Processing of DNA and RNA substrates by endonuclease V

Erik Sebastian Vik

Thesis for the degree of Philosophiae Doctor (PhD)

Oslo, November 2013

Department of microbiology
Department of medical biochemistry
Oslo University Hospital
University of Oslo
Norway

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 1736

ISBN 978-82-8264-720-5

Cover: Inger Sandved Anfinsen.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

Acknowledgements

The work presented in this dissertation was conducted at Department of Microbiology and the Department of Medical Biochemistry at Oslo University Hospital, Rikshospitalet and University of Oslo during the period from May 2010 to November 2013. Financial support was provided by the University of Oslo and Anders Jahres fond for vitenskapens fremme.

First of all, I wish to express my gratitude towards my three supervisors. My main supervisor Dr. Ingrun Alseth has been an excellent guidance during my work with the thesis. I have really appreciated that you always are available for questions and discussion. You have really taught me a lot during the last years.

Professor Magnar Bjørås, head of research at the Department of Microbiology, introduced me to the exiting field of DNA repair when I carried out my bachelor thesis at the department in 2007. Thank you for giving me the opportunity to work in the lab and perform a PhD. Your knowledge in science is really admirable and I think everybody gets inspired by the commitment and enthusiasm you show for science (and sport). Thanks for always turning negative results into something positive and useful, and also for our trip to Giens in France.

My third supervisor, Dr. Bjørn Dalhus, has also been a great support during my PhD. Always open for questions and help when needed. I also thank you for the introduction to geocaching at our two interesting and enjoyable conference trips to Amsterdam and Giens.

I would also like to thank present and former members of the “pombe-group”, Cathrine, Meh Sameen, Hanne and Line for scientific support and conversations on everything from science to everyday nonsense.

Thanks to all the co-authors of the papers, for both experimental work and preparation of the manuscripts. I would also like to thank the rest of my colleagues at the department for making a great working environment. In particular Ragnhild, my “lab-bench-partner”.

Finally, I wish to thank my entire family for their support, kindness and encouragement. Especially Anne Margrethe, Elise and Jonas, you are the best!

Oslo, February 2014

Erik Sebastian Vik

Table of Contents

List of papers

The thesis is based upon the following three papers, which will be referred to by their Roman numbers throughout the text.

I **The human homolog of *Escherichia coli* endonuclease V is a nucleolar protein with affinity for branched DNA structures.**
Fladeby C*, **Vik ES***, Laerdahl JK, Gran Neurauter C, Heggelund JE, Thorgaard E, Strøm-Andersen P, Bjørås M, Dalhus B and Alseth I.
PLoS One, Volume 7, Issue 11, November 2012.

II **Endonuclease V cleaves at inosines in RNA.**
Vik ES, Nawaz MS, Strøm-Andersen P, Fladeby C, Bjørås M, Dalhus B and Alseth I.
Nature communications, 4:2271, August 5th, 2013

III **Structural basis of DNA loop recognition by endonuclease V.**
Rosnes I, Rowe AD, **Vik ES**, Forstrøm RJ, Alseth I, Bjørås M and Dalhus B.
Structure 21, 257-265, February 5th 2013.

* The authors have contributed equally to the work.

Abbreviations

1-meA	1-methyladenine
3-meC	3-methylcytosine
5-hmU	5-hydroxymethyluracil
5-meC	5-methylcytosine
5-ohU	5-hydroxyuracil
8-oxoG	7,8-dihydro-8-oxoguanine
8-oxodGTP	7,8-dihydro-8-oxodeoxyguanine triphosphate
A	adenine
A-to-I	adenosine to inosine
AAG	alkyladenine DNA glycosylase
ADAR	adenosine deaminases acting on RNA
AID	activation-induced deamination
AlkA	3-methyladenine DNA glycosylase
AMP	adenosine monophosphate
AP-site	apurinic/apyrimidinic site
BER	base excision repair
BS	Bloom syndrome
C	cytosine
CRC	colorectal cancer
CS	Cockayne syndrome
DDB2	DNA-binding protein 2
DDR	DNA damage response
DNA	deoxyribonucleic acid
dITP	deoxyinosine triphosphate
dNTP	deoxynucleoside triphosphate
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DSB	double-strand break
dUTP	deoxyuridine triphosphate
dUMP	deoxyuridine monophosphate
dXTP	deoxyxanthosine triphosphate
E. coli	*Escherichia coli*
EndoV	Endonuclease V
G	guanine
GFP	green fluorescent protein
GGR	global-genomic repair
GMP	guanosine monophosphate
HCC	hepatocellular carcinoma
HR	homologous recombination
Hx	hypoxanthine
I	inosine
IDLs	insertion/deletion loops
IMP	inosine monophosphate
ITPase	inosine triphosphatase
KO	knockout
LS	Lynch syndrome
MBD4	methyl-CpG-binding domain 4
MMR	mismatch repair
Mug	mismatch-specific uracil-DNA glycosylase
mRNA	messenger RNA
miRNA	micro RNA
NEIL1	endonuclease VIII-like 1 protein
NHEJ	non-homologous end joining
NMP	nucleoside monophosphate
OH	hydroxyl
P	phosphoryl
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAP	RNA polymerase
RNS	reactive nitrogen species
ROS	reactive oxygen species
rRNA	ribosomal RNA
RT-PCR	real time polymerase chain reaction
SG	stress granules
SMUG1	single stranded selective monofunctional uracil-DNA glycosylase
S. pombe	*Schizosaccharomyces pombe*
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
T	thymine
TCR	transcription-coupled repair
TDG	thymine-DNA glycosylase
TENR	testis-expressed nuclear RNA-binding protein
TG	thymine glycol
T. maritima	*Thermotoga maritima*
tRNA	transfer RNA
TTD	trichothiodystrophy
U	uracil
UTRs	untranslated regions
UV	ultraviolet
WS	Werner syndrome
X. laevis	*Xenopus laevis*
XP	xeroderma pigmentosum

Summary

The DNA molecule, which comprises the genetic information in all organisms, is constantly challenged by endogenous and exogenous agents, threatening the genetic stability. If the damage are not repaired, mutations might accumulate, and possible consequences are cancer, aging and neurodegenerative disease. To protect the cells and preserve the DNAs integrity, a number of repair pathways have evolved. Deamination of bases in DNA can arise spontaneously, after chemical exposure or by enzymatic processes. Deamination of adenine leads to hypoxanthine, cytosine to uracil and guanine to both xanthine and oxanine. The deamination products are miscoding and potentially mutagenic. Several enzymes are involved in the repair of deaminated bases and endonuclease V (EndoV) is identified as the main enzyme for initiation of hypoxanthine repair in prokaryotes. EndoV hydrolyzes the second phosphodiester bond 3' to the lesion, leaving a 3' hydroxyl and a 5' phosphoryl group, using Mg^{2+} as a cofactor. Any downstream processes are not known. EndoV is highly conserved and homologous are found from bacteria to humans, suggesting an important function for the enzyme.

The eukaryotic homologous of EndoV are poorly characterized and we aimed to identify substrates and function for human EndoV (hEndoV). In paper I, we characterized hEndoV by bioinformatics, gene expression and *in vitro* analyses. We identified transcripts with different exon-configurations and found hEndoV to have affinity for various branched DNA substrates like flap, pseudo-Y, three-way junction, fork-structure and Holliday junction. However, under the assay condition used, we did not find any activity or affinity for hypoxanthine in DNA.

Adenines in RNA are also subjected to deamination and inosine is the most common editing event in RNA (A-to-I editing). Both tRNAs, mRNAs, microRNAs and Alu-containing transcripts are deaminated by enzymes called "adenosine deaminases that act on RNA". Many A-to-I deamination targets are found in the central nervous system and proper editing is a determinant for correct protein function. To further characterize hEndoV, we assayed with different RNA substrates and found hEndoV to cleave RNA with inosine. The substrate specificity for inosine in RNA could point to a function for EndoV in RNA metabolism rather than DNA repair.

The structure of *Thermotoga maritima* EndoV (TmEndoV) has been solved in complex with DNA with hypoxanthine. The structure revealed a conserved wedge motif (Pro79, Tyr80, Ile81 and Pro82) which separates the two DNA strands at the lesion. To further investigate the role of the wedge, the crystal structure of TmEndoV in complex with a one-nucleotide loop was solved (paper III). The structure shows how the wedge separates the two DNA strands exactly at the helical distortion and a normal adenine is flipped into the recognition pocket.

1. Introduction

"The double helix is indeed a remarkable molecule. Modern man is perhaps 50,000 years old, civilization has existed for scarcely 10,000 years and the United States for only just over 200 years; but DNA and RNA have been around for at least several billion years. All that time the double helix has been there, and active, and yet we are the first creatures on Earth to become aware of its existence."

Francis Crick (1916–2004).

DNA, a macromolecule essential for life, was first described in the late 1860s by the Swiss doctor Friedrich Miescher (Dahm 2008) and has undergone extensive research during the last decades. Two of the major highlights have been the discovery of the double helical structure by Watson and Crick in 1953 (Watson et *al.*, 1953) and the first complete sequencing of a human genome in 2001 (Venter et *al.*, 2001). Accurate maintenance of the genetic information stored in the DNA is vital for proper life function and survival for all living organisms. Both endogenous and exogenous agents challenge the genomes integrity and if left unrepaired the damage may impede cellular processes such as replication and transcription and the consequence for an organism may be cancer, premature aging or inheritable diseases (Hakem 2008).

1.1 DNA damage and repair

Cells of all organisms are continuously challenged by exogenous DNA damaging agents, also known as environmental agents, which cause a variety of lesions. The consequence of DNA damage can be illustrated by the increasing problem in many western countries with growth in skin cancer after excessive sun exposure (Kanavy et *al.*, 2011). Ultraviolet (UV) light from sunlight is in fact the most prominent and widespread physical carcinogen in our natural environment (de Gruijl 1999). UV-light from the sun and sun beds mainly induce bulky photoproducts between adjacent thymine (T) residues (Altieri et *al.*, 2008). Other sources of environmental DNA damage are X-rays and other sources of ionizing radiation and viral infections. DNA

damaging agents are also introduced from genotoxins in food and cigarettes and also chemotherapeutic agents (Hoeijmakers 2001).

While we at least partly can protect ourselves against environmental toxic compounds, exposures from endogenous agents are inevitable. The intracellular environment can be quite hostile to DNA. Cellular metabolites, reactive oxygen species (ROS) and water may interact with and damage DNA. Mistakes made by DNA polymerases may generate base mismatches by incorporation of non-complementary nucleotides during replication or insertion/deletion loops (IDLs). IDLs are caused by replication-associated strand slippage, most commonly found in repetitive sequences (Hsieh et *al.*, 2008; Kunkel et *al.*, 2000; Kunkel et *al.*, 2005; Pena-Diaz et *al.*, 2012). The most frequent DNA damage is single-strand breaks with ~50 000 lesions pr cell pr day (Tice et *al.*, 1985). Another common damage is apurinic/apyrimidinic sites (AP-sites) with as much as 10 000 events pr human cell per day (Lindahl 1993). The brain is the most affected organ and purines are lost more easily than pyrimidines (Altieri et *al.*, 2008). AP-sites are generated via hydrolysis, a spontaneous reaction with water, by DNA glycosylases during removal of damaged bases and also by some chemicals, like free radicals, which promotes the release of bases (Dianov et *al.*, 2003). Hydrolysis is also responsible for deamination of bases, with deamination of adenine being the most relevant modification for this thesis (discussed in detail in section 1.1.1). Oxidation of DNA by ROS is responsible for other common groups of DNA damage including 7,8-dihydro-8-oxoguanin (8-oxoG), thymine glycol (TG), formamidopyrimidine (faPyA and faPyG) and others. ROS are made under normal physiological conditions and from exogenous factors (Dalhus et *al.*, 2009b).

To make sure the genetic information is accurately duplicated and transferred to next generation, cells employ different strategies to detect, signal and respond to DNA damage. These responses are termed the DNA damage response (DDR) and includes cell cycle checkpoint activation, DNA repair, senescence and apoptosis and are highly dependent on post-translational protein modifications (Huen et *al.*, 2008). Checkpoint activation involves different DNA damage sensors, signal transducers and effector pathways which can arrest the cell in distinct phases of the cell cycle. After arrest, cells of multicellular eukaryotes may take different actions which will be decisive for a cells fate. One option for a cell is to go into senescence, a process which is related to protection against cancer and aging because it obstruct excessive or

aberrant cellular proliferation (Collado et *al.*, 2007). Another possibility is DNA damage repair through different DNA repair pathways. Repair may restore the DNA making it possible for the cell to progress in the cell cycle. The different DNA repair pathways will be discussed in more detail in the succeeding sections. A third response is programmed cell death, named apoptosis, where the cell signals its own death. Apoptosis may be the best solution to avoid transfer of e.g. mutations to the next generation and avoid propagation of genetically deviating cells (Friedberg et *al.*, 2006; Hakem 2008).

1.1.1 Deamination

Deamination of DNA is the process were one of the bases adenine (A), cytosine (C), guanine (G) or 5-methylcytosine (5-meC) loses its exocyclic amino group, and the amino group is being replaced by a keto group (figure 1).

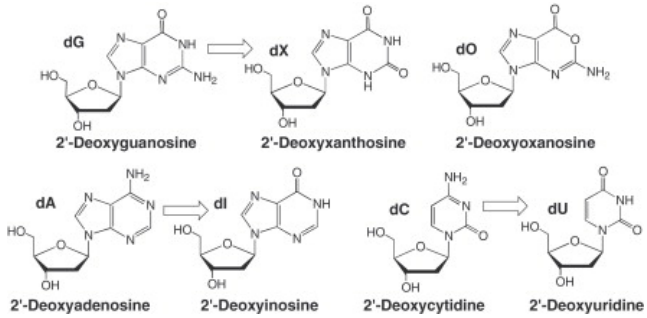


Figure 1: Examples of deamination products. Deamination of deoxyguanosine yields deoxyxanthosine or deoxyoxanosine, deoxyadenosine yields deoxyinosine and deoxycytidine yields deoxyuridine **(Dedon et *al.*, 2006)**.

Deamination may change the pairing properties of the bases in DNA and therefore represents a threat to the genetic stability. Deaminated adenine yields hypoxanthine (the nucleoside is termed inosine (I)) which is read as guanine by the replication machinery, if not repaired, a mutagenic A-T to G-C transition mutation will be fixed in the genome (Hill-Perkins et *al.*, 1986). Xanthine and oxanine are both products of guanine deamination (Lucas et *al.*, 1999; Lucas et *al.*, 2001; Suzuki et *al.*, 1997). Xanthine is able to make weak base pairs with both cytosine and thymine, where thymine is preferred, leading to G-C to A-T transition mutations (Eritja et *al.*, 1986). Oxanine can also base pair with both cytosine and thymine, with preference for thymine (Suzuki et *al.*, 1998). Deamination of cytosine forms uracil which gives U-G

mispairs that after a round of replication ends up as C-G to T-A transition mutations (Lindahl et *al.*, 1974). In contrast to deamination of adenine, cytosine and guanine, deamination of 5-meC does not change the base to a non-canonical base, but to one of the normal bases in DNA, thymine. 5-meC is deaminated three to four times more frequent than normal cytosine (Shen et *al.*, 1994) and the deamination gives a T-G mismatch. Many cytosine residues that are methylated to 5-meC in both the *E. coli* and human genome are considered as mutational "hot-spots" with increased risk for fixation of mutations. There are two reasons for 5-meC to be mutational "hot-spots"; first is the more frequent deamination rate and second is that T-G mismatches are repaired less efficiently than U-G mismatches (Duncan et *al.*, 1980; Lutsenko et *al.*, 1999). Additional deamination products may be generated from oxidative stress, such as 5-hydroxymethyluracil (5-hmU) and 5-hydroxyuracil (5-ohU) (Kow 2002; Lindahl 1993; Suzuki et *al.*, 2000).

To counteract the mutagenic effect of deaminated bases in DNA, both prokaryotic and eukaryotic cells have several DNA repair enzymes that remove deaminated bases. Lesion specific DNA glycosylases are the main enzymes involved, removing the bases through the base excision repair (BER) pathway (Kow 2002). In *E. coli*, hypoxanthine, xanthine and oxanine are removed by 3-methyladenine DNA glycosylase (AlkA) (Saparbaev et *al.*, 1994; Terato et *al.*, 2002), mismatch-specific uracil-DNA glycosylase (Mug) (Lee et *al.*, 2010b; O'Neill et *al.*, 2003), endonuclease VIII (Nei; xanthine and oxanine only) (Terato et *al.*, 2002) and EndoV (Gates, III et *al.*, 1977; He et *al.*, 2000; Hitchcock et *al.*, 2004; Yao et *al.*, 1994a; Yao et *al.*, 1994b). In mammalian cells, oxanine is removed by single-strand selective monofunctional uracil-DNA glycosylase (SMUG1) and endonuclease VIII-like 1 protein (NEIL1) (Dong et *al.*, 2008) whereas alkyladenine DNA glycosylase (AAG) are processing both hypoxanthine, xanthine and oxanine (Saparbaev et *al.*, 1994; Terato et *al.*, 2002).

The enzyme responsible for removing uracil from DNA, uracil-DNA glycosylase (UDG), catalyzes the hydrolytic cleavage of the *N*-glycosylic bond. UDG was first identified in *E. coli* (Lindahl 1974) and later two different forms were found in mammals (Slupphaug et *al.*, 1993). Other enzymes with uracil in DNA as a substrate have also been identified. In *E. coli* Mug is removing uracil (Gallinari et *al.*, 1996) and three human DNA glycosylases share uracil as substrate; thymine-DNA

glycosylase (TDG) (Neddermann et *al.*, 1994), methyl-CpG-binding domain 4 protein (MBD4) (Hendrich et *al.*, 1999) and SMUG1 (Haushalter et *al.*, 1999).
Deaminated bases in DNA can be generated in four ways: hydrolytic-, nitrosative- and enzymatic deamination and direct incorporation of deaminated nucleotides in DNA.

Hydrolytic deamination

Hydrolytic deamination may occur through two different reactions, either by an attack from a hydroxide or via a water molecule (figure 2). These reactions might occur spontaneously under normal cellular conditions, but they are considerably enhanced by ROS or agents such as nitrous acid (HNO_2) or nitric oxide (NO) (Gates 2009; Shapiro et *al.*, 1968; Shapiro et *al.*, 1969).

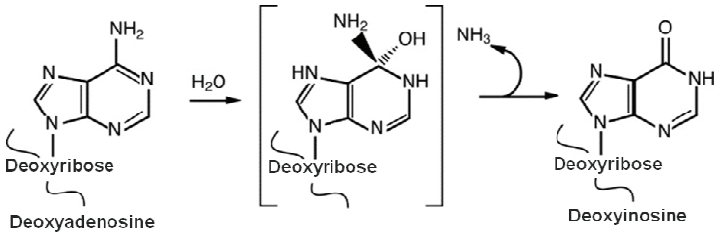


Figure 2: Hydrolytic deamination of deoxyadenosine to deoxyinosine. A simplified illustration of hydrolytic deamination of deoxyadenosine to deoxyinosine, with the release of ammonia (NH_3). Modified from: **(Schaub et *al.*, 2002)**

Deamination occurs faster in single-stranded DNA (ssDNA) compared to double-stranded DNA (dsDNA), since the amino group is less accessible in a double-stranded configuration. The bases with the highest deamination rate are cytosine and 5-meC. It is estimated that ~100-500 cytosine residues are deaminated to uracil in each human cell per day (Lindahl et *al.*, 1974; Shen et *al.*, 1994). Adenine and guanine are deaminated only 2-3% of the frequency of cytosine deamination (Karran et *al.*, 1980; Lindahl et *al.*, 1974).

Nitrosative deamination

Environmental agents and cellular metabolism are both considerable sources to deamination. During anaerobic respiration, nitrate (NO_3^-) and nitrite (NO_2^-) are the preferred electron acceptors in many bacteria. The metabolism of nitrate and nitrite

generates NO and HNO_2 which can be further processed into the mutagenic nitrosating agent nitrous anhydride (N_2O_3), which is believed to be the most important contributor to deamination of bases in DNA (Dedon et *al.*, 2006; Lewis et *al.*, 1995; Weiss 2006). NO is an important cellular signaling molecule in mammalian cells. One example is in the immune and inflammatory response system, where it is produced by macrophages and released, together with other substances, as part of the response. Experiments with NO scavengers shows that the cytostatic effect of macrophages is blocked and it is therefore believed that NO is the key mediator of macrophage-induced cytostasis (Burney et *al.*, 1999; MacMicking et *al.*, 1997). The production of NO by activated macrophages in chronic inflammation exposes the surrounding epithelial cells to elevated levels of reactive nitrogen species (RNS) which might damage DNA. Chronic inflammation is consequently regarded a risk factor for the development of different human cancers (Ohshima et *al.*, 1994).

Enzymatic deamination

There are different enzymes which can deaminate bases in both DNA and RNA (RNA deamination will be discussed in section 1.2). Activation-induced deaminase (AID) is an example, that can deaminate cytosine to uracil in ssDNA (Bransteitter et *al.*, 2003; Muramatsu et *al.*, 2000). AID deamination initiates the three antibody diversification processes in B-cells: somatic hypermutation, gene conversion and class switch recombination (Arakawa et *al.*, 2002; Muramatsu et *al.*, 2000). Genetically inherited defects in AID have been shown to give hyper IgM syndrome, a rare human immunodeficiency disease characterized by profound predisposition for bacterial infections (Revy et *al.*, 2000).

Incorporation of deaminated nucleotides in DNA

Free purine nucleotides used for DNA synthesis may be modified trough deamination, oxidation and other processes. If these non-canonical deoxynucleoside triphosphates (dNTP) are not degraded to their corresponding monophosphates (dNMP), they might be incorporated in the DNA. In *E. coli* deoxyuridine triphosphate (dUTP) is the most common non-canonical dNTP. dUTP is hydrolyzed by deoxyuridine triphosphatase (dUTPase) to dUMP, reducing the number of misincorporated dUTPs in DNA (Galperin et *al.*, 2006). Deoxyinosine- and deoxyxanthosine triphosphates (dITP/dXTP) are both removed from the precursor pool by inosine triphosphatase (ITPase) (Chern et *al.*, 1969; Liakopouloul et *al.*, 1964). Homologous of ITPase are

found from prokaryotes to mammals and cells lacking the enzyme display severe genomic instability. *E. coli* cells are not viable when mutated in both the *rdgB* gene encoding the ITPase homologue and the *recA* gene that encodes recombination repair protein RecA (Clyman et *al.*, 1987). Further, ITPase knockout mice show growth retardation and die before two weeks of age (Behmanesh et *al.*, 2009).

1.1.2 DNA repair mechanisms

DNA damage have led to a selection for enzymes and mechanisms to give cells and organisms tools to combat the multitude of DNA damage and to ensure the correct transcription and copying of the DNA blueprint. More than 150 mammalian DNA repair genes are involved in DNA repair (Wood et *al.*, 2005). The repair mechanisms are categorized as either direct reversal of damage, excision repair or double-strand break repair. Some proteins are involved in more than one pathway, but for simplification, the pathways will be described as six distinct pathways (figure 3) (Kovtun et *al.*, 2007).

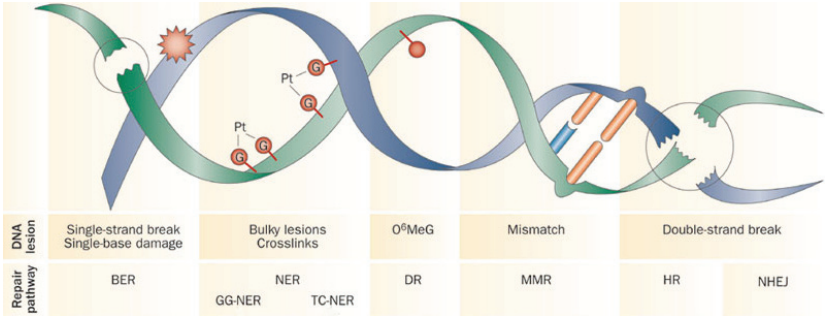


Figure 3: Principal DNA lesions and repair pathways. Single-strand breaks and simple base modifications, like oxidation and deamination, are mostly repaired by base excision repair (BER). Lesions that alter the structure of the DNA strand are corrected by nucleotide excision repair (NER), either through global-genomic NER (GG-NER) or transcription-coupled NER (TC-NER), whereas simple base modifications might undergo direct reversal (DR). The mismatch repair machinery repairs base mismatches and the highly toxic double-strand breaks are handled by either homologous recombination or non-homologous end joining (NHEJ). O^6MeG; O^6-methylguanine, Pt; platinum. The picture is adapted from **(Postel-Vinay et *al.*, 2012)**.

Direct reversal repair

The simplest and most cost efficient way for a cell to restore a damaged base to its native form is to use a single enzyme in a one-step reaction. The methylated bases 1-

methyladenine (1-meA) and 3-methylcytosine (3-meC) are repaired by direct reversal by an important group of enzymes which include *E. coli* AlkB (Falnes et *al.*, 2002; Trewick et *al.*, 2002) and the human homologues hABH2 and hABH3 (Aas et *al.*, 2003; Duncan et *al.*, 2002). AlkB belongs to a family of proteins which is dependent on iron and 2-oxoglutarate to restore bases such as 1-meA and 3-meC through oxidative demethylation (Falnes et *al.*, 2002; Trewick et *al.*, 2002).

Alkylating agents may methylate guanine to O^6-methylguanine which can be reversed by O^6-methylguanine DNA methyltransferase (MGMT), another direct reversal repair mechanism. MGMT transfers the methyl group to a cystein residue, but in contrast to many other enzymes, MGMT is a "suicide" protein which remain bound to the methyl group and therefore cannot regenerate after the reaction (Demple et *al.*, 1982; Lindahl et *al.*, 1982; Olsson et *al.*, 1980).

Base excision repair

Base excision repair is probably the mechanism which removes the highest number of damage from DNA, and the mechanism is conserved from *E. coli* to human (Friedberg et *al.*, 2006). The substrates for BER are simple base lesions such as base modifications, single-strand breaks and AP-sites, and the process is initiated by a group of enzymes named DNA glycosylases. So far 11 human DNA glycosylases are identified (Dianov et *al.*, 2013). A DNA glycosylase recognizes a damaged base and cleaves the N-glycosidic bond between the base and the 2'-deoxyribose, leaving an AP-site. The DNA glycosylases can either be mono- or bifunctional. Monofunctional glycosylases are dependent on an AP-endonuclease for 5'-incision of the AP-site whereas the bifunctional glycosylases have an additional AP-lyase activity which cleaves 3' to the AP-site. The following process proceeds through one of two sub-pathways. The first option is short patch BER where only a single nucleotide is removed, the other is long patch BER where up to 12 bases are removed before DNA polymerase replaces the bases and DNA ligase finalizes the repair process by re-sealing the strand (Altieri et *al.*, 2008; Dalhus et *al.*, 2009b; Dianov et *al.*, 2013).

Nucleotide excision repair

In contrast to BER, that is dependent on enzymes with affinity for a limited number of substrates, nucleotide excision repair (NER) is much more versatile. More than 30 enzymes are involved in human NER (Fagbemi et *al.*, 2011) whereas the prokaryotic counterpart only consist of three proteins; UvrA, B and C (Kisker et *al.*, 2013).

Substrates for NER are distortions in the DNA strand caused by chemicals that covalently bind DNA and makes bulky adducts or cross-linking agents which forms two covalent bonds with DNA, cross-linking within a strand or between the two DNA strands in the helix. NER is also crucial for removal of UV-induced damage like cyclobutane pyrimidine dimers and (6-4) pyrimidine pyrimidone photoproducts (Nouspikel 2009). NER can be divided in two sub-pathways: transcription-coupled repair (TCR) and global-genomic repair (GGR). During transcription, TCR may operate on the transcribed strand to repair damage that, if left unrepaired, may lead to up- or down regulation of the gene expression. Such damage may also act as an obstacle for the RNA polymerase (RNAP) complex and either release the polymerase too early with the unfinished mRNA or stop transcription completely. A stalled RNAP will recruit the tumor suppressor p53 which acts as a signal for apoptosis (Tornaletti 2009). Global genomic repair is not dependent on actively transcribed DNA and repairs damage throughout the genome. The recognition part of NER is different in TCR and GGR. In GGR the xeroderma pigmentosum group C (XPC) complex (XPC, HR23B and centrin 2) senses the distortion and initiates the repair (Araki et *al.*, 2001). In TCR, Cockayne syndrome type A and B proteins (CSA and CSB) might be involved in damage recognition but their function remains unclear (Nouspikel 2009). The downstream processes are the same for GGR and TCR and includes incision on both sides of the lesion by two structure specific endonucleases, 5' by ERCC1 and 3' by XPG, removal of 24-32 nucleotides, resynthesis of the gap with the non-damaged strand as a template, followed by ligation by DNA ligase (Fagbemi et *al.*, 2011).

Mismatch repair

Mismatches may be generated during replication by insertion of incorrect bases by DNA polymerases. These mismatches, together with small unpaired loops caused by replication slippage in repetitive sequences, might be missed by the proofreading function of the DNA polymerase and end up as substrates for the mismatch repair system (MMR) (Jiricny 2006; Li 2008). The enzymes involved are conserved from bacteria to humans and was first identified in *E. coli* where the genes were designated "Mut" because of their hypermutator knockout phenotype (Altieri et *al.*, 2008). The MutS protein initiates MMR in *E. coli* by binding to the mismatch followed by recruitment of MutL. This activates the MutH endonuclease, that can discriminate the

template from the newly synthesized strand because of the methylation pattern, and incises the newly synthesized strand at a hemi-methylated GATC sequence (Au et *al.*, 1992; Welsh et *al.*, 1987). Next, UvrD helicase unwinds the strand and one of several exonucleases excise nucleotides from the nick and past the mismatch while single-strand binding protein (SSB) stabilizes the gap. DNA polymerase III and DNA ligase completes the reaction (Jiricny 2006; Li 2008). In mammals, the repair is initiated by either one of two heterodimeric complexes; MutSα (MSH2 and MSH6) predominantly recognizing base mismatches or IDLs with one or two extra bases and MutSβ (MSH2 and MSH3) for larger IDLs (Acharya et *al.*, 1996; Drummond et *al.*, 1995; Palombo et *al.*, 1995). Four MutL homologues are indentified in humans; MLH1, MLH3, PMS1 and PMS2. These proteins combine and operate as three different heterodimers, where MutLα (MLH1 and PMS2) is most important for MMR (Jiricny 2006). No MutH homologues have been identified in eukaryotes and the exact downstream mechanisms are not yet fully understood. What is known is that proliferating nuclear antigen (PCNA), replication factor C (RFC), exonuclease I (EXO1), replication protein A (RPA), DNA polymerase δ and DNA ligase 1 are involved (Pena-Diaz et *al.*, 2012). Discrimination of the newly snthesized strand from the template could be from pre-existing discontinuities at the termini of Okazaki fragments or after the removal of the RNA primer by RNAse H (Fukui 2010; Li 2008).

Homologous recombination and non-homologous end joining

DNA double-strand breaks (DSB) are highly toxic and the most lethal form of DNA damage. They can be generated from exogenous sources like ionizing radiation and genotoxic chemicals. The main contributor to endogenous DSBs is fork collapse, that occur when DNA replication forks are confronted with unrepaired DNA lesions (Chapman et *al.*, 2012; Goodarzi et *al.*, 2013). There are two repair pathways for DSBs: homologous recombination (HR) and non-homologous end joining (NHEJ). More than 90% of the mammalian DSBs are processed by NHEJ. In bacteria and yeast, on the other hand, HR is the preferred repair mechanism (Hakem 2008). HR is the most accurate pathway of the two, using an intact homologous DNA strand as a template, usually located on the sister chromatide. Because of the need for a template strand, HR generally takes place in the S and G2 phase of the cell cycle, when a sister chromatide is available. HR is initiated by the resection of the double-strand, making

ssDNA tails. This single-stranded tails are coated with RPA which the RAD51 recombinase assembles on and then invades the homologous DNA which will be used as a template for repair (Chapman et *al.*, 2012; Goodarzi et *al.*, 2013).

Compared to HR, NHEJ is an easier way to repair a DSB, and NHEJ is not dependent on a homologous template. NHEJ is initiated by the Ku70-Ku80 heterodimeric complex which binds both DNA ends, protecting the DNA from degradation and recruits the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). The DNA ends are then processed, something that might give a small deletion, before they are ligated (Chapman et *al.*, 2012; Lieber 2010; Symington et *al.*, 2011).

1.1.3 Human disease – consequences of repair defects

A restricted number of hereditary diseases, which symptoms are caused by genetic defects in different repair pathways, are described. The symptoms are not all the same, but they share many features like increased risk for different cancers, growth retardation, premature aging and neurological degeneration. Research on these syndromes is important because they can be regarded as models for similar, but less severe, diseases in the "normal" population (Knoch et *al.*, 2012; Thoms et *al.*, 2007). In this section, some of these syndromes will be described briefly, with focus on clinical features and genes involved.

Inheritable mutations in 13 genes involved in NER have so far been associated with 11 distinct syndromes including the autosomal recessive diseases xeroderma pigmentosum (XP), Cockayne syndrome (CS) and trichothiodystrophy (TTD) (Niedernhofer et *al.*, 2011). About 50% of the XP patients are characterized by extreme sensitivity to sunlight with severe sunburn after exposure, whereas the rest develops different forms of skin pigmentation. A large fraction of the patients does also show neurological degeneration (DiGiovanna et *al.*, 2012; Knoch et *al.*, 2012; Niedernhofer et *al.*, 2011). XP patients have a 10 000 fold risk of developing non-melanoma skin cancer (Bradford et *al.*, 2011), the most prevalent human malignancy (Wehner et *al.*, 2012). The extensiveness of XP is not large, with 2.3 out of 1 000 000 people affected in the western European population (Kleijer et *al.*, 2008). XP is caused by mutations in any of eight genes with gene products named XPA through XPG and DNA polymerase η (Lehmann et *al.*, 2011).

CS is caused by mutations in the Cockayne syndrome genes CSA or CSB which participates in TC-NER, but mutations in XPB, XPD or XPG are also connected with CS. The most severe CS symptoms include mental and developmental retardation, photosensitivity and premature aging (Knoch et *al.*, 2012).

Werner (WS) and Bloom (BS) syndrome are two diseases with mutations in genes involved in DSB repair. The WRN and BLM genes are both members of the RecQ helicase family (Bohr 2010). WS patients are normal from birth, but develop aging symptoms after puberty and the median life expectancy is 47 years (Dominguez-Gerpe et *al.*, 2008; Knoch et *al.*, 2012).

Defects in MMR are also connected with rare hereditary diseases. Worldwide, more than 1 000 000 people will be diagnosed with colorectal cancer (CRC) each year. Of these patients ~3% will have the most common hereditary CRC syndrome; the Lynch syndrome (LS). LS patients usually have mutations in MMR genes *MLH1* or *MLH2* (Lynch et *al.*, 2009).

1.2 RNA editing

After transcription, the process where DNA is transcribed to RNA by RNA polymerases, RNA molecules go through post-transcriptional processes such as folding, polyadenylation, splicing and editing (Gott et *al.*, 2000). Adenosine to inosine (A-to-I) editing is the most common editing event in mammalian cells and is found in different RNA molecules like messenger RNA (mRNA) (Tang et *al.*, 2012), transfer RNA (tRNA) (Su et *al.*, 2011), non-coding RNAs like micro RNA (miRNA) (Gommans 2012) and Alu-containing transcripts (Athanasiadis et *al.*, 2004; Kim et *al.*, 2004; Levanon et *al.*, 2004). A-to-I editing might affect RNA molecules in different ways and is an important factor for the generation of the complexity found in higher eukaryotes. A-to-I editing can generate or destroy splicing sites and in exons the editing might lead to codon changes since I is read as G. Such changes have the potential to generate structurally and functionally different isoforms of proteins (Mallela et *al.*, 2012). There are only identified a few genes (~30) with site-specific editing sites in coding sequences, most in ion channels and neurotransmitter receptors, consequently most of the editing is conducted in non-coding sequences (Nishikura 2010).

1.2.1 mRNA editing

Adenosine deamination in mRNA is mediated by a family of enzymes named "adenosine deaminases that act on RNA" (ADARs). The A-to-I activity was first demonstrated by Bass and Weintraub by examination of extracts from the frog *Xenopus laevis* (Bass et *al.*, 1988) and later, four members of the ADAR family has been described; ADAR1, ADAR2, ADAR3 and testis-expressed nuclear RNA-binding protein (TENR). It is predicted that the ADAR family has evolved from tRNA adenosine deaminases (Tad1/ADAT1) by the addition of dsRNA-binding domains (Gerber et *al.*, 2001). ADAR1 was the initial protein to be purified with A-to-I activity on long dsRNA (Hough et *al.*, 1994; Kim et *al.*, 1994b; Kim et *al.*, 1994a; O'Connell et *al.*, 1994). The enzyme is omnipresent and well conserved in many organisms (Keegan et *al.*, 2004). ADAR1 is essential for embryonic development in mice, as the *Adar1*$^{-/-}$ knockout animals have massive liver disintegration and die at embryonic day 12.5 (Hartner et *al.*, 2004; Wang et *al.*, 2004). Despite great effort, the RNA substrates and non-editing functions of ADAR1, which are so crucial for animal development, are not well understood (Hogg et *al.*, 2011; Wang 2011).

In contrast to ADAR1, many specific targets are identified for ADAR2, which is mostly expressed in the central nervous system (CNS) (Hogg et *al.*, 2011). The *Adar2*$^{-/-}$ knockout mice live 20 days after birth with an ascending predisposition for epileptic seizures. One of ADAR2s substrates is the *GluR-B* transcript where one single A-to-I editing event in the mRNA changes a glutamine to arginine (the Q/R site) in the translated protein. The phenotype of the knockout mice can be completely reversed by introduction of a pre-edited version of GluR-B, demonstrating the importance of mRNA editing (Higuchi et *al.*, 2000).

By now, there is no report on ADAR3 activity, but the protein has shown to be an efficient competitor for ADAR1 and 2 *in vitro* by binding to the same transcripts and impede deamination of the substrates (Chen et *al.*, 2000).

TENR lacks conserved residues important in catalysis and is probably not active, but male *Tenr*$^{-/-}$ mice have low sperm count and are sterile, indicating a role for TENR in sperm morphogenesis (Connolly et *al.*, 2005).

1.2.2 tRNA editing

To generate mature tRNA molecules, pre-tRNA has to go through many post-transcriptional changes. Some of the tRNAs are edited in the wobble position

(position 34, the first base in the anticodon triplet) of the anticodon. A-to-I editing increases the number of possible codons that can be recognized, as I can pair with both A, C and U (figure 4).

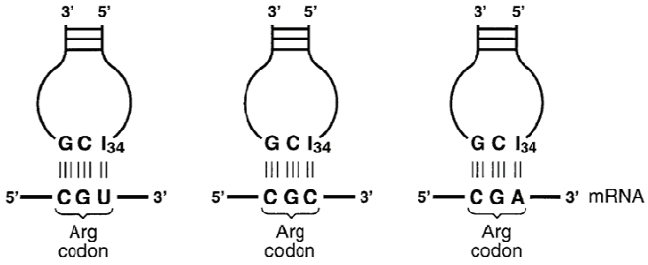


Figure 4: Anticodon arm of tRNA^{Arg}. The tRNA^{Arg} anticodon with I in the 34th position may base pair with both U, C and A in the third position of the corresponding codon **(Schaub et *al.*, 2002)**.

There is only one identified tRNA with I in the wobble position in prokaryotes, tRNA^{Arg}(ACG) which is deaminated by Tad1 (Wolf et *al.*, 2002). In higher eukaryotes, I in the 34th position, is present in eight tRNAs (seven in yeast) (Schaub et *al.*, 2002; Su et *al.*, 2011) and the deamination is accomplished by the heterodimeric ADAT2-ADAT3 complex (Gerber et *al.*, 1999).

1.2.3 miRNA editing

RNA interference (RNAi) is the process whereby RNA molecules regulate protein expression via destruction or inhibition of mRNA molecules. One type of RNA molecule involved in RNAi is miRNA (Mallela et *al.*, 2012). There are thousands of annotated miRNA genes and it is shown that one miRNA can affect hundreds of protein encoding genes and consequently a vast number of genes might be controlled by miRNAs (Selbach et *al.*, 2008). The transcribed pre-miRNA folds into several double-stranded species that are cut into ~75 nucleotides long hairpins by Drosha nuclease before they are exported to the cytoplasm. Further processing is performed by the endoribonuclease Dicer to form 20-24 nucleotides long duplexes. One strand is incorporated into the RNA-induced silencing complex (RISC) and is then used to silence complementary mRNAs. The double-stranded stem-loop structures of pre-miRNA is a good substrate for the ADAR enzymes and it has been predicted that 16% of pre-miRNAs in human brain are edited (Kawahara et *al.*, 2008) A-to-I editing of

pre-miRNAs might change the dsRNA structure as I:U base pair are less stable than A:U. The editing might therefore inhibit the Drosha processing and consequently reduce the number of mature miRNAs (Yang et *al.*, 2006). Editing of miRNAs can also increase the number of substrates each miRNA molecule can target because of changed base pairing properties (Gommans 2012; Hogg et *al.*, 2011; Nishikura 2010).

One of the components of the RISC complex is Tudor staphylococcal nuclease (Tudor-SN) (Caudy et *al.*, 2003). Tudor-SN is shown to interact with A-to-I hyper-edited double-stranded RNA (dsRNA) molecules and promote cleavage at inosines in I:U pairs of dsRNA and hence be involved in RNAi (Li et *al.*, 2008; Scadden et *al.*, 2005; Scadden 2005). Both Tudor-SN and ADAR1 has been shown to localize to stress granules after oxidative stress in mammalian cells (Weissbach et *al.*, 2012). A possible link between Tudor-SN and endonuclease V will be discussed in section 3.

1.2.4 Hyperediting of long dsRNA

Long dsRNA molecules are potent substrates for the ADAR enzymes and it is shown that up to 50% of the adenosine residues might be edited (hyper-editing) *in vitro* (Bass et *al.*, 1988; Nishikura et *al.*, 1991). Most of the ADAR editing found in mammals are in non-coding sequences like Alu elements (Levanon et *al.*, 2004). Alu repeat elements are exclusively found in primates in introns and untranslated regions (UTRs). The non-coding sequences are ~300 nucleotides and form long dsRNA structures with Alu elements oriented in opposite directions (Sie et *al.*, 2011). A transcriptome-wide investigation showed that 44% of all expressed transcripts contained Alu sequences, either partial or full-length. Alu elements belong to a group of genetic elements which have the possibility to amplify themselves, so called retrotransposons. Retrotransposons are able to affect gene expression in various ways, e.g. by making new transcription start sites or polyadenylation sites, introduce binding sites for miRNA or RNA binding proteins or interfere with splicing patterns if inserted in an intron (Moolhuijzen et *al.*, 2010). About 10% of the human genome consists of Alu elements (Batzer et *al.*, 2002) and Alu-containing transcripts are good substrates for ADAR enzymes (Athanasiadis et *al.*, 2004).

Viral RNA is also believed to be targeted by ADARs. ADAR editing patterns have been found in the RNA of viruses like hepatitis delta virus, human immunodeficiency virus, measles virus and influenza virus. Editing of viral RNA has

been shown to be both pro- and antiviral depending on the combination of host and virus [reviewed in (Samuel 2011)].

1.2.4 Human disease – consequences of editing defects

Dysfunctional A-to-I editing is associated with different human diseases. There are reports on diseases with abnormal editing at specific sites or deviant editing patterns of more general character. One example is the under-editing of Alu-containing transcripts that have been reported in tumors from brain, prostate, lung, kidney and testis (Paz et *al.*, 2007). Another example is from a recent study which connects dysfunctional expression of ADAR1 (overexpression) and ADAR2 (downregulation) to one of the most common types of cancer; human hepatocellular carcinoma (HCC) (Chan et *al.*, 2013). A specific editing site has been identified in patients with the neurological disease sporadic amyotrophic lateral sclerosis (ALS), where underediting of pre-mRNAs of the Q/R site in glutamate receptor GluR-B has been connected with death of motor neurons (Kawahara et *al.*, 2004). A-to-I editing is also connected with psychological dysfunction. The serotonin receptor 5-HT_{2C}R pre-mRNA have five A-to-I editing sites within the coding sequence called A, B, C', C and D. The editing patterns have been investigated in suicide victims with a known history of depression. Among the findings were increased editing of site C' and decreased editing of site D. Moreover, mice treated with antidepressant drug (Prozac^{TM}) showed opposite editing patterns than the suicide victims. Together this links A-to-I editing of 5-HT_{2C}R receptor to depression (Gurevich et *al.*, 2002).

1.3 Endonuclease V

1.3.1 *Escherichia coli* endonuclease V

Endonuclease V (EndoV) was first identified in *E. coli* by Gates and Linn more than three decades ago. The enzyme was described as a possible DNA repair enzyme and *E. coli* EndoV (EcEndoV) was found to be active against DNA treated with UV-light or under acidic conditions after osmium tetroxide (OsO_4) exposure (Gates et *al.*, 1977). Years later, a deoxyinosine 3' endonuclease with *in vitro* affinity for hypoxanthine, uracil, AP-sites, urea, base mismatches and flap and pseudo-Y structures was described by Kow and colleagues (Yao et *al.*, 1994a; Yao et *al.*, 1994b;

Yao et *al.*, 1994c; Yao et *al.*, 1995; Yao et *al.*, 1996; Yao et *al.*, 1997). This versatile enzyme was identified as EndoV (Guo et *al.*, 1997; Yao et *al.*, 1997) and hypoxanthine was considered to be the main substrate (Weiss 2008; Yao et *al.*, 1997). The hypothesis of hypoxanthine as the main substrate is emphasized by the fact that the EcEndoV mutant (*nfi*$^-$) displays a increased mutation frequency after exposure to nitrous acid compared to the wild type (Guo et *al.*, 1998; Schouten et *al.*, 1999). EcEndoV has also been shown to recognize the deamination products xanthosine (He et *al.*, 2000) and oxanosine (Hitchcock et *al.*, 2004). EndoV is dependent on a divalent metal ion like Mg^{2+} or Mn^{2+} and pH between 6.0 and 9.5 to be active (Yao et *al.*, 1994b; Yao et *al.*, 1997), and cleaves the second phosphodiester bond 3' to the lesion, leaving a 3' hydroxyl (OH) and a 5' phosphoryl (P) group (figure 5) (Yao et *al.*, 1994b).

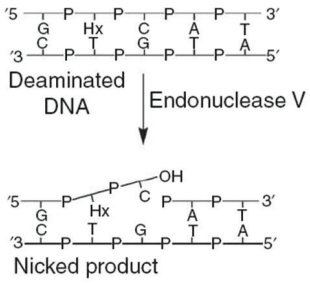


Figure 5: Incision by endonuclease V on hypoxanthine containing DNA. Endonuclease V cleaves the second phosphodiester bond 3' to the hypoxanthine (Hx) lesion leaving a 3' OH and 5' phosphoryl (P) group **(Dalhus et *al.*, 2009a)**.

After incision, EndoV remains bound to the hypoxanthine containing DNA, but do not remove the damage (Yao et *al.*, 1994b), indicating that other enzymes are involved in the EndoV initiated process. At present, the individual steps in the downstream process are not completely understood, but hypoxanthine repair can be fully reconstituted in an *in vitro* assay using recombinant EcEndoV, DNA polymerase I, the four dNTPs and DNA ligase (Lee et *al.*, 2010a). This has led to the proposal of a model where EcEndoV initiates the repair of hypoxanthine by introducing a nick in DNA followed by the 3'-5' exonuclease proofreading mechanism of polymerase I to remove at least 3 nucleotides (Lee et *al.*, 2013). Another model, where EcEndoV first cleaves 3' to the lesion followed by recruitment of downstream proteins which induce

a conformational change in EcEndoV, has been proposed (Feng et *al.*, 2005). The conformational change could possibly give EcEndoV a 3' exonuclease activity that might eliminate nucleotides from the 3' side to the 5' side of the damaged base, but this is so far just a hypothesis and is not shown *in vivo*.

1.3.2 *Thermotoga maritima* endonuclease V

EndoV from the hyperthermophilic bacterium *Thermotoga maritima* share many characteristics with EcEndoV including affinity for hypoxanthine, oxanine, mismatch, AP-site and uracil (Hitchcock et *al.*, 2004; Huang et *al.*, 2001; Mi et *al.*, 2011). The crystal structure of TmEndoV in complex with DNA was solved in 2009 (figure 6) (Dalhus et *al.*, 2009a).

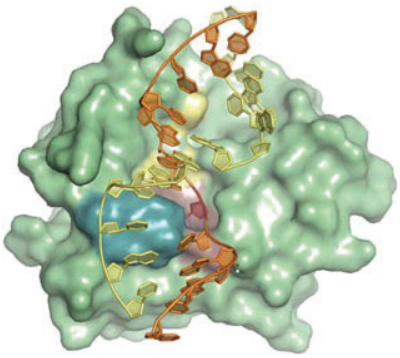


Figure 6: Crystal structure of *Thermotoga maritima* Endonuclease V. *Thermotoga maritima* endonuclease V in complex with DNA with a hypoxanthine lesion (red base) in the recognition pocket. The strand separating wedge is shown in cyan, the site which binds the metal ion and cuts the phosphodiester bond in yellow and the recognition pocket is shown in pink **(Dalhus et *al.*, 2009a)**.

The enzyme is similar to the RNase H-like superfamily of proteins with respect to a RNase H-like motif (Dalhus et *al.*, 2009a), including RNase H from *E. coli* (Katayanagi et *al.*, 1990; Yang et *al.*, 1990), Holliday junction resolvase RuvC (Ariyoshi et *al.*, 1994) and the PIWI domain of *Pyrococcus furiosus* Argonaute (Song et *al.*, 2004). TmEndoV has a conserved wedge motif; PYIP (Pro79, Tyr80, Ile81 and Pro82) that act as a minor groove damage sensor and can separate the two DNA strands at the lesion. The hypoxanthine base is flipped ~90 ° into a pocket and the enzyme incises the second phosphodiester bond 3' to the damage, using a divalent metal ion bound in the active site next to the base recognition pocket. Following

cleavage, the DNA ends are stabilized by tight binding to the pocket and hydrogen bonds with surrounding residues (Dalhus et *al.*, 2009a).

1.3.3 Mammalian endonuclease V

Even if prokaryotic EndoVs have undergone extensive research for many years, little is known about the eukaryotic orthologues. There is one report on mouse EndoV (Moe et *al.*, 2003) and one on the human orthologue (Mi et *al.*, 2012), both indicating a similar function as the prokaryotic homologues with incision activity at hypoxanthine in DNA in *in vitro* assays. Furthermore, Kuraoka and co-workers published a paper on hEndoV back-to-back with paper II (Morita et *al.*, 2013). These findings will be revised in section 3.

2. Present investigation

2.1 Aims of the study

Endonuclease V is conserved through all domains of life and has been under thorough research. The prokaryotic orthologues are known to incise the second phosphodiester bond 3' to deaminated adenine in DNA, initiating DNA repair. Prokaryotic EndoV has also been shown to recognize other DNA damage such as mismatches, uracil, AP-sites and many DNA structures with helical distortions. Despite this knowledge, little is known about the eukaryotic orthologues. The main goals of this doctoral thesis have been to increase our knowledge about this enigmatic enzyme with respect to substrate specificity, structural features and *in vivo* function.

Previous studies have revealed only weak nicking activity on DNA substrates containing hypoxanthine for hEndoV. We initiated our study with a broader characterization of hEndoV. In paper I, we used bioinformatics tools, analyses of gene expression in different cell lines together with biochemical assays to characterize hEndoV.

We did not find any activity for hEndoV on DNA with hypoxanthine (paper I), suggesting that we were not using the correct substrate. As inosine also is present in RNA, we changed the focus to RNA in paper II.

The highly conserved wedge motif of TmEndoV has previously been shown to separate the two DNA strands at the inosine lesion and is therefore believed to be important for recognition. To broaden the understanding of this wedge motif, we aimed to determine the structure of TmEndoV in complex with DNA substrates with helical distortions. The crystal structure of TmEndoV was solved in complex with a one-nucleotide loop in paper III.

2.2 Abstract of papers

Paper I: The human homolog of Escherichia coli endonuclease V is a nucleolar protein with affinity for branched DNA structures

Loss of amino groups from adenines in DNA results in the formation of hypoxanthine bases with miscoding properties. In *Escherichia coli*, the primary enzyme for initiation of DNA repair at deaminated adenine is endonuclease V (EndoV), encoded by the *nfi* gene. Endonuclease V orthologues are widespread in nature and belong to a family of highly conserved proteins. Whereas prokaryotic EndoV enzymes are well characterized, the function of the eukaryotic homologues remains obscure. Here we describe the human EndoV orthologue and show with bioinformatics and experimental analyses that a large number of transcript variants exist for the human endonuclease V gene (*ENDOV*), many of which are unlikely to be translated into functional proteins. Full-length hEndoV is encoded by 8 evolutionary conserved exons covering the core region of the enzyme, in addition to one or two 3'-exons encoding an unstructured and poorly conserved C-terminus. In contrast to the *E. coli* enzyme, we find recombinant hEndoV neither to incise nor bind inosine-containing DNA. While both enzymes have strong affinity for several branched DNA substrates, cleavage is observed only with *E. coli* EndoV. We find that hEndoV is localized in the cytoplasm and nucleoli of human cells. As nucleoli harbor the rRNA genes, this may suggest a role for the protein in rRNA gene metabolism such as DNA replication or RNA transcription.

Paper II: Endonuclease V cleaves at inosines in RNA

Endonuclease V orthologues are highly conserved proteins found in all kingdoms of life. While the prokaryotic enzymes are DNA repair proteins for removal of deaminated adenosine (inosine) from the genome, no clear role for the eukaryotic counterparts has hitherto been described. Here we report that human endonuclease V and also *Escherichia coli* endonuclease V are highly active ribonucleases specific for inosine in RNA. Inosines are normal residues in certain RNAs introduced by specific deaminases. Adenosine-to-inosine editing is essential for proper function of these transcripts and defects are linked to various human disease. Here we show that human hEndoV cleaves an RNA substrate containing inosine in a position corresponding to a biologically important site for deamination in the Gabra-3 transcript of the $GABA_A$

neurotransmitter. Further, human hEndoV specifically incises transfer RNAs with inosine in the wobble position. This previously unknown RNA incision activity may suggest a role for endonuclease V in normal RNA metabolism.

Paper III: Structural basis of DNA loop recognition by endonuclease V

The DNA repair enzyme endonuclease V recognizes and cleaves DNA at deaminated adenine lesions (hypoxanthine). In addition, EndoV cleaves DNA containing various helical distortions such as loops, hairpins, and flaps. To understand the molecular basis of EndoV's ability to recognize and incise DNA structures with helical distortions, we solved the crystal structure of *Thermotoga maritima* EndoV in complex with DNA containing a one-nucleotide loop. The structure shows that a strand-separating wedge is crucial for DNA loop recognition, with DNA strands separated precisely at the helical distortion. The additional nucleotide forming the loop rests on the surface of the wedge, while the normal adenine opposite the loop is flipped into a base recognition pocket. Our data show a different principle for DNA loop recognition and cleavage by EndoV, in which a coordinated action of a DNA-intercalating wedge and a base pocket accommodating a flipped normal base facilitate strand incision.

3. Results and discussion

3.1 Characterization of human endonuclease V transcripts

EndoV is regarded as the principal enzyme for repair of hypoxanthine in DNA in *E. coli* (Guo et *al.*, 1997; Guo et *al.*, 1998; Yao et *al.*, 1994b; Yao et *al.*, 1995). EcEndoV has also affinity for a range of substrates like uracil in DNA, AP-sites, urea residues, base mismatches, different DNA structures and helical distortions (Yao et *al.*, 1994a; Yao et *al.*, 1994b; Yao et *al.*, 1994c; Yao et *al.*, 1996; Yao et *al.*, 1997). The eukaryotic orthologues are on the other hand poorly described with only one published paper for the mouse orthologue (Moe et *al.*, 2003) and one for the human orthologue (Mi et *al.*, 2012). Both show only weak affinity for EndoV on deaminated adenines. Bioinformatics searches in available public databases (paper I) show that EndoV orthologues are present in most plants and green algae, vertebrates and sponges, but are apparently completely missing in insects. EndoV is found in the fungi *Schizosaccharomyces pombe* (*S. pombe*), but is almost absent in all other fungi like *Saccharomyces cerevisiae*. Even if the EndoV gene (*nfi*) is missing in some eukaryotic species, it is highly conserved among those who have the gene, and therefore an important function is likely.

Our bioinformatics searches indicate that there is a high degree of alternative splicing of the human orthologues. The mRNA consists of up to ten exons and alternative splicing in both the 5' and 3' ends leads to many different isoforms, e.g. a version with long exon 9 and without exon 10 and another with short exon 9 followed by exon 10. Bioinformatics searches revealed that most transcripts were without exon 3, something that we confirmed in human fibroblasts, kidney and colon cell lines (paper I). Exon 3 is probably essential as the exon makes up a significant fraction of EndoVs core domain and includes residues involved in damage recognition, DNA strand cleavage and the strand separating wedge motif (Dalhus et *al.*, 2009a). The redundant share of exon 3-lacking hEndoV transcripts is difficult to explain as they would probably not be translated and folded into functional proteins. However, if hEndoV without exon 3 is translated, we cannot rule out that this form of the enzyme might have a function different from what we know about prokaryotic EndoV, e.g. serveing as a signaling or recruitment factor. Attempts to produce recombinant

hEndoV without exon 3 were not successful. We have tried to precipitate endogenous hEndoV to identify which of the isoforms that are produced. The endogenous level of hEndoV appears to be low and so far we have not succeeded to precipitate the enzyme.

3.2 Biochemical properties of endonuclease V

EcEndoV mediated repair of hypoxanthine in DNA has been reconstituted *in vitro* with purified EcEndoV, DNA polymerase I, DNA ligase and the four dNTPs (Lee et *al.*, 2010a). The reconstitution does not exclude the need for other downstream proteins as one could potentially repair many DNA damage by mixing a lesion-specific DNA repair enzyme with a suitable endo- or exonuclease, DNA polymerase, DNA ligase and dNTPs. We have not found any robust activity for hEndoV on DNA substrates with hypoxanthine, neither on uracil or AP-sites which also are substrates for EcEndoV (Gates et *al.*, 1977; Gates, III et *al.*, 1977; Yao et *al.*, 1994a). What we did find however, was a strong affinity for various DNA substrates with helical distortions: 3'-flap, 5'-flap, 3-way junction, fork, pseudo-Y and Holliday junction. Albeit no endonuclease activity for these substrates were detected (paper I). The lack of endonuclease activity could be due to wrong *in vitro* conditions like pH, ions or substrate or because of missing cofactors, protein partners or posttranslational modifications. EndoV shares a RNase H-like motif with the yeast Holliday junction resolvase RuvC (Ariyoshi et *al.*, 1994; Dalhus et *al.*, 2009a) and the affinity for branched DNA substrates might indicate a function in recombination repair, but so far, no other evidence points in that direction.

3.3 Impact of endonuclease V in hypoxanthine repair

Prokaryotic EndoV has catalytic activity for hypoxanthine *in vitro* and is regarded the principle enzyme for hypoxanthine repair, but the biological role remains ambiguous. One of the strongest evidence for EndoV to be a repair enzyme is the increase in A-T to G-C transition mutations after nitrous acid exposure in an *nfi* mutant in *E. coli* (Schouten et *al.*, 1999). Other reports show only modest to low mutation frequencies after stress by deaminating agents and the *nfi* mutant does not display increased spontaneous mutation frequency (Guo et *al.*, 1998; Moe et *al.*, 2003; Weiss 2001). It

is not surprising that the *nfi* mutant does not have a strong mutator phenotype. First, there is a low number of hypoxanthine in DNA under normal cellular conditions in *E. coli*, with only ~1.2 dI pr 10^6 nucleotides which equals ~5 dI pr genome (Pang et *al.*, 2012). Second, other enzymes are active on hypoxanthine, like the DNA repair enzymes Mug (O'Neill et *al.*, 2003) and AlkA which also recognize hypoxanthine in DNA (Saparbaev et *al.*, 1994; Terato et *al.*, 2002). EndoV, AlkA and Mug could have an overlapping function that contributes to the non-mutagenic phenotype of the *nfi* mutant. It would be interesting to see if mutations accumulate in a triple mutant with *nfi*, *alkA* and *Mug*.

Another source of hypoxanthine in DNA is incorporation of deaminated dATP during DNA replication. At least *in vitro*, dITP is preferentially incorporated opposite cytosine (Bessman et *al.*, 1958) and since hypoxanthine is interpreted as a guanine, this incorporation is non-mutagenic. The non-mutagenity of dITP incorporation is demonstrated in *E. coli* strains with mutations in *nfi* and *rdgB* genes. *RdgB* encodes an ITPase which cleans out dITP from the precursor pool. The double mutant does not show any elevated levels of neither spontaneous nor nitrous acid-induced mutations (Budke et *al.*, 2006). In untreated *E. coli* cells, the *rdgB* single mutant has a ~10 fold increase in the levels of hypoxanthine in DNA whereas the *nfi* mutant is comparable with the wild type. The double mutant does neither have increased levels of deoxyinosine in DNA nor inosine in RNA compared to the single mutants (Pang et *al.*, 2012). In survival assays, the *nfi* mutant exhibits increased survival compared to the wild type after nitrous acid treatment (Guo et *al.*, 1998). The enhanced survival might be connected to the strand breaks EndoV makes at hypoxanthine. The lethality of EndoV mediated strand breaks is further demonstrated in an *E. coli* mutant where *rdgB* is mutated in combination with recombination repair genes *recA* or *recBC*. Both *rdgB recA (*Clyman et *al.,* 1987*)* and *rdgB recBC* double mutant cells are inviable and *rdgB recBC* has increased level of chromosomal fragmentation. The lethality can be suppressed by deleting *nfi* in the two double mutants (Bradshaw et *al.*, 2003). The lethality and chromosomal fragmentation may be the consequence of the elevated level of EndoV induced strand breaks at hypoxanthines that are incorporated into DNA due to the *rdgB* mutation and cannot be repaired in cells with impaired recombination repair. One could imagine it would be beneficial for the cells if EndoV could discriminate between hypoxanthines arising from deamination of adenines and hypoxanthines which have been incorporated into DNA. EcEndoV is equally active

on I:T as I:C substrates *in vitro* (He et *al.*, 2000)(unpublished data) and together with the *nfi* suppression of the lethality of the *rdgB recA* and *rdgB recBC* double mutants, it does not seem like EndoV is able to do such a discrimination.

In summary, the biological function of EcEndoV and its potential role in repair of hypoxanthine still remains unclear as the *nfi* mutant has no distinct phenotype. The possible association of EndoV with chromosomal fragmentation is interesting and could possibly be viewed together with the findings of Foti and colleagues. They have found that overexpression of the translesion polymerase DinB (polymerase IV) is cytotoxic in *E. coli* cells (Foti et *al.*, 2012). Din B incorporates 8-oxodeoxyguanosine triphosphate (8-oxodGTP) more frequently than other polymerases. Overproduction of DinB is lethal as more 8-oxodGTPs are incorporated than the cell can handle by DNA repair. Closely spaced 8-oxoG lesions may disrupt the possibility for proper base excision repair and consequently more lethal DSBs will be generated. The involvement of EcEndoV in chromosomal fragmentation and cell death could become more evident by genetic studies of DinB or other translesion polymerases together with EndoV, RdgB and recombination repair proteins.

Despite our results, the role for hEndoV in hypoxanthine DNA repair is still unclear and taken together with the uncertain activity shown by others, we aimed to discover other, possibly better, substrates for hEndoV.

3.4 The search for a new substrate for endonuclease V

The localization of hEndoV to cytoplasm and nucleoli (paper I) and the structural similarities with the RNase H-like superfamily (Dalhus et *al.*, 2009a) led our focus to RNA. Inosine is the most common RNA modification and adenosine are deaminated to inosine by ADAT in tRNA and by ADAR in mRNA and different non-coding RNAs [reviewed in (Gerber et *al.*, 2001; Sie et *al.*, 2011) respectively]. Nucleoli are structural compartments, composed of proteins and nucleic acids and they are found in the nucleus of eukaryotic cells. Even if nucleoli mostly consist of ribosomal RNA (rRNA), where inosines are unusual (Cantara et *al.*, 2011), hEndoV could have a function in nucleoli, e.g. in relationship with rRNA transcription. In paper II, a clear association between hEndoV and RNA was found. Both EcEndoV and hEndoV efficiently cleaved ss- and dsRNA with inosine under reaction conditions where

hEndoV did not cleave corresponding DNA substrates. The activity for EndoV on RNA suggests a so far unknown function in RNA metabolism. Our results concur with the results of Isao Kuraoka and his group who also demonstrate that EndoV cleaves RNA with inosine (Morita et *al.*, 2013). Currently, the biological function for EndoV is yet not understood, but there are several potential RNA substrates for EndoV which will be discussed in the following section.

3.5 Potential *in vivo* substrates for endonuclease V

Adenosine in the wobble position of the anticodon loop of tRNAs (position 34) can be deaminated to increase the number of possible codons a specific tRNA can base pair with during translation (Gustilo et *al.*, 2008). In procaryotes, $tRNA^{Arg}$ is the only identified tRNA with inosine in the wobble position (Wolf et *al.*, 2002), whereas eight different tRNAs have this modification in higher eukaryotes (seven in yeast) (Schaub et *al.*, 2002; Su et *al.*, 2011). We show that both EcEndoV and hEndoV cleaves a $tRNA^{Arg}$ substrate with I in the wobble position and we also demonstrate cleavage of $tRNA^{Ser}$, $tRNA^{Leu}$ and $tRNA^{Arg}$ from total tRNA isolated from human cell extracts, all having A/I in position 34 (paper II). Cleavage of the anti-codon loop of tRNAs have been reported in both prokaryotic and eukaryotic organisms as a response to stress [reviewed in (Phizicky et *al.*, 2010; Thompson et *al.*, 2009)]. To our knowledge there is not published anything on cleavage of inosine in tRNAs. tRNAs with inosine in the wobble position could be a target for EndoV, but we cannot rule out that EndoV has a more universal function in degradation of deaminated RNA transcripts, like the RNA quality control function that has been attributed to some DNA repair proteins; AlkB/ABH3 (Aas et *al.*, 2003), SMUG1 (Jobert et *al.*, 2013), APE1 (Berquist et *al.*, 2008; Vascotto et *al.*, 2009) and TDP2 (Virgen-Slane et *al.*, 2012).

Messenger RNA deamination is one factor which contributes to the complexity of higher eukaryotes. The mouse neurotransmitter $GABA_A$ is a chloride-permeable receptor which is formed from at least 16 different subunits. $GABA_A$ receptors are regarded the main mediators of fast inhibitory neurotransmissions in mammalian CNS. The Gabra-3 transcript of $GABA_A$ codes for the α3 subunit and has an A-to-I editing site, referred to as the I/M site (Ohlson et *al.*, 2007). The I/M site of Gabra-3 is conserved from chicken to human and is believed to be important in

balancing the ratios of α1 and α3 subunits and their location during development (Daniel et *al.*, 2011). We made an oligonucleotide of a part of the Gabra-3 transcript, harboring the I/M site, and found EndoV to incise the substrate (paper II). This result could point to a function for hEndoV in counteracting ADAR deamination by destruction of deaminated transcripts. However, more data is needed to strengthen this hypothesis. Many of the ADAR targets are found in the CNS, and editing of a single site could be decisive for e.g. ion channel permeability wherein the A-to-I editing could work as an on-off switch. Because of the significance of A-to-I editing in CNS, it is of course of major importance to have a proper balance of edited versus non-edited forms of transcripts and hEndoV could have an important role fine-tuning this process.

There is one known protein that interacts with I in RNA; Tudor-SN. Tudor-SN has been shown to be important for cleavage of A-to-I hyper-edited dsRNAs in extracts from *Xenopus laevis* (*X. laevis*) and HeLa cells (Scadden et *al.*, 2001; Scadden et *al.*, 2005; Scadden 2005), but it is not known which enzyme that catalyzes the reaction. We suggest that hEndoV is the catalytic active enzyme. In paper II, we demonstrate that the catalytic activity of hEndoV on deoxyinosine is dependent on a ribonucleotide 3' to the lesion. The same substrate is also used by Scadden and O'Connel to demonstrate nicking in *X. laevis* extracts (Scadden et *al.*, 2005). Another point is that cleavage activity on hyper-edited RNA substrate is stimulated by the addition of recombinant Tudor-SN to the extracts (Scadden 2005). We see a 2-3 fold increase in hEndoV activity on RNA with inosine with the addition of equimolar amounts of Tudor-SN (unpublished data) and similar results are also demonstrated by Morita and colleagues (Morita et *al.*, 2013). In addition, the ribonuclease activity found in HeLa extracts is limited to the cytoplasmic fraction (Scadden et *al.*, 2001) and Tudor-SN was mainly found to be present in cytoplasm in cells from both *Drosophila melongaster* and human (Caudy et *al.*, 2003). The fractionated activity and cellular localization fits with (Morita et *al.*, 2013) and our localization data which show that hEndoV predominantly localizes to cytoplasm (paper I).

Different stress conditions, like oxidative stress, viral infections and radiation might act as triggers for recruitment of stress response proteins like Tudor-SN. The stress will activate a response and reprogram the translational machinery in the cell to only express genes important for survival. mRNAs from fundamental "house-keeping" genes will be assigned different functions and assemble at cytoplasmic foci

called stress granules (SG) (Anderson et *al.*, 2002; Kedersha et *al.*, 2002) together with close to hundred different proteins (Thomas et *al.*, 2011). SG are thought to be important in mRNA metabolism and could be decisive for a mRNAs fate; storage, degradation or translation [reviewed in (Anderson et *al.*, 2009)]. Scadden and co-workers have demonstrated that Tudor-SN and ADAR1 localizes to stress granules after arsenite induced oxidative stress (Scadden 2007; Weissbach et *al.*, 2012). We have found hEndoV to colocalize with SG after arsenit treatment (figure 7) and that hEndoV also colocalization with Tudor-SN in HeLa cells (figure 8) (unpublished data).

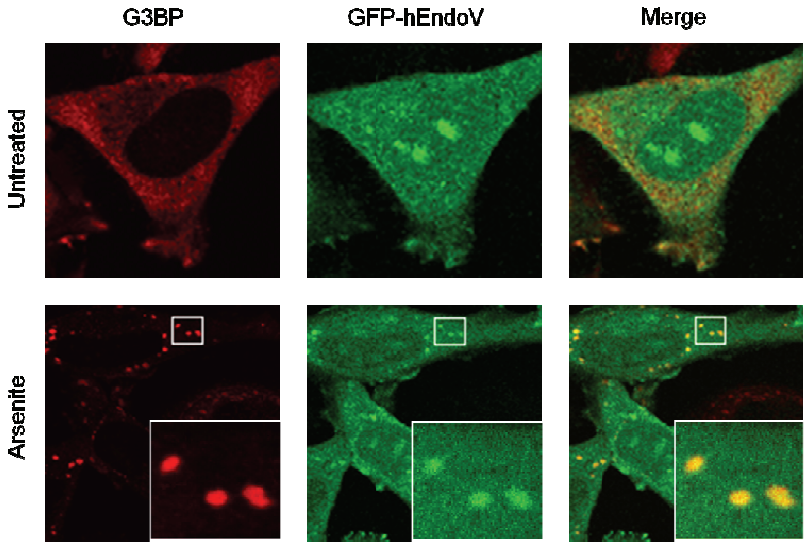


Figure 7: hEndoV colocalize with stress granules after arsenite induced stress. Confocal imaging of HeLa cells transfected with GFP-hEndoV (green) and stained with anti-G3BP (red, stress granule marker) shows colocalization (merge/yellow).

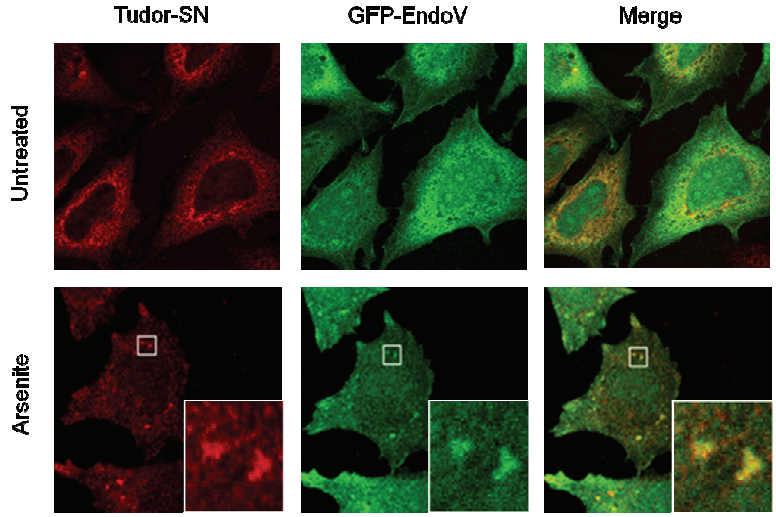


Figure 8: hEndoV colocalize with Tudor-SN after arsenite induced stress. Confocal imaging of HeLa cells transfected with GFP-hEndoV (green) and stained with anti-Tudor-SN (red) shows colocalization of the two proteins (merge/yellow).

The inosine specific ribonuclease activity we describe for hEndoV resembles the activity on hyper-edited RNA Scadden and colleagues have shown for Tudor-SN. The biochemical results together with our preliminary colocalization data on hEndoV and Tudor-SN does point to a function for hEndoV in RNA metabolism.

3.6 Structural features of endonuclease V

The structure of EndoV in complex with DNA incorporating hypoxanthine nucleobase was solved for the *T. maritima* orthologue (Dalhus et *al.*, 2009a). EndoV was found to resemble a RNase H-like superfamily fold, which among others are shared with *E. coli* RNase H (Katayanagi et *al.*, 1992; Yang et *al.*, 1990), Holliday junction resolvase RuvC (Ariyoshi et *al.*, 1994) and the PIWI domain of *Pyrococcus furiosus* Argonaut that is part of RISC (Song et *al.*, 2004). These three enzymes are involved in recognition and cleavage of DNA or RNA. The TmEndoV structure revealed a conserved PYIP motif (Pro79-Tyr80-Ile81-Pro82) with a wedge-shape, which separates the DNA strands and dislocates the base opposite the lesion. This wedge is believed to be the key in EndoVs broad substrate specificity. The unusual

cleavage site at the second phosphodiester bond 3' to lesion is shown to be due to a physical barrier between the catalytic site and the lesion recognition pocket (Dalhus et *al.*, 2009a).

IDLs might lead to frameshift mutations in DNA and they are usually repaired by MMR, but there are also reports on loop repair activity independent of MMR (Corrette-Bennett et *al.*, 1999; Corrette-Bennett et *al.*, 2001; Fang et *al.*, 2003; Littman et *al.*, 1999; McCulloch et *al.*, 2003a; McCulloch et *al.*, 2003b). With respect to EcEndoVs affinity for DNA loops (Yao et *al.*, 1996), the nick-directed repair of DNA loops observed in *E. coli* and human cells (Fang et *al.*, 2003; Littman et *al.*, 1999; McCulloch et *al.*, 2003a), is interesting. By solving the crystal structure of TmEndoV in complex with DNA containing an ID-loop, we show that the PYIP wedge separates the DNA at the loop, without disturbing the hydrogen-bonds of the neighboring G:C base pairs. An adenine base, opposite the loop, is flipped into the recognition pocket and the wedge obstructs hydrogen-bonding between the adenine and the two thymine bases (paper III). The structure of TmEndoV with loop DNA shows that the lesion recognition pocket is not restricted to bases with modifications like hypoxanthine; EndoV is also able to flip a normal base into the pocket, particularly in a context of a small helical distortion like an A:TT loop. Even if EndoV can flip undamaged bases into the pocket, the weak affinity for undamaged DNA (paper III) makes it unlikely for the enzyme to flip bases with the usual Watson-Crick base pairing into the pocket. Hence, the PYIP wedge seems to be decisive for an accurate identification of helical distortions or other inherent weak points in DNA which alters the base pairing properties, like mismatches.

The wedge motif of EndoV and the way the wedge interacts with the sugar-phosphate backbone of DNA and detects damage in a sequence-independent manner is uncommon, but not unique. DNA-binding protein 2 (DDB2) participates in NER and uses a mechanism similar to EndoV where it separates the DNA strands with a wedge-like mechanism, bends it by ~40° and flips the damaged base into a pocket (Scharer et *al.*, 2009; Scrima et *al.*, 2008). The wedges used for damage detection by EndoV and DDB2 are highly conserved among their orthologues and the wedge seems to be crucial in lesion sensoring. The structure of hEndoV has yet not been solved, but we have solved a structure of mEndoV (figure 9) (unpublished data).

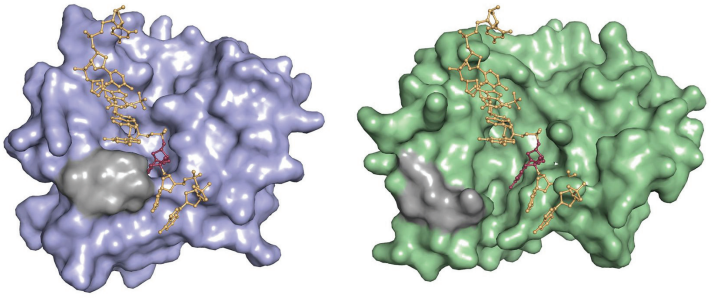


Figure 9: Molecular surfaces of the crystal structures of TmEndoV and mEndoV. The crystal structure of TmEndoV (blue), the enzyme binds an A:TT loop substrate (only the shorter strand is shown) with adenine (red), placed opposite the two thymines, in the recognition pocket. The structure of mEndoV (green) with the DNA strand superposed from the TmEndoV structure shows differences in the surface and wedge motif (gray) compared to TmEndoV.

The structure of mEndoV is solved without DNA or RNA, but it gives a good picture of similarities and differences between the eukaryotic and prokaryotic structures. In general, the two structures share many characteristics, but there are also some clear differences. mEndoV is 113 amino acids longer than TmEndoV. Most of the additional amino acids are located in the unstructured C-terminus but some are also within the enzyme core, most notable is an extra alpha-helix made out of amino acids 167-175 (figure 10).

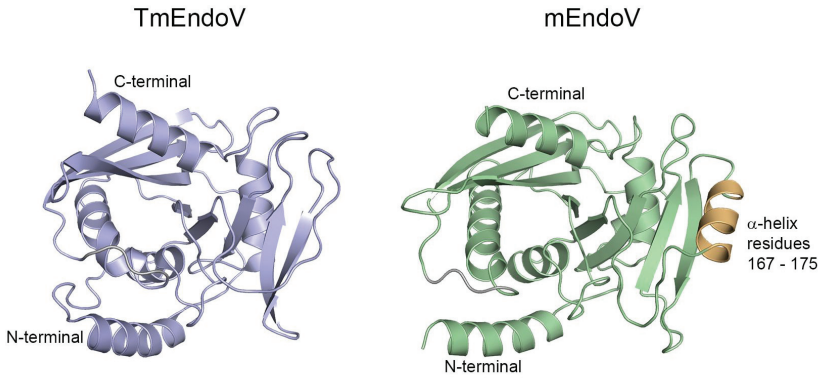


Figure 10: Protein fold of TmEndoV and mEndoV. Protein fold of TmEndoV (blue) and mEndoV (green) shows an overall similarity between the two enzymes. The most notable difference is an alpha-helix in mEndoV (yellow, amino acids 167-175).

The amino acids are much more fixed in a helix configuration and could thus serve as a domain for an interaction partner for mEndoV. The metal ion, which is placed in the active site, seems to be able to position in the same way in the two structures. The wedge motifs PYIP in TmEndoV and PYVS (Pro90-Tyr91-Val92-Ser93) in mEndoV (and hEndoV), that are part of exon 3, are positioned differently (figure 11). The aromatic side chains of amino acid Tyr80 and Pro82 in the TmEndoV wedge are facing towards the DNA strand that will be cleaved. Tyr80 fills the vacant position where the base that has been flipped into the recognition pocket used to be and binds to the phosphate backbone through hydrogen-bonding. The base 5' to the lesion stacks against Pro82 which again stacks against Tyr80 and opens the DNA duplex together with Ile81 (Dalhus et *al.*, 2009a). The wedge in the mEndoV structure has a different configuration where the aromatic side chain of Tyr91 (corresponding to Tyr80 in TmEndoV), points away from the DNA, and proline in position 82 is changed to a serine in position 93, which do not interact with Tyr91 (figure 11).

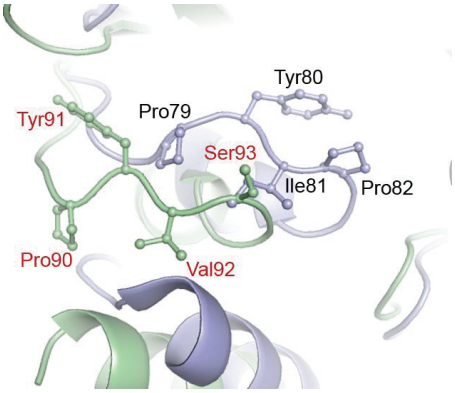


Figure 11: Close-up of the wedge in TmEndoV and mEndoV. The four residues making up the wedge in TmEndoV (blue) and mEndoV (green) shows the difference in amino acids and their respective positioning.

The difference in the wedge configuration makes it difficult to explain how mEndoV would use the wedge for damage detection and strand separation. As the structure of mEndoV is without a substrate, one cannot rule out that the enzyme will make a conformational change when it is bound to DNA or RNA. Such a conformational

change is not seen in the TmEndoV structures and is therefore not very likely for mEndoV. mEndoV prefers, as hEndoV, inosine in RNA as substrate (unpublished results). The difference in affinity for DNA and RNA between prokaryotic and eukaryotic EndoV cannot be entirely explained by the two structures, but we have proposed a model. In paper II we show that the OH-group in the ribose of the base next to the deaminated adenine is essential for activity. We also present a homology model of hEndoV where we show that the OH-group of the ribose in RNA might replace a water molecule in the active site which coordinates the Mg^{2+} that links the 3'- and 5'-ends in the incised product. If the eukaryotic orthologues lacks this water molecule, which is found in the TmEndoV structure, the OH-group would be essential for activity. However, the homology model in paper II was based upon the TmaEndoV fold, and the later mEndoV structure shows substantial differences in protein surfaces (figures 9 and 10), suggesting differences in DNA/RNA binding beyond the accuracy of the homology model.

4. Concluding remarks and future perspectives

The results presented in this thesis have broadened the knowledge of endonuclease V and given insights into new substrates, structural features and possible new functions for the enzyme. Paper I reveals that EndoV is highly conserved from bacteria to humans and that there exists several splicing variants of hEndoV. hEndoV does not seem to be active on DNA with deaminated adenine, as the prokaryotic orthologues are. Instead, hEndoV has inosine in RNA, the most common RNA edit, as substrate (paper II). The RNA activity taken together with colocalization to stress granules and Tudor-SN after oxidative stress (unpublished data), points out a new function for eukaryotic EndoV. Prokaryotic EndoV are involved in DNA repair, but our results suggest a role in RNA metabolism for the eukaryotic orthologues.

It will of course be very interesting to pursue the possible role of eukaryotic EndoV in RNA metabolism and we have obtained the *nfi* knockout (KO) mouse, which should provide a powerful tool in the further characterization of EndoV. One of the top priorities for the future would of course be to find an *in vivo* substrate for eukaryotic EndoV. RNA sequencing of different organs from KO and wild type mice could reveal a substrate. Another approach is to cross-link EndoV with RNA, precipitate EndoV and do MS analyses to identify substrates.

As many A-to-I editing sites are in the CNS, it would be interesting to conduct behavioral studies with the KO mouse that can reveal deviating behavior when it comes to e.g. anxiety, memory or learning skills. A more general characterization of the KO mouse should also be performed where the mice could be monitored over time to reveal deviating phenotypes. We have, in our preliminary work with the KO mouse, noticed an imbalance between the sexes, where there apparently is born fewer male pups than females, these observations should will of course be followed up.

The link between hEndoV, Tudor-SN and stress granules should also be looked further into and the interaction could be confirmed by co-immunoprecipitation of cell extracts. Immunoprecipitation could also reveal other partners for hEndoV that we so far do not know of.

At present, we have not been able to identify endogenously expressed hEndoV. Since hEndoV mainly localize to cytoplasm, fractionated extracts could be obtained to get a higher concentration of the enzyme. Identification of endogenous

hEndoV would give valuable information with respect to all the different transcripts that have been identified. Since bioinformatics predict that hEndoV transcripts without exon 3 cannot be translated into functional proteins (paper I), it would be very interesting to see if both transcripts with and without exon 3 in fact are translated.

5. References

Aas, P. A., Otterlei, M., Falnes, P. O.*et al*., 2003, "Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA", *Nature*, vol. 421, no. 6925, pp. 859-863.

Acharya, S., Wilson, T., Gradia, S.*et al*., 1996, "hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6", *Proc.Natl.Acad.Sci.U.S.A*, vol. 93, no. 24, pp. 13629-13634.

Altieri, F., Grillo, C., Maceroni, M.*et al*., 2008, "DNA damage and repair: from molecular mechanisms to health implications", *Antioxid.Redox.Signal.*, vol. 10, no. 5, pp. 891-937.

Anderson, P. & Kedersha, N. 2002, "Stressful initiations", *J.Cell Sci.*, vol. 115, no. Pt 16, pp. 3227-3234.

Anderson, P. & Kedersha, N. 2009, "Stress granules", *Curr.Biol.*, vol. 19, no. 10, p. R397-R398.

Arakawa, H., Hauschild, J., & Buerstedde, J. M. 2002, "Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion", *Science*, vol. 295, no. 5558, pp. 1301-1306.

Araki, M., Masutani, C., Takemura, M.*et al*., 2001, "Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair", *J.Biol.Chem.*, vol. 276, no. 22, pp. 18665-18672.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H.*et al*., 1994, "Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from E. coli", *Cell*, vol. 78, no. 6, pp. 1063-1072.

Athanasiadis, A., Rich, A., & Maas, S. 2004, "Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome", *PLoS.Biol.*, vol. 2, no. 12, p. e391.

Au, K. G., Welsh, K., & Modrich, P. 1992, "Initiation of methyl-directed mismatch repair", *J.Biol.Chem.*, vol. 267, no. 17, pp. 12142-12148.

Bass, B. L. & Weintraub, H. 1988, "An unwinding activity that covalently modifies its double-stranded RNA substrate", *Cell*, vol. 55, no. 6, pp. 1089-1098.

Batzer, M. A. & Deininger, P. L. 2002, "Alu repeats and human genomic diversity", *Nat.Rev.Genet.*, vol. 3, no. 5, pp. 370-379.

Behmanesh, M., Sakumi, K., Abolhassani, N.*et al*., 2009, "ITPase-deficient mice show growth retardation and die before weaning", *Cell Death.Differ.*, vol. 16, no. 10, pp. 1315-1322.

Berquist, B. R., McNeill, D. R., & Wilson, D. M., III 2008, "Characterization of abasic endonuclease activity of human Ape1 on alternative substrates, as well as effects of ATP and sequence context on AP site incision", *J.Mol.Biol.*, vol. 379, no. 1, pp. 17-27.

Bessman, M. J., Lehman, I. R., Adler, J.*et al*., 1958, "Enzymatic synthesis of deoxyribonucleic acid. III. The incorporation of pyrimidine and purine analogues into deoxyribonucleic acid", *Proc.Natl.Acad.Sci.U.S.A*, vol. 44, no. 7, pp. 633-640.

Bradford, P. T., Goldstein, A. M., Tamura, D.*et al*., 2011, "Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair", *J.Med.Genet.*, vol. 48, no. 3, pp. 168-176.

Bradshaw, J. S. & Kuzminov, A. 2003, "RdgB acts to avoid chromosome fragmentation in Escherichia coli", *Mol.Microbiol.*, vol. 48, no. 6, pp. 1711-1725.

References

Bransteitter, R., Pham, P., Scharff, M. D.*et al*., 2003, "Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase", *Proc.Natl.Acad.Sci.U.S.A*, vol. 100, no. 7, pp. 4102-4107.

Budke, B. & Kuzminov, A. 2006, "Hypoxanthine incorporation is nonmutagenic in Escherichia coli", *J.Bacteriol.*, vol. 188, no. 18, pp. 6553-6560.

Burney, S., Caulfield, J. L., Niles, J. C.*et al*., 1999, "The chemistry of DNA damage from nitric oxide and peroxynitrite", *Mutat.Res.*, vol. 424, no. 1-2, pp. 37-49.

Cantara, W. A., Crain, P. F., Rozenski, J.*et al*., 2011, "The RNA Modification Database, RNAMDB: 2011 update", *Nucleic Acids Res.*, vol. 39, no. Database issue, p. D195-D201.

Caudy, A. A., Ketting, R. F., Hammond, S. M.*et al*., 2003, "A micrococcal nuclease homologue in RNAi effector complexes", *Nature*, vol. 425, no. 6956, pp. 411-414.

Chan, T. H., Lin, C. H., Qi, L.*et al*., 2013, "A disrupted RNA editing balance mediated by ADARs (Adenosine DeAminases that act on RNA) in human hepatocellular carcinoma", *Gut*.

Chapman, J. R., Taylor, M. R., & Boulton, S. J. 2012, "Playing the end game: DNA double-strand break repair pathway choice", *Mol.Cell*, vol. 47, no. 4, pp. 497-510.

Chen, C. X., Cho, D. S., Wang, Q.*et al*., 2000, "A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains", *RNA.*, vol. 6, no. 5, pp. 755-767.

Chern, C. J., MacDonald, A. B., & Morris, A. J. 1969, "Purification and properties of a nucleoside triphosphate pyrophosphohydrolase from red cells of the rabbit", *J.Biol.Chem.*, vol. 244, no. 20, pp. 5489-5495.

Clyman, J. & Cunningham, R. P. 1987, "Escherichia coli K-12 mutants in which viability is dependent on recA function", *J.Bacteriol.*, vol. 169, no. 9, pp. 4203-4210.

Collado, M., Blasco, M. A., & Serrano, M. 2007, "Cellular senescence in cancer and aging", *Cell*, vol. 130, no. 2, pp. 223-233.

Connolly, C. M., Dearth, A. T., & Braun, R. E. 2005, "Disruption of murine Tenr results in teratospermia and male infertility", *Dev.Biol.*, vol. 278, no. 1, pp. 13-21.

Corrette-Bennett, S. E., Mohlman, N. L., Rosado, Z.*et al*., 2001, "Efficient repair of large DNA loops in Saccharomyces cerevisiae", *Nucleic Acids Res.*, vol. 29, no. 20, pp. 4134-4143.

Corrette-Bennett, S. E., Parker, B. O., Mohlman, N. L.*et al*., 1999, "Correction of large mispaired DNA loops by extracts of Saccharomyces cerevisiae", *J.Biol.Chem.*, vol. 274, no. 25, pp. 17605-17611.

Dahm, R. 2008, "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research", *Hum.Genet.*, vol. 122, no. 6, pp. 565-581.

Dalhus, B., Arvai, A. S., Rosnes, I.*et al*., 2009a, "Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair", *Nat.Struct.Mol.Biol.*, vol. 16, no. 2, pp. 138-143.

Dalhus, B., Laerdahl, J. K., Backe, P. H.*et al*., 2009b, "DNA base repair--recognition and initiation of catalysis", *FEMS Microbiol.Rev.*, vol. 33, no. 6, pp. 1044-1078.

Daniel, C., Wahlstedt, H., Ohlson, J.*et al*., 2011, "Adenosine-to-inosine RNA editing affects trafficking of the gamma-aminobutyric acid type A (GABA(A)) receptor", *J.Biol.Chem.*, vol. 286, no. 3, pp. 2031-2040.

References

de Gruijl, F. R. 1999, "Skin cancer and solar UV radiation", *Eur.J.Cancer*, vol. 35, no. 14, pp. 2003-2009.

Dedon, P. C., Barth, M., Chen, B., De Mott, M., Dendroulkais, V., Dong, M., Kalinga, S., Elmquist, E., Margolin, Y., & Zhou, X. 2006, "Diverse Mechanisms of Endogenous Nucleobase Deamination in DNA and RNA," in *Advances in Molecular Toxicology*, Elsevier B.V., pp. 25-63.

Demple, B., Jacobsson, A., Olsson, M.*et al*., 1982, "Repair of alkylated DNA in Escherichia coli. Physical properties of O6-methylguanine-DNA methyltransferase", *J.Biol.Chem.*, vol. 257, no. 22, pp. 13776-13780.

Dianov, G. L. & Hubscher, U. 2013, "Mammalian base excision repair: the forgotten archangel", *Nucleic Acids Res.*, vol. 41, no. 6, pp. 3483-3490.

Dianov, G. L., Sleeth, K. M., Dianova, I. I.*et al*., 2003, "Repair of abasic sites in DNA", *Mutat.Res.*, vol. 531, no. 1-2, pp. 157-163.

DiGiovanna, J. J. & Kraemer, K. H. 2012, "Shining a light on xeroderma pigmentosum", *J.Invest Dermatol.*, vol. 132, no. 3 Pt 2, pp. 785-796.

Dominguez-Gerpe, L. & raujo-Vilar, D. 2008, "Prematurely aged children: molecular alterations leading to Hutchinson-Gilford progeria and Werner syndromes", *Curr.Aging Sci.*, vol. 1, no. 3, pp. 202-212.

Dong, L., Meira, L. B., Hazra, T. K.*et al*., 2008, "Oxanine DNA glycosylase activities in mammalian systems", *DNA Repair (Amst)*, vol. 7, no. 1, pp. 128-134.

Drummond, J. T., Li, G. M., Longley, M. J.*et al*., 1995, "Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells", *Science*, vol. 268, no. 5219, pp. 1909-1912.

Duncan, B. K. & Miller, J. H. 1980, "Mutagenic deamination of cytosine residues in DNA", *Nature*, vol. 287, no. 5782, pp. 560-561.

Duncan, T., Trewick, S. C., Koivisto, P.*et al*., 2002, "Reversal of DNA alkylation damage by two human dioxygenases", *Proc.Natl.Acad.Sci.U.S.A*, vol. 99, no. 26, pp. 16660-16665.

Eritja, R., Horowitz, D. M., Walker, P. A.*et al*., 1986, "Synthesis and properties of oligonucleotides containing 2'-deoxynebularine and 2'-deoxyxanthosine", *Nucleic Acids Res.*, vol. 14, no. 20, pp. 8135-8153.

Fagbemi, A. F., Orelli, B., & Scharer, O. D. 2011, "Regulation of endonuclease activity in human nucleotide excision repair", *DNA Repair (Amst)*, vol. 10, no. 7, pp. 722-729.

Falnes, P. O., Johansen, R. F., & Seeberg, E. 2002, "AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli", *Nature*, vol. 419, no. 6903, pp. 178-182.

Fang, W. H., Wang, B. J., Wang, C. H.*et al*., 2003, "DNA loop repair by Escherichia coli cell extracts", *J.Biol.Chem.*, vol. 278, no. 25, pp. 22446-22452.

Feng, H., Dong, L., Klutz, A. M.*et al*., 2005, "Defining amino acid residues involved in DNA-protein interactions and revelation of 3'-exonuclease activity in endonuclease V", *Biochemistry*, vol. 44, no. 34, pp. 11486-11495.

Foti, J. J., Devadoss, B., Winkler, J. A.*et al*., 2012, "Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics", *Science*, vol. 336, no. 6079, pp. 315-319.

Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D., Schultz, R. A., & Ellenberger, T. 2006, *DNA repair and mutagenesis*, 2nd edn, ASM Press.

References

Fukui, K. 2010, "DNA mismatch repair in eukaryotes and bacteria", *J.Nucleic Acids*, vol. 2010.

Gallinari, P. & Jiricny, J. 1996, "A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase", *Nature*, vol. 383, no. 6602, pp. 735-738.

Galperin, M. Y., Moroz, O. V., Wilson, K. S.*et al*., 2006, "House cleaning, a part of good housekeeping", *Mol.Microbiol.*, vol. 59, no. 1, pp. 5-19.

Gates, F. T. & Linn, S. 1977, "Endonuclease from Escherichia coli that acts specifically upon duplex DNA damaged by ultraviolet light, osmium tetroxide, acid, or x-rays", *J.Biol.Chem.*, vol. 252, no. 9, pp. 2802-2807.

Gates, F. T., III & Linn, S. 1977, "Endonuclease V of Escherichia coli", *J.Biol.Chem.*, vol. 252, no. 5, pp. 1647-1653.

Gates, K. S. 2009, "An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals", *Chem.Res.Toxicol.*, vol. 22, no. 11, pp. 1747-1760.

Gerber, A. P. & Keller, W. 1999, "An adenosine deaminase that generates inosine at the wobble position of tRNAs", *Science*, vol. 286, no. 5442, pp. 1146-1149.

Gerber, A. P. & Keller, W. 2001, "RNA editing by base deamination: more enzymes, more targets, new mysteries", *Trends Biochem.Sci.*, vol. 26, no. 6, pp. 376-384.

Gommans, W. M. 2012, "A-to-I editing of microRNAs: regulating the regulators?", *Semin.Cell Dev.Biol.*, vol. 23, no. 3, pp. 251-257.

Goodarzi, A. A. & Jeggo, P. A. 2013, "The repair and signaling responses to DNA double-strand breaks", *Adv.Genet.*, vol. 82, pp. 1-45.

Gott, J. M. & Emeson, R. B. 2000, "Functions and mechanisms of RNA editing", *Annu.Rev.Genet.*, vol. 34, pp. 499-531.

Guo, G., Ding, Y., & Weiss, B. 1997, "nfi, the gene for endonuclease V in Escherichia coli K-12", *J.Bacteriol.*, vol. 179, no. 2, pp. 310-316.

Guo, G. & Weiss, B. 1998, "Endonuclease V (nfi) mutant of Escherichia coli K-12", *J.Bacteriol.*, vol. 180, no. 1, pp. 46-51.

Gurevich, I., Tamir, H., Arango, V.*et al*., 2002, "Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims", *Neuron*, vol. 34, no. 3, pp. 349-356.

Gustilo, E. M., Vendeix, F. A., & Agris, P. F. 2008, "tRNA's modifications bring order to gene expression", *Curr.Opin.Microbiol.*, vol. 11, no. 2, pp. 134-140.

Hakem, R. 2008, "DNA-damage repair; the good, the bad, and the ugly", *EMBO J.*, vol. 27, no. 4, pp. 589-605.

Hartner, J. C., Schmittwolf, C., Kispert, A.*et al*., 2004, "Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1", *J.Biol.Chem.*, vol. 279, no. 6, pp. 4894-4902.

Haushalter, K. A., Todd Stukenberg, M. W., Kirschner, M. W.*et al*., 1999, "Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors", *Curr.Biol.*, vol. 9, no. 4, pp. 174-185.

He, B., Qing, H., & Kow, Y. W. 2000, "Deoxyxanthosine in DNA is repaired by Escherichia coli endonuclease V", *Mutat.Res.*, vol. 459, no. 2, pp. 109-114.

References

Hendrich, B., Hardeland, U., Ng, H. H.*et al*., 1999, "The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites", *Nature*, vol. 401, no. 6750, pp. 301-304.

Higuchi, M., Maas, S., Single, F. N.*et al*., 2000, "Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2", *Nature*, vol. 406, no. 6791, pp. 78-81.

Hill-Perkins, M., Jones, M. D., & Karran, P. 1986, "Site-specific mutagenesis in vivo by single methylated or deaminated purine bases", *Mutat.Res.*, vol. 162, no. 2, pp. 153-163.

Hitchcock, T. M., Gao, H., & Cao, W. 2004, "Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V", *Nucleic Acids Res.*, vol. 32, no. 13, pp. 4071-4080.

Hoeijmakers, J. H. 2001, "DNA repair mechanisms", *Maturitas*, vol. 38, no. 1, pp. 17-22.

Hogg, M., Paro, S., Keegan, L. P.*et al*., 2011, "RNA editing by mammalian ADARs", *Adv.Genet.*, vol. 73, pp. 87-120.

Hough, R. F. & Bass, B. L. 1994, "Purification of the Xenopus laevis double-stranded RNA adenosine deaminase", *J.Biol.Chem.*, vol. 269, no. 13, pp. 9933-9939.

Hsieh, P. & Yamane, K. 2008, "DNA mismatch repair: molecular mechanism, cancer, and ageing", *Mech.Ageing Dev.*, vol. 129, no. 7-8, pp. 391-407.

Huang, J., Lu, J., Barany, F.*et al*., 2001, "Multiple cleavage activities of endonuclease V from Thermotoga maritima: recognition and strand nicking mechanism", *Biochemistry*, vol. 40, no. 30, pp. 8738-8748.

Huen, M. S. & Chen, J. 2008, "The DNA damage response pathways: at the crossroad of protein modifications", *Cell Res.*, vol. 18, no. 1, pp. 8-16.

Jiricny, J. 2006, "The multifaceted mismatch-repair system", *Nat.Rev.Mol.Cell Biol.*, vol. 7, no. 5, pp. 335-346.

Jobert, L., Skjeldam, H. K., Dalhus, B.*et al*., 2013, "The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control", *Mol.Cell*, vol. 49, no. 2, pp. 339-345.

Kanavy, H. E. & Gerstenblith, M. R. 2011, "Ultraviolet radiation and melanoma", *Semin.Cutan.Med.Surg.*, vol. 30, no. 4, pp. 222-228.

Karran, P. & Lindahl, T. 1980, "Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus", *Biochemistry*, vol. 19, no. 26, pp. 6005-6011.

Katayanagi, K., Miyagawa, M., Matsushima, M.*et al*., 1990, "Three-dimensional structure of ribonuclease H from E. coli", *Nature*, vol. 347, no. 6290, pp. 306-309.

Katayanagi, K., Miyagawa, M., Matsushima, M.*et al*., 1992, "Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution", *J.Mol.Biol.*, vol. 223, no. 4, pp. 1029-1052.

Kawahara, Y., Ito, K., Sun, H.*et al*., 2004, "Glutamate receptors: RNA editing and death of motor neurons", *Nature*, vol. 427, no. 6977, p. 801.

Kawahara, Y., Megraw, M., Kreider, E.*et al*., 2008, "Frequency and fate of microRNA editing in human brain", *Nucleic Acids Res.*, vol. 36, no. 16, pp. 5270-5280.

Kedersha, N. & Anderson, P. 2002, "Stress granules: sites of mRNA triage that regulate mRNA stability and translatability", *Biochem.Soc.Trans.*, vol. 30, no. Pt 6, pp. 963-969.

References

Keegan, L. P., Leroy, A., Sproul, D.*et al*., 2004, "Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes", *Genome Biol.*, vol. 5, no. 2, p. 209.

Kim, D. D., Kim, T. T., Walsh, T.*et al*., 2004, "Widespread RNA editing of embedded alu elements in the human transcriptome", *Genome Res.*, vol. 14, no. 9, pp. 1719-1725.

Kim, U., Garner, T. L., Sanford, T.*et al*., 1994a, "Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts", *J.Biol.Chem.*, vol. 269, no. 18, pp. 13480-13489.

Kim, U., Wang, Y., Sanford, T.*et al*., 1994b, "Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 24, pp. 11457-11461.

Kisker, C., Kuper, J., & Van, H. B. 2013, "Prokaryotic nucleotide excision repair", *Cold Spring Harb.Perspect.Biol.*, vol. 5, no. 3, p. a012591.

Kleijer, W. J., Laugel, V., Berneburg, M.*et al*., 2008, "Incidence of DNA repair deficiency disorders in western Europe: Xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy", *DNA Repair (Amst)*, vol. 7, no. 5, pp. 744-750.

Knoch, J., Kamenisch, Y., Kubisch, C.*et al*., 2012, "Rare hereditary diseases with defects in DNA-repair", *Eur.J.Dermatol.*, vol. 22, no. 4, pp. 443-455.

Kovtun, I. V. & McMurray, C. T. 2007, "Crosstalk of DNA glycosylases with pathways other than base excision repair", *DNA Repair (Amst)*, vol. 6, no. 4, pp. 517-529.

Kow, Y. W. 2002, "Repair of deaminated bases in DNA", *Free Radic.Biol.Med.*, vol. 33, no. 7, pp. 886-893.

Kunkel, T. A. & Bebenek, K. 2000, "DNA replication fidelity", *Annu.Rev.Biochem.*, vol. 69, pp. 497-529.

Kunkel, T. A. & Erie, D. A. 2005, "DNA mismatch repair", *Annu.Rev.Biochem.*, vol. 74, pp. 681-710.

Lee, C. C., Yang, Y. C., Goodman, S. D.*et al*., 2013, "The excision of 3' penultimate errors by DNA polymerase I and its role in endonuclease V-mediated DNA repair", *DNA Repair (Amst)*.

Lee, C. C., Yang, Y. C., Goodman, S. D.*et al*., 2010a, "Endonuclease V-mediated deoxyinosine excision repair in vitro", *DNA Repair (Amst)*, vol. 9, no. 10, pp. 1073-1079.

Lee, H. W., Brice, A. R., Wright, C. B.*et al*., 2010b, "Identification of Escherichia coli mismatch-specific uracil DNA glycosylase as a robust xanthine DNA glycosylase", *J.Biol.Chem.*, vol. 285, no. 53, pp. 41483-41490.

Lehmann, A. R., McGibbon, D., & Stefanini, M. 2011, "Xeroderma pigmentosum", *Orphanet.J.Rare.Dis.*, vol. 6, p. 70.

Levanon, E. Y., Eisenberg, E., Yelin, R.*et al*., 2004, "Systematic identification of abundant A-to-I editing sites in the human transcriptome", *Nat.Biotechnol.*, vol. 22, no. 8, pp. 1001-1005.

Lewis, R. S., Tamir, S., Tannenbaum, S. R.*et al*., 1995, "Kinetic analysis of the fate of nitric oxide synthesized by macrophages in vitro", *J.Biol.Chem.*, vol. 270, no. 49, pp. 29350-29355.

Li, C. L., Yang, W. Z., Chen, Y. P.*et al*., 2008, "Structural and functional insights into human Tudor-SN, a key component linking RNA interference and editing", *Nucleic Acids Res.*, vol. 36, no. 11, pp. 3579-3589.

References

Li, G. M. 2008, "Mechanisms and functions of DNA mismatch repair", *Cell Res.*, vol. 18, no. 1, pp. 85-98.

Liakopouloul, A. & ALIVISATOS, S. G. 1964, "Distribution of nucleoside triphosphatases in human erythrocytes", *Biochim.Biophys.Acta*, vol. 89, pp. 158-161.

Lieber, M. R. 2010, "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway", *Annu.Rev.Biochem.*, vol. 79, pp. 181-211.

Lindahl, T. 1974, "An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues", *Proc.Natl.Acad.Sci.U.S.A*, vol. 71, no. 9, pp. 3649-3653.

Lindahl, T. 1993, "Instability and decay of the primary structure of DNA", *Nature*, vol. 362, no. 6422, pp. 709-715.

Lindahl, T., Demple, B., & Robins, P. 1982, "Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase", *EMBO J.*, vol. 1, no. 11, pp. 1359-1363.

Lindahl, T. & Nyberg, B. 1974, "Heat-induced deamination of cytosine residues in deoxyribonucleic acid", *Biochemistry*, vol. 13, no. 16, pp. 3405-3410.

Littman, S. J., Fang, W. H., & Modrich, P. 1999, "Repair of large insertion/deletion heterologies in human nuclear extracts is directed by a 5' single-strand break and is independent of the mismatch repair system", *J.Biol.Chem.*, vol. 274, no. 11, pp. 7474-7481.

Lucas, L. T., Gatehouse, D., Jones, G. D.*et al*., 2001, "Characterization of DNA damage at purine residues in oligonucleotides and calf thymus DNA induced by the mutagen 1-nitrosoindole-3-acetonitrile", *Chem.Res.Toxicol.*, vol. 14, no. 2, pp. 158-164.

Lucas, L. T., Gatehouse, D., & Shuker, D. E. 1999, "Efficient nitroso group transfer from N-nitrosoindoles to nucleotides and 2'-deoxyguanosine at physiological pH. A new pathway for N-nitrosocompounds to exert genotoxicity", *J.Biol.Chem.*, vol. 274, no. 26, pp. 18319-18326.

Lutsenko, E. & Bhagwat, A. S. 1999, "Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications", *Mutat.Res.*, vol. 437, no. 1, pp. 11-20.

Lynch, H. T., Lynch, P. M., Lanspa, S. J.*et al*., 2009, "Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications", *Clin.Genet.*, vol. 76, no. 1, pp. 1-18.

MacMicking, J., Xie, Q. W., & Nathan, C. 1997, "Nitric oxide and macrophage function", *Annu.Rev.Immunol.*, vol. 15, pp. 323-350.

Mallela, A. & Nishikura, K. 2012, "A-to-I editing of protein coding and noncoding RNAs", *Crit Rev.Biochem.Mol.Biol.*, vol. 47, no. 6, pp. 493-501.

McCulloch, S. D., Gu, L., & Li, G. M. 2003a, "Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells", *J.Biol.Chem.*, vol. 278, no. 6, pp. 3891-3896.

McCulloch, S. D., Gu, L., & Li, G. M. 2003b, "Nick-dependent and -independent processing of large DNA loops in human cells", *J.Biol.Chem.*, vol. 278, no. 50, pp. 50803-50809.

Mi, R., Abole, A. K., & Cao, W. 2011, "Dissecting endonuclease and exonuclease activities in endonuclease V from Thermotoga maritima", *Nucleic Acids Res.*, vol. 39, no. 2, pp. 536-544.

Mi, R., ford-Zappala, M., Kow, Y. W.*et al*., 2012, "Human endonuclease V as a repair enzyme for DNA deamination", *Mutat.Res.*, vol. 735, no. 1-2, pp. 12-18.

References

Moe, A., Ringvoll, J., Nordstrand, L. M.*et al*., 2003, "Incision at hypoxanthine residues in DNA by a mammalian homologue of the Escherichia coli antimutator enzyme endonuclease V", *Nucleic Acids Res.*, vol. 31, no. 14, pp. 3893-3900.

Moolhuijzen, P., Kulski, J. K., Dunn, D. S.*et al*., 2010, "The transcript repeat element: the human Alu sequence as a component of gene networks influencing cancer", *Funct.Integr.Genomics*.

Morita, Y., Shibutani, T., Nakanishi, N.*et al*., 2013, "Human endonuclease V is a ribonuclease specific for inosine-containing RNA", *Nat.Commun.*, vol. 4, p. 2273.

Muramatsu, M., Kinoshita, K., Fagarasan, S.*et al*., 2000, "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme", *Cell*, vol. 102, no. 5, pp. 553-563.

Neddermann, P. & Jiricny, J. 1994, "Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 5, pp. 1642-1646.

Niedernhofer, L. J., Bohr, V. A., Sander, M.*et al*., 2011, "Xeroderma pigmentosum and other diseases of human premature aging and DNA repair: molecules to patients", *Mech.Ageing Dev.*, vol. 132, no. 6-7, pp. 340-347.

Nishikura, K. 2010, "Functions and regulation of RNA editing by ADAR deaminases", *Annu.Rev.Biochem.*, vol. 79, pp. 321-349.

Nishikura, K., Yoo, C., Kim, U.*et al*., 1991, "Substrate specificity of the dsRNA unwinding/modifying activity", *EMBO J.*, vol. 10, no. 11, pp. 3523-3532.

Nouspikel, T. 2009, "DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility", *Cell Mol.Life Sci.*, vol. 66, no. 6, pp. 994-1009.

O'Connell, M. A. & Keller, W. 1994, "Purification and properties of double-stranded RNA-specific adenosine deaminase from calf thymus", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 22, pp. 10596-10600.

O'Neill, R. J., Vorob'eva, O. V., Shahbakhti, H.*et al*., 2003, "Mismatch uracil glycosylase from Escherichia coli: a general mismatch or a specific DNA glycosylase?", *J.Biol.Chem.*, vol. 278, no. 23, pp. 20526-20532.

Ohlson, J., Pedersen, J. S., Haussler, D.*et al*., 2007, "Editing modifies the GABA(A) receptor subunit alpha3", *RNA.*, vol. 13, no. 5, pp. 698-703.

Ohshima, H. & Bartsch, H. 1994, "Chronic infections and inflammatory processes as cancer risk factors: possible role of nitric oxide in carcinogenesis", *Mutat.Res.*, vol. 305, no. 2, pp. 253-264.

Olsson, M. & Lindahl, T. 1980, "Repair of alkylated DNA in Escherichia coli. Methyl group transfer from O6-methylguanine to a protein cysteine residue", *J.Biol.Chem.*, vol. 255, no. 22, pp. 10569-10571.

Palombo, F., Gallinari, P., Iaccarino, I.*et al*., 1995, "GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells", *Science*, vol. 268, no. 5219, pp. 1912-1914.

Pang, B., McFaline, J. L., Burgis, N. E.*et al*., 2012, "Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA", *Proc.Natl.Acad.Sci.U.S.A*, vol. 109, no. 7, pp. 2319-2324.

Paz, N., Levanon, E. Y., Amariglio, N.*et al*., 2007, "Altered adenosine-to-inosine RNA editing in human cancer", *Genome Res.*, vol. 17, no. 11, pp. 1586-1595.

References

Pena-Diaz, J. & Jiricny, J. 2012, "Mammalian mismatch repair: error-free or error-prone?", *Trends Biochem.Sci.*, vol. 37, no. 5, pp. 206-214.

Phizicky, E. M. & Hopper, A. K. 2010, "tRNA biology charges to the front", *Genes Dev.*, vol. 24, no. 17, pp. 1832-1860.

Postel-Vinay, S., Vanhecke, E., Olaussen, K. A.*et al*., 2012, "The potential of exploiting DNA-repair defects for optimizing lung cancer treatment", *Nat.Rev.Clin.Oncol.*, vol. 9, no. 3, pp. 144-155.

Revy, P., Muto, T., Levy, Y.*et al*., 2000, "Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)", *Cell*, vol. 102, no. 5, pp. 565-575.

Samuel, C. E. 2011, "Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral", *Virology*, vol. 411, no. 2, pp. 180-193.

Saparbaev, M. & Laval, J. 1994, "Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 13, pp. 5873-5877.

Scadden, A. D. 2005, "The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage", *Nat.Struct.Mol.Biol.*, vol. 12, no. 6, pp. 489-496.

Scadden, A. D. 2007, "Inosine-containing dsRNA binds a stress-granule-like complex and downregulates gene expression in trans", *Mol.Cell*, vol. 28, no. 3, pp. 491-500.

Scadden, A. D. & O'Connell, M. A. 2005, "Cleavage of dsRNAs hyper-edited by ADARs occurs at preferred editing sites", *Nucleic Acids Res.*, vol. 33, no. 18, pp. 5954-5964.

Scadden, A. D. & Smith, C. W. 2001, "Specific cleavage of hyper-edited dsRNAs", *EMBO J.*, vol. 20, no. 15, pp. 4243-4252.

Scharer, O. D. & Campbell, A. J. 2009, "Wedging out DNA damage", *Nat.Struct.Mol.Biol.*, vol. 16, no. 2, pp. 102-104.

Schaub, M. & Keller, W. 2002, "RNA editing by adenosine deaminases generates RNA and protein diversity", *Biochimie*, vol. 84, no. 8, pp. 791-803.

Schouten, K. A. & Weiss, B. 1999, "Endonuclease V protects Escherichia coli against specific mutations caused by nitrous acid", *Mutat.Res.*, vol. 435, no. 3, pp. 245-254.

Scrima, A., Konickova, R., Czyzewski, B. K.*et al*., 2008, "Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex", *Cell*, vol. 135, no. 7, pp. 1213-1223.

Selbach, M., Schwanhausser, B., Thierfelder, N.*et al*., 2008, "Widespread changes in protein synthesis induced by microRNAs", *Nature*, vol. 455, no. 7209, pp. 58-63.

Shapiro, R. & Pohl, S. H. 1968, "The reaction of ribonucleosides with nitrous acid. Side products and kinetics", *Biochemistry*, vol. 7, no. 1, pp. 448-455.

Shapiro, R. & Shiuey, S. J. 1969, "Reaction of nitrous acid with alkylaminopurines", *Biochim.Biophys.Acta*, vol. 174, no. 1, pp. 403-405.

Shen, J. C., Rideout, W. M., III, & Jones, P. A. 1994, "The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA", *Nucleic Acids Res.*, vol. 22, no. 6, pp. 972-976.

Sie, C. P. & Kuchka, M. 2011, "RNA editing adds flavor to complexity", *Biochemistry (Mosc.)*, vol. 76, no. 8, pp. 869-881.

References

Slupphaug, G., Markussen, F. H., Olsen, L. C.*et al*., 1993, "Nuclear and mitochondrial forms of human uracil-DNA glycosylase are encoded by the same gene", *Nucleic Acids Res.*, vol. 21, no. 11, pp. 2579-2584.

Song, J. J., Smith, S. K., Hannon, G. J.*et al*., 2004, "Crystal structure of Argonaute and its implications for RISC slicer activity", *Science*, vol. 305, no. 5689, pp. 1434-1437.

Su, A. A. & Randau, L. 2011, "A-to-I and C-to-U editing within transfer RNAs", *Biochemistry (Mosc.)*, vol. 76, no. 8, pp. 932-937.

Suzuki, T., Kanaori, K., Tajima, K.*et al*., 1997, "Mechanism and intermediate for formation of 2'-deoxyoxanosine", *Nucleic Acids Symp.Ser.* no. 37, pp. 313-314.

Suzuki, T., Yamada, M., Ide, H.*et al*., 2000, "Identification and characterization of a reaction product of 2'-deoxyoxanosine with glycine", *Chem.Res.Toxicol.*, vol. 13, no. 4, pp. 227-230.

Suzuki, T., Yoshida, M., Yamada, M.*et al*., 1998, "Misincorporation of 2'-deoxyoxanosine 5'-triphosphate by DNA polymerases and its implication for mutagenesis", *Biochemistry*, vol. 37, no. 33, pp. 11592-11598.

Symington, L. S. & Gautier, J. 2011, "Double-strand break end resection and repair pathway choice", *Annu.Rev.Genet.*, vol. 45, pp. 247-271.

Tang, W., Fei, Y., & Page, M. 2012, "Biological significance of RNA editing in cells", *Mol.Biotechnol.*, vol. 52, no. 1, pp. 91-100.

Terato, H., Masaoka, A., Asagoshi, K.*et al*., 2002, "Novel repair activities of AlkA (3-methyladenine DNA glycosylase II) and endonuclease VIII for xanthine and oxanine, guanine lesions induced by nitric oxide and nitrous acid", *Nucleic Acids Res.*, vol. 30, no. 22, pp. 4975-4984.

Thomas, M. G., Loschi, M., Desbats, M. A.*et al*., 2011, "RNA granules: the good, the bad and the ugly", *Cell Signal.*, vol. 23, no. 2, pp. 324-334.

Thompson, D. M. & Parker, R. 2009, "Stressing out over tRNA cleavage", *Cell*, vol. 138, no. 2, pp. 215-219.

Thoms, K. M., Kuschal, C., & Emmert, S. 2007, "Lessons learned from DNA repair defective syndromes", *Exp.Dermatol.*, vol. 16, no. 6, pp. 532-544.

Tice, R. R. & Setlow, R. B. 1985, "DNA repair and replication in aging organisms and cells," in *Handbook of the biology of aging*, Van Nostrand Reinhold, pp. 173-224.

Tornaletti, S. 2009, "DNA repair in mammalian cells: Transcription-coupled DNA repair: directing your effort where it's most needed", *Cell Mol.Life Sci.*, vol. 66, no. 6, pp. 1010-1020.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P.*et al*., 2002, "Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage", *Nature*, vol. 419, no. 6903, pp. 174-178.

Vascotto, C., Fantini, D., Romanello, M.*et al*., 2009, "APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process", *Mol.Cell Biol.*, vol. 29, no. 7, pp. 1834-1854.

Venter, J. C., Adams, M. D., Myers, E. W.*et al*., 2001, "The sequence of the human genome", *Science*, vol. 291, no. 5507, pp. 1304-1351.

Virgen-Slane, R., Rozovics, J. M., Fitzgerald, K. D.*et al*., 2012, "An RNA virus hijacks an incognito function of a DNA repair enzyme", *Proc.Natl.Acad.Sci.U.S.A*, vol. 109, no. 36, pp. 14634-14639.

Wang, Q. 2011, "RNA editing catalyzed by ADAR1 and its function in mammalian cells", *Biochemistry (Mosc.)*, vol. 76, no. 8, pp. 900-911.

References

Wang, Q., Miyakoda, M., Yang, W.*et al*., 2004, "Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene", *J.Biol.Chem.*, vol. 279, no. 6, pp. 4952-4961.

Watson , J. D. & CRICK, F. H. 1953, "The structure of DNA", *Cold Spring Harb.Symp.Quant.Biol.*, vol. 18, pp. 123-131.

Wehner, M. R., Shive, M. L., Chren, M. M.*et al*., 2012, "Indoor tanning and non-melanoma skin cancer: systematic review and meta-analysis", *BMJ*, vol. 345.

Weiss, B. 2001, "Endonuclease V of Escherichia coli prevents mutations from nitrosative deamination during nitrate/nitrite respiration", *Mutat.Res.*, vol. 461, no. 4, pp. 301-309.

Weiss, B. 2006, "Evidence for mutagenesis by nitric oxide during nitrate metabolism in Escherichia coli", *J.Bacteriol.*, vol. 188, no. 3, pp. 829-833.

Weiss, B. 2008, "Removal of deoxyinosine from the Escherichia coli chromosome as studied by oligonucleotide transformation", *DNA Repair (Amst)*, vol. 7, no. 2, pp. 205-212.

Weissbach, R. & Scadden, A. D. 2012, "Tudor-SN and ADAR1 are components of cytoplasmic stress granules", *RNA.*, vol. 18, no. 3, pp. 462-471.

Welsh, K. M., Lu, A. L., Clark, S.*et al*., 1987, "Isolation and characterization of the Escherichia coli mutH gene product", *J.Biol.Chem.*, vol. 262, no. 32, pp. 15624-15629.

Wolf, J., Gerber, A. P., & Keller, W. 2002, "tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli", *EMBO J.*, vol. 21, no. 14, pp. 3841-3851.

Wood, R. D., Mitchell, M., & Lindahl, T. 2005, "Human DNA repair genes, 2005", *Mutat.Res.*, vol. 577, no. 1-2, pp. 275-283.

Yang, W., Chendrimada, T. P., Wang, Q.*et al*., 2006, "Modulation of microRNA processing and expression through RNA editing by ADAR deaminases", *Nat.Struct.Mol.Biol.*, vol. 13, no. 1, pp. 13-21.

Yang, W., Hendrickson, W. A., Crouch, R. J.*et al*., 1990, "Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein", *Science*, vol. 249, no. 4975, pp. 1398-1405.

Yao, M., Hatahet, Z., Melamede, R. J.*et al*., 1994a, "Deoxyinosine 3' endonuclease, a novel deoxyinosine-specific endonuclease from Escherichia coli", *Ann.N.Y.Acad.Sci.*, vol. 726, pp. 315-316.

Yao, M., Hatahet, Z., Melamede, R. J.*et al*., 1994b, "Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from Escherichia coli", *J.Biol.Chem.*, vol. 269, no. 23, pp. 16260-16268.

Yao, M. & Kow, Y. W. 1994c, "Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3'-endonuclease from Escherichia coli", *J.Biol.Chem.*, vol. 269, no. 50, pp. 31390-31396.

Yao, M. & Kow, Y. W. 1995, "Interaction of deoxyinosine 3'-endonuclease from Escherichia coli with DNA containing deoxyinosine", *J.Biol.Chem.*, vol. 270, no. 48, pp. 28609-28616.

Yao, M. & Kow, Y. W. 1996, "Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3'-endonuclease from Escherichia coli", *J.Biol.Chem.*, vol. 271, no. 48, pp. 30672-30676.

Yao, M. & Kow, Y. W. 1997, "Further characterization of Escherichia coli endonuclease V. Mechanism of recognition for deoxyinosine, deoxyuridine, and base mismatches in DNA", *J.Biol.Chem.*, vol. 272, no. 49, pp. 30774-30779.

I

OPEN ACCESS Freely available online

PLOS ONE

The Human Homolog of *Escherichia coli* Endonuclease V Is a Nucleolar Protein with Affinity for Branched DNA Structures

Cathrine Fladeby[1,3☯], Erik Sebastian Vik[1,2,3☯], Jon K. Laerdahl[1,3], Christine Gran Neurauter[1,3], Julie E. Heggelund[2,3], Eirik Thorgaard[1,3], Pernille Strøm-Andersen[2,3], Magnar Bjørås[1,3], Bjørn Dalhus[1,2,3], Ingrun Alseth[1,3]*

1 Department of Microbiology, Oslo University Hospital HF and University of Oslo, Rikshospitalet, Oslo, Norway, **2** Department of Medical Biochemistry, Oslo University Hospital HF and University of Oslo, Rikshospitalet, Oslo, Norway, **3** Centre for Molecular Biology and Neuroscience (CMBN), Oslo University Hospital HF and University of Oslo, Rikshospitalet, Oslo, Norway

Abstract

Loss of amino groups from adenines in DNA results in the formation of hypoxanthine (Hx) bases with miscoding properties. The primary enzyme in *Escherichia coli* for DNA repair initiation at deaminated adenine is endonuclease V (endoV), encoded by the *nfi* gene, which cleaves the second phosphodiester bond 3′ of an Hx lesion. Endonuclease V orthologs are widespread in nature and belong to a family of highly conserved proteins. Whereas prokaryotic endoV enzymes are well characterized, the function of the eukaryotic homologs remains obscure. Here we describe the human endoV ortholog and show with bioinformatics and experimental analysis that a large number of transcript variants exist for the human endonuclease V gene (*ENDOV*), many of which are unlikely to be translated into functional protein. Full-length ENDOV is encoded by 8 evolutionary conserved exons covering the core region of the enzyme, in addition to one or more 3′-exons encoding an unstructured and poorly conserved C-terminus. In contrast to the *E. coli* enzyme, we find recombinant ENDOV neither to incise nor bind Hx-containing DNA. While both enzymes have strong affinity for several branched DNA substrates, cleavage is observed only with *E. coli* endoV. We find that ENDOV is localized in the cytoplasm and nucleoli of human cells. As nucleoli harbor the rRNA genes, this may suggest a role for the protein in rRNA gene transactions such as DNA replication or RNA transcription.

Citation: Fladeby C, Vik ES, Laerdahl JK, Gran Neurauter C, Heggelund JE, et al. (2012) The Human Homolog of *Escherichia coli* Endonuclease V Is a Nucleolar Protein with Affinity for Branched DNA Structures. PLoS ONE 7(11): e47466. doi:10.1371/journal.pone.0047466

Editor: Zhongjun Zhou, The University of Hong Kong, Hong Kong

Received August 3, 2012; **Accepted** September 11, 2012; **Published** November 5, 2012

Funding: The study was supported by The Norwegian Research Council and The Norwegian Cancer Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ingrun.alseth@rr-research.no

☯ These authors contributed equally to this work.

Introduction

The genomes of all organisms are constantly challenged by agents, produced inside the cell or in the environment, that cause damage to the DNA. DNA base damage may lead to errors in replication and transcription, compromising the integrity of the genome. Three of the four bases present in DNA (cytosine, adenine, and guanine) contain an exocyclic amino group. Loss of this group by deamination occurs spontaneously under physiological conditions via a hydrolytic reaction [1,2]. This process is greatly enhanced by agents such as reactive oxygen radicals, UV radiation, heat, ionizing radiation, nitrous acid, nitric oxide, and sodium bisulfite [3–7].

It is estimated that a few hundred amino groups are lost from the DNA bases spontaneously in each cell every day, most frequently from cytosine bases. Adenine deamination occurs only at a rate of 2–3% compared to that of cytosine [8]. Deamination of cytosine and adenine produces uracil and hypoxanthine (Hx), respectively, both having miscoding properties. In addition, Hx in DNA might be the result of misincorporation of 2′-deoxyinosine triphosphate (dITP) during DNA replication [9]. In this case dITP is incorporated opposite cytosine and is also read as guanine by the DNA polymerases. Thus, at least in *Escherichia coli*, dITP incorporation is nonmutagenic [10].

Whereas uracil in DNA is removed by uracil DNA glycosylases [11], the principal enzyme for removal of Hx in *E. coli* is endonuclease five (endoV) encoded by the *nfi* gene [12]. This enzyme binds to and cleaves the second phosphodiester bond 3′ to Hx in an Mg^{2+} dependent manner generating 3′-OH and 5′-P termini [13,14]. Endonuclease V does not on its own remove the damage from DNA and additional proteins are thus required to complete repair. This process is poorly understood but has been shown to be reconstituted with recombinant endoV, DNA polymerase I and DNA ligase [15]. *E. coli* cells lacking endoV have a normal spontaneous mutation frequency, however upon exposure to nitrous acid *nfi*$^-$ cells are mutators showing elevation in AT→GC and GC→AT transition as well as GC→CG transversion mutations [16]. *E. coli* endoV is a rather promiscuous enzyme acting on different substrates including uracil [17,18], xanthine (deaminated guanine) [19], apurinic/apyrimidinic (AP)

sites [14], urea residues [14], mismatches [20] and also structure substrates such as insertion and deletion loops, 5′-flaps, hairpins and pseudo-Y structures [21]. The ability of *E. coli* endoV to recognize all three deamination products in DNA is unique and is not shared by any of the other known repair enzymes. Finally, is has been shown that endoV from *Thermotoga maritima* (*Tma*) possesses both 5′ and 3′ exonuclease activities and a potential role for these activities in end-processing after Hx incision was suggested [22]. The 3-dimensional structure of *Tma* endoV in complex with Hx-containing DNA was recently determined [23]. The structure reveals the presence of a wedge motif (PYIP) involved in damage detection and DNA strand separation at the site of the lesion. The deaminated adenine lesion is rotated approximately 90° into a recognition pocket where it is tightly coordinated by hydrogen-bonding interactions.

Homologs of endoV are widespread in nature and are found in all three domains of life [23]. In addition to *E. coli*, endoV homologs have been characterized from *Archaeglobus fulgidus* [24], *T. maritima* [25], *Ferroplasma acidarmanus* (in fusion with O^6-alkylguanine-DNA alkyltransferase active site domain) [26] and *Salmonella typhimurium* [27], however knowledge about the eukaryotic counterparts is sparse. cDNA for endoV from mice has been cloned, however no robust enzyme activity for Hx or other tested substrates were found [28]. In this work we have characterized the human variant of endonuclease V by identification of isoforms, subcellular localization and biochemical assays.

Materials and Methods

Ethics statement

A commercially available tissue array was used and ethical principles maintained by the manufacturer (Origene) (http://www.origene.com/Tissue/Tissue_QC.aspx).

Bioinformatics analysis

Protein and mRNA derived sequences were obtained from GenBank [29] and other NCBI database resources [30] and from the Ensembl project [31]. The large number of human transcript variants was also investigated in the Ensembl and the UCSC [32] genome browsers. Protein structural disorder was predicted with DISOPRED2 [33].

A multiple sequence alignment of human ENDOV and *Tma* endoV as well as 13 and 8 additional eukaryotic and bacterial homologs, respectively, was generated with Muscle [34]. An alignment of the human (target) and *Tma* (template) sequences based on this multiple sequence alignment was manually edited in order to move insertions and deletions out of secondary structure elements in the structural modeling template from Dalhus *et al.* [23] (Protein database (PDB) identifier 2W35). With this alignment, a model of human ENDOV was generated with SwissModel [35] employing standard homology modeling.

Cell culture and transfection

Human embryonic (HE) fibroblasts were obtained from the National Institute of Public Health (Folkehelsa, Oslo, Norway) and cell lines HCT116 (human colon epithelial cells), ACHN (human kidney epithelial cells) and HeLa S3 (human cervix epithelial cells) were purchased from American Type Culture Collection (ATCC). HE cells were cultured in a 1:1 mix of minimal essential medium (MEM; Gibco, Life Technologies, Carlsbad, CA, USA) and Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (Standard quality FBS, PAA lab, Austria), 1× GlutaMAX (200 mM, Gibco), and 1× penicillin-streptomycin (10000 U/ml, Lonza, Basel, Switzerland). HCT116, ACHT and HeLa cells were cultured in DMEM supplemented with 10% FBS, 1× GlutaMAX, and 1× penicillin-streptomycin. Transient transfections were performed with FUGENE (Invitrogen, Life Technologies), according to the supplied protocol.

Cell cycle synchronization and analysis by flow cytometry

Synchronization of the cells in G0 phase was achieved by culturing cells as a confluent layer for 72 h followed by serum starvation (0.2% serum) for 72 h. The cells were released from G0 by trypsination (Trypsin-EDTA 200 mg/l, Lonza) for 4 min at 37°C and cultivated in standard growth medium at 25% confluence. Cells were harvested by trypsination at indicated time points, washed in ice-cold PBS and stored at −20°C. Cells used for phase analysis were resuspended in PBS and fixed by addition of ice-cold 100% ethanol to a final concentration of 70% and stored at −20°C. For FACS analysis, the cells (about 1 mill/ml) were stained with 50 µg/ml propidium iodide (Sigma-Aldrich, St.Louis, MO, USA) in 4 mM Na-citrate buffer containing 0.1 mg/ml RNaseA (Molzyme GmbH & Co, Bremen, Germany) and 0.1% Triton X-100 (Sigma-Aldrich) for 10 min at 37°C and put on ice. Cells were subjected to flow cytometry analysis (BD LSRII flow cytometer, Becton Dickinson, San Jose, California, USA) and the results were analysed with CellQuest software (Becton Dickinson).

Total RNA isolation, cDNA synthesis and quantitative real-time RT-PCR

Total RNA was isolated from frozen cell pellets using RNeasy kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. cDNA was generated from total RNA samples using the High-capacity cDNA reverse transcription kit (Applied Biosystem, Life Technologies). Human *ENDOV* mRNA levels was determined with primers amplifying exons 2 to 3 or exons 6 to 8 (Table S1; Eurofins MWG Operon, Ebersberg, Germany) using the Power SYBR Green PCR master mix and the Step One Plus Real-Time PCR system (Applied Biosystem) according to the kit and system instructions. All samples were run in triplicate, and melting point analyses were performed to confirm the specificity of the PCR reaction. *GAPDH* (Table S1, primers 5 and 6) was used as the reference gene for normalization, and G0 as the reference sample for RQ calculation.

For measurement of human *ENDOV* expression in normal and cancer tissue, TissueScan™ Cancer Survey cDNA Arrays (CSRT303; OriGene Technologies, Rockville, USA) were used. These arrays consisted of cDNA prepared from pathologist-verified human tumor tissue obtained from 18 different tissues normalised against β-*ACTIN*. Primers covering exons 6 to 8 were used (Table S1). Additional clinical information for each sample can be found at http://www.origene.com/qPCR/Tissue-qPCR-Arrays. Results are shown for 13 of the tissues, whereas 5 were omitted due to high uncertainty in the values obtained.

Nothern Blot analysis

Total RNA was isolated from HE cells using TRIzol Reagent (Ambion, Applied Biosystems) according to manufacturer's instructions. mRNA was isolated from total RNA using the MicroPoly(A)Purist Kit (Ambion) and 5 µg/lane was subjected to 1% denaturing agarose gel electrophoresis at 5 V/cm. mRNA was transferred to an BrightStar-Plus membrane (Ambion) by downward transfer from gel and crosslinked to the membrane at 120 mJ/cm^2 in a CL-1000 UV-Crosslinker (UVP, Upland, California, USA). The Northern Max kit (Ambion) was used in the blotting, with prehybridization/hybridization and washing steps performed as described by the manufacturer. Primers for

PCR amplification of the probes are listed in Table S1. The β-*ACTIN* cDNA probe was from Clontech (Takara Bio, Otsu, Japan). The probes (178 bp for exon 3, 378 bp for exons 4–8 and 342 bp for exon 10) were labeled using Rediprime II Random Prime labeling system (Amersham Biosciences, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) and EasyTide dCTP (α-^{32}P) (PerkinElmer, Massachusetts, USA). Hybridization signals were detected and quantified by phosphorimaging (Typhoon 9410) and ImageQuant TL software (Molecular Dynamics, California, USA). The amount of *ENDOV* transcripts was calculated relative to β-*ACTIN*.

Immunofluorescence microscopy

Human *ENDOV* (exons 1–10) was subcloned into pEGFP N- and C-vectors (Clontech) by Genescript (Piscataway, New Jersey, USA). HeLa cells were grown to 60–70% confluence on 2-well chamber slides and transfected with the different pEGFP constructs using the FuGENE6 (Roche, Mannheim, Germany) reagent according to the manufacturer's instructions. Cells were fixed for 15 min in PBS containing 4% paraformaldehyde, washed and quenched in 20 mM glycine in PBS for 10 min. Permeabilization was performed in 0.1% Triton X-100 in PBS for 10 min followed by blocking with 10% FBS in PBS for 30 min. All labelling steps were carried out in the blocking buffer. Cells were incubated with monoclonal primary antibodies against fibrillarin (Abcam, Ab4566) for 1 h, washed, and further incubated for 1 h with Alexa 595–conjugated anti-mouse antibodies (Molecular Probes Europe, Life Technologies). Cells were then washed in PBS and coverslips were mounted with Mowiol (Sigma-Aldrich). Confocal images were acquired with Carl Zeiss LSM 510 CLSM laser scanning microscope (Jena, Germany).

Protein extracts for western analysis were made by adding ice-cold RIPA lysis buffer containing protease inhibitor cocktail (P8340, Sigma-Aldrich) to the transfected cells. Cells were collected, sonicated for 2×20 sec and spun down at 12.000 *g* for 10 min. 30 µg cell extract were boiled with NuPAGE LDS sample buffer (Invitrogen) and subjected to SDS-PAGE using 10% NuPAGE gels (Invitrogen). Proteins were transferred to PVDF membrane using iBlot® Gel Transfer Device (Life Technologies), detected by incubating the membrane ON at 4°C with anti-GFP antibody (Abcam, Ab290) followed by incubation for 1 h at RT with HRP-conjugated secondary antibody and visualised by Immun-Star WesternC chemiluminescence kit (BioRad Laboratories, California, USA) using the ChemiDoc MP System (BioRad).

Design of constructs for recombinant protein expression

The nucleotide sequence for the human *ENDOV* transcript (exons 1–9, 309 residues) was synthesised by Genscript Inc. (Piscataway, New Jersey, USA) with optimal codon usage for expression in *E. coli* in the pET28b vector (Novagen, Darmstadt Germany) using the NdeI and EcoRI restriction sites. No protein was produced from this construct upon induction of *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene, Agilent technologies, California, USA) and an N-terminal maltose binding protein (MBP)-ENDOV fusion construct was made as follows: The codon optimised sequence was amplified by PCR using primers 5′-ATATCCATGGCACTGGAAGCCGCCGGC-3′ and 5′-ATATGGATCCTTACTGCCAATCTTTACCCGCCTGTTC-C-3′ for subcloning into the vector pETM-41 (EMBL, Heidelberg, Germany) using NcoI (underlined) and BamHI (underlined) to give a construct with ENDOV fused to an N-terminal MBP tag separated by a tobacco etch virus (TEV) protease cleavage site (pETM-41-MBP-TEV-ENDOV-Exon9). The fusion protein also contained an N-terminal hexahistidine tag in front of the MBP protein. This construct was further used as a template to design the corresponding fusion between MBP and the shorter isoform of human ENDOV (exon 1–10, 282 residues) by site-specific mutagenesis using the forward primer 5′-GGCGATTCTGGT-GAAAGCTCTGCGCTGTGTTAGCCGCCGCAGGATCAC-TCTCCG-3′ together with its corresponding reverse and complementary oligonucleotide. A total of 6 nucleotide mutations in the underlined region transformed the amino acid sequence for ENDOV exon 9 from 280…GEGQ…283 to 280…ALC-Stop…. to give pETM-41-MBP-TEV-ENDOV-Exon10. The mutants ENDOV-RK, ENDOV-Y91A and ENDOV-Wg were designed by site-specific mutagenesis using the forward primers 5′-GCAGTCGTGAACATATCGACGATAGCCTGGGTCTGC-CGGGTCC-3′, 5′- CTGACGGCGCCGGCGGTTAGCGGC-TTTC-3′ and 5′- CGTATGGTGAGCCTGACGGCGGGGG-GCGGTGGCGGCTTTCTGGCCTTCCGTGAAGTGCC-3′ with their corresponding reverse primers, respectively.

Protein expression and purification

The pETM-41 based human ENDOV exon 10 WT and mutant constructs were transformed into chemically competent *E. coli* BL21-Codon Plus (DE3)-RIPL cells. The cells were grown in LB medium supplemented with 100 mg/l kanamycin at 37°C with shaking until the $OD_{600\ nm}$ reached ~0.8. The temperature was lowered to 18°C before induction of the protein expression by 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After overnight expression at 18°C, cells were harvested by centrifugation and the cell pellets were resuspended in buffer A, containing 50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole and 10 mM β-mercaptoethanol (β-ME). Cells were lysed by ultrasonication (3×30 sec) followed by centrifugation for 30 min at 27000 *g* at 4°C. The recombinant His-MBP-ENDOV fusion proteins were extracted from the supernatant by Ni-NTA affinity chromatography, using 50 mM and 300 mM imidazole versions of buffer A for elution. Fractions rich in ENDOV were pooled, concentrated and dialysed at 4°C against a TEV buffer (50 mM Tris pH 8.0, 0.5 mM EDTA, 1 mM DTT). TEV protease with an N-terminal hexahistidine tag produced from a pSC563 plasmid (Courtesy of Prof. M Ehrmann, Cardiff University, UK) was added to the fusion proteins in ratio 1:100 and incubated at 12°C overnight. After proteolysis, the protein mixtures were dialysed against buffer A, and the free His-MBP and TEV proteins were separated from ENDOV by a second Ni-NTA purification step. The untagged ENDOV proteins were collected in the flow-through and wash fractions, concentrated and applied to a Superdex 75 size-exclusion chromatography column (GE Healtcare) equilibrated with 50 mM NaCl, 50 mM Tris pH 8.0 and 10 mM β-ME. The purified human ENDOV was concentrated and stored at 4°C.

Oligonucleotides and ^{32}P labeling

The DNA substrates were made by combining the oligonucleotides (Eurofins MGW Operon; Table S2) in the following way (asterisk indicates the ^{32}P labeled oligonucleotides): undamaged DNA: 1+2*, double stranded with hypoxanthine: 1+3*, double stranded with uracil: 4*+5, loop: 6*+7 hairpin: 8*+9, 3′-flap: 10+11+14*, 5′-flap: 10+12+13*, 3-way: 10*+13+14, pseudo-Y: 10*+13, fork: 10+11+14*+15 and Holliday junction: 16+17*+18+19. The DNA substrates were 5′ end labeled using T4 polynucleotide kinase (New England BioLabs, Hitchin, UK) in the presence of [γ-^{32}P]ATP (Amersham Biosciences). Radioactive labeled oligonucleotides were annealed to their respective complementary strands by heating the solution to 90°C for 2 min and

slowly cooling to room temperature. The DNA substrates were separated by 10% native PAGE, excised from the gel, eluted by diffusion in H_2O and stored at 4°C.

DNA nicking activity

Various amounts of ENDOV were mixed with 10 fmol substrate DNA, 5 ng pQE31 competitor DNA (Qiagen, QIAGEN, Hilden, Germany) and reaction buffer (10 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 50 mM KCl, 5% glycerol and 1 mM DTT) in a total volume of 10 µl, and incubated at 37°C for 30 min. The reactions were stopped by adding formamide loading buffer (90% formamide, 0.1% xylene cyanol, 0.1% bromphenol blue) and heating to 85°C for 3 min. Cleavage products were analysed by 20% PAGE (Long Ranger, 7 M urea, 1× taurin), visualised by phosphorimaging and quantified by ImageQuant TL.

Electrophoretic mobility shift assay

The affinity of human ENDOV for damaged DNA was analysed by electrophoretic mobility shift assay (EMSA). Enzyme (amounts given in figure legends), 10 fmol DNA substrates and 5 ng pQE31 competitor DNA were incubated in a 10 µl reaction volume (5 mM $CaCl_2$, 10 mM HEPES-KOH pH 7.4, 1 mM DTT and 20% glycerol) at 4°C for 15 minutes. DNA loading buffer (Fermentas) were added and the samples were separated by 10% native PAGE (Long Ranger, 1× taurin, 5 mM $CaCl_2$) on ice. Results were visualised by phosphorimaging and quantified with ImageQuant TL software.

Results

Bioinformatics analysis of ENDOV demonstrates numerous splice variants

The endonuclease V gene has a scattered distribution in eukaryotes, most compatible with independent gene loss in multiple lineages. Publicly available sequence information shows that *ENDOV* orthologs are present in most plants and green algae, in echinoderms (*Strongylocentrotus purpuratus*) and in all three subphyla of the chordates, the vertebrates, tunicates, and cephalochordates. It is also found in sponges (*Amphimedon queenslandica*), cnidarians (*e.g. Nematostella vectensis*), and amoebozoans such as *Entamoeba histolytica* and *Dictyostelium discoideum*. Among the arthropods the gene appears to be completely missing in the largest group, the insects, but it is found in the genomes of at least some crustaceans (*e.g. Daphnia pulex* and *Caligus rogercresseyi*). Within the fungi, an *ENDOV* ortholog is found in the genome of *Schizosaccharomyces pombe* and the other three sequenced *Schizosaccharomyces* species, but is otherwise nearly, or possibly completely, absent. Endonuclease V also appears to be missing in apicomplexan protozoans such as *Plasmodium falciparum* and *Cryptosporidium parvum*. In summary, the ENDOV sequence is highly conserved in eukaryotes, but the gene itself appears to have been lost in a large fraction of the eukaryotic species.

The mouse endonuclease V variant described by Moe *et al.* [28] comprises nine exons and translates into a protein with 338 residues (validated RefSeq [36] record NP_001158108). At least 20 publicly available expressed sequence tag (EST) sequences, representing processed mRNA from various tissues and life stages, confirm that this is the major splice variant for EndoV in mice (Figure 1A). The chicken genome [37] encodes an ENDOV ortholog that is spliced identically as in mouse, while the frog *Xenopus tropicalis* [38] ortholog is spliced identically for the seven 5′ introns (Figure 1A). Also for chicken and frog, this splicing pattern is supported by a number of published EST sequences. Frog ENDOV has a shorter C-terminus due to a stop codon in exon 8. The correctness of this stop codon is supported by both the genomic sequence and all ten available EST sequences spanning the 3′ region of the transcript. The pattern of splicing of exons 1–8 described above is conserved, also including the intron phases, in ENDOV orthologs in other mammals such as pig and rat, and in the more distantly related sea urchin *S. purpuratus* (Figure 1A). All protein sequences are listed in Figure S1.

The conserved splicing pattern results in metazoan endonuclease V proteins with a high degree of similarity with prokaryotic endoV orthologs. For example, the core region of mouse ENDOV, *i.e.* residues 1–250, encoded by exons 1–8 (Figure 1), aligns with full-length *E. coli* (223 residues) and *T. maritima* (225 residues) endoV with ~32% and ~29% sequence identity, respectively. Given this degree of similarity, it is likely that the 3D structures of metazoan ENDOV enzymes are very similar to the bacterial homologs (See Ref. [23] as well as PDB identifiers 3HD0, 3GOC, and 3GA2). The C-terminal segment of the above bilaterian ENDOV homologs, *e.g.* residues 251–338 of mouse ENDOV, is predicted to be structurally disordered (Figure 1 and S1). This part of ENDOV also appears to be under no selective pressure (Table S3). The ratio of non-synonymous (*Ka*) and synonymous (*Ks*) substitution rates for pairwise comparisons of ENDOV segments from mouse, rat, and Chinese hamster gives an average value of $Ka/Ks = 0.19$ for the core region, showing, as expected, purifying selection ($Ka/Ks < 1$). For the C-terminal segment, average $Ka/Ks = 1.49$, strongly suggesting that during evolution there is no selective pressure for conserving the sequence of the C-terminal tail (*i.e.* all mutations are equally tolerated).

Splicing of human *ENDOV* transcripts has not previously been analysed in detail, but judging from the sequence data available in public databases there is a high degree of alternative splicing of this gene, resulting in a multitude of isoforms. The human protein with RefSeq identifier NP_775898 corresponds to an mRNA where exons 1–9 are spliced in the same fashion as in other bilateria (see above), but with an additional 3′ exon (Figure 1B). This transcript consists of 2858 nucleotides (nt) and will in the following be referred to as full-length (1–10). However, only one single (AK096344) of nine human *ENDOV* mRNAs available in GenBank is spliced in a similar fashion, while the remaining eight mRNAs all are lacking exon 3. There are more than 70 human ESTs, representing mRNAs from various tissues, available for *ENDOV*, but among the ~30 ESTs comprising exons 1, 2 and 4, only 9 contains exon 3. Thus, the major fraction of available spliced human *ENDOV* transcript sequences is lacking exon 3. It is, however, highly unlikely that these transcripts will be translated into a folded and functional protein, since exon 3 builds a large part of the ENDOV core domain and contains residues involved in damage recognition, DNA strand cleavage, as well as the DNA strand separating PYIP motif [23]. While many of the human *ENDOV* transcripts comprise exons 5–8, splicing at the 3′ end results in more than eight alternative isoforms with less than ten of nearly fifty transcripts from the 3′ end of the gene having identical splicing as in mouse and chicken (Figure 1). None of the alternative 3′ region splice variants appears to be evolutionary conserved or to encode a structured protein domain. In conclusion, the majority of human *ENDOV* transcripts available in public databases are unlikely to produce functional protein. The C-terminal tail of ENDOV is not evolutionary conserved in metazoa and is found in many variants due to alternative, seemingly random, splicing in human cells.

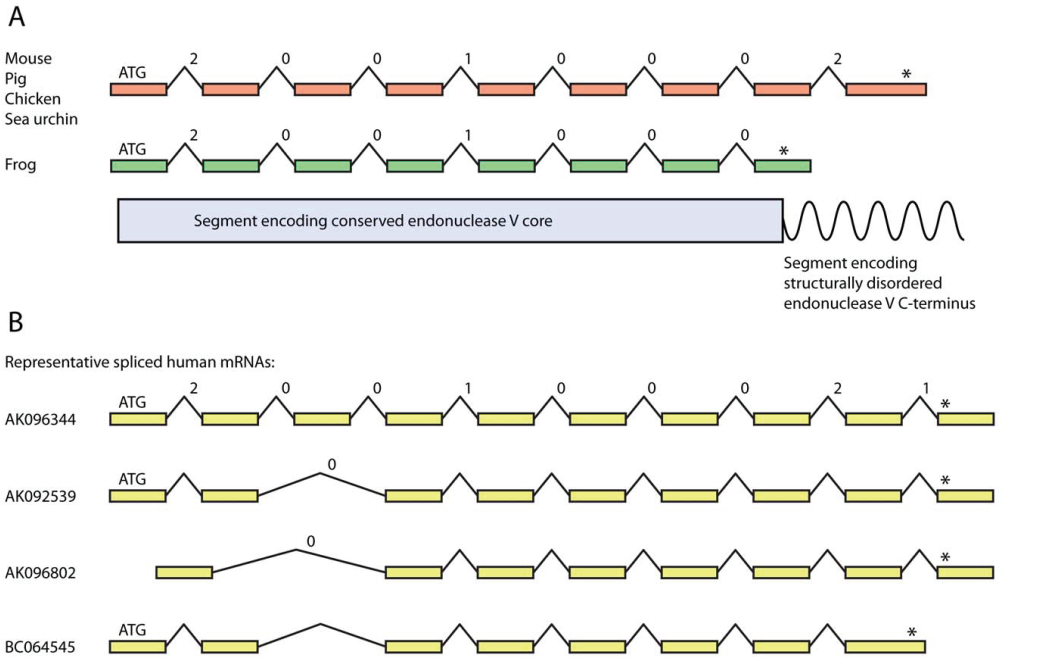


Figure 1. Schematic representation of splicing of human, other tetrapod and sea urchin *ENDOV* mRNAs. (**A**) The segments of the *ENDOV* transcripts encoding the conserved protein core are spliced identically, including the intron phases, in rodents, pig, chicken, frog and the echinoderm sea urchin. The position of the start (ATG) and stop (asterisk) codons are indicated above the relevant exons (not shown to scale), while intron phases are shown above the introns. These are defined as the position of the intron within a codon, with phase 0, 1, and 2 placed before the first nucleotide, after the first nucleotide, and after the second nucleotide, respectively. The structurally disordered and poorly conserved C-terminus of ENDOV is encoded by one or more 3′ exons and has variable length, *e.g.* ~85 and ~5 residues in mouse and frog, respectively. (**B**) Most previously published spliced human mRNAs are lacking one or several 5′ exons, in particular exon 3, while alternative splicing at the 3′ end results in multiple variants. Four representative full length mRNAs are shown with accession numbers from GenBank [29].
doi:10.1371/journal.pone.0047466.g001

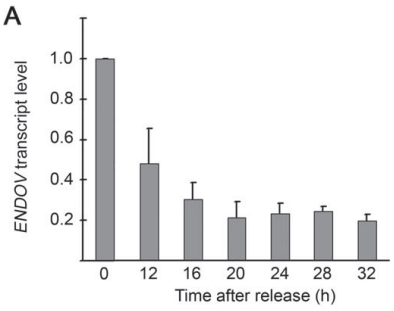
Characterization of *ENDOV* transcripts in human cells

As the sequence databases revealed a multitude of human *ENDOV* transcript variants, we aimed to verify their presence experimentally. Initially we tried to detect full-length *ENDOV* cDNA from different human cell types. Total RNA was purified from primary human fibroblasts, kidney and colon cell lines and subjected to cDNA synthesis and PCR analysis using exon 1 and exon 8 or 9-specific primers. With these primer sets we only amplified *ENDOV* transcripts lacking exon 3 (data not shown). However exon 3 containing transcripts were detected in these cells using exon 3 specific primers (data not shown). Based on these observations we assumed that full-length transcripts containing exon 3 most likely are expressed at low levels in the cells. This was further demonstrated in 5′ RACE experiments using Marathon-Ready cDNA from brain (Clontech) in which only 2 of 30 transcripts sequenced contained exon 3 (data not shown). Using Real-Time qPCR technique we sought to quantify the amount of exon 3 containing transcripts relative to the total level of *ENDOV* transcripts in brain tissue and primary fibroblasts. For this purpose we designed primers specific for exon 3 (Table S1: primers 1 and 2) and exon 7 (representing the total amount of transcript, Table S1: primers 3 and 4), and standard curve measurements confirmed an equal amplification efficiency of ~95% for the two primer sets. It was shown that in cDNA from brain tissue and HE fibroblasts the amount of transcripts containing exon 3 was approximately 30% and 50% relative to exon 7 containing transcripts, respectively. These data support the findings from the sequence databases analyses showing that human *ENDOV* transcripts without exon 3 are more abundant than the transcripts that contain exon 3.

ENDOV expression during cell cycle

To examine whether *ENDOV* transcription is regulated during cell-cycle progression, cultured human fibroblasts were arrested in G0 by contact inhibition and serum deprivation. Flow cytometry showed that 83% of the cells were in G0/G1 phase (Figure 2A). Cells were released from the arrest by culturing at subconfluence in serum supplemented complete medium and harvested at time points as indicated. Total RNA was isolated and Real-Time qRT-PCR experiments were performed with exon 3 specific primers (Table S1: primers 1 and 2). The results revealed highest level of *ENDOV* mRNA in G0 arrested cells and a reduction upon cell cycle progression (Figure 2A). After 20–28 h, where most of the cells were in S or G2/M phases, the *ENDOV* transcript level had decreased 4–5 folds.

To further investigate *ENDOV* transcripts, northern blot analysis was performed with mRNA isolated from HE fibroblasts synchronised as above. A probe spanning exons 4–8 gave a strong



Hours	0	12	16	20	24	28	32
G0/G1	83	75	75	45	21	17	41
S	2.1	1.6	3.2	28	21	6.3	5
G2/M	10	13	12	14	34	51	34

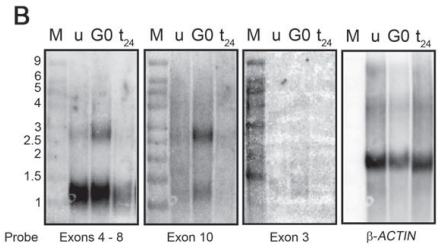


Figure 2. Upregulation of *ENDOV* transcription during quiescence. (**A**) Human embryonic fibroblasts were arrested in G0 by serum starvation at confluence and released by replating 1:4 in culture medium with serum. *ENDOV* transcript levels (exons 2 to 3) were measured during cell cycle progression after G0 release at indicated time points by qRT-PCR. G0 cells were used as the reference for calculations. The average of 3 parallels (same RNA) was calculated and standard deviation is shown. Cell cycle distribution was monitored using propidium iodide staining followed by flow cytometry after release from the block. The percentage of cells in each cell cycle is presented in the table. The experiment was repeated twice with similar results. (**B**) Nothern blot analysis of *ENDOV* mRNA. mRNA was isolated from human fibroblasts that were unsynchronised (u), G0 arrested (G0) and allowed to proliferate for 24 hours (t_{24}), separated by electrophoresis and transferred to a nylon membrane. Hybridisation signals with probes spanning exons 4–8, exon 10, exon 3 of *ENDOV* and for β-*ACTIN* are shown. M is the RNA size standard as indicated (in kilobases).
doi:10.1371/journal.pone.0047466.g002

signal corresponding to transcripts of 1200–1400 nucleotides (nt) in addition to a weaker band of 2800 nt (Figure 2B). A transcript of 2800 nt corresponds to the full-length transcript (1–10) which was verified with an exon 10 specific probe (Figure 2B). This signal was highest in G0 arrested cells (3–4 fold), confirming the qRT-PCR data. However, the majority of the transcripts (~70%) detected with the exon 4–8 probe, was less than 1500 nt and appeared to lack exon 10. With an exon 3 specific probe, two faint bands of ~1200 and ~1400 nt were detected, but none of 2800 nt (Figure 2B). As the signals were very weak, it might be that the amount of full-length transcripts is below the detection limit with this specific probe. Thus, our expression analyses show that the *ENDOV* mRNA level is steady in cycling cells but peaks upon quiescence. Further, different *ENDOV* transcript variants exist and the majority appears to lack exon 10.

In order to clarify which protein variants are produced by the cells, we attempted to immunoprecipitate endogenous ENDOV by the use of an in-house rabbit polyclonal antibody raised against ENDOV mixed with protein extract made from G0 arrested HE fibroblasts. However, we were not able to precipitate ENDOV as evidenced by lack of signals in western blot analysis (data not shown). It appears that the expression of ENDOV is below the detection limit of this method.

ENDOV expression in various tumor samples

We also investigated the expression level of *ENDOV* in several tumor types and their normal counterparts by using qPCR arrays containing cDNA from diseased and normal tissues and primers for exons 6/7 to 8. Although we observed high variation within the samples from the same tissue, the results showed that tumor samples in general did not display altered expression of *ENDOV* compared to the corresponding normal tissues (Figure 3). In addition, this analysis showed that kidney, pancreas and testis tissue express highest amounts of *ENDOV* whereas adrenal gland, cervix and colon are the tissues with lowest level of *ENDOV*.

ENDOV localizes to nucleolus

Cellular localization of ENDOV was examined in HeLa cells transiently transfected with full-length *ENDOV* (1–10) cloned into pEGFPN/C vectors. The GFP-ENDOV fusion protein was found in the cytoplasm and in nucleolus, confirmed by colocalization with the nucleolar protein fibrillarin (Figure 4 A–C). The same

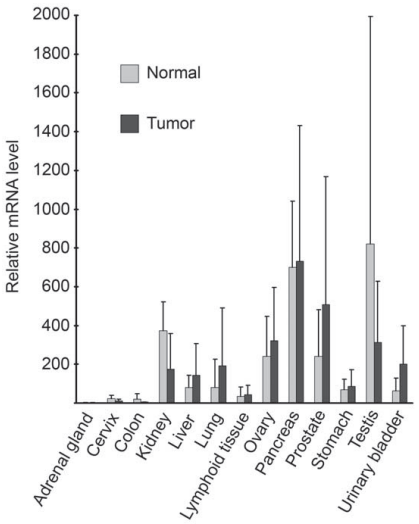


Figure 3. Expression of *ENDOV* in human cancers. Transcript profiling was performed using a cancer tissue qPCR array and primers amplifying exons 6 to 8 of *ENDOV*. cDNA levels are normalised to β-*ACTIN*. *ENDOV* mRNA levels were calculated relative to the tissue with lowest expression level (adrenal gland). Tumour and normal tissue were grouped according to their origin and the average and standard deviation were calculated for each group. The experiment was performed twice with similar results.
doi:10.1371/journal.pone.0047466.g003

localization pattern was also seen when the GFP-tag was placed at the C-terminal end of the protein (data not shown). The GFP protein itself was evenly distributed throughout the cell (Figure 4D). Western blot analysis of protein extracts from the GFP-ENDOV expressing cells probed with a GFP antibody, showed that GFP-ENDOV existed as fusion proteins migrating slightly faster than the expected sizes (MW GFP: 27 kD, ENDOV: 30.8 kD) (Figure S2). These data demonstrate that full-length ENDOV may exert its function in nucleolus and possibly also in cytoplasm.

ENDOV has no endonuclease activity but binds branched DNA structures

The primary activity of prokaryotic endoV enzymes is incision of the second phosphodiester bond 3′ to Hx residues in the DNA. To test whether this also is true for the human homolog, we performed endonuclease activity assays with recombinant ENDOV (1–10) and an Hx containing DNA substrate. Unexpectedly, ENDOV had no activity towards Hx under the same condition where *E. coli* endoV cleaved efficiently (Figure S3). Different reaction conditions were tested by varying pH (6.5–8.5) and salt ions (Mg^{2+}, Mn^{2+}) without any effect on ENDOV endonuclease activity. *E. coli* endoV is also active on uracil [17] and AP site [14] containing DNA, however no cleavage was obtained when ENDOV was tested against these substrates (data not shown).

In addition to deaminated bases, *E. coli* endoV is also active upon different branched DNA structures such as 5′-flaps, hairpins and insertion/deletions loops where it cleaves 3′ to the branching point [21]. Human ENDOV was tested against a panel of similar DNA substrates including 5′-flap, 3′-flap, 3-way junction, pseudo-Y, fork and Holliday junction, but no endonuclease activity was detected on these substrates either (data not shown). Next, we tested whether ENDOV could recognize and bind any of these substrates. Only very weak and unspecific binding was found for the Hx containing DNA and for the undamaged double stranded DNA substrate (Figure 5A, B). In contrast, ENDOV had high affinity for all 6 branched DNA substrates (Figure 5C–H). The affinities of *E. coli* endoV for these substrates were comparable to Hx containing DNA. To our knowledge, binding of *E. coli* endoV to 3′-flap, fork, 3-way and Holliday junction has not previously been demonstrated. For *E. coli* endoV with control double stranded DNA, a smear rather than a distinct shift was observed, possibly reflecting weak interactions between the DNA and enzyme that cease during gel electrophoresis. To biochemically map the binding between ENDOV and branched DNA, three different site-specific mutants were designed and the corresponding proteins purified after heterologous expression in *E. coli*. All three mutant proteins were soluble and expressed to the same level as the wild type protein (Figure S4). The Tyr73/80 in *E. coli* and *T. maritima* endoV enzymes are key residues for Hx strand recognition [39], however substitution of the corresponding ENDOV Y91 to an alanine had no effect on the binding affinity to branched substrates (Figure 6A–F), suggesting that the Tyr residue of the wedge structure is not involved in recognition of branched DNA structures. However, the wedge mutant ENDOV-Wg, with the entire DNA strand-separating motif PYVS (residue 90–93) replaced with 4 glycin residues, was strongly compromised in DNA binding. It appears that removal of the entire wedge domain introduces major structural distortions that abolish accommodation of DNA. Further, the ENDOV double mutant RK, with amino acids Arg248 and Lys249, predicted to be involved in sugar-phosphate binding based on comparison with the structure of *T. maritima* endoV (Figure 7), replaced by glutamate residues, completely lost the DNA binding ability. In sum, these data suggest that human ENDOV employ the same mechanism as its prokaryotic counterparts to accommodate branched DNA, whereas the ability to recognize Hx is severely diminished.

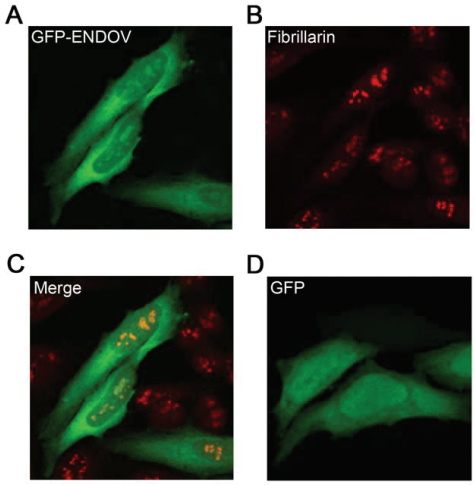


Figure 4. The EGFP-ENDOV fusion protein localises to nucleoli and cytoplasm. Confocal imaging of HeLa cells overexpressing GFP-ENDOV (**A**) probed with an anti-fibrillarin antibody (**B**), an overlay of the two (**C**) or GFP only (**D**).
doi:10.1371/journal.pone.0047466.g004

Discussion

Deamination of adenine to Hx in DNA occurs endogenously at a low but significant level and may result in A:T to G:C transition mutations. Removal of Hx is therefore necessary for the maintenance of genomic integrity and the primary enzyme involved in this process in *E. coli* is endoV. Homologs of endoV are present in all kingdoms of life and represent a family of highly conserved proteins (Figure S1). The high degree of evolutionary conservation clearly demonstrates that these proteins have an important functional role in the cells. Nevertheless, the functions of the eukaryotic endoV homologs remain unknown. Here we have characterized human ENDOV by gene expression, subcellular localization and biochemical analysis.

Information available in public databases shows that human *ENDOV* transcripts are highly variable. Surprisingly, most of the transcripts, experimentally determined by many different groups, are highly unlikely to encode functional protein. Our gene expression experiments confirmed the existence of multiform, truncated and probably misspliced transcripts in human cells. The lack of exon 3 in many transcripts could be due to an unfavorable splice donor site in intron 3 (GT substituted with GC) which may allow splicing of exon 4 directly to exon 2. The nonstandard GC-AG intron 3 is conserved in eutherian mammals, except rodents, but not in marsupials and other vertebrates. Whether this nonstandard intron contributes to *ENDOV* transcript regulation in eutherian cells is at present unknown. By RT-PCR, we demonstrated that transcripts containing the expected exon 2-exon 3 junction are present in human cells, although at significantly lower level than total *ENDOV* mRNA. Nevertheless, at least some full-length transcripts appear to be present in most cell types.

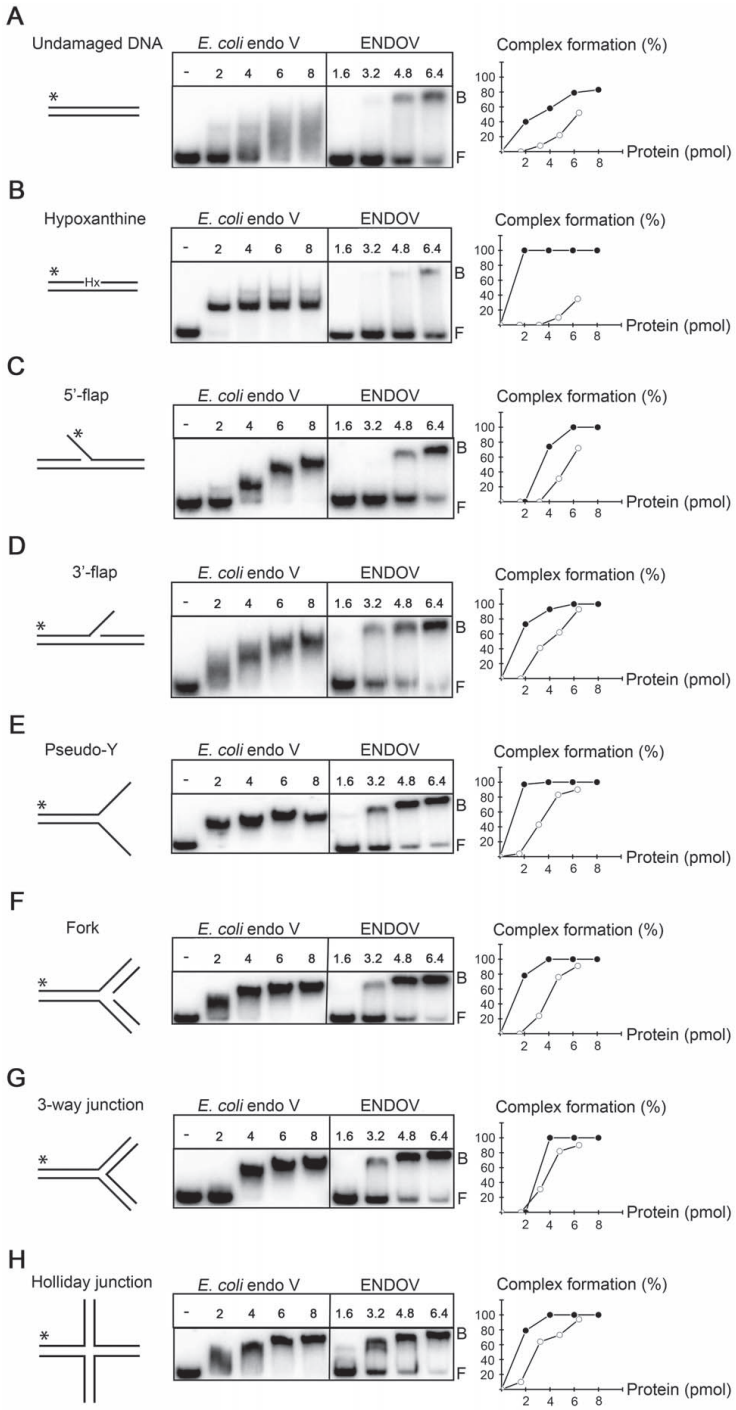

A
Undamaged DNA
E. coli endo V
ENDOV
- 2 4 6 8 1.6 3.2 4.8 6.4
B
F
Complex formation (%)
100 80 60 40 20
2 4 6 8
Protein (pmol)
B
Hypoxanthine
Hx
C
5'-flap
D
3'-flap
E
Pseudo-Y
F
Fork
G
3-way junction
H
Holliday junction

Figure 5. Human ENDOV binds branched DNA substrates. The affinities of *E. coli* and human endonuclease V for different DNA substrates were tested by electrophoretic mobility shift assay. Substrates used were: (**A**): Undamaged DNA, (**B**): Hypoxanthine, (**C**): 5′-flap, (**D**): 3′-flap, (**E**): pseudo-Y, (**F**): fork, (**G**): 3-way junction and (**H**): Holliday junction (asterisk indicates the ^{32}P-labelled strands). 2–8 pmol *E. coli* endoV and 1.6–6.4 pmol ENDOV enzymes were assayed with 10 fmol substrates as indicated. All experiments were repeated 2 to 3 times and a representative assay is shown. Bound substrates relative to free were quantified and are shown to the right. F = free DNA, B = bound DNA, - = no enzyme added, filled circles = *E. coli* endoV; open circles = ENDOV.
doi:10.1371/journal.pone.0047466.g005

Northern blot analysis showed that the majority of transcripts were 1200–1400 nt long. This probably corresponds to an *ENDOV* transcript containing exon 9 and lacking exon 10, with identical splicing as for example in mouse, pig, and sea urchin (Figure 1A). Less abundant is a larger transcript hybridizing to an exon 10 specific probe which could correspond to an isoform harboring a short exon 9 in addition to exon 10 (2800 nt). Variation in the 3′ end of the *ENDOV* transcript allows for different C-termini of the expressed protein. Whereas amino acids 1–250, encoded by exons 1–8, constitute the structurally ordered core domain required for ENDOV function in prokaryotes, the C-terminus is structurally flexible and apparently evolving neutrally. Prokaryotic proteins, as well as the orthologs from frog, *S. pombe* and several other eukaryotes (Figure S1) do not possess the extended C-terminus, suggesting that it may have no significant function. Both the biochemical analyses and localization studies were performed with the two ENDOV isoforms 1–9 (data not shown) and 1–10, always with similar results, supporting a non-essential role for the C-terminus.

Prokaryotic endonuclease V proteins are considered the principal enzyme for Hx repair and quite surprisingly we were unable to demonstrate such an activity for human ENDOV. It was previously shown that mouse ENDOV processes Hx in single stranded DNA. However, to get cleavage of a double stranded Hx substrate, micromolar concentrations of the enzyme was needed [28], questioning the efficiency of the Hx activity also for mouse ENDOV. *E. coli* endoV and the other characterized bacterial endoV enzymes have clearly a high and robust Hx nicking activity *in vitro*, but their role *in vivo* is more unclear. For instance, *E. coli* cells lacking endoV have no pronounced phenotype except an increased mutation frequency when exposed to nitrous acid [16]. In *E. coli*, another enzyme, the AlkA DNA glycosylase, also acts on Hx in DNA [40], suggesting redundancy in the repair of Hx. Conversely, deletion of *alkA* in an *nfi* mutant does not increase the mutation frequency of the *nfi* cells questioning the assumption of redundancy [16]. Another distinct role for *E. coli* endoV was found by analysis of the deoxyinosine triphosphatase *rdgB* mutant, which in combination with a *recA* mutant is lethal. RgdB functions in degradation of dITP and hence, an *rdgB* mutant has increased levels of dITP in the nucleotide pool and consequently also Hx in the DNA. The lethality of the double mutant is believed to be associated with endoV induced strand breaks at Hx which cannot be repaired in the recombination defect *recA* background. Simultaneous deletion of *nfi* suppresses the lethality, pointing to a clear role for endoV under these circumstances. It appears that DNA strand incision is important, but none of these studies demonstrate that Hx actually is removed from DNA. In fact, a recent study shows that the genes affecting the Hx levels in DNA all belong to the purine nucleotide metabolism whereas *nfi* has no effect [41].

In this work we were not able to demonstrate endonuclease activity for recombinant human ENDOV. We cannot exclude that this is due to a missing factor or protein partner interacting with ENDOV. The dependence on associated proteins for activity has been shown for several endonucleases such as MUS81, SLX1 and XPF interacting with EMM1 [42–44], SLX4 [44–46] and ERCC1 [47], respectively. Interestingly, these are all structure-specific endonucleases active upon the types of DNA substrates used in this work. We may speculate if the sorting to the nucleoli and the affinity for specific DNA structures such as flaps and branching points could hint to a role of ENDOV in replication or transcription of ribosomal DNA. The rDNA genes are organized in long tandem repeats in hundreds of copies that are dynamic in size (reviewed in [48]). Tight control of replication is required to

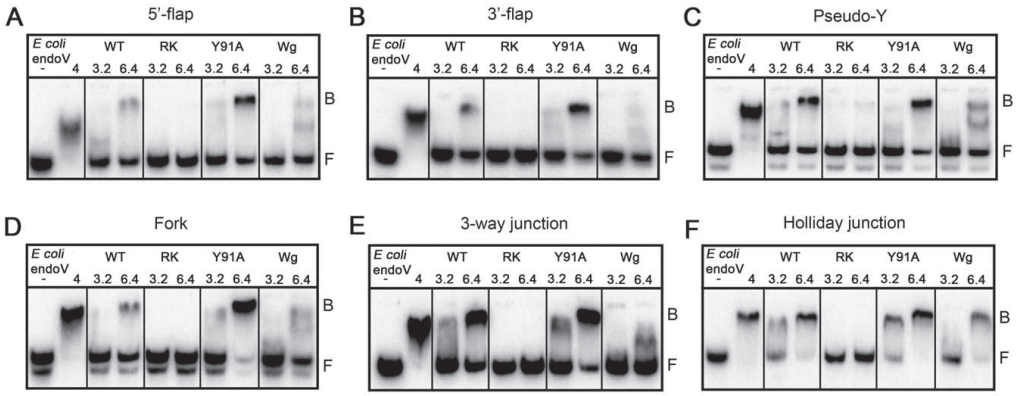


Figure 6. The RK ENDOV mutant has lost its affinity for branched DNA substrates. Three ENDOV site specific mutants, RK (R248E/K249E double mutant), Y91A, and Wg (residues P90–S93 replaced with 4 glycins), wild type enzyme (3.2 and 6.4 pmol) and *E. coli* endoV (4 pmol) were tested for their ability to bind branched DNA substrates. Substrates tested were: (**A**): 5′-flap, (**B**): 3′-flap, (**C**): pseudo-Y, (**D**), fork, (**E**): 3-way junction and (**F**) Holliday junction. F = free DNA, B = bound DNA, - = no enzyme added.
doi:10.1371/journal.pone.0047466.g006

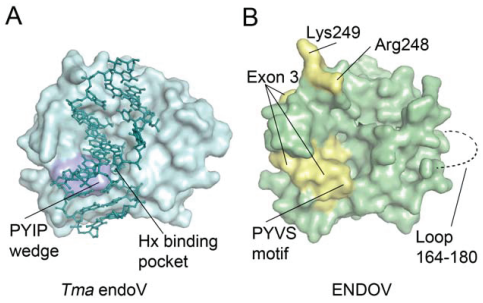


Figure 7. Location of structural elements in endonuclease V. (**A**) Structure of *Tma* endoV binding to deaminated DNA (PDB code 2W35 [23]) showing the central location of the strand-separating wedge (PYIP motif) close to the damage recognition pocket. (**B**). Homology model of human ENDOV showing the location of mutated residues Arg248 and Lys249 (yellow), as well as residues forming the PYVS motif in exon 3 (yellow), which were also mutated. A loop comprising residues 164–180 could not be reliably modelled and is not included (dashed line).
doi:10.1371/journal.pone.0047466.g007

avoid unwanted expansions or contractions. Moreover, replication fork barriers are present in nontranscribed regions which prevent collisions between replication forks and the transcription machinery (reviewed in [49]). Both these processes have the potential to create DNA structures that are possible substrates for ENDOV.

T. maritima endoV has much stronger affinity for Hx-DNA than non-damaged DNA and the protein crystal structure shows that the conserved PYIP wedge is important for presenting the Hx base into the lesion recognition pocket [23]. It appears that human ENDOV have lost the ability to efficiently recognize Hx in DNA although the wedge structure and the base recognition pocket appear to be conserved. However, the binding to branched DNA remains intact in human ENDOV, suggesting that the structural fold may have been evolutionary modified to avoid incision at Hx in DNA in mammalian cells.

During preparation of this report, Mi *et al.* [50] published a paper describing biochemical properties of recombinant ENDOV. In contrast to our data, they find Hx nicking activity, albeit weak, for human ENDOV. We do not know the reason for this discrepancy, but the use of a fusion protein of ENDOV with thioredoxin in their study may influence the results.

Supporting Information

Figure S1 Multiple sequence alignment of 14 eukaryotic endonuclease V homologs. The sequences are from human (RefSeq [36] identifier NP_775898), mouse (*Mus musculus*, NP_001158108), hamster (*Cricetulus griseus*, XP_003496919), rat (*Rattus norvegicus*, GenBank [29] identifier EDM06795), pig (*Sus scrofa*, XP_003131183), chicken (*Gallus gallus*, XP_420082), frog (*Xenopus tropicalis*), medaka ricefish (*Oryzias latipes*), sea urchin (*Strongylocentrotus purpuratus*, XP_794487), hemichordate acorn worm (*Saccoglossus kowalevskii*, XP_002731652), sponge (*Amphimedon queenslandica*, XP_003386872), *Trichoplax adhaerens* (XP_002110086), the fission yeast *Schizosaccharomyces pombe* (NP_594332), and *Arabidopsis thaliana* (NP_567868). In addition, the sequences of the bacterial homologs from *E. coli* (NP_418426) and *Thermotoga maritima* (NP_229661) are shown. The medaka sequence was generated from Ensembl [31] protein ENSORLP00000002194 (exons 1–7), but with the C-terminus derived from EST sequences [29] BJ009743 and BJ023154. The full-length *X. tropicalis* sequence was generated by combining data from IMAGE cDNA clone sequences BC154886 and BC087745 and several ESTs (*e.g.* DN064695). The alignment was generated with Muscle [34]. Conserved and functionally important residues are highlighted above and below the alignment for the human (black) and *T. maritima* (dark blue) endoV homologs, respectively. These includes the residues of the DDD-motif of the catalytic triad (human residues Asp52, Asp126, and Asp240) which together with Glu100 are complexing the divalent cation of the catalytic site, the residues forming the lesion recognition pocket (Tyr91, Gly94, Leu96, Gly127, Asn128, His132, Gly137, and Leu158), as well as the active site stabilizing Lys155. See Dalhus *et al.* [23] for more details on lesion recognition and the catalytic mechanism of endoV.
(JPG)

Figure S2 GFP-ENDOV exists as a fusion protein. Protein extracts were prepared from HeLa cells overexpressing (from left) GFP alone or GFP-ENDOV. Proteins were separated on 10% SDS-PAGE in 1 × MOPS and transferred to PVDF membranes as described in Material and methods. The membrane was probed with a GFP antibody. M is the molecular weight marker as indicated.
(JPG)

Figure S3 Human ENDOV does not cleave a hypoxanthine containing DNA substrate. *E. coli* endoV (2–8 pmol) and ENDOV (1.6–6.4 pmol) were tested for activity towards hypoxanthine DNA. Reaction products were separated by PAGE and visualised by Phosphorimaging. S = substrate, C = cleaved substrate, - = no enzyme added.
(JPG)

Figure S4 SDS-PAGE analysis of purified *E. coli* endoV, human ENDOV wildtype (WT) and mutant (RK, Y91A, Wg) proteins. 1 μg of each protein was analysed by SDS-PAGE. M is the molecular weight marker as indicated.
(JPG)

Table S1 Oligonucleotides for probes and real time-PCR.
(PDF)

Table S2 Oligonucleotides for DNA substrates.
(PDF)

Table S3 Evaluation of the evolutionary selective pressure acting on ENDOV.
(PDF)

Author Contributions

Conceived and designed the experiments: CF JKL BD MB IA. Performed the experiments: CF ESV JEH ET PSA IA CGN. Analyzed the data: CF ESV JKL MB BD IA CGN. Contributed reagents/materials/analysis tools: CF JEH PSA JKL. Wrote the paper: CF ESV JKL BD MB IA.

References

1. Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362: 709–715.
2. Lindahl T, Nyberg B (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry 13: 3405–3410.
3. Chen H, Shaw BR (1993) Kinetics of bisulfite-induced cytosine deamination in single-stranded DNA. Biochemistry 32: 3535–3539.
4. Shapiro R, Pohl SH (1968) The reaction of ribonucleosides with nitrous acid. Side products and kinetics. Biochemistry 7: 448–455.

5. Shapiro R, Shiuey SJ (1969) Reaction of nitrous acid with alkylaminopurines. Biochim Biophys Acta 174: 403–405.
6. Lee DH, Pfeifer GP (2003) Deamination of 5-methylcytosines within cyclobutane pyrimidine dimers is an important component of UVB mutagenesis. J Biol Chem 278: 10314–10321.
7. Green M, Barner HD, Cohen SS (1957) Studies on the biosynthesis of bacterial and viral pyrimidines. V. Hydrogenation of 5-hydroxymethylpyrimidines. J Biol Chem 228: 621–631.
8. Karran P, Lindahl T (1980) Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. Biochemistry 19: 6005–6011.
9. Myrnes B, Guddal PH, Krokan H (1982) Metabolism of dITP in HeLa cell extracts, incorporation into DNA by isolated nuclei and release of hypoxanthine from DNA by a hypoxanthine-DNA glycosylase activity. Nucleic Acids Res 10: 3693–3701.
10. Budke B, Kuzminov A (2006) Hypoxanthine incorporation is nonmutagenic in *Escherichia coli*. J Bacteriol 188: 6553–6560.
11. Sousa MM, Krokan HE, Slupphaug G (2007) DNA-uracil and human pathology. Mol Aspects Med 28: 276–306.
12. Guo G, Ding Y, Weiss B (1997) *nfi*, the gene for endonuclease V in *Escherichia coli* K-12. J Bacteriol 179: 310–316.
13. Gates FT, Linn S (1977) Endonuclease from *Escherichia coli* that acts specifically upon duplex DNA damaged by ultraviolet light, osmium tetroxide, acid, or x-rays. J Biol Chem 252: 2802–2807.
14. Yao M, Hatahet Z, Melamede RJ, Kow YW (1994) Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3′ endonuclease, from *Escherichia coli*. J Biol Chem 269: 16260–16268.
15. Lee CC, Yang YC, Goodman SD, Yu YH, Lin SB, et al. (2010) Endonuclease V-mediated deoxyinosine excision repair in vitro. DNA Repair 9: 1073–1079.
16. Schouten KA, Weiss B (1999) Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. Mutat Res 435: 245–254.
17. Gates FT, III, Linn S (1977) Endonuclease V of *Escherichia coli*. J Biol Chem 252: 1647–1653.
18. Yao M, Kow YW (1997) Further characterization of *Escherichia coli* endonuclease V. Mechanism of recognition for deoxyinosine, deoxyuridine, and base mismatches in DNA. J Biol Chem 272: 30774–30779.
19. He B, Qing H, Kow YW (2000) Deoxyxanthosine in DNA is repaired by *Escherichia coli* endonuclease V. Mutat Res 459: 109–114.
20. Yao M, Kow YW (1994) Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3′-endonuclease from *Escherichia coli*. J Biol Chem 269: 31390–31396.
21. Yao M, Kow YW (1996) Cleavage of insertion/deletion mismatches, flap and pseudo-Y DNA structures by deoxyinosine 3′-endonuclease from *Escherichia coli*. J Biol Chem 271: 30672–30676.
22. Mi R, Abole AK, Cao W (2011) Dissecting endonuclease and exonuclease activities in endonuclease V from *Thermotoga maritima*. Nucleic Acids Res 39: 536–544.
23. Dalhus B, Arvai AS, Rosnes I, Olsen OE, Backe PH A, et al. (2009) Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. Nat Struct Mol Biol 16: 138–143.
24. Liu J, He B, Qing H, Kow YW (2000) A deoxyinosine specific endonuclease from hyperthermophile, *Archaeoglobus fulgidus*: a homolog of *Escherichia coli* endonuclease V. Mutat Res 461: 169–177.
25. Huang J, Lu J, Barany F, Cao W (2001) Multiple cleavage activities of endonuclease V from *Thermotoga maritima*: recognition and strand nicking mechanism. Biochemistry 40: 8738–8748.
26. Kanugula S, Pauly GT, Moschel RC, Pegg AE (2005) A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits O6-alkylguanine-DNA alkyltransferase and endonuclease V activities. Proc Natl Acad Sci USA 102: 3617–3622.
27. Feng H, Klutz AM, Cao W (2005) Active site plasticity of endonuclease V from *Salmonella typhimurium*. Biochemistry 44: 675–683.
28. Moe A, Ringvoll J, Nordstrand LM, Eide L, Bjoras M, et al. (2003) Incision at hypoxanthine residues in DNA by a mammalian homologue of the *Escherichia coli* antimutator enzyme endonuclease V. Nucleic Acids Res 31: 3893–3900.
29. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, et al. (2012) GenBank. Nucleic Acids Res 40: D48–D53.
30. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, et al. (2012) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 40: D13–D25.
31. Flicek P, Amode MR, Barrell D, Beal K, Brent S, et al. (2012) Ensembl 2012. Nucleic Acids Res 40: D84–D90.
32. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, et al. (2012) The UCSC Genome Browser database: extensions and updates 2011. Nucleic Acids Res 40: D918–D923.
33. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337: 635–645.
34. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.
35. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195–201.
36. Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res 40: D130–D135.
37. International chicken sequencing consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695–716.
38. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, et al. (2010) The genome of the Western clawed frog *Xenopus tropicalis*. Science 328: 633–636.
39. Huang J, Lu J, Barany F, Cao W (2002) Mutational analysis of endonuclease V from *Thermotoga maritima*. Biochemistry 41: 8342–8350.
40. Saparbaev M, Laval J (1994) Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. Proc Natl Acad Sci USA 91: 5873–5877.
41. Pang B, McFaline JL, Burgis NE, Dong M, Taghizadeh K, et al. (2012) Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. Proc Natl Acad Sci USA 109: 2319–2324.
42. Boddy MN, Gaillard PH, McDonald WH, Shanahan P, Yates JR III, et al. (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. Cell 107: 537–548.
43. Kaliraman V, Mullen JR, Fricke WM, Bastin-Shanower SA, Brill SJ (2001) Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. Genes Dev 15: 2730–2740.
44. Mullen JR, Kaliraman V, Ibrahim SS, Brill SJ (2001) Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in *Saccharomyces cerevisiae*. Genetics 157: 103–118.
45. Fekairi S, Scaglione S, Chahwan C, Taylor ER, Tissier A, et al. (2009) Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. Cell 138: 78–89.
46. Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, et al. (2009) Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell 138: 63–77.
47. Park CH, Bessho T, Matsunaga T, Sancar A (1995) Purification and characterization of the XPF-ERCC1 complex of human DNA repair excision nuclease. J Biol Chem 270: 22657–22660.
48. Kobayashi T (2006) Strategies to maintain the stability of the ribosomal RNA gene repeats–collaboration of recombination, cohesion, and condensation. Genes Genet Syst 81: 155–161.
49. Grummt I (2003) Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes Dev 17: 1691–1702.
50. Mi R, Ford-Zappala M, Kow YW, Cunningham RP, Cao W (2012) Human endonuclease V as a repair enzyme for DNA deamination. Mutat Res 735: 12–18.

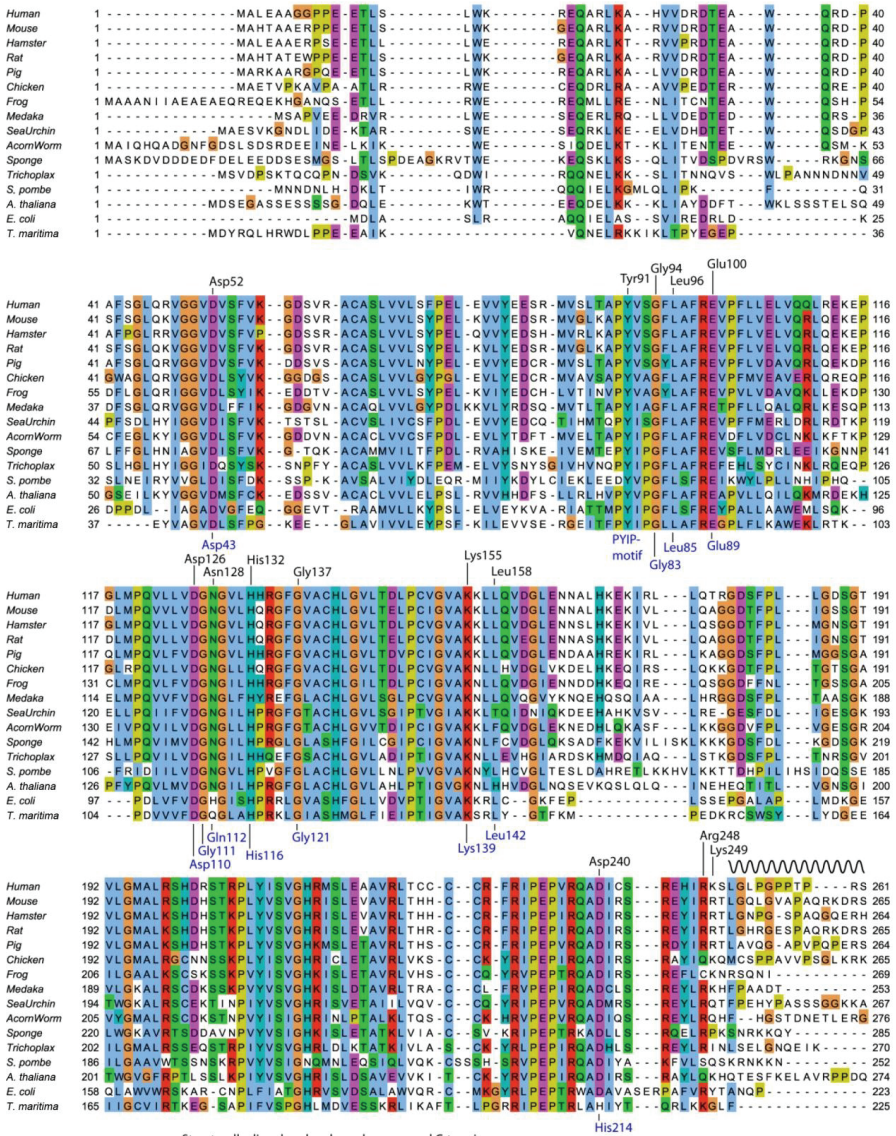

Asp52
Tyr91
Gly94
Leu96
Glu100
PYIP-motif
Leu85
Glu89
Gly83
Asp43
Asp126
Asn128
His132
Gly137
Lys155
Leu158
Gln112
Gly111
Asp110
His116
Gly121
Lys139
Leu142
Arg248
Lys249
Asp240
His214
Structurally disordered and poorly conserved C-terminus
Human
Mouse
Hamster
Rat
Pig
Chicken
Frog
Medaka
SeaUrchin
AcomWorm
Sponge
Trichoplax
S. pombe
A. thaliana
E. coli
T. maritima

Figure S1

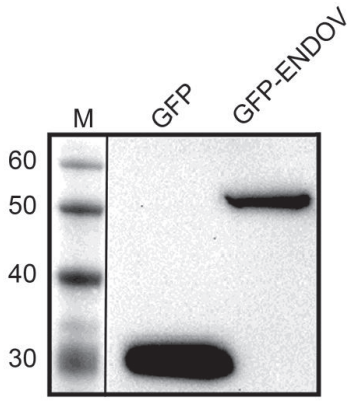


Figure S2

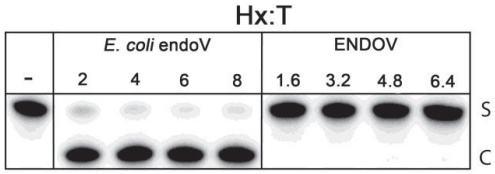


Figure S3

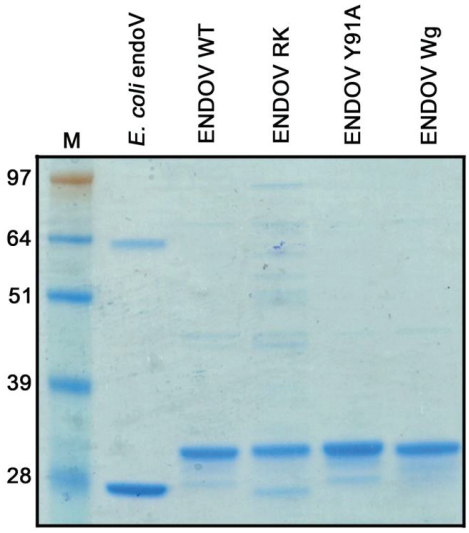


Figure S4

Table S1. Oligonucleotides for probes and real time-PCR

Primer:	***Sequence 5' → 3':***	***Description:***
1	TTGACGTGTCCTTCGTGAAA	Fwd exon 2 RT-PCR
2	ATGCGGCTCTCCTCATACAC	Rev exon 3 RT-PCR
3	ACTGTCCTGGGAATGGC	Fwd exon 6/7 RT-PCR, tissue-scan
4	ACTTGCGGATGTGCTCTCG	Rev exon 8 RT-PCR, tissue-scan
5	CCACATCGCTCAGACACCAT	Fwd GAPDH RT-PCR
6	GCGCCCAATACGACCAAAT	Rev GAPDH RT-PCR
7	GTGGTGTATGAGGAGAGCCG	Fwd exon 3 Nothern
8	CTGGGGCATGAGGCCCG	Rev exon3 Nothern
9	CCTTCTTGTGGATGGAAACGGGGTAC	Fwd exon 4 Nothern
10	AGCGACTTGCGGATGTGCTCTCG	Rev exon 8 Nothern
11	TGGTGGTGAGAGCACACGTC	Fwd exon 10 Northern
12	GGACCGTCTTCAGCTGGATG	Rev exon 10 Northern

Table S2. Oligonucleotides for DNA substrates

Primer:	***Sequence 5′ → 3′:***	***Description:***
1	CGGTGACCGATCTGTAGCTCTACGG	Complementary oligo for ds control and Hx
2	CCGTAGAGCTACAGATCGGTCACCG	ds control*
3	CCGTAGAGCTAC[Hx]GATCGGTCACCG	Hypoxanthine*
4	GCATGCCTGCACGG[U]CATGGCCAGATCCCCGGGTACCGAG	Uracil*
5	CTCGGTACCCGGGGATCTGGCCATGGCCGTGCAGGCATGC	Complementary G for uracil
6	CCGTAGAGCGACAGATCGGTCACCG	Complementary for loop (A)*
7	CGGTGACCGATCTTGTCGCTCTACGG	Loop (TT)
8	TCAACTCTGGAATAAGTGCGTGGTCGGT	Complementary oligo for hairpin*
9	ACCGACCACGCACTGCGCGTTTTCGCGCTATTCCAGAGTTGA	Hairpin
10	GGATACGTAACAACGCTTATGCATCGCCGCCGCTACATCCCTGAGCTGAC	3′-flap, 5′-flap, 3-way*, pseudo-Y*, fork
11	ATGCATAAGCGTTGTTACGTATCC	3′-flap, fork
12	GTCAGCTCAGGGATGTAGCGGCGG	5′-flap
13	TGTGTTCGATCTCGATCAGAATGACGATGCATAAGCGTTGTTACGTATCC	5′-flap*, 3-way, pseudo-Y
14	GTCAGCTCAGGGATGTAGCGGCGGAGTCATTCTGATCGAGATCGAACACA	3′-flap*, 3-way, fork*
15	TGTGTTCGATCTCGATCAGAATGA	fork
16	CCGCTACCAGTGATCACCAATGGATTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC	HJ_1
17	TGGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAATCCATTGTCTATGACGTCAAGCT	HJ_2*
18	GAGCTTGACGTCATAGACAATGGATTGCTAGGACATCTTTGCCGTCTTGTCAATATCGGC	HJ_3
19	TGCCGATATTGACAAGACGGCAAAGATGTCCTAGCAATCCATTGGTGATCACTGGTAGCGG	HJ_4

*=^{32}P labelled strand

Table S3. Evaluation of the evolutionary selective pressure acting on ENDOV. Ratio of amino acid replacing (*Ka*) and silent (*Ks*) mutation rates for pairwise comparisons of the endonuclease V core regions and C-terminal tails from mouse (GenBank [1] identifier NM_001164636), rat (Ensembl [2] transcript ENSRNOT00000054986) and hamster (XM_003496871) *ENDOV*. These species were selected for analysis because the mRNAs are processed by identical splicing of the same nine exons in the three rodents, simplifying analysis. *Ks* is the number of synonymous substitutions per synonymous site and *Ka* is the number of amino acid changing substitutions per non-synonymous site [3]. The calculation of the *Ka*/*Ks* ratio is a powerful tool for evaluating evolutionary selective pressure, with *Ka*/*Ks* < 1 indicating purifying (negative) selection, while *Ka*/*Ks* ~ 1 is indicating neutral evolution of the protein segment.

	Ka/*Ks*[a]	
	Mm ENDOV core	*Rn* ENDOV core
Rn ENDOV core	0.129 (0.123)	
Cg ENDOV core	0.233 (0.230)	0.214 (0.210)

	Ka/*Ks*[b]	
	Mm ENDOV C-term	*Rn* ENDOV C-term
Rn ENDOV C-term	1.424 (1.457)	
Cg ENDOV C-term	1.087 (1.137)	1.752 (1.880)

[a]*Ka*/*Ks* ratios for pairwise comparisons of ENDOV from mouse (*Mm*), rat (*Rn*), and Chinese hamster (*Cg*) for the core region (residues 1-250) calculated according to the model of Goldman and Yang [4] and the model averaging method of Zhang *et al.* (in parenthesis) [5] with the KaKs_Calculator [6].
[b]*Ka*/*Ks* ratios for pairwise comparisons of the C-terminal segments (codon 251 – stop) of ENDOV for the three rodents calculated as described above. Sites corresponding to insertions/deletions were removed.

Supporting Table S3 Reference List

1. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, and Sayers EW (2012) GenBank. Nucleic Acids Res 40: D48-D53.

2. Flicek P, Amode MR, Barrell D, Beal K, Brent S, *et al.* (2012) Ensembl 2012. Nucleic Acids Res 40: D84-D90.

3. Hurst LD (2002) The *Ka*/*Ks* ratio: diagnosing the form of sequence evolution. Trends Genet 18: 486-487.
4. Goldman N and Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11: 725-736.

5. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J (2006) KaKs_Calculator: calculating *K*a and *K*s through model selection and model averaging. Genomics Proteomics Bioinformatics 4: 259-263.
6. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8: 77-80.

II

nature COMMUNICATIONS

ARTICLE

Received 13 Mar 2013 | Accepted 9 Jul 2013 | Published 5 Aug 2013

DOI: 10.1038/ncomms3271 OPEN

Endonuclease V cleaves at inosines in RNA

Erik Sebastian Vik[1,2], Meh Sameen Nawaz[1], Pernille Strøm Andersen[1,2], Cathrine Fladeby[1], Magnar Bjørås[1,2], Bjørn Dalhus[1,2] & Ingrun Alseth[1]

Endonuclease V orthologues are highly conserved proteins found in all kingdoms of life. While the prokaryotic enzymes are DNA repair proteins for removal of deaminated adenosine (inosine) from the genome, no clear role for the eukaryotic counterparts has hitherto been described. Here we report that human endonuclease V (ENDOV) and also *Escherichia coli* endonuclease V are highly active ribonucleases specific for inosine in RNA. Inosines are normal residues in certain RNAs introduced by specific deaminases. Adenosine-to-inosine editing is essential for proper function of these transcripts and defects are linked to various human disease. Here we show that human ENDOV cleaves an RNA substrate containing inosine in a position corresponding to a biologically important site for deamination in the *Gabra-3* transcript of the $GABA_A$ neurotransmitter. Further, human ENDOV specifically incises transfer RNAs with inosine in the wobble position. This previously unknown RNA incision activity may suggest a role for endonuclease V in normal RNA metabolism.

[1] Department of Microbiology, Oslo University Hospital HF and University of Oslo, Rikshospitalet, P.O. Box 4950 Nydalen, 0424 Oslo, Norway. [2] Department of Medical Biochemistry, Oslo University Hospital HF and University of Oslo, Rikshospitalet, P.O. Box 4950 Nydalen, 0424 Oslo, Norway. Correspondence and requests for materials should be addressed to I.A. (email: ingrun.alseth@rr-research.no).

RNA editing covalently alter the nucleotide sequence of RNA transcripts relative to that of the encoding DNA, thereby contributing to gene diversity[1]. One of the most prevalent RNA modifications is adenosine (A) deamination, which results in conversion of the 6-aminopurine ring of A to the 6-oxopurine ring of inosine (I)[2]. Inosine has different base pairing properties to A and is interpreted as guanosine (G) in cells. Enzymes catalyzing this conversion are adenosine deaminases acting on RNA (ADARs)[3], which are conserved enzymes found in most multicellular organisms[4]. Important ADAR targets are mRNAs for neurotransmitter receptors in mammals and editing is critical for normal brain development and behaviour[5]. In all cases, adenosine deamination results in a dramatic alteration of the receptor function[5]. Improper function of ADARs has been correlated with serious neurological and mental human disorders[6]. However, the vast majority of A to I conversions are found in non-coding regions where they are involved in controlling the activity of small RNAs such as short interfering (si)RNA and micro (mi)RNAs[7,8].

Inosine is also a central component of transfer RNA (tRNA) where it is found in the wobble position (I_{34}) in the anticodon loop of certain tRNAs[9]. In *E. coli*, only $tRNA^{Arg}$(ACG) undergoes A-to-I editing, whereas in eukaryotes seven to eight tRNAs contain I_{34} (ref. 10). Deamination of A to I is performed by adenosine-specific deaminases acting on tRNA (ADATs), which are homologues of the ADARs. I_{34} is crucial for decoding during protein synthesis as wobble I allow for base pairing with C, T and A[9,11]. Inosine in ribosomal RNA is unusual[12].

While I is a normal and essential residue in RNA (rI), I in DNA (dI) is regarded as damage because of its miscoding properties. Inosine in DNA is a result of spontaneous or nitrosative stress-induced deamination of dA[13]. To counteract such threats, cells express DNA repair proteins specific for inosine. In *E. coli*, the primary enzyme for the repair of dI is endonuclease V (EndoV)[14], which is encoded by the *nfi* gene[15]. EndoV initiates repair by Mg^{2+}-dependent cleavage of the second phosphodiester bond 3′ to the lesion generating 3′-OH and 5′-P termini[14,16,17]. EndoV incises DNA without removing the inosine nucleotide (nt), thus completion of repair depends on additional proteins, a process that is currently poorly understood. Deaminated adenosines can also be repaired by the base excision repair pathway[18,19]. Homologues of EndoV are found in most species, and despite strong sequence conservation robust dI activity has only been demonstrated for the prokaryotic enzymes[20–26].

Recently, we characterized human ENDOV without detecting any dI incision[21]. Interestingly, when fused to the green fluorescent protein, ENDOV was not found in the nucleus of HeLa cells as expected for a DNA repair protein. Rather, ENDOV localized to nucleoli and the cytoplasm, which are the compartments for RNA.

In the present study, we show that human ENDOV has a strong and specific incision activity on RNA substrates containing rI. Also EndoV from *E. coli*, which is an inosine-specific DNA repair enzyme, cleaves at inosines in RNA. Human ENDOV has a preference for single-stranded substrates, whereas *E. coli* EndoV is equally active on both single- and double-stranded RNA. These

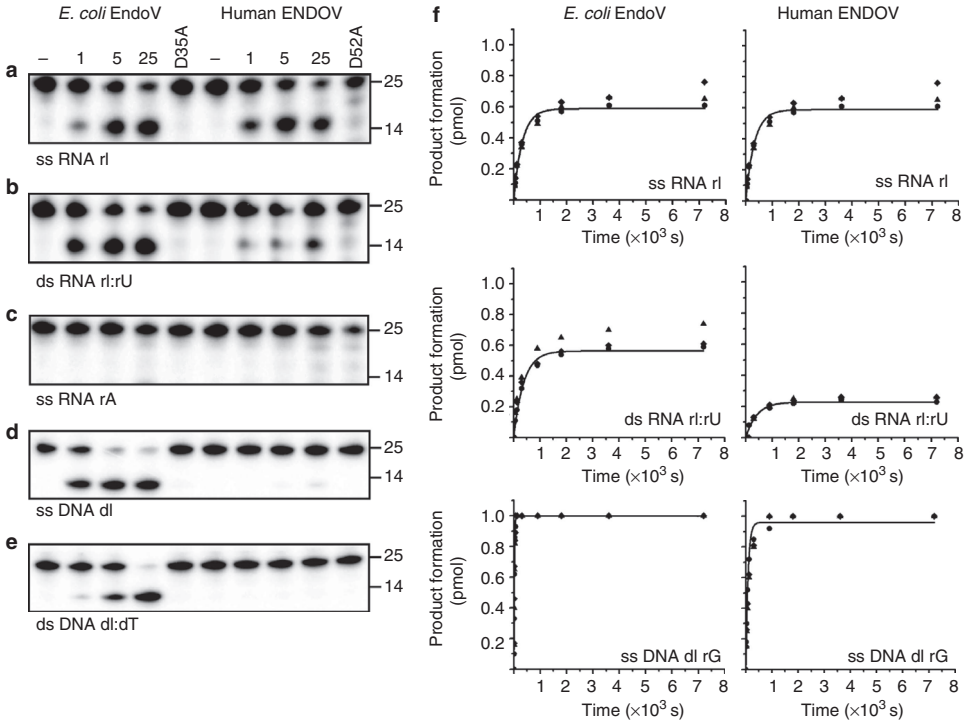


Figure 1 | Various incision activities for *E. coli* and human endonuclease V at inosines in DNA and RNA. The substrates (**a**) single-stranded (ss) RNA rI, (**b**) double-stranded (ds) RNA rI:rU, (**c**) ss RNA rA, (**d**) ssDNA dI, (**e**) ds DNA dI:dT were incubated with the wild-type enzymes (1–25 nM, as indicated) or the two site-specific mutants *E. coli* D35A and human D52A (25 nM) and reaction products analyzed by PAGE. RNA sizes (in nt) are indicated, – = no enzyme added, r = ribonucleotide and d = deoxynucleotide. (**f**) Single-turnover kinetic analysis with enzyme and substrates as indicated.

are the novel findings implying a role for endonuclease V in RNA metabolism.

Results

Human ENDOV is an inosine-specific ribonuclease. ENDOV proteins are highly conserved[16], a feature that normally reflects conserved function—in this case incision at dI residues in DNA. However, human ENDOV appears to be inactive towards dI[21]. As inosines are not only found in DNA, but are also abundant in RNA, we hypothesized that inosines in RNA could be the substrate for ENDOV. Furthermore, the protein structure of endonuclease V from *Thermotoga maritima* reveals an 'RNase H-like motif' supporting link to RNA[16]. Finally, RNA as a substrate for ENDOV is also consistent with the observation of ENDOV–green fluorescent protein fusion proteins in the cytoplasm and nucleoli of HeLa cells[21]. Therefore, human ENDOV was purified as described and analyzed for activity on single- and double-stranded RNA oligonucleotide substrates with centrally located rI residues (Supplementary Table S1). Interestingly, human, as well as *E. coli* endonuclease V, efficiently cleaved both the single- (rI) and double-stranded (rI:rU) RNA substrates (Fig. 1a,b). Both endonucleases were most effective at the highest pH tested (9.5) when Mg^{2+} (5 mM) was used in the reaction, however, with Mn^{2+} as the divalent ion, both enzymes were most active at pH 7.5 (Supplementary Fig. S1a). At pH 7.5, both enzymes were active over a broad range of Mn^{2+} concentrations (0.25–5 mM) (Supplementary Fig. S1b). Neither of the enzymes incised an RNA substrate with a cognate rA instead of rI (Fig. 1c). Single- and double-stranded DNA substrates with inosine were cleaved by the *E. coli* enzyme, but not by the human ENDOV (Fig. 1d,e) as previously described[21].

Mapping of the exact position of cleavage revealed that both enzymes nicked the RNA substrate at the second phosphodiester bond 3′ to the deaminated base as expected (Supplementary Fig. S2a). Single-turnover kinetic analyses revealed a 1.3-fold higher turnover rate (k_{obs}) for human ENDOV on single-stranded than on double-stranded RNA, whereas *E. coli* EndoV showed the same turnover rate for incision of both the substrates (Fig. 1f and Supplementary Table S2). Inosine-containing DNA appears to be the preferred substrate for *E. coli* EndoV, as the turnover rate on inosine-containing single-stranded DNA was twice that of single-stranded RNA and five times higher for double-stranded DNA versus double-stranded RNA (Supplementary Table S2).

RNAs are transcribed as single-stranded molecules but fold spontaneously under physiological conditions to adopt a secondary structure[27]. Thus, all four combinations of base pairs with rI may form, of which rI:rC is the most stable pair[28]. As double-stranded RNA with a rI:rU pair had already been tested (Fig. 1b), the three remaining base pair combinations (rI:rG, rI:rA, rI:rC) were used as substrates in activity assays. *E. coli* EndoV cleaved all double-stranded RNA substrates with lowest affinity for the rI:rC substrate (Supplementary Fig. S2b–d). As is the case for the rI:rU substrate (Fig. 1b), human ENDOV also incised the three other double-stranded RNA substrates (Supplementary Fig. S2b–d).

Site-specific ENDOV mutants were tested for ribonuclease activity. Mutants of the conserved catalytic aspartates (*E. coli*: D35A and human: D52A) could not incise the rI substrates (Fig. 1a–e, Supplementary Fig. S2b–d). Tyrosine 91 of human ENDOV corresponds to Y80 of *T. maritima* EndoV, which is a key residue for dI recognition[16]. The ENDOV Y91A mutant displayed wild-type affinity for branched DNA substrates[21],

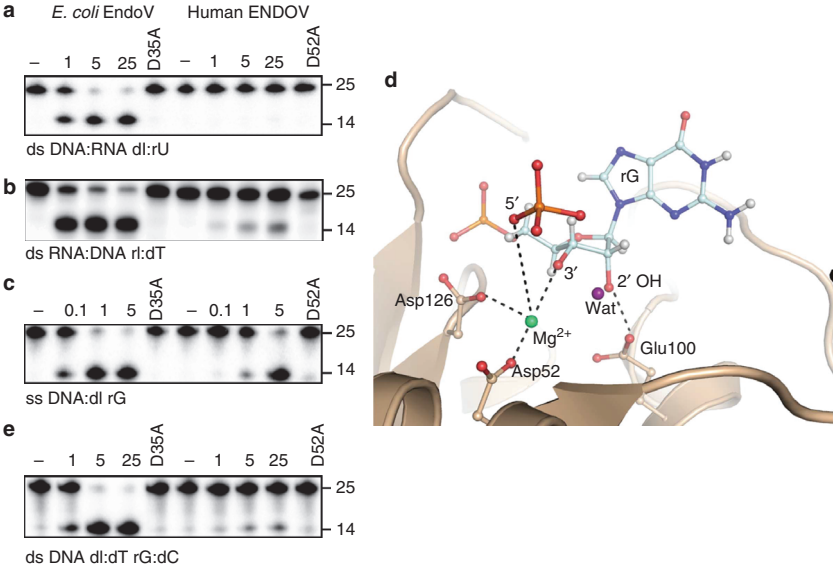


Figure 2 | Incision activities for human ENDOV on RNA depends on a ribonucleotide 3′ to inosine. The substrates (**a**) ds DNA:RNA dI:rU, (**b**) ds RNA:DNA rI:dT, (**c**) ss DNA dI rG were incubated with the wild-type enzymes (0.1–25 nM, as indicated) or the two site-specific mutants *E. coli* D35A and human D52A (25 nM) and reaction products analyzed by PAGE. (**d**) Structural model of the active site of human ENDOV with a ribonucleotide. The 2′ OH group is close to the conserved glutamate (E100), as well as the active site water molecule (Wat) coordinating the Mg^{2+} cofactor, which bridges the 3′- and 5′-ends in the incised product. (**e**) The endonuclease V enzymes were tested for activity against the ds DNA dI:dT rG:dC substrate as in **a**–**c**). RNA sizes (in nt) are indicated, – = no enzyme added, r = ribonucleotide and d = deoxynucleotide.

however, rI incision was totally abolished (Supplementary Fig. S2e). The human mutants RK (R248A, K249A) and Wedge (the four residues PYVS[90–93] mutated to four glycines) previously shown to have defective DNA binding[21] were also inactive on RNA (Supplementary Fig. S2e). These results confirm that endonuclease V enzymes indeed are inosine-specific enzymes, yet the affinity for DNA versus RNA differs.

DNA with a ribonucleotide 3′ to dI is incised by ENDOV. Endonuclease V belongs to the same structural family as RNase H enzymes, which specifically cleaves the RNA strand of RNA:DNA hybrids[16,29]. RNA:DNA hybrids are formed during DNA replication, RNA transcription and reverse transcription, and adopt a different conformation to double-stranded DNA[30]. To test whether ENDOV has activity for mixed RNA:DNA substrates, two hybrid substrates RNA:DNA- and DNA:RNA-containing rI or dI, respectively, were designed. Human ENDOV cleaved only the hybrid with rI in the RNA strand, whereas *E. coli* EndoV cleaved both the substrates (Fig. 2a,b). From these data, we conclude that human ENDOV cleavage is strictly dependent on ribonucleotides in the inosine containing strand. Specifically, we find that ENDOV is highly active on single-stranded DNA when a single ribonucleotide is present directly following the dI residue (Fig. 2c). Hence, the 2′ OH group in this particular ribose seems to be critical and sufficient for incision activity. A model of human ENDOV with rG in this position reveals that the 2′ OH group may interact with the conserved catalytic glutamate (E100) and possibly replace an active site water molecule (Fig. 2d). This assumption is based on comparison with the corresponding coordination sphere around Mg^{2+} in the active site of *T. maritima* EndoV in complex with the cleavage product[16]. The difference in activity of human ENDOV between single- and double-stranded DNA with rG next to dI (Figs 1f and 2c,e), suggests that the nucleic acid helical structure is also critical for substrate processing; incision activity is only observed for double-stranded inosine substrates having A-form helices, like double-stranded RNA and RNA:DNA hybrids[31], but not for double-stranded B-form DNA. A B-form DNA may have steric conflicts with surface-exposed residues in human ENDOV.

***In vivo* and *in vivo*-like RNAs as substrates for ENDOV.** The substrates used above are all random oligonucleotide sequences, and to test for activity towards an *in vivo* deamination target, an oligonucleotide corresponding to a part of the mouse *(m)Gabra-3* transcript of the neurotransmitter $GABA_A$ was synthesized. This 38 nt RNA has an A in position 27, known to be deaminated by adenosine deaminases ADAR1 and ADAR2, and is referred to as the I/M site[32]. Upon folding of this RNA, rI will form a stable base pair with cytosine (C), whereas an unedited A will form a mismatch with C[32]. Human ENDOV, as well as the *E. coli* enzyme, generated an incision product of the expected size of 28 nt when assayed with the rI-containing mGabra3 substrate, whereas no activity was seen with the unedited variant (Fig. 3a,b). This result suggests that ENDOV could be involved in antagonizing the effect of the ADAR enzymes by destruction of rI-containing transcripts.

Next, a substrate corresponding to the anticodon loop in *E. coli* $tRNA^{Arg}$(ACG), wherein the wobble base A34 is the only known A-to-I deaminated tRNA position in prokaryotes[33], was tested. Cleavage within the anticodon loop of tRNAs is a well-documented response to different stress conditions, such as starvation and oxidation, in different organisms[34]. The rI-containing anticodon substrate, but not the unedited A-containing control substrate, was incised by both enzymes, though most efficiently by human ENDOV (Fig. 3c,d). Further, total tRNA isolated from human U373 cells, was incubated with ENDOV and examined for

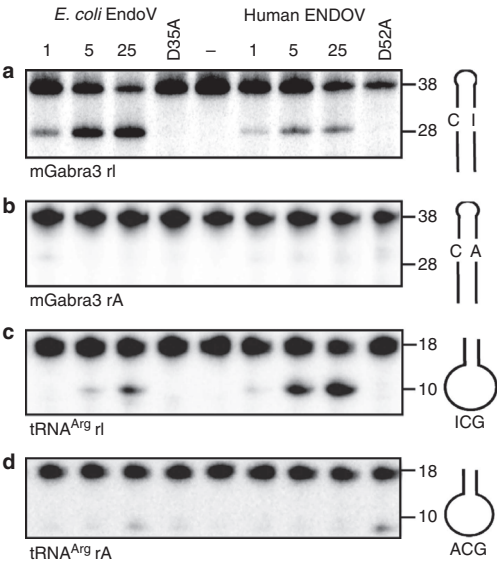


Figure 3 | Human ENDOV cleaves at inosines in *in vivo*-like RNA substrates. The substrates (**a**) mGabra3 rI, (**b**) mGabra3 rA, (**c**) $tRNA^{Arg}$ rI, (**d**) $tRNA^{Arg}$ rA were incubated with the endonuclease V enzymes (1–25 nM, as indicated) and reaction products analyzed by PAGE. RNA sizes (in nt) are indicated, – = no enzyme added, r = ribonucleotide and glyphs illustrate the different substrates.

cleavage. Individual tRNAs were detected by Northern blot hybridization using labelled oligonucleotides complementary to the 5′-fragment of the different tRNAs as probes. Full-length processed tRNAs are 71–82 nt long and cleavage next to the wobble A/I would result in products of 35 (5′-fragment) and 36–47 (3′-fragment) nt. Specific cleavage in the anticodon loop of $tRNA^{Ser}$(AGA), $tRNA^{Leu}$(AAG) and $tRNA^{Arg}$(ACG), all known to have rI in the anticodon loop, was demonstrated (Fig. 4a–c). In contrast, probes for $tRNA^{Asp}$(GTC), $tRNA^{Glu}$(CTC) and $tRNA^{Lys}$(CTT) lacking A/I_{34} editing, revealed no incision products (Fig. 4d–f). No general degradation of the tRNA was seen as total tRNA remained intact after incubation with ENDOV (Supplementary Fig. S2f). These data could imply a role for ENDOV in fragmentation of tRNAs containing rI.

Discussion

In this report, we demonstrate efficient and specific cleavage at single rI residues in RNA by human ENDOV. ENDOV has activity for both single- and double-stranded RNAs, as well as for edited tRNA anticodon loops containing rI. To our knowledge, ribonuclease activity has not been previously described for endonuclease V, and we propose that ENDOV may have an important role in RNA metabolism.

We have previously reported that human ENDOV binds to DNA with higher affinity for branched DNA structures than linear DNA[21]. Here we find that the nucleic acid helical structure is also of importance for ENDOV activity (Fig. 2c–e). It appears that the helical conformation (A- or B-form) is more important for binding than the type of nucleic acid (DNA or RNA), whereas catalytic activity is critically dependent on a ribonucletide in the correct position with a 2′ hydroxyl group close to the active site metal cofactor. At present, we do not know whether ENDOV

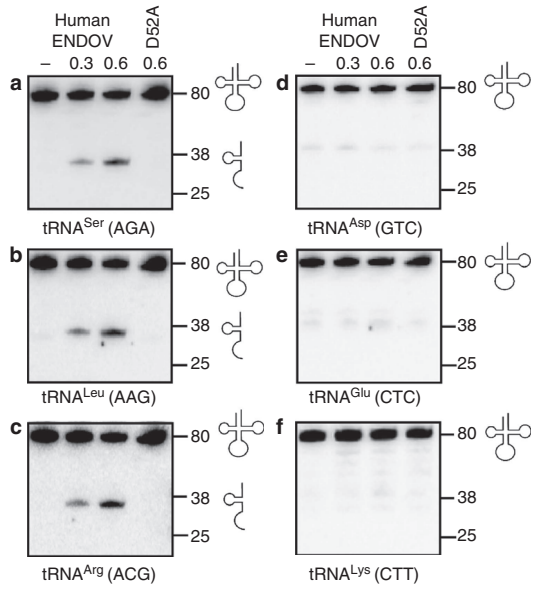


Figure 4 | Human ENDOV cleaves at inosines in tRNA. Northern blots of tRNA isolated from human U373 cells incubated with human wild-type (0.3 and 0.6 μM) or mutant D52A (0.6 μM) ENDOV enzymes, hybridized with DNA probes complementary to the 5′-termini of (**a**) $tRNA^{Ser}$(AGA), (**b**) $tRNA^{Leu}$(AAG), (**c**) $tRNA^{Arg}$(ACG), (**d**) $tRNA^{Asp}$(GTC), (**e**) $tRNA^{Glu}$(CTC) and (**f**) $tRNA^{Lys}$(CTT). Size markers (80, 38 and 25 nt) are shown and glyphs indicate full-length and fragment tRNA species.

binding to DNA is biologically important or whether it simply reflects a general affinity of ENDOV for nucleic acids.

Several reports demonstrate that deamination of adenosine to inosine by the ADAR enzymes has a central function *in vivo*[5,6]. We may speculate that the ribonuclease activity of ENDOV could antagonize the effect of the ADAR enzymes by specific cleavage and destruction of edited transcripts. As many ADAR targets are neurotransmitters, fine-tuning of receptor activity by regulation of the level of edited versus non-edited forms could be of major importance for optimal brain function. It should be mentioned that the Tudor-SN nuclease, which is part of the RISC complex, has also been coupled to cleavage of inosines in RNA[35,36]. It is proposed that Tudor-SN cleaves or promotes the cleavage of hyperedited double-stranded RNA by acting as an activator, however, its specific enzymatic properties have not been thoroughly investigated[35,36].

Cleavage of tRNAs within the anticodon loop is part of the cellular response to stress conditions such as oxidation and starvation in many organisms[37]. Nucleases responsible for this cleavage have been identified, and defects in tRNA fragmentation are associated with apoptosis, cancer and disease progression[34]. As far as we know, specific cleavage of tRNAs at inosines has not previously been described, but an analogous system is the killer toxin zymocin from *Kluyveromyces lactis*, which depends on the wobble uridine modification 5-methoxy-carbonyl-methyl for tRNA cleavage[38].

Interestingly, *E. coli* EndoV was also active on inosines in RNA, demonstrating that EndoV is more than a DNA repair enzyme. Enzymatic deamination of adenosines in RNA appears to be uncommon in prokaryotes, and we may speculate that a more general role exists for endonuclease V enzymes in removal of damaged/deaminated RNA transcripts. Recently, several DNA repair proteins including AlkB/ABH2, SMUG1, APE1 and TDP2 have been shown to possess activity on damaged RNA[39–43] and this is suggested to function as a quality control for RNA. In any case, tight regulation of the ENDOV ribonuclease activity to prevent aberrant cleavage, for example, by compartmentalization, post-translational modification or interacting partners, will be of great importance.

Methods

Expression and purification of endonuclease V enzymes. Wild-type and mutants human ENDOV proteins were expressed in *E. coli* BL21-Codon Plus (DE3)-RIPL cells (Agilent Technologies) as fusion proteins with an N-terminal His-MBP tag[21]. After induction of protein expression with isopropyl β-D-1 thiogalactopyranoside (0.25 mM), bacteria were grown at 18 °C over night. Cells were harvested by centrifugation and the pellet resuspended in 50 mM Tris HCl pH 8.0, 300 mM NaCl, 10 mM imidazole and 10 mM β-mercaptoethanol (β-ME) (buffer A). Cells were lysed by sonication and the cleared lysate applied to Ni-nitrilotriacetic acid affinity chromatography. Recombinant His-MBP-ENDOV was eluted with 300 mM imidazole in buffer A. Peak fractions were pooled and dialyzed at 4 °C in 50 mM Tris HCl pH 8.0, 0.5 mM EDTA, 1 mM dithiothreitol (TEV buffer). TEV protease was added (ratio 1:100) and dialysis continued at 12 °C over night. After proteolysis, the protein mixtures were dialysed against buffer A and the free His-MBP and TEV proteins were separated from ENDOV by a second Ni-NTA purification step. The untagged ENDOV proteins were collected in the flow-through and wash fractions, concentrated and applied to a Superdex 75 size-exclusion chromatography column (GE Healthcare) equilibrated with 50 mM Tris HCl pH 8.0, 50 mM NaCl and 10 mM β-ME. Purified ENDOV was concentrated and stored at − 20 °C. The two inactive endonuclease V mutants *E. coli* D35A and human D52A were made by site-specific mutagenesis using the forward primers 5′-ACCGGATCTGATCGCCGGAGCCGCTGTCGGGTTTGAGCAGGGC-3′ for D35A and 5′-CAGCGTGTGGGCGGTGTGGCTGTTAGTTTCGTGAAAG GTG-3′ for D52A with their corresponding reverse and complementary oligonucleotides (mutated codons are underlined).

DNA/RNA substrates. All DNA/RNA substrates were oligonucleotides synthesized by The Midland Certified Reagent Company, and are listed in Supplementary Table S1. The oligonucleotides were 5′-end labelled with [γ-^{32}P]ATP (Perkin Elmer) using T4 polynucleotide kinase (New England BioLabs). Polynucleotide kinase was inactivated by the addition of 10 mM EDTA. Double-stranded substrates were made by adding the complementary strand and heating to 55 °C followed by slowly cooling to room temperature. The DNA and RNA substrates were separated by 20% native polyacrylamide gel electrophoresis (PAGE; Long-Ranger Gel Solution, Lonza, 0.5xTBE), excised from the gel, eluted by diffusion in H_2O and stored at 4 °C.

DNA and RNA incision activities and single-turnover analysis. The DNA and RNA nicking assays were performed with 1 nM substrate, 0.1–25 nM of enzyme (as indicated in the figures) and standard reaction buffer (10 mM Tris-HCl pH 7.5, 0.5 mM $MnCl_2$, 50 mM KCl, 1 mM dithiothreitol and 5% glycerol) in a total volume of 10 μl or with changes are indicated. To adjust for pH, Tris-HCl, pH 7.5, 8.5 or 9.5 was used and divalent ions kept at 5 mM. When testing Mn^{2+} concentrations, Tris-HCl pH 7.5 was used and different amounts of $MnCl_2$ added (0.25–5 mM). The samples were incubated at 37 °C for 30 min and formamide loading buffer (90% formamide, 0.1% xylene xyanol and 0.1% bromphenol blue) was added to terminate the reactions. Samples were heated to 50 °C for 5 min before separation of substrate and product by 15% denaturating PAGE (Long Ranger, 7 M urea and 1x taurine). To map ENDOV cleavage position, samples were run on a 20% sequencing gel together with ^{32}P-labelled RNA oligonucleotides of defined lengths (13–16 nt; Supplementary Table S1, primers 16–19). The results were visualized by phosphorimaging (Typhon 9410, Amersham Biosciences) and quantified with ImageQuant TL (Molecular Dynamics). Full size images of all panels presented are found in Supplementary Fig. S3 in the order they appear in the manuscript.

For the single-turnover assays, 5 nM of endonuclease V was incubated with 1 nM substrate in 100-μl reaction volume using the same buffer as above. Samples were withdrawn after 3, 8, 15, 30, 60, 120, 300, 900, 1,800, 3,600 and 7,200 s and reactions stopped and analyzed as described above. Quenched samples were kept on ice until all time points were collected. Results were visualized and quantified as above. For the calculation of the catalytic turnover rate k_{obs} (s^{-1}), a one phase association model was fitted to three parallel data sets.

Modelling. A model of human ENDOV in complex with DNA dI rG was constructed by superposing the homology model of human ENDOV[21] with the *T. maritima* EndoV in complex with incised DNA[44] (PDB id 2W35) and

subsequently replacing the dG 3′ to the inosine with an rG nt retrieved from a 0.95-Å resolution X-ray structure of a double-helix RNA fragment[45] (PDB id 3NJ6).

Cell culture and Northern blot analysis. Human epithelial glioblastoma cells U373-MG (American Type Culture Collection) were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% fetal calf serum (PAA lab), 1x GlutaMAX (200 mM, Gibco) and 1 × penicillin–streptomycin (Lonza) at 37 °C in 5% CO_2 atmosphere. Total tRNAs were isolated using RNAzol RT (Molecular Research Center) according to the manufacturer's recommendations. Gel electrophoresis showed that the isolates contained mainly tRNA. Total tRNA (4.5 nM) was incubated with ENDOV (0.3 and 0.6 μM) as described under activity assays. After heat denaturation at 55 °C for 5 min, samples were separated by 15% PAGE in 1x taurine at 200 V for 50 min. The tRNA was transferred to a nylon membrane (Hybond XL, GE Healthcare) by electroblotting in 1x taurine at 4 °C, 5 V and over night. RNA was UV-crosslinked to the membranes (120 mJ cm^{-2} in a CL-1000 UV-Crosslinker, UVP). The Northern Max kit (Ambion, Applied Biosystems) was used for prehybridization, hybridization and washing steps as described by the manufacturer. ^{32}P 5′-end-labelled oligonucleotides (Eurofins; Supplementary Table S3) complementary to the 5′-ends of $tRNA^{Ser}$(AGA), $tRNA^{Leu}$(AAG), $tRNA^{Arg}$(ACG), $tRNA^{Asp}$(GTC), $tRNA^{Glu}$(CTC) and $tRNA^{Lys}$(CTT) were used as probes. Hybridization signals were analyzed by phosphorimaging and ImageQuant TL software. If endonuclease V incises the tRNA at the wobble I, the expected size of the 5′-cleavage product is 35 nt. Markers used were ^{32}P labelled RNA oligonucleotides (25 and 38 nt), which will migrate slightly faster than unlabelled fragments due to the negative charge of the phosphate label. Hybridized probes were removed from the filters by boiling in 0.1% SDS. Full size images of the Northern blots presented are found in Supplementary Fig. S3n–s.

References

1. Gray, M. W. Evolutionary origin of RNA editing. *Biochemistry* **51,** 5235–5242 (2012).
2. Keegan, L. P., Leroy, A., Sproul, D. & O'Connell, M. A. Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome Biol.* **5,** 209 (2004).
3. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71,** 817–846 (2002).
4. Jin, Y., Zhang, W. & Li, Q. Origins and evolution of ADAR-mediated RNA editing. *IUBMB. Life* **61,** 572–578 (2009).
5. Rosenthal, J. J. & Seeburg, P. H. A-to-I RNA editing: effects on proteins key to neural excitability. *Neuron* **74,** 432–439 (2012).
6. Maas, S., Kawahara, Y., Tamburro, K. M. & Nishikura, K. A-to-I RNA editing and human disease. *RNA. Biol.* **3,** 1–9 (2006).
7. Nishikura, K. Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat. Rev. Mol. Cell Biol.* **7,** 919–931 (2006).
8. Orlandi, C., Barbon, A. & Barlati, S. Activity regulation of adenosine deaminases acting on RNA (ADARs). *Mol. Neurobiol.* **45,** 61–75 (2012).
9. Agris, P. F., Vendeix, F. A. & Graham, W. D. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366,** 1–13 (2007).
10. Su, A. A. & Randau, L. A-to-I and C-to-U editing within transfer RNAs. *Biochemistry (Mosc.)* **76,** 932–937 (2011).
11. Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F. & Ochoa, S. Synthetic polynucleotides and the amino acid code. V. *Proc. Natl Acad. Sci. USA* **48,** 613–616 (1962).
12. Cantara, W. A. *et al.* The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39,** D195–D201 (2011).
13. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
14. Yao, M., Hatahet, Z., Melamede, R. J. & Kow, Y. W. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3' endonuclease, from *Escherichia coli*. *J. Biol. Chem.* **269,** 16260–16268 (1994).
15. Guo, G., Ding, Y. & Weiss, B. *nfi*, the gene for endonuclease V in *Escherichia coli* K-12. *J. Bacteriol.* **179,** 310–316 (1997).
16. Dalhus, B. *et al.* Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **16,** 138–143 (2009).
17. Gates, F. T. & Linn, S. Endonuclease from Escherichia coli that acts specifically upon duplex DNA damaged by ultraviolet light, osmium tetroxide, acid, or x-rays. *J. Biol. Chem.* **252,** 2802–2807 (1977).
18. Hardeland, U., Bentele, M., Jiricny, J. & Schar, P. The versatile thymine DNA-glycosylase: a comparative characterization of the human, *Drosophila* and fission yeast orthologs. *Nucleic Acids Res.* **31,** 2261–2271 (2003).
19. Saparbaev, M., Mani, J. C. & Laval, J. Interactions of the human, rat, *Saccharomyces cerevisiae* and *Escherichia coli* 3-methyladenine-DNA glycosylases with DNA containing dIMP residues. *Nucleic Acids Res.* **28,** 1332–1339 (2000).
20. Feng, H., Klutz, A. M. & Cao, W. Active site plasticity of endonuclease V from *Salmonella Typhimurium*. *Biochemistry* **44,** 675–683 (2005).
21. Fladeby, C. *et al.* The human homolog of *Escherichia coli* endonuclease V is a nucleolar protein with affinity for branched DNA structures. *PLoS ONE* **7,** e47466 (2012).
22. Huang, J., Lu, J., Barany, F. & Cao, W. Multiple cleavage activities of endonuclease V from *Thermotoga Maritima*: recognition and strand nicking mechanism. *Biochemistry* **40,** 8738–8748 (2001).
23. Kanugula, S., Pauly, G. T., Moschel, R. C. & Pegg, A. E. A bifunctional DNA repair protein from *Ferroplasma acidarmanus* exhibits O6-alkylguanine-DNA alkyltransferase and endonuclease V activities. *Proc. Natl Acad. Sci. USA* **102,** 3617–3622 (2005).
24. Liu, J., He, B., Qing, H. & Kow, Y. W. A deoxyinosine specific endonuclease from hyperthermophile, *Archaeoglobus fulgidus*: a homolog of *Escherichia coli* endonuclease V. *Mutat. Res.* **461,** 169–177 (2000).
25. Mi, R., ford-Zappala, M., Kow, Y. W., Cunningham, R. P. & Cao, W. Human endonuclease V as a repair enzyme for DNA deamination. *Mutat. Res.* **735,** 12–18 (2012).
26. Moe, A. *et al.* Incision at hypoxanthine residues in DNA by a mammalian homologue of the *Escherichia coli* antimutator enzyme endonuclease V. *Nucleic Acids Res.* **31,** 3893–3900 (2003).
27. Onoa, B. & Tinoco, Jr I. RNA folding and unfolding. *Curr. Opin. Struct. Biol.* **14,** 374–379 (2004).
28. Vendeix, F. A., Munoz, A. M. & Agris, P. F. Free energy calculation of modified base-pair formation in explicit solvent: a predictive model. *RNA.* **15,** 2278–2287 (2009).
29. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276,** 1494–1505 (2009).
30. Shaw, N. N. & Arya, D. P. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* **90,** 1026–1039 (2008).
31. Noy, A., Perez, A., Marquez, M., Luque, F. J. & Orozco, M. Structure, recognition properties, and flexibility of the DNA:RNA hybrid. *J. Am. Chem. Soc.* **127,** 4910–4920 (2005).
32. Daniel, C., Veno, M. T., Ekdahl, Y., Kjems, J. & Ohman, M. A distant cis acting intronic element induces site-selective RNA editing. *Nucleic Acids Res.* **40,** 9876–9886 (2012).
33. Grosjean, H., de Crecy-Lagard, V. & Marck, C. Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.* **584,** 252–264 (2010).
34. Thompson, D. M. & Parker, R. Stressing out over tRNA cleavage. *Cell* **138,** 215–219 (2009).
35. Scadden, A. D. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat. Struct. Mol. Biol.* **12,** 489–496 (2005).
36. Yang, W. *et al.* Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* **13,** 13–21 (2006).
37. Phizicky, E. M. & Hopper, A. K. tRNA biology charges to the front. *Genes Dev.* **24,** 1832–1860 (2010).
38. Lu, J., Huang, B., Esberg, A., Johansson, M. J. & Bystrom, A. S. The *Kluyveromyces lactis* gamma-toxin targets tRNA anticodons. *RNA* **11,** 1648–1654 (2005).
39. Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421,** 859–863 (2003).
40. Berquist, B. R., McNeill, D. R. & Wilson, III D. M. Characterization of abasic endonuclease activity of human Ape1 on alternative substrates, as well as effects of ATP and sequence context on AP site incision. *J. Mol. Biol.* **379,** 17–27 (2008).
41. Jobert, L. *et al.* The human base excision repair enzyme SMUG1 directly interacts with DKC1 and contributes to RNA quality control. *Mol. Cell* **49,** 339–345 (2013).
42. Vascotto, C. *et al.* APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process. *Mol. Cell Biol.* **29,** 1834–1854 (2009).
43. Virgen-Slane, R. *et al.* An RNA virus hijacks an incognito function of a DNA repair enzyme. *Proc. Natl Acad. Sci. USA* **109,** 14634–14639 (2012).
44. Dalhus, B. *et al.* Structural insight into repair of alkylated DNA by a new superfamily of DNA glycosylases comprising HEAT-like repeats. *Nucleic Acids Res.* **35,** 2451–2459 (2007).
45. Kiliszek, A., Kierzek, R., Krzyzosiak, W. J. & Rypniewski, W. Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res.* **38,** 8370–8376 (2010).

Acknowledgements

We acknowledge financial support from the Norwegian Research Council and the Norwegian Cancer Society. E.S.V., B.D. and I.A. have additionally received support from the MLS^{UiO} program for Molecular Life Science research at the University of Oslo. M.B. and B.D. have received support from the South-Eastern Norway Regional Health Authority (grants no. 2009100, 2011040 and 2012085) for establishing the Regional Core Facility for Structural Biology and Bioinformatics. We thank Rune J. Forstrøm for technical assistance and Alexander Rowe for editing the manuscript.

Author contributions

B.D., M.B. and I.A. conceived and planned the study. P.S.A. purified proteins, E.S.V. and M.S.N. performed the activity assays, B.D. made the 3D homology model. E.S.V., B.D.,

M.B., C.F. and I.A. wrote the paper. All authors discussed the results and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

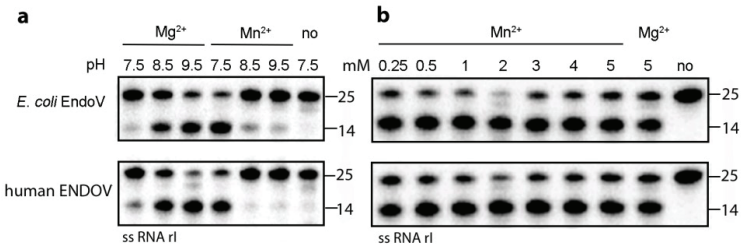
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Vik, E. S. *et al.* Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* 4:2271 doi: 10.1038/ncomms3271 (2013).

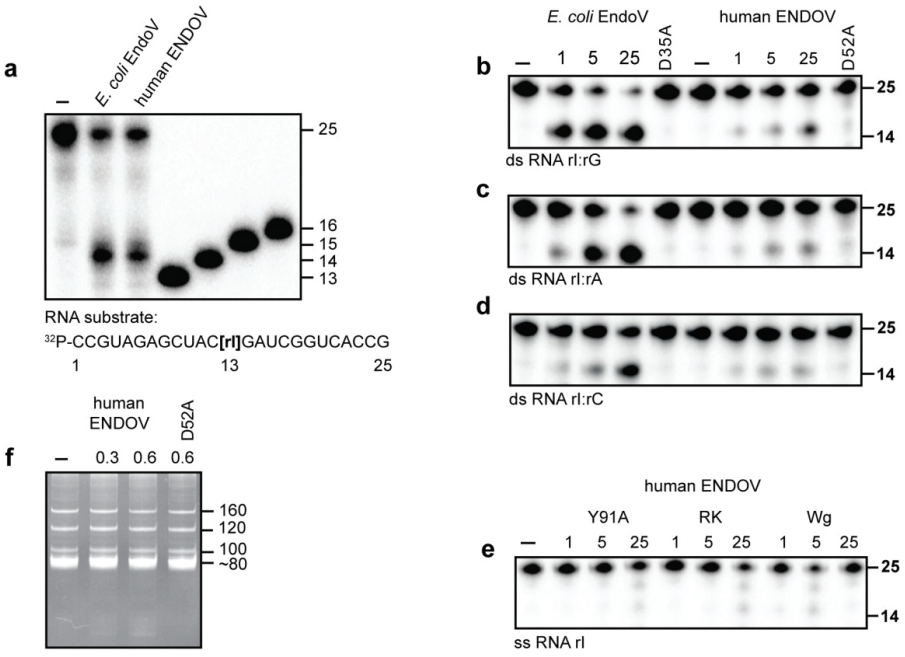
Supplementary information

Supplementary Figure-S1 (Alseth)



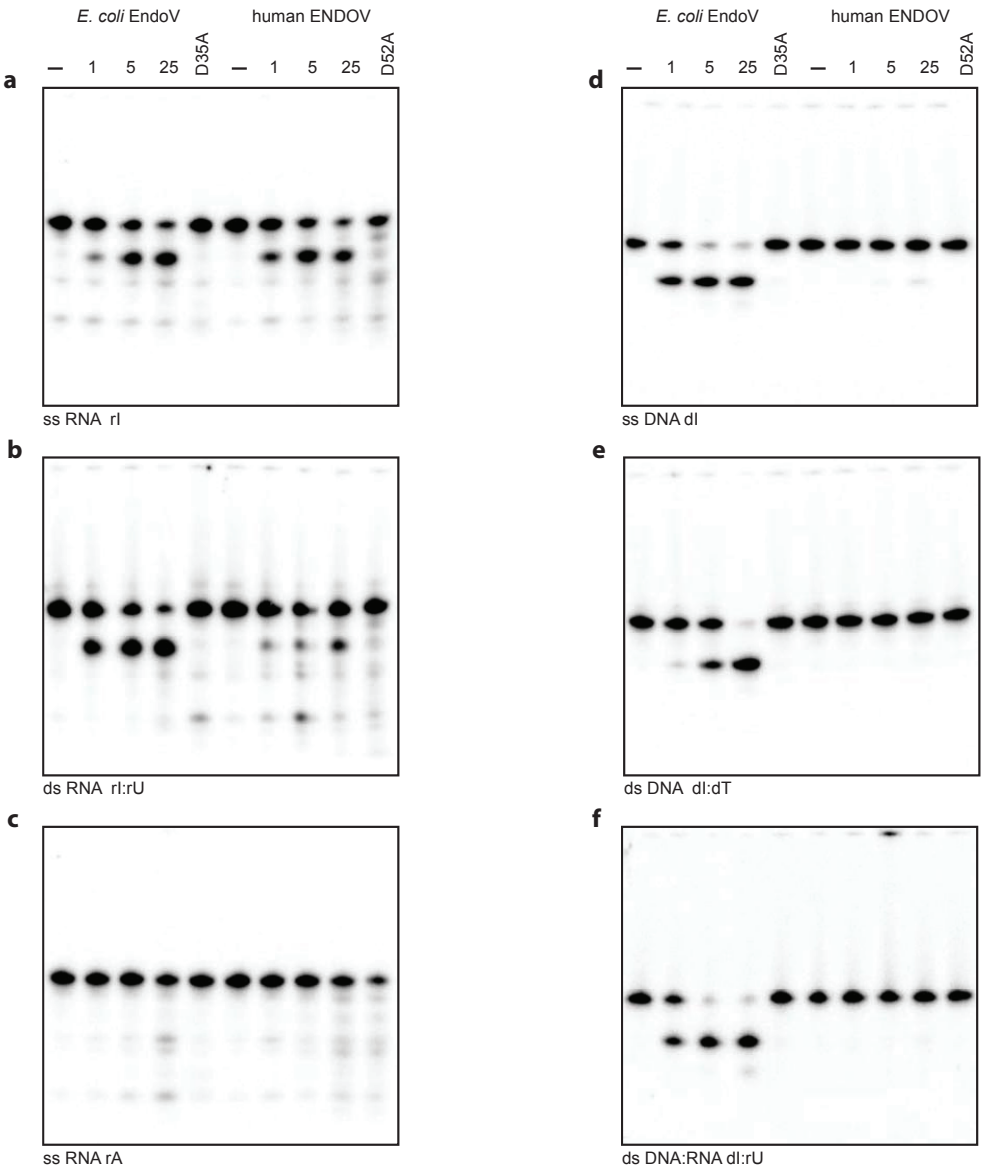
Effect of pH, Mg^{2+} and Mn^{2+} concentrations on endonuclease V activity. (**a**) Single-stranded RNA substrate with rI was incubated with endonuclease V enzymes (25 nM) in buffers with varying pH (7.5-9.5) and divalent ions (5 mM) as indicated and cleavage products were separated by 15% denaturating PAGE. Term "no" indicates same amount of enzyme in the sample, but neither Mn^{2+} nor Mg^{2+} was added. (**b**) Increasing amounts of Mn^{2+} (0.25-5 mM) was added to a buffer of pH 7.5 together with *E. coli* or human endonuclease V (25 nM) and activity on ss RNA with rI was assayed as above. For comparison, a sample with 5 mM Mg^{2+}, pH 8.5 was included. In this figure "no" indicates that no enzyme was added. Size markers (nt) are indicated, r = ribonucleotide.

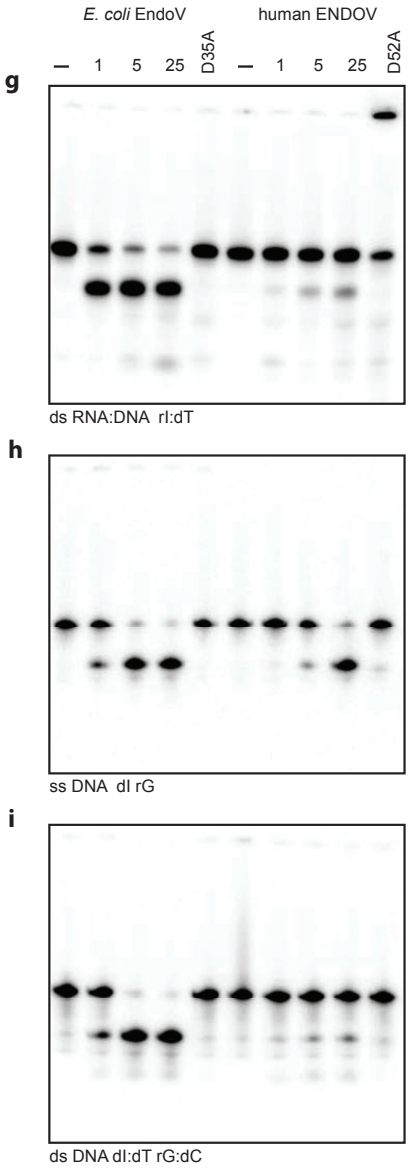
Supplementary Figure-S2 (Alseth)

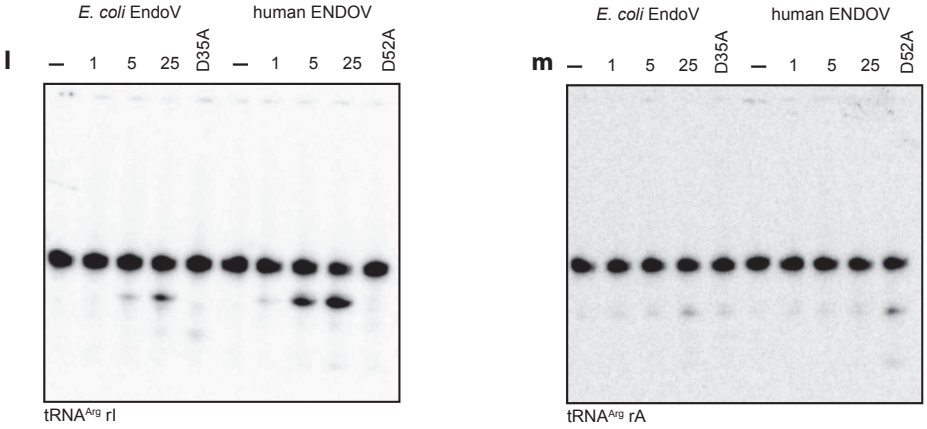


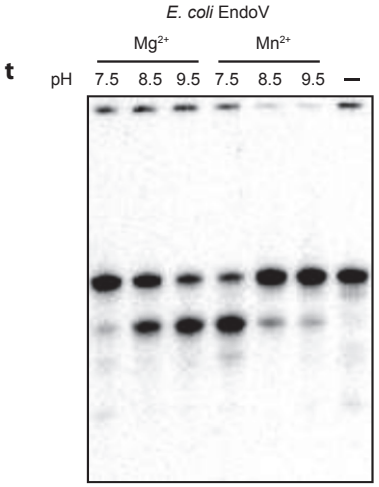
Characterization of incision activities for *E. coli* and human ENDOV on various inosine containing RNA substrates. (**a**) A ss RNA substrate with rI in position 13 (as indicated) was incubated with the endonuclease V enzymes and cleavage products were separated on a 20% sequencing gel next to RNA oligonucleotides of defined lengths (13-16 nucleotides; Supplementary Table S1). (**b-e**) The substrates (**b**) ds RNA rI:rG, (**c**) ds RNA rI:rA and (**d**) ds RNA rI:rC were incubated with increasing amounts of wild type enzymes as indicated (in nM) or the two mutants *E. coli* D35A and human D52A (25 nM). (**e**) The three human ENDOV mutants Y91A, RK and Wedge (Wg) (1-25 nM) were incubated with the ss RNA rI substrate. Reaction products were separated by 15% denaturating PAGE. (**f**) Total tRNA (4.5 nM) isolated from U373 cells was treated with human wild type (0.3 and 0.6 μM) or mutant D52A (0.6 μM) ENDOV enzymes and separated by 15% denaturating PAGE. The gel was stained with ethidium bromide. Size markers (nt) are indicated, – = no enzyme added, r = ribonucleotide.

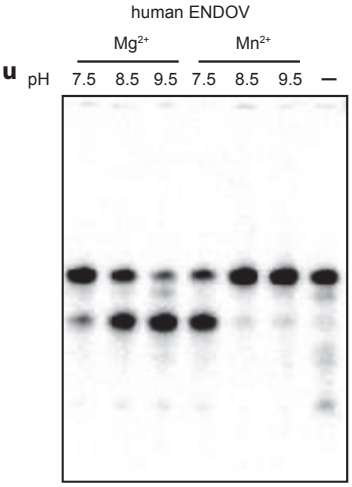
Supplementary Figure-S3 (Alseth)

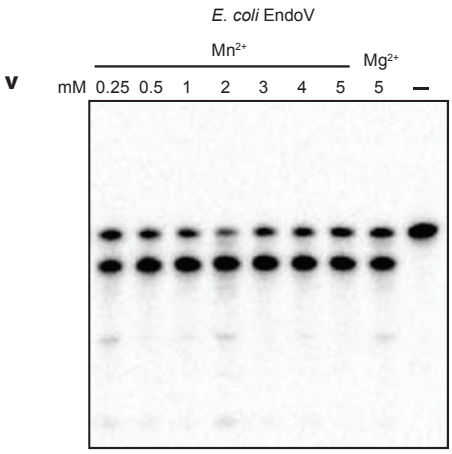


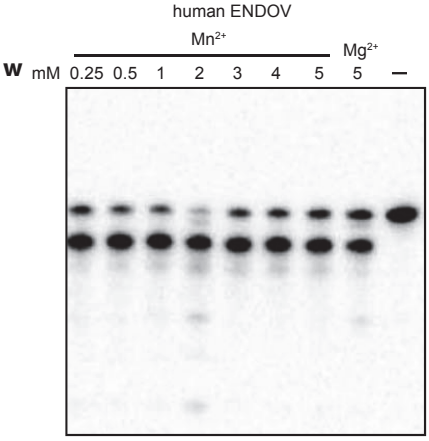

g
E. coli EndoV
human ENDOV
—
1
5
25
D35A
—
1
5
25
D52A
ds RNA:DNA rI:dT
h
ss DNA dI rG
i
ds DNA dI:dT rG:dC
j
E. coli EndoV
human ENDOV
—
1
5
25
D35A
—
1
5
25
D52A
mGabra rI
k
mGabra rA


l
E. coli EndoV
human ENDOV
– 1 5 25 D35A – 1 5 25 D52A
tRNA^Arg rI
m
E. coli EndoV
human ENDOV
– 1 5 25 D35A – 1 5 25 D52A
tRNA^Arg rA
n
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Ser (AGA)
o
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Leu (AAG)
p
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Arg (ACG)
q
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Asp (GTC)
r
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Glu (CTC)
s
human ENDOV D52A
– 0.3 0.6 0.6
tRNA^Lys (CTT)

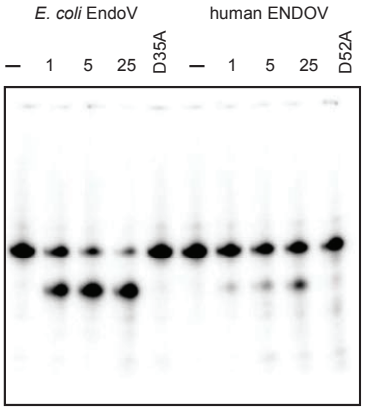

t
E. coli EndoV
Mg2+
Mn2+
pH 7.5 8.5 9.5 7.5 8.5 9.5 –


u
human ENDOV
Mg2+
Mn2+
pH 7.5 8.5 9.5 7.5 8.5 9.5 –


v
E. coli EndoV
Mn2+
Mg2+
mM 0.25 0.5 1 2 3 4 5 5 –

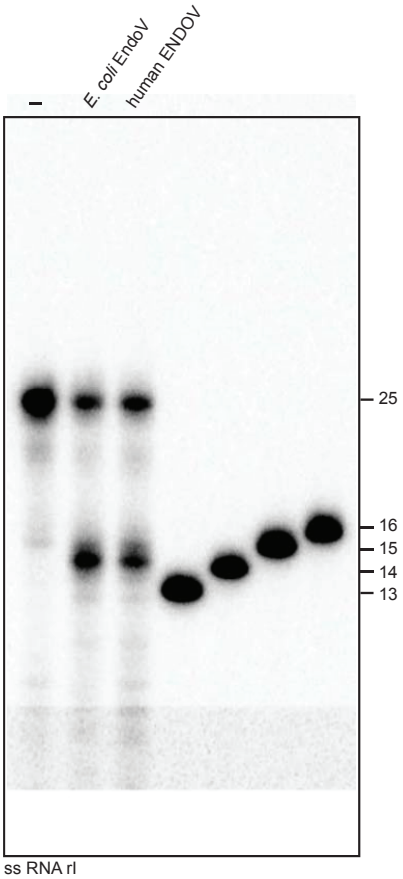

w
human ENDOV
Mn2+
Mg2+
mM 0.25 0.5 1 2 3 4 5 5 –

x

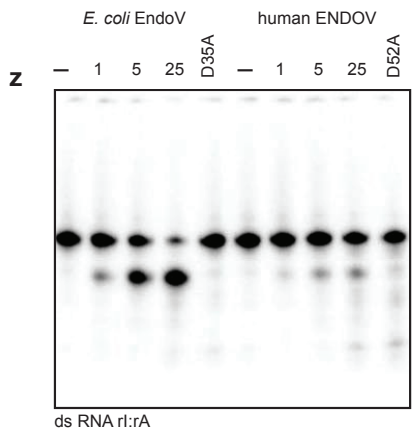


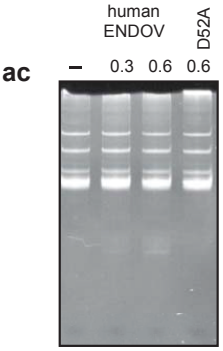
ds RNA rI:rG

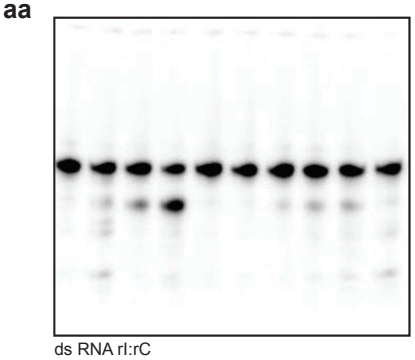
y

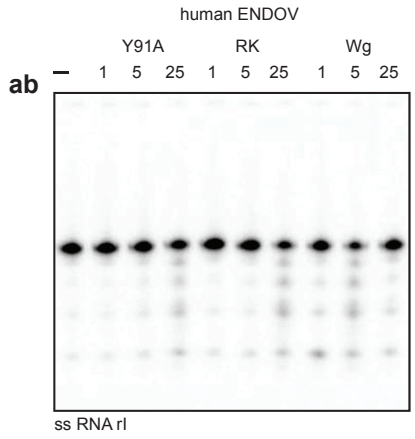


ss RNA rI









Supplementary Figure-S3

Full size images of all gels and blots presented in the manuscript (a-z + aa-ac).

Supplementary Table S1

DNA and RNA oligonucleotides used as substrates in activity assays.

Primer:	5'→3' sequence:	Description:
1	CCGTAGAGCTAC**[dI]**GATCGGTCACCG	DNA - dI*
2	CCGTAGAGCTACAGATCGGTCACCG	DNA - control*
3	CGGTGACCGATCTGTAGCTCTACGG	DNA - Complementary oligo for ds control and I
4	CCGUAGAGCUAC**[rI]**GAUCGGUCACCG	RNA - rI*
5	CCGUAGAGCUACAGAUCGGUCACCG	RNA - control*
6	CGGUGACCGAUC**[U]**GUAGCUCUACGG	RNA - Complementary oligo for ds control and I
7	CGGUGACCGAUC**[C]**GUAGCUCUACGG	RNA - Complementary oligo for ds control and I
8	CGGUGACCGAUC**[A]**GUAGCUCUACGG	RNA - Complementary oligo for ds control and I
9	CGGUGACCGAUC**[G]**GUAGCUCUACGG	RNA - Complementary oligo for ds control and I
10	CCGTAGAGCTAC**[rI]**GATCGGTCACCG	DNA - rI*
11	CCGTAGAGCTAC**[dI][rG]**ATCGGTCACCG	DNA - dI rG*
12	UACGCGACGGCCAUGGACUGGUUCAUAGCCGUCUGUUA	RNA - mGabra-3 control*
13	UACGCGACGGCCAUGGACUGGUUCAU**[rI]**GCCGUCUGUUA	RNA - mGabra-3 rI*
14	ACUCGGCUACGAACCGAG	RNA - $tRNA^{Arg}$ *E. coli* control*‡
15	ACUCGGCU**[rI]**CGAACCGAG	RNA - $tRNA^{Arg}$ *E. coli* rI*‡
16	CCGUAGAGCUAC**[rI]**	RNA - 13 mer marker*
17	CCGUAGAGCUAC**[rI]**G	RNA - 14 mer marker*
18	CCGUAGAGCUAC**[rI]**GA	RNA - 15 mer marker*
19	CCGUAGAGCUAC**[rI]**GAU	RNA - 16 mer marker*

*= 5'P labeled strand
‡=underlined bases correspond to the anticodon triplet

Supplementary Table S2

Kinetic parameters for incision activities by *E. coli* and human endonuclease V.

Enzyme	*E. coli* EndoV	Human ENDOV
Substrate	**k_{obs} (s^{-1})**	**k_{obs} (s^{-1})**
ss RNA rI	0.0024 ± 0.0003	0.0031 ± 0.0004
ds RNA rI:rU	0.0026 ± 0.0004	0.0024 ± 0.0004
ss DNA dI	0.0047 +/- 0.0001	-
ds DNA dI:dT	0.0140 +/- 0.0007	-
ss DNA dI rG	0.058 ± 0.003	0.012 ± 0.001

The catalytic turnover rates (k_{obs}) were measured under single-turnover conditions for the substrates ss and ds RNA and DNA with I, and ss DNA with one ribonucleotide 3' to dI as shown in Figure 1f. A single phase association model was fitted to data from three independent series of experiments for each substrate to calculate values for k_{obs} (s^{-1}).

Supplementary Table S3

DNA oligonucleotides used as probes in northern analysis.

Primer:	5'→3' sequence:	Description:
1	CATCGCCTTAACCACTCGGCCACGACTAC	chr6.trna147-tRNA^{Ser}(AGA)
2	CCTTAGACCGCTCGGCCACGCTACC	chr5.trna3-tRNA^{Leu}(AAG)
3	ACGCGTTATCCATTGCGCCACTGGCCC	chr6.trna8-tRNA^{Arg}(ACG)
4	AGGCGGGGATACTCACCACTATACTAACGAGGA	chr6.trna48-tRNA^{Asp}(GTC)
5	GCCGAATCCTAACCACTAGACCACCAGGGA	chr6.trna87-tRNA^{Glu}(CTC)
6	ATGCTCTACCGACTGAGCTAGCCGGGC	chr15.trna2-tRNA^{Lys}(CTT)

III