shown in Figure 3.6 to design an expression construct more likely to succeed. This model suggests that residues 1-51 of Rex1 form a disordered region, which may interrupt the regular crystal packing that is required for crystal formation. Two truncation mutants were designed to this end: the 1-51Δ truncation mutant was designed to remove the full length of the disordered region as predicted by the Alphafold model, and the 1-47Δ truncation was designed to give 5 extra amino acids of leeway to account for potential modelling errors.

When studying a protein using structural biology, it is preferable to try and purify proteins without a tag, using a short tag, or with a tag that can be cleaved off using a protease. With this in mind, for each truncation a primer was designed to amplify a tag-less open reading frame (ORF) or an N-terminally 6xHis-tagged ORF. Each Rex1 construct featured the H360A mutation to preserve native-like cofactor binding, as discussed in section 4.2.1. This cloning of this tagged construct into pRSETb (Schoepfer, 1993) is outlined in Figure 4.10, with confirmatory restriction and Sanger sequencing reactions performed, but not shown.

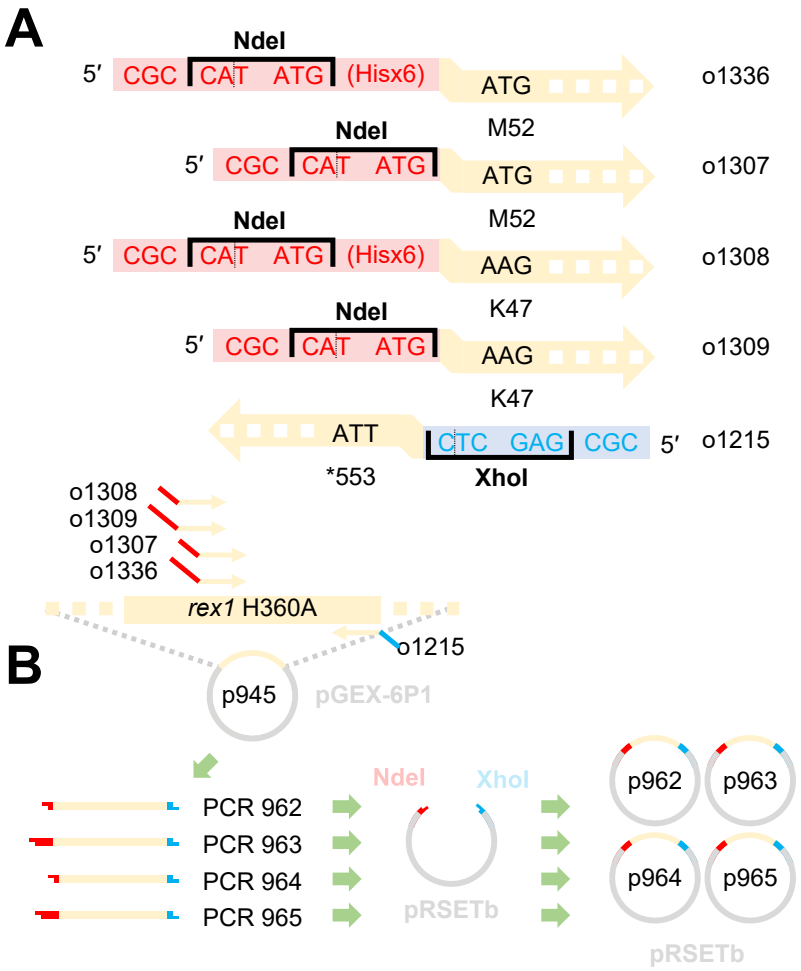


Figure 4.10: Cloning scheme for generation of Rex1 crystallisation constructs.
A: Extended primer design for cloning of truncated *rex1* into pRSETb, with and without a primer-encoded 6xHis tag. **B:** Flowchart outlining the use of extended primers to clone *rex1* truncations into pRSETb.

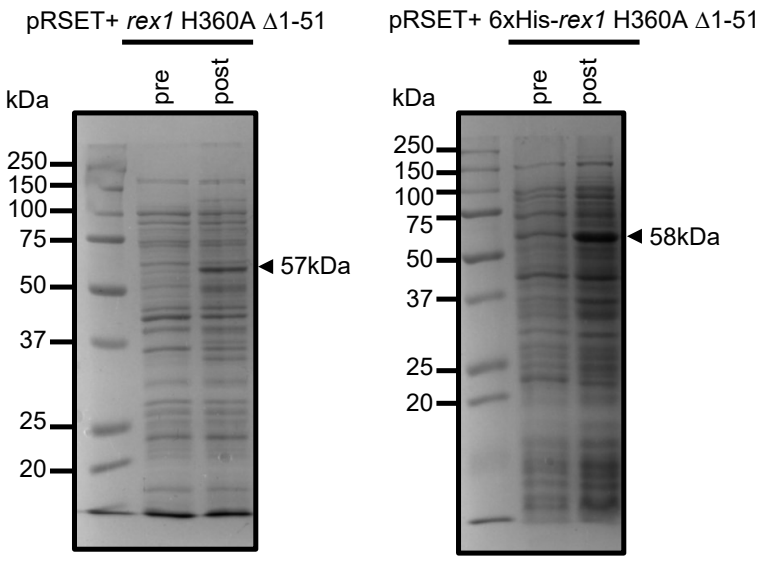


Figure 4.11: Autoinduction of Rex1 crystallisation constructs
Both panels show 12% denaturing acrylamide gels, imaged with Coomassie stain. Expected sizes for each Rex1 mutant are indicated with a black arrow. Whole-cell lysate representing BL21 *E. coli* autoinduction 2L cultures. Equivalent cell mass taken before and after overnight incubation at 18°C. Each lane pair representative of four 500ml cultures.

4.2.5 Expression and purification of Rex1 H360A Δ1-51 with and without 6xHis-tag

Two of the four *rex1* crystallisation constructs generated were used during this project: the untagged and 6xHis-tagged *rex1* H360A 1-51Δ. Both constructs were transformed into BL21 *E. coli* to give 2 litres of culture each, and demonstrated visible overexpression as shown in Figure 4.11. His-tagged Rex1 seemed to demonstrate the highest expression of the two, with untagged Rex1 showing comparatively lower expression.

Before attempting large-scale purification, small-scale bead purifications were attempted in order to establish the resin-binding properties of Rex1, as shown in Figure 4.12. Of the resins tested, Rex1 demonstrated a clear affinity for heparin, which was expected given heparin's known bias towards binding nucleic acid-binding proteins. Nickel affinity purification of pooled heparin fractions seemed to offer limited success, with the first elution's band distribution closely resembling the input. A more usual purification approach is to begin with affinity purification then use fractionation, however the reverse order was used in Figure 4.12D and E. This purification order was used to account for the extremely low solubility of Rex1 upon elution from nickel, an example of which is shown in 4.12F.

Large scale purifications of Rex1 H360A 1-51Δ with and without a 6xHis tag were performed by Svetlana Sedelnikova at the University of Sheffield's protein purification facility, as shown in Figure 4.13 and 4.14. The untagged construct was purified first, with the hope that the enrichment offered by the heparin column may be sufficient to overcome the low expression level. The end result however was a relatively heterogeneous preparation, lacking a monodisperse peak at gel filtration. As bands were visible both above and below the target band size by SDS-PAGE Coomassie staining, it is likely these bands represent co-purifying contaminant proteins. Throughout the purification, a solubility limit of roughly 3mg/ml was evident, with each significant precipitation event leading to a loss of enrichment for the 57 kDa Rex1 band compared to the background. This low apparent solubility limit of Rex1 was not

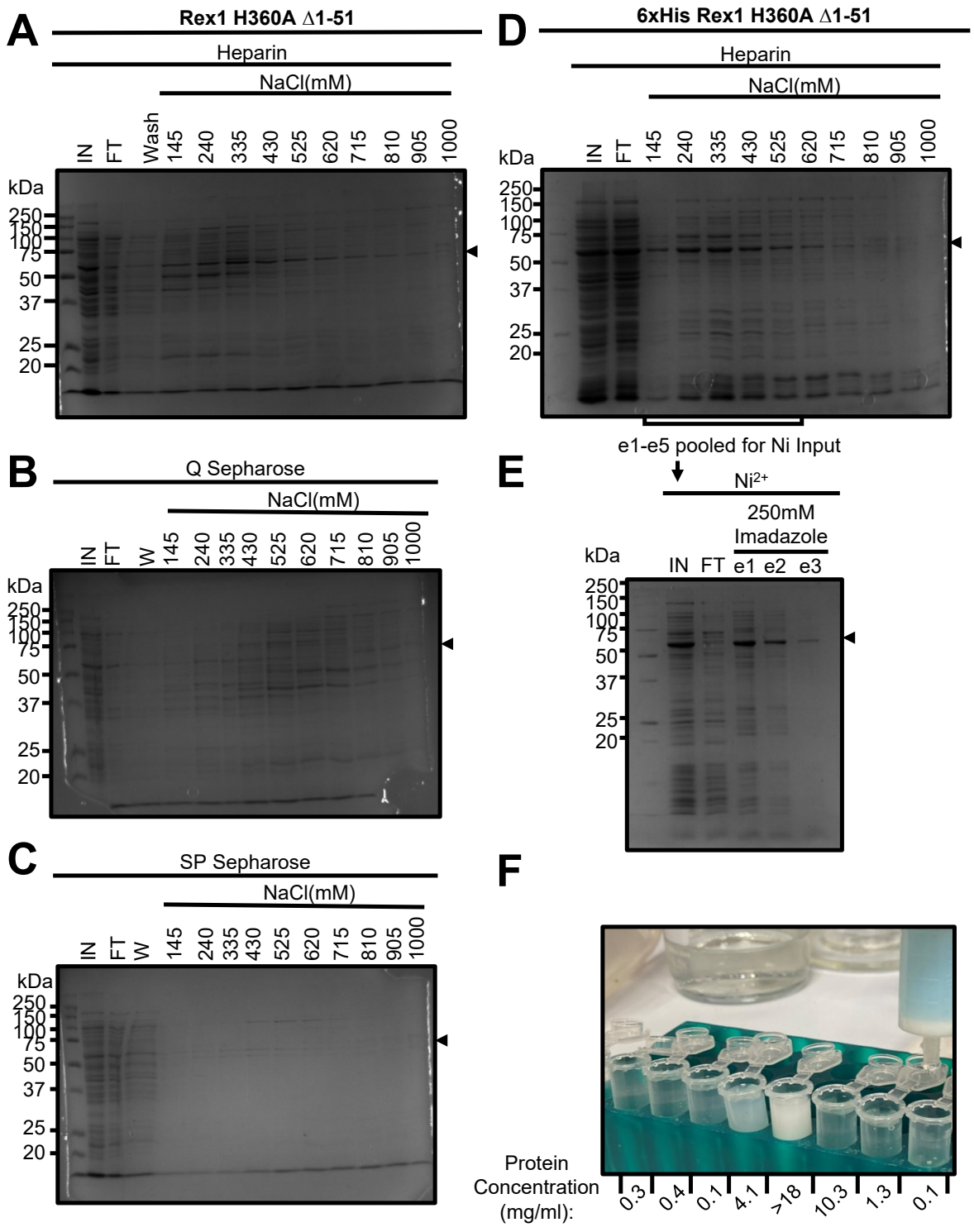


Figure 4.12: Small scale purification optimisation of Rex1 crystallisation mutants
All panels show 12% denaturing acrylamide gels, imaged with Coomassie stain. The expected size of Rex1 mutants are indicated with a black triangle. **A-C:** Fractionation of Rex1 H360A Δ1-51 –expressing *E. coli* 1ml lysates using 100ul of Heparin, Q Sepharose, or SP Sepharose beads. **D,E:** 2-step fractionation of 6xHis Rex1 H360A Δ1-51-expressing *E. coli* 1ml lysate using 100ul Heparin beads, then 100ul Ni^{2+}-NTA beads. **F:** Photograph of 6xHis Rex1 H360A Δ1-51 precipitation during a later large-scale Ni^{2+}-NTA pulldown. Concentrations measured using Bradford assay. Key: Ins = insoluble fraction of lysate, CFE = cell free extract, FT = flowthrough, e = elution fraction, GF = gel filtration, AS = ammonium sulphate precipitated protein.

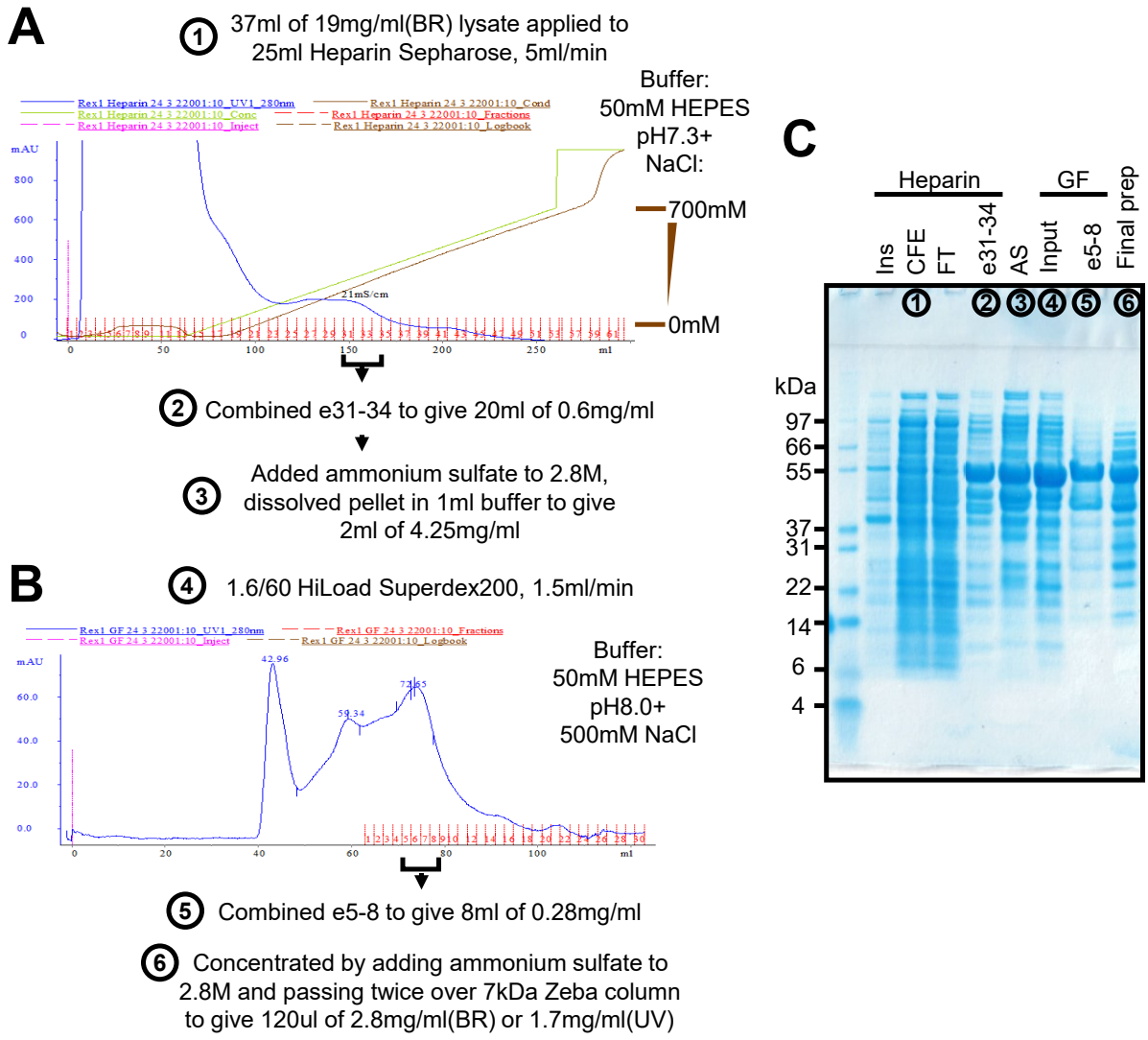


Figure 4.13: AKTA purification of untagged Rex1 1-51Δ

All data shown in this figure was collected by Sveta Sedelnikova. Purification scheme featuring circled numbers corresponding to lanes in the Coomassie-stained SDS-PAGE shown in C. **A:** AKTA purifier trace demonstrating fractionation of Rex1 1-51Δ by heparin affinity. BR = Bradford assay for protein concentration. **B:** Downstream size-exclusion chromatography of pooled heparin fractions. **C:** SDS-PAGE analysis of protein samples taken at the points in the purification schematic shown in A and B.

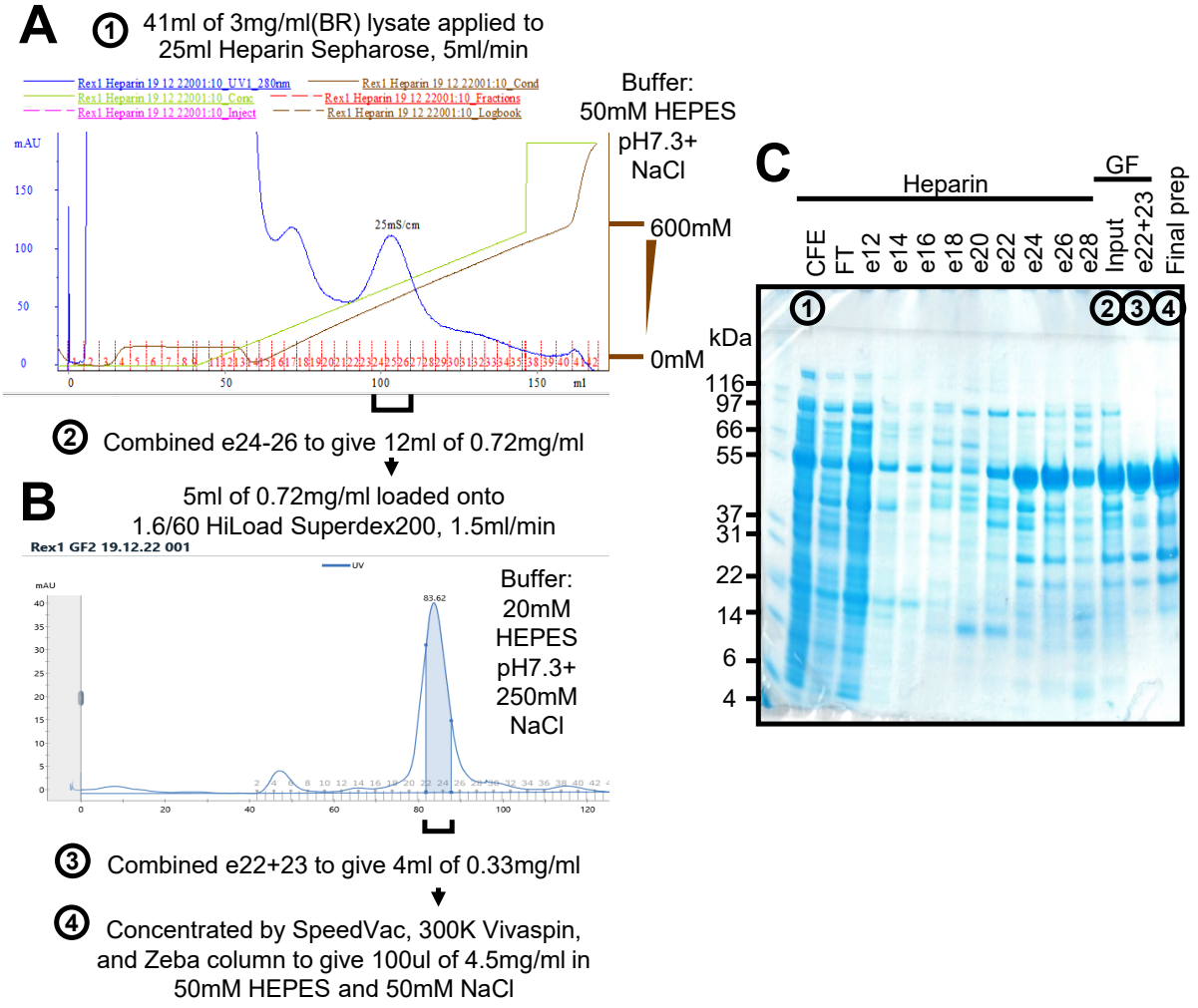


Figure 4.14: AKTA purification of 6xHis-Rex1 1-51Δ

All data shown in this figure was collected by Sveta Sedelnikova. Purification scheme featuring circled numbers corresponding to lanes in the Coomassie-stained SDS-PAGE shown in C. **A:** AKTA purifier trace demonstrating fractionation of Rex1 1-51Δ by heparin affinity. BR = Bradford assay for protein concentration. **B:** Downstream size-exclusion chromatography of pooled heparin fractions. **C:** SDS-PAGE analysis of protein samples taken at the points in the purification schematic shown in A and B.

Table 4.1: Crystallisation screening of 6xHis-1-51ΔRex1H360A and 1-51ΔRex1H360A
Crystallisation screens seeded for recombinant preparation of 1-51ΔRex1H360A (Figure 4.13) and 6xHis-1-51ΔRex1H360A (Figure 4.14) without or with the addition of 5mM $MgCl_2$ and 5mM AMP. Dates in combination with the screen name and Rex1 construct serve as unique identifiers for each crystallisation tray. JCSG, PACT, $AmSO_4$, and ProPlex screens used pre-made from Molecular Dimensions, 'Custom 1' and 'Custom 2' screens were prepared using a Formulatrix Formulator as described in Figure 4.15.

Screen:	1-51ΔRex1H360A	6xHis-1-51ΔRex1H360A	6xHis-Rex1 H360A 1-51Δ + 5mM $MgCl_2$, 5mM AMP
JCSG	3/3/2023, 24/3/2023	5/12/2022, 19/12/2022	7/12/2022, 19/12/2022
PACT	3/3/2023, 24/3/2023	5/12/2022, 19/12/2022	7/12/2022, 19/12/2022
$AmSO_4$	3/3/2023, 24/3/2023		
ProPlex		5/12/2022, 19/12/2022	7/12/2022, 19/12/2022
'Custom 1'		6/12/2022	
'Custom 2'		7/12/2022	

seen in previous experiments, likely due to the presence of the GST-tag. During this purification it was observed that Rex1 could be precipitated with 2.8M ammonium sulphate in agreement with previously published observations (Frank *et al*., 1999), at the cost of decreased purity as shown in Figure 4.13C lane 'AS'.

The 6x-His tagged construct offered greater purity through a higher degree of overexpression, however a nickel column step was avoided due to high levels of precipitation in a previously attempted purification as shown in Figure 4.12F. As a result, the same purification pathway as the untagged construct was followed, although this time this led to a monodisperse peak at gel filtration, indicating better purity. The SDS-PAGE analysis of the end product revealed only bands below the expected size, which taken with the gel filtration trace indicates the protein may have undergone degradation.

4.2.6 Crystallisation Trials of Rex1

The purified Rex1 constructs described in Figure 4.13 and Figure 4.14 were each used for crystallisation screening as outlined in Table 4.1. Screens were selected based on two known properties of Rex1: its ability to be reversibly precipitated by ammonium sulphate, and its inactivation by glycerol as observed in previous *in vitro* assays in the Mitchell Lab (Hama Soor, 2017). The initial untagged-Rex1 preparation shown in Figure 4.13 was screened without the addition of cofactors to each condition, whereas the 6xHis-Rex1 preparation shown in Figure 4.14 was tested in the presence of 5mM $MgCl_2$ and in some cases with 5mM AMP. Crystallisation screens for the untagged-Rex1 preparation failed to yield crystallised protein, but two conditions gave crystals with protein-like diffraction patterns: 0.15 M DL-Malic acid with 20 % w/v PEG 3350 (JCSG G8) and 0.2 M Potassium sodium tartrate tetrahydrate with 20 % w/v PEG 3350 (PACT E9). Based on these conditions, custom screens were prepared using a Formulatrix Formulator robot, with conditions shown in Figure 4.15. No conditions produced protein crystals that gave diffraction patterns at sufficient resolution for identification. The only crystals with diffraction patterns of sufficient resolution were used up by diffraction testing at Diamond Light Source, preventing identification by mass spectrometry.

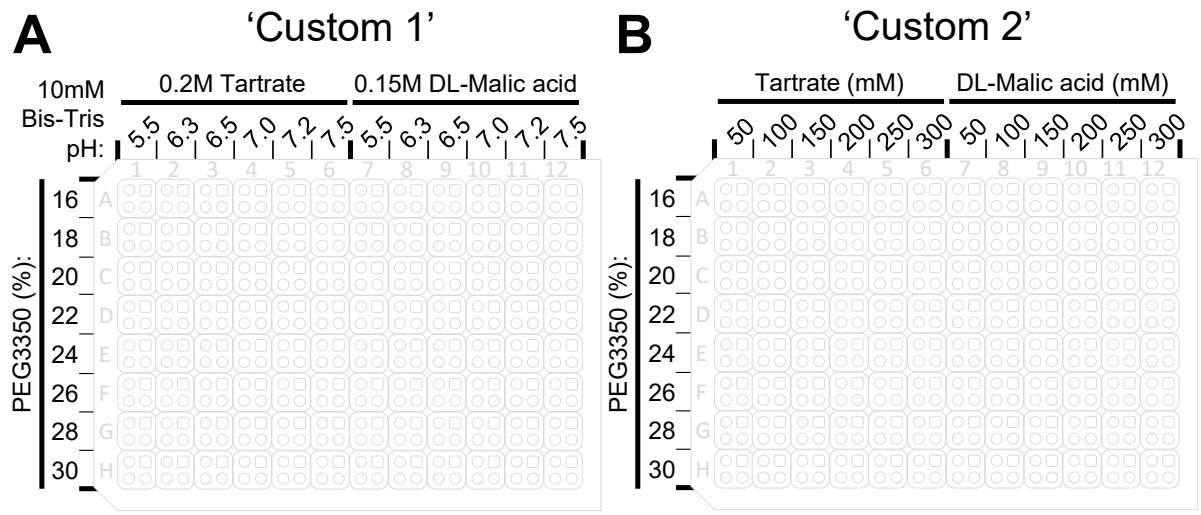


Figure 4.15: Optimisation of crystallisation screens for 6xHis-1-51ΔRex1H360A
Schematic of custom optimisation screens for crystallisation of 6xHis-1-51ΔRex1H360A, based screen conditions JCSG G8 (Molecular Dimensions) and PACT E9 (Molecular Dimensions). Each condition is optimised using half of the 96-well sitting 3-drop plate. **A:** The 'Custom 1' screen features the concentrations of Potassium sodium tartrate tetrahydrate (abbreviated to 'Tartrate') and DL-Malic acid, used in PACT E9 and JCSG G8 respectively, varied against 8 PEG3350 concentrations from 16-30% and 6 pHs of 10mM Bis-Tris from 5.5 to 7.5. **B:** The The 'Custom 2' screen features varied concentrations of Potassium sodium tartrate tetrahydrate (abbreviated to 'Tartrate') and DL-Malic acid from 50mM to 300mM, varied against 8 PEG3350 concentrations from 16-30%.

4.3 Discussion

The results presented in this chapter represent a starting point for the detailed biochemical analysis of Rex1. With a preliminary estimation of optimal buffer conditions, follow-up experiments and additional replication will allow a thorough biochemical assessment of Rex1's exonuclease activity, including the measurement of dissociation constants.

The data presented here suggest further reaction conditions that may be informative. The use of different divalent cations may reveal further properties of Rex activity: Rex1 may be able to use Mn^{2+} as observed for RNase T, Rrp6, and REXO5, which seems to relax substrate binding requirements owing to relaxed H_2O coordination geometry. Conversely, it is likely that Ca^{2+} would lead to non-productive binding of substrates. Given the high degree of precipitation seen during purification of the crystallisation construct after Ni^{2+}-NTA purification, it could be interesting to see whether the presence of Ni^{2+} has an inhibitory effect. With the prediction of the putative alkaline phosphatase-like second domain in Chapter 3, it may be informative to include Zn^{2+} as a cofactor to account for the possibility that cation binding is conserved.

It may be possible to examine the impact of the second domain more directly, based on the prediction that it can specifically recognise the base paired 5´ terminal nucleotide-(tri-)phosphate of an RNA duplex. The yeast 5´ exonucleases Xrn1 and Rat1 can recognise 5´ phosphates and are known to be inhibited by nucleotide bisphosphates (pNp; Dichtl *et al*., 1997), it would be interesting to see whether Rex1's ability to process double-stranded nucleic acids with 3´ and not 5´ overhangs could be inhibited by pNp, while leaving its ability to process single-stranded nucleic acids intact. It is very likely however

that the presence of mononucleotide analogues alone would be inhibitory to all *in vivo* Rex1 activity as has been observed for nucleases including Rrp6 (Axhemi *et al*., 2020), although this inhibition would be worthwhile examining in its own right.

The use of DNA oligos has enabled an initial estimation of the parameters for *in vitro* Rex1 activity, which can next be used for more physiological substrates such as RNA oligos, or substrates that form stemloops. Given Rex1's predicted preference for base paired 5´ nucleotide stemloops, it may be wise when designing such substrates to use ^{32}P as a 5´ label if designing a hairpin molecule, or to have two strands that anneal to give a blunt end more than 9nt away from the overhang end if a 5´ fluorescent dye is used. Given the Cy3 dye's behaviour in the data shown here and its tendency to base-stack with nucleotides (Norman *et al*., 2000), it may be best to avoid cyanin-based dyes altogether in favour of less disruptive dyes such as fluorescein.

The DNA oligo substrates used in this chapter are a more stable alternative to RNA oligos, although the missing 2´-OH groups and the corresponding protein interactions may result in an underestimation of substrate binding ability. The absence of this 2´-OH reduces steric limitation of the phosphodiester backbone; ssDNA is thought to be more flexible than ssRNA (Pal and Levy, 2019). This lack of rigidity may also lead to protein-phosphodiester backbone interactions for DNA oligos that would be sterically forbidden for an equivalent RNA oligo.

Early biochemical characterisation of Rex1 took place under the apparent observation of RNase H activity (the ability to cleave RNA in an RNA:DNA duplex) co-purifying with a protein of roughly 70 kDa (Karwan *et al*., 1983), leading to the name RNH70 (Frank *et al*., 1999). RNase H activity in this case was measured by the ability to liberate radio-labelled nucleotides from a poly rA:dT hybrid (Frank *et al*., 1993), made according to Sarngadharan *et al*. (1975). Interestingly, the lack of an S1 nuclease step means that the poly rA:dT duplex could possess overhangs, potentially providing Rex1 an opportunity to release nucleotides from hybrids that had annealed leaving staggered ends. This initial characterisation revealed a preference for neutral pH and inhibition at pH 8, as seen in the preliminary analysis presented here in Figure 4.9.

Although DEDD domains possesses an RNase H-like fold (Yang, 2011), no RNase H activity has been observed for Rex1 *in vivo* so far. The only experiment undertaken to this end showing that deletion of *REX1* led to no detectable reduction in the RNase H activity of whole-cell extract (Frank *et al*., 1999). It would be interesting to see if Rex1 displays any preference for 3´ RNA overhangs over 3´ DNA overhangs *in vitro*, which could be examined using annealed oligos. If Rex1 shows a preference for RNA over DNA, this could provide evidence towards potential RNAse H-like activity.

While investigating the nuclease activity of Rex1, Karwan *et al*. (1983) made the surprising observation that a labelled poly rU substrate wasn't degraded by Rex1. If this is the case, then one possible explanation could be that Rex1 requires a duplexed 5´-terminal stemloop to effectively engage with an RNA. This could also be a result of contaminating proteins, as downstream analysis revealed a degree of contamination with other proteins known to bind RNA in the initial preparations (Frank *et al*., 1999). Interestingly, the renaturation gel assay used by Frank *et al*. (1999) in this latter study suggests that Rex1 may be able to renature in the presence of 10mM $MgCl_2$, which wasn't attempted for the precipitated Rex1 crystallisation construct, but may have enabled recovery of protein.

Oligos serve as a useful analogue to physiological substrates, but it may be more informative to measure the binding of Rex1 to its substrates directly. It is straightforward to purchase tRNAs commer-

cially, and the Mitchell lab has already established a scheme to purify extended 5S RNPs from mature ribosomes (Hama Soor, 2017; Daniels *et al*., 2022) based on Blobel (1971) and Steitz *et al*. (1988). As a protocol exists for the fractionation of 5S based on the number of 5´ phosphates using hydroxyapatite chromatography (Soave *et al*., 1973), it would be interesting to see if there is any impact of 5´ tri- or di-phosphorylation on binding affinity to Rex1.

The crystallography screen presented in this chapter failed to yield crystallised Rex1. Crystallisation can fail for any number of reasons, but a potential barrier is the innate flexibility of the protein. Given the Alphafold DB model for Rex1 (Jumper *et al*., 2021; Varadi *et al*., 2022) shown in Chapter 3, there seems to be several predicted features that may imply flexibility: the three 'Loop' structures demonstrate a degree of disorder; a hinge is found at the base of the central cleft; and the two domain fold demonstrates limited mutual interaction predictions. In order to account for flexibility, one approach could be to co-crystallise Rex1 with substrate. The use of tRNA may be preferable to 5S rRNA, as tRNA possesses a comparatively compact fold, and doesn't require the association of ribosomal proteins such as RPL5 (Steitz *et al*., 1988). As a flexible fold interrupts crystal packing, an alternative approach could be the use of electron cryo-microscopy, although if a protein's flexibility is too great this can lead to challenges at the data-processing stage.

With the release of Alphafold2 (Jumper *et al*., 2021) and the Alphafold DB (Varadi *et al*., 2022), a frequently expressed sentiment is whether a protein that has a high-confidence predicted structure, as Rex1 does, requires any further structural study. While the AlphafoldDB entry for Rex1 gives a number of hints for mutations that could assist with crystallisation efforts, such as the truncation of the disordered NLS performed in this chapter, many details that would normally be captured by an experimentally-derived structure are missing. In the case of Rex1, a key question that needs addressing is whether there are additional cofactor requirements beyond the two $MgCl_2$ coordinated by the DEDD active site, for example the binding of Zn^{2+} ions at the predicted alkaline phosphatase domain (by coincidence, the mechanisms for both of these cofactor binding sites are illustrated side by side in Steitz and Steitz, 1993). Further details that aren't captured in the Alphafold DB model include the conformation flexibility between the two predicted domains, and the specific structural features of Rex1 that are required for substrate engagement. Educated guesses can be made in the absence of a protein-substrate structural model, such as those made in Chapter 3, the impact of which on Rex1 function are examined in the following Chapter.

Chapter 5

Mutational analysis of Rex1 Sequence Features

5.1 Introduction

This project aims to not only examine the biochemistry of Rex1 *in vitro* activity against model substrates, but also to contextualise this activity to the level of whole-cell phenotypes through the analysis of mutants. The phenotype of *REX1* loss was first examined in the *rna82-1* mutant (Piper *et al*., 1983), identified during a temperature sensitivity screen for small RNA processing defects by the accumulation of extended 5S rRNA. This mutant was later identified by sequencing as a *rex1* W433* truncation mutant (van Hoof *et al*., 2000a). *REX1* was then characterised to be required for full maturation of 25S rRNA (Kempers-Veenstra *et al*., 1986), and the dicistronic Arg-Asp tRNA (Piper and Stråby, 1989).

A precise deletion of the yeast *REX1* locus was generated by van Hoof and colleagues (van Hoof *et al*., 2000a), which combined with *rex2*Δ, *rex3*Δ, and *rrp6*Δ mutants revealed multiple synergetic substrate processing defects including in 5.8S rRNA and the RNase P catalytic RNA, and a synthetic lethality relationship in the case of the *rex1*Δ/*rrp6*Δ mutant. This mutant has since served the basis of subsequent studies examining tRNA metabolism, revealing a role for Rex1 in the maturation of the intron-containing tRNAs Lys (UUU) and Ser (CGA; Copela *et al*., 2008) and a role for Rex1 in the turnover of hypomethylated Met tRNA (ATG; Ozanick *et al*., 2009). An earlier study of *rex1*Δ/*rrp*47Δ mutants complemented with hypomorphic *rrp47* mutants in the Mitchell laboratory has enabled observation of the RNA processing defects that may underpin the *rex1/rrp6* synthetic lethality relationship, which is suggested to be a result of the backup processing of snoRNAs and CUTs (cryptic unstable transcripts) in the absence of Rrp6 or Rex1 saturating the poly-adenylation machinery, leading to global dysregulation of gene expression (Garland *et al*., 2013).

The work presented in this chapter aims to measure the contribution of conserved sequence features to Rex1 function. A range of *rex1* mutants lacking features predicted in Chapter 3 were cloned and the effect of these mutations on Rex1 function were examined. The functionality of mutants were first examined *in vitro* using the purification schemes and biochemical assays optimised in Chapter 4, then examined *in vivo* as plasmid-borne expression constructs in *rex1*Δ and *rex1*Δ/*rrp*47Δ yeast backgrounds. This plasmid shuffle assay was applied with a yeast expression vector designed for the expression of human DEDDh enzymes, with this initial analysis revealing lack of complementation with

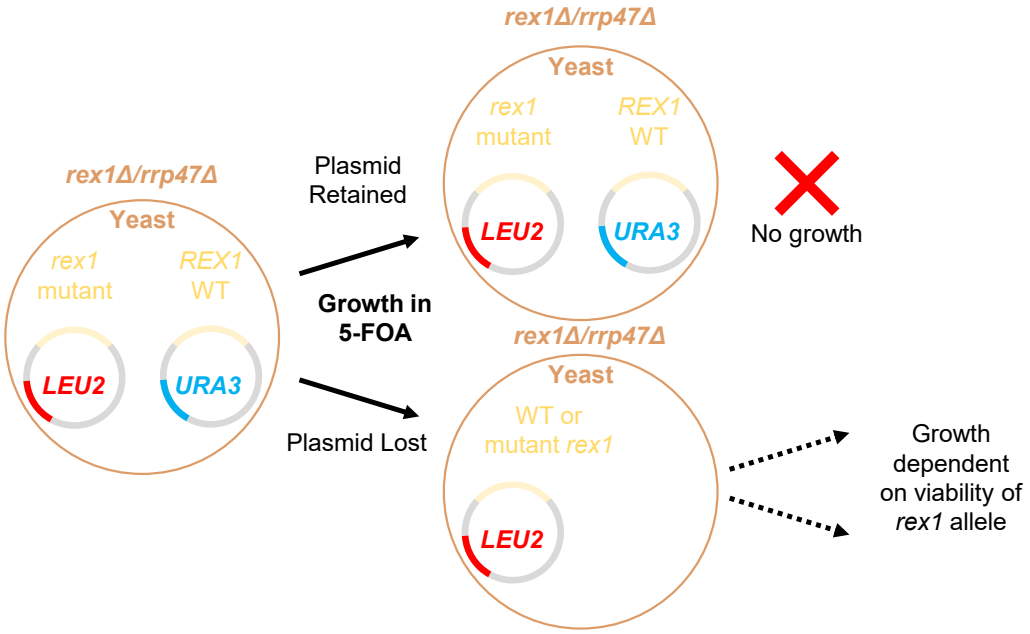


Figure 5.1: Schematic of yeast plasmid shuffle assay.
Diagram demonstrating plasmid shuffle assay measuring *Rex1* activity. Yeast bearing an active *URA3* gene on a plasmid are unable to grow in the presence of 5-FOA, meaning only yeast that are able to lose the *URA3*-bearing plasmid during cell division are able to grow. The yeast is now dependent on the rex1 allele found on a non-*URA3*-bearing plasmid. The viability of this *Rex1* mutant dictates the growth rate of the yeast, reading out as the amount of colony growth for spots on a plate.

human *REXO5*.

5.2 Results

5.2.1 N- and C- terminal Rex1 sequences are required for Rex1 function

There are multiple DEDD exonucleases with close homology within the exonuclease domain to Rex1 in the yeast genome, however Rex1 is unique amongst them in its ability to mature 5S rRNA (Piper *et al.*, 1983), and to compensate for the absence of Rrp6 (van Hoof *et al.*, 2000a). It was proposed at the outset of the predecessor PhD project undertaken by Taib Hama Soor that the extended N- and C-terminal sequences unique to Rex1 orthologues may confer this ability (Hama Soor, 2017). A range of N- and C-terminal truncation mutants were cloned during this predecessor project, as a set of yeast expression plasmids with both low copy number (CENP; Sikorski and Hieter, 1989) and high copy number (2μ; Christianson *et al.*, 1992) replication origins (Hama Soor, 2017).

The level of Rex1 function for these mutants was examined using a plasmid shuffle assay. This assay, summarised in Figure 5.1, overcomes the inability to transform plasmids into a synthetic lethal yeast mutant by supporting the strain with a plasmid coding for *REX1*, and a *URA3* marker. The *rex1*Δ/*rrp*47Δ + p(*REX1*/ *URA3*) can be transformed with a plasmid bearing a desired *rex1* allele, which can be rendered the sole *Rex1* expression plasmid through counterselection on media containing 5-fluoroorotic acid (5-FOA), a precursor compound that is converted to toxic 5-fluorouricil in cells expressing the *URA3* gene.

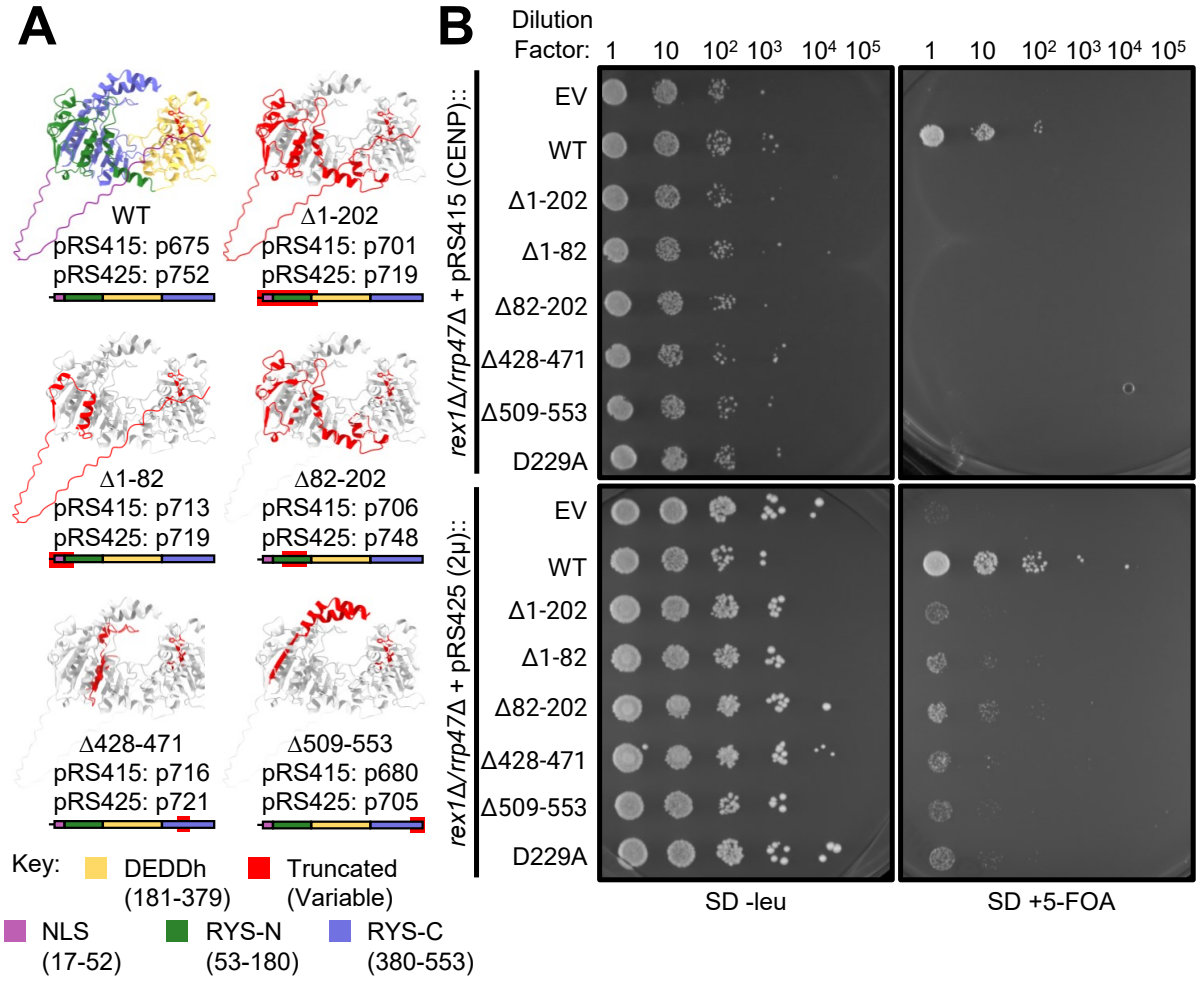


Figure 5.2: N- and C- termini are required for Rex1 function *in vivo*

A: Description of each truncation mutant. The truncated region is indicated on a ORF diagram of Rex1 by a red box. The composite N- and C-terminal RYS domain is represented in green and purple respectively, the DEDD domain is represented in yellow, and the NLS is represented in purple. The corresponding region on the Rex1 AlphafoldDB model is highlighted in red. Each mutant is expressed on a low copy number (pRS415; 'CENP') or high copy number (pRS425; '2μ') plasmid, with the corresponding plasmid name given in each case. **B:** Spot growth plates of *rex1* truncation mutants after 5 days at 30ºC. Each plate represents three technical replicates. The 'SD –leu' plates represent a selection-free control that demonstrates equal seeding of cells into each lane. 'EV' = empty vector control.

The data collected by Taib Hama Soor revealed that each truncation mutant showed loss of Rex1 activity, which is replicated by this current study in Figure 5.2B (Daniels *et al*., 2022). There seems to be a mild compensatory effect conferred by the increased expression accompanying the high copy number plasmid, however the growth visible in the empty vector control may indicate a degree of background growth. This background may be the result of comparably lower 5-FOA activity, or an example of *rex1*Δ*/rrp*47Δ synthetic slow growth as observed in one case in the literature (Copela *et al*., 2008).

The truncation mutants examined were designed based on secondary structure predictions available at the time, however high-confidence structure prediction by Alphafold2 have since become available as discussed in Chapter 3. When considered together, the regions deleted by these mutants in each case are predicted to disrupt a composite β-sheet that forms the core of the putative RYS domain as shown in Figure 5.2A. The most likely explanation for the loss of Rex1 function in light of this is that the deletion of these regions interrupts correct folding of the RYS domain, with the removal of any of the examined N- and C-terminal sequences resulting in a total loss of function. This may serve as evidence in support of the Alphafold DB-predicted model for Rex1 domain structure, despite the fact these mutants were designed long before it became available (Jumper *et al*., 2021; Varadi *et al*., 2022).

5.2.2 Measurement of RNA binding to the Rex1 N- and C- termini using RNA crosslinking and 5´ radio labelling

Previous work in the Mitchell laboratory has shown that RNA crosslinking could be detected for Rex1 *in vivo* (Daniels *et al*., 2022). This crosslinking scheme is outlined as follows: mid log-phase yeast cultures expressing zz-tagged Rex1 were pelleted and resuspended in fresh media. This media was dispensed into a lidless 6cm Petri dish to give a liquid depth of 1-2mm, then crosslinked with 12000J of UV-C on an ice-cooled glass plate for a total of 5 minutes. Cells were harvested and lysed before IgG bead purification, which included a stringent 2M $MgCl_2$ wash to remove any loosely associated protein or nucleic acids (Allmang *et al*., 1999a; Mitchell *et al*., 2003). The purified zz-Rex1-RNA complex was treated with RNase A, and the resulting 5´-OH nucleic acid ends are conjugated to radiolabelled-phosphate using polynucleotide kinase and $\gamma[^{32}P]$-ATP. The radiolabelled Rex1-RNA crosslinked species are then analysed by SDS-PAGE and Western blotting, giving a PhosphoImager signal coincident with a PαP antibody signal for sites of crosslinking.

This crosslinking was not detectable for wild type zz-Rex1, and could only be detected in the D229A zz-Rex1; the only active site point mutant of the mutants examined (Daniels *et al*., 2022). By combining the D229A mutant with N- and C- terminal Rex1 truncation mutants, no crosslinking could be detected. This indicates an essential role for the N- and C-terminal regions in Rex1's ability to bind its substrates in a manner detectable by this assay (Daniels *et al*., 2022). It was unclear whether this is a result of incorrect protein folding, or a loss of substrate binding ability. In order to address the question of RNA binding more directly, an aim of this project was to localise the sites of crosslinking observed for full-length Rex1 D229A mutant to specific domains through site-specific protease cleavage of the D229A Rex1 protein. To this end I generated three *rex1* D229A constructs that each contained a PreScission protease cleavage site, LEVLFQ/GP, flanked on each side by a BglII-encoded RS dipeptide as described in Figure 5.3. Three target sites were chosen: one N terminal site at H220 before the exonuclease domain, and two C-terminal sites; one immediately after the exonuclease domain at A366, and one at the position corresponding to the 509-553Δ mutant at A511. The specific position of these sites were

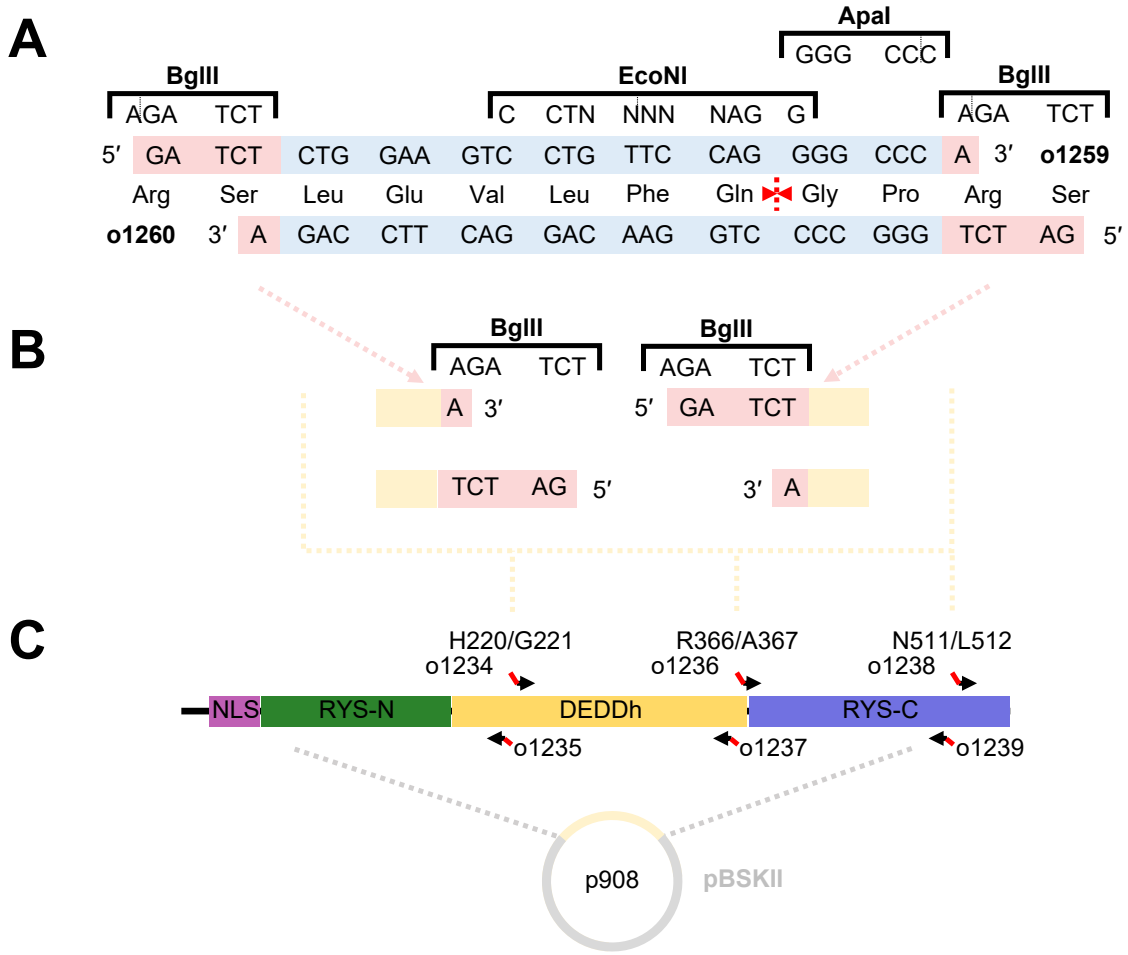


Figure 5.3: Design of PreScission protease site insertion

A: The full sequence of a protease site oligo linker flanked with BglII sites, including ApaI and EcoNI sites for screening. 'N' refers to a position in a restriction site that can be any nucleotide. Red cross indicates position in the protease site amino acid sequence where cleavage takes place. **B:** Complementary BglII sticky ends for the destination vector, intended to allow in-frame insertion of the linker into the *REX1* ORF. **C:** Diverging primer SDM for introduction of BglII sites at positions throughout the *Rex1* ORF. Domain diagram drawn to scale, boundaries are defined according to the Alphafold model shown in Figure 3.6. Precise insertion sites are indicated above each primer pair, with individual primer names listed beneath.

selected based on Phyre2-based secondary structure prediction available at the time (Kelley *et al*., 2015).

As the protease site sequence was 8 amino acids long, linker insertion was judged to be an appropriate cloning method. For each intended integration site within the Rex1 ORF, a unique BglII site was introduced using diverging extended primer PCR. A short plasmid-borne copy of wild type *Rex1* was used as the PCR template, the product of which was used for the linker insertion step before sub-cloning into the pRS313 yeast expression vector.

At this stage the cloning process suffered from concatenation of the linker sequence for the H220 and A366 insertion site constructs. This concatenation was resolved using the linker's internal EcoNI site to condense the concatemer down to a single copy of the repeating unit. A further intervention was required for the H220-G221, which featured an FQGP to LQDFQ in-frame substitution, which was corrected using a further site-directed mutagenesis step. Once each PreScission site *Rex1* mutant had been generated, a final SDM step was performed to introduce the D229A mutation, which accompanied a silent SphI site for re-circularisation and screening. This cloning process is outlined in Figure 5.4.

While the RNA crosslinking experiment requires the use of *rex1* D229A mutants, the intermediate generation of *Rex1* mutants with the full DEDDh motif allowed the impact of the protease-cleavable sites on Rex1 function to be assessed *in vivo* using a plasmid shuffle assay, as shown in Figure 5.5. The later Alphafold DB prediction (Jumper *et al*., 2021; Varadi *et al*., 2022) suggested that the exonuclease domain boundaries used may interfere with structured regions (see Figure 5.5A), however this did not correlate with the results observed. The Alphafold2 model suggests that only the H220-G221 insertion did not interrupt an alpha helix, instead residing in a predicted solvent-exposed loop at the periphery of the fold. In a subversion of this prediction, only the H220-PsP-G221 mutant demonstrated loss of Rex1 activity in the plasmid shuffle assay, with the other mutants retaining wild type levels of Rex1 activity. Each mutant demonstrated variable expression levels, with only the L511-R512 mutant showing lower expression than a genomic HTP(6xHis-TEV-Protein A)-tagged *REX1*, shown in Figure 5.5B.

An in-frame derivative of the H220-G221 PsP cleavage site was generated by mistake during cloning (p927 in Figure 5.4), resulting in the replacement of the LEVLFQGP motif with LEVLLQDFQR. This mutant demonstrated wild type levels of growth in the plasmid shuffle assay, despite being 3 amino acids longer. The primary difference between these mutants is the lack of a C-terminal proline, which may cause misfolding when introduced into the exonuclease domain loop that contains H220 and G221.

The D229A *rex1* mutant-bearing plasmids were next transformed into *rex1*Δ yeast and a preliminary purification and cleavage experiment was attempted in Figure 5.6. The parameters that were varied were denaturing conditions using 1M urea and 1M arginine, increased NaCl concentration from 150mM to 500mM, and the addition of 0.4M $MgCl_2$, based on observations in the literature that PreScission protease is able to tolerate highly denaturing conditions (Ullah *et al*., 2016). The results indicated that high salt and denaturing conditions were tolerated by PreScission protease, but that the addition of $MgCl_2$ had an inhibitory effect on this substrate. The native cleavage condition for the H220-PsP-G221 mutant lacked any chemiluminescence signal, which was mostly likely due to an experimental error at or after the binding of the yeast lysate to the beads.

Based on the observation that non-denaturing conditions were sufficient for PreScission protease cleavage, a full crosslinking, radiolabelling, and cleavage experiment was performed. This PreScission protease cleavage step was integrated into the crosslinking protocol, with PreScission protease treatment performed overnight at 4°C after the PNK labelling step. The resulting supernatant from the protease

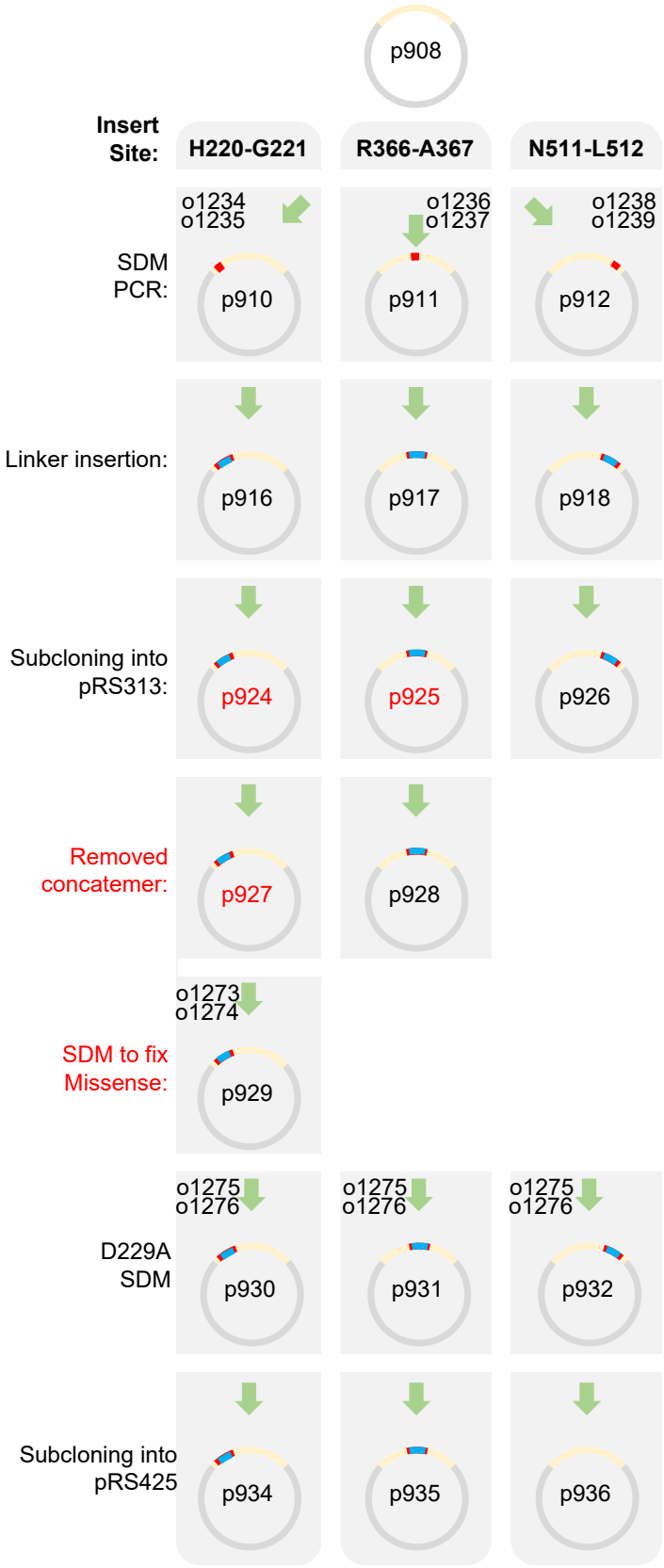


Figure 5.4: Cloning schematic for PreScission protease site insertion *rex1* mutants
Full cloning scheme used to generate *rex1* protease site insertion mutants. Red text indicates steps where unintended mutations required correcting. Plasmid cartoons represent the *rex1* ORF in beige, the plasmid backbone in grey, the BglII restriction site in red, and the PreScission protease site in blue. PCR steps include the names of primers used. Red plasmid names indicate plasmids featuring sequence errors that required correction.

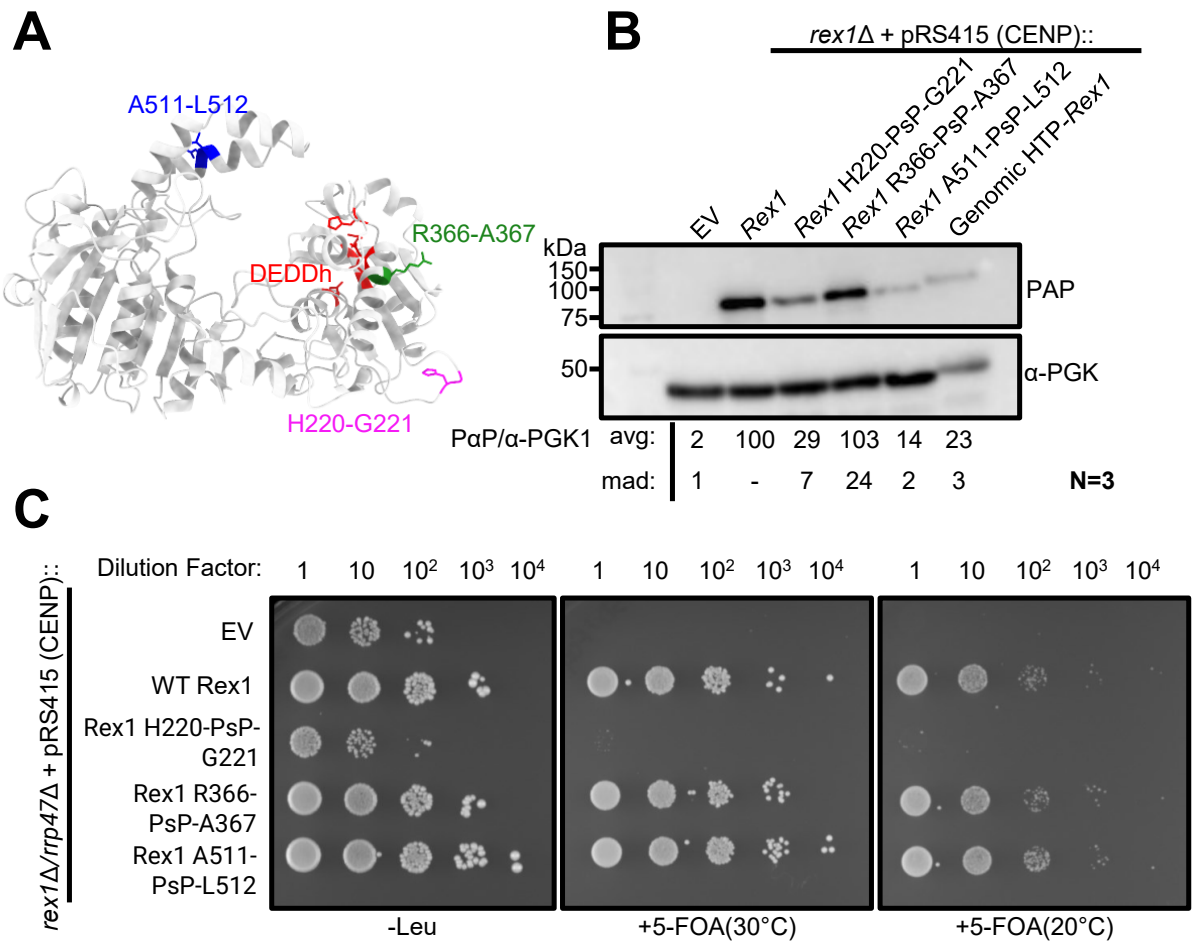


Figure 5.5: Expression and viability of active PreScission-cleavable Rex1 mutants
A: PreScission site insertion positions on Alphafold2-predicted Rex1 model indicated in blue, green, and magenta, with the DEDDh active site indicated in red. **B:** Western blot analysis of whole-cell extracts of *rex1*Δ yeast transformed with pRS415 expressing various *rex1* mutants or WT *REX1*, with an endogenously tagged HTP-*REX1* expressing strain for comparison. Each blot representative of three technical replicates, non-merged blot images analysed with ImageJ. 'avg' = average, 'mad' = mean absolute deviation. **C:** Plasmid shuffled assay of PreScission-cleavable rex1 mutants. Plates were imaged after 5 days, each plate image representative of 3 technical replicates. 'EV' = empty vector control.

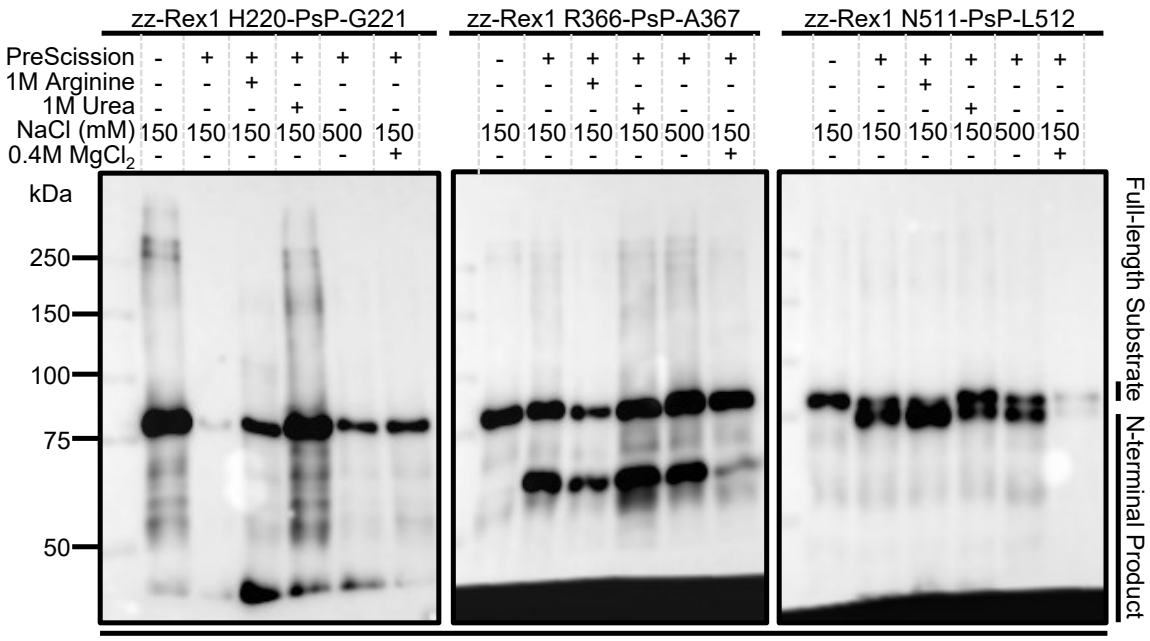


Figure 5.6: Establishing compatible buffer conditions for Rex1 PreScission protease cleavage
PreScission cleavage of zz-Rex1 on IgG beads. Each lane represents a 1 hour 25ul bead pulldown on 660ul of lysate at 4°C, representing 166ml of culture (1L split 6 ways). Each 1L culture grown to 1.0 OD_{600nm} of *rex1*Δ yeast transformed with pRS415::*rex1*D229A+precision site insertion. Beads were incubated with PreScission protease overnight at 4°C in 50mM HEPES pH 7.4 along with the indicated buffer components. Samples were analysed by SDS-PAGE and Western blotting.

treatment was retained along with an equivalent volume of acetic acid elution. Both eluates were analysed by SDS-PAGE followed by Western blotting and radiography, although the acetic acid elution recovered a higher proportion of protein (protease supernatant not shown).

The results of this experiment, shown in Figure 5.7, indicate binding at both the N- and C- termini to some extent. The A511-PsP-L512 mutant did not seem to produce crosslinking at detectable levels across the three replicates, but both the H220-PsP-G221 and A366-PsP-R367 mutants produced bands that resembled the pattern predicted for RNA binding at both termini.

While most ^{32}P bands seemed equal in intensity, a bright outlier reliably appeared in the H220-PsP-G221 lane. The experiment lacked a control that would allow comparison of the relative labelling efficiencies for each sample, as the intention of the experiment was the binary detection of crosslinking at the N- and C- termini. As a result, we were unable to determine whether this increased intensity represents increased RNA binding, or artefactual increased radio-labelling of the fragment before protease cleavage. With regards to the binary detection of RNA crosslinking however, this experiment demonstrates a positive observation of RNA binding outside of the exonuclease domain of zz-tagged Rex1, spanning both the N- and C- termini.

5.2.3 Cloning of *rex1* mutants lacking predicted substrate-binding sequence features

With the observation that the N- and C- termini of Rex1 can crosslink substrates *in vivo*, the next aim of this project was to identify specific Rex1 sequence features that mediate substrate recognition. A range of structural features within and beyond the exonuclease domain were identified by the bioinformatic

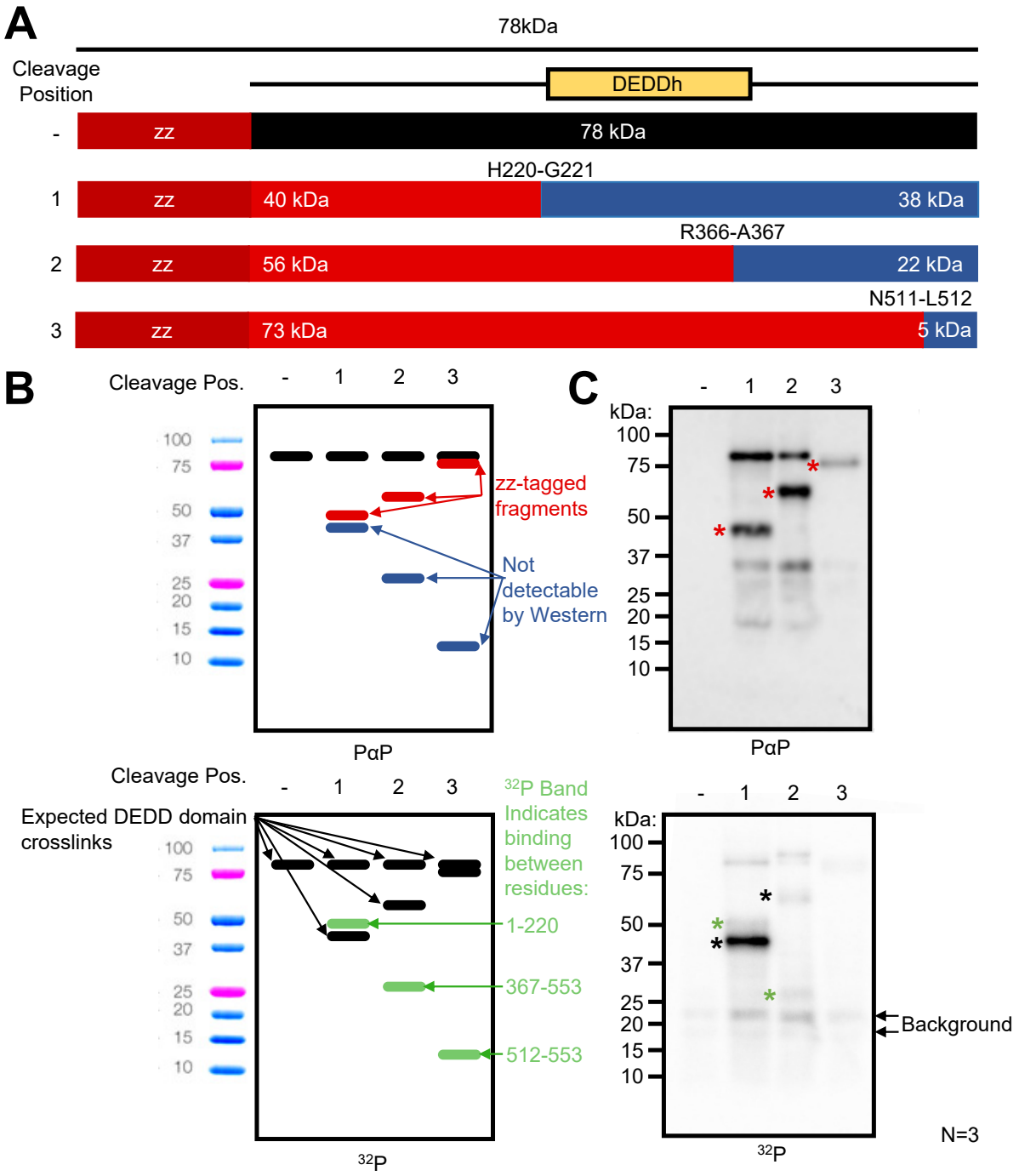

A
78kDa
Cleavage Position
DEDDh
-
zz
78 kDa
H220-G221
1
zz
40 kDa
38 kDa
R366-A367
2
zz
56 kDa
22 kDa
N511-L512
3
zz
73 kDa
5 kDa
B
Cleavage Pos.
- 1 2 3
100
75
50
37
25
20
15
10
zz-tagged fragments
Not detectable by Western
PαP
Cleavage Pos.
- 1 2 3
Expected DEDD domain crosslinks
32P Band Indicates binding between residues:
1-220
367-553
512-553
32P
C
kDa:
- 1 2 3
100
75
50
37
25
20
15
10
PαP
kDa:
- 1 2 3
Background
N=3
32P

Figure 5.7: Crosslinking, cleavage and ^{32}P labelling reveals both Rex1 N- and C- termini contact RNA *in vivo*.
A: Scale domain diagram of zz-tagged Rex1, with post PreScission cleavage band sizes indicated. **B:** Schematic of predicted cleavage pattern for PreScission protease incubation for each mutant. Green bands and arrows indicate positions where RNA crosslinking would indicate substrate binding beyond the exonuclease domain. Red bands indicate bands that can be detected by PAP antibody chemiluminescence, with grey bands not visible. Black bands indicate fragments containing the exonuclease domain, including uncut full-length protein. **C:** PreScission cleavage of zz-Rex1 on IgG beads. Each lane represents a 1 hour 100ul bead pulldown on 500ul of lysate at 4°C, representing 250ml of culture. Each 250ml culture grown to 1.0 OD_{600nm} of *rex1*Δ yeast transformed with pRS415::*rex1*D229A+precision site insertion. Beads were incubated with PreScission protease overnight at 4°C, supernatants were retained and analysed but not shown, remaining bead-bound fractions were eluted with 210ul 0.5M acetic acid and analysed. Visible bands corresponding to predicted bands are indicated with coloured asterisks, background signals are indicated with arrows. Blots representative of three biological replicates.

analysis described in Chapter 3. Within the exonuclease domain, 5 residues were predicted to make side chain interactions with RNA: Y272, H308, N312, K340, and S342. For each of these residues, an alanine substitution mutant was cloned by diverging extended PCR SDM in the same manner as H360A in Figure 4.1. These mutants were sub-cloned into the *E. coli* expression vector pGEX-6P-1 (Smith and Johnson, 1988) and the yeast expression vector pRS313 (Sikorski and Hieter, 1989). A further N312A/K340A/S342A triple mutant was cloned, based on the later observation that these mutants individually had less impact on Rex1 activity and that these residues may each make contributions to substrate binding with an additive effect. Outside the exonuclease domain, the three distinctive loops shown in Figure 3.6 were replaced with short glycine-serine rich linkers. This served as the basis of a Masters project undertaken by Sophie Kelly, who succeeded in cloning all three mutants in the yeast expression vector pRS313 (Christianson *et al.*, 1992). I subcloned these mutants into the pGEX-6P1 *E. coli* expression vector (Smith and Johnson, 1988) and a pRS313 yeast expression vector (Sikorski and Hieter, 1989). I then used overlapping PCR to make every combination of the three loop mutants in the pGEX-6P-1 vector as described in Figure 5.8.

5.2.4 Expression and purification of Rex1 mutants

For every GST-Rex1 mutant available at least one batch of autoinduction culture was prepared, although an obvious overexpression band couldn't be seen by Coomassie staining for most mutants as shown in Figure 5.9. For Rex1 point mutants within the exonuclease domain, 1-3 batches of each Rex1 mutant were prepared. An initial batch for all 5 mutants was prepared using 100ml autoinduction cultures, with two 500ml culture batches prepared for Y272A, H308A, and N312A. Only a single 500ml batch was prepared of S342A, and no further batches of K340A could be prepared due to time constraints. A single 500ml culture batch was prepared for every 'loop' mutant, and the N312A/K340A/S342A combined mutant, henceforth referred to as 'NA/KA/SA'.

The initial batch of GST-Rex1 point mutant purifications was prepared using a one-step glutathione purification scheme, with all later preparations using a two-step heparin-glutathione purification scheme. Both purification schemes are shown in Figure 5.10. The elutions of these purifications, as with the preparation of wild type GST-Rex1, included the detergent TWEEN-20 at 0.1% for enhanced protein stability and reduced adsorption to the inner surface of the microfuge tube (Chou *et al.*, 2005). As

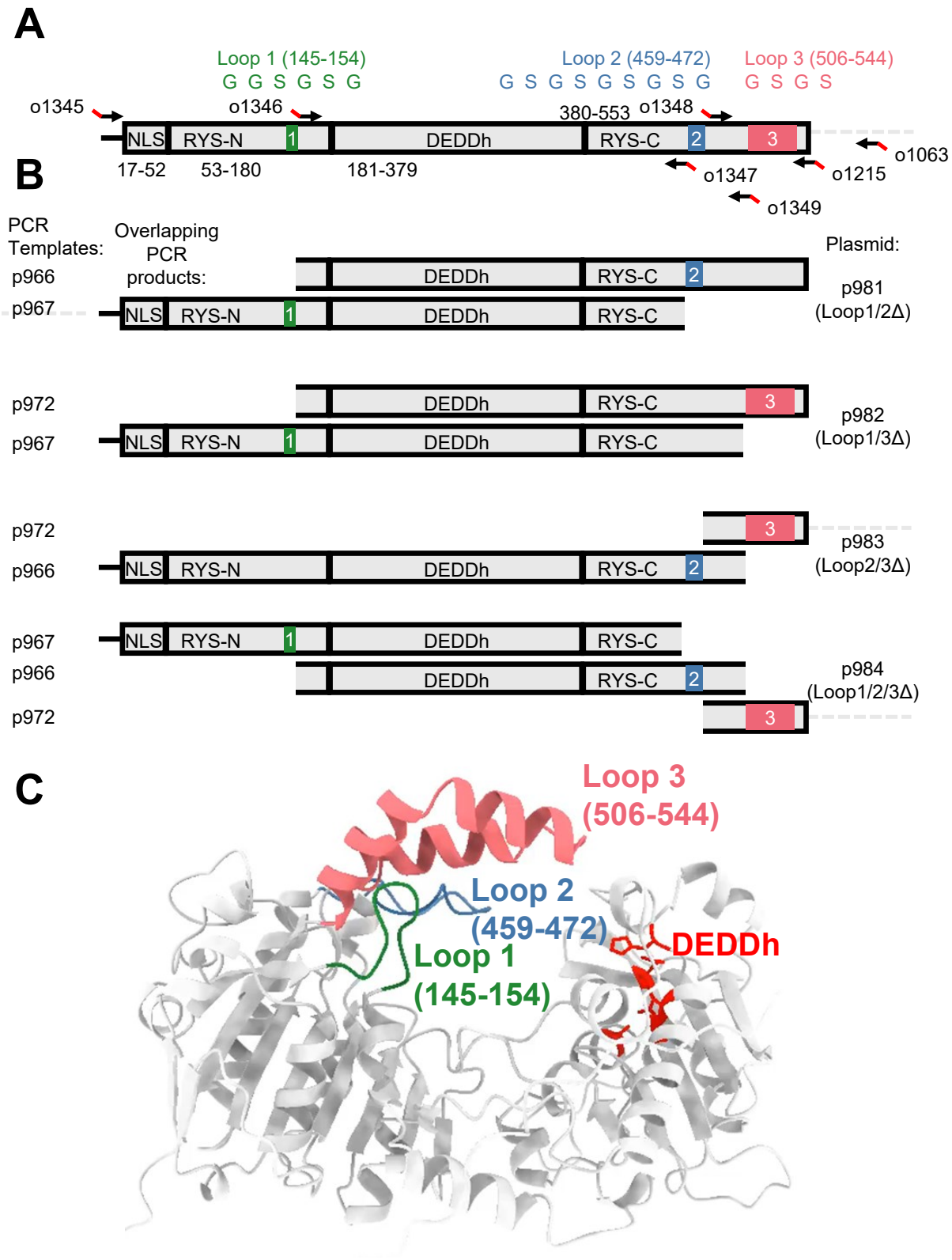


Figure 5.8: Cloning scheme for *rex1* combined loop mutants.
A: Primer positions for the generation of overlapping PCR products. The position of Loop substitutions is indicated with the replacement linker sequences. **B:** Overlapping PCR products for each combination of loop mutants, aligned horizontally with primer positions in A. The plasmids used as templates for the first step PCR reactions are indicated on the left. The overlapping PCR products are combined in second PCR step to give the plasmids indicated on the right. **C:** ChimeraX model of the positions for each loop mutant. The DEDDh residue side chains are shown, and highlighted in red.

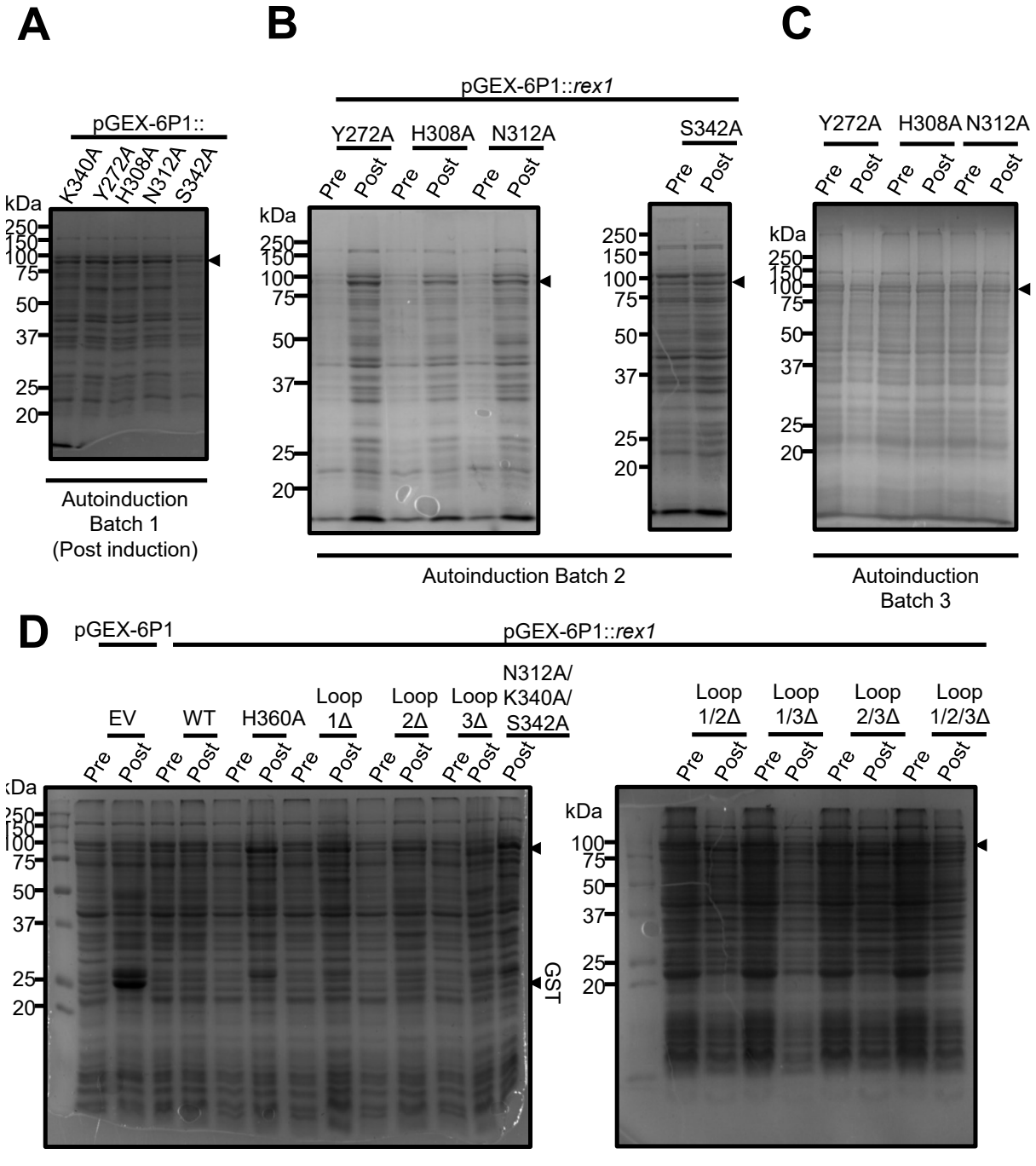


Figure 5.9: Autoinduction of recombinant GST-Rex1 mutants in *E. coli* shows limited overexpression in most cases

All panels show 12% denaturing acrylamide gels, imaged with Coomassie stain. The expected size of full-length GST-Rex1 is 89kDa, indicated with a black triangle. **A:** Whole-cell lysates of BL21 *E. coli* autoinduction 100ml cultures ('batch 1'). Pre-induction samples were not retained **B:** Whole-cell lysates of BL21 *E. coli* autoinduction 500ml cultures ('batch 2'). "Pre" indicates samples taken at between 0.6-0.9 OD_{600nm} before cultures were shifted to 18°C overnight. "Post" indicates samples taken when harvesting the next day. **C,D:** A Whole-cell lysates of BL21 *E. coli* autoinduction 500ml cultures. "Pre" indicates samples taken at between 0.4-0.7 OD_{600nm} before cultures were shifted to 18°C overnight. "Post" indicates samples taken when harvesting the next day. The size of Rex1 loop mutants ranges from 84-89 kDa. For N312A/K340A/S342A, only the post-induction sample were loaded.

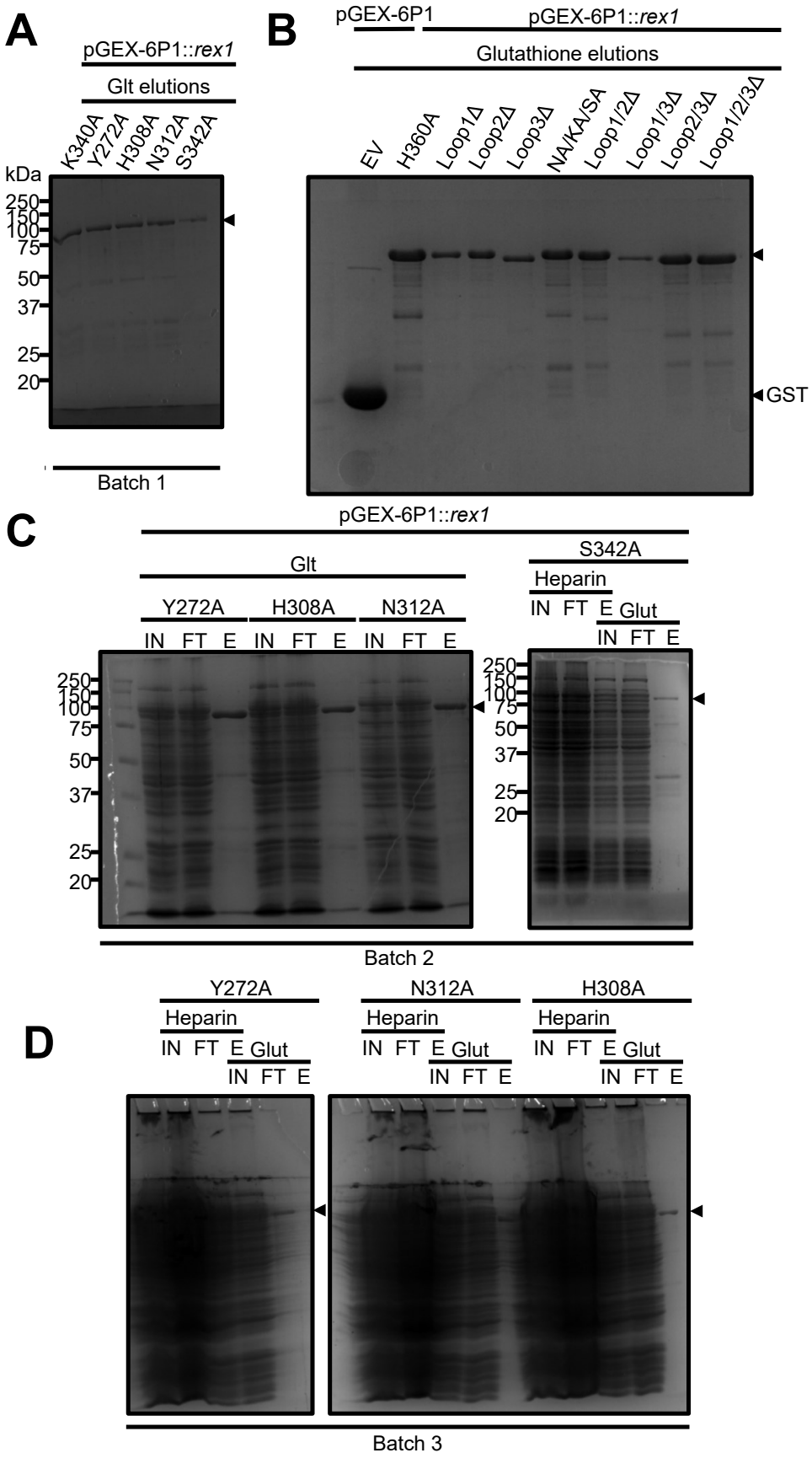

A
pGEX-6P1::rex1
Glt elutions
K340A
Y272A
H308A
N312A
S342A
kDa
250
150
100
75
50
37
25
20
Batch 1
B
pGEX-6P1
pGEX-6P1::rex1
Glutathione elutions
EV
H360A
Loop1Δ
Loop2Δ
Loop3Δ
NA/KA/SA
Loop1/2Δ
Loop1/3Δ
Loop2/3Δ
Loop1/2/3Δ
GST
C
pGEX-6P1::rex1
Glt
Y272A
H308A
N312A
IN FT E IN FT E IN FT E
S342A
Heparin
IN FT E Glut
IN FT E
250
150
100
75
50
37
25
20
Batch 2
D
Y272A
N312A
H308A
Heparin
IN FT E Glut
IN FT E
Batch 3

Figure 5.10: Purification of recombinant GST-tagged Rex1 mutants.
All panels show 12% denaturing acrylamide gels, imaged with Coomassie stain. The expected size of full-length GST-Rex1 is 89kDa, indicated with a black triangle. The size of Rex1 loop mutants ranges from 84-89 kDa. **A:** Elutions from 1-step glutathione purification on lysates of 100ml autoinduction cultures ("batch 1"). **B:** Elutions from 2-step Heparin-Glutathione purifications on 500ml of autoinduction culture. Degraded protein ladder prevented accurate sizing. NA/KA/SA shorthand used for the combined N312A/K340A/S342A mutant. **C,D:** Elutions from 1- or 2- step glutathione purification on lysates of 500ml autoinduction cultures ("batch 2", "batch 3"). Glut = Glutathione, IN = Input, FT = Flowthrough, E = elution.

this prevented the use of the Bradford assay to estimate protein concentration, and because the elution concentration was too low for detection by UV absorbance, anti-GST Western blotting signal was used to normalise all protein preparations to a standard concentration of 0.7μM for the full-length species as shown in Figure 5.11. Although more time-consuming, this method had the benefit of quantifying the concentration specifically of the full-length Rex1 product. Although variable amounts of degradation product can be seen for mutants such as S342A, the lack of Rex1 activity for the various truncation mutants indicate that only full-length Rex1 should confer activity. As the least concentrated preparation, the second batch of S342A was used as a standard to which all other mutants were standardised against.

5.2.5 The contribution of Rex1 sequence features to *in vitro* DNA exonuclease activity

Purified GST-Rex1 mutants were used to assess the impact of the loss of each structural feature on the *in vitro* exonuclease activity of GST-Rex1. The same set of substrates as described in 4.2.3 were used for *in vitro* degradation assays: a 5´-^{32}P-labelled oligo of mixed sequence in Figure 5.12, a 5´-^{32}P-labelled oligo dT in Figure 5.13, and a 5´-Cy3-labelled oligo dT in Figure 5.14 (quantified in Figure 5.15). Not all mutants were tested against all substrates; only the Y272A, H308A, N312A, and S342A were tested against all three substrates, the K340A mutant was only tested against the 5´-^{32}P-labelled oligo of mixed sequence, and the 'loop' mutants were only tested against the 5´-Cy3-labelled oligo dT.

It was expected that RNA-binding residues may contribute minimally to substrate binding individually, as multiple residues seem to form the positively charged binding patches that mediate substrate binding seen in structural models (see Table 3.1). A mutations caused a reduction in Rex1 activity to some extent against at least one substrate: the 'loop' mutants, H308A, and NA/KA/SA mutants each seem to almost fully ablate Rex1 activity, the Y272A, demonstrated minimally detectable activity, with intermediate activity evident for the K340A, and N312A mutants. The S342A mutant demonstrated WT-like activity against the oligo dT-based substrates, but showed a small decrease in activity against the mixed sequence oligo. The intermediate and full activity evident in the initial assays of N312A, K340A, and S342A shown in Figure 5.12 served as the basis for designing the NA/KA/SA mutant, which demonstrates a large reduction in activity against the 5´-Cy3-labelled oligo dT in Figure 5.14.

The assays shown in Figure 5.12 and Figure 5.13 demonstrated undetectable levels of background activity, however the Cy3-labelled substrate assays in Figure 5.14 demonstrated a consistent level of smearing, evident in the buffer-only control. Low levels of Rex1 activity could howeber be observed for the NA/KA/SA mutant against this background. The impact of this background is that low levels of activity in the 64 and 128 minute time points may be concealed for the Cy3 substrate assays.

It was expected that the Y272A mutant would have a large impact on Rex1 function *in vitro*, as

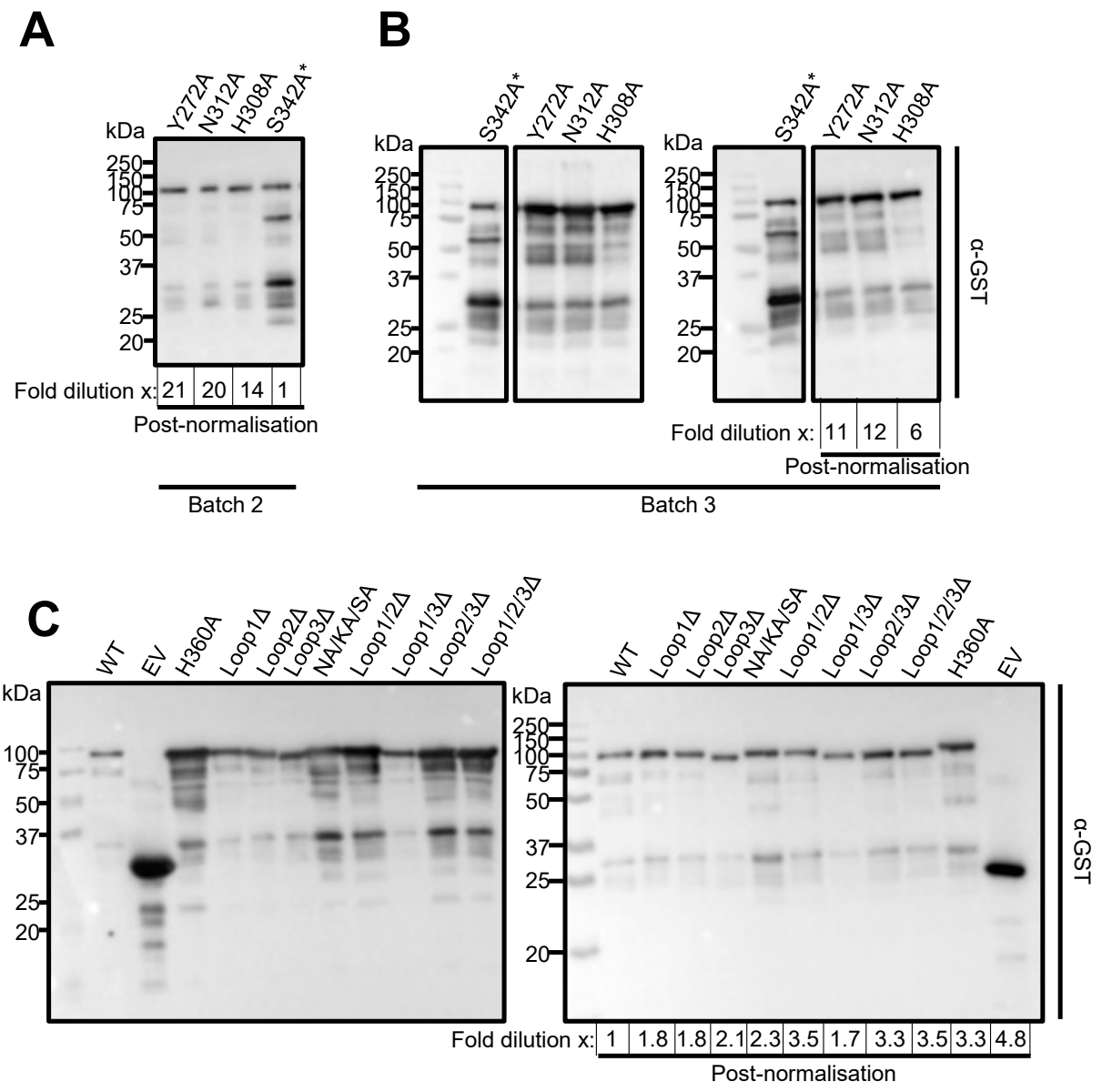


Figure 5.11: Western analysis and concentration normalisation of GST-tagged Rex1 mutants. All panels show 12% denaturing acrylamide gels, transferred onto nitrocellulose membrane probed with anti-GST ('α-GST') primary antibody and goat-anti-mouse peroxidase secondary antibody. The expected size of full-length GST-Rex1 is 89kDa. The size of Rex1 loop mutants ranges from 84-89 kDa. * indicates all shown S342A preps were elutions from the same 2-step Heparin-Glutathione purification. **A:** Elutions from 1-step glutathione purification on lysates of 500ml autoinduction cultures ('batch 2'), normalized based on quantification of the Coomassie-stained gel in Figure. **B:** Elutions from 2-step Heparin-Glutathione purifications on 500ml of autoinduction culture ('batch 3'). **C:** Elutions from 2-step heparin-glutathione purification on 500ml of autoinduction culture. NA/KA/SA shorthand used for the combined N312A/K340A/S342A mutant.

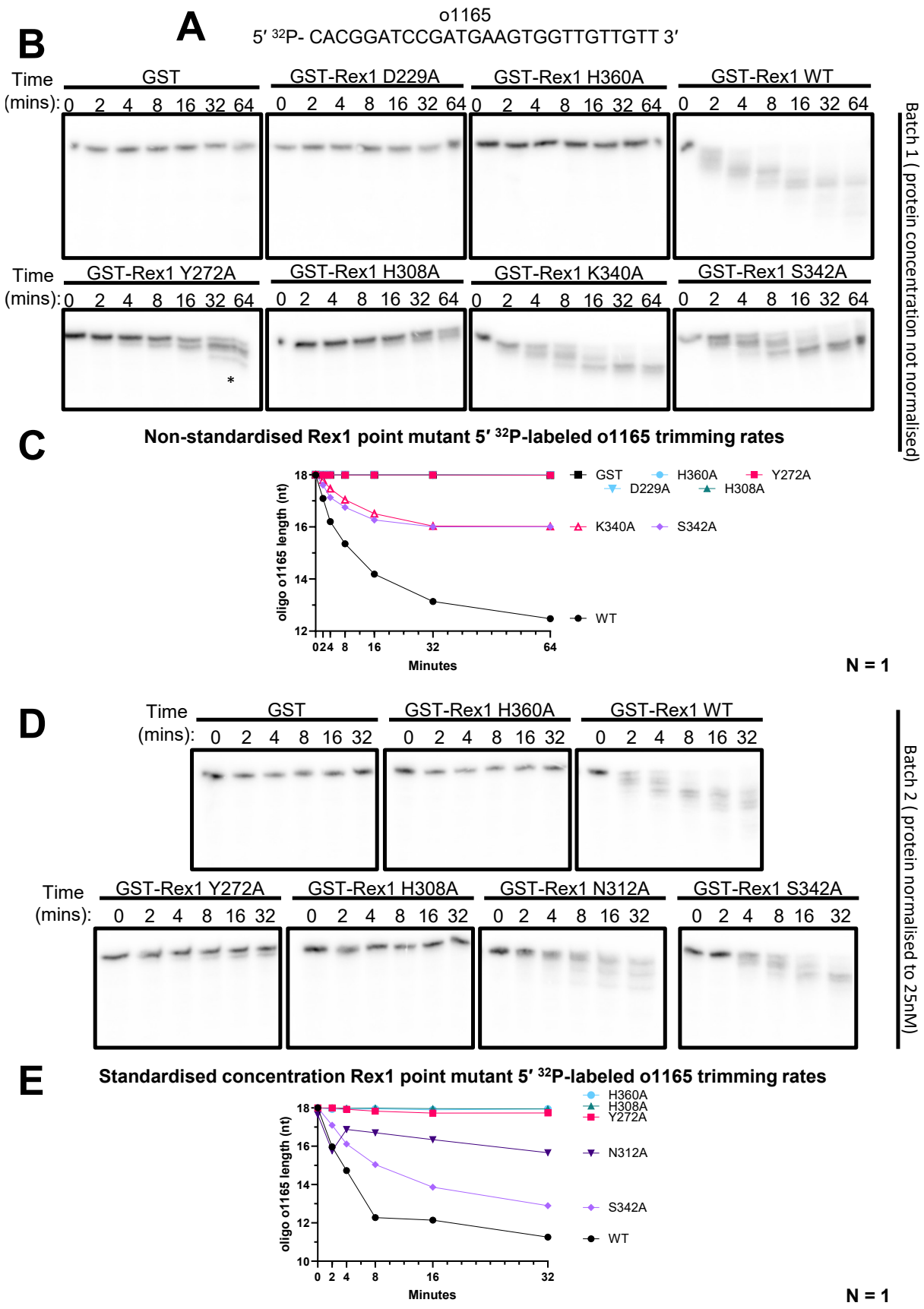
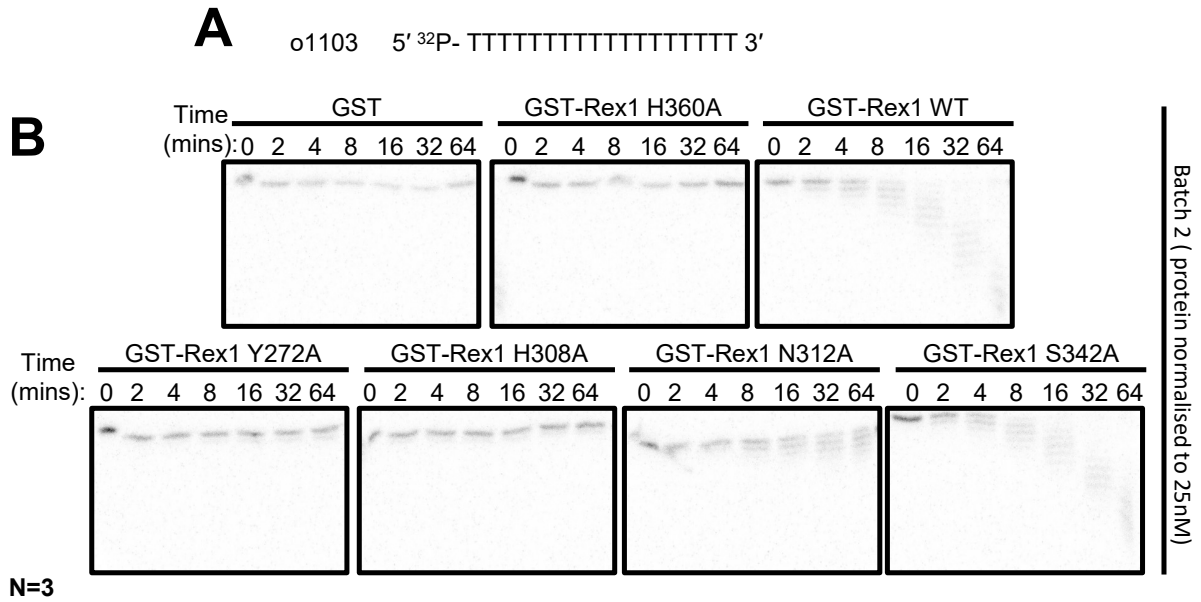

A
o1165
5′ 32P- CACGGATCCGATGAAGTGGTTGTTGTT 3′
B
Time (mins):
GST
GST-Rex1 D229A
GST-Rex1 H360A
GST-Rex1 WT
0 2 4 8 16 32 64
GST-Rex1 Y272A
GST-Rex1 H308A
GST-Rex1 K340A
GST-Rex1 S342A
*
Batch 1 (protein concentration not normalised)
C
Non-standardised Rex1 point mutant 5′ 32P-labeled o1165 trimming rates
oligo o1165 length (nt)
Minutes
GST
H360A
Y272A
D229A
H308A
K340A
S342A
WT
N = 1
D
Time (mins):
GST
GST-Rex1 H360A
GST-Rex1 WT
0 2 4 8 16 32
GST-Rex1 Y272A
GST-Rex1 H308A
GST-Rex1 N312A
GST-Rex1 S342A
Batch 2 (protein normalised to 25nM)
E
Standardised concentration Rex1 point mutant 5′ 32P-labeled o1165 trimming rates
oligo o1165 length (nt)
Minutes
H360A
H308A
Y272A
N312A
S342A
WT
N = 1

Figure 5.12: Degradation of a 5´ ^{32}P-labelled mixed sequence oligo by Rex1 mutants
A: Sequence composition of 5´ ^{32}P -labelled DNA oligo substrate o1165 **B:** Degradation time courses. Batch 1 of GST-Rex1 purifications were incubated with 3.3nM substrate at 30°C. Rex1 concentrations used were not measured, with relative amounts visible in Figure 5.10A. * indicates two gel lanes with a broken wall in-between. **C** (and **E**): Band intensity for each nucleotide length was measured using SAFA software (Das *et al.*, 2005), allowing calculation of average oligo length at each time point. Plots show average measured oligo length over time. **D:** Degradation time courses. Batch 2 of GST-Rex1 purifications were used at 25nM final concentration with 3.3nM substrate.



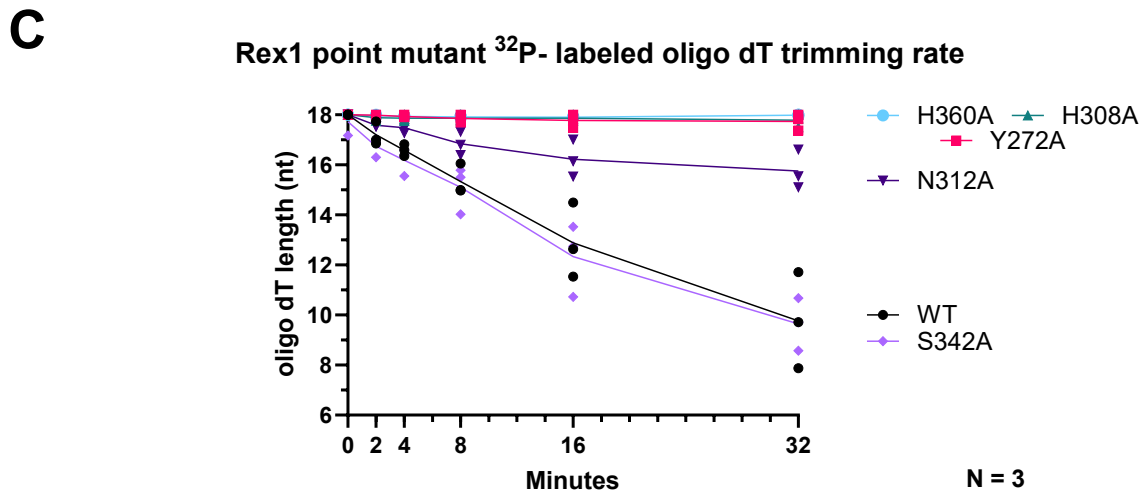


Figure 5.13: Degradation of a 5´ ^{32}P-labelled sequence oligo dT by Rex1 mutants
A: Sequence composition of 5´ ^{32}P -labelled DNA oligo substrate o1103 **B:** Degradation time courses. Batch 2 of GST-Rex1 purifications were used at 25nM final concentration with 3.3nM substrate. Phosphoimages representative of three technical replicates (three assays run using the same batch of protein). **C:** Quantification of triplicate time courses. Band intensity for each nucleotide length was measured using SAFA software (Das *et al.*, 2005), allowing calculation of average oligo length at each time point. Plotted lines show average measured oligo length over time, with all measured lengths from three replicates shown.

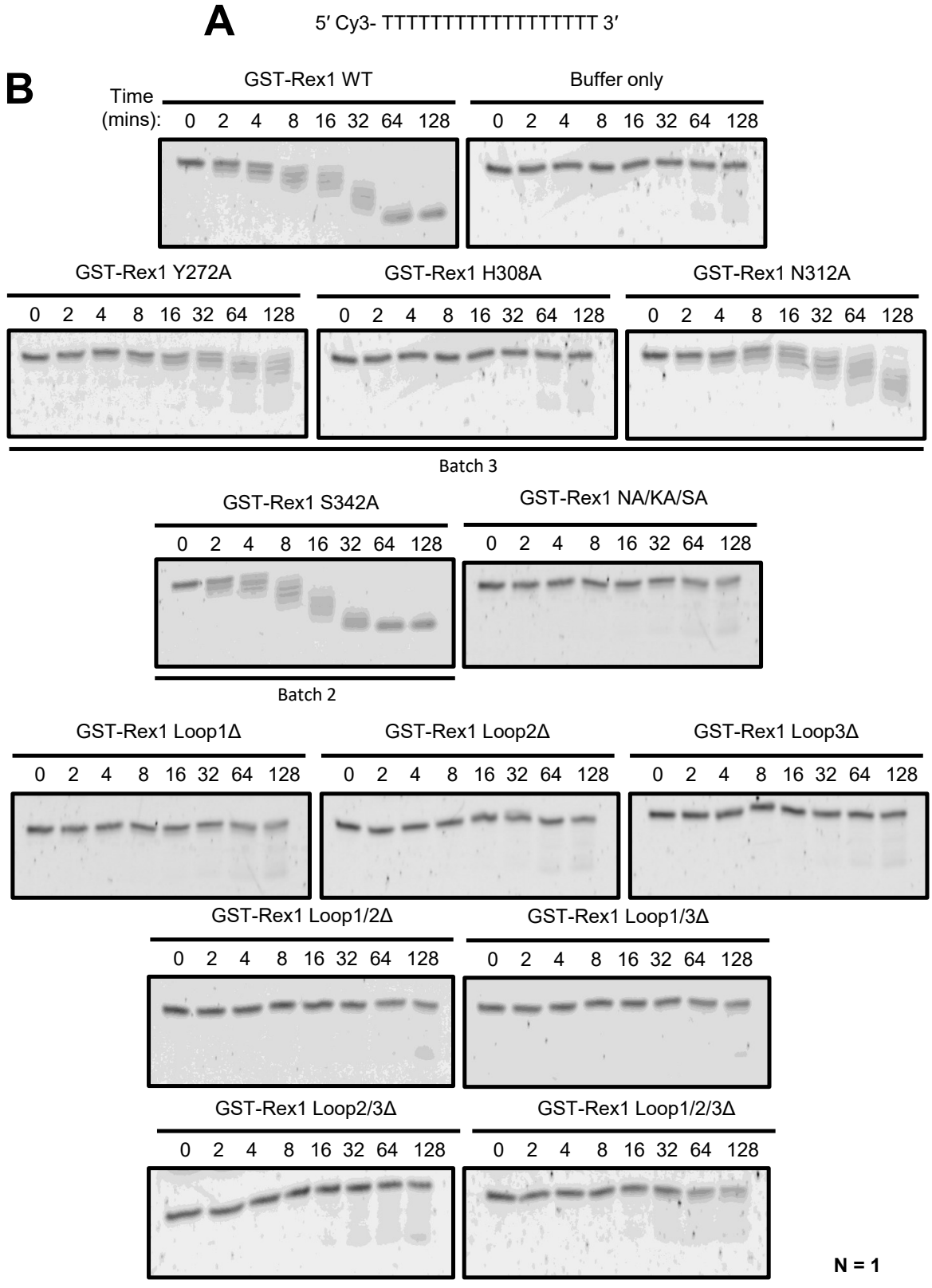


Figure 5.14: Degradation of a 5´ Cy3-labeled sequence oligo dT by Rex1 mutants
A: Sequence composition of 5´ Cy3-labelled DNA oligo substrate **B:** Degradation time courses. 25nM of protein was incubated with 3.3nM 5´ Cy3-labelled substrate in 30μl. Purification batch 3 of GST-Rex1 mutant Y272A, H308A and N312A were used, alongside purification batch 2 GST-Rex1 S342A. All time course images are in-gel fluorescence scans. Each assay was quantified, shown in Figure 5.15.

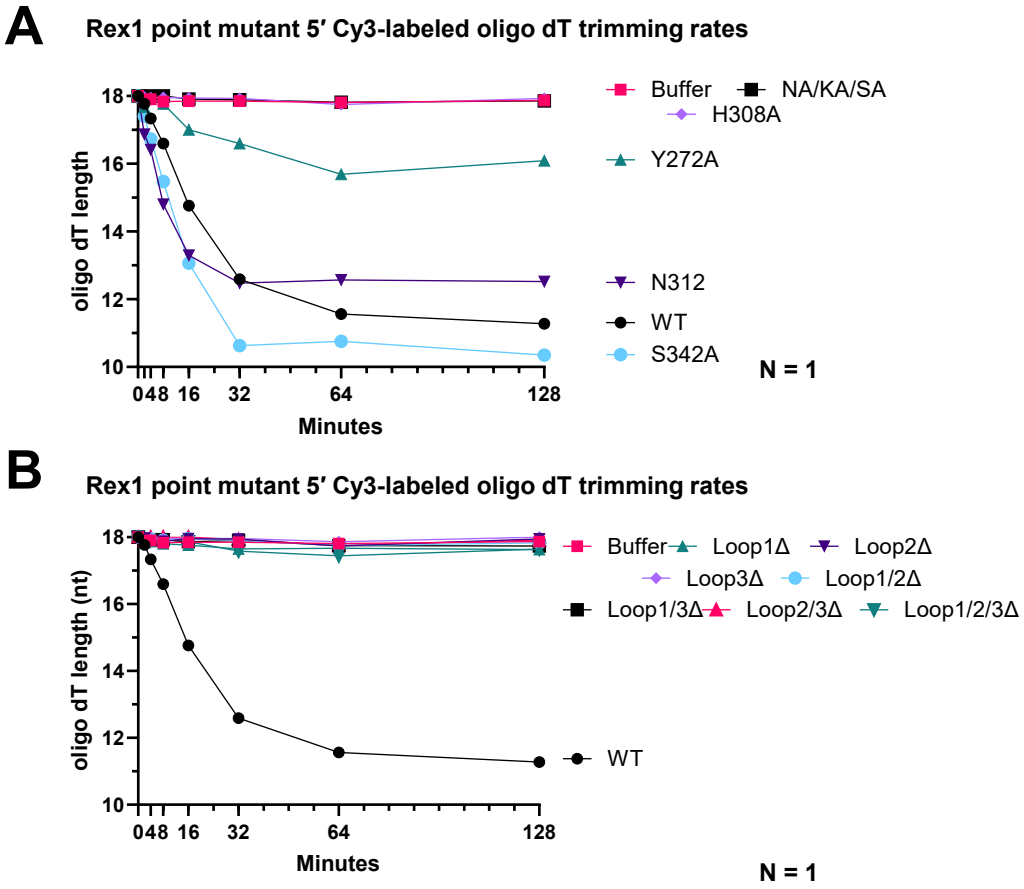


Figure 5.15: Quantification for degradation of a 5´ Cy3-labeled sequence oligo dT by Rex1 mutants Quantification of the degradation time courses shown in Figure 5.14. Band intensity for each nucleotide length was measured using SAFA software (Das *et al.*, 2005), allowing calculation of average oligo length at each time point. Plots show average measured oligo length over time**.** **A**: Degradation time courses showing relative activity of rex1 exonuclease domain point mutants. **B**: Degradation time courses shown relative activity of rex1 RYS domain mutants.

this residue is predicted to serve as a 'stopper' mechanism that uses an aromatic ring to stack with the 3´-most base of a substrate as recently described by Schäfer *et al*. (2019) and Tang *et al*. (2019) in a pair of structural studies on yeast Pan2. The Y272A mutant appears to demonstrate enhanced activity against the 5´-Cy3- labelled substrate, but it is worth noting that different purified batches of GST-rex1 Y272A were used for the 5´-^{32}P- labelled (Figure 5.13; batch 2) and 5´-Cy3- labelled (Figure 5.14; batch 3).

The most complete loss of Rex1 function among the point mutants (with the exception of H360A) was for the H308A mutant. The equivalent residues in *A. thaliana* SDN1 (H223; Chen *et al*. 2018) and PARN (H280; Wu *et al*. 2009) haven't been mutated to examine the impact on their respective proteins, but in each protein the contacts that the histidine make with the substrate seem to be shared redundantly with neighbouring residues: H223, S224, and the main chain of P257 interact with the third 3´-most nucleotide's 2´OH in SDN1 (Chen *et al*., 2018), and H280, K319, and the main chain of L336 share an interaction with one of the phosphates of the m7GpppG 5´ cap (Wu *et al*., 2009). The total loss of activity seen for the Rex1 H308A mutant by comparison suggests a catastrophic misfolding of the active DEDD site, or at least misfolding of the loop harbouring the EXOII cation-binding aspartate D313.

5.2.6 The contribution of Rex1 sequence features to *in vivo* Rex1 function

In vitro exonuclease activity against single stranded oligos represents a narrow view of Rex1 function. In order to relate the contribution of Rex1 sequence features to broader whole-cell phenotypes, plasmid-borne *rex1* mutants were used to complement the phenotypes of two yeast strain backgrounds: *rex1*Δ, and *rex1*Δ/*rrp47*Δ (Costello *et al*., 2011). Only a subset of *rex1* mutants were available in yeast expression vectors: H360A, Y272A, H308A, N312A, K340A, S342A, Loop1Δ, Loop2Δ, and Loop3Δ. Each plasmid contained a CENP replication origin (Christianson *et al*., 1992), the promoter region of the *RRP4* gene (Mitchell *et al*., 1996), an N-terminally zz-tagged *rex1* ORF and the 3´ UTR of the genomic *REX1* locus (Hama Soor, 2017).

The expression of each plasmid-encoded *rex1* mutant was assessed in the *rex1*Δ background by Western blotting of whole cell lysates, as shown in Figure 5.16. All mutants demonstrated comparable levels of expression to the plasmid-encoded *zz-REX1* wild type, with the exception of zz-*rex1* H308A, which demonstrated a clear and reproducible reduction in expression at 15% the signal of the plasmid-expressed wild type zz-Rex1.

5.2.6.1 *Rex1* mutants demonstrate variable ability to complement a *rex1*Δ/*rrp47*Δ plasmid shuffle strain

The same set of mutant zz-*Rex1* plasmids were transformed into the *rex1*Δ/*rrp47*Δ plasmid shuffle strain to assess their ability to complement the synthetic lethality phenotype, as shown in Figure 5.17. All mutants demonstrated reduced colony growth in comparison to the wild type control. For the five point mutants, the N312A, K340A, and S342A each demonstrated a modest reduction in growth rate, the Y272A demonstrated a large reduction in growth rate, while the H308A demonstrated an almost complete loss of growth. This distribution of activities correlates closely with the *in vitro* activity of these mutants shown in Figure 5.12, with N312A, K340A, and S342 demonstrating the most Rex1 activity, and the Y272A and H308A demonstrating the least.

All three loop truncation mutants demonstrated a modest reduction in growth, with the Loop1Δ mutant demonstrating the slowest growth rate. Unlike the correlation seen for the H308A mutant's *in*

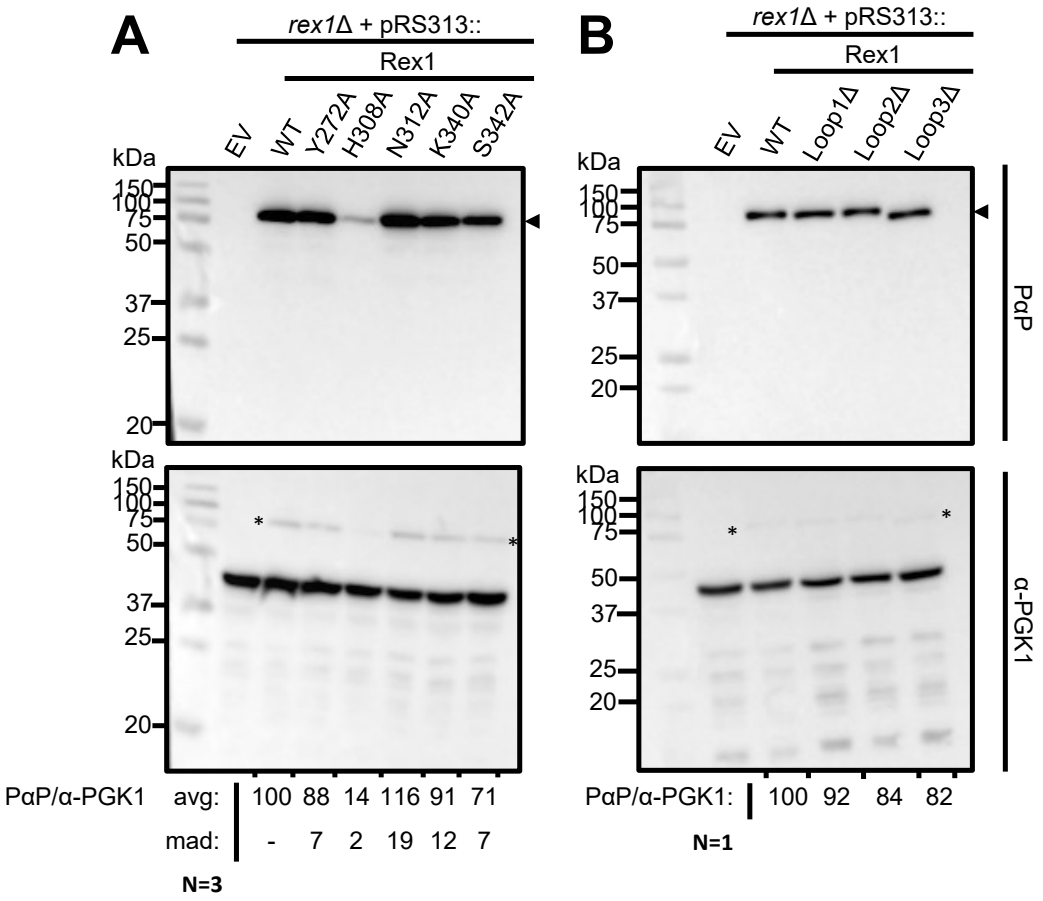


Figure 5.16: Plasmid expression of zz-tagged *rex1* mutants in *rex1*Δ yeast.
Both panels show 12% denaturing acrylamide gels transferred onto Protran nitrocellulose membrane. The expected size of full-length zz-Rex1 is 77kDa, indicated with a black triangle. The expected size of the Rex1 loop 3 mutant is 73kDa. Each lane loaded with whole-cell extract equivalent to $0.5OD_{600nm}$ of yeast. '*' indicates residual PαP antibody activity from incomplete inactivation by NaN_3 **A:** Expression of *Rex1* alanine missense mutants. Band intensities from raw chemiluminescence images of each PαP blot were quantified for the 3 replicates , normalised by the band intensities of the α-PGK1. Average signal intensities are shown normalised against wild type Rex1 (100), with (mad) values. **B:** Expression of *rex1* loop truncation mutants. Loop1Δ refers to truncation of residues 146-154, Loop2Δ refers to truncation of residues 459-472, and Loop3Δ refers to truncation of residues 506-544. All truncated loops were replaced with short serine-glycine-rich linkers. Only a single Western blot was performed, with band intensities from raw chemiluminescence image shown, normalised against the band intensities of the α-PGK1 blot.

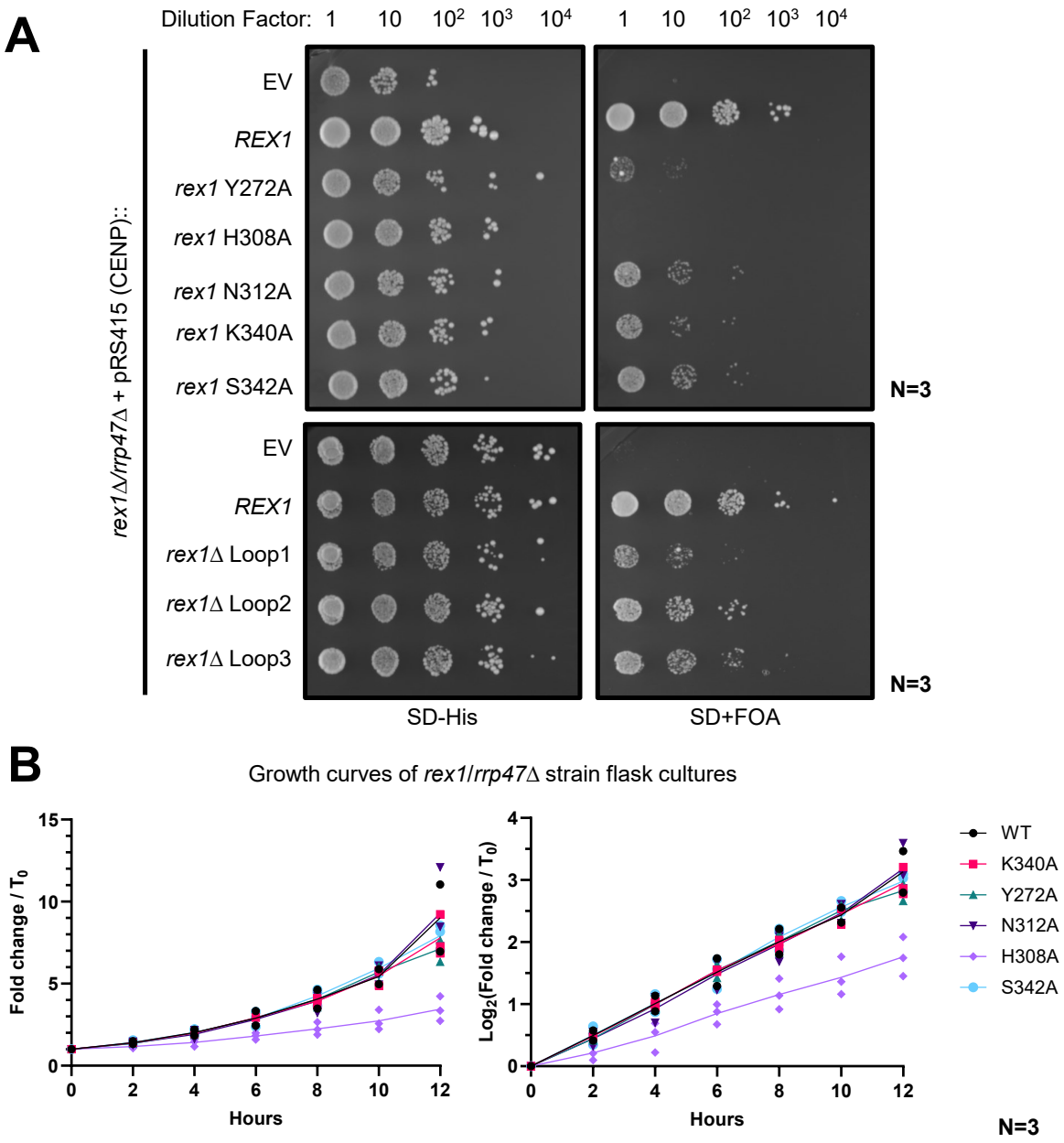


Figure 5.17: Plasmid shuffle assay of *rex1* mutant viability demonstrates variable levels of *Rex1* function in *rex1*Δ/*rrp47*Δ yeast.
A: Plasmid shuffle assay of zz-*Rex1* mutant activity. All plates were imaged 5-7 days post-seeding. Each plate was seeded as three technical replicates. **B**: Growth curves of 3 5-FOA isolate 50ml SD cultures from each *rex1* point mutant, with 2 5-FOA isolates of wild type *REX1*. Growth was tracked for 12 hours after reaching an OD_{600nm} of at least 0.1, cultures were diluted 1/10 upon reaching an OD_{600nm} of 1.0 to preserve logarithmic growth. Both normalised (/T_0: divided by time point zero) OD_{600nm} and Log2(normalised OD_{600nm}) are shown on the left and right graph respectively.

vitro and *in vivo* Rex1 activity, the lack of *in vitro* activity seen for the Loop mutants does not seem to result in as dramatic a loss of Rex1 function as measured by its ability to complement the synthetic lethality phenotype. It is unclear how this could be the case, as the generic exonuclease activity of Rex1 is presumably what enables it to complement the *rex1*Δ/*rrp47*Δ plasmid shuffle strain. Based on these data, the loss of these Loop structures results in recombinant Rex1 that is unable to trim single-stranded oligo dT, but is still capable of sufficiently trimming RNAs, most likely snoRNAs (Garland *et al*., 2013), to rescue the synthetic lethality of the *rex1*Δ/*rrp47*Δ genotype. This lack of correlation between *in vivo* Rex1 activity and *in vitro* trimming activity against oligo dT indicates that DNA oligos may not be representative of Rex1's physiological substrate processing activity.

For all transformants with the exception of the empty vector control clonal isolates could be obtained, referred to herein as 'FOA isolates'. As 5-FOA is a mutagen, three isolates were taken for each condition to account for the possibility of mutagenesis, although time limitations led to the analysis of only one isolate per mutant.

5.2.6.2 Northern blotting reveals two subtypes of 5S rRNA misprocessing by *rex1* mutants

In order to assess the RNA processing defects that arise from the loss of each Rex1 sequence feature, flask cultures were grown for the *rex1*Δ background complemented with *Rex1* mutant plasmids, and for FOA isolates obtained from the plasmid shuffle assay. Whole cell RNA lysates were prepared, and 2µg of each were resolved on 8% Acrylamide 0.5x TBE gel electrophoreses until either tRNA, or 5S rRNA had migrated to the bottom, as shown in Figure 5.18A and B respectively. The two different run lengths allowed better resolution of two aspects of 5S rRNA processing defects: the short run gel enabled detection of the truncated 5S rRNA species seen in exosome mutants with *REX1* (Kadaba *et al*., 2006; Schneider, Kudla, Wlotzka, Tuck and Tollervey, 2012; Garland *et al*., 2013; Feigenbutz *et al*., 2013a; Han and van Hoof, 2016; Delan-Forino *et al*., 2017), and the long run gel enabled better resolution of 3´ 5S rRNA extensions that may accumulate from incomplete processing by Rex1 mutants (Piper *et al*., 1983; van Hoof *et al*., 2000a).

The 8% acrylamide gels were first imaged using ethidium bromide staining as shown in Figure 5.18 before transfer onto N+ nylon membrane for Northern hybridisation. For a full loss of Rex1 function in a *rex1*Δ strain, a compact band corresponding to a 3 nucleotide extension was expected (Piper *et al*., 1983; van Hoof *et al*., 2000a). In the *rex1*Δ/*rrp47*Δ background no exacerbation of the 5S processing defect in the absence of *REX1* was expected, but the presence of wild type Rex1 activity instead leads to the accumulation of a truncated species (Copela *et al*., 2008), which has been validated as a 3´ processing product previously in the Mitchell lab through mapping with probes at the predicted truncation site (Garland *et al*., 2013).

Of the five point mutants, the H308A mutant demonstrates the clearest 5S processing phenotypes: in the *rex1*Δ background a mixed pool of fully processed and unprocessed 5S rRNA can be seen. Unexpectedly, *rex1*Δ/*rrp47*Δ complemented with *rex1*H308A seemed to have a slightly higher proportion of mature 5S rRNA compared to *rex1*Δ complemented with *rex1*H308A, with the loss of *RRP47* potentially suppressing the *rex1*Δ 5S rRNA processing defect. A milder, but still visible by Northern blot equivalent of this defect can be seen for the N312A and K340A mutants, which are similarly suppressed in the *rex1*Δ/*rrp47*Δ background. The formation of the 5S truncation product in the *rex1*Δ/*rrp47*Δ background seems to broadly correlate with the pattern of Rex1 activity: the activity of the Y272A, N312A, K340A,

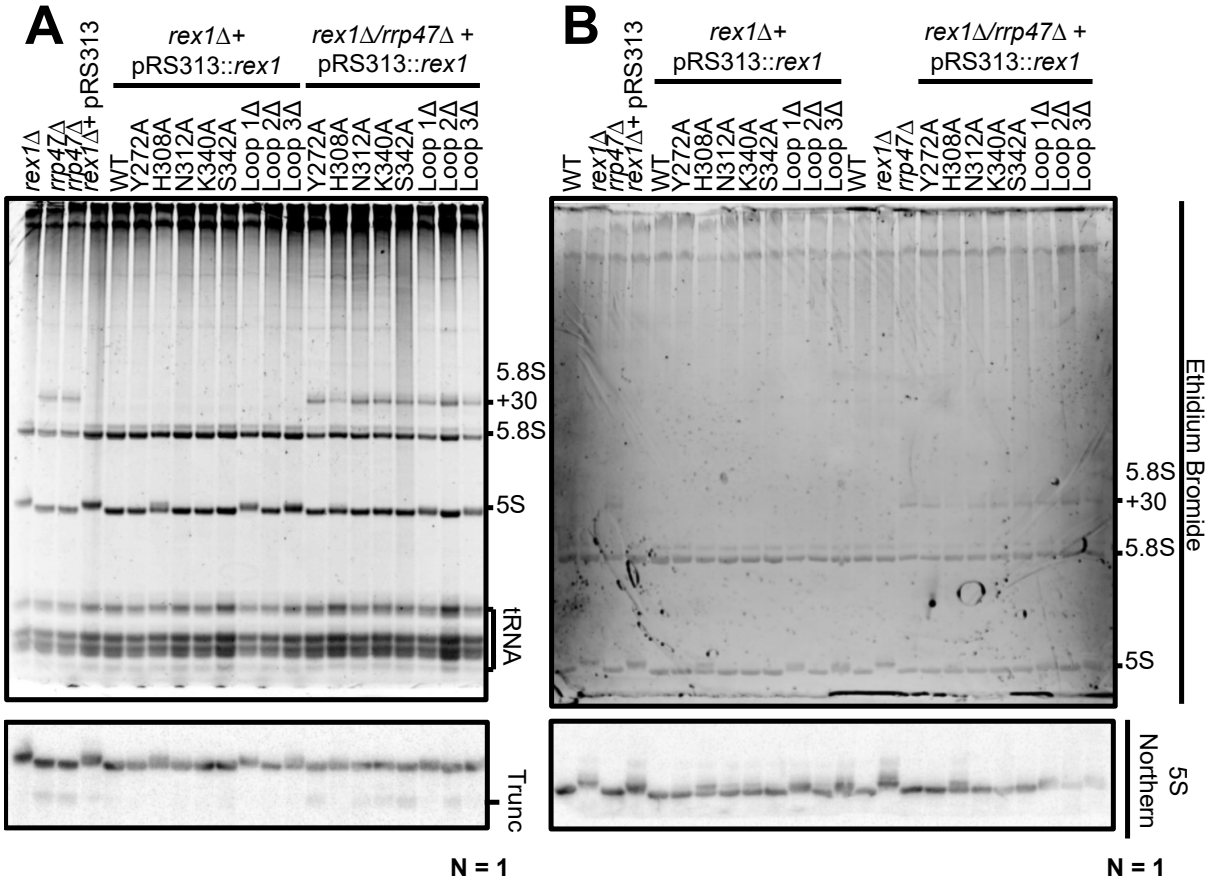


Figure 5.18: Ethidium bromide and Northern blot analysis reveal rRNA processing defects in *Rex1* mutants

Ethidium-stained 8% acrylamide 0.5% TBE gel of whole-cell RNA extracts from logarithmically growing yeast. The following strains are shown: 'WT' BY4741, '*rex1*Δ' P550 transformed with a HIS3-bearing plasmid (pRS313), '*rex1*Δ/*rrp47*Δ' P1604 plasmid shuffle strain isolated from 5-FOA plates after 7 days of growth. Each isolate was taken in triplicate, isolate 1 for each mutant is shown. For the ethidium-stained gels, the most abundant RNA species are indicated. Panels marked '5S' represent the same gels transferred onto N+ membrane and probed with ^{32}P-labelled oligo o925. **A** and **B** represent largely the same samples, with B run for twice as much time as A to allow better resolution of bands below and above the 5S species respectively. Two identical *rrp47*Δ samples were loaded in A due to a labelling error.

and S342 seems sufficient for the formation of a 5S truncation product, which is seemingly equal in intensity for each mutant in spite of their relative activities. This may suggest that even the same residual activity in these mutants that's sufficient for the formation of mature 5S may be sufficient to enable the misprocessing of 5S in *rrp47*Δ strains.

Of the Rex1 Loop mutants, Loop1Δ and Loop3Δ have the greatest impact on 5S processing, with a similar correlation to the formation of truncated products as shown by the point mutants. The Loop1Δ and Loop3Δ mutants both demonstrate variably extended 5S rRNA pool that lack a mature band size, which suggests a different mechanism of misprocessing in comparison to the H308A mutant. The Rex1 H308A mutant is seemingly only able to process a small fraction of 5S, likely due to the low expression evident in the Western blot in figure 5.16, but the 5S that is processed seems to be processed to completion. The Rex1 Loop1Δ and Loop3Δ mutants by contrast still seem to be capable of engaging the 3 nucleotide extension that's inaccessible to other exonucleases, but the resulting processing seems to fall short of completion. This indicates a different basis underpinning the low functionality of Rex1 H308A compared to Rex1 Loop1Δ and Loop2Δ: the Rex1 H308A mutation seems to result in a small fraction of fully functional Rex1, while the loss of Loop1 and Loop3 seem to remove a mechanistic feature of Rex1 that is required for the trimming of nucleotides found in the closest proximity to the 5S terminal stemloop.

5.2.6.3 Northern analysis of representative Rex1-targeted tRNAs reveals 3´ extension in *rex1* mutants that are suppressed by *rrp47*Δ

Yeast strains lacking *REX1* are known to be defective in the processing of certain tRNAs. One tRNA that seems to be dependent on Rex1 for maturation is $tRNA^{Arg3}$, which is transcribed as a dicistronic $tRNA^{Arg3}$-$tRNA^{Asp}$ transcript. RNase P cleavage of this transcript at the 5´ of the $tRNA^{Asp}$ leaves an unusual 3´ trailer for $tRNA^{Arg3}$ that can't seem to be processed by the endonuclease activity later ascribed to tRNase Z (Engelke *et al*., 1985; Piper and Stråby, 1989; van Hoof *et al*., 2000a; Skowronek *et al*., 2014). Based on the initial mapping of processing intermediates (Engelke *et al*., 1985), it is predicted that there may be two processing steps that result in the formation of RNase T-like substrates: before its removal by RNase P, the 5´ trailer is initially able to base pair with the first 5 nucleotides of the 10 nucleotide 3´ trailer, resulting in an extended acceptor stem of the tRNA (Engelke *et al*., 1985; Piper and Stråby, 1989), which upon removal results in the usual acceptor stem structure that resembles RNase T substrates. It is known that deletion of *REX1* leads to the accumulation of precursors containing the final 5 nucleotides, however the other $tRNA^{Arg3}$ precursors that are visible by Northern blotting of *rex1*Δ yeast remain to be conclusively identified (van Hoof *et al*., 2000a).

According to the database RNA Central as of September 2023, only 4 of the 11 yeast $tRNA^{Arg}_{(UCU)}$ genes are expressed in this manner; 6 are expressed as monocistronic transcripts, and one expressed as the downstream cistron of a dicistronic $tRNA^{Asp}_{(GUC)}$-$tRNA^{Arg}_{(UCU)}$ transcript. As a result, probes that are complementary to the mature tRNA will detect the mature species, even if no mature length $tRNA^{Arg}_{(UCU)}$ species are produced from the 4 dicistronic loci. This probe design may be a consequence of the short length of the 3´ trailer sequence, which is 10 nucleotides long.

A further tRNA that was observed to be processed by Rex1 in wild type cells is the intron-containing $tRNA^{Lys}_{(UUU)}$ (Copela *et al*., 2008). The impact of *REX1* loss on this tRNA can be seen with the accumulation of a precursor that corresponds to a mature 5´ terminus, but an unprocessed 3´ terminus

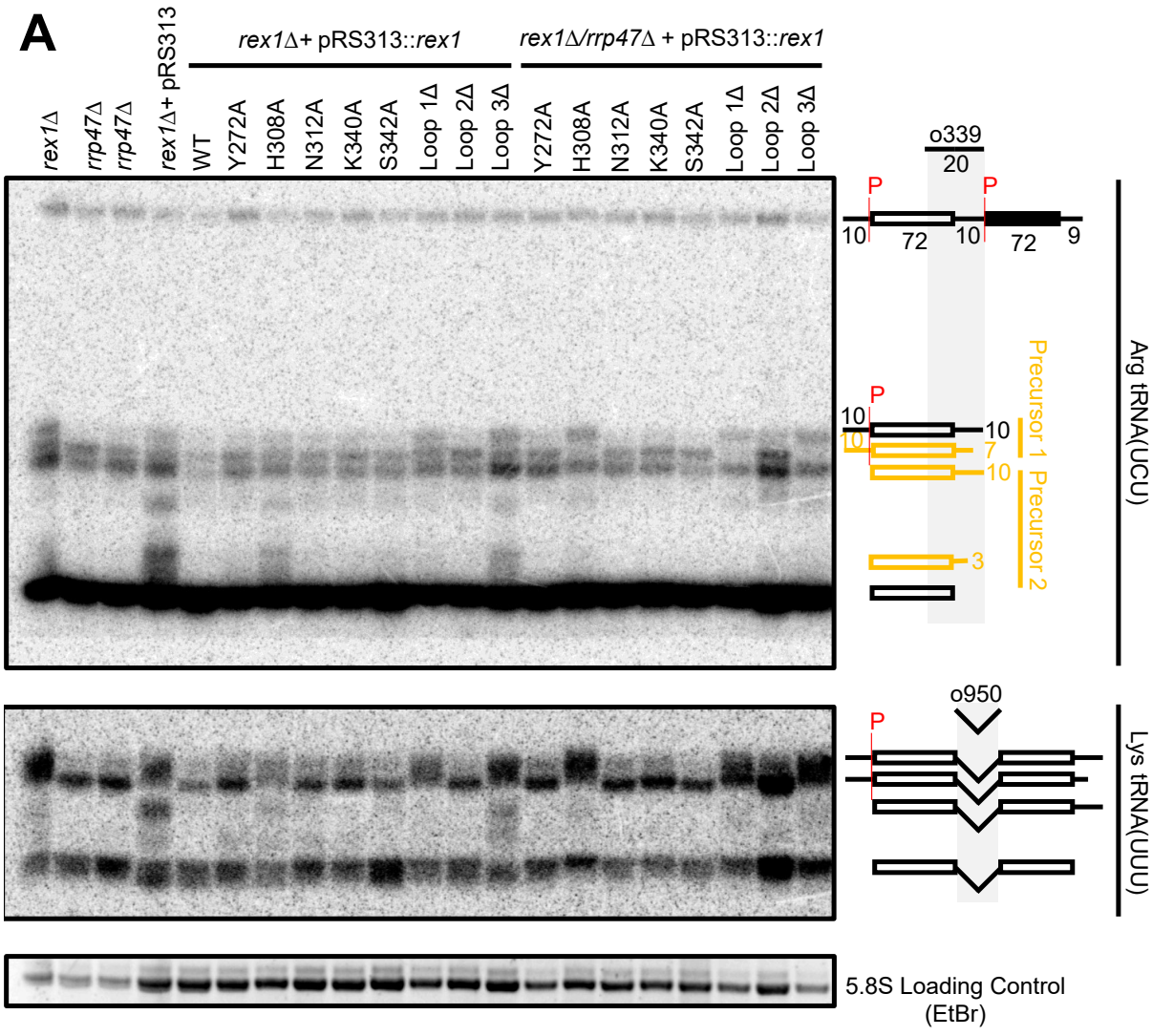

A
rex1Δ
rrp47Δ
rrp47Δ
rex1Δ+ pRS313
rex1Δ+ pRS313::rex1
rex1Δ/rrp47Δ + pRS313::rex1
WT
Y272A
H308A
N312A
K340A
S342A
Loop 1Δ
Loop 2Δ
Loop 3Δ
o339
20
P
10
72
10
72
9
Precursor 1
Precursor 2
Arg tRNA(UCU)
o950
Lys tRNA(UUU)
5.8S Loading Control
(EtBr)
N = 1
B
Precursor 1
RNase P
Rex1
Precursor 2
Key:
- Precursor sequence
- Mature sequence tRNA Arg
- Rex1 Substrate

Figure 5.19: Northern blot analysis of *rex1* tRNA processing defects.
A: Northern blot of N+ membranes transferred with 8% acrylamide 0.5% TBE gel of whole-cell RNA extracts from logarithmically growing yeast. The following strains are shown: '*rex1*Δ' P550 transformed with a HIS3-bearing plasmid (pRS313), '*rex1*Δ/*rrp47*Δ' P1604 plasmid shuffle strain isolated from 5-FOA plates after 7 days of growth. Each isolate was taken in triplicate, isolate 1 for each mutant is shown. Diagrams of probe annealing positions are shown, with grey boxes indicating annealed sequence. $tRNA^{Arg3}_{(UCU)}$ diagrams based on the source of the probe, van Hoof *et al*. (2000a); black circles indicate the presence of each precursor for each condition, with the *rex1*Δ/*rrp47*Δ condition approximated by the hypomorphic *rex1* mutants in the *rrp47*Δ background. $tRNA^{Lys}_{(UUU)}$ diagrams based on the source of the probe, Skowronek *et al*. (2014). Red 'P's indicate sites of RNase P cleavage. The black box shown in the $tRNA^{Arg3}$ cartoon represents the downstream $tRNA^{Asp}$ cistron. Yellow gene cartoons represent estimated precursors based on the folding of $tRNA^{Arg3}_{(UCU)}$ shown in B. Ethidium bromide (EtBr)- stained 5.8S rRNA reproduced from Figure 5.18A as a loading control. **B:** Schematic of $tRNA^{Arg3}_{(UCU)}$ reproduced from Figure 1.5, with the cartoons of known (black) and estimated (yellow) precursor sequences matched to corresponding 3´ cleavage positions by vertical lines.

(Copela *et al*., 2008). The impact of *RRP6* deletion has also been examined for this tRNA (Skowronek *et al*., 2014), revealing a small increase in the proportion of 3´ and 5´ extended species indicating a general delay in processing compared to wild type.

The 'short run' gel shown in Figure 5.18A was probed with an oligo spanning the mature $tRNA^{Arg}_{(UCU)}$ (van Hoof *et al*., 2000a) followed by the 3´ trailer, and an oligo covering the intron of the intron-containing $tRNA^{Lys}_{(UUU)}$ as shown in Figure 5.19. While the specific identities of $tRNA^{Arg}_{(UCU)}$ precursors are uncertain, there seems to be a consistent pattern for strains lacking *REX1* activity. In strains with an active *RRP47* gene, the previously reported 5´-processed and 3´-unprocessed species in both RNAs (Copela *et al*., 2008; van Hoof *et al*., 2000a) can be seen for the empty vector control, the H308A mutant, and the Loop3Δ mutant. A notable exception to this is the *rex1* strain, which doesn't seem to present this band; this is likely a blotting artefact which seems to be evident in a horizontal fade in the smaller species. Interestingly, the $tRNA^{Arg}_{(UCU)}$ was able to detect the mature length species, which wasn't seen in the study the probe originated from (van Hoof *et al*., 2000a); this indicates that the post hybridisation washes used here may be less stringent.

A heterogeneous 3´ extension can be seen clearly with the $tRNA^{Arg}_{(UCU)}$ probe for the H308A and Loop3Δ mutants, but with a less intense signal for the Loop1Δ mutant and the other point mutants. The Loop1Δ mutant differs from the four functioning point mutants in the presence of an additional precursor; this may suggest that Loop1 has a role in the processing event that leads to the loss of this intermediate, estimated here as the unprocessed product of the 3´ RNase P cleavage, but not the trimming of the final 3´ extension that instead seems to require the action of Loop3. The difference between the intensity of these extensions is a little narrow, and may warrant repeating with a longer exposure during radiography. The 3´ extension and fully unprocessed 3´ species can be seen more clearly in the $tRNA^{Lys}_{(UUU)}$ blot, suggesting the role of Rex1 in processing these intermediates may be a general feature of exonucleolytic tRNA processing that relies on Loop3, whereas the processing of the raw 3´ RNase P cleavage product specific to dicistronic tRNA maturation seems to depend on both Loop1 and Loop3.

For tRNA processing in the *rrp47*Δ background, there seems to be a suppression effect mediated against the misprocessing phenotypes seen in the presence of active *RRP47*. One explanation for this

may be that the full absence of Rex1 and Rrp6 activity lead to degradation of the transcripts instead of the accumulation of unprocessed precursors, although there isn't an obvious reduction in the intensity of the mature species of both tRNAs that would corroborate this. Another explanation may be that the total absence of exonuclease activity may leave intact the uridine tract required for Lhp1 binding (Yoo and Wolin, 1997), allowing processing by tRNase Z via the endonucleolytic pathway (Skowronek *et al*., 2014). It is worth noting that endonucleolytic cleavage of discistronic $tRNA^{Arg}_{(UCU)}$ may require prior removal of the 5´ leader, as the stemloop-adjacent polyU tract may be base paired as part of the predicted precursor extended acceptor stem (Engelke *et al*., 1985). If this is the case, then this may indicate the identities for the middle two intermediates that accumulate in the *rex1*H308A/*rrp47*Δ, *rex1*Loop1Δ/*rrp47*Δ, and *rex1*Loop3Δ/*rrp47*Δ mutants: the raw product of the 3´ RNase P cleavage site, and the subsequent cleavage product of the 5´ RNase P cleavage site. It is unlikely that these bands represent maturation intermediates from non dicistronic $tRNA^{Arg}_{(UCU)}$, as the 3´ 10 nucleotides of the 20 nucleotide probe wouldn't anneal with the trailer sequence.

5.2.6.4 Northern analysis of snoRNAs and snRNAs reveals exacerbated processing defects in *rex1*Δ/*rrp47*Δ compared to *rex1*Δ

Previous work in the Mitchell lab has observed that the joint loss of Rex1 and Rrp47 activity doesn't lead to exacerbated 5S or tRNA processing defects, but does result in exacerbated processing defects for snoRNAs including snR38, snR13, and U14, and the snRNA U6 (Costello *et al*., 2011; Garland *et al*., 2013). The strains demonstrating this synergetic defect were variants on a *rex1*Δ/*rrp47*Δ strain supported by hypomorphic Rrp47 function taking two forms: i) *rrp47*ΔC, a truncation mutant of *rrp47* removing the RNA-binding domain, but leaving the Rrp6 and exosome-interacting C1D domain intact- giving an exosome lacking the Rrp47 RNA-binding domain (Costello *et al*., 2011); ii) and the *rrp47*ΔC combined with a galactose-inducible Rrp6 N-terminal domain decoy fragment that inhibits the association of Rrp6 with the exosome- effectively an inducible depletion of Rrp6's association with the exosome (Garland *et al*., 2013). It is however worth noting that the N- termini of Rrp6 and Rrp47 found in this latter hypomorphic complex are still sufficient to mediate the exosome's interaction with the TRAMP complex component Mtr4 (Schuch *et al*., 2014).

An increase in polyadenylation was observed for each of the aforementioned RNA species, leading to the model that the *rex1*Δ/*rrp47*Δ synthetic lethality relationship may result from the titration of polyA-based RNA surveillance machinery by a broad increase in defective snoRNA production (Garland *et al*., 2013). These defective processing events were a result of hypomorphic exosome function; in the analysis presented here, it is possible to observe the defective processing events that conversely result from hypomorphic Rex1 function.

The wild type 3´ maturation of the box C/D snoRNA snR38 is thought to be primarily performed by the exosome, helped in part by the ability of Rrp47 to engage with both the protein and RNA components of the nascent RNP (Costello *et al*., 2011). In the absence of Rrp6/Rrp47, a mixed population of 3 nucleotide-extended and mature length snR38 accumulates (Mitchell *et al*., 2003), as seen in the lane 2 and 3 of Figure 5.20A. In a rare case where the phenotypes of *rrp6*Δ and *rrp47*Δ differ, *rrp6*Δ mutants accumulate the 3 nucleotide-extended species with no mature species (Mitchell *et al*., 2003; Stead *et al*., 2007). The steady-state snR38 population of a *rex1*Δ mutant differs little from wild type, but when combined with a hypomorphic exosome mutant lacking the C-terminus of Rrp47 (Garland *et al*., 2013),

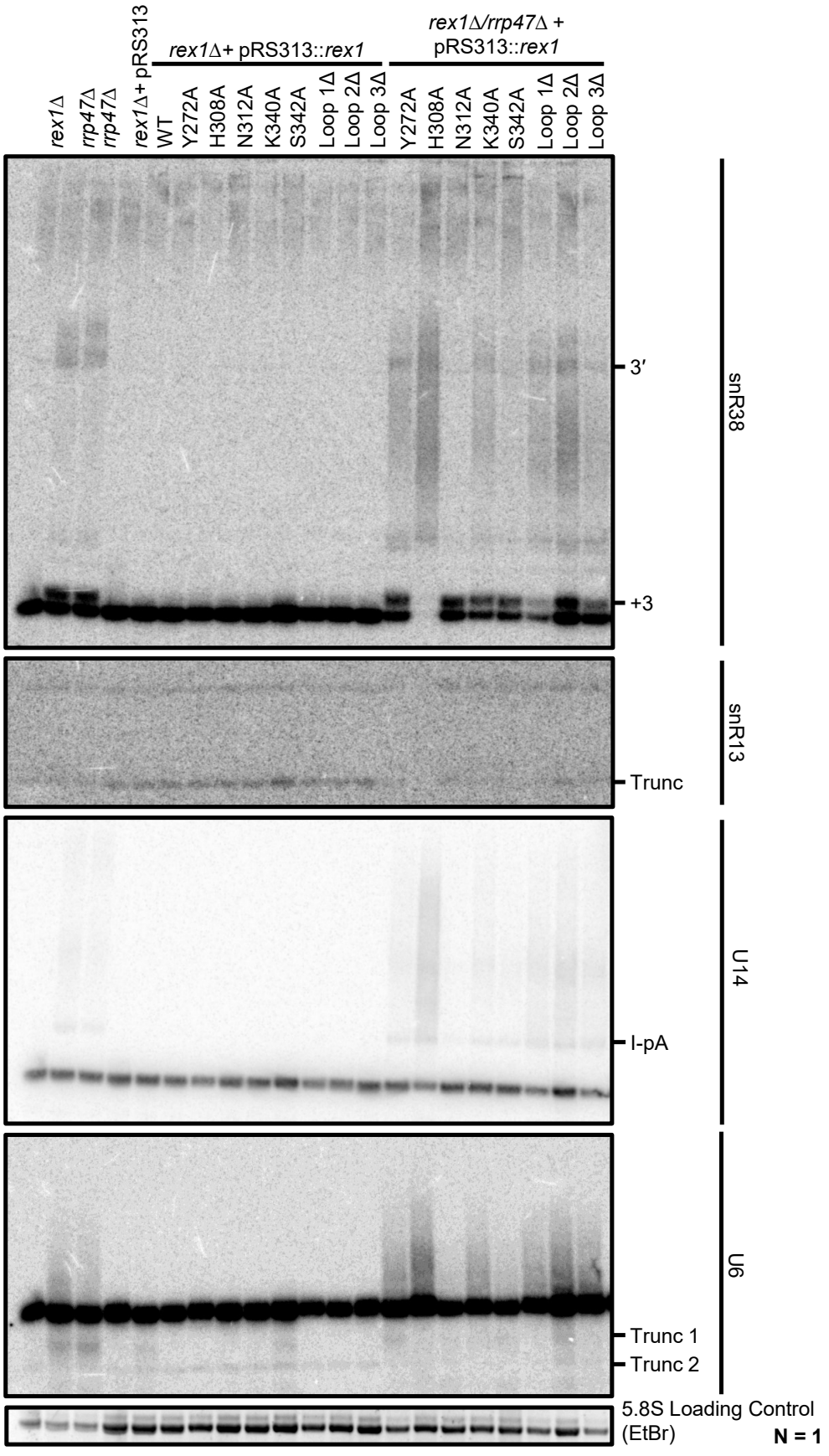

rex1Δ
rrp47Δ
rrp47Δ
rex1Δ+ pRS313
rex1Δ+ pRS313::rex1
rex1Δ/rrp47Δ + pRS313::rex1
WT
Y272A
H308A
N312A
K340A
S342A
Loop 1Δ
Loop 2Δ
Loop 3Δ
Y272A
H308A
N312A
K340A
S342A
Loop 1Δ
Loop 2Δ
Loop 3Δ
3′
+3
snR38
Trunc
snR13
I-pA
U14
Trunc 1
Trunc 2
U6
5.8S Loading Control (EtBr)
N = 1

Figure 5.20: Northern blot analysis of *rex1* snRNA and snoRNA processing defects.
Northern blots of an N+membrane onto which was transferred an 8% acrylamide 0.5% TBE gel of whole-cell RNA extracts from logarithmically growing yeast. The following strains are shown: '*rex1*Δ' P550 transformed with a HIS3-bearing plasmid (pRS313), '*rex1*Δ/*rrp47*Δ' P1604 plasmid shuffle strain isolated from 5-FOA plates after 7 days of growth. Each isolate was taken in triplicate, isolate 1 for each mutant is shown. 'Trunc' = truncation, 'I-pA' = the primary polyadenylation site. The following probes were used: snR38 - o243, snR13 - o821, U14 - o238, U6 - o517. Ethidium bromide (EtBr)- stained 5.8S rRNA reproduced from Figure 5.18A as a loading control.

an intense and highly heterogeneous 3´ extension defect is observed, although the strain still manages to accumulate comparable levels of mature-length snR38 to the wild type.

Hypomorphic *rex1* mutants seem identical to wild type in strains with an active copy of *RRP47* in terms of snR38 processing, but when combined with *rrp47*Δ a heavy processing defect is observed. The *rex1*ΔLoop1 shows a large decrease in the +3 species, with modest decrease in the mature species. By contrast the *rex1*ΔLoop3 mutant seems to be largely comparable to the *rrp47*Δ/*REX1* phenotype, with a minor reduction in the accumulation of the +3 species. The most intense processing defect is seen for the *rex1*H308A mutant, which features an almost total loss of both +3 and mature snR38 leaving a barely detectable doublet, with a comparable increase in the polyadenylated species. This distribution of phenotypes bears a resemblance to the complemented growth rates seen in the plasmid shuffle assay in Figure 5.17, which shows *rex1*ΔLoop1 mutant as the slowest growing, followed by *rex1*ΔLoop3, with the best growing mutant being *rex1*ΔLoop2. There seems to be an increase in polyadenylated species for the *rex1*ΔLoop2 mutant across the board, but this is most likely a result of the uneven loading that is visible in the 5.8S band of the ethidium bromide-stained gel in Figure 5.18A.

The maturation of another box C/D snoRNA, snR13, shows a matching set of processing defects in *rrp6*Δ and *rrp47*Δ strains: a short 3´ extension of the mature species, and a 5´ truncation product (Mitchell *et al*., 2003). In a *rex1*Δ/*rrp47*ΔC strain, this truncation product disappears, and the 3´ extension grows longer (Garland *et al*., 2013). When this *rex1*Δ/*rrp47*ΔC strain is combined with galactose-driven depletion of Rrp6 from the exosome, a high degree of polyadenylation can be seen with a modest decrease in the mature species (Garland *et al*., 2013). The Northern analysis attempted in Figure 5.20B suffered from very low signal during radiography, however a barely detectable reduction in the mature and truncated species of snR13 can be seen in the *rex1*H308A/*rrp47*Δ strain.

Processing of the U14 snoRNA is another case where the *rrp6*Δ and *rrp47*Δ phenotypes differ slightly (Mitchell *et al*., 2003): a U14+3 intermediate accumulates in the *rrp6*Δ strain that is fainter in the *rrp47*Δ strain, but both strains demonstrate the same degree of polyadenylation (Allmang *et al*., 1999b; Mitchell *et al*., 2003; van Hoof *et al*., 2000b), resulting from polyadenylation at either the primary termination site I-pA, or the backup termination site II-pA (Grzechnik and Kufel, 2008; Garland *et al*., 2013). Of the hypomorphic *rex1* mutants examined, only the *rex1*H308A/*rrp47*Δ strain demonstrates a more pronounced phenotype than the *REX1*/*rrp47*Δ strain with an increase in the degree of polyadenylation seen.

While snR38, snR14, and U14 are transcribed by RNAPII, U6 snRNA is transcribed by RNAPIII (Brow and Guthrie, 1990). The U6 processing defects of *rrp6*Δ consist of TRAMP-mediated polyadenylation (Wyers *et al*., 2005), with a similar phenotype seen in *rrp47*Δ (Garland *et al*., 2013). A defined truncation product is seen in hypomorphic *rrp47* strains that is lost when combined with *rex1*Δ

Table 5.1: Correlation of *rex1* mutant phenotypes between *in vitro* and *in vivo* activity
Key: '+++' Full activity, '++' Intermediate activity, '+' Minimal activity, '-' Absent activity. For Northern analysis, additional designations: '- -' Detrimental Activity, '- - -' Loss of mature species. 'n/a' refers to cases where FOA isolates could not be obtained, due to the synthetic lethality of *rex1*Δ/*rrp47*Δ

	***In vitro* Assay**	**Plasmid Shuffle**	**Northern Blot**							
			5S rRNA		**tRNA**		**snR38**	**snR13**	**U6**	**U14**
Strain:	n/a	*rex1*Δ/ *rrp47*Δ	*rex1*Δ	*rex1*Δ/ *rrp47*Δ	*rex1*Δ	*rex1*Δ/ *rrp47*Δ	*rex1*Δ/ *rrp47*Δ	*rex1*Δ/ *rrp47*Δ	*rex1*Δ/ *rrp47*Δ	*rex1*Δ/ *rrp47*Δ
no Rex1	-	-	-	n/a	+++	n/a	n/a	n/a	n/a	n/a
WT	+++	+++	+++	+++	+++	+++	+	+	-	-
Y272A	+	+	++	+++	+++	+++	+	+	-	-
H308A	-	-	+	+	- -	-	- - -	-	- -	- -
N312A	++	++	++	+++	+++	+++	++	+	+	+
K340A	+++	++	++	+++	+++	+++	+	+	-	-
S342A	+++	++	+++	+++	+++	+++	++	+	+	+
Loop1Δ	+	+	-	-	-	+	-	+	-	-
Loop2Δ	-	++	+	+	+++	+++	+	+	-	-
Loop3Δ	-	++	-	-	-	+	+	+	-	-

(Garland *et al*., 2013). When examining the U6 profiles in Figure 5.20D, it seems to be that two distinct truncation products emerge: truncation 1 is the previously observed Rex1-dependent species seen in *rrp47*Δ mutants, but truncation 2 seems to accumulate in *RRP47* strains. The U6 probe used anneals to the middle of the U6 sequence, meaning these truncation products could be truncated at either the 3´ or 5´ end. Neither truncation product seems to accumulate in the hypomorphic *rex1*/*rrp47*Δ strains, which instead show varying degrees of polyadenylation; an exception is seen for *rex1*Y272A/*rrp47*Δ, which presents the Rex1-dependent truncation 1 species. The most pronounced polyadenylation is seen in *rex1*H308A/*rrp47*Δ. The *rex1*ΔLoop2/*rrp47*Δ strain also demonstrates enhanced polyadenylation, but this again may be a consequence of uneven loading. The *rex1*N312A/*rrp47*Δ and *rex1*S342A/*rrp47*Δ seem to demonstrate a lower degree of polyadenylation than the *REX1*/*rrp47*Δ strain, which may suggest that the overexpression of wild type-like Rex1 may be able to suppress the U6 processing defect of *rrp47*Δ.

5.2.6.5 Correlation of the *in vitro* and *in vivo* Rex1 function of *rex1* mutants

The *rex1* mutants generated by this project have received three broad levels of characterisation: the *in vitro* activity of mutant protein against ssDNA oligos, the degree of complementation in the *rex1*Δ/*rrp47*Δ plasmid shuffle strain, and the complementation of RNA defects in *rex1* and *rex1*/*rrp47*Δ strains. The phenotype of each mutant is compiled in Table 5.1.

For the point mutants within the exonuclease domain, the *in vitro* processing defects broadly agree with the degree of *rex1*Δ/*rrp47*Δ complementation, and to some extent with the defects seen with and without wild type *RRP47* by Northern blotting. For 5S processing, the H308A mutant demonstrated a clear defect. A minor accumulation of the extended 5S could be observed for Y272A, N312A, and K340A that seemed equal in strength, despite the differing levels of Rex1 functionality as measured by

in vitro trimming and the plasmid shuffle assay. The point mutants didn't impact tRNA processing, with the exception of the H308A mutant. A better stratification of function was seen for snoRNA processing in the *rrp47*Δ background: the degree of polyadenylation seen correlates with the growth rate of the mutants in the plasmid shuffle assay, with the most extreme defect shown by H308A leading to the loss of mature snR38.

For the Loop mutants in the Rex1 RYS domain, there seems to be a consistent level of polyadenylation for the three mutants, when taking into account the overloading of the *rex1*ΔLoop2/*rrp47*Δ RNA evident in the 5.8S band of Figure 5.18A. There was however a degree of stratification visible in the processing of snR38: the *rex1*ΔLoop1/*rrp47*Δ mutant grew the slowest, and showed the greatest depletion for the mature and +3 snR38 species. The *rex1*ΔLoop2/*rrp47*Δ and *rex1*ΔLoop3*rrp47*Δ mutants by contrast accumulate similar levels of these two RNA species as the *REX1*/*rrp47*Δ strain, and grew at a comparably higher rate.

These data served as additional evidence supporting the proposed model that the synthetic lethality relationship between *rex1*Δ and *rrp47*Δ are a result of global snoRNA processing defects leading to RNA metabolism disruptions through saturation of polyA machinery (Garland *et al*., 2013). The worst growth rates were seen for mutants such as *rex1*H308A and *rex1*ΔLoop1 that in the case of snR38 lead to loss of the mature species, whereas the intermediate growth rates seen for *rex1*ΔLoop2, *rex1*ΔLoop3, and *rex1*Y272A demonstrated intermediate polyadenylation phenotypes throughout all snoRNAs examined. This model is further supported by observations made in a *REXO5* mutant model in *D. melanogaster* (Gerstberger *et al*., 2017), which used small RNA sequencing to reveal a global polyadenylation of snoRNAs in the absence of *REXO5*.

Given that REXO5 is likely to serve a similar role in the maturation of snoRNAs, this project's next aim was to test whether expression of human REXO5 is able to rescue the RNA processing defects seen in a yeast *rex1*Δ/*rrp47*Δ strain, or the intermediate defects seen in the hypomorphic *rex1*/*rrp47*Δ strains generated here.

5.2.7 Complementation of *rex1*Δ and hypomorphic *Rex1* mutants by REXO5

A long term aim of this project is to apply our understanding of Rex1 in yeast to a functional homologue in humans. Based on the analysis presented in Chapter 3 REXO5 is likely the closest human homologue of yeast Rex1 in terms of evolutionary distance, and REXO5 in *D. melanogaster* seems to demonstrate Rex1 function (Gerstberger *et al*., 2017). It is not necessarily the case however that REXO5 is the human functional homologue of Rex1, especially given that a mouse study examining murine and human REXO5 did not observe Rex1 activity (Silva *et al*., 2017).

In order to identify any functional homologues of Rex1 amongst the human DEDDh exonucleases identified in Chapter 3, we sought to design a yeast expression cassette with compatible restriction sites as outlined in Figure 5.21. In addition to a compatible multiple cloning site, an expression cassette was designed with the following features: a galactose-inducible promoter, an N-terminal-zz tag, the 3´ UTR of *ADH1* - a highly expressed housekeeping gene, and a site between the ORF insertion site and the N-terminal tag into which the SV40 nuclear localisation sequence and/or a PreScission protease cleavage site can be inserted as oligo linkers. Due to time constraints, the vector was cloned without a protease-cleavable site or NLS. In order to generate this vector with sites compatible with these human DEDDh ORFs, it was necessary to remove an AgeI site from the GAL promoter using the Klenow

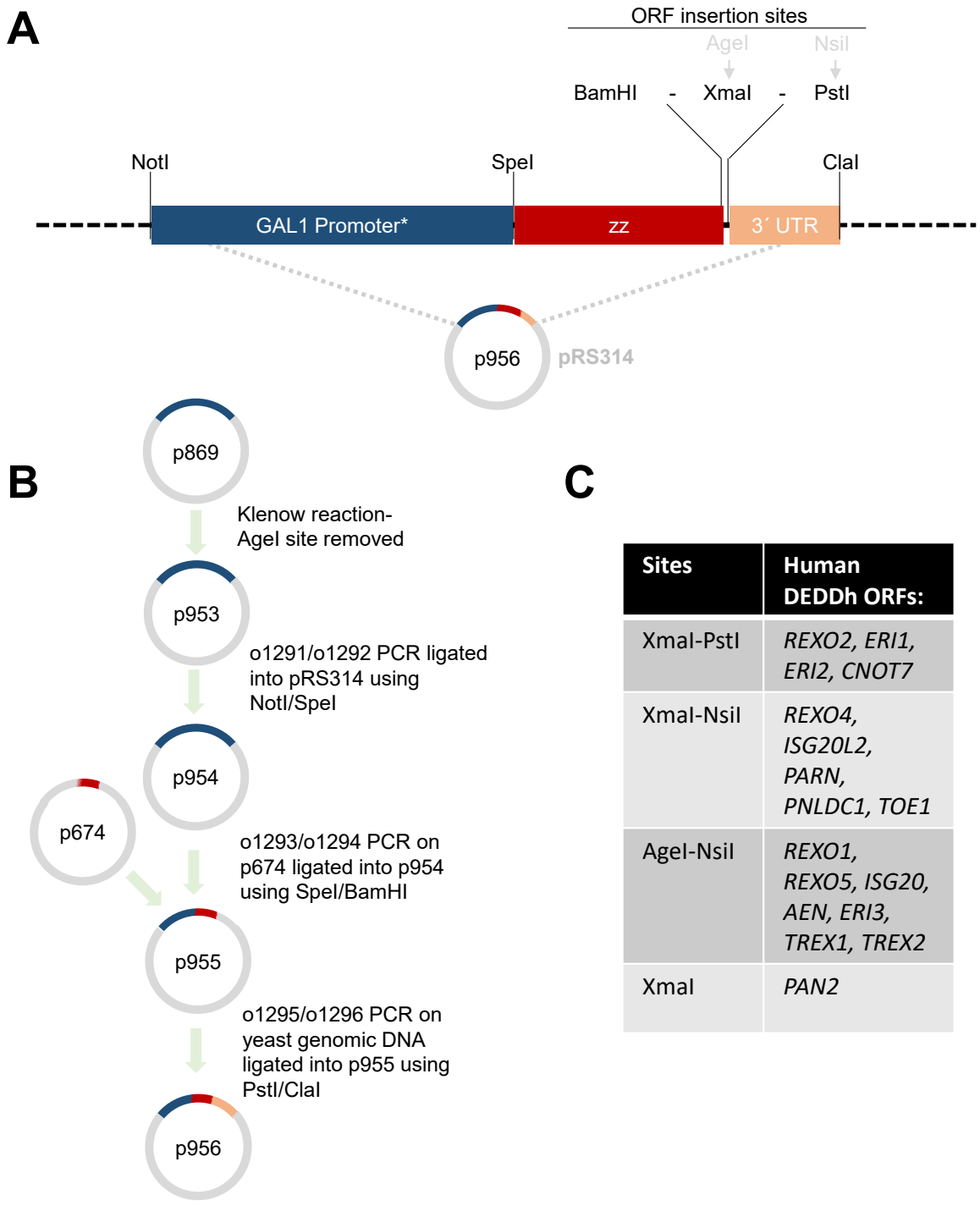


Figure 5.21: Construction of a yeast expression vector compatible with human DEDDh ORFs
A: Scale diagram of the expression cassette. Greyed out restriction sites indicate sites with compatible sticky ends to XmaI and PstI. "3´ UTR" refers to the 3´ untranslated region of Adh1. **B:** Flowchart of vector construction. **C:** List of human DEDDh ORFs, grouped by the sites available for cloning into p956.

reaction. This sequence alteration was observed to lie outside the critical sites of transcription factor binding ($CGGN_{11}CCG$; Marmorstein *et al*., 1992; Liang *et al*., 1996), and the preservation of wild type-like GAL promoter activity was confirmed using a test induction of HA-Mtr4 and Western blotting (data not shown).

As a pilot attempt at identifying a human functional homologue of Rex1, the *REXO5* open reading frame was cloned from an RNA extract of human HCT116 cells gifted by Ang Li from the laboratory of Stuart Wilson. The Sanger sequencing of the cloned ORF revealed an identical nucleotide sequence to the NCBI reference sequence NM_030941.3, which is annotated as the reference isoform. This open reading frame was cloned into the pRS314-derived yeast expression construct described in Figure 5.21.

The *REXO5* expression plasmid was transformed into *rex1*Δ yeast and inoculated into minimal media precultures containing 2% raffinose as a carbon source before addition to larger 50ml culture volumes containing either 2% glucose or galactose. A high level of expression can be induced with the addition of galactose, with undetectable expression with the addition of glucose as demonstrated by Western blot in Figure 5.22A. Unexpectedly, a galactose-induced increase in expression could also be seen in this Western blot for the pRS313::*REX1* positive control, where *REX1* expression was under the control of the RRP4 promoter. Despite high levels of inducible expression, the *REXO5* expression construct failed to rescue the 5S rRNA processing defect in the *rex1*Δ background, and failed to rescue the synthetic lethality phenotype in the *rex1*Δ/*rrrp47*Δ plasmid shuffle strain as shown in Figure 5.22B and C respectively. Based on these results, we can conclude that zz-*REXO5* expression alone is insufficient to rescue the tested *rex1* phenotypes at a detectable level.

We aimed to detect any Rex1 activity conferred by the *REXO5* expression plasmid with a higher degree of sensitivity. To this end, we examined whether *REXO5* expression could rescue the slowed growth phenotypes evident in FOA isolates taken from the plasmid shuffle assay in Figure 5.17, as shown in Figure 5.23. A spot growth assay was used in place of growth curves due to time constraints, however there were no obvious improvements in growth rates for the *rex1* mutants examined.

The failure to detect complementation of yeast *rex1*Δ or *rex1*Δ/*rrp47*Δ by human *REXO5* does not rule out Rex1 function. There are multiple reasons that cross-species complementation fail, particularly in cases where a protein requires specific localisation to mediate its function. Specifically in the case of REXO5, there is a published precedent for variable localisation behaviour in cross-species expression (Silva *et al*., 2017): human REXO5 was observed to accumulate in the nucleolus of HeLa cells, while murine REXO5 was observed to localise to the nucleoplasm. It was the original intention of this project to express human REXO5 from a construct including an SV40 NLS, which would be a logical next step for future study.

5.3 Discussion

The work presented in this chapter demonstrates that Rex1 sequence features both within and outside the exonuclease domain contribute to Rex1 function, particularly as measured by *in vitro* exonuclease activity against DNA oligos, and *in vivo* complementation of the *rex1*Δ/*rrp47*Δ synthetic lethality phenotype.

Taking together the observation that RNA wass able to crosslink *in vivo* to both the N-terminal and C-terminal halves of the RYS domain and the observation that loss of the RYS domain N-terminal Loop1

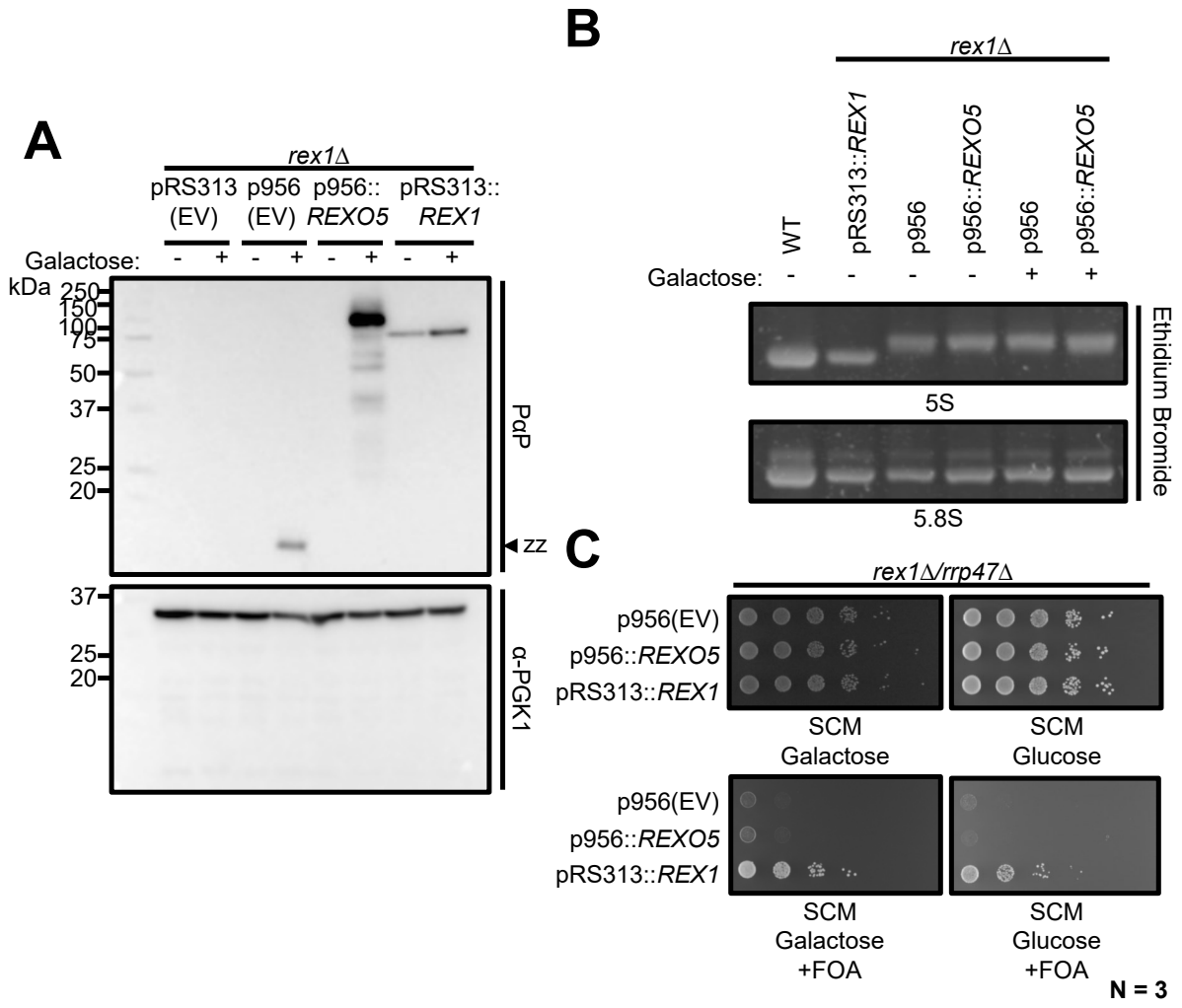


Figure 5.22: Human *REXO5* is unable to complement *rex1*Δ phenotypes
All galactose induction cultures replaced 2% glucose with 2% galactose, both cultures were inoculated with precultures using 2% raffinose. **A:** 12% SDS-acrylamide gel transferred onto nitrocellulose, probed with PαP antibody, and anti-PGK1 as a loading control. **B:** 8% Acrylamide 0.5% TBE gel stained with ethidium bromide, cropped to show 5S rRNA, and 5.8S rRNA as a loading control. **C:** Spot growth assay of *REX1* activity for human *REXO5*. Each plate representative of three technical replicates, imaged after 5 days at 30°C for glucose plates, or 7 days for galactose.

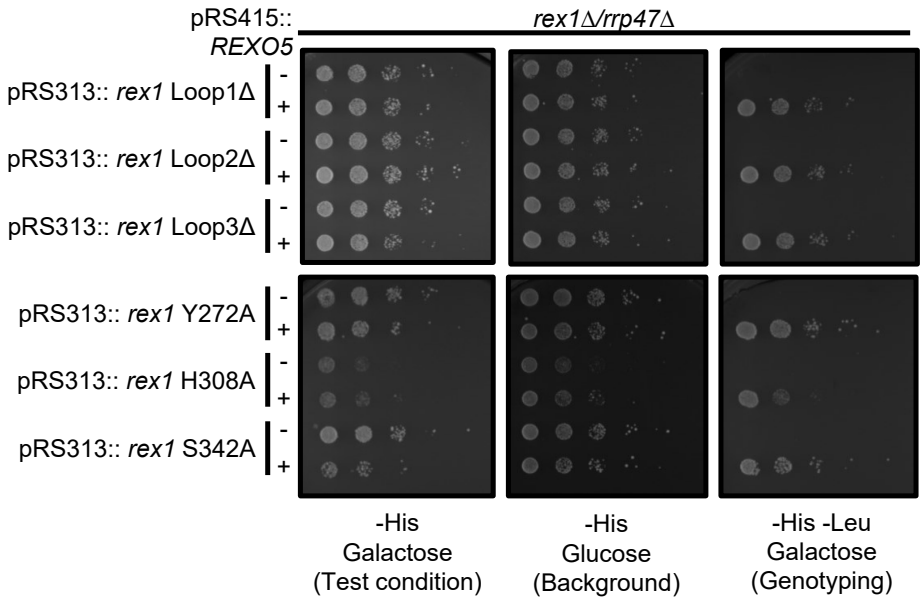


Figure 5.23: *REXO5* is unable to complement hypomorphic *rex1* mutant growth rate
Colony growth assays of *rex1* mutant FOA isolates derived from the plasmid shuffle assay shown in Figure 5.17. Each isolate was transformed with a galactose-inducible *REXO5* expression vector before seeding indicated by a '+', or seeded without transformation as indicated by a '-'. Any successful growth rate complementation would be visible as an increased growth on the galactose compared to the glucose, labelled as 'test' and 'background' respectively. A further plate lacking both His and Leu is included to indicate the presence of the pRS415::REXO5 plasmid. Each plate is representative of two technical replicates for two different FOA isolate clones.

and the C-terminal Loop 3 lead to RNA processing defects, these data suggest a substrate recognition role for the Rex1 RYS domain that is required for full Rex1 function.

The specific processing defects that are seen in the *rex1*ΔLoop1 and *rex1*ΔLoop3 mutants may give hints as to the function of the large, flexible loop structures of the RYS domain: Loop3 seems to be required for the accurate maturation of exonucleolytically processed tRNAs, whereas Loop1 seems to be more important for the production of mature snR38.

It is worth noting that the bioinformatic analysis presented in Chapter 3 suggests that Loop 2 and Loop 3 are likely to be missing in the REXO5 of *D. melanogaster*, an organism where loss of *REXO5* leads to snoRNA processing defects and lethality (Gerstberger *et al*., 2017), but a small structure that resembles Loop 3 of the '3 Loop' RYS domain arrangement may be preserved. In light of the fact that REXO5 bearing the 'AP-like' RYS domain configuration is still able to mediate essential functions in *D. melanogaster* (Gerstberger *et al*., 2017), it is likely that there are essential substrate binding mechanisms of the RYS domain beyond these three Loop structures.

With the exception of H308, the individual residues examined in the exonuclease domain seem to make minimal impact to the processing of 5S rRNA and tRNA. It was expected that each of these mutants may make minor individual contributions to the binding of substrates in the active site, which may become more pronounced when jointly lost. The combined N312A/K340A/S342A mutant seems to confirm this by *in vitro* processing of ssDNA oligos, but a yeast expression construct for this mutant would allow the physiological relevance of this distributive binding activity to be seen.

The H308A mutant behaved unexpectedly *in vivo*: a lower steady state expression level suggests that this Rex1 mutant may be highly unstable. This may suggest that H308 plays an important role for coordinating correct folding of the protein. The relative expression of Rex1H308A seems to be much higher in *E. coli* than in *S. cerevisiae*, which may be due to the deletion of lon and ompT proteases in the BL21 *E. coli* strain that would otherwise degrade misfolded protein (Studier and Moffatt, 1986; Daegelen *et al*., 2009). If sufficient quantities of Rex1H308A can be purified, it would be simple to measure the extent of misfolding using biophysical methods such as differential scanning fluorimetry (Senisterra and Finerty, 2009). It remains a possibility that H308 may also serve a substrate-binding role on top of a fold-stabilising role, however this may require the solution of a substrate-bound Rex1 structure to serve as evidence rather than the generation of a H308 missense mutation.

An important future experiment for the Rex1 mutants generated by this project will be examining their *in vitro* ability to turn over the extended 5S RNP that can be isolated from the mature ribosomes of *rex1*Δ yeast. Two categories of 5S misprocessing were observed *in vivo* for the most hypomorphic *rex1* mutants: the *rex1*ΔLoop1 and *rex1*ΔLoop3 mutants demonstrated a 5S pool that was shorter than the *rex1*Δ phenotype, but not as short as the matured 5S seen in *REX1* strains; *rex1*H308A by contrast demonstrated an even split between fully processed, and fully unprocessed 5S. Based on these results it may be that the Loop 1 and Loop 3 features of the RYS domain enable Rex1 to trim the final precursor nucleotides in a distributive manner, or that Loop 1 and Loop 3 enable highly processive trimming of the trailer, that become inaccessible to Rex1 upon premature dissociation. The wild type Rex1 activity that is observed against the extended 5S RNP *in vitro* did not detect any processing intermediates between the 3 nucleotide extension and the mature product (Hama Soor, 2017; Daniels *et al*., 2022); by applying the hypomorphic Rex1 mutants presented here in this assay, it may be possible to determine the degree and cause of processivity in 5S processing by Rex1.

It is interesting to note that there may be a narrow window for Rex1 to be able to mature 5S rRNA *in vivo*: previous work in the Mitchell lab (Hama Soor, 2017) revealed that recombinant Rex1 is unable to process the 3´ extensions of the 5S rRNA when part of the mature 60S subunit, despite solvent accessibility of the 5S rRNA 3´ end as visible in structural studies (PDB entry 5GAK; Schmidt *et al*., 2016). Some reports suggest that the integration of the 5S RNP into the ribosome takes place at an early stage in yeast, specifically into the 90S precursor (Zhang *et al*., 2007; reviewed in Ciganda and Williams, 2011). If this is the case, then the degree of 5S misprocessing seen in hypomorphic *rex1* mutants may be acutely sensitive to the rate of ribosome biogenesis; it will be interesting to see how consistent the intensities of these defects are in future Northern blots of replicate yeast cultures.

Looking beyond 5S processing, it would be informative to see the degree of processing by recombinant Rex1 mutants for other model substrates. A set of T7 transcribed tRNA substrates have been designed by Ozanick *et al*. (2009), which would be of particular interest for expanding on the tRNA processing defects evident in the *rex1*ΔLoop3 mutant. Other potential *in vitro* substrates include model snoRNAs, although these would require the association of proteins to recreate the physiological snoRNP.

Chapter 6

Conclusions and future work

Taken as a whole, this work represents an initial characterisation of the Rex1 RYS domain. These data demonstrate an *in vivo* RNA binding role for both the N- and C-terminal RYS domain components using *in vivo* RNA crosslinking, and observe defects in the maturation of specific RNA species for *rex1* mutants lacking conserved RYS domain features: the Loop 1 region (146-154) seems to be required for the efficient processing of the box C/D snoRNA snR38, and the Loop 3 region (506-544) seems to be required for efficient exonucleolytic processing of tRNA. The bioinformatic analysis presented here indicates that the RYS domain is preserved in the sequence orthologues of Rex1 found in all eukaryotes, and analysis of models compiled in the Alphafold Database (Jumper *et al*., 2021; Varadi *et al*., 2022) has identified three distinct configurations: an 'Alkaline Phosphatase-like' configuration found in invertebrates that consists of few features beyond the core composite β-sheet; a '2-RRM' configuration found in vertebrates that features a Puf60-releated RNA recognition motif pair fused to the core; and a '3-loop' configuration found in all other eukaryotes including plants and fungi, which consists of the core domain adorned with three flexible loops, including a large coiled-coil motif termed 'Loop 3'.

This project aimed to determine a structural model of Rex1 using X-ray crystallography, however attempts to crystallise Rex1 in the presence of AMP and $MgCl_2$ failed to yield crystals of sufficient quality. It may be that the large Loop structures predicted by the Alphafold2 model for Rex1 (Jumper *et al*., 2021; Varadi *et al*., 2022) are sufficiently flexible to disrupt crystal packing. There are multiple potential solutions to this problem: the flexibility of these structures may be reduced through co-crystallisation with a well-structured substrate such as tRNA; the loop structures could be truncated, indeed this project has already generated a *rex1* mutant lacking all three Loop structures; or limited proteolysis could be attempted as an unbiased approach to removing flexible regions, although if the Alphafold2 model (Jumper *et al*., 2021; Varadi *et al*., 2022) is accurate this may result in the proteolytic separation of the exonuclease domain from the RYS domain. An alternative may be to use orthologues of Rex1 that are predicted to have a more compact fold such as those found in *D. melanogaster* and *C. elegans*, which posses the 'AP-like' RYS domain configuration that lacks the disordered regions seen in the '3-loop' and '2-RRM' RYS domain configurations.

While the Loop structures are visually striking, the existence of RYS domains lacking these structures in organisms where REXO5 is essential to function (Gerstberger *et al*., 2017) suggest a fundamental substrate recognition mechanism residing within the RYS domain core. In order to more thoroughly characterise the RYS domain, it may be worthwhile to derive point mutations of conserved residues found within the RYS core fold, particularly those conserved in *D. melanogaster* REXO5.

The exonuclease domain point mutants were designed based on a Phyre2-based (Kelley *et al*., 2015) model of Rex1 threaded onto Pan2 in complex with an RNA substrate (Tang *et al*., 2019). There is a degree of uncertainty in how Phyre2 calculates the orientation of the side chains described in its models (Kelley *et al*., 2015), which is less uncertain in Alphafold2 models (Jumper *et al*., 2021). In light of this increased side chain confidence inspection of the binding-cleft exposed residues reveals additional residues in the exonuclease domain that may contribute to substrate binding: L239, L269, H333, F339, K344, Q356, and K344. In addition to mediating generic polar interactions with the phosphodiester backbone and stacking with base groups, it may be the case that residues in the active site may contribute to substrate preferences such as the 'C-effect' that is seen both in Rex1 (see Figure 4.7), Rrp6 (Axhemi *et al*., 2020), and RNase T (Zuo and Deutscher, 2002b).

It is difficult to overstate the impact the release of the Alphafold Database had on this project; the high-confidence prediction generated for Rex1 has enabled prediction of Rex1 sequence features that would otherwise have remained unseen in the absence of a structural model, and enabled the identification of a substrate recognition domain, the RYS domain, which is conserved throughout eukaryotes spanning plants to humans. There is limited information however that a static prediction can provide, a full characterisation of Rex1's substrate recognition mechanisms will require the investigation of current and additional mutants through *in vitro* and *in vivo* measurements of Rex1 activity. The best evidence towards a mechanistic model of Rex1 activity would be an experimentally-derived structure of Rex1 in complex with both its substrates and divalent cation cofactors.

It may one day be the case that a detailed mechanistic understanding of Rex1 is achieved. Knowledge of Rex1's substrate recognition mechanisms alone is useful in understanding which substrates it targets, however a true understanding of Rex1 function in the cell will require an integrated knowledge of how Rex1 interacts and overlaps with other mediators of RNA metabolism. A primary competitor with Rex1 at the 3´ end of its substrates is Lhp1, which serves to protect substrates from over-trimming before they have had the chance to fold correctly (Copela *et al*., 2008). A more complete understanding of Rex1's activity *in vivo* will require a knowledge of competing activity at the 3´ by RNA binding proteins such as Lhp1 and the action of other nucleases such as the exosome complex. It will be important however not to ignore the impact of the RNA 5´ on Rex1 activity, as trends in the substrates that are best processed by Rex1 seem to suggest a requirement for a base-paired 5´ terminus for efficient processing of stemloop overhangs as seen in RNase T (Li *et al*., 1998, 1999). It remains a remote possibility that Rex1 may be able to process these equivalent stemloop structures in dsDNA, or even in DNA:RNA hybrids as detected in the initial studies of apparent Rex1 RNase H-like activity (Karwan *et al*., 1983; Frank *et al*., 1999).

A long term aim of this project is to apply the mechanistic knowledge of Rex1 in *S. cerevisiae* to its sequence orthologue in humans, REXO5. REXO5 represents a fascinating subject for study in humans, as REXO5 expressed seems to be heavily overexpressed in the testes (Silva *et al*., 2017). It remains to be seen however to what extent REXO5 is active in the rest of the human body owing to low levels of mRNA expression, and to what extent it shares its activity with the human sequence orthologue of RNase T TREX1, which has recently been discovered to harbour cryptic RNA exonuclease activity capable of mediating RNase T-like substrate processing against model substrates *in vitro* (Yuan *et al*., 2015). Future study of REXO5 in human cells will need to first establish the expression status of REXO5 at the protein level by Western blotting or mass spectrometry, then contend with the potential overlapping

exoribonuclease activity of TREX1 through selective depletion or deletion of the alternative exon that confers RNA binding. One potentially interesting avenue for investigation would be to reintroduce *REXO5* expression into the Jurkat cell line, a T-cell acute lymphoblastic leukaemia cell line that is unique of the common cell lines in demonstrating no expression of *REXO5* mRNA (see Figure 1.7B), potentially representing an oncogenic adaptation.

The dysregulation of RNA metabolism underpins a range of human diseases. Disease-linked mutations have been observed in the exosome complex specifically (for review see; Fasken *et al*., 2020), which can lead to developmental defects such as pontocerebellar hypoplasia (Wan *et al*., 2012), with mutations in the human Rrp44 homologue DIS3 observed in an estimated 11% of multiple myeloma cases (Fasken *et al*., 2020). The mutations observed in these diseases result in a hypomorphic exosome complex; a mediator of RNA metabolism that is essential in humans to the point where its surprising that cells are viable. REXO5 is one of several exonucleases that may be able to mediate backup processing to compensate for this hypomorphic exosome activity. It is unlikely that this complementation would be based on snoRNA processing as seen in yeast (Garland *et al*., 2013), but could involve the 3´ maturation of ribosomes as observed in *D. melanogaster* (Gerstberger *et al*., 2017), or perhaps a role in the regulation of pervasive transcripts. The relevance of REXO5 to exosome-related disease just one hypothetical real-world application of research into the mechanisms of Rex1 in *S. cerevisiae*. As with all blue sky research into fundamental biology, the realistic impact of this research is hard to predict. Given the importance of RNase T-like activity to the RNA metabolism of most model organisms, understanding how these equivalent events take place in humans will at the very least give us a fuller understanding of how stable RNAs are accurately produced in humans.

Bibliography

Abou Elela, S., Igel, H. and Ares, M. (1996), 'RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site', *Cell* **85**(1), 115–124.

Acker, J., Conesa, C. and Lefebvre, O. (2013), 'Yeast RNA polymerase III transcription factors and effectors', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**(3-4), 283–295.

Agrawal, A., Mohanty, B. K. and Kushner, S. R. (2014), 'Processing of the seven valine tRNAs in *Escherichia coli* involves novel features of RNase P', *Nucleic Acids Research* **42**(17), 11166–11179.

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S. L., Hughes, T. R., Grayhack, E. J. and Phizicky, E. M. (2006), 'Rapid tRNA Decay Can Result from Lack of Nonessential Modifications', *Molecular Cell* **21**(1), 87–96.

Allison, D. S. and Hall, B. D. (1985), 'Effects of alterations in the 3′ flanking sequence on *in vivo* and *in vitro* expression of the yeast SUP4-o tRNATyr gene.', *The EMBO Journal* **4**(10), 2657–2664.

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E. and Tollervey, D. (1999b), 'Functions of the exosome in rRNA, snoRNA and snRNA synthesis.', *The EMBO Journal* **18**(19), 5399–5410.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D. and Mitchell, P. (1999a), 'The yeast exosome and human PM-Scl are related complexes of 3′ -5′ exonucleases', *Genes Development* **13**(16), 2148–2158.

Alting-Mees, M. A. and Short, J. M. (1989), 'pBluescript II: gene mapping vectors', *Nucleic acids research* **17**(22), 9494.

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977), 'Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes.', *Proceedings of the National Academy of Sciences of the United States of America* **74**(12), 5350–5354.

Ansel, K. M., Pastor, W. A., Rath, N., Lapan, A. D., Glasmacher, E., Wolf, C., Smith, L. C., Papadopoulou, N., Lamperti, E. D., Tahiliani, M., Ellwart, J. W., Shi, Y., Kremmer, E., Rao, A. and Heissmeyer, V. (2008), 'Mouse Eri1 interacts with the ribosome and catalyzes 5.8S rRNA processing', *Nature Structural Molecular Biology* **15**(5), 523–530.

Aravind, L. and Koonin, E. V. (2000), 'SAP - A putative DNA-binding motif involved in chromosomal organization', *Trends in Biochemical Sciences* **25**(3), 112–114.

Arimbasseri, A. G., Rijal, K. and Maraia, R. J. (2013), 'Transcription termination by the eukaryotic RNA polymerase III', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**(3-4), 318–330.

Axhemi, A., Wasmuth, E. V., Lima, C. D. and Jankowsky, E. (2020), 'Substrate selectivity by the exonuclease Rrp6p', *Proceedings of the National Academy of Sciences of the United States of America* **117**(2), 982–992.

Badhwar, P., Khan, S. H. and Taneja, B. (2022), 'Three-dimensional structure of a mycobacterial oligoribonuclease reveals a unique C-terminal tail that stabilizes the homodimer', *The Journal of Biological Chemistry* **298**(12), 102595.

Barany, F. (1988), 'The TaqI "star" reaction: strand preferences reveal hydrogen-bond donor and acceptor sites in canonical sequence recognition', *Gene* **65**(2), 149–165.

Bechhofer, D. H. and Deutscher, M. P. (2019), 'Bacterial ribonucleases and their roles in RNA metabolism', *Critical Reviews in Biochemistry and Molecular Biology* **54**(3), 242–300.

Beese, L. S. and Steitz, T. A. (1991), 'Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism.', *The EMBO Journal* **10**(1), 25–33.

Bernad, A., Blanco, L., Lázaro, J. M., Martín, G. and Salas, M. (1989), 'A conserved 3′-5′ exonuclease active site in prokaryotic and eukaryotic DNA polymerases', *Cell* **59**(1), 219–228.

Bhagwat, M., Hobbs, L. J. and Nossal, N. G. (1997), 'The 5′-exonuclease activity of bacteriophage T4 RNase H is stimulated by the T4 gene 32 single-stranded DNA-binding protein, but its flap endonuclease is inhibited', *Journal of Biological Chemistry* **272**(45), 28523–28530.

Bimboim, H. C. and Doly, J. (1979), 'A rapid alkaline extraction procedure for screening recombinant plasmid DNA.', *Nucleic Acids Research* **7**(6), 1513–1523.

Blaszczyk, J., Gan, J., Tropea, J. E., Court, D. L., Waugh, D. S. and Ji, X. (2004), 'Noncatalytic assembly of ribonuclease III with double-stranded RNA', *Structure* **12**(3), 457–466.

Blobel, G. (1971), 'Isolation of a 5S RNA-protein complex from mammalian ribosomes', *Proceedings of the National Academy of Sciences of the United States of America* **68**(8), 1881–1885.

Bogenhagen, D. F. and Brown, D. D. (1981), 'Nucleotide sequences in Xenopus 5S DNA required for transcription termination', *Cell* **24**(1), 261–270.

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J. and Madden, T. L. (2012), 'Domain enhanced lookup time accelerated BLAST', *Biology Direct* **7**(1), 1–14.

Braglia, P., Percudani, R. and Dieci, G. (2005), 'Sequence Context Effects on Oligo(dT) Termination Signal Recognition by Saccharomyces cerevisiae RNA Polymerase III', *Journal of Biological Chemistry* **280**(20), 19551–19562.

Bresson, S. and Tollervey, D. (2018), 'Surveillance-ready transcription: nuclear RNA decay as a default fate', *Open Biology* **8**(3), 170270.

Briggs, M. W., Burkard, K. T. and Butler, J. S. (1998), 'Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3′ end formation', *Journal of Biological Chemistry* **273**(21), 13255–13263.

Brow, D. A. and Guthrie, C. (1990), 'Transcription of a yeast U6 snRNA gene requires a polymerase III promoter element in a novel position', *Genes Development* **4**(8), 1345–1356.

Burkard, K. T. D. and Butler, J. S. (2000), 'A Nuclear 3′-5′ Exonuclease Involved in mRNA Degradation Interacts with Poly(A) Polymerase and the hnRNA Protein Npl3p', *Molecular and Cellular Biology* **20**(2), 604–616.

Burnette, W. N. (1981), '"Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A', *Analytical Biochemistry* **112**(2), 195–203.

Callahan, K. P. and Butler, J. S. (2010), 'TRAMP Complex Enhances RNA Degradation by the Nuclear Exosome Component Rrp6', *The Journal of Biological Chemistry* **285**(6), 3540–3547.

Cannavo, E. and Cejka, P. (2014), 'Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks', *Nature* **514**(7520), 122–125.

Castle, C. D., Cassimere, E. K., Lee, J. and Denicourt, C. (2010), 'Las1L Is a Nucleolar Protein Required for Cell Proliferation and Ribosome Biogenesis', *Molecular and Cellular Biology* **30**(18), 4404–4414.

Castle, C. D., Sardana, R., Dandekar, V., Borgianini, V., Johnson, A. W. and Denicourt, C. (2013), 'Las1 interacts with Grc3 polynucleotide kinase and is required for ribosome synthesis in *Saccharomyces cerevisiae*', *Nucleic Acids Research* **41**(2), 1135–1150.

Chanfreau, G., Elela, S. A., Ares, M. and Guthrie, C. (1997), 'Alternative 3′-end processing of U5 snRNA by RNase III', *Genes Development* **11**(20), 2741–2751.

Chédin, S., Riva, M., Schultz, P., Sentenac, A. and Caries, C. (1998), 'The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination', *Genes Development* **12**(24), 3857–3871.

Chen, J., Liu, L., You, C., Gu, J., Ruan, W., Zhang, L., Gan, J., Cao, C., Huang, Y., Chen, X. and Ma, J. (2018), 'Structural and biochemical insights into small RNA 3′ end trimming by *Arabidopsis* SDN1', *Nature Communications* **9**(1), 3585.

Chou, D. K., Krishnamurthy, R., Randolph, T. W., Carpenter, J. F. and Manning, M. C. (2005), 'Effects of Tween 20® and Tween 80® on the Stability of Albutropin During Agitation', *Journal of Pharmaceutical Sciences* **94**(6), 1368–1381.

Chou, K. M. and Cheng, Y. C. (2002), 'An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3′ mispaired DNA', *Nature* **415**(6872), 655–659.

Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. and Hieter, P. (1992), 'Multifunctional yeast high-copy-number shuttle vectors', *Gene* **110**(1), 119–122.

Chu, L. Y., Agrawal, S., Chen, Y. P., Yang, W. Z. and Yuan, H. S. (2019), 'Structural insights into nanoRNA degradation by human Rexo2', *RNA* **25**(6), 737–746.

Ciganda, M. and Williams, N. (2011), 'Eukaryotic 5S rRNA biogenesis', *Wiley Interdisciplinary Reviews: RNA* **2**(4), 523–533.

Connelly, S. and Manley, J. L. (1988), 'A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II.', *Genes Development* **2**(4), 440–452.

Copela, L. A., Fernandez, C. F., Sherrer, R. L. and Wolin, S. L. (2008), 'Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation', *RNA* **14**(6), 1214–1227.

Costello, J. L., Stead, J. A., Feigenbutz, M., Jones, R. M. and Mitchell, P. (2011), 'The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3′-maturation', *Journal of Biological Chemistry* **286**(6), 4535–4543.

Czech, A., Wende, S., Mörl, M., Pan, T. and Ignatova, Z. (2013), 'Reversible and Rapid Transfer-RNA Deactivation as a Mechanism of Translational Repression in Stress', *PLoS Genetics* **9**(8), 1003767.

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. and Kim, J. F. (2009), 'Tracing Ancestors and Relatives of *Escherichia coli* B, and the Derivation of B Strains REL606 and BL21(DE3)', *Journal of Molecular Biology* **394**(4), 634–643.

Daniels, P., Hama Soor, T., Levicky, Q., Hettema, E. H. and Mitchell, P. (2022), 'Contribution of domain structure to the function of the yeast DEDD family exoribonuclease and RNase T functional homologue, Rex1', *RNA* **28**(4), 493–507.

Das, G., Henning, D., Wright, D. and Reddy, R. (1988), 'Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III', *The EMBO journal* **7**(2), 503–512.

Das, M., Zattas, D., Zinder, J. C., Wasmuth, E. V., Henri, J. and Lima, C. D. (2021), 'Substrate discrimination and quality control require each catalytic activity of TRAMP and the nuclear RNA exosome', *Proceedings of the National Academy of Sciences of the United States of America* **118**(14), e2024846118.

Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. and Altman, R. B. (2005), 'SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments', *RNA* **11**(3), 344–354.

Daugeron, M. C., Mauxion, F. and Séraphin, B. (2001), 'The yeast *POP2* gene encodes a nuclease involved in mRNA deadenylation', *Nucleic Acids Research* **29**(12), 2448–2455.

de Boer, H. A., Comstock, L. J. and Vasser, M. (1983), 'The tac promoter: a functional hybrid derived from the trp and lac promoters.', *Proceedings of the National Academy of Sciences of the United States of America* **80**(1), 21–25.

Delan-Forino, C., Schneider, C. and Tollervey, D. (2017), 'Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex', *PLoS Genetics* **13**(3), e1006699.

Deutscher, M. P. and Marlor, C. W. (1985), 'Purification and characterization of *Escherichia coli* RNase T', *Journal of Biological Chemistry* **260**(11), 7067–7071.

Deutscher, M. P., Marlor, C. W. and Zaniewski, R. (1984), 'Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA.', *Proceedings of the National Academy of Sciences of the United States of America* **81**(14), 4290–4293.

Dichtl, B., Stevens, A. and Tollervey, D. (1997), 'Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes.', *The EMBO Journal* **16**(23), 7184–7195.

Dieci, G., Conti, A., Pagano, A. and Carnevali, D. (2013), 'Identification of RNA polymerase III-transcribed genes in eukaryotic genomes', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**(3-4), 296–305.

Dominski, Z., Yang, X. C., Kaygun, H., Dadlez, M. and Marzluff, W. F. (2003), 'A 3′ exonuclease that specifically interacts with the 3′ end of histone mRNA', *Molecular Cell* **12**(2), 295–305.

Donze, D., Adams, C. R., Rine, J. and Kamakaka, R. T. (1999), 'The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*', *Genes Development* **13**(6), 698–708.

Douet, J. and Tourmente, S. (2007), 'Transcription of the 5S rRNA heterochromatic genes is epigenetically controlled in *Arabidopsis thaliana* and *Xenopus laevis*', *Heredity* **99**(1), 5–13.

Dupasquier, M., Kim, S., Halkidis, K., Gamper, H. and Hou, Y. M. (2008), 'tRNA Integrity Is a Prerequisite for Rapid CCA Addition: Implication for Quality Control', *Journal of Molecular Biology* **379**(3), 579–588.

Dziembowski, A., Lorentzen, E., Conti, E. and Séraphin, B. (2006), 'A single subunit, Dis3, is essentially responsible for yeast exosome core activity', *Nature Structural Molecular Biology* **14**(1), 15–22.

El Hage, A., Koper, M., Kufel, J. and Tollervey, D. (2008), 'Efficient termination of transcription by RNA polymerase I requires the 5′ exonuclease Rat1 in yeast', *Genes Development* **22**(8), 1069–1081.

Engelke, D. R., Gegenheimer, P. and Abelson, J. (1985), 'Nucleolytic processing of a tRNAArg-tRNAAsp dimeric precursor by a homologous component from *Saccharomyces cerevisiae*.', *Journal of Biological Chemistry* **260**(2), 1271–1279.

Faber, A. W., Van Dijk, M., Raué, H. A. and Vos, J. C. (2002), 'Ngl2p is a Ccr4p-like RNA nuclease essential for the final step in 3′-end processing of 5.8S rRNA in *Saccharomyces cerevisiae*.', *RNA* **8**(9), 1095â1101.

Fasken, M. B., Morton, D. J., Kuiper, E. G., Jones, S. K., Leung, S. W. and Corbett, A. H. (2020), 'The RNA Exosome and Human Disease', *Methods in Molecular Biology* **2062**, 3–33.

Fatica, A., Oeffinger, M., Dlakić, M. and Tollervey, D. (2003), 'Nob1p Is Required for Cleavage of the 3′ End of 18S rRNA', *Molecular and Cellular Biology* **23**(5), 1798–1807.

Fatica, A., Tollervey, D. and Dlakić, M. (2004), 'PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA', *RNA* **10**(11), 1698–1701.

Feigenbutz, M., Garland, W., Turner, M. and Mitchell, P. (2013a), 'The Exosome Cofactor Rrp47 Is Critical for the Stability and Normal Expression of Its Associated Exoribonuclease Rrp6 in *Saccharomyces cerevisiae*', *PLoS ONE* **8**(11), e80752.

Feigenbutz, M., Jones, R., Besong, T. M., Harding, S. E. and Mitchell, P. (2013b), 'Assembly of the Yeast Exoribonuclease Rrp6 with Its Associated Cofactor Rrp47 Occurs in the Nucleus and Is Critical for the Controlled Expression of Rrp47', *The Journal of Biological Chemistry* **288**(22), 15959–15970.

Felici, F., Cesareni, G. and Hughes, J. M. X. (1989), 'The most abundant small cytoplasmic RNA of *Saccharomyces cerevisiae* has an important function required for normal cell growth', *Molecular and Cellular Biology* **9**(8), 3260–3268.

Ferat, J. L. and Michel, F. (1993), 'Group II self-splicing introns in bacteria', *Nature* **364**(6435), 358–361.

Feunteun, J., Jordan, B. R. and Monier, R. (1972), 'Study of the maturation of 5 s RNA precursors in *Escherichia coli*', *Journal of Molecular Biology* **70**(3), 465–474.

Foretek, D., Wu, J., Hopper, A. K. and Boguta, M. (2016), 'Control of *Saccharomyces cerevisiae* pre-tRNA processing by environmental conditions', *RNA* **22**(3), 339–349.

Frank, P., Braunshofer-Reiter, C., Karwan, A., Grimm, R. and Wintersberger, U. (1999), 'Purification of Saccharomyces cerevisiae RNase H(70) and identification of the corresponding gene', *FEBS Letters* **450**(3), 251–256.

Frank, P., Cazenave, C., Albert, S. and Toulmé, J. J. (1993), 'Sensitive Detection of Low Levels of Ribonuclease H Activity by an Improved Renaturation Gel Assay', *Biochemical and Biophysical Research Communications* **196**(3), 1552–1557.

Gabel, H. W. and Ruvkun, G. (2008), 'The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi', *Nature Structural Molecular Biology* **15**(5), 531–533.

Galperin, M. Y. and Koonin, E. V. (2012), 'Divergence and convergence in enzyme evolution', *Journal of Biological Chemistry* **287**(1), 21–28.

Galperin, M. Y., Koonin, E. V. and Bairoch, A. (1998), 'A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases', *Protein Science* **7**(8), 1829–1835.

Garland, W., Feigenbutz, M., Turner, M. and Mitchell, P. (2013), 'Rrp47 functions in RNA surveillance and stable RNA processing when divorced from the exoribonuclease and exosome-binding domains of Rrp6', *RNA* **19**(12), 1659–1668.

Ge, Q., Frank, M. B., O'Brien, C. and Targoff, I. N. (1992), 'Cloning of a complementary DNA coding for the 100-kD antigenic protein of the PM-Scl autoantigen.', *Journal of Clinical Investigation* **90**(2), 559–570.

Geerlings, T. H., Vos, J. C. and Raue, H. A. (2000), 'The final step in the formation of 25S rRNA in *Saccharomyces cerevisiae* is performed by 5′–>3′ exonucleases.', *RNA* **6**(12), 1698–1703.

Gerlach, P., Garland, W., Lingaraju, M., Salerno-Kochan, A., Bonneau, F., Basquin, J., Jensen, T. H. and Conti, E. (2022), 'Structure and regulation of the nuclear exosome targeting complex guides RNA substrates to the exosome', *Molecular Cell* **82**(13), 2505–2518.

Gerstberger, S., Meyer, C., Benjamin-Hong, S., Rodriguez, J., Briskin, D., Bognanni, C., Bogardus, K., Steller, H. and Tuschl, T. (2017), 'The Conserved RNA Exonuclease Rexo5 Is Required for 3′ End Maturation of 28S rRNA, 5S rRNA, and snoRNAs', *Cell Reports* **21**(3), 758–772.

Ghodge, S. V. and Raushel, F. M. (2015), 'Discovery of a previously unrecognized ribonuclease from *Escherichia coli* that hydrolyzes 5′-phosphorylated fragments of RNA', *Biochemistry* **54**(18), 2911–2918.

Ghosh, R. K. and Deutscher, M. P. (1978), 'Identification of an *Escherichia coli* nuclease acting on structurally altered transfer RNA molecules', *Journal of Biological Chemistry* **253**(4), 997–1000.

Ghosh, S. and Deutscher, M. P. (1999), 'Oligoribonuclease is an essential component of the mRNA decay pathway', *Proceedings of the National Academy of Sciences of the United States of America* **96**(8), 4372–4377.

Gietz, D., Jean, A. S., Woods, R. A. and Schiestl, R. H. (1992), 'Improved method for high efficiency transformation of intact yeast cells', *Nucleic Acids Research* **20**(6), 1425.

Goodenbour, J. M. and Pan, T. (2006), 'Diversity of tRNA genes in eukaryotes', *Nucleic Acids Research* **34**(21), 6137–6146.

Grzechnik, P. and Kufel, J. (2008), 'Polyadenylation Linked to Transcription Termination Directs the Processing of snoRNA Precursors in Yeast', *Molecular Cell* **32**(2-3), 247–258.

Gudipati, R. K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L. M., Jacquier, A. and Libri, D. (2012), 'Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells', *Molecular Cell* **48**(3), 409–421.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983), 'The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme', *Cell* **35**(3), 849–857.

Hama Soor, T. (2017), Mutational and Biochemical Analysis of the Yeast Ribonuclease Rex1, PhD thesis, The University of Sheffield.

Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L. and Dixon, N. E. (2002), 'Structural basis for proofreading during replication of the *Escherichia coli* chromosome', *Structure* **10**(4), 535–546.

Han, J. and van Hoof, A. (2016), 'The RNA exosome channeling and direct access conformations have distinct in vivo functions', *Cell Reports* **16**(12), 3348–3358.

Han, Z., Moore, G. A., Mitter, R., Dirac, A. B., Rueda, D. S., Svejstrup Correspondence, J. Q., Lopez Martinez, D., Wan, L., Barbara, A., Svejstrup, D. and Svejstrup, J. Q. (2023), 'DNA-directed termination of RNA polymerase II transcription', *Molecular Cell* **83**(18), Epub ahead of print.

Hanekamp, T. and Thorsness, P. E. (1999), '*YNT20*, a bypass suppressor of *yme1 yme2*, encodes a putative 3′-5′ exonuclease localized in mitochondria of *Saccharomyces cerevisiae*', *Current Genetics* **34**(6), 438–448.

Hayashi, S., Mori, S., Suzuki, T., Suzuki, T. and Yoshihisa, T. (2019), 'Impact of intron removal from tRNA genes on *Saccharomyces cerevisiae*', *Nucleic Acids Research* **47**(11), 5936–5949.

Hayne, C. K., Sekulovski, S., Hurtig, J. E., Stanley, R. E., Trowitzsch, S. and van Hoof, A. (2023), 'New Insights into RNA processing by the eukaryotic tRNA Splicing Endonuclease', *Journal of Biological Chemistry* **299**(9), 105138.

Hoefig, K. P., Rath, N., Heinz, G. A., Wolf, C., Dameris, J., Schepers, A., Kremmer, E., Ansel, K. M. and Heissmeyer, V. (2012), 'Eri1 degrades the stem-loop of oligouridylated histone mRNAs to induce replication-dependent decay', *Nature Structural Molecular Biology* **20**(1), 73–81.

Holm, L., Laiho, A., Törönen, P. and Salgado, M. (2023), 'DALI shines a light on remote homologs: One hundred discoveries', *Protein Science* **32**(1), e4519.

Hopper, A. K. (2013), 'Transfer RNA Post-Transcriptional Processing, Turnover, and Subcellular Dynamics in the Yeast *Saccharomyces cerevisiae*', *Genetics* **194**(1), 43–67.

Hopper, A. K., Banks, F. and Evangelidis, V. (1978), 'A yeast mutant which accumulates precursor tRNAs', *Cell* **14**(2), 211–219.

Horio, T., Murai, M., Inoue, T., Hamasaki, T., Tanaka, T. and Ohgi, T. (2004), 'Crystal structure of human ISG20, an interferon-induced antiviral ribonuclease', *FEBS Letters* **577**(1-2), 111–116.

Hsiao, Y. Y., Duh, Y., Chen, Y. P., Wang, Y. T. and Yuan, H. S. (2012), 'How an exonuclease decides where to stop in trimming of nucleic acids: Crystal structures of RNase T-product complexes', *Nucleic Acids Research* **40**(16), 8144–8154.

Hsiao, Y. Y., Fang, W. H., Lee, C. C., Chen, Y. P. and Yuan, H. S. (2014), 'Structural Insights Into DNA Repair by RNase T-An Exonuclease Processing 3′ End of Structured DNA in Repair Pathways', *PLoS Biology* **12**(3), e1001803.

Hsiao, Y. Y., Yang, C. C., Lin, C. L., Lin, J. L., Duh, Y. and Yuan, H. S. (2011), 'Structural basis for RNA trimming by RNase T in stable RNA 3′-end maturation', *Nature Chemical Biology* **7**(4), 236–243.

Huang, S., Li, B., Gray, M. D., Oshima, J., Mian, I. S. and Campisi, J. (1998), 'The premature ageing syndrome protein, WRN, is a 3′-5′ exonuclease', *Nature Genetics* **20**(2), 114–116.

Hurtig, J. E. and van Hoof, A. (2022), 'Yeast Dxo1 is required for 25S rRNA maturation and acts as a transcriptome-wide distributive exonuclease', *RNA* **28**(5), 657–667.

Iben, J. R. and Maraia, R. J. (2012), 'tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote', *RNA* **18**(7), 1358–1372.

Iqbal, A., Wang, L., Thompson, K. C., Lilley, D. M. and Norman, D. G. (2008), 'The structure of cyanine 5 terminally attached to double-stranded DNA: Implications for FRET studies', *Biochemistry* **47**(30), 7857–7862.

Iverson, T. M., Panosian, T. D., Birmingham, W. R., Nannemann, D. P. and Bachmann, B. O. (2012), 'Molecular differences between a mutase and a phosphatase: Investigations of the activation step in *Bacillus cereus* phosphopentomutase', *Biochemistry* **51**(9), 1964–1975.

Jain, C. (2020), 'RNase AM, a 5′ to 3′ exonuclease, matures the 5′ end of all three ribosomal RNAs in *E. coli*', *Nucleic Acids Research* **48**(10), 5616–5623.

Jobe, A. and Bourgeois, S. (1972), 'lac repressor-operator interaction: VI. The natural inducer of the lac operon', *Journal of Molecular Biology* **69**(3), 397–408.

Jonas, S. and Hollfelder, F. (2009), 'Mapping catalytic promiscuity in the alkaline phosphatase superfamily', *Pure and Applied Chemistry* **81**(4), 731–742.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. and Hunter, S. (2014), 'InterProScan 5: genome-scale protein function classification', *Bioinformatics* **30**(9), 1236–1240.

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., 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**(7873), 583–589.

Kadaba, S., Krueger, A., Trice, T., Krecic, A. M., Hinnebusch, A. G. and Anderson, J. (2004), 'Nuclear surveillance and degradation of hypomodified initiator tRNAMet in *S. cerevisiae*', *Genes Development* **18**(11), 1227–1240.

Kadaba, S., Wang, X. and Anderson, J. T. (2006), 'Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA', *RNA* **12**(3), 508–521.

Karwan, R., Wintersberger, U. and Blutsch, H. (1983), 'Physical Association of a DNA Polymerase Stimulating Activity with a Ribonuclease H Purified from Yeast', *Biochemistry* **22**(24), 5500–5507.

Kawauchi, J., Mischo, H., Braglia, P., Rondon, A. and Proudfoot, N. J. (2008), 'Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination', *Genes Development* **22**(8), 1082–1092.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. (2015), 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols* **10**(6), 845–858.

Kempers-Veenstra, A. E., Oliemans, J., Offenberg, H., Dekker, A. F., Piper, P. W., Planta, R. J. and Klootwijk, J. (1986), '3′-End formation of transcripts from the yeast rRNA operon.', *The EMBO Journal* **5**(10), 2703–2710.

Kennedy, S., Wang, D. and Ruvkun, G. (2004), 'A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*', *Nature* **427**(6975), 645–649.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F. and Buratowski, S. (2004), 'The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II', *Nature* **432**(7016), 517–522.

Kim, S. K., Lormand, J. D., Weiss, C. A., Eger, K. A., Turdiev, H., Turdiev, A., Winkler, W. C., Sondermann, H. and Lee, V. T. (2019), 'A dedicated diribonucleotidase resolves a key bottleneck for the terminal step of RNA degradation', *eLife* **21**(8), e46313.

Kishor, A., Fritz, S. E. and Hogg, J. R. (2019), 'Nonsense-mediated mRNA decay: the challenge of telling right from wrong in a complex transcriptome', *Wiley Interdisciplinary Reviews: RNA* **10**(6), e1548.

Kögel, A., Keidel, A., Bonneau, F., Schäfer, I. B. and Conti, E. (2022), 'The human SKI complex regulates channeling of ribosome-bound RNA to the exosome via an intrinsic gatekeeping mechanism', *Molecular Cell* **82**(4), 756–769.

Kufel, J., Allmang, C., Verdone, L., Beggs, J. D. and Tollervey, D. (2002), 'Lsm Proteins Are Required for Normal Processing of Pre-tRNAs and Their Efficient Association with La-Homologous Protein Lhp1p', *Molecular and Cellular Biology* **22**(14), 5248–5256.

Kufel, J., Dichtl, B. and Tollervey, D. (1999), 'Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3′ ETS but not the 5′ ETS.', *RNA* **5**(7), 909–917.

Kufel, J. and Tollervey, D. (2003), '3′-processing of yeast tRNATrp precedes 5′-processing', *RNA* **9**(2), 202–208.

Kupsco, J. M., Wu, M. J., Marzluff, W. F., Thapar, R. and Duronio, R. J. (2006), 'Genetic and biochemical characterization of *Drosophila* Snipper: A promiscuous member of the metazoan 3′hExo/ERI-1 family of 3′ to 5′ exonucleases', *RNA* **12**(12), 2103–2117.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. and Tollervey, D. (2005), 'RNA degradation by the exosome is promoted by a nuclear polyadenylation complex', *Cell* **121**(5), 713–724.

Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. and Allen, K. N. (2003), 'The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction', *Science* **299**(5615), 2067–2071.

Lan, P., Tan, M., Zhang, Y., Niu, S., Chen, J., Shi, S., Qiu, S., Wang, X., Peng, X., Cai, G., Cheng, H., Wu, J., Li, G. and Lei, M. (2018), 'Structural insight into precursor tRNA processing by yeast ribonuclease P', *Science* **362**(6415), eaat6678.

Lebreton, A., Tomecki, R., Dziembowski, A. and Séraphin, B. (2008), 'Endonucleolytic RNA cleavage by a eukaryotic exosome', *Nature* **456**(7224), 993–996.

Lee, A., Hansen, K. D., Bullard, J., Dudoit, S. and Sherlock, G. (2008), 'Novel Low Abundance and Transient RNAs in Yeast Revealed by Tiling Microarrays and Ultra High-Throughput Sequencing Are Not Conserved Across Closely Related Yeast Species', *PLoS Genetics* **4**(12), e1000299.

Lee, J. Y., Rohlman, C. E., Molony, L. A. and Engelke, D. R. (1991), 'Characterization of *RPR1*, an essential gene encoding the RNA component of *Saccharomyces cerevisiae* nuclear RNase P.', *Molecular and Cellular Biology* **11**(2), 721–730.

Lemay, J. F. and Bachand, F. (2015), 'Fail-safe transcription termination: Because one is never enough', *RNA Biology* **12**(9), 927–932.

Leshin, J. A., Heselpoth, R., Belew, A. T. and Dinman, J. D. (2011), 'High throughput structural analysis of yeast ribosomes using hSHAPE', *RNA Biology* **8**(3), 478–487.

Letunic, I. and Bork, P. (2019), 'Interactive Tree Of Life (iTOL) v4: recent updates and new developments', *Nucleic Acids Research* **47**(W1), W256–W259.

Leung, E., Schneider, C., Yan, F., Mohi-El-Din, H., Kudla, G., Tuck, A., Wlotzka, W., Doronina, V. A., Bartley, R., Watkins, N. J., Tollervey, D. and Brown, J. D. (2014), 'Integrity of SRP RNA is ensured by La and the nuclear RNA quality control machinery', *Nucleic Acids Research* **42**(16), 10698–10710.

Li, Z. and Deutscher, M. P. (1994), 'The role of individual exoribonucleases in processing at the 3′ end of *Escherichia coli* tRNA precursors.', *Journal of Biological Chemistry* **269**(8), 6064–6071.

Li, Z. and Deutscher, M. P. (1995), 'The tRNA processing enzyme RNase T is essential for maturation of 5S RNA.', *Proceedings of the National Academy of Sciences of the United States of America* **92**(15), 6883â6886.

Li, Z. and Deutscher, M. P. (1996), 'Maturation pathways for *E. coli* tRNA precursors: A random multienzyme process in vivo', *Cell* **86**(3), 503–512.

Li, Z. and Deutscher, M. P. (2002), 'RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors.', *RNA* **8**(1), 97–109.

Li, Z., Jaroszewski, L., Iyer, M., Sedova, M. and Godzik, A. (2020), 'FATCAT 2.0: Towards a better understanding of the structural diversity of proteins', *Nucleic Acids Research* **48**(1), W60–W64.

Li, Z., Pandit, S. and Deutscher, M. P. (1998), '3′ Exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* **95**(6), 2856–2861.

Li, Z., Pandit, S. and Deutscher, M. P. (1999), 'Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T.', *RNA* **5**(1), 139–146.

Liang, S. D., Marmorstein, R., Harrison, S. C. and Ptashne, M. (1996), 'DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA.', *Molecular and Cellular Biology* **16**(7), 3773–3780.

Lindahl, T. (2013), 'My Journey to DNA Repair', *Genomics, Proteomics Bioinformatics* **11**(1), 2–7.

Lindahl, T., Gally, J. A. and Edelman, G. M. (1969), 'Properties of deoxyribonuclease 3 from mammalian tissues.', *Journal of Biological Chemistry* **244**(18), 5014–5019.

Lindström, M. S., Bartek, J. and Maya-Mendoza, A. (2022), 'p53 at the crossroad of DNA replication and ribosome biogenesis stress pathways', *Cell Death and Differentiation* **29**(5), 972–982.

Liu, Q., Greimann, J. C. and Lima, C. D. (2006), 'Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome', *Cell* **127**(6), 1223–1237.

Losson, R. and Lacroute, F. (1979), 'Interference of nonsense mutations with eukaryotic messenger RNA stability.', *Proceedings of the National Academy of Sciences of the United States of America* **76**(10), 5134–5137.

Lygerou, Z., Mitchell, P., Petfalski, E., Séraphin, B. and Tollervey, D. (1994), 'The *POP1* gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins.', *Genes Development* **8**(12), 1423–1433.

Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A. and Lopez, R. (2022), 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic Acids Research* **50**(W1), W276–W279.

Makino, D. L., Schuch, B., Stegmann, E., Baumgärtner, M., Basquin, C. and Conti, E. (2015), 'RNA degradation paths in a 12-subunit nuclear exosome complex', *Nature* **524**(7563), 54–58.

Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A. and Ross, J. (1981), 'Unstable β-globin mRNA in mRNA-deficient β0 thalassemia', *Cell* **27**(3), 543–553.

Marck, C. and Grosjean, H. (2002), 'tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features', *RNA* **8**(10), 1189–1232.

Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S. C. (1992), 'DNA recognition by GAL4: structure of a protein-DNA complex', *Nature* **356**(6368), 408–414.

Martin, F. H. and Tinoco, I. (1980), 'DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription', *Nucleic Acids Research* **8**(10), 2295–2300.

Mayes, A. E., Verdone, L., Legrain, P. and Beggs, J. D. (1999), 'Characterization of Sm-like proteins in yeast and their association with U6 snRNA.', *The EMBO Journal* **18**(15), 4321–4331.

Mazur, D. J. and Perrino, F. W. (1999), 'Identification and Expression of the *TREX1* and *TREX2* cDNA Sequences Encoding Mammalian 3′-5′ Exonucleases', *Journal of Biological Chemistry* **274**(28), 19655–19660.

Mazur, D. J. and Perrino, F. W. (2001), 'Excision of 3′ Termini by the Trex1 and TREX2 3′-5′ Exonucleases: CHARACTERIZATION OF THE RECOMBINANT PROTEINS', *Journal of Biological Chemistry* **276**(20), 17022–17029.

McHenry, C. S. (1985), 'DNA polymerase III holoenzyme of *Escherichia coli*: Components and function of a true replicative complex', *Molecular and Cellular Biochemistry* **66**(1), 71–85.

Milligan, L., Decourty, L., Saveanu, C., Rappsilber, J., Ceulemans, H., Jacquier, A. and Tollervey, D. (2008), 'A Yeast Exosome Cofactor, Mpp6, Functions in RNA Surveillance and in the Degradation of Noncoding RNA Transcripts', *Molecular and Cellular Biology* **28**(17), 5446–5457.

Minskaia, E., Hertzig, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B. and Ziebuhr, J. (2006), 'Discovery of an RNA virus 3′-5′ exoribonuclease that is critically involved in coronavirus RNA synthesis', *Proceedings of the National Academy of Sciences of the United States of America* **103**(13), 5108–5113.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. C., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D. and Bateman, A. (2021), 'Pfam: The protein families database in 2021', *Nucleic Acids Research* **49**(D1), D412–D419.

Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M. and Tollervey, D. (2003), 'Rrp47p is an exosome-associated protein required for the 3′ processing of stable RNAs', *Molecular and Cellular Biology* **23**(19), 6982–6992.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D. (1997), 'The exosome: A conserved eukaryotic RNA processing complex containing multiple 3′-5′ exoribonucleases', *Cell* **91**(4), 457–466.

Mitchell, P., Petfalski, E. and Tollervey, D. (1996), 'The 3′ end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism', *Genes Development* **10**(4), 502–513.

Miyazaki, M. (1974), 'Studies on the Nucleotide Sequence of Pseudouridine-containing 5S RNA from *Saccharomyces cerevisiae*', *The Journal of Biochemistry* **75**(6), 1407–1410.

Mohanty, B. K. and Kushner, S. R. (2007), 'Ribonuclease P processes polycistronic tRNA transcripts in *Escherichia coli* independent of ribonuclease E', *Nucleic Acids Research* **35**(22), 7614–7625.

Mori, Y. and Ichiyanagi, K. (2021), 'melRNA-seq for Expression Analysis of SINE RNAs and Other Medium-Length Non-Coding RNAs', *Mobile DNA* **12**(1), 15.

Morozov, V., Mushegian, A. R., Koonin, E. V. and Bork, P. (1997), 'A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases', *Trends in Biochemical Sciences* **22**(11), 417–418.

Motley, A. M., Nuttall, J. M. and Hettema, E. H. (2012), 'Pex3-anchored Atg36 tags peroxisomes for degradation in *Saccharomyces cerevisiae*', *The EMBO Journal* **31**(13), 2852–2868.

Neil, H., Malabat, C., D'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M. and Jacquier, A. (2009), 'Widespread bidirectional promoters are the major source of cryptic transcripts in yeast', *Nature* **457**(7232), 1038–1042.

Nguyen, L. H., Erzberger, J. P., Root, J. and Wilson, D. M. (2000), 'The human homolog of *Escherichia coli* Orn degrades small single-stranded RNA and DNA oligomers', *Journal of Biological Chemistry* **275**(34), 25900–25906.

Nguyen, T. H. D., Galej, W. P., Bai, X. C., Savva, C. G., Newman, A. J., Scheres, S. H. and Nagai, K. (2015), 'The architecture of the spliceosomal U4/U6.U5 tri-snRNP', *Nature* **523**(7558), 47–52.

Nielsen, S., Yuzenkova, Y. and Zenkin, N. (2013), 'Mechanism of eukaryotic RNA polymerase III transcription termination', *Science* **340**(6140), 1577–1580.

Niyogi, S. K. and Datta, A. K. (1975), 'A novel oligoribonuclease of *Escherichia coli*. I. Isolation and properties', *Journal of Biological Chemistry* **250**(18), 7307–7312.

Norman, D. G., Grainger, R. J., Uhrín, D. and Lilley, D. M. (2000), 'Location of Cyanine-3 on Double-Stranded DNA: Importance for Fluorescence Resonance Energy Transfer Studies', *Biochemistry* **39**(21), 6317–6324.

O'Connor, J. P. and Peebles, C. L. (1991), 'In vivo pre-tRNA processing in *Saccharomyces cerevisiae*.', *Molecular and Cellular Biology* **11**(1), 425–439.

Ozanick, S. G., Wang, X., Costanzo, M., Brost, R. L., Boone, C. and Anderson, J. T. (2009), 'Rex1p deficiency leads to accumulation of precursor initiator tRNAMet and polyadenylation of substrate RNAs in *Saccharomyces cerevisiae*', *Nucleic Acids Research* **37**(1), 298–308.

Pal, A. and Levy, Y. (2019), 'Structure, stability and specificity of the binding of ssDNA and ssRNA with proteins', *PLoS Computational Biology* **15**(4), e1006768.

Patterson, B. and Guthrie, C. (1987), 'An essential yeast snRNA with a U5-like domain is required for splicing in vivo', *Cell* **49**(5), 613–624.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., Sigrist, C. J., Sillitoe, I., Thanki, N., Thomas, P. D., Tosatto, S. C., Wu, C. H. and Bateman, A. (2023), 'InterPro in 2022', *Nucleic Acids Research* **51**(D1), D418–D427.

Peart, N., Sataluri, A., Baillat, D. and Wagner, E. J. (2013), 'Non mRNA 3′ End Formation: How the other half lives', *Wiley Interdisciplinary Reviews: RNA* **4**(5), 491–506.

Peng, W. T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W., Mitsakakis, N., Ryan, O. W., Datta, N., Jojic, V., Pal, C., Canadien, V., Richards, D., Beattie, B., Wu, L. F., Altschuler, S. J., Roweis, S., Frey, B. J., Emili, A., Greenblatt, J. F. and Hughes, T. R. (2003), 'A panoramic view of yeast noncoding RNA processing', *Cell* **113**(7), 919–933.

Petrov, A. S., Bernier, C. R., Gulen, B., Waterbury, C. C., Hershkovits, E., Hsiao, C., Harvey, S. C., Hud, N. V., Fox, G. E., Wartell, R. M. and Williams, L. D. (2014), 'Secondary Structures of rRNAs from All Three Domains of Life', *PLoS ONE* **9**(2), e88222.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. and Ferrin, T. E. (2021), 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', *Protein Science* **30**(1), 70–82.

Philippsen, P., Thomas, M., Kramer, R. A. and Davis, R. W. (1978), 'Unique arrangement of coding sequences for 5 S, 5.8 S, 18 S and 25 S ribosomal RNA in *Saccharomyces cerevisiae* as determined by R-loop and hybridization analysis', *Journal of Molecular Biology* **123**(3), 387–404.

Phillips, S. and Butler, J. S. (2003), 'Contribution of domain structure to the RNA 3′ end processing and degradation functions of the nuclear exosome subunit Rrp6p', *RNA* **9**(9), 1098–1107.

Phizicky, E. M. and Hopper, A. K. (2023), 'The life and times of a tRNA', *RNA* **29**(7), 898–957.

Piper, P. W., Bellatin, J. A. and Lockheart, A. (1983), 'Altered maturation of sequences at the 3′ terminus of 5S gene transcripts in a *Saccharomyces cerevisiae* mutant that lacks a RNA processing endonuclease.', *The EMBO journal* **2**(3), 353–359.

Piper, P. W. and Stråby, K. B. (1989), 'Processing of transcripts of a dimeric tRNA gene in yeast uses the nuclease responsible for maturation of the 3′ termini upon 5 S and 37 S precursor rRNAs', *FEBS Letters* **250**(2), 311–316.

Porrua, O. and Libri, D. (2013), 'RNA quality control in the nucleus: The Angels' share of RNA', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1829**(6-7), 604–611.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H. and Jensen, T. H. (2008), 'RNA exosome depletion reveals transcription upstream of active human promoters', *Science* **322**(5909), 1851–1854.

Reichow, S. L., Hamma, T., Ferré-D'Amaré, A. R. and Varani, G. (2007), 'The structure and function of small nucleolar ribonucleoproteins', *Nucleic Acids Research* **35**(5), 1452–1464.

Reinhold-Hurek, B. and Shub, D. A. (1992), 'Self-splicing introns in tRNA genes of widely divergent bacteria', *Nature* **357**(6374), 173–176.

Reuven, N. B. and Deutscher, M. P. (1993), 'Multiple exoribonucleases are required for the 3′ processing of *Escherichia coli* tRNA precursors in vivo', *The FASEB Journal* **7**(1), 143–148.

Sarngadharan, M. G., Leis, J. P. and Gallo, R. C. (1975), 'Isolation and characterization of a ribonuclease from human leukemic blood cells specific for ribonucleic acid of ribonucleic acid deoxyribonucleic acid hybrid molecules', *Journal of Biological Chemistry* **250**(2), 365–373.

Schaeffer, D., Tsanova, B., Barbas, A., Reis, F. P., Dastidar, E. G., Sanchez-Rotunno, M., Arraiano, C. M. and Van Hoof, A. (2009), 'The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities', *Nature Structural Molecular Biology* **16**(1), 56–62.

Schäfer, I. B., Yamashita, M., Schuller, J. M., Schüssler, S., Reichelt, P., Strauss, M. and Conti, E. (2019), 'Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylase', *Cell* **177**(6), 1619–1631.

Schilders, G., Raijmakers, R., Raats, J. M. and Pruijn, G. J. (2005), 'MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation', *Nucleic Acids Research* **33**(21), 6795–6804.

Schillewaert, S., Wacheul, L., Lhomme, F. and Lafontaine, D. L. J. (2012), 'The Evolutionarily Conserved Protein LAS1 Is Required for Pre-rRNA Processing at Both Ends of ITS2', *Molecular and Cellular Biology* **32**(2), 430–444.

Schmidt, C., Becker, T., Heuer, A., Braunger, K., Shanmuganathan, V., Pech, M., Berninghausen, O., Wilson, D. N. and Beckmann, R. (2016), 'Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome', *Nucleic Acids Research* **44**(4), 1944–1951.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012), 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods 2012 9:7* **9**(7), 671–675.

Schneider, C., Anderson, J. T. and Tollervey, D. (2007), 'The Exosome Subunit Rrp44 Plays a Direct Role in RNA Substrate Recognition', *Molecular Cell* **27**(2), 324–331.

Schneider, C. and Bohnsack, K. E. (2023), 'Caught in the act-Visualizing ribonucleases during eukaryotic ribosome assembly', *Wiley Interdisciplinary Reviews: RNA* **14**(4), e1766.

Schneider, C., Kudla, G., Wlotzka, W., Tuck, A. and Tollervey, D. (2012), 'Transcriptome-wide Analysis of Exosome Targets', *Molecular Cell* **48**(3-3), 422–433.

Schneider, C., Leung, E., Brown, J. and Tollervey, D. (2009), 'The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome', *Nucleic Acids Research* **37**(4), 1127–1140.

Schoepfer, R. (1993), 'The pRSET family of T7 promoter expression vectors for *Escherichia coli*', *Gene* **124**(1), 83–85.

Schuch, B., Feigenbutz, M., Makino, D. L., Falk, S., Basquin, C., Mitchell, P. and Conti, E. (2014), 'The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase', *The EMBO Journal* **33**(23), 2829–2846.

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A. and Sharp, P. A. (2008), 'Divergent transcription from active promoters', *Science* **322**(5909), 1849–1851.

Senisterra, G. A. and Finerty, P. J. (2009), 'High throughput methods of assessing protein stability and aggregation', *Molecular BioSystems* **5**(3), 217–223.

Shapiro, A. L., Viñuela, E. and V. Maizel Jr., J. (1967), 'Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels', *Biochemical and Biophysical Research Communications* **28**(5), 815–820.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. (2011), 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology* **7**, 539–539.

Sikorski, R. S. and Hieter, P. (1989), 'A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces cerevisiae*', *Genetics* **122**(1), 19–27.

Silva, S., Homolka, D. and Pillai, R. S. (2017), 'Characterization of the mammalian RNA exonuclease 5/NEF-sp as a testis-specific nuclear 3′ - 5′ exoribonuclease', *RNA* **23**(9), 1385–1392.

Sim, G. Y., Kehling, A. C., Park, M. S., Secor, J., Divoky, C., Zhang, H., Malhotra, N., Bhagdikar, D., Abd El-Wahab, E. W. and Nakanishi, K. (2022), 'Manganese-dependent microRNA trimming by 3′-5′ exonucleases generates 14-nucleotide or shorter tiny RNAs', *Proceedings of the National Academy of Sciences of the United States of America* **119**(51), e2214335119.

Skowronek, E., Grzechnik, P., Späth, B., Marchfelder, A. and Kufel, J. (2014), 'tRNA 3′ processing in yeast involves tRNase Z, Rex1, and Rrp6', *RNA* **20**(1), 115–130.

Smith, D. B. and Johnson, K. S. (1988), 'Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase', *Gene* **67**(1), 31–40.

Soave, C., NUCCA, R., Sala, E., VIOTTI, A. and Galante, E. (1973), '5-S RNA: Investigation of the Different Extent of Phosphorylation at 5′ Terminus', *European Journal of Biochemistry* **32**(2), 392–400.

Staals, R. H., Bronkhorst, A. W., Schilders, G., Slomovic, S., Schuster, G., Heck, A. J., Raijmakers, R. and Pruijn, G. J. (2010), 'Dis3-like 1: A novel exoribonuclease associated with the human exosome', *EMBO Journal* **29**(14), 2358–2367.

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006), 'BioGRID: a general repository for interaction datasets.', *Nucleic acids research* **1**(34), D535–D539.

Stead, J. A., Costello, J. L., Livingstone, M. J. and Mitchell, P. (2007), 'The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein', *Nucleic Acids Research* **35**(16), 5556–5567.

Steitz, J. A., Berg, C., Hendrick, J. P., La Branche-Chabot, H., Metspalu, A., Rinke, J. and Yario, T. (1988), 'A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells.', *Journal of Cell Biology* **106**(3), 545–556.

Steitz, T. A. and Steitz, J. A. (1993), 'A general two-metal-ion mechanism for catalytic RNA', *Proceedings of the National Academy of Sciences of the United States of America* **90**(14), 6498–6502.

Stevens, A. and Niyogi, S. K. (1967), 'Hydrolysis of oligoribonucleotides by an enzyme fraction from *Escherichia coli*', *Biochemical and Biophysical Research Communications* **29**(4), 550–555.

Studier, F. W. (2005), 'Protein production by auto-induction in high density shaking cultures.', *Protein Expression and Purification* **41**(1), 207–234.

Studier, F. W. and Moffatt, B. A. (1986), 'Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes', *Journal of Molecular Biology* **189**(1), 113–130.

Szewczyk, M., Malik, D., Borowski, L. S., Czarnomska, S. D., Kotrys, A. V., Klosowska-Kosicka, K., Nowotny, M. and Szczesny, R. J. (2020), 'Human REXO2 controls short mitochondrial RNAs generated by mtRNA processing and decay machinery to prevent accumulation of double-stranded RNA', *Nucleic Acids Research* **48**(10), 5572–5590.

Tang, T. T., Stowell, J. A., Hill, C. H. and Passmore, L. A. (2019), 'The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases', *Nature Structural Molecular Biology* **26**(6), 433–442.

Thompson, D. M. and Parker, R. (2007), 'Cytoplasmic Decay of Intergenic Transcripts in *Saccharomyces cerevisiae*', *Molecular and Cellular Biology* **27**(1), 92–101.

Thomson, E. and Tollervey, D. (2010), 'The Final Step in 5.8S rRNA Processing Is Cytoplasmic in *Saccharomyces cerevisiae*', *Molecular and Cellular Biology* **30**(4), 976–984.

Tomecki, R., Kristiansen, M. S., Lykke-Andersen, S., Chlebowski, A., Larsen, K. M., Szczesny, R. J., Drazkowska, K., Pastula, A., Andersen, J. S., Stepien, P. P., Dziembowski, A. and Jensen, T. H. (2010), 'The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L', *The EMBO Journal* **29**(14), 2342.

Turowski, T. W., Leśniewska, E., Delan-Forino, C., Sayou, C., Boguta, M. and Tollervey, D. (2016), 'Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts', *Genome Research* **26**(7), 933–944.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I. M., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A. K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., Von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Von Heijne, G., Nielsen, J. and Pontén, F. (2015), 'Tissue-based map of the human proteome', *Science* **347**(6220), 1260419.

Ullah, R., Ali Shah, M., Tufail, S., Ismat, F., Imran, M., Iqbal, M., Mirza, O. and Rhaman, M. (2016), 'Activity of the Human Rhinovirus 3C Protease Studied in Various Buffers, Additives and Detergents Solutions for Recombinant Protein Production', *PLoS ONE* **11**(4), e0153436.

van Hoof, A., Lennertz, P. and Parker, R. (2000a), 'Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5. RNase MRP and RNase P RNAs in yeast', *EMBO Journal* **19**(6), 1357–1365.

van Hoof, A., Lennertz, P. and Parker, R. (2000b), 'Yeast Exosome Mutants Accumulate 3′-Extended Polyadenylated Forms of U4 Small Nuclear RNA and Small Nucleolar RNAs', *Molecular and Cellular Biology* **20**(2), 441–452.

van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J. and Steinegger, M. (2023), 'Fast and accurate protein structure search with Foldseek', *Nature Biotechnology* .

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G. and Keller, W. (2005), 'A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control', *PLoS Biology* **3**(6), 0986–0997.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. and Velankar, S. (2022), 'AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research* **50**(D1), D439–D444.

Villa, T., Porrua, O., Villa, C. T. and Porrua, O. (2023), 'Pervasive transcription: a controlled risk', *The FEBS Journal* **290**(15), 3723–3736.

Viswanathan, M., Dower, K. W. and Lovett, S. T. (1998), 'Identification of a Potent DNase Activity Associated with RNase T of *Escherichia coli*', *Journal of Biological Chemistry* **273**(52), 35126–35131.

Viswanathan, M., Lanjuin, A. and Lovett, S. T. (1999), 'Identification of RNase T as a high-copy suppressor of the UV sensitivity associated with single-strand DNA exonuclease deficiency in *Escherichia coli*.', *Genetics* **151**(3), 929–934.

Wan, J., Yourshaw, M., Mamsa, H., Rudnik-Schöneborn, S., Menezes, M. P., Hong, J. E., Leong, D. W., Senderek, J., Salman, M. S., Chitayat, D., Seeman, P., Von Moers, A., Graul-Neumann, L., Kornberg, A. J., Castro-Gago, M., Sobrido, M. J., Sanefuji, M., Shieh, P. B., Salamon, N., Kim, R. C., Vinters, H. V., Chen, Z., Zerres, K., Ryan, M. M., Nelson, S. F. and Jen, J. C. (2012), 'Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration', *Nature Genetics* **44**(6), 704–708.

Wang, J. Y., Tuck, O. T., Skopintsev, P., Soczek, K. M., Li, G., Al-Shayeb, B., Zhou, J. and Doudna, J. A. (2023), 'Genome expansion by a CRISPR trimmer-integrase', *Nature* **618**(7966), 855–866.

Wang, Q., Du, J., Hua, S. and Zhao, K. (2022), 'TREX1 plays multiple roles in human diseases', *Cellular Immunology* **375**, 104527.

Wasmuth, E. V., Januszyk, K. and Lima, C. D. (2014), 'Structure of an Rrp6-RNA exosome complex bound to polyA RNA', *Nature* **511**(7510), 435–439.

Wasmuth, E. V., Zinder, J. C., Zattas, D., Das, M. and Lima, C. D. (2017), 'Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that mpp6 stimulates RNA decay and recruits the Mtr4 helicase', *eLife* **25**(6), e29062.

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. and Barton, G. J. (2009), 'Jalview Version 2-a multiple sequence alignment editor and analysis workbench', *Bioinformatics* **25**(9), 1189–1191.

Wellner, K., Betat, H. and Mörl, M. (2018), 'A tRNA's fate is decided at its 3′ end: Collaborative actions of CCA-adding enzyme and RNases involved in tRNA processing and degradation', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1861**(4), 433–441.

West, S., Gromak, N. and Proudfoot, N. J. (2004), 'Human 5′ –> 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites', *Nature* **432**(7016), 522–525.

West, S., Proudfoot, N. J. and Dye, M. J. (2008), 'Molecular dissection of mammalian RNA polymerase II transcriptional termination', *Molecular Cell* **29**(5), 600–610.

Whipple, J. M., Lane, E. A., Chernyakov, I., D'Silva, S. and Phizicky, E. M. (2011), 'The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA', *Genes Development* **25**(11), 1173–1184.

Wilusz, J. E., Whipple, J. M., Phizicky, E. M. and Sharp, P. A. (2011), 'tRNAs marked with CCACCA are targeted for degradation', *Science* **334**(6057), 817–821.

Wlotzka, W., Kudla, G., Granneman, S. and Tollervey, D. (2011), 'The nuclear RNA polymerase II surveillance system targets polymerase III transcripts', *The EMBO Journal* **30**(9), 1790–1803.

Wong, Y. and Rosindell, J. (2022), 'Dynamic visualisation of million-tip trees: The OneZoom project', *Methods in Ecology and Evolution* **13**(2), 303–313.

Wu, M., Nilsson, P., Henriksson, N., Niedzwiecka, A., Lim, M. K., Cheng, Z., Kokkoris, K., Virtanen, A. and Song, H. (2009), 'Structural Basis of m7GpppG Binding to Poly(A)-Specific Ribonuclease', *Structure* **17**(2), 276–286.

Wu, M., Reuter, M., Lilie, H., Liu, Y., Wahle, E. and Song, H. (2005), 'Structural insight into poly(A) binding and catalytic mechanism of human PARN', *The EMBO Journal* **24**(23), 4082–4093.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D. and Jacquier, A. (2005), 'Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase', *Cell* **121**(5), 725–737.

Xie, J., Aiello, U., Clement, Y., Haidara, N., Girbig, M., Schmitzova, J., Pena, V., Müller, C. W., Libri, D. and Porrua, O. (2022), 'An integrated model for termination of RNA polymerase III transcription', *Science Advances* **8**(28), eabm9875.

Xie, J., Libri, D. and Porrua, O. (2023), 'Mechanisms of eukaryotic transcription termination at a glance', *Journal of Cell Science* **136**(1), jcs259873.

Xu, M. Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A. and Shub, D. A. (1990), 'Bacterial Origin of a Chloroplast Intron: Conserved Self-Splicing Group I Introns in Cyanobacteria', *Science* **250**(4987), 1566–1570.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W. and Steinmetz, L. M. (2009), 'Bidirectional promoters generate pervasive transcription in yeast', *Nature* **457**(7232), 1033–1037.

Yang, W. (2011), 'Nucleases: diversity of structure, function and mechanism', *Quarterly Reviews of Biophysics* **44**(1), 1–93.

Yang, X. C., Purdy, M., Marzluff, W. F. and Dominski, Z. (2006), 'Characterization of 3′hExo, a 3′ exonuclease specifically interacting with the 3′ end of histone mRNA', *Journal of Biological Chemistry* **281**(41), 30447–30454.

Yoo, C. J. and Wolin, S. L. (1997), 'The yeast La protein is required for the 3′ endonucleolytic cleavage that matures tRNA precursors', *Cell* **89**(3), 393–402.

Yoshihisa, T., Yunoki-Esaki, K., Ohshima, C., Tanaka, N. and Endo, T. (2003), 'Possibility of Cytoplasmic pre-tRNA Splicing: the Yeast tRNA Splicing Endonuclease Mainly Localizes on the Mitochondria', *Molecular Biology of the Cell* **14**(8), 3266–3279.

Yuan, F., Dutta, T., Wang, L., Song, L., Gu, L., Qian, L., Benitez, A., Ning, S., Malhotra, A., Deutscher, M. P. and Zhang, Y. (2015), 'Human DNA Exonuclease TREX1 Is Also an Exoribonuclease That Acts on Single-stranded RNA', *The Journal of Biological Chemistry* **290**(21), 13344–13353.

Zhang, G., Dai, J., Wang, L., Dunaway-Mariano, D., Tremblay, L. W. and Allen, K. N. (2005), 'Catalytic cycling in β-phosphoglucomutase: A kinetic and structural analysis', *Biochemistry* **44**(27), 9404–9416.

Zhang, J. and Deutscher, M. P. (1988), 'Transfer RNA is a substrate for RNase D *in vivo*', *Journal of Biological Chemistry* **263**(34), 17909–17912.

Zhang, J., Harnpicharnchai, P., Jakovljevic, J., Tang, L., Guo, Y., Oeffinger, M., Rout, M. P., Hiley, S. L., Hughes, T. and Woolford, J. L. (2007), 'Assembly factors Rpf2 and Rrs1 recruit 5S rRNA and ribosomal proteins rpL5 and rpL11 into nascent ribosomes', *Genes Development* **21**(20), 2580–2592.

Zinder, J. C. and Lima, C. D. (2017), 'Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors', *Genes Development* **31**(2), 88–100.

Zinder, J. C., Wasmuth, E. V. and Lima, C. D. (2016), 'Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3', *Molecular Cell* **64**(4), 734–745.

Zuo, Y. and Deutscher, M. P. (1999), 'The DNase activity of RNase T and its application to DNA cloning', *Nucleic Acids Research* **27**(20), 4077–4082.

Zuo, Y. and Deutscher, M. P. (2001), 'Exoribonuclease superfamilies: Structural analysis and phylogenetic distribution', *Nucleic Acids Research* **29**(5), 1017–1026.

Zuo, Y. and Deutscher, M. P. (2002a), 'Mechanism of action of RNase T: I. Identification of residues required for catalysis, substrate binding, and dimerization', *Journal of Biological Chemistry* **277**(51), 50155–50159.

Zuo, Y. and Deutscher, M. P. (2002b), 'The physiological role of RNase T can be explained by its unusual substrate specificity.', *The Journal of Biological Chemistry* **277**(33), 29654–29661.

Zuo, Y. and Deutscher, M. P. (2002c), 'Mechanism of action of RNase T: II. A structural and functional model of the enzyme', *Journal of Biological Chemistry* **277**(51), 50160–50164.

Zuo, Y., Wang, Y. and Malhotra, A. (2005), 'Crystal structure of *Escherichia coli* RNase D, an exoribonuclease involved in structured RNA processing', *Structure* **13**(7), 973–984.

Zuo, Y., Zheng, H., Wang, Y., Chruszcz, M., Cymborowski, M., Skarina, T., Savchenko, A., Malhotra, A. and Minor, W. (2007), 'Crystal Structure of RNase T, an Exoribonuclease Involved in tRNA Maturation and End Turnover', *Structure* **15**(4), 417–428.

Appendix A

Full multiple alignment of the DEDDh exonuclease domain

Table A.1: ClustalX conserved residue colour scheme
Colour scheme for aligned residues in Jalview, using the ClustalX colour scheme, based on https://www.jalview.org/help/html/colourSchemes/clustal.html [accessed 2023/09/26].

Category	Colour	Residue at position	Threshold for colour, residue groups
Hydrophobic	**Blue**	A, I, L, M, F, W, V	{>60%, WLVIMAFCHP}
		C	{>60%, WLVIMAFCHP}
Positive	**Red**	K, R	{>60%,KR},{>80%, K,R,Q}
Negative	**Magenta**	E	{>60%,KR}, {>50%,QE}, {>85%,E,Q,D}
		D	{>60%,KR}, {>85%, K,R,Q}, {>50%,ED}
Polar	**Green**	N	{>50%, N}, {>85%, N,Y}
		Q	{>60%,KR}, {>50%,QE}, {>85%,Q,E,K,R}
		S, T	{>60%, WLVIMAFCHP}, {>50%, TS}, {>85%,S,T}
Cysteines	**Pink**	C	{>85%, C}
Glycines	**Orange**	G	{>0%, G}
Prolines	**Yellow**	P	{>0%, P}
Aromatic	**Cyan**	H, Y	{>60%, WLVIMAFCHP}, {>85%, W,Y,A,C,P,Q,F,H,I,L,M,V}
Unconserved	**White**	Any / gap	If none of the above criteria are met

Figure A.1: Full-scale alignment of the DEDDh exonuclease domain

Uncropped alignment described in Figure 3.2 and Figure 3.3. Clustal Omega (Sievers *et al.*, 2011) multi-alignment of DEDDh exonuclease domains, generated in Jalview v2.11.2.7 (Waterhouse *et al.*, 2009). Clades are shown in coloured boxes that correspond to positions in the phylogenetic tree Figure 3.3. Alignment defined by Rex1 exonuclease domain amino acid sequence (225-373), the corresponding row is highlighted with a dotted red line. The positions of key residues in *S. cerevisiae* Rex1 are indicated beneath the alignment: in black for the three EXO motifs, in red for DEDDh residues, and in blue for predicted RNA-binding residues. Amino acids are coloured based on the Clustal X colour scheme, shown in Figure A.1. Alignment % for highlighting is calculated per phylogenetic tree-defined clade.

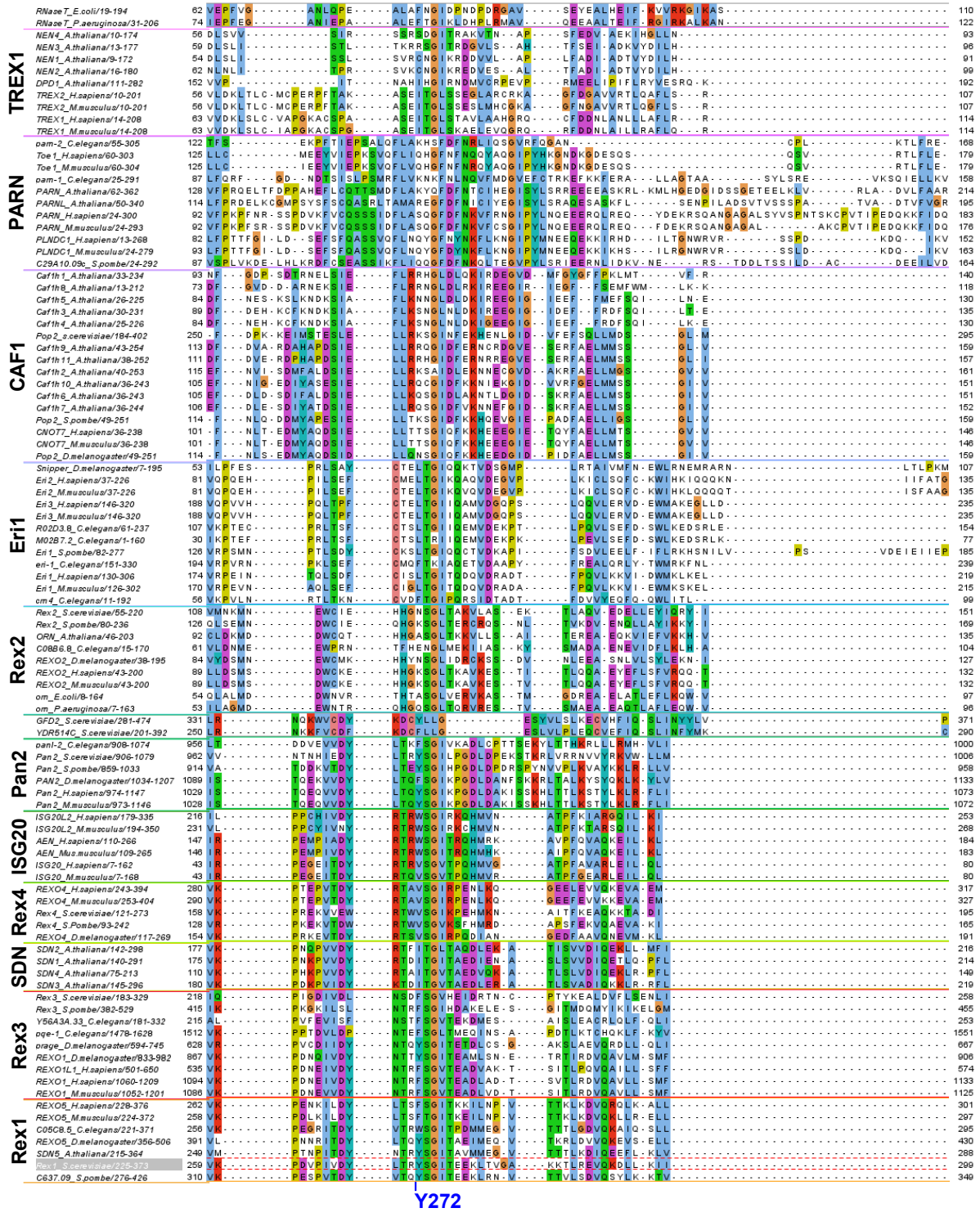


Figure A.1 continued (2/4)

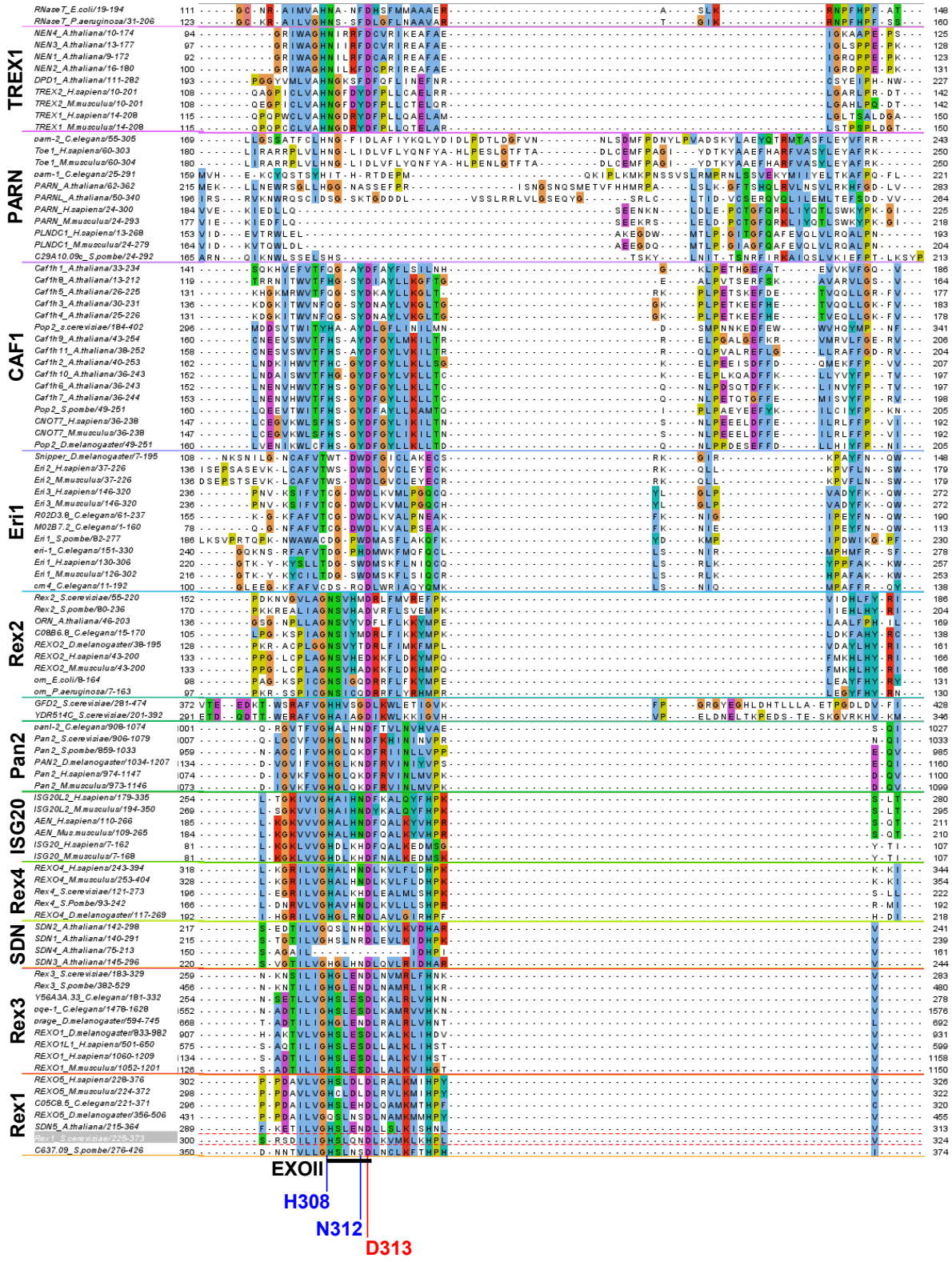


Figure A.1 continued (3/4)

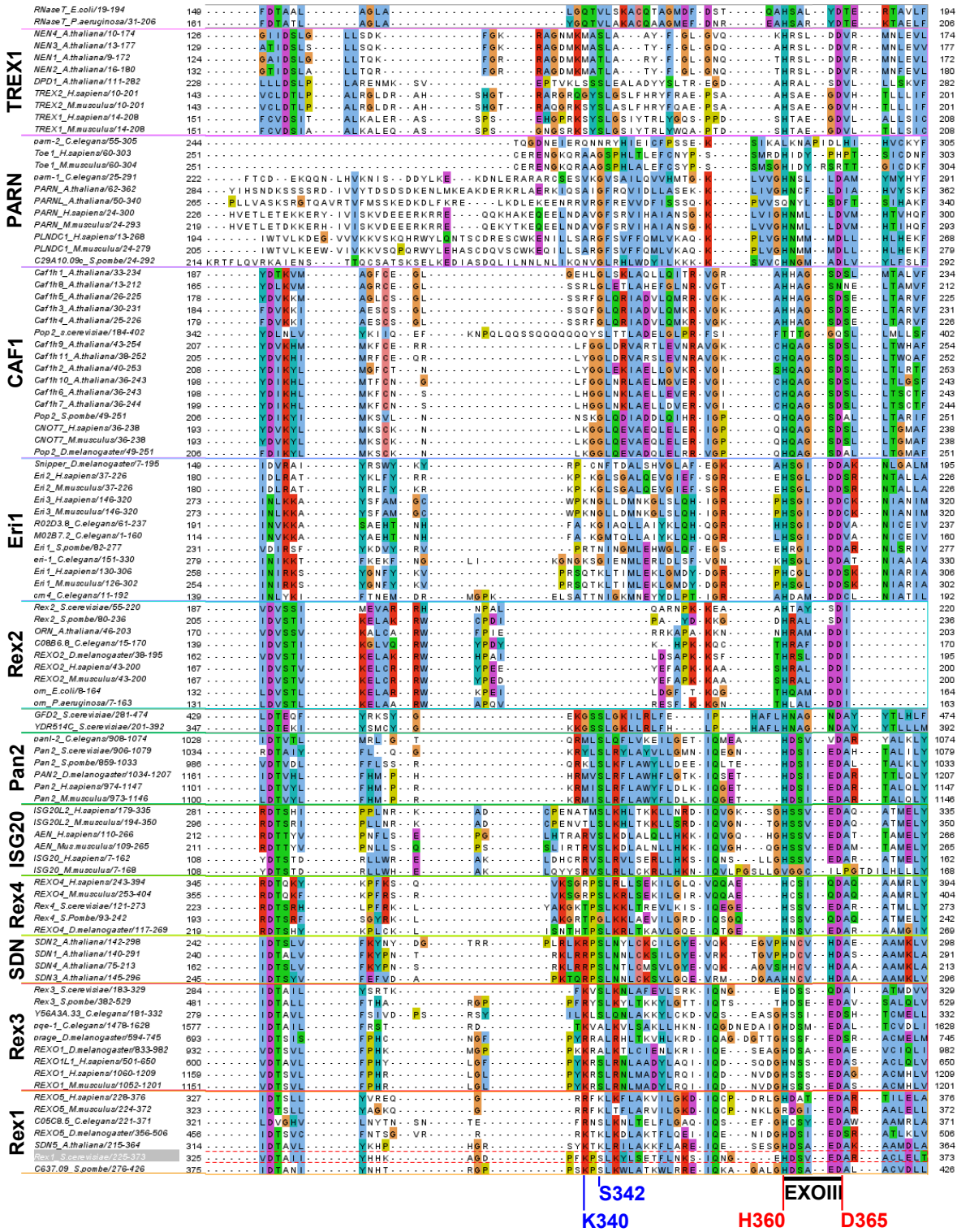


Figure A.1 continued (4/4)

Appendix B

InterPro nuclease superfamilies

Table B.1: List of InterPro-defined nuclease superfamilies

Interpro (Paysan-Lafosse *et al.*, 2023) entries defined as 'Homologous Superfamilies', returned in a search of 'nucleases' on ebi.ac.uk/interpro. Families organised alphabetically, search accessible at https://www.ebi.ac.uk/interpro/entry/InterPro/?search=nuclease&type=homologous_superfamily [accessed 2023-09-26].

Accession	Description
IPR036279	5'-3' exonuclease, C-terminal domain superfamily
IPR047007	5'-3' exoribonuclease 1, D1 domain superfamily
IPR047008	5'-3' exoribonuclease 1, SH3-like domain superfamily
IPR036104	Bifunctional nuclease superfamily
IPR041963	BsuBI/PstI restriction endonuclease, C-terminal domain superfamily
IPR041962	BsuBI/PstI restriction endonuclease, N-terminal domain superfamily
IPR038233	Colicin D/E5 nuclease domain superfamily
IPR036725	Colicin E3-like ribonuclease domain superfamily
IPR038234	Colicin E5 ribonuclease domain superfamily
IPR042206	CRISPR-associated endonuclease Cas1, C-terminal domain
IPR042211	CRISPR-associated endonuclease Cas1, N-terminal domain
IPR042564	CRISPR-associated endoribonuclease Cas6/Csy4, subtype I-F/YPEST superfamily
IPR037057	DNA mismatch repair MutH/Type II restriction endonuclease superfamily
IPR044929	DNA/RNA non-specific endonuclease superfamily
IPR038691	DNA-entry nuclease inhibitor, ComJ superfamily
IPR043086	EME1, nuclease domain, subdomain 1
IPR043087	EME1, nuclease domain, subdomain 2
IPR049173	Endonuclease NucS, N-terminal domain superfamily
IPR036691	Endonuclease/exonuclease/phosphatase superfamily
IPR038241	Endoribonuclease antitoxin GhoS superfamily
IPR037227	Endoribonuclease EndoU-like
IPR038649	Exonuclease I, SH3-like domain superfamily
IPR037004	Exonuclease VII, small subunit superfamily
IPR036345	Exoribonuclease, PH domain 2 superfamily
IPR035901	GIY-YIG endonuclease superfamily
IPR027434	Homing endonuclease
IPR044930	Homing endonuclease, His-Me finger superfamily
IPR042034	Host-nuclease inhibitor Gam superfamily
IPR038372	Influenza RNA-dependent RNA polymerase subunit PA, endonuclease domain
IPR043004	MvaI/BcnI restriction endonuclease, catalytic domain
IPR043005	MvaI/BcnI restriction endonuclease, recognition domain
IPR036587	Nuclease A inhibitor-like superfamily

Table B.1 continued: List of InterPro-defined nuclease superfamilies
Bolded - the RNase H and RNase H-like superfamilies include the the DEDD exonuclease clade, including Rex1

Accession	Description
IPR012296	Nuclease, putative, TT1808
IPR011604	PD-(D/E)XK endonuclease-like domain superfamily
IPR008947	Phospholipase C/P1 nuclease domain superfamily
IPR037074	Protein of unknown function DUF1780, putative endonuclease superfamily
IPR038563	Recombination endonuclease VII superfamily
IPR012297	Restriction endonuclease EcoO109IR, catalytic domain superfamily
IPR038374	Restriction endonuclease ThaI superfamily
IPR011335	Restriction endonuclease type II-like
IPR043118	Restriction endonuclease, type II SfiI, domain 1
IPR043117	Restriction endonuclease, type II SfiI, domain 2
IPR043091	Restriction endonuclease, type II, AvaI/BsoBI, helical domain
IPR011338	Restriction endonuclease, type II, BamHI/BglIII/BstY
IPR043121	Restriction endonuclease, type II, BglI superfamily
IPR011336	Restriction endonuclease, type II, EcoRI/MunI
IPR038373	Restriction endonuclease, type II, HindIII superfamily
IPR038402	Restriction endonuclease, type II, PvuII superfamily
IPR036816	Ribonuclease A-like domain superfamily
IPR036704	Ribonuclease E inhibitor RraA/RraA-like superfamily
IPR036397	**Ribonuclease H superfamily**
IPR012337	**Ribonuclease H-like superfamily**
IPR037056	Ribonuclease H1, N-terminal domain superfamily
IPR023160	Ribonuclease HII, helix-loop-helix cap domain superfamily
IPR038133	Ribonuclease III superfamily, archaeal
IPR036389	Ribonuclease III, endonuclease domain superfamily
IPR042173	Ribonuclease J, domain 2
IPR036980	Ribonuclease P/MRP subunit Rpp29 superfamily
IPR036430	Ribonuclease T2-like superfamily
IPR038573	Ribonuclease toxin BrnT superfamily
IPR036866	Ribonuclease Z/Hydroxyacylglutathione hydrolase-like
IPR016191	Ribonuclease/ribotoxin
IPR024796	T4 endonuclease V
IPR036309	T4 recombination endonuclease VII, dimerisation domain superfamily
IPR011856	tRNA endonuclease-like domain superfamily
IPR036167	tRNA intron endonuclease, catalytic domain-like superfamily
IPR036740	tRNA intron endonuclease, N-terminal domain superfamily
IPR037038	tRNA nuclease HepT-like superfamily
IPR038476	UvrC, RNAse H endonuclease domain superfamily

Appendix C

Contribution of domain structure to the function of the yeast DEDD family exoribonuclease and RNase T functional homolog, Rex1

Contribution of domain structure to the function of the yeast DEDD family exoribonuclease and RNase T functional homolog, Rex1

PETER W. DANIELS,[1] TAIB HAMA SOOR,[1] QUENTIN LEVICKY, EWALD H. HETTEMA, and PHIL MITCHELL

Department of Molecular Biology and Biotechnology, The University of Sheffield, S10 2TN Sheffield, United Kingdom

ABSTRACT

The 3′ exonucleolytic processing of stable RNAs is conserved throughout biology. Yeast strains lacking the exoribonuclease Rex1 are defective in the 3′ processing of stable RNAs, including 5S rRNA and tRNA. The equivalent RNA processing steps in *Escherichia coli* are carried out by RNase T. Rex1 is larger than RNase T, the catalytic DEDD domain being embedded within uncharacterized amino- and carboxy-terminal regions. Here we report that both amino- and carboxy-terminal regions of Rex1 are essential for its function, as shown by genetic analyses and 5S rRNA profiling. Full-length Rex1, but not mutants lacking amino- or carboxy-terminal regions, accurately processed a 3′ extended 5S rRNA substrate. Crosslinking analyses showed that both amino- and carboxy-terminal regions of Rex1 directly contact RNA in vivo. Sequence homology searches identified YFE9 in *Schizosaccharomyces pombe* and SDN5 in *Arabidopsis thaliana* as closely related proteins to Rex1. In addition to the DEDD domain, these proteins share a domain, referred to as the RYS (Rex1, YFE9 and SDN5) domain, that includes elements of both the amino- and caroxy-terminal flanking regions. We also characterize a nuclear localization signal in the amino-terminal region of Rex1. These studies reveal a novel dual domain structure at the core of Rex1-related ribonucleases, wherein the catalytic DEDD domain and the RYS domain are aligned such that they both contact the bound substrate. The domain organization of Rex1 is distinct from that of other previously characterized DEDD family nucleases and expands the known repertoire of structures for this fundamental family of RNA processing enzymes.

Keywords: RNA processing; nuclease; 5S rRNA; yeast

INTRODUCTION

Stable RNAs are processed from initial transcripts by the action of one or more ribonucleases, which either cleave the phosphodiester backbone endonucleolytically or shorten the RNA exonucleolytically from the 5′- or 3′ end. The 3′–5′ exonucleolytic trimming of extended transcripts to mature stable RNAs is universally observed throughout biology. The genomes of most organisms encode a number of 3′–5′ exoribonucleases and, in the cellular systems that have been well characterized, these have been shown to have overlapping, redundant substrate specificities, and cellular functions (Deutscher 2006; Houseley and Tollervey 2009).

Eight distinct 3′ exoribonucleases have been characterized in *Escherichia coli* (Zuo and Deutscher 2001; Bechhofer and Deutscher 2019). Despite the strong functional overlap of these enzymes, there is a unique requirement for one enzyme, RNase T, in the final 3′ end trimming of some stable RNAs. Mutants lacking RNase T accumulate a major form of 5S rRNA that is extended at its 3′ end by 2 nt (Li and Deutscher 1995), while 3′ extended forms of 23S rRNA are observed that are predominantly one or 3 nt longer than the species observed in wild-type cells (Li et al. 1999). The 3′ ends of 5S and 23S rRNA share a common structural feature, where base-pairing between the 5′ and 3′ ends of the RNA generate a double-stranded region with short 3′ extensions of one or two U residues, respectively. The requirement for RNase T in the final maturation of 5S and 23S rRNA reflects its unique ability to remove nucleotides close to regions of double-stranded RNA (Zuo and Deutscher 2002c). Notably, RNase T was initially identified as the enzyme necessary for the removal of the 3′ terminal adenosine residue of deacylated tRNA (Deutscher et al. 1985) that also has a terminal stem with a short 3′ extension.

[1]These authors contributed equally to this work.

Corresponding author: p.j.mitchell@sheffield.ac.uk

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.078939.121. Freely available online through the *RNA* Open Access option.

RNase T is a member of the DEDD family of 3′ exonucleases (Zuo and Deutscher 2001). The catalytic domains of these enzymes contain four conserved acidic residues (hence the term, DEDD domain) within three exonuclease motifs that coordinate binding of two Mg^{2+} cations at the active site. Mutational analyses of ribonucleases within this family have shown that each of these four residues is required for catalytic activity (Zuo and Deutscher 2002a; Phillips and Butler 2003). The DEDD family also includes the 3′ exonuclease proof-reading domains of DNA polymerase I and the ε subunit of DNA polymerase III, and the catalytic domain of DNA exonuclease I (Moser et al. 1997). It is not clear what features of the catalytic domain enable DEDD family enzymes to act on RNA or DNA but several, including RNase T, can degrade both.

Members of the DEDD family of exonucleases have distinct structural features that extend the substrate binding path of the catalytic domain. RNase T is a homodimeric protein, each subunit containing a cluster of conserved basic residues within its DEDD domain that forms a nucleotide binding site (NBS). The NBS of one monomer is positioned adjacent to the catalytic site of the other monomer, providing an extended binding site for the 3′ end of the RNA (Zuo and Deutscher 2002b; Zuo et al. 2007). In contrast, RNase D is a monomer that contains two α-helical HRDC domains in addition to the DEDD catalytic domain. The DEDD and HRDC domains of RNase D are arranged in a funnel structure that has been proposed to contribute to substrate specificity (Zuo et al. 2005).

The genome of the budding yeast *Saccharomyces cerevisiae* encodes multiple DEDD family exoribonucleases that have known functions in RNA processing and/or degradation. The best characterized of these proteins is Rrp6, a catalytic subunit of the nuclear exosome RNase complex (Allmang et al. 1999; Zinder and Lima 2017). The amino-terminal PMC2NT domain of Rrp6 forms a stable heterodimeric complex with the small basic protein Rrp47 (Stead et al. 2007; Schuch et al. 2014) and genetic depletion of either Rrp6 or Rrp47 causes a decreased level of expression of the other protein (Feigenbutz et al. 2013; Stuparevic et al. 2013). A DEDD domain is also observed in members of the yeast RNA exoribonuclease (Rex) family of proteins that includes Rex1. Rex1 (also known as Rna82 or Rnh70) is a nuclear protein (Frank et al. 1999) required for the final 3′ end maturation of 5S rRNA and is implicated in the maturation of the 3′ end of 25S rRNA, the 3′ processing of tRNAs and the renewal of the CCA terminal sequence of tRNA (Piper et al. 1983; Kempers-Veenstra et al. 1986; van Hoof et al. 2000; Copela et al. 2008; Ozanick et al. 2009). Yeast *rex1* mutants accumulate forms of 5S rRNA that are extended at their 3′ end by 2 or 3 nt (Piper et al. 1983; van Hoof et al. 2000). Rex1 therefore carries out a similar set of RNA processing reactions to those mediated by RNase T in *E. coli*. However, the two proteins are structurally distinct; the DEDD domain of yeast Rex1 is flanked by extended amino- and carboxy-terminal regions and lacks the residues that comprise the hydrophobic interaction surface or the NBS of RNase T. None of the genes encoding a DEDD family exoribonuclease in yeast is essential for cell growth. However, null alleles of the *REX1* gene are synthetic lethal with deletions of the genes encoding Rrp6 or its associated protein Rrp47 (Peng et al. 2003; van Hoof et al. 2000).

To identify features of Rex1 that are required for its function in vivo, we generated a set of *rex1* mutants and analyzed the effect on cell growth in *rrp47Δ* mutants. The *rex1* mutants were also screened for their impact on protein expression and localization, 5S rRNA processing, Rex1 RNA binding activity in vivo and exonuclease activity in vitro. These analyses reveal a requirement for both the amino- and carboxy-terminal flanking regions of Rex1 for stable expression of the DEDD domain and for substrate binding, and characterize a bipartite nuclear localization signal close to the amino terminus of the protein.

RESULTS AND DISCUSSION

Rex1 function is dependent upon its amino- and carboxy-terminal regions, as well as specific features in the DEDD domain

The four Rex proteins from *S. cerevisiae* (Rex1, Rex2, Rex3, and Rex4) vary in size from 31–63 kDa, each containing a DEDD domain of approximately 160 amino acids as the only protein domain annotated in the *Saccharomyces* Genome Database (https://www.yeastgenome.org [accessed, 7-19-21]) (Fig. 1A). Rex1 is the largest of these proteins, its catalytic domain being flanked by comparably sized amino- and carboxy-terminal regions. The unique ability of Rex1 among the yeast Rex proteins to complement for the lack of Rrp6 (van Hoof et al. 2000) may in principle be due to functions associated with the amino- and/or carboxy-terminal regions of the protein or due to specific features of its catalytic domain. To address this, we first generated domain swap mutants that expressed Rex1 variants where the DEDD domain had been substituted for that from either the distantly related Rex2 or the very closely related Rex3 protein (Gerstberger et al. 2017). To assess the effect of mutations on Rex1 expression levels, wild-type and mutant Rex1 proteins were expressed as amino-terminal fusion proteins that harbor two copies of the z domain of protein A from *Staphylococcus aureus* (denoted zz). We also generated yeast strains expressing *rex1-TAP* or *rex1-HTP* (His-TEV-Protein A) alleles from the chromosomal *REX1* locus that bear the same duplicated z domain epitope (Puig et al. 1998; Granneman et al. 2009). The *rex1-TAP* and *rex1-HTP* alleles supported growth of an *rrp6Δ* mutant and did not exhibit the 5S rRNA processing phenotype characteristic of *rex1Δ* mutants, providing indicators of

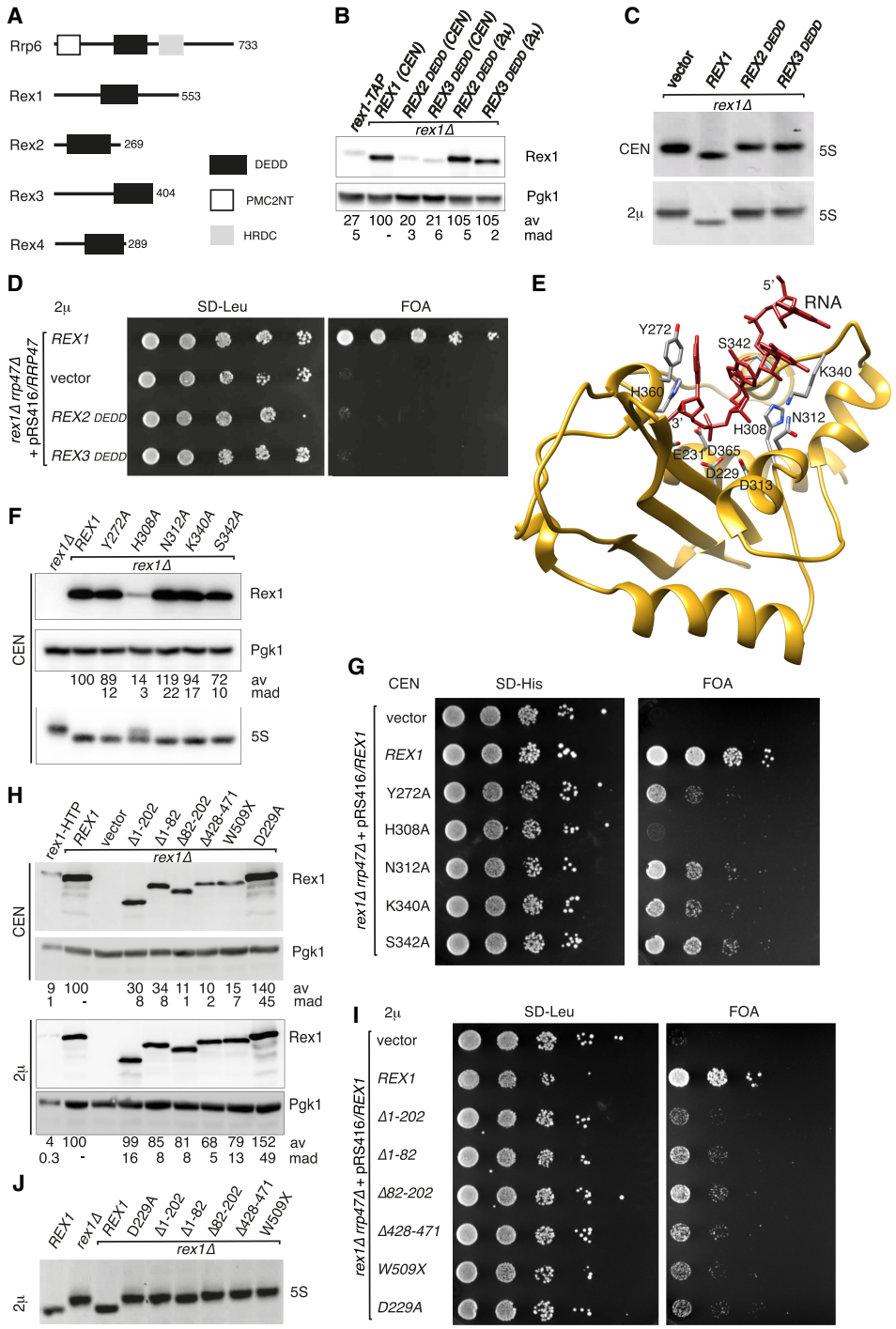


FIGURE 1. Rex1 function requires both the DEDD domain and flanking regions. (*A*) Domain organization of the yeast Rex enzymes and Rrp6. Protein lengths (amino acid residues) are indicated. (*B*) Western analysis of Rex1 fusion proteins. Rex1 was expressed either from the integrated *rex1-TAP* allele or from plasmids encoding the wild-type protein or domain swap mutants containing the DEDD domain from either Rex2 or Rex3. (*C*) Ethidium-stained RNA gels showing the relative mobility of 5S rRNA species from a *rex1Δ* mutant expressing either wild-type Rex1 or the domain swap mutants. (*D*) Plasmid shuffle assay comparing wild-type and *rex1* domain swap mutants. (*E*) Threaded model of the DEDD domain of Rex1 (residues 227–372) with bound oligoribonucleotide, derived from the Pan2/RNA structure (PDB 6R9M). Side chains of residues implicated in catalysis or substrate binding are shown. (*F*) Western analysis and 5S rRNA northern analysis of wild-type and *rex1* point mutants. (*G*) Plasmid shuffle assay of the *rex1* point mutants. (*H*) Western analysis of wild-type and *rex1* deletion mutants. (*I*) Plasmid shuffle assay of the *rex1* deletion mutants. (*J*) Northern analysis of 5S rRNA from the *rex1* deletion mutants. Expression from low copy number, centromeric (CEN) or high copy number 2 micron (2μ) plasmids is indicated in each panel. Expression levels of mutant Rex1 proteins, normalized to Pgk1 and expressed as a percentage of the wild-type protein, are indicated *beneath* each lane. Values given are the mean average (av) and the mean absolute deviation (mad) of three independent replicates.

Rex1 expression levels at or above a functional activity threshold.

The observed levels of the chimeric Rex1 fusion proteins were lower than the wild-type Rex1 protein when expressed from centromeric plasmids in a *rex1Δ* strain but comparable to the Rex1 expression levels from a related *rex1-TAP* strain (Fig. 1B). Expression of the domain swap mutants in the *rex1Δ* strain failed to complement the 5S rRNA processing phenotype (Fig. 1C). To address whether this was due to the lower expression levels of the mutant proteins relative to the wild-type protein, we also expressed the wild-type and domain swap proteins from a high copy number, 2 micron plasmid. Using this vector, steady state expression levels of the chimeric domain swap proteins were about fourfold higher than the Rex1-TAP protein and close to that of the plasmid-encoded wild-type protein (Fig. 1B). However, the mutants still failed to complement the 5S rRNA processing phenotype (Fig. 1C). Furthermore, while expression of the wild-type *REX1* control supported growth of the *rex1Δ rrp47Δ* double mutant, transformants expressing either chimeric protein from the high copy number plasmid failed to grow (Fig. 1D). These data suggest that specific structural features of the Rex1 DEDD domain are required for the overall structural folding or stability of the protein and/or may contribute to substrate binding.

Threading of the yeast Rex1 DEDD domain sequence (residues 227–382) to the available structure of Pan2 in association with bound oligonucleotide (Tang et al. 2019) provided an excellent fit, with a RMSD for the Cα atoms of 0.75 Å. Residues involved in RNA binding in the Pan2 structure could therefore be easily mapped within the threaded Rex1 model (Fig. 1E). The aromatic ring of Y975 in yeast Pan2 stacks onto the base of the terminal 3′ nt bound in the active site of the enzyme. A tyrosine residue is also found at the equivalent position in Rex1 from *S. cerevisiae* (Y272) and in the functionally homologous Rexo5 from *D. melanogaster*, while a conservative substitution to phenylalanine is seen in human Rexo5, yeast Rex3, and RNase T. Pan2 makes several contacts with the phosphodiester backbone of the bound substrate. Side chain interactions are mediated by N1019 and S1048 of Pan2 (Tang et al. 2019), both residues of which are also found at the equivalent positions in yeast Rex1 (N312 and S342, respectively). Y1046 in Pan2 makes main chain interactions with the ribose phosphate backbone of the bound substrate, with the side chain being directed away from the RNA. The side chain of the equivalent residue in Rex1, K340, is directed toward the bound substrate in the threaded model.

To address a potential requirement for these residues in Rex1 function, alanine substitutions were generated and the mutants were assayed for complementation of the *rex1Δ rrp47Δ* growth phenotype and the 5S rRNA processing defect of the *rex1Δ* mutant. Substitution of residues Y272, N312, K340, and S342 had a minor effect on the expression of Rex1 and the corresponding mutants exhibited no 5S rRNA processing phenotype (Fig. 1F). However, these substitutions supported growth of the *rex1Δ rrp47Δ* double mutant much less effectively than the wild-type protein (Fig. 1G). Given the contacts between equivalent residues in Pan2 and bound RNA, the slow growth rate of the Y272A, N312A, and S342A *rex1 rrp47Δ* double mutants most probably reflects defects in substrate binding. We anticipate that these residues act synergistically in substrate binding, and that loss of individual contacts is not sufficient to block substrate binding completely.

We also generated an alanine substitution of H308 within the Exo II motif, the equivalent residue in mouse PARN (H280) and SDN1 from *A. thaliana* (H236) being implicated in substrate interaction (Wu et al. 2009; Chen et al. 2018). The H308A mutation caused a strong reduction in Rex1 expression levels (Fig. 1F) and a correspondingly strong impact on growth in the plasmid shuffle assay (Fig. 1G). Nevertheless, 5S rRNA processing was only partially affected in this mutant (Fig. 1F). These two observations are not inconsistent; the synthetic lethality of the *rex1Δ rrp47Δ* mutant is not due to defective 5S rRNA processing, as the block in 5S rRNA processing observed in the *rex1Δ* mutant is not exacerbated in a conditional *rex1Δ rrp47* double mutant (Garland et al. 2013). One possibility is that 5S rRNA is more readily processed by Rex1 than some of its other substrates. Alternatively, partial rather than complete processing of a Rex1 substrate other than 5S rRNA may be growth limiting in the absence of Rrp47.

To address whether regions of Rex1 other than the catalytic domain are required for its function, we generated a series of amino- and carboxy-terminal deletion mutants. We constructed three amino-terminal deletion mutants; a large deletion that essentially lacks all the polypeptide sequence upstream of the catalytic domain (Δ1–202) and two smaller deletions thereof (Δ1–82 and Δ82–202). Deletion of the carboxy-terminal region of Rex1 was predicted to cause loss of function, since allele linkage and sequence analysis of the loss of function *rna82-1* allele (Piper et al. 1983) revealed it to be congenic with *REX1* and to have a premature termination codon (W433X) downstream from the catalytic domain (van Hoof et al. 2000). We generated two carboxy-terminal *rex1* mutants; one is a truncation that lacks the final 45 amino acid residues (W509X) and the other is a deletion downstream from the catalytic domain (Δ428–471). We also mutated one of the four conserved acidic residues within the DEDD domain to an alanine (D229A). This mutation was predicted to block catalytic activity, based on previous studies on the equivalent mutation in RNase T, PARN, and Rrp6 (Ren et al. 2002; Zuo and Deutscher 2002a; Phillips and Butler 2003) and the analysis of the D229A, E231A double mutant in Rex1 (Ozanick et al. 2009).

The steady state expression levels of the Δ82–202, Δ428–471, and W509X *rex1* mutants were markedly lower

than the wild-type protein when expressed from a centromeric plasmid, yet nonetheless comparable to that of the integrated *rex1* allele (Fig. 1H). To ensure that observed effects were not due to limiting expression, the mutants were also expressed from a high copy number vector. Expression levels of all mutants from this vector were comparable to that of the wild-type protein and notably higher than expression from the *rex1-HTP* allele (Fig. 1H, lower panel). Strikingly, all the *rex1* deletion mutants gave rise to only very small colonies in the plasmid shuffle assay and failed to complement the 5S rRNA processing defect of the *rex1Δ* mutant when expressed from either the centromeric plasmid or the high copy number plasmid (Fig. 1I,J). Since only very low expression levels of Rex1 are sufficient to facilitate maturation of at least some 5S rRNA in vivo (Fig. 1F), we infer that all the *rex1* deletion mutants are severely compromised in their ability to process 5S rRNA. Taken together, the data show that the structural folding and/or function of Rex1 is dependent upon features of both the amino-terminal and carboxy-terminal regions of the protein, as well as specific characteristics of its catalytic domain. As predicted from previous studies, the D229A *rex1* mutant failed to complement the 5S rRNA processing phenotype of the *rex1Δ* single mutant (Fig. 1J).

Residues 17–52 of Rex1 constitute an NLS

To determine whether the deletion mutants affect the nuclear localization of Rex1, we generated plasmid constructs encoding wild-type and mutant GFP-tagged Rex1 fusion proteins and compared their subcellular distribution by fluorescence microscopy. Western analysis showed that the wild-type GFP–Rex1 protein was stably expressed (Fig. 2A) and its expression in a *rex1Δ* mutant complemented the 5S rRNA processing phenotype (Fig. 2B). Furthermore, this construct genetically complemented the growth phenotype of the *rex1Δ rrp47Δ* double mutant (Fig. 2C). Consistent with the nuclear localization of Rex1 (Frank et al. 1999), GFP–Rex1 localized to a region coincident with the distribution of the nucleoporin RFP–Nic96 (Fig. 2D; Grandi et al. 1993). The carboxy-terminal Δ428–471 and W509X mutants also showed a localized subcellular distribution, while the amino-terminal Δ1–202 deletion gave a diffuse, nonlocalized signal (Fig. 2E). Further analysis of the amino-terminal mutants revealed that the Δ1–82 mutant showed a nonlocalized distribution throughout the cell, while the distribution of the Δ82–202 mutant was restricted in a similar manner to the wild-type protein. This data suggests that the first 82 residues of Rex1 contain a signal that mediates the nuclear localization of Rex1.

NLS prediction webserver tools indicated a potential lysine-rich NLS spanning residues 42–50 and a longer, overlapping bipartite signal between residues 17–51 (Kosugi et al. 2009; Nguyen Ba et al. 2009). We generated GFP fusion *rex1* deletion mutants that lacked residues 42–52 or 17–52 and analyzed their subcellular distribution. The Δ42-52 mutant showed a normal nuclear distribution, while the larger Δ17-52 mutant showed a nonlocalized signal throughout the cell. Addition of the NLS from the large T antigen of SV40 virus to the Δ17–52 mutant restored nuclear localization of the protein (Fig. 2E). Furthermore, addition of residues 17–52 of Rex1 to the carboxyl terminus of GFP was sufficient to direct localization of the fusion protein to the nucleus, as shown by its colocalization with Nic96 (Fig. 2F). Western analyses of GFP–Rex1, the Δ17–52 deletion mutant and the GFP–NLS fusion protein demonstrated that the localized proteins were expressed as full-length, intact polypeptides (Fig. 2A). These data demonstrate that residues 17–52 of Rex1 constitute an NLS. The amino-terminal region of the mouse and human proteins has previously been shown to harbor an NLS (Silva et al. 2017).

Notably, the Δ17–52 mutant supported growth of the *rex1Δ rrp47Δ* double mutant strain (Fig. 2C). These data suggest that, in the absence of the identified NLS, sufficient Rex1 can be localized to the nucleus through one or more additional mechanisms to support cell growth in the absence of Rrp47. A similar observation was previously made for the *rrp6-15* mutant that lacks a functional NLS and causes delocalization of the Rrp6 protein but nevertheless complements the temperature-sensitive growth phenotype of an *rrp6Δ* mutant (Phillips and Butler 2003).

Rex1 is both phosphorylated and ubiquitylated in vivo. Strikingly, the sites of phosphorylation (S24, T26, S27, and T34) are clustered within the NLS, while a nearby lysine residue (K58) is a known site of ubiquitylation (Albuquerque et al. 2008; Holt et al. 2009; Swaney et al. 2013). Phosphorylation at S24 is reduced during the environmental stress response (MacGilvray et al. 2020) and Rex1 activity has been reported to be limiting for 3′ end processing of specific tRNAs during growth at high temperature and upon nutritional deprivation (Foretek et al. 2016). This suggests Rex1 activity may be regulated in part through post-translational modifications that alter its nucleocytoplasmic distribution, in a manner similar to the RNA polymerase III repressor Maf1 (Oficjalska-Pham et al. 2006; Roberts et al. 2006).

The amino- and carboxy-terminal regions of Rex1 are required for 5S rRNA processing in vitro

To address whether the mutant Rex1 ribonucleases are compromised in their ability to accurately process RNA, Rex1 GST fusion proteins were expressed in *E. coli*, purified by affinity chromatography and incubated with 5S RNP complexes isolated from a *rex1Δ* mutant. The substrate and reaction products were then resolved through denaturing acrylamide/urea gels, together with total cellular RNA from a wild-type strain to serve as a marker for mature 5S RNA.

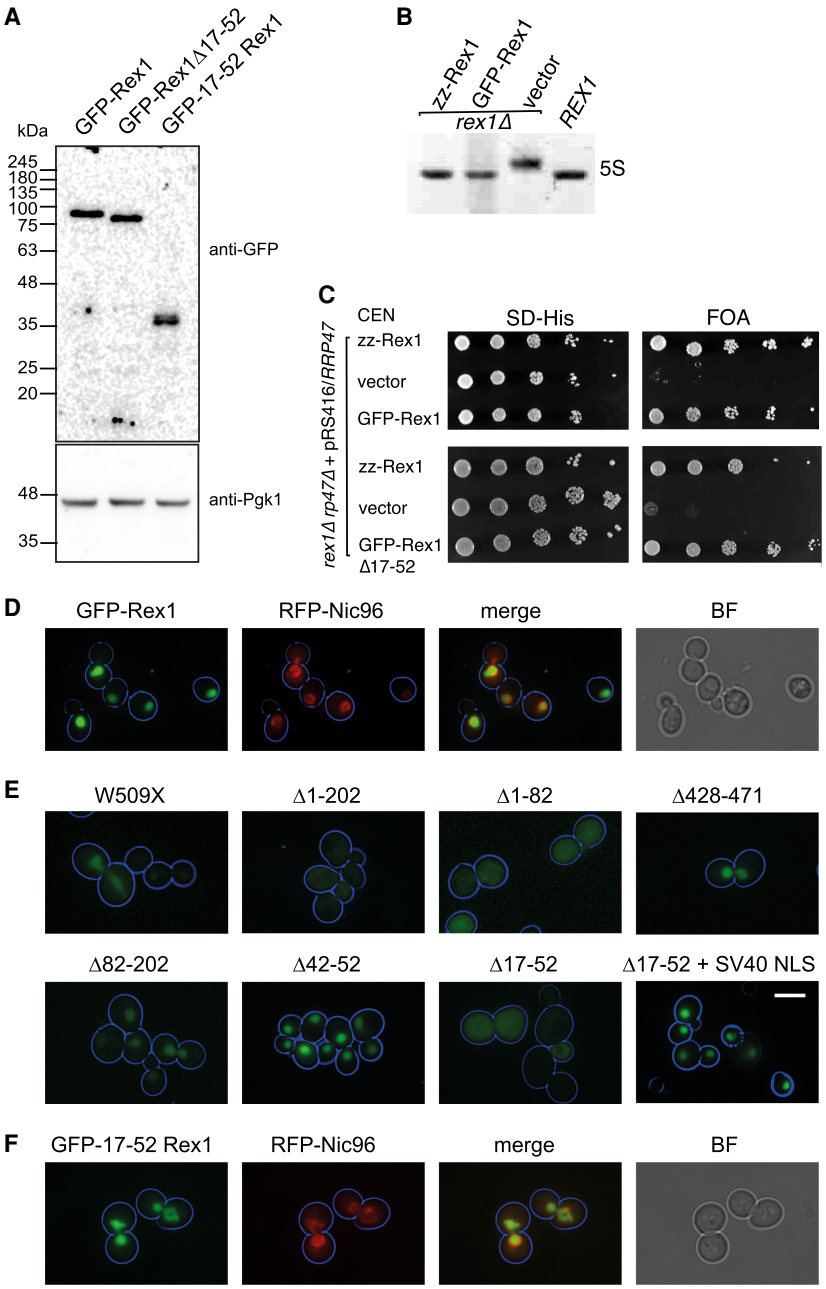


FIGURE 2. Identification of a nuclear localization signal in Rex1. (*A*) Western analysis of whole-cell extracts from a *rex1Δ* mutant expressing GFP–Rex1 fusion proteins. (*B*) Ethidium-stained RNA gel showing the relative migration of 5S rRNA species from a *REX1* wild-type strain and *rex1Δ* mutants expressing either Rex1 zz or GFP fusions. (*C*) Plasmid shuffle assays on *rex1Δ rrp47Δ* double mutants expressing Rex1 GFP and zz fusion proteins. (*D–F*) Epifluorescence microscopy images of cells expressing versions of GFP–Rex1 and mRFP–Nic96. (*D*) Colocalization of GFP–Rex1 and mRFP–Nic96. (*E*) Localization of GFP–Rex1 deletion mutants. (*F*) Colocalization of GFP fused to residues 17–52 of Rex1 and mRFP–Nic96. Cell circumferences are indicated in blue. Bar, 5 µm. (BF) Brightfield, (SV40-NLS) simian vacuolating virus 40 nuclear localization signal.

Upon incubation with the wild-type Rex1 protein, the majority of the 3′ extended 5S rRNA was converted to a product of the same size as the mature 5S rRNA by the first time point (Fig. 3A). In contrast, very little of the substrate was shortened upon incubation with either the Δ1–203 mutant or the W509X mutant throughout the whole time course. The same observations were made upon assaying Rex1 wild-type and mutant zz fusion proteins purified from yeast extracts. These data show that Rex1 exhibits exoribonuclease activity when expressed in *E. coli* and that its ability to process a physiologically relevant substrate, 3′ extended 5S rRNA, is dependent upon both the amino- and carboxy-terminal regions of Rex1.

We also assayed the nuclease activity of recombinant Rex1 proteins using a DNA oligonucleotide substrate that is predicted to have negligible secondary structure. Wild-type Rex1 exhibited DNase activity in vitro, while the D229A mutant had only a very low activity similar to that seen for the analogous active site mutant of RNase T (Fig. 3B; Zuo and Deutscher 2002a). The Rex1 DNase activity was less processive than the 5S RNP processing activity, with discrete intermediates detected that are progressively shortened during the time course. The Δ1–203 and W509X mutants showed some catalytic activity upon incubation with DNA but the substrate was shortened by both mutants at a slower rate than that seen upon incubation with the wild-type protein.

Rex1 makes multiple contacts with RNA

To address whether the amino- or carboxy-terminal region of Rex1 contributes to substrate binding, we compared the yield of covalent Rex1/substrate complexes generated upon UV irradiation of intact, growing cells expressing wild-type or mutant Rex1 fusion proteins. Cross-linked protein/RNA complexes were visualized by protein purification from cell lysates under stringent conditions, on-bead RNase digestion and 5′ ^{32}P labeling of RNP complexes, followed by western blotting and phosphorImaging of the eluted proteins. The Rex1 mutants were

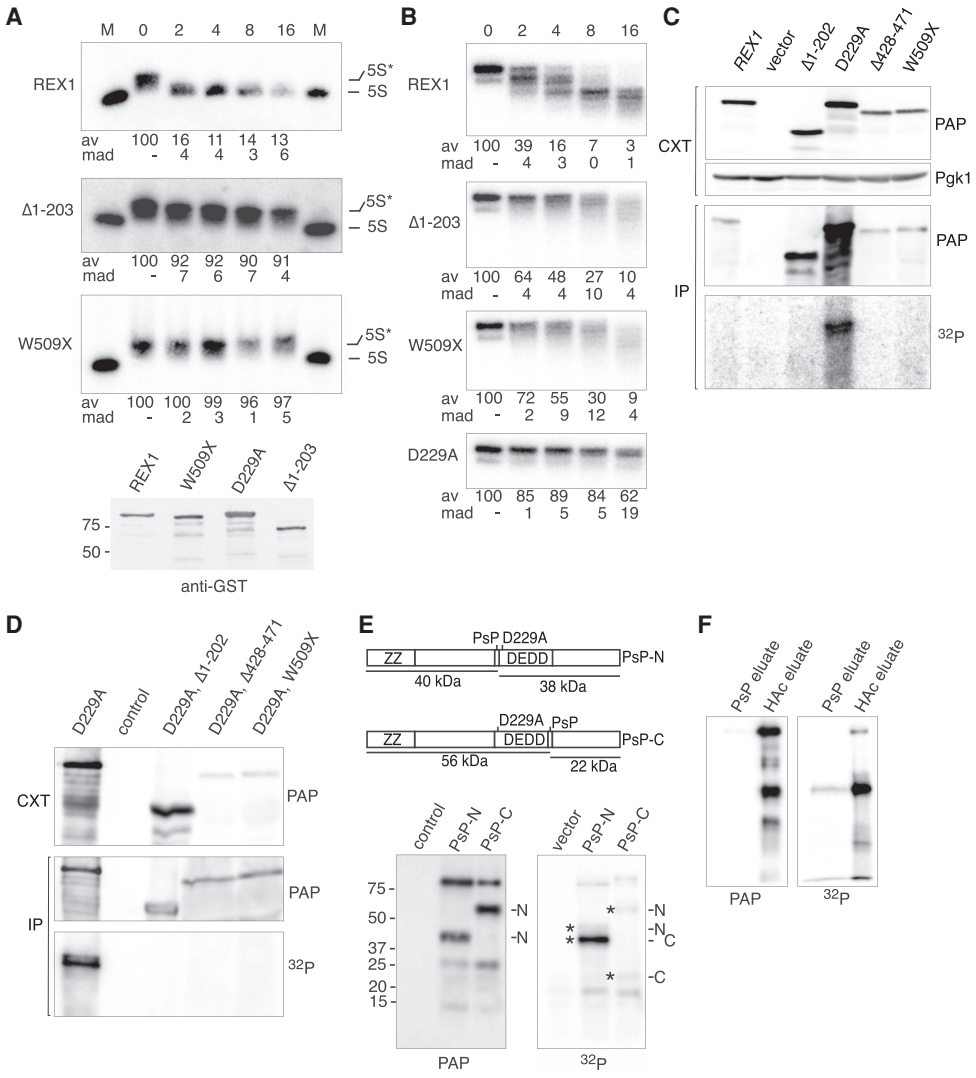


FIGURE 3. The amino- and carboxy-terminal regions of Rex1 are required for RNA binding. (*A*,*B*) In vitro nuclease assays. (*A*) 5S rRNA processing assays. (*B*) Assays using a 5′ ^{32}P-labeled DNA oligonucleotide. Time points (minutes) are indicated *above* each panel. Lanes labeled M contain cellular RNA from a wild-type strain. The amount of full-length substrate (as a mean average [av] percentage of the total) remaining at each time-point is indicated *below*, together with the mean absolute deviation (mad) values ($n \geq 2$). The *lower* panel shows a western analysis of the purified proteins assayed. (*C–F*) In vivo RNA crosslinking analyses. Analyses are shown of whole-cell extracts (CXT) and purified proteins (IP). Westerns show zz-tagged proteins (labeled PAP) or the Pgk1 protein. PhosphoImaged blots are labeled ^{32}P. Control samples are from a strain expressing nontagged Rex1. (*C*) Analysis of wild-type and mutant Rex1 proteins. (*D*) Analysis of deletion mutants bearing the active site mutation. (*E*) Proteolytic cleavage of Rex1/RNA complexes. PsP cleavage sites and the predicted fragments are indicated. Analyses shown are of the acid eluates after proteolysis. Asterisks denote cleavage products. (*F*) comparison of PsP eluates and acetic acid (HAc) eluates of the PsP-N construct. A fivefold relative excess of the PsP eluate over the acid eluate was analyzed.

expressed from multicopy plasmids to ensure expression levels comparable to that of the wild-type protein.

All fusion proteins were expressed comparably and readily recovered upon purification from lysed, irradiated cells, but crosslinking to RNA was only consistently observed for the D229A mutant (Fig. 3C). This suggests that inhibition of catalysis is required for efficient crosslinking under the experimental conditions used. We therefore introduced the D229A mutation into the Δ1–202, Δ428–471, and W509X deletion mutants and assayed the resulting double mutants, together with the D229A single mutant. Comparable amounts of each fusion protein were recovered upon purification from lysates of irradiated cells, but crosslinking to RNA was observed only for the full-length D229A mutant (Fig. 3D). We conclude from these data that both the amino- and carboxy-terminal regions of Rex1 are required, either directly or indirectly, for substrate binding in vivo.

To determine whether direct contacts are made between the amino- and carboxy-terminal regions of Rex1 and bound substrate, crosslinking experiments were

performed on cells expressing amino-terminal zz fusions of Rex1 containing an internal PreScission protease (PsP) cleavage site either at the amino- or carboxy-terminal end of the DEDD domain (Fig. 3E). Crosslinked RNP complexes were treated as above and digested with PsP prior to acid elution from the beads.

Western analyses of the acid eluates revealed bands corresponding to the amino-terminal 40 and 56 kDa epitope-tagged cleavage products (labeled N in the PAP blot shown in Fig. 3E), along with residual, nondigested full-length Rex1. PhosphorImager analyses of the same blots revealed pairs of ^{32}P-labeled bands corresponding to both the amino- and carboxy-terminal polypeptide fragments of each construct (denoted with asterisks and labeled N or C). The electrophoretic migration of the ^{32}P-labeled PsP-N digestion products was slightly retarded in the SDS-PAGE gels, relative to the western signal from the same fragment, due to the additional mass of the crosslinked oligonucleotide. Direct comparison of the PsP-N digestion products eluted upon enzymatic cleavage with the acid eluates (Fig. 3F) confirmed that the more intensely ^{32}P-labeled, faster migrating PsP-N product was the non-tagged, carboxy-terminal fragment. Only a small amount of the carboxy-terminal cleavage product was solubilized upon PsP digestion. This fragment was retained on the beads after stringent washing with buffer containing 2 M $MgCl_2$, which readily dissociates stable multimeric protein complexes (Allmang et al. 1999). These observations suggest that the amino-terminal flanking region of Rex1 has very stable intramolecular interactions with either the DEDD domain or carboxy-terminal flanking region, rather than comprising a structurally distinct, modular domain. Taken together, the data show that Rex1 makes direct contacts with its RNA substrates within both the amino- terminal and carboxy-terminal regions of the protein in vivo and are consistent with loss of 5S rRNA processing activity observed in the deletion mutants due to a decrease in substrate binding.

We consistently observed a higher level of ^{32}P associated with the PsP-N carboxy-terminal cleavage product than would be expected, based on the observed efficiency of PsP cleavage. Preliminary studies showed that digestion using a range of buffer and incubation conditions under which PsP is active (Ullah et al. 2016) did not allow complete digestion, suggesting that a fraction of the purified PsP-N fusion protein is not amenable to cleavage. We interpret these observations to suggest that the fraction of crosslinked protein that is poorly accessible for digestion by PsP is also a poor substrate for polynucleotide kinase.

Rex1 amino- and carboxy-terminal flanking sequences constitute a composite domain

The AlphaFold Protein Structure Database (Jumper et al. 2021) was released during preparation of this report. The predicted structure for Rex1 consists of the catalytic DEDD domain and an adjacent domain that we refer to here as the RYS (Rex1, YFE9, and SDN5) domain (Fig. 4A). The catalytic center lies on the internal surface of the DEDD domain. A striking feature of the model is an extended α-helical arch within the RYS domain that is directed toward the DEDD domain. The model has a very high confidence score (pLDDT mean score of 91.6% for residues 53–553), with only the amino-terminal region and two loops (residues 203–209 and 462–469) that have low confidence scores. The largely unstructured amino-terminal 52 residues of Rex1 are not shown in Figure 4. The two domains are aligned by interactions involving D337, which makes both a main chain interaction with N161 and a side chain interaction with K420. The domain interface is further constrained through an interaction between K340 and N197. Whether K340 is involved in both intramolecular interactions and substrate binding is currently unclear.

Two terminal nucleotides linked by the scissile phosphate were mapped to the active site by superposition of the threaded structure of the Rex1 DEDD domain with associated substrate onto the AlphaFold model (Fig. 4A). The two models aligned well, with an RMSD of 2.0 Å for the complete length of the threaded DEDD domain. The orientation of the modeled nucleotides in the active site strongly suggests that the RNA substrate enters the active site from the front, as viewed in Figure 4. Mature 5S rRNA has a terminal stem with a single 3′ nt overhang. Molecular docking studies suggest that there is sufficient distance between the two domains in the model to accommodate base-paired nucleotides, consistent with the enzyme being able to trim the short 3′ overhang found in 5S rRNA. However, a base pair is too bulky to fit through the entrance to the active site (Fig. 4B). This strongly suggests that there is realignment of the two domains upon substrate binding. The electrostatic potential map of the model reveals an extended, positively charged surface on the front side of the protein that may function in substrate recruitment (Fig. 4B). In marked contrast, the rear side of the protein is largely negatively charged. We speculate that the two domains are positioned further apart than shown in the model when bound to the substrate such that the RNA can interact with the basic patch on the surface of the enzyme.

The eight-stranded, extended β-sheet and associated α-helices at the core of the RYS domain integrates polypeptide sequences from both the amino- and carboxy-terminal regions of Rex1 (Fig. 4A). It follows that the structural integrity of the RYS domain is predicted to be perturbed in the amino- and carboxy-terminal *rex1* deletion mutants. Two of the mutations that showed the greatest effect on Rex1 expression levels (Fig. 1H) are restricted to elements of the RYS domain; the Δ428–471 deletion spans a central β-strand and the flexible loop adjacent to the arch, while the W509X truncation removes the arch and the β-strand at the carboxyl terminus. Both mutations result in loss of

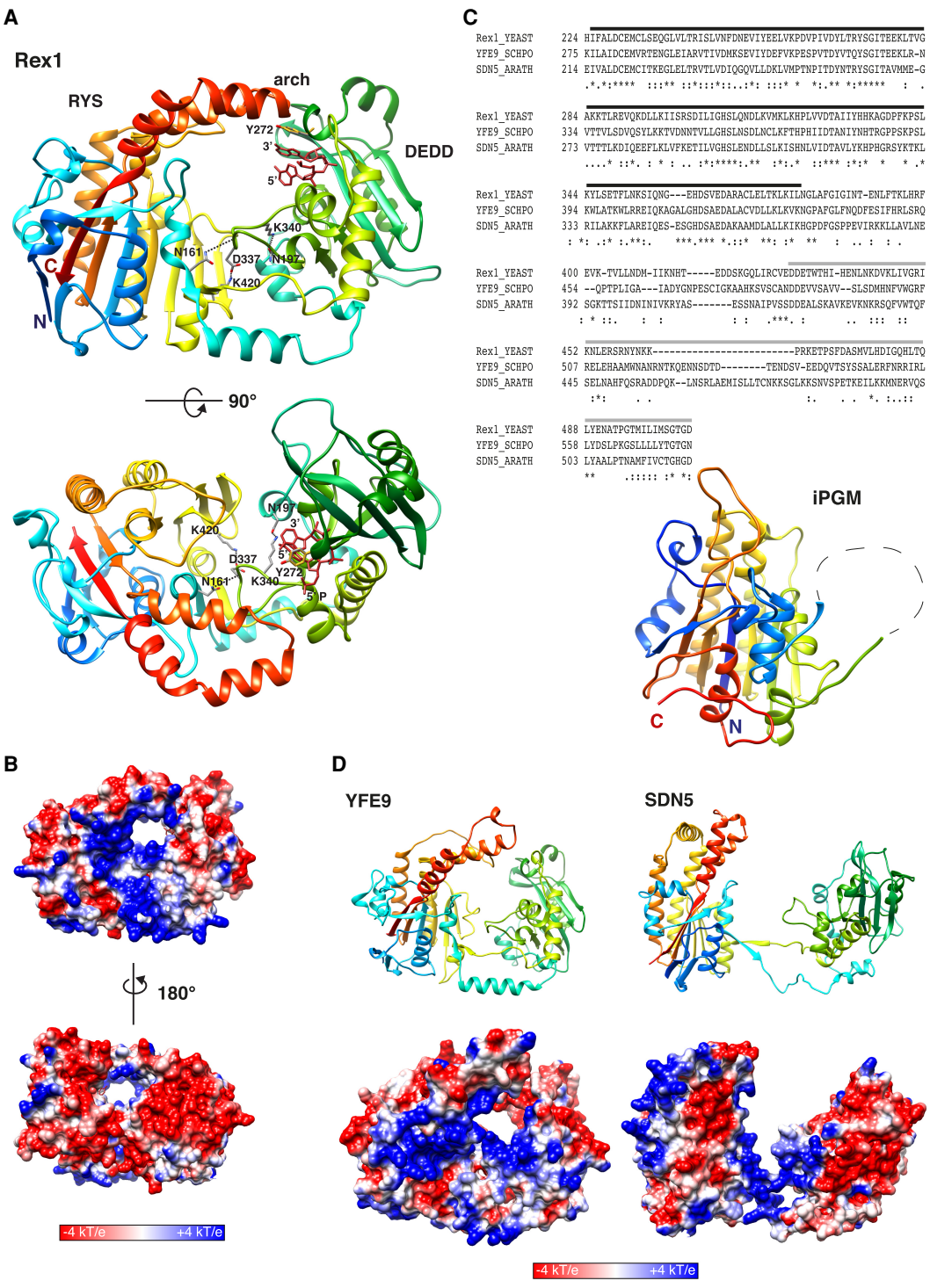


FIGURE 4. The Dual domain structure of Rex1. (*A*) Ribbon structure of the AlphaFold model of Rex1 (residues 53–553) with bound substrate. The terminal dinucleotide from the threaded Rex1/RNA model (Fig. 1E) was superposed onto the Rex1 AlphaFold model by manual alignment with residue Y272. The Y272 side chain is shown in both models (gray and gold) to indicate the alignment. Front and top views are shown. Coloring is from blue (amino terminus) to red (carboxyl terminus). The DEDD domain, the RYS domain and the helical arch are labeled. Interactions between specified residues are shown. The 3′ and 5′ nt are labeled. The 5′ phosphate group is labeled in the *lower* panel. (*B*) Electrostatic potential map of Rex1. Front and rear views are shown. (*C*) BLAST sequence alignment of Rex1, YFE9, and SDN5. The DEDD domain is overscored in black. iPGM- and PPM-related sequences are overscored in gray. The fold of the substrate binding domain of iPGM from *B. stearothermophilus* (residues 2–78 and 309–511) is shown. (*D*) Ribbon structures and electrostatic potential maps of (*left* panels) YFE9 (residues 87–631) and (*right* panels) SDN5 (residues 41–567). Amino-terminal sequences of the Rex1, YFE9, and SDN5 AlphaFold models with low confidence scores are excluded. Orientations of iPGM, YFE9, and SDN5 are shown as for Rex1 in *A*.

Rex1 function in vivo (Fig. 1I,J), while the W509X mutation causes a loss of 5S RNP processing activity in vitro (Fig. 3). This strongly suggests that the RYS domain is critical for expression of a catalytically active DEDD domain.

Rex1, YFE9, and SDN5 share a similar dual domain structure

BLAST searches for homologs of Rex1 revealed that YFE9 from *S. pombe* and SDN5 from *A. thaliana* show sequence homology that extends throughout the DEDD domain and substantially into the carboxy-terminal region (Fig. 4C). Phyre2 searches using Rex1, YFE9, and SDN5 aligned a homologous region within each protein (indicated with a gray bar in Fig. 4C) downstream from the DEDD domain to the three-dimensional structure of the substrate binding domain of cofactor-independent phosphoglycerate mutases (iPGMs) and phosphopentomutases (PPMs) (Jedrzejas et al. 2000; Panosian et al. 2011), two enzymes belonging to the alkaline phosphatase superfamily. iPGMs and PPMs have a similar domain organization to Rex1, consisting of a central domain and a composite domain comprising the amino- and carboxy-terminal polypeptide sequences. Strikingly, the overall fold of the composite substrate-binding domain of iPGM from *Bacillus stearothermophilus* (Fig. 4C) and PPM from *Bacillus cereus* is very similar to that predicted for the RYS domain of Rex1 but lacks the extended helical arch. Furthermore, AlphaFold models of YFE9 and SDN5 (pLDDT mean scores of 88.7% and 87.3%, respectively, for the residues shown) predict dual domain structures similar to Rex1 that include a carboxy-terminal helical arch connecting the equivalent strands within the central beta sheet of the composite domain (Fig. 4D). The YFE9 model has a positively charged surface around the predicted entrance to the active site that potentially offers an extended substrate binding surface, as in the case of yeast Rex1. There are no interactions between the DEDD and RYS domains in the AlphaFold model of SDN5 equivalent to those involving D337 in Rex1 or in YFE9, and hence the domains are not well aligned. We anticipate, however, that these domains are more closely aligned in the protein structure. Nevertheless, there is a localized surface positive charge in the SDN5 model that lines the base of the two domains. We conclude that Rex1, YFE9, and SDN5 share a common structural domain in addition to the DEDD domain, referred to here as the RYS (Rex1, YFE9, and SDN5) domain, the fold of which is related to the substrate binding domain of iPGMs and PPMs. Furthermore, the predicted structures of Rex1, YEF9, and SDN5 suggest the presence of a substrate binding site at the entrance to the catalytic center of the enzymes.

The data reported here analyze the function of a novel protein domain found within DEDD family ribonucleases. We report that the RYS domain is required for stable expression of Rex1, contributes to substrate interaction and is critical for RNA processing. The RYS and DEDD domains are predicted to cradle the binding site for the 3′ end of RNA. A conserved basic patch in Rex1-related enzymes suggests an extended interaction surface leading to the catalytic center. The dual domain structure of Rex1 and the homodimeric arrangement of RNase T therefore represent remarkably distinct structural architectures that facilitate the 3′ end processing of equivalent RNA substrates.

MATERIALS AND METHODS

Plasmids and strains

Yeast sequences were amplified by PCR on genomic DNA from BY4741-related strains and cloned using standard molecular biology procedures. Constructs and strains generated in this study are given in the Supplemental Data. All constructs were validated by sequencing. Centromeric and high copy number yeast expression vectors were derived from the pRS series (Sikorski and Hieter 1989; Christianson et al. 1992).

The wild-type *REX1* ORF and 3′-UTR were cloned into a construct that allows the expression of amino-terminal zz fusion proteins from the *RRP4* promoter (Mitchell et al. 1996). Site-directed mutagenesis reactions to introduce point mutations, restriction sites or cleavage sites were done using QuikChange kits (Agilent) or by PCR with divergent primer pairs using Q5 DNA polymerase (New England BioLabs). Internal deletions were created after introducing pairs of restriction sites. The *REX2* DEDD and *REX3* DEDD domain constructs were generated by homologous recombination in yeast; the zz fusion, wild-type *REX1* expression construct was linearized with *Xcm*I and cotransformed with PCR amplicons of the *REX2* and *REX3* DEDD domains that were generated using *REX1*/*REX2* and *REX1*/*REX3* splint primers. The DEDD catalytic domain of Rex1 (residues A227–L372) was replaced with that of Rex2 (V56-Q226) and Rex3 (S245-V388). DNA encoding GFP was amplified from plasmid pFA6a-GFP (S65T)-kanMX6 (Longtine et al. 1998) as an *Eco*RV-*Eco*RI fragment and substituted for the sequence encoding the zz tag after introduction of an *Eco*RV site at the second and third codon. To introduce the SV40 NLS, oligonucleotides encoding the peptide sequence IPKKKRKVD were blunt end ligated into the GFP–Rex1ΔNLS construct after digestion with *Eco*RV. The GFP–REX1–NLS construct was generated by ligating annealed, overlapping oligonucleotides encoding Rex1 residues 17–52 into the *Eco*RI site at the 3′ end of the GFP ORF. A peptide encoding the PreScission Protease site was inserted into a unique *Bgl*II site that had initially been engineered up- or downstream from the DEDD domain. The resulting constructs harbor the peptide sequence RSLEVLFQGPRS between residues H220/G221 and A366/R367 of Rex1. For expression of Rex1 proteins in *E. coli*, inserts from the yeast expression vectors were amplified and subcloned as *Bam*HI-*Eco*RI fragments into pGEX-6P1. The amino-terminal deletion construct generated (Δ1–203) expresses an in-frame fusion but lacks the *Bam*HI site. The *NIC96* ORF and 3′-UTR was amplified as an *Nsi*I-*Xho*I fragment and cloned into an mRFP expression construct (a kind gift from the Hurt laboratory, University of Heidelberg) (Ulbrich et al. 2009).

The *rex1Δ* strain was obtained from Euroscarf (University of Frankfurt). The *rex1Δ rrp47Δ* strain has been reported previously (Costello et al. 2011). The *rex1-TAP::URA3* and *rex1-HTP::URA3* strains were generated by standard homologous recombination, using amplicons derived from pBS1539 (Puig et al. 1998) or the HTP-tagging cassette (Granneman et al. 2009), respectively.

Yeast methods

Yeast strains were routinely grown in standard rich YPD medium (2% glucose, 1% yeast extract, 2% bacto-peptone) or selective SD medium (2% glucose, 0.5% ammonium sulfate, 0.17% yeast nitrogen base and appropriate amino acids and bases but lacking either histidine or leucine) at 30°C. Transformations were carried out using standard lithium acetate protocols. Complementation analyses were performed using plasmid shuffle assays on *rex1Δ rrp47Δ* strains that are supported by a plasmid bearing either the wild-type *RRP47* gene or encoding the zz-Rex1 fusion protein. These plasmids also carry the counter-selectable *URA3* gene. Plasmid shuffle assays were performed as previously described (Feigenbutz et al. 2013). Briefly, *rex1Δ rrp47Δ* strains were transformed with a second plasmid encoding a test *rex1* allele and transformants were selected by growth on appropriate drop-out medium. Freshly grown precultures were normalized to a standard OD at 600 nm and 10-fold serial dilutions were generated in minimal selective medium. Aliquots were pinned to the surface of agar plates containing either selective SD medium or complete minimal medium containing uracil (50 µg/ml) and 1 mg/mL 5′-fluoroorotic acid (FOA) to select for loss of the initial plasmid. Plates were incubated at 30°C for 3 to 5 d. Assays were performed on at least three biological replicates.

Western analyses were performed on cell lysates prepared under alkaline conditions as previously described (Motley et al. 2012), except that the TCA-precipitated protein was resuspended in 15 mM Tris-HCl pH 8.7, 50% urea before adding SDS-PAGE loading buffer. Blots of yeast cell extracts were probed with peroxidase/anti-peroxidase (PAP) conjugate (P1291, Sigma-Aldrich) to detect the zz fusion proteins. Anti-Pgk1 (clone 22C5D8, Life Technologies) or anti-GFP (clones 7.1 and 13.1, Sigma-Aldrich) mouse monoclonal antibodies followed by HRP-linked goat anti-mouse secondary antibody (1706516, Bio-Rad Laboratories) were used to detect the levels of Pgk1 and GFP fusion proteins, respectively. GST-tagged proteins expressed in *E. coli* were detected with a rabbit anti-GST antibody (G7781, Merck) and a goat anti-rabbit HRP conjugate (A4914, Sigma-Aldrich). Proteins were visualized by ECL using an iChemi XL GelDoc system fitted with GeneSnap software (SynGene) and quantified using ImageJ (NIH).

Total cellular RNA was prepared from yeast using a standard phenol/guanidinium hydrochloride extraction. RNA was resolved through 8% or 6% acrylamide/urea gels and visualized by staining with ethidium bromide or hybridization of northern blots with a 5S rRNA-specific probe (Schuch et al. 2014). RNA analyses were performed on at least three biological replicates.

For epifluorescence microscopy, overnight cultures were diluted to an OD_{600} of ~0.1 and grown for 5 to 7 h in selective glucose-based medium. The cell pellet from 1mL of log phase culture was resuspended in 100 µL fresh medium and the cell suspensions were analyzed with epifluorescence, as previously described (Motley et al. 2015). Briefly, images were acquired using an Axiovert 200M (Carl Zeiss) equipped with an Exfo X-cite 120 excitation light source, band pass filters (Carl Zeiss and Chroma Technology Corp.), Plan Apochromat 63× 1.4 NA objective lens (Carl Zeiss), and a digital camera (Orca ER, Hamamatsu Photonics). Image acquisition was performed using Volocity software (PerkinElmer). Fluorescence images were collected as 0.5 mm Z stacks, merged into one plane in Openlab (PerkinElmer) and processed further in Photoshop (Adobe) by adjustment of levels. Brightfield images were collected in one plane at the center of cells, processed where necessary to highlight just the circumference of the cells and pasted into the blue channel of Photoshop.

5S rRNA processing assays

5S RNP complexes containing 3′ extended 5S rRNA were purified from the *rex1Δ* strain by sucrose density gradient ultracentrifugation using buffers containing EDTA (Blobel 1971; Steitz et al. 1988). Briefly, cell pellets were lysed with glass beads in a buffer comprising 50 mM HEPES pH 7.4, 50 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA. After clarification of the cell extract, EDTA was added to a final concentration of 3 mM and the lysate was left on ice for 15 min. The lysates were then loaded onto 10%–50% sucrose density gradients in a buffer comprising 10 mM HEPES pH 7.4, 50 mM KCl, 1 mM EDTA and centrifuged in an SW41 rotor at 36,000 rpm for 210 min. The sucrose gradients were aliquoted and $MgCl_2$ added to each fraction to a final concentration of 10 mM. The fractions were then aliquoted, frozen in liquid nitrogen and stored at −80°C. The gradient fractions were screened for 5S and 5.8S rRNA by electrophoresis through acrylamide/urea gels. Fractions were used for subsequent processing assays that contained 5S rRNA but lacked detectable 5.8S rRNA.

Expression of Rex1 GST fusion proteins in *E. coli* was autoinduced using Terrific broth supplemented with 0.05% glucose and 0.2% lactose. Cultures were grown to an OD_{600} of ~0.5 at 37°C and then transferred to 23°C and incubated overnight before harvesting. Cells were lysed in 10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween 20 containing 100 µM PMSF, 10 µM leupeptin, and 10 µM pepstatin A, and the proteins purified by affinity chromatography using glutathione sepharose beads. After washing with lysis buffer, bound protein was eluted by incubation in lysis buffer containing 20 mM glutathione and aliquots were frozen in liquid nitrogen. The relative yields of GST fusion proteins were estimated by SDS-PAGE analysis and equivalent amounts of the full-length proteins were used in the degradation assays.

5S RNP complexes were mixed with purified GST-tagged proteins and incubated at 30°C. Reaction mixtures contained 5 mM Tris-HCl pH 7.4, 5 mM HEPES pH 7.4, 82 mM NaCl, 25 mM KCl, 0.05% Tween 20, 10 mM glutathione, 0.5 mM EDTA, 5 mM $MgCl_2$, ~10 nM 5S RNP complex, and ~20 nM Rex1 protein. Aliquots were removed at time-points and the reaction quenched by addition of formamide gel loading buffer. The incubation mixtures were resolved through 6% acrylamide/urea gels, together with total cellular RNA from a wild-type strain, and the RNA detected by hybridization of northern blots using a 5S RNA specific probe. RNA was visualized by PhosphoImaging using a Typhoon FLA7000 (GE Healthcare) and quantified using ImageJ.

DNase assays were carried out on GST fusion proteins using a 5′ [^{32}P]-labeled oligonucleotide (o1165; CACGGATCCGATGAAG

TGGTTGTTGTT) predicted to have negligible secondary structure (a free energy of −1.6 kcal/mol and frequency of 17% at 30°C). Assays were carried out in mixtures containing 9 mM Tris-HCl pH 7.4, 135 mM NaCl, 0.1% Tween 20, 18 mM glutathione, ~20 nM DNA substrate and ~30 nM Rex1 protein. Incubation mixtures recovered at different time points were resolved through 16% acrylamide/urea gels, transferred to Hybond-N^+ membranes and the DNA was visualized by PhosphoImaging, as above. Nuclease assays using RNA and DNA substrates were performed on multiple technical repeat time courses using at least two independent samples of each purified protein.

Crosslinking assays

Cells expressing Rex1 zz fusion proteins from high copy number plasmids were harvested in mid-log growth, resuspended in fresh medium and transferred to petri dishes to give a depth of 1–2 mm. The petri dishes were placed on an ice-cooled glass plate and irradiated with UVC light for 5 min with a pulse time of 30 sec every minute, using a CL1000 Crosslinker (UVP Plastics). Irradiated cells were harvested by centrifugation and the cell pellets stored at −80°C.

The zz fusion proteins were purified from lysates of irradiated cells by affinity chromatography using IgG sepharose beads. Cells were lysed in a buffer consisting of 50 mM HEPES pH 7.6, 150 mM NaCl, 1 mM EDTA containing 1 mM PMSF and a yeast protease inhibitor cocktail (Melford Laboratories). Clarified lysates were mixed with IgG sepharose beads for 1 h at 4°C. After washing the beads with lysis buffer, nonspecific and indirectly bound protein was eluted in lysis buffer containing 2 M $MgCl_2$ (Allmang et al. 1999; Mitchell et al. 2003). The beads were then washed again with lysis buffer and incubated with buffer containing 1 µg RNase A for 30 min at 37°C. After washing with PNK buffer (50 mM HEPES pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), beads were incubated with 5 units of polynucleotide kinase and 3 pmol $\gamma[^{32}P]$-ATP (PerkinElmer). Nonincorporated ATP was removed by washing, and protein was recovered from the beads by elution with 0.5 M acetic acid. Eluted protein fractions were concentrated by extraction with n-butanol and then precipitated in n-butanol at −80°C. After centrifugation at 15,000*g* for 30 min, protein pellets were dried and then resuspended in SDS-PAGE loading buffer containing 50% urea. Crosslinking assays were carried out on two sets of biological replicates in the wild-type and D229A background.

For PsP cleavage of the crosslinked proteins, Rex1 was purified from irradiated cells and the bound RNA was digested and radiolabeled, as described above. After removal of unbound ATP, the beads were washed in PsP cleavage buffer (50 mM HEPES pH 7.6, 150 mM NaCl) and incubated with PsP overnight at 4°C. The solubilized fraction and subsequent acetic acid eluate were then concentrated and precipitated using butanol, as described above. Samples were resolved by SDS-PAGE and transferred to Protran membranes (GE Healthcare). Radiolabeled, crosslinked proteins were detected by PhosphoImaging and the zz fusion proteins were identified by western analyses. PsP was expressed in *E. coli* (using an expression plasmid kindly provided by Stuart Wilson, University of Sheffield) and purified by glutathione sepharose affinity chromatography, as described above. PsP cleavage assays were carried out on three sets of biological replicates.

Bioinformatics

Yeast protein sequences were obtained from the *S. cerevisiae* Genome Database (SGD) (Engel et al. 2014). Multiple sequence alignments were generated using Clustal Omega (Madeira et al. 2019). The access code for PDB files were as follows: iPGM from *B. stearothermophilus*, 1EJJ; Pan2 from *S. cerevisiae* with bound oligonucleotide, 6R9M; PPM from *B. cereus*, 3M8W. PDB files of the AlphaFold predicted protein structures were accessed at the EBI webpage (https://www.ebi.ac.uk) using Uniprot codes (Rex1, P53331; YFE9, O94443; SDN5, Q8L7M4). The DEDD domain sequence of Rex1 (SMART domain SM00479, residues 224–382) (Letunic et al. 2021) was fitted to the three-dimensional structure of yeast Pan2 with bound oligonucleotide (Tang et al. 2019) using the one-to-one threading tool of the Phyre2 web portal (Kelley et al. 2015). RMSD values between the threaded Rex1 DEDD domain model and the Pan2 substrate complex or Rex1 AlphaFold model were determined using the pairwise alignment tool hosted on the RSCB Protein Databank website (https://www.rcsb.org/alignment) by using the jFCAT (rigid) method and default parameter settings (Li et al. 2020). Molecular graphics were generated using UCSF Chimera (Pettersen et al. 2004). The terminal dinucleotide of the bound substrate in the Rex1 DEDD domain threaded model was manually superposed onto the AlphaFold model by alignment with the Y272 residue. An RNA duplex structure (PDB 1QCU) (Klosterman et al. 1999) was used for manual molecular docking of RNA base pairs to the AlphaFold model for Rex1 in UCSF Chimera. Oligonucleotide secondary structure for the DNA substrate was predicted with the RNAfold webserver (Lorenz et al. 2011), using energy parameters for DNA and a temperature of 30°C.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

T.H.S. was supported by a scholarship from the Iraqi Higher Committee for Education Development. P.D. and Q.L. were awarded PhD studentships funded by the White Rose BBSRC doctoral training program. We thank Joshua Kenway, Samuel Lewin, Marios Panayi, William Royle, and Sopida Wongwas for technical support, Mike Williamson and Stuart Wilson for critical reading of the manuscript, and John Rafferty for support in molecular graphics analyses.

Received August 4, 2021; accepted January 11, 2022.

REFERENCES

Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H. 2008. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics* **7:** 1389–1396. doi:10.1074/mcp.M700468-MCP200

Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P. 1999. The yeast exosome and human PM-Scl are related complexes of 3′→5′ exonucleases. *Genes Dev* **13:** 2148–2158. doi:10.1101/gad.13.16.2148

Bechhofer DH, Deutscher MP. 2019. Bacterial ribonucleases and their roles in RNA metabolism. *Crit Rev Biochem Mol Biol* **54:** 242–300. doi:10.1080/10409238.2019.1651816

Blobel G. 1971. Isolation of a 5S RNA-protein complex from mammalian ribosomes. *Proc Natl Acad Sci* **68:** 1881–1885. doi:10.1073/pnas.68.8.1881

Chen J, Liu L, You C, Gu J, Ruan W, Zhang L, Gan J, Cao C, Huang Y, Chen X, et al. 2018. Structural and biochemical insights into small RNA 3′ end trimming by *Arabidopsis* SDN1. *Nat Commun* **9:** 3585. doi:10.1038/s41467-018-05942-7

Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. 1992. Multifunctional yeast high-copy-number shuttle vectors. *Gene* **110:** 119–122. doi:10.1016/0378-1119(92)90454-W

Copela LA, Fernandez CF, Sherrer RL, Wolin SL. 2008. Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA* **14:** 1214–1227. doi:10.1261/rna.1050408

Costello JL, Stead JA, Feigenbutz M, Jones RM, Mitchell P. 2011. The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3′-maturation. *J Biol Chem* **286:** 4535–4543. doi:10.1074/jbc.M110.162826

Deutscher MP. 2006. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* **34:** 659–666. doi:10.1093/nar/gkj472

Deutscher MP, Marlor CW, Zaniewski R. 1985. RNase T is responsible for the end-turnover of tRNA in *Escherichia coli*. *Proc Natl Acad Sci* **82:** 6427–6430. doi:10.1073/pnas.82.19.6427

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, et al. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)* **4:** 389–398. doi:10.1534/g3.113.008995

Feigenbutz M, Garland W, Turner M, Mitchell P. 2013. The exosome cofactor Rrp47 is critical for the stability and normal expression of its associated exoribonuclease Rrp6 in *Saccharomyces cerevisiae*. *PLoS One* **8:** e80752. doi:10.1371/journal.pone.0080752

Foretek D, Wu J, Hopper AK, Boguta M. 2016. Control of *Saccharomyces cerevisiae* pre-tRNA processing by environmental conditions. *RNA* **22:** 339–349. doi:10.1261/rna.054973.115

Frank P, Braunshofer-Reiter C, Karwan A, Grimm R, Wintersberger U. 1999. Purification of *Saccharomyces cerevisiae* RNase H(70) and identification of the corresponding gene. *FEBS Lett* **450:** 251–256. doi:10.1016/S0014-5793(99)00512-8

Garland W, Feigenbutz M, Turner M, Mitchell P. 2013. Rrp47 functions in RNA surveillance and stable RNA processing when divorced from the exoribonuclease and exosome-binding domains of Rrp6. *RNA* **19:** 1659–1668. doi:10.1261/rna.039388.113

Gerstberger S, Meyer C, Benjamin-Hong S, Rodriguez J, Briskin D, Bognanni C, Bogardus K, Steller H, Tuschl T. 2017. The conserved RNA exonuclease Rexo5 is required for 3′ end maturation of 28S rRNA, 5S rRNA, and snoRNAs. *Cell Rep* **21:** 758–772. doi:10.1016/j.celrep.2017.09.067

Grandi P, Doye V, Hurt EC. 1993. Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. *EMBO J* **12:** 3061–3071. doi:10.1002/j.1460-2075.1993.tb05975.x

Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci* **106:** 9613–9618. doi:10.1073/pnas.0901997106

Holt LJ, Tuch BB, Villen J, Johnson AD, Gygi SP, Morgan DO. 2009. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325:** 1682–1686. doi:10.1126/science.1172867

Houseley J, Tollervey D. 2009. The many pathways of RNA degradation. *Cell* **136:** 763–776. doi:10.1016/j.cell.2009.01.019

Jedrzejas MJ, Chander M, Setlow P, Krishnasamy G. 2000. Structure and mechanism of action of a novel phosphoglycerate mutase from *Bacillus stearothermophilus*. *EMBO J* **19:** 1419–1431. doi:10.1093/emboj/19.7.1419

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596:** 583–589. doi:10.1038/s41586-021-03819-2

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10:** 845–858. doi:10.1038/nprot.2015.053

Kempers-Veenstra AE, Oliemans J, Offenberg H, Dekker AF, Piper PW, Planta RJ, Klootwijk J. 1986. 3′-end formation of transcripts from the yeast rRNA operon. *EMBO J* **5:** 2703–2710. doi:10.1002/j.1460-2075.1986.tb04554.x

Klosterman PS, Shah SA, Steitz TA. 1999. Crystal structures of two plasmid copy control related RNA duplexes: an 18 base pair duplex at 1.20 Å resolution and a 19 base pair duplex at 1.55 Å resolution. *Biochemistry* **38:** 14784–14792. doi:10.1021/bi9912793

Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci* **106:** 10171–10176. doi:10.1073/pnas.0900604106

Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res* **49:** D458–D460. doi:10.1093/nar/gkaa937

Li Z, Deutscher MP. 1995. The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. *Proc Natl Acad Sci* **92:** 6883–6886. doi:10.1073/pnas.92.15.6883

Li Z, Pandit S, Deutscher MP. 1999. Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. *RNA* **5:** 139–146. doi:10.1017/S1355838299981669

Li Z, Jaroszewski L, Iyer M, Sedova M, Godzik A. 2020. FATCAT 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res* **48:** W60–W64. doi:10.1093/nar/gkaa443

Longtine MS, McKenzie A III, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14:** 953–961. doi:10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U

Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6:** 26. doi:10.1186/1748-7188-6-26

MacGilvray ME, Shishkova E, Place M, Wagner ER, Coon JJ, Gasch AP. 2020. Phosphoproteome response to dithiothreitol reveals unique versus shared features of *Saccharomyces cerevisiae* stress responses. *J Proteome Res* **19:** 3405–3417. doi:10.1021/acs.jproteome.0c00253

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47:** W636–W641. doi:10.1093/nar/gkz268

Mitchell P, Petfalski E, Tollervey D. 1996. The 3′ end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev* **10:** 502–513. doi:10.1101/gad.10.4.502

Mitchell P, Petfalski E, Houalla R, Podtelejnikov A, Mann M, Tollervey D. 2003. Rrp47p is an exosome-associated protein required for the 3′ processing of stable RNAs. *Mol Cell Biol* **23:** 6982–6992. doi:10.1128/MCB.23.19.6982-6992.2003

Moser MJ, Holley WR, Chatterjee A, Mian IS. 1997. The proofreading domain of *Escherichia coli* DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res* **25:** 5110–5118. doi:10.1093/nar/25.24.5110

Motley AM, Nuttall JM, Hettema EH. 2012. Pex3-anchored Atg36 tags peroxisomes for degradation in *Saccharomyces cerevisiae*. *EMBO J* **31:** 2852–2868. doi:10.1038/emboj.2012.151

Motley AM, Galvin PC, Ekal L, Nuttall JM, Hettema EH. 2015. Reevaluation of the role of Pex1 and dynamin-related proteins in peroxisome membrane biogenesis. *J Cell Biol* **211:** 1041–1056. doi:10.1083/jcb.201412066

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10:** 202. doi:10.1186/1471-2105-10-202

Oficjalska-Pham D, Harismendy O, Smagowicz WJ, de Peredo A G, Boguta M, Sentenac A, Lefebvre O. 2006. General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. *Mol Cell* **22:** 623–632. doi:10.1016/j.molcel.2006.04.008

Ozanick SG, Wang X, Costanzo M, Brost RL, Boone C, Anderson JT. 2009. Rex1p deficiency leads to accumulation of precursor initiator $tRNA^{Met}$ and polyadenylation of substrate RNAs in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **37:** 298–308. doi:10.1093/nar/gkn925

Panosian TD, Nannemann DP, Watkins GR, Phelan VV, McDonald WH, Wadzinski BE, Bachmann BO, Iverson TM. 2011. *Bacillus cereus* phosphopentomutase is an alkaline phosphatase family member that exhibits an altered entry point into the catalytic cycle. *J Biol Chem* **286:** 8043–8054. doi:10.1074/jbc.M110.201350

Peng W-T, Robinson MD, Mnaimneh S, Krogan NJ, Cagney G, Morris Q, Davierwala AP, Grigull J, Yang X, Zhang W, et al. 2003. A panoramic view of yeast noncoding RNA processing. *Cell* **113:** 919–933. doi:10.1016/S0092-8674(03)00466-5

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605–1612. doi:10.1002/jcc.20084

Phillips S, Butler JS. 2003. Contribution of domain structure to the RNA 3′ end processing and degradation functions of the nuclear exosome subunit Rrp6p. *RNA* **9:** 1098–1107. doi:10.1261/rna.5560903

Piper PW, Bellatin JA, Lockheart A. 1983. Altered maturation of sequences at the 3′ terminus of 5S gene transcripts in a *Saccharomyces cerevisiae* mutant that lacks a RNA processing endonuclease. *EMBO J* **2:** 353–359. doi:10.1002/j.1460-2075.1983.tb01430.x

Puig O, Rutz B, Luukkonen BG, Kandels-Lewis S, Bragado-Nilsson E, Seraphin B. 1998. New constructs and strategies for efficient PCR-based gene manipulations in yeast. *Yeast* **14:** 1139–1146. doi:10.1002/(SICI)1097-0061(19980915)14:12<1139::AID-YEA306>3.0.CO;2-B

Ren YG, Martinez J, Virtanen A. 2002. Identification of the active site of poly(A)-specific ribonuclease by site-directed mutagenesis and Fe^{2+}-mediated cleavage. *J Biol Chem* **277:** 5982–5987. doi:10.1074/jbc.M111515200

Roberts DN, Wilson B, Huff JT, Stewart AJ, Cairns BR. 2006. Dephosphorylation and genome-wide association of Maf1 with Pol III-transcribed genes during repression. *Mol Cell* **22:** 633–644. doi:10.1016/j.molcel.2006.04.009

Schuch B, Feigenbutz M, Makino DL, Falk S, Basquin C, Mitchell P, Conti E. 2014. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J* **33:** 2829–2846. doi:10.15252/embj.201488757

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122:** 19–27. doi:10.1093/genetics/122.1.19

Silva S, Homolka D, Pillai RS. 2017. Characterization of the mammalian RNA exonuclease 5/NEF-sp as a testis-specific nuclear 3′→5′ exoribonuclease. *RNA* **23:** 1385–1392. doi:10.1261/rna.060723.117

Stead JA, Costello JL, Livingstone MJ, Mitchell P. 2007. The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. *Nucleic Acids Res* **35:** 5556–5567. doi:10.1093/nar/gkm614

Steitz JA, Berg C, Hendrick JP, La Branche-Chabot H, Metspalu A, Rinke J, Yario T. 1988. A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. *J Cell Biol* **106:** 545–556. doi:10.1083/jcb.106.3.545

Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR. 2013. Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. *J Biol Chem* **288:** 31816–31829. doi:10.1074/jbc.M113.491290

Swaney DL, Beltrao P, Starita L, Guo A, Rush J, Fields S, Krogan NJ, Villen J. 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* **10:** 676–682. doi:10.1038/nmeth.2519

Tang TTL, Stowell JAW, Hill CH, Passmore LA. 2019. The intrinsic structure of poly(A) RNA determines the specificity of Pan2 and Caf1 deadenylases. *Nat Struct Mol Biol* **26:** 433–442. doi:10.1038/s41594-019-0227-9

Ulbrich C, Diepholz M, Bassler J, Kressler D, Pertschy B, Galani K, Bottcher B, Hurt E. 2009. Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell* **138:** 911–922. doi:10.1016/j.cell.2009.06.045

Ullah R, Shah MA, Tufail S, Ismat F, Imran M, Iqbal M, Mirza O, Rhaman M. 2016. Activity of the human rhinovirus 3C protease studied in various buffers, additives and detergents solutions for recombinant protein production. *PLoS One* **11:** e0153436. doi:10.1371/journal.pone.0153436

van Hoof A, Lennertz P, Parker R. 2000. Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J* **19:** 1357–1365. doi:10.1093/emboj/19.6.1357

Wu M, Nilsson P, Henriksson N, Niedzwiecka A, Lim MK, Cheng Z, Kokkoris K, Virtanen A, Song H. 2009. Structural basis of m^7GpppG binding to poly(A)-specific ribonuclease. *Structure* **17:** 276–286. doi:10.1016/j.str.2008.11.012

Zinder JC, Lima CD. 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* **31:** 88–100. doi:10.1101/gad.294769.116

Zuo Y, Deutscher MP. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29:** 1017–1026. doi:10.1093/nar/29.5.1017

Zuo Y, Deutscher MP. 2002a. Mechanism of action of RNase T. I. Identification of residues required for catalysis, substrate binding, and dimerization. *J Biol Chem* **277:** 50155–50159. doi:10.1074/jbc.M207706200

Zuo Y, Deutscher MP. 2002b. Mechanism of action of RNase T. II. A structural and functional model of the enzyme. *J Biol Chem* **277:** 50160–50164. doi:10.1074/jbc.M207707200

Zuo Y, Deutscher MP. 2002c. The physiological role of RNase T can be explained by its unusual substrate specificity. *J Biol Chem* **277:** 29654–29661. doi:10.1074/jbc.M204252200

Zuo Y, Wang Y, Malhotra A. 2005. Crystal structure of *Escherichia coli* RNase D, an exoribonuclease involved in structured RNA processing. *Structure* **13:** 973–984. doi:10.1016/j.str.2005.04.015

Zuo Y, Zheng H, Wang Y, Chruszcz M, Cymborowski M, Skarina T, Savchenko A, Malhotra A, Minor W. 2007. Crystal structure of RNase T, an exoribonuclease involved in tRNA maturation and end turnover. *Structure* **15:** 417–428. doi:10.1016/j.str.2007.02.004

See the following page for ***Meet the First Author***