Functional characterization of hypothetical proteins and their role in radiation tolerance of *Deinococcus radiodurans*

By

Anubrata Das

LIFE01200704001

Bhabha Atomic Research Centre

A thesis submitted to the

Board of Studies in Life Sciences

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

October 2012

HOMI BHABHA NATIONAL INSTITUTE

RECOMMENDATIONS OF THE VIVA VOCE BOARD

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Shri Anubrata D. Das entitled **'Functional characterization of hypothetical proteins and their role in radiation tolerance of *Deinococcus radiodurans*'** and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ **Date: 26/09/13**

Chairman – Dr. S.F.D'Souza

_____ **Date: 26/09/13**

Guide Dr. H. S. Misra

_____ **Date: 26/09/13**

Member Dr. (Mrs) J. K. Sainis

_____ **Date: 26/09/13**

Member Dr. M. V. Hosur

_____ **Date: 26/09/13**

External Examiner – Prof. D. Narasimha Rao

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI. I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 26th September 2013

Place: Mumbai

(H.S. Misra)
Research Guide

STATEMENT BY THE AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

(Anubrata Das)

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

(Anubrata Das)

DEDICATED TO

MY FAMILY

ACKNOWLEDGEMENTS

I feel highly privileged to express my heartfelt sincere gratitude to my esteemed guide, Prof. H.S. Misra (HBNI), Head, Moleular Genetics Section, Molecular Biology Division for providing me an opportunity to register for Ph. D. It is difficult to find the words to express my gratitude and respect but I want to thank him for constantly keeping me focussed and providing me with keen powers of observation. Indeed, I am honored to be one of his students.

It is my great privilege to acknowledge Dr. S.F.D'souza, Associate Director (A), BMG and Head, NA&BTD for his encouragement, moral support and valuable guidance which inspired me to do my best. I am also grateful to Dr. (Mrs) J.K.Sainis, Raja Ramanna Fellow and Dr. M.V.Hosur, Head, PCS, SSPD, members of my HBNI doctoral committee for providing valuable suggestions for the completion of this work.

It has been my proud privilege and immense pleasure to extend my sincere gratitude for Dr. K. B. Sainis, Director, Bio-Medical Group, BARC for his encouragement and support in carrying out this work. With utter pleasure I acknowledge the role of Dr.S.K.Apte, Dean, Life sciences, HBNI in guiding me to a career in research.

I wish to thank my lab colleagues Shri Y.S.Rajpurohit, Mrs. Swati Kota, Mrs. Nivedita P. Khairnar, Shri C. Vijay Kumar, Mrs. Shruti Desai, Mrs. Vidya Kamble and Ms. Kruti Mehta and Shri Sunil D. Dhotre for their constant support. I am also thankful to all my erstwhile colleagues in Environmental Biotechnology Section, Dr. (Mrs) V.P.Joshi and members of MBD and my project trainees for their support at different times.

I am deeply obliged to Mamoni and Bapi who undertook lot of hardships to provide a comfortable life for me. This Ph.D would have been impossible without the steadfast support and the sacrifice of Mantu and am grateful for being blessed with Teetul who taught me to work with dedication. I also wish to thank Dr.P.Venubabu for standing by me during times of difficulty.

September, 2012

MBD, BARC *Anubrata Das*

Table of Contents

HOMI BHABHA NATIONAL INSTITUTE

PH. D. PROGRAMME

1. Name of the Student:	**Anubrata Das**
2. Name of the Constituent Institution:	**Bhabha Atomic Research Centre**
3. Enrolment No. :	**LIFE01200704001**
4. Title of the Thesis:	Functional characterization of hypothetical proteins and their role in radiation tolerance of *Deinococcus radiodurans*
5. Board of Studies:	**Life Sciences**

SYNOPSIS

Functional characterization of hypothetical proteins and their role in radiation tolerance of *Deinococcus radiodurans*.

Maintaining genetic integrity is the key to survival. When organisms are exposed to various stresses such as ionizing and non-ionizing radiation, numerous types of DNA damage such as single strand breaks, double strand breaks, base modifications etc occur by the direct deposition of radiation energy on DNA molecules and by the free radicals generated in the aqueous environment. It is believed that organisms, which sustain high doses of these stressors, have evolved processes to either protect the biomolecules from the incident damage and/or have efficient repair mechanisms. *D. radiodurans* is amongst those few organisms known to tolerate up to 15 kGy of γ radiation which produces more than 200 double strand breaks and more than 3000 single strand breaks [1]. It achieves this by a strong oxidative stress tolerance mechanism consisting of high quality antioxidant enzymes and antioxidant metabolites like caretenoids[2,3], pyrroloquinoline quinone [4,5] and a higher ratio of Mn^{2+}/Fe^{2+} , along with an efficient error free DNA strand break repair mechanism [6]. In an effort to understand the extraordinary radiation resistance of *D. radiodurans*, the genome of this bacterium was sequenced [14] and it shows two classes of proteins 1) those that are homologous to known proteins and 2) those that satisfy *in silico* criteria for being ORFs but have no known homologs, or even any physical evidence. These ORFs are termed as 'Orphans' and the encoded proteins are labelled as 'hypothetical proteins'. Over the past decade around 60 proteins from *D.*

radiodurans have been characterized and 25% of these have been hypothetical proteins. Some of these hypothetical proteins like DdrA, PprI have been shown to play an important role in DNA repair [7]. These 'hypothetical proteins', assume importance because majority of the known proteins in *D. radiodurans* are similar to their counterparts in sensitive bacterium like *E. coli* and thus they do not necessarily explain the remarkable capacity of *D. radiodurans* in reparing damaged DNA [8]. Earlier studies have shown that *D. radiodurans* cells exposed to γ radiation produce numerous breaks in DNA, which this organism could reassemble by two distinct mechanisms occurring in 2 phases. Phase I normally occurs between the first 2.5 to 4.5 hrs after irradiation when the cells are grown in complex medium after irradiation. If the cells are grown in nutritionally deficient mineral medium the repair process has been shown to slow down. In this phase, by the Extended Synthesis Dependent Strand Annealing (ESDSA) mechanism, double stranded DNA ends are first extensively recessed to form single stranded DNA with 3'-OH overhangs. These overhangs invade complementary template strands and fresh DNA synthesis takes place [9]. In phase II RecA dependent homologous recombination pathway then follows phase I to finally generate intact chromosomes.

The genesis of this study began with an observation of *D. radiodurans* cells undergoing post irradiation recovery (PIR), after exposure to a 4kGy dose of γ radiation. When the cell lysate of the cells, obtained 1hour after irradiation was fractionated by heparin column followed by ion exchange chromatography, a fraction was obtained, which showed ATP inhibited nucleolytic degradation of double stranded DNA [10]. At a 4kGy dose of γ radiation the DNA inside the cell shatters into >100 DNA pieces and these pieces are accurately reassembled back by a poorly known network of the proteins which execute

ESDSA [11]. Since, the above regulated nucleolytic function was seen in the first hour of repair after irradiation, we got curious to find out the possible nucleases/kinases in this fraction. To identify the possible nucleases/kinases, this fraction was analyzed by Fourier Transform Ion Cyclotron Resonance-Mass Spectroscopy (FTICR-MS), but no known nuclease or kinase was found. However 11 "hypothetical proteins" were also found and an initial domain search showed that 6 (DR0041, DR0390, DR0461, DR1654, DR2417m, DRA0282) of them had possible domains for DNA metabolic functions.

In this thesis we have attempted to predict the functions of four proteins from these six hypothetical proteins for possible nuclease/kinase activity using bioinformatics and functionally characterized two of these proteins DR2417 and DRA0282 both *in vitro* and *in vivo*. Total work described in this thesis has been accomplished through following objectives:

1. Theoretical prediction of functional domains in the above ORFs with the help of bioinformatics tools and online databases.
2. *In vitro* characterization of recombinant DR2417m and DRA0282 for their predicted functions and their effect on the response of transgenic cells to DNA damaging agents.
3. Molecular genetic studies to understand the role of DR2417 and DRA0282 in DNA repair and radiation resistance of *Deinococcus radiodurans.*

The work carried out to address the above objectives, will be presented in this thesis as following chapters:

Chapter 1. General Introduction and review of literature

This chapter describes the general information both on *D. radiodurans*, the various types of damage to DNA and the repair systems available in *D. radiodurans*. It further discusses the role of hypothetical proteins in its DNA repair and various bioinformatics approaches available to predict their structure and function.

Chapter 2. Materials and Experimental Procedures

This chapter describes the details of materials used with their sources and common experimental methods used in this study. Different techniques and approaches used in this study will be described under the category of (1) Bioinformatics –use of available tools BLAST, PSI-BLAST, ClustalX, Pfam database, Treeview, Swissmodel server, python etc. (2) biochemical techniques including protein purification, Electrophoretic Mobility Shift Assay, Sequencing gel, RNase assays (3) recombinant DNA technology and molecular biology techniques including cloning, generation of deletion mutants and their confirmation.

Chapter 3. Results

This chapter starts with a preamble describing the hypotheses and logical anticipation from this study. The results obtained to test the hypotheses along with other new findings during the course of this investigation will be summarized. Total results obtained from this study have been presented in three distinct sections. Each starts with a brief introduction, experimental procedures and results obtained from this specific study, followed by a brief discussion on the findings specific to results included in this section.

3.1 Bioinformatics analysis of selected hypothetical proteins identified from the pool of DNA binding proteins exhibiting ATP sensitive nucleolytic function

We have used online tools to find possible nuclease/kinase amongst the hypothetical proteins. The features studied were (1) protein sequence analysis (2) genomic context (3) subcellular location (4) peptide abundance in FTICR-MS data (5) theoretically modeled structure (6) phylogenetic relationships. DRA0282 (84 peptides), DR0390 (45 peptides), DR2417m (36 peptides), DR1654 (39 peptides) were abundant in the fraction amongst all ORFs encoding hypothetical protein detected, while DR0041 (5 peptides) and DR0461 (4 peptides) were below the significance threshold and hence we focussed on these four abundant proteins. Initial BLAST analysis showed DR2417m has a match with a β-CASP family nuclease- Artemis, DRA0282 with a DNA binding repair protein –Ku 70/80, DR0390 with DAK family kinases and DR1654 with DEAD box helicases. A critical histidine residue of the DEA/X-DH motif was missing in DR1654 and hence, a further study on this was postponed. Ku70/80 and Artemis proteins form the critical components of Non Homologous End Joining (NHEJ) repair Pathway and also initial matches showed that DR2417m could be a nuclease, and hence we selected DR2417m and DRA0282 for further studies. The ORF encoding DR2417 was reported to have a genuine frameshift mutation and hence the protein was annotated as DR2417m.

β-CASP family nucleases are involved in diverse functions like 3’ end processing of RNA precursors, NHEJ repair, V(D)J recombination etc. These have several key conserved motifs, motifs 1,2,3,4 of the catalytic metallo-β lactamse domain and motifs A, B, C of the substrate specificity determining CASP domain. DR2417m and its β-CASP family

homologs obtained from a PSI-BLAST search were aligned in ClustalX, to look for conserved motifs. Sequence alignments showed that DR2417m had all these conserved motifs and an unusually arginine and proline rich C-terminus as per annotated sequence. The distance tree of DR2417m with its homologs showed that DR2417m is closer to Dgeo2150 from *D.geothermalis* but not its paralog DRA0069.

To find the function of DRA0282, we did a BLAST search with RefSeq database to find putative homologs, however none were found. We then did a PSI-BLAST with DRA0282, with the keywords like DNA, repair etc, as DRA0282 was found in a PIR fraction. Mostly Ku homologues were obtained as the nearest matches and we aligned DRA0282 with these sequences. The multiple sequence alignment showed that DRA0282 had an overall homology with these proteins. Since DNA binding was a structural property, we built a 3-D model of DRA0282 with the structure of Ku80 from Protein Data Bank (PDB code-1JEY). A model with RMSD of 4.0 was built. We observed that DRA0282 had a remarkable homology with Ku80. The N terminal von Willebrand A domain, which binds nucleotides and the DNA binding β-barrel domain were seen. Interestingly, DRA0282 showed a closer match with eukaryotic Ku, than with prokaryotic Ku proteins

3.2 Functional characterization of DRA0282 for its role in radiation resistance in bacteria.

Computational analyses suggested DRA0282 of *D. radiodurans* was homologous to human Ku80. It was thus checked for its similarity to human Ku80 in in vitro activity and

its in vivo role in radiation resistance in both *E. coli* and *D. radiodurans*. The *drA0282* ORF was cloned and expressed in *E. coli*. Recombinant DRA0282 was purified to homogeneity and its interaction with DNA and protection of double stranded DNA from exonuclease III degradation were checked. It preferred Mn^{2+} for its DNA binding activity, and its ability to bind DNA was inhibited by ATP. Electrophoretic Mobility Shift Assay experiments showed that it has the highest affinity (Kd=2.93±0.5 nM) with covalently closed circular double stranded DNA followed by linear single stranded DNA (Kd = 60.22±9.1 nM) and linear double stranded DNA with (Kd=196.9±30.8 nM). These results suggested that DRA0282 is indeed a DNA binding protein like human Ku80 while it was unique in its substrate preference. [12]

Transgenic *E. coli* cells expressing DRA0282 showed enhanced resistance to UVC and γ radiation in wild type background. The *recA*$^{-}$ *E. coli* JC1553 cells expressing DRA0282 did not show UVC and γ radiation protection. The RecA dependent protection of *E. coli* cells expressing DRA0282, was found to be independent of RecA's role through SOS response as the *E. coli* DM49 (lexA) cells having a full-fledged homologous recombination pathway but deficient in SOS response, continued showing UVC protection in presence of DRA0282. These results suggested the possible role of DRA0282 in processes requiring functional RecA at least in *E. coli.*

The role of DRA0282 was further confirmed in *D. radiodurans*. The *drA0282* deletion mutant of *D. radiodurans* was generated and γ radiation response of the mutant was compared with wild type cells. The mutant showed no visible effect on γ radiation treatment. This could be possible if DRA0282 has other functional homologues in *D. radiodurans*. Earlier two proteins, DdrA and PprA were shown for having in vitro

functional similarities with DRA0282. Therefore, *drA0282* gene was deleted from the pprA deletion mutant of *D. radiodurans*. The *drA0282pprA* double mutant was evaluated for γ radiation resistance and compared with single pprA and *drA0282* mutant and wild type. The double mutant did not add any effect to *pprA* single mutant suggesting that *pprA* and *drA0282* possible functions through common mechanisms in *D. radiodurans*.

These results suggested that DRA0282 is an important DNA repair protein similar to human Ku80, which exhibits a role in heterlogous system that does not have its functional homologues. On the other hand, *D. radiodurans* seems to have other proteins which provide functional redundancy to *drA0282* deletion or simply our screen was inadequate.

3.3. Characterization of DR2417 for its predicted functions and its role in bacterial resistance to γ radiation

Earlier DR2417m (m stands for mutation) was predicted to be a member of the β-CASP family of nucleases based on its N-terminal homologies and presence of defined motifs in these proteins except for motif C [13]. The C-terminal of the annotated polypeptide is unusually rich for arginine and proline amino acids, which is normally not observed in genuine proteins and for these reasons it was earmarked as a hypothetical protein. Since, the full length gene was cloned and a recombinant protein of the desired size was made as well and since this protein was detected in a pool of DNA binding proteins, the possibility of a wrong annotation of a frame shift in this ORF was looked into. Complete nucleotide sequencing was carried out with the PCR amplified products from genomic DNA

templates. It was observed that the reported frame shift was wrongly annotated. This information has been reported to the NCBI database and revised sequence of *dr2417* has been submitted vide accession number JQ432552 and subsequently we refer to the protein as DR2417. In vitro functional characterization of this protein was also carried out. Recombinant DR2417 was purified from *E. coli* BL21 and checked for its preferences to different types of DNA and RNA substrates and other requirements. Purified DR2417 was found to be a Mn^{2+} dependent nuclease. It acted both as a potent DNase and a weak RNase. It was inhibited by ATP and GTP but not by AMP or GMP. At a gross level, DR2417 degrades both ØX174 virion DNA (single stranded DNA) and ØX174 RFII DNA (covalently closed double stranded DNA). It degrades PCR DNA (linear double stranded DNA), when the ends of the PCR fragment have been phosphorylated by treatment with polynucleotide kinase. To understand the mechanism, we tested it with defined oligonucleotides. We found that it acts on double stranded DNA from 3' end. It does not degrade single stranded DNA. It has a strict requirement for a 5' phosphate, without which it does not act upon its substrate. Although DR2417, has a 40% overall identity with RNaseJ from *T.thermophilus* and 100% conservation at the active site residues, it nevertheless has a poor RNase activity. These characteristics make this enzyme distinct from other members of β-CASP family proteins and offers the novelty in this enzyme.

The *in vivo* effect of DR2417 was studied both in a heterologous host *E. coli* and in *D. radiodurans*. DR2417 was expressed both in *recA*$^-$ and wild type background of *E. coli*. No significant change in cell survival upon UVC and γ radiation treatment was observed in case of *E. coli*. Attempts to delete the DR2417 gene from the *D. radiodurans* genome were always unsuccessful. We monitored the deletion process by checking for

amplification with primers flanking 500 bp upstream and downstream to the *dr2417* ORF. The homozygous knockout of *dr2417* was never seen, however heterozygous clones amplifying both 1.7 kb *dr2417* and 0.8kb *nptII* marker were obtained. These heterozygous clones generally grew poorly as compared to the wild type, and also were more sensitive to U.V and γ radiation. Interestingly, we observed that this gene shows dose effect on growth of this bacterium. The transformants were grown for several generations and at each stage, the reduction in copy number of *dr2417* was monitored. We observed a direct correlation between copy number and growth inhibition. We did real time PCR to quantify the genomic copies of *dr2417* present in these clones and found that the population with the least number of copies of *dr2417* was also the most sensitive. This suggested that deletion of *dr2417* from *D. radiodurans* was deleterious for the survival of this bacterium and thus *dr2417* seems to be an essential gene.

Chapter 4. Discussion

In this chapter the findings on functional characterization of these two proteins in the context of various hypotheses proposed for explaining the efficient DNA strand break repair in this bacterium has been discussed. *D. radiodurans* has been extensively studied for its ability to tolerate UVC, ionizing radiation, desiccation etc, which generate a huge number of DSBs, such that it would debilitate many other organisms. Interestingly, this bacterium is not tolerant to near UV (beyond 254nm). Early reports indicated novel DNA repair mechanisms to be the key to survival, but sequencing of genomes showed that its repair machinery is unremarkable, save a few oddities [14]. Newer studies now hint to an

efficient system which protects proteins from oxidative damage and may be the key to its resistance [15]. But till date not much is known how after suffering a 10 kGy dose of γ radiation, growth remains halted, damaged DNA and proteins are recycled, while 200 random fragments of DNA are accurately paired back, when evidence exists for increased transposon activity, and all this with the same set of repair genes, found in other radiosensitive bacteria ![16] Thus the role of 'hypothetical proteins/predicted proteins' or proteins with 'putative function', becomes relevant, for e.g. at 3 hours PIR 30% of over-expressed proteins are 'hypothetical proteins' [17]. The broad outlines of the repair process involves regulated nucleolytic processing of double stranded DNA during early period of post irradiation recovery, followed by the synthesis of >10 kb long single stranded DNA and then an accurate patchwork assembly of old and new DNA [9]. Molecular mechanisms of nucleolytic processing from indiscriminate to productive DNA processing and their regulation is not clearly understood. Initial work from our laboratory has however, indicated the involvement of high energy nucleotide triphosphates in these processes [10]. It was during these studies with various PIR fractions that ATP inhibited nuclease activity was seen and DR2417 and DRA0282 were selected as interesting candidates.

Initial bioinformatics showed that DR2417 has homology to Artemis, a β-CASP family nuclease, and DRA0282 has structural similarity with Ku80, a double strand break sensing protein. These proteins were then expressed in *E. coli* and the purified proteins were characterized for their predicted functions. Both are found to be DNA interacting proteins and DRA0282 showed protection of linear double stranded DNA from exonuclease

degradation but DRA0282 had higher affinity to linear single stranded DNA and supercoiled double stranded DNA than linear double stranded DNA. The later observation was different from human Ku80 type proteins. On the other hand, *dr2417m* was annotated as a genuine frameshift in *D. radiodurans* genome. This protein was detected in various studies as a functional component induced upon irradiation [10,18]. Also C-terminal was unnaturally rich with proline and arginine, a normal anomaly of translating non-coding genes. Complete nucleotide sequencing confirmed the wrong annotation of this ORF in genome sequence. The polypeptide derived from new gene sequence was confirmed by MALDI-MS analysis. The peptide mass fingerprints with both N-terminal and C-terminal peptides were completely in agreement with the new sequence and differed grossly at C-terminal of earlier annotation. Recombinant DR2417 was characterized as a novel β-CASP family nuclease. Like other members of the β-CASP family, DR2417 needed a 5'phosphate end and was inhibited by ATP, but unique to it was the exonucleolytic degradation from 3' end, and specificity for double stranded DNA or junction substrates and a weak RNase activity. The bioinformatics function prediction based on corrected sequence indicated DR2417 to be an RNase, as the translated protein would have histidine at motif C. All earlier studies had argued that this would make the protein RNA specific, as histidine is seen in RNases and valine is seen in DNAses.[13] Based on the structural differences between our 3-D model of DR2417 and the structure of RNase J, a well characterized member of β-CASP family, we suspect that absence of the β sheets, β6 and β7 in DR2417 which makes the channel for binding nucleic acids wide, thus fitting well to DNA, but loosely fitting to RNA [19]. Also, the structure of RNase J shows that binding to nucleic acids is by sugar phosphate backbone only. Thus, we assume that overall

structure and not a single residue determines substrate specificity. Recombinant DRA0282, was characterized as a DNA binding protein which protected DNA ends from ExoIII nuclease activity like Ku, but further studies showed it actually preferred single stranded and covalently closed circular DNA over linear double stranded DNA unlike Ku.

The role of both these proteins in some common processes of DSB repair may be speculated. It gets strong supports from the facts that both are abundantly present in same pool of DNA binding proteins isolated from 1hour PIR sample and transcription of both genes are concurrently induced at 1hour of post γ irradiation. Interestingly, DNA interaction of both these proteins was inhibited with high-energy phosphates like ATP and GTP. However, the exact mechanism(s) of ATP inhibition is not clear yet. Since DRA0282 and DR2417 are structurally very close to human Ku80 and Artemis respectively, the possibility of these proteins associated with Non Homologous End Joining (NHEJ) repair of double strand breaks in this bacterium cannot be ruled out. Since patch work assembly is seen during the repair process, the possibility of joining of 2 double stranded DNA molecule ends by a mechanism similar to NHEJ is an open question and needs to be cleared up. NHEJ has been demonstrated in *Bacillus subtilis* and *Mycobacterium tuberculosis* in prokaryotes [20]. Molecular genetic studies with these proteins showed interesting phenotypes. DR2417 did not grossly affect the survival of *E. coli* cells expressing this protein while *E. coli* expressing DRA0282 showed improved resistance to UVC and γ radiation. *D. radiodurans* cells deleted for *drA0282* showed no discernable effect and found to contribute in pathways also supported by other proteins with similar functions. On the other hand, an attempt to create a complete deletion mutant

of *dr2417* failed and the partial knockouts, which were obtained, were debilitated. Thus DR2417 seems to have an essential role in growth and survival of *D. radiodurans*, while DRA0282 is needed possibly in an accessory function. Nevertheless, our findings hint to the fact that in *D. radiodurans* DNA metabolic proteins may be the same as that of radiosensitive organisms but they act in different combinations in unique pathways.

Chapter 5. Summary, Conclusion and Future Perspective

This chapter will summarize briefly the work presented in this thesis by highlighting the potential observations that will include the answers to the hypotheses on the basis the work that has been undertaken and unanswered questions / new hypotheses generated during course of this study. The conclusion drawn from this study will be deciphered. Along with several interesting findings reported in this thesis, two hypothetical proteins in *D. radiodurans* genome have been assigned novel functions and one misled annotation has been corrected during the course of this study. We appreciated from this study also that structure –function prediction using bioinformatics tools guide well but require validation by wet biology. While this study has answered some of the questions and hypotheses raised both from experimental data of earlier work and bioinformatics analyses of certain proteins, it has left several unanswered questions. The immediate ones which require attention include (i) how these proteins work inside *D. radiodurans* cells, (ii) who are their interacting partners under *in vivo* conditions, (iii) how does ATP/GTP influence the interaction of these proteins with DNA and what is the functional significance of strong

binding of these proteins to supercoiled DNA. Also highlighted are the needs for more accurate algorithms for prediction of function of the ever increasing number of hypothetical proteins.

List of Publications

[A] Papers Published in Peer-reviewed International Journal

1. Das, AD and Misra, HS (2011). Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage.

Microbiology (SGM) 157: 2196-2205

2. Das, AD and Misra, HS (2012). DR2417m, a hypothetical protein characterized as a novel β- CASP family nuclease in radiation resistant bacterium, *Deinococcus radiodurans.*

Biochim. Biophys. Acta (General Subjects) 1820:1052-1061

3. Das, AD and Misra, HS (2013). Hypothetical proteins present during recovery phase of radiation resistant bacterium *Deinococcus radiodurans* are under purifying selection.

J. Mol. Evol 77:31-42

[B] Abstracts Presented in National conferences

1. Das A. D. and Misra H. S. "Functional characterization of a hypothetical protein from *Deinococcus radiodurans* and its role in bacterial resistance to DNA damage" *in* 34th All India Cell Biology Conference and symposium on quantitative biology : from molecules to cells, Abstract No. 64, p.142, December 4-6, 2010, Kolkata.

2. Das A. D. and Misra H. S. "A novel member of the beta CASP family of nuclease in involved in the survival of *Deinococcus radiodurans*" in 52nd Annual Conference of Association of Microbiologists of India (AMI-2011), Abstract No. MM111, p392, November 3-6, 2011, Punjab University, Chandigarh.

[C] Genbank Submission

A. D. Das and H. S. Misra (2012) *Deinococcus radiodurans* strain ATCC 13939 DR2417 gene. Accession No. JQ432552.

[D] Award and other recognition

A. D. Das received Dr. Rana Memorial Best Poster Award during 52nd Annual Conference of Association of Microbiologists of India (AMI-2011), held at Punjab University, Chandigarh, November 3-6, 2011.

References

1. Battista JR (2000) Radiation resistance: the fragments that remain. Curr Biol 10: R204-205.

2. Tian B, Sun Z, Xu Z, Shen S, Wang H, et al. (2008) Carotenoid 3',4'-desaturase is involved in carotenoid biosynthesis in the radioresistant bacterium *Deinococcus radiodurans*. Microbiology 154: 3697-3706.

3. Markillie LM, Varnum SM, Hradecky P, Wong KK (1999) Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants. J Bacteriol 181: 666-669.

4. Misra HS, Khairnar NP, Barik A, Indira Priyadarsini K, Mohan H, et al. (2004) Pyrroloquinoline-quinone: a reactive oxygen species scavenger in bacteria. FEBS Lett 578: 26-30.

5. Rajpurohit YS, Gopalakrishnan R, Misra HS (2008) Involvement of a protein kinase activity inducer in DNA double strand break repair and radioresistance of *Deinococcus radiodurans*. J Bacteriol 190: 3948-3954.

6. Slade D, Lindner AB, Paul G, Radman M (2009) Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. Cell 136: 1044-1055.

7. Harris DR, Tanaka M, Saveliev SV, Jolivet E, Earl AM, et al. (2004) Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1. PLoS Biol 2: e304.

8. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, et al. (2001) Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics. Microbiol Mol Biol Rev 65: 44-79.

9. Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, et al. (2006) Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature 443: 569-573.

10. Kamble VA, Rajpurohit YS, Srivastava AK, Misra HS (2010) Increased synthesis of signaling molecules coincides with reversible inhibition of nucleolytic activity during postirradiation recovery of *Deinococcus radiodurans*. FEMS Microbiol Lett 303: 18-25.

11. Misra HS, Khairnar NP, Kota S, Shrivastava S, Joshi VP, et al. (2006) An exonuclease I-sensitive DNA repair pathway in Deinococcus radiodurans: a major determinant of radiation resistance. Mol Microbiol 59: 1308-1316.

12. Ristic D, Modesti M, Kanaar R, Wyman C (2003) Rad52 and Ku bind to different DNA structures produced early in double-strand break repair. Nucleic Acids Res 31: 5229-5237.

13. Callebaut I, Moshous D, Mornon JP, de Villartay JP (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res 30: 3592-3601.

14. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, et al. (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286: 1571-1577.

15. Slade D, Radman M (2011) Oxidative stress resistance in *Deinococcus radiodurans*. Microbiol Mol Biol Rev 75: 133-191.

16. Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, et al. (2010) Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. PLoS Genet 6: e1000799.

17. Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, et al. (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Proc Natl Acad Sci U S A 100: 4191-4196.

18. Kota S, Misra HS (2008) Identification of a DNA processing complex from *Deinococcus radiodurans*. Biochem Cell Biol 86: 448-458.

19. Li de la Sierra-Gallay I, Zig L, Jamalli A, Putzer H (2008) Structural insights into the dual activity of RNase J. Nat Struct Mol Biol 15: 206-212.

20. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, et al. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297: 1686-1689.

Publications:

a. <u>Published</u>

1. Das, AD and Misra, HS (2011). Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage. Microbiology (SGM) 157: 2196-2205

2. Das, AD and Misra, HS (2012). DR2417m, a hypothetical protein characterized as a novel β- CASP family nuclease in radiation resistant bacterium, *Deinococcus radiodurans.* Biochim Biophys Acta (General Subjects) 1820:1052-1061

3. Das, AD and Misra, HS (2013). Hypothetical proteins present during recovery phase of radiation resistant bacterium *Deinococcus radiodurans* are under purifying selection.

J. Mol. Evol 77:31-42

Signature of Student:

Date:

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. S.F D'Souza	Chairman		
2.	Dr. H.S.Misra	Guide		
3.	Dr.M.V.Hosur	Member		
4.	Dr.(Mrs) J.K.sainis	Member		

List of figures

List of Tables

List of Abbreviations

>	greater than
1 kGy	absorption 1 joule of ionizing radiation by 1 kg of matter
AMP	Adenosine monophosphate
AP	Apurinic/apyrimidinic
ATM	Ataxia-telangiectasia
ATP	Adenosine triphosphate
ATR	ATM and Rad3 related
BIR	Break induced repair
bp	base pair
BrdU	Bromodeoxyuridine
CCC	Covalently closed circular DNA
DDBJ	DNA databank of Japan
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid

dNTP	Deoxyribonucleotides
DSB	Double strand break
DSBR	Double strand break repair
dsDNA	Double stranded DNA
EDTA	Ethylenediamine tetraacetic acid
EEO	electroendoosmosis
EMBL	European nucleotide sequence database
ESDSA	Extended synthesis dependent stand annealing
FTICR-MS	Fourier transform ion cyclotron resonance-mass spectroscopy
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
Kd	dissociation constant –a equilibrium constant
MALDI-MS	Matrix assisted lase desorption/ionization- Mass spectroscopy
MB	Mega basepair
MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information

NER	Nucleotide exicision repair
NHEJ	Non homologous end joining
nm	nanometre
ORF	Open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank
PIR	Post irradiation recovery
PMSF	phenylmethanesulfonylfluoride
PQQ	Pyrroloquinoline-quinone
PSI-BLAST	Position specific iterated-Basic local alignment search tool
RMSD	Root mean square deviation
ROS	Reactive oxygen species
RPM	Rounds per minute
SDS	sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing

SOS	save our soul- in this thesis it refers to SOS repair pathway
SSB	Single strand break
ssDNA	Single stranded DNA
TCA	Tricholoroacetic acid
TYG	Typtone-yeast extract-glucose medium for culture of *D. radiodurans*
UV	Ultraviolet
UVC	short wavelenght UV radiation between 250-270 nm

Chapter 1

General Introduction and Review of Literature

Deinococcus radiodurans has been the most widely studied organism for understanding the molecular mechanism of resistance to ionizing radiation and repair of DNA damaged by ionizing radiation. Although in the last five decades lot information has been gained, yet the exact details of this accurate repair process are sketchy. This work started from an observation of a novel nuclease activity in *D. radiodurans* during the early hours of the repair process. Herein we provide the basis of undertaking this study on an aspect of the important initial phase of DNA repair in *D. radiodurans.*

1.1 Radiation Resistance and *Deinococcus radiodurans*

D. radiodurans is a species among the estimated 11 million species which inhabit the earth and its oceans [1]. It is a gram positive, red pigmented, nonspore forming bacterium which belongs to a genus whose members have been isolated from diverse environments like canned meat [2], desert soils[3], radiation contaminated soils and air 10 kms above

the earth's surface[4]. It occurs in diads and tetrads with an average cell diameter of 1 µM. It is a mesophile and is dessication tolerant. Its 3.28 MB genome is GC rich (66% GC content) and made up of 2 chromosomes and 2 plasmids. It is an organotrophic organism and amino acids are the preffered carbon energy source. What sets it apart from other organisms is its unusual resistance to ionizing radiation [5]. Ionizing radiation like γ rays (1-10 MeV) and X-rays (100-10 KeV) have very high energies which can knock off electrons from nearby atoms. When electrons get knocked off, bonds are broken and atoms with unpaired electrons called *radicals* are generated, in a process called Ionization. These radicals are highly reactive and spontaneously interact with double bonds and break them. In biological systems the predominant moiety is water and these radiation ionize the water molecules to generate very reactive radicals like $OH^{\cdot}$, O_2^{-} and H_2O_2 collectively called as Reactive Oxygen Species (ROS). These ROS cause protein carbonylation, lipid peroxidation and several damages to DNA including the formation of abasic sites, single strand breaks (SSBs), base modifications and double strand breaks (DSBs) [6]. Due to the systemic nature of the damage affecting several subcellular components, most species are susceptible to ionizing radiation but several reports have shown that a key factor to the survival from ionizing radiation is the ability to repair DSBs in the damaged DNA [7]. In this milieu it is interesting to note that *D. radiodurans* easily survives a dose of 15 kGy of γ radiation whereas most humans are killed by a dose 1000 times less. It shares this property with but a handful of organisms including prokaryotes like other members of the Deinococcace family, a cyanobacterium *Chroococcidiopsis*, *Rubrobacter* spp and several hyperthermophiles like *Pyrococcus furiosus* and a few eukaryotes like the slime mold *Dictyostelium discoideum*, various

fungi like *Filobasidium* and the bdelloid rotifer *Adineta vaga* [5]. Ever since its discovery in 1956, lot of efforts have been made to understand the reason for *D. radiodurans*'s radioresistance as it is would help to understand the cellular ROS protection systems, the DNA repair processes as well as find potential applications to protect against the damaging effects of radiation. The origin of this phenomenal resistance to ionizing radiation is mysterious. As the phylogenetic tree below shows (Fig 1.1.1.1) *Deinococcus-Thermus* belong to the same lineage, but while *Deinococci* are resistant to ionizing radiation and desiccation, members of *Thermus* species are radiation and desiccation sensitive but are thermotolerant.

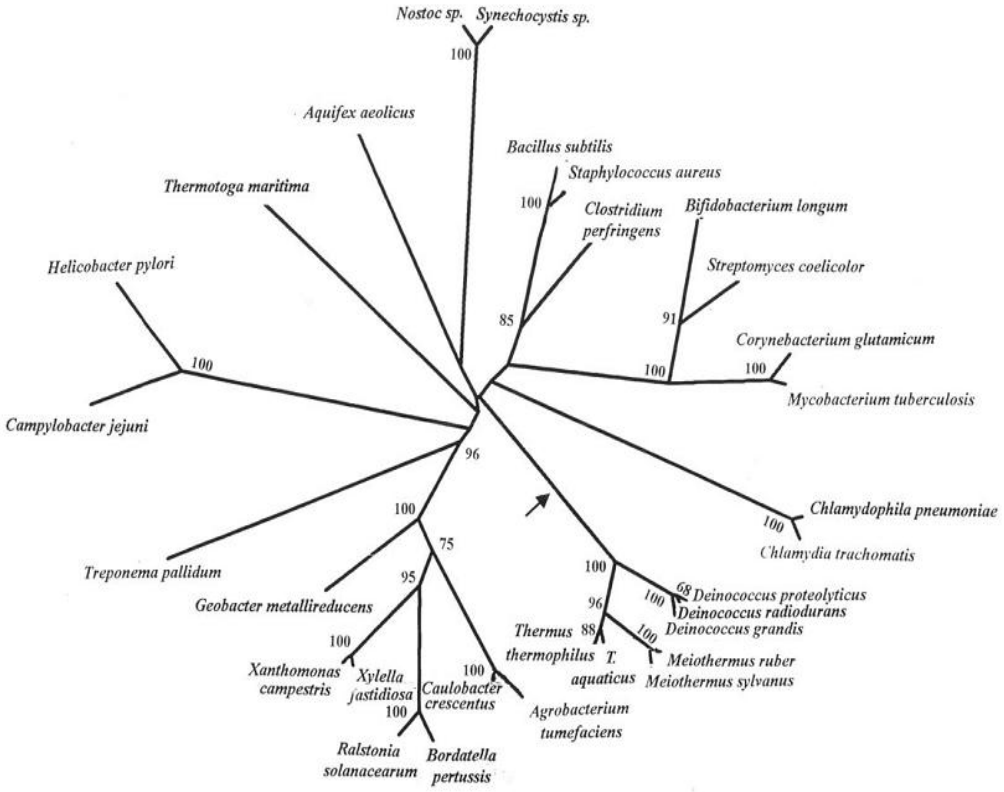


Fig.1.1.1.1. An unrooted phylogenetic tree of *Deinococcus* spp. based on sigma 70 subunit [8]. Note that *Deinococcus* and *Thermus* are closely related and they both differ from organisms of other phyla.

There are evidences of extensive 'Horizontal Gene Transfer' in this lineage which has given its members their unique properties [9]. It is assumed that genes responsible for radiation resistance have been imported into *D. radiodurans* during evolution but for the purpose of desiccation tolerance. This is so because on Earth there are no known sources of radiation which have or had levels of radiation upto which *D. radiodurans* easily tolerates but there are many arid environments and mostly deinococci have been isolated from these environments. Desiccation causes extensive double strand breaks in DNA like radiation does and deinococci are desiccation tolerant too.

Till date no single factor has be pinpointed in deinococci or other radioresistant organisms as the sole source of resistance. Indeed the present view is that several factors coalesce to provide for this phenomenon [10]. At first, *D. radiodurans* has metabolic mechanisms to quench ROS generation. It has fewer Fe-S cluster respiratory enzymes to prevent 'iron toxicity'. It induces the glyoxylate cycle in response to radiation, which allows generation of reductants like NADH which can quench the free radicals formed. It has 3 different alleles of superoxide dismutase, 3 alleles of catalase and 2 alleles of peroxidase to scavenge free radicals and also the basal level of catalase activity is 15 fold more than that of *E. coli* [5]. In rat mitochondria even 0.1 nM concentration of superoxide radical (O_2^-) inactivates Aconitase, a key Tricarboxylic Acid Cycle enzyme [11]. Superoxide dismutase converts the superoxide radical to oxygen and hydrogen

peroxide and catalase further decomposes the toxic hydrogen peroxide to oxygen and water. Peroxidases also assist in scavenging ROS by converting hydrogen peroxide to water. *D. radiodurans* also has pyrroloquinoline-quinone (PQQ) coenzyme which efficiently scavenges ROS [12]. PQQ, also known as methoxatin, was identified as a cofactor of bacterial dehydrogenases and has been shown to be a more effcient scavenger of ROS, as compared to other antioxidants. PQQ obtained from *Deinococcus* has been shown to protect *E. coli* from oxidative damage [5,13]. Deinoxanthin, a major carotenoid found in *D. radiodurans* has good ROS scavenging property.A study of *D.radiophilus* lipid extracts showed that these lipids have antioxidant properties [14]. It is reasonable to expect that *D. radiodurans* also has similar lipid protection mechanism which prevents its membranes from getting damaged by radiation. *D. radiodurans* also has several Mn^{2+} complexes which are known to scavenge ROS [15]. Next, it has specific pathways to prevent protein damage by oxidation. Studies have shown that a small molecule mediated mechanism in *D. radiodurans* prevents protein carbonylation which can be caused by ionizing radiation [16,17].

Also a study showed that proteins get selectively recycled after irradiation, which means that damaged proteins are degraded by proteases and replaced with fresh proteins [17]. This protease activity, besides helping in protein recycling may also be involved in cell signaling. It has earlier been shown that proteases can act as irreversible signaling and regulating systems. For example, caspase 3 cleaves the ICAD protein, which releases the Caspase Activated DNase, an endonuclease which works in the apoptosis pathway[18]. Next, *D. radiodurans* has an extensive cell cleaning mechanism comprised of Nudix phosphohydrolases and nucleotidases which clean up oxidized derivatives which are

formed due to the free radicals generated by radiation. In fact the Nudix family is highly expanded in *D. radiodurans* with 23 members [19].

However, *curiously*, all this elaborate mechanism protects the membrane lipids and proteins but cannot protect the DNA inside the cell, which bears the brunt of ionizing radiation. One reason might be that DNA is the longest polymer in the cell and hence an elaborate target for radiation. A dose of 7 kGy shatters the genome into >200 pieces, with at least 10 times more single strand breaks (SSBs) as well as modifying bases on the DNA molecule. Such massive damage halts the replisome and eventually the cell dies. To put it into perspective, *E. coli* barely withstands 3-4 double strand breaks (DSBs) per genome and thus critical to the survival of *D. radiodurans* is its ability to accurately splice back this damaged DNA.

1.2 Repair Systems in *Deinococcus radiodurans*

1.2.1 A general Study of Repair Systems

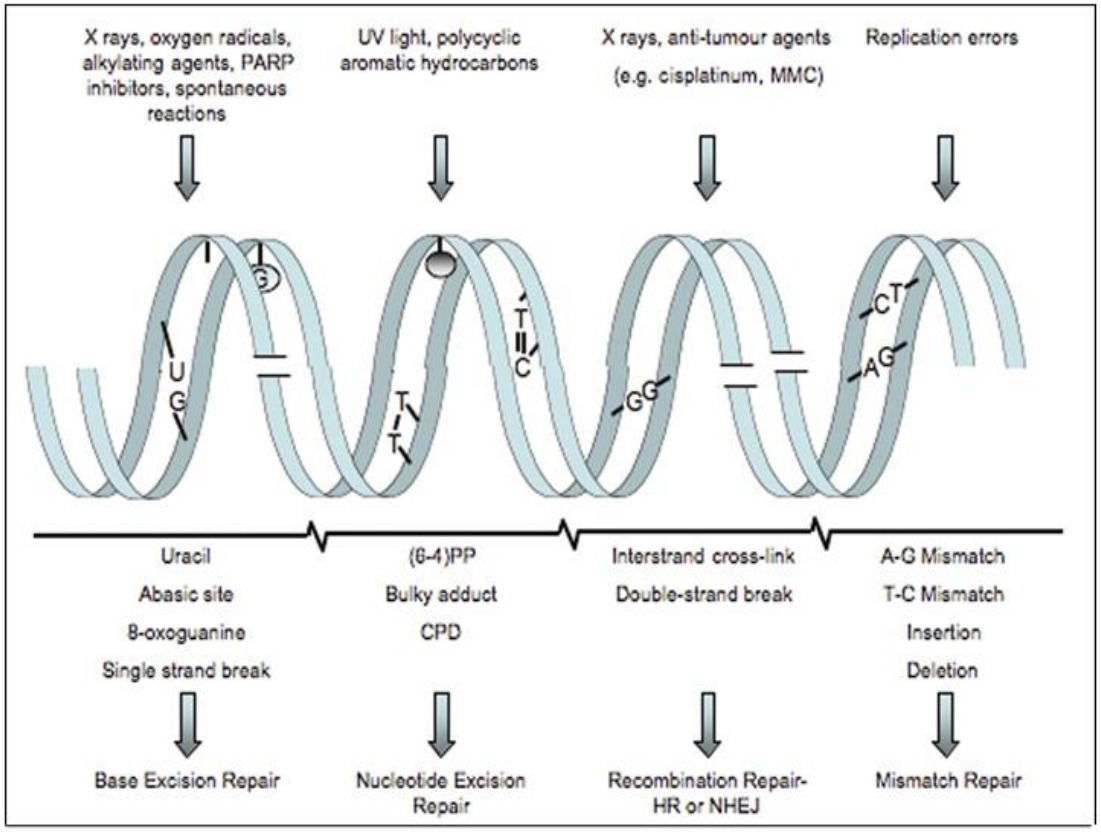


Fig.1.2.1.1. Schematic showing various damaging agents and DNA repair pathways in Living Systems. (Adapted from Martin *et al*., 2008) [20]

The figure above summarises the various repair mechanisms found in living beings (Fig 1.2.1.1).The evolution of repair pathways show that it has evolved independently in different phases. Mechanisms can be as direct as in Light Repair or 'Photoreactivation', where photolyases detect pyrimidine dimers caused due to UVC rays and with the help of

chromophores which convert light energy to chemical energy they cleave these pyrimidine dimer bonds.

A more elaborate pathway is the Base Excision Repair pathway. Single bases get damaged through oxidation, alkylation and deamination and hang out of the DNA helix. For example, guanine can be oxidized to 8-oxoguanine, adenine can be methylated to 3'methyl adenine and cytosine can be deaminated to uracil. If left unattended, these will cause mispairing during replication and thus cause mutations. DNA glycosylases recognize these bases and remove them. The AP endonucleases then cleave the sugar phosphate backbone which is then resealed by polymerases like polX and ligases.

UV rays cause photoproducts like pyrimidine dimers and these bulky adducts cannot be repaired by the Base Exicision Repair pathway. The Nucleotide Excision Repair (NER) comprising UvrA, UvrB, UvrC and UvrD are employed to repair these bulky lesions. The dimer lesions are recognized by UvrA-B. UvrA then leaves the complex making way for UvrC. UvrB-C then makes a 12 nucleotide incision spanning the dimer segment. UvrD, a helicase removes the damaged segment and DNA Pol I and ligase then fill the gap. A specialized addition to NER is Transcription Coupled Repair, wherein RNA polymerases get stalled at the dimers and recruit a protein TCRF, which uses ATP hydrolysis to bypass the RNA polymerase over the lesion and also recruits UvrA at the damaged site. NER is a major repair pathway which is well distributed in living systems and repairs the bulk of UV induced damage. In eukaryotes, around 30 genes are involved in this pathway while in bacteria this is accomplished by 4 genes [21].

Another repair pathway is the Mismatch Repair pathway , which is invoked due to errors caused by incorporation of mismatched bases during replication and recombination.

Mismatch repair pathway is strand specific and corrects the errors on the newly synthesized strand by recognizing the state of its methylation. Usually the template strand is methylated and the newly synthesized strand is unmethylated. The Mismatch Repair pathway comprises of MutS, MutH and MutL which specifically targets the newly synthesized strand. The MutS dimer recognizes the mismatch. Then MutS forms a MutS-DNA complex. The MutL dimer then binds to this complex and this activates MutH to nick the newly sysnthesized strand. This nicked DNA segment having the mismatch is then removed by UvrD helicase along with exonucleases. The resulting gap is filled by DNA pol III and DNA ligase.

Bacteria also possess the SOS Repair pathway as studied in *E. coli*.This is a coordinated response to the challenge of DNA damage. The SOS repair pathway is made up of RecA, the LexA repressor, Sul A and repair genes of the NER pathway and a translesion DNA synthesis system comprising UmuDC. The LexA repressor binds to a regulatory element, the SOS-box, upstream to the genes involved in SOS response. When DNA is damaged, it causes replication fork arrest at the damaged site and single stranded DNA is released. RecA binds to this single stranded DNA and gets activated as a protease. It then cleaves the LexA repressor, and the proteins involved in SOS response are now expressed. Two things happen-1) Sul A stops cell division by interacting with the cytoskeleton protein FtsZ and 2) UmuDC initiates DNA synthesis bypassing the damaged site. This causes mutations to be maintained and hence the SOS response is considered to be a 'error prone repair' system. However in most cases the cells survive the DNA damage rather than dying due to replication fork arrest [22].

All these above mechanisms can repair DNA only in presence of a complementary DNA strand and hence cannot tackle a double strand break. Double strand breaks occur either via endonuclease mediated mechanisms or by oxidative stress or directly by ionizing radiation. Cells have SOS repair and Transcription Coupled Repair mechanism to bypass SSBs but none to bypass DSBs and hence DSBs are lethal. As shown in the figure below (Fig 1.2.1.2) The Recombination Repair pathway is invoked for DSBs and, recombination repair can be of 2 types (1) Homologous Recombination and (2) Non-Homologous End Joining (NHEJ).

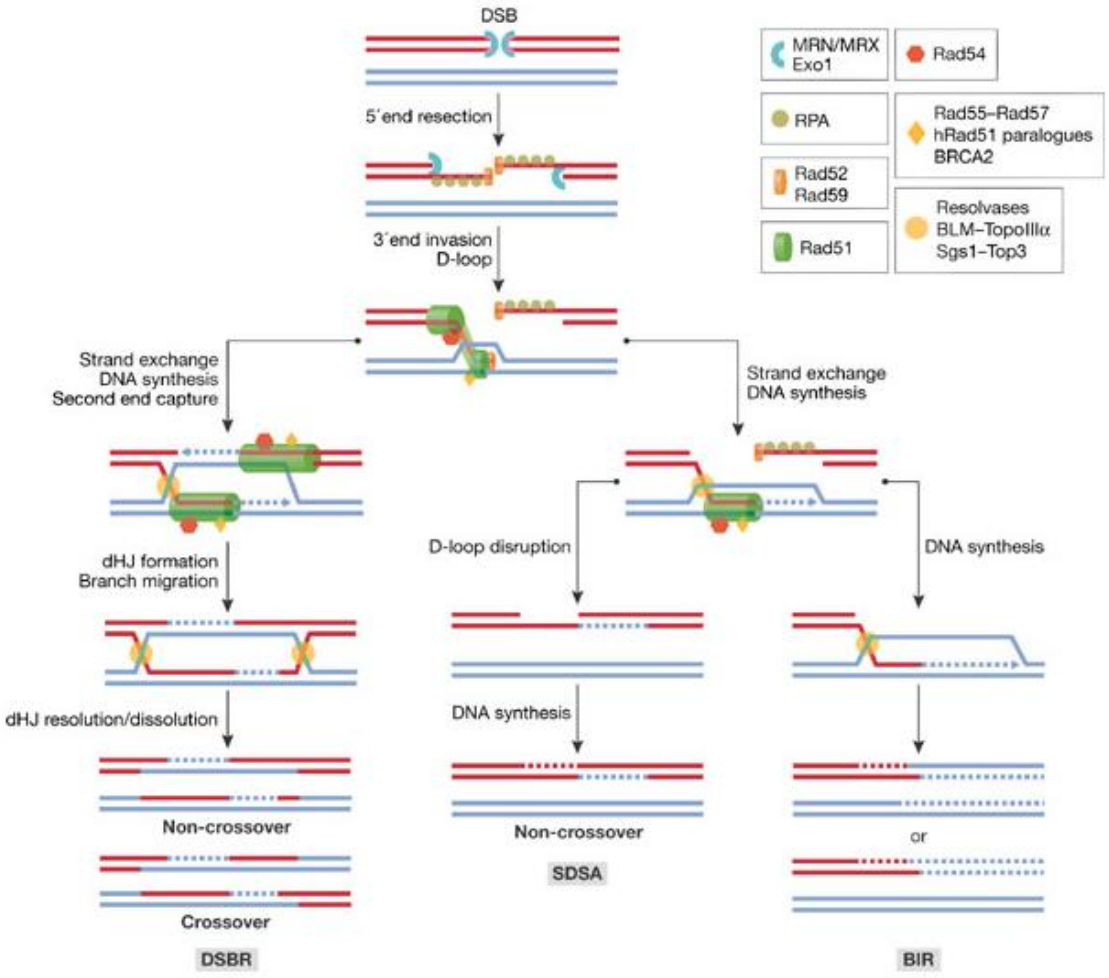


Fig. 1.2.1.2. Double Strand Break Repair Pathway showing the possible ways in which the breaks can be repaired. SDSA and BIR are predominantly seen in eukaryotes while DSBR is a universal mechanism. (Adapted from Aguilera and Boulton.,2007) [23].

The prerequisite of homologous recombination is a complementary DNA strand. It is seen in both eukaryotes as well as prokaryotes. In eukaryotes members of RAD52 epistasis group carry out homologous recombination while in prokaryotes 2 pathways, namely RecBCD and RecFOR pathways are known. RecBCD has been well characterized in *E. coli*. RecBCD is a sequence regulated DNA helicase-nuclease [24]. It binds to a blunt DSB and unwinds the DNA in an ATP dependent fashion and acts as a weak 3'-5' nuclease. On meeting a chi (**χ)** sequence-5'-GCTGGTGG-3', it alteres its activity and becomes a 5'-3' double stranded nuclease -helicase and generates free 3'-OH ends. It also assists RecA to bind to these ends for further steps of recombination. In *D. radiodurans*, RecBC is absent and the major pathway of DSB repair is the RecFOR pathway. Here, RecQ helicase unwinds the DSB and RecJ exonuclease degrades the 5' strand leaving the 3'-OH strand free. RecA then binds this free strand, and then this nucleoprotein complex invades a homologous partner to form a D-loop. Pol III then extends this D-loop by extending the 3' strand. Due to this process a junction of interlinked strands is created called the *Holliday Junction*. The RuvABC resolvase participates in resolving this junction and creating 2 chimeras. A modification of this pathway is Synthesis Dependent Strand Annealing (SDSA) where the resultant chimera is not formed. This mechanism is seen in yeast and *D. radiodurans*. The NHEJ pathway is a major repair pathway in eukaryotes and also seen in mycobacteria and bacilli [25]. This

pathway has mostly been characterized in yeast, mice and humans. NHEJ is active all the time in cell cycle and especially during G1 phase. Here, the key player is the Ku protein. The bacterial protein is a homodimer while the eukaryotic protein is a hetrodimer made of Ku70/80. Ku identifies a DSB and holds the ends of the 2 strands together as a clamp. Then it recruits DNA-PKcs and then this complex now becomes the DNA-PK complex. DNA-PK recruits Artemis, a multifunctional nuclease which blunts the 2 juxtaposed ends and also XRCC4/Ligase whiles seals the ends. The choice between homologous recombination and non-homologous recombination is regulated by a signalling pathway comprising ATM/ATR kinases and DNA-PK and also by the nature of the DSB. Blunt or nearly blunt ends are processed by NHEJ while ends with overhangs are processed by homologous recombination.

1.2.2 Double Strand Break Repair in D. radiodurans

As discussed above, *D. radiodurans* has a mammoth task of reassembling its shattered genome. Certain physical aspects have evolved in *D. radiodurans* to assist the reassembly of its genome. It is multigenomic, and multiple copies help in the homology assisted recombination repair pathway. Based on electron microscope and epifluoresence studies, Minsky and co-workers proposed an interesting hypothesis. They observed that DNA in *D. radiodurans* is aggregated into compact toroids which they hypothesized would help NHEJ pathway proteins to repair the DNA [26].The necessity and importance of this toroid formation in radioresistance has been refuted in several other studies [27], but it is

agreed that compaction of the genome could help the repair process. Also, the small molecule antioxidant proteome comprising peptides and nucleosides in complex with Mn^{2+} polyphosphate complex helps protect key DNA repair enzymes [28]. Bioinformatics studies done with the *D. radiodurans* genome showed that there were no new potential DNA repair proteins and *D. radiodurans* shares the same set of repair enzymes seen in *E. coli* [19]. Further, it has been shown that *E. coli* pol I is able to restore a *polA* strain of *D. radiodurans* hinting that ROS scavenging is more important in radioresistance and in the cellular milieu any repair system can efficiently repair damaged DNA, even extremely damaged DNA, provided ROS generated by ionizing radiation is removed from the system [29].

Genome sequence studies also showed that it lacks photolyases for the direct repair of thymine dimers and also a viable SOS response although it has 2 alleles of lexA. *D. radiodurans* has the Base Excision Repair, Nucleotide Excision Repair and Recombination Repair Pathways pathways. Both Base Exicision Repair and NER seem to be efficient since single strand breaks are repaired within 90 minutes of ionizing radiation damage [30,31].

Studies have shown that recombination repair pathway in *D. radiodurans* is mediated by RecJ exonuclease along with components of RecFOR pathway and either RecQ helicase or UvrD helicase to load RecA and carry out homologous recombination [32]. Seminal studies by Radman and co-workers has shown that *D. radiodurans* may work with ordinary proteins but the mechanism of reassembly is unique [33,34] and they have named this process as Extended Synthesis Dependent Strand Annealing (ESDSA) as it is similar to the DSB repair mechanism Synthesis Dependent Strand Annealing (SDSA) in

yeast. ESDSA occurs in the initial 2-3 hours of reassembly of the deinococcal genome after a dose of irradiation, when it is grown in a complex medium (TGY medium-1% Tryptone, 1% glucose, 0.5% Yeast Extract). During this time, first there is a massive degradation of DNA. Amongst the possible candidate nucleases are RecJ and SbcCD. RecJ has been shown to have 5'-3' exonuclease activity and also that it is an essential gene, as conditional mutants which don't express RecJ die away [32]. This is a regulated exonucleolytic phase where single stranded DNA with free 3'OH ends are generated which can invade complementary strands of DNA assisted by RecA. Interestingly, studies have shown that RecA stimulates this exonuclease activty. These single strands then invade the complementary strands after which an extensive round of synthesis takes place. Unique to ESDSA is the extent of this synthesis, where upto 20-30 kb DNA is freshly synthesized. This amount of fresh synthesis is at least 4 times more than what has been reported for synthesis dependent strand annealing (SDSA) in yeast and drosophila where DSBs are repaired by SDSA [35,36]. In SDSA, the D loop formed is transient, rather it is a bubble and thus extension of a strand needs several rounds of invasion. Also, as the D loop is not extensive, resolvases cannot pair the template and freshly synthesized strand, and these new strands slip off and the recombinases join the complementary new strands together without the need for resolvases. It is believed that the mechanism is similar in *D. radiodurans* and several such freshly synthesized overlapping strands exist and in the second phase RecA dependent recombination repair pathways recombine these freshly synthesized strands into intact molecules. Thus the reconstituted genome is a mosaic of old and new patches of DNA. This patchwork nature was clearly visible when *D. radiodurans* cells were grown in the presence of the thymine analog BrdU. BrdU then

is incorporated into the newly synthesized DNA and when these cells are exposed to a few kilojoules of UV light, it causes DSBs.The presence of DSBs thus confirm both the fresh synthesis of DNA and nonreciprocal crossover events [33].

Is ESDSA a prerequisite for repair of DNA having extensive double strand breaks? Till other radioresistant bacteria are studied in detail, it would be difficult to guess as SDSA or ESDSA is not prevalent in bacteria, although SDSA has been implicated in homing of group I introns in phage T4 [37]. How this eukaryotic repair mechanism was adopted by *Deinococci* remains to be discovered. It is interesting to note that in several fungi where SDSA is a predominant mitotic DSB repair mechanism, do exhibit radiation resistance, for e.g the D_{10} for *C. albicans* is 1.1-2.3 kGy. Also, members of spp *Filobasidium, Aspergillus, Curvularia, Ustilago* have been found to be radiation resistant [38]. A study was done in *S.cerevisiae* to study the genes required for ionizing radiation resistance by generating a library of deletion mutants. This study showed that members of the RAD52 epistasis group play a critical role in radiation resistance. Members of this group can be broadly subdivided into the MRN (nbs1) complex, which are involved in processing of DSBs in repair pathways and the RAD51 subgroup which is involved in all homologous recombination processes [39]. An earlier study showed that sbcCD in *D. radiodurans* functions in a fashion similar to the MRN complex [40].Again *D. radiodurans* has an homolog of RAD51 (RecA) , RAD52 (DdrA) and RAD54(DR1259) but further investigations are needed to find homologs of RAD 55 and RAD 57 which have been shown to be essential for both crossover mediated recombination repair as well as SDSA.

Curiously, *S.cerevisiae* also encodes certain key proteins of the NHEJ pathway, including the critical protein Ku70/80 (*YKU70* and *YKU80*) and a specialized ligase DLN4 but till

date no homolog of a DNA dependent protein kinase and Artemis, which are also components of NHEJ has been observed [41]. Thus yeast in spite of being a eukaryote like mammals, uses NHEJ as a backup pathway while the major repair pathway is by homologous recombination and this preference has a notable similarity with *D. radiodurans*. Other than accuracy, there may be yet another reason why *D. radiodurans* may have evolved the ESDSA pathway as a major repair pathway. Single stranded DNA has been shown to be signal for cell cycle checkpoint in eukaryotes and all present evidence shows that while repair is in progress, the cell cycle in *D. radiodurans* remains arrested [42]. Thus the copious amount of single stranded DNA generated during ESDSA may additionally act as a cell cycle arrest signal. A preference for error free homologous recombination explains how in spite of apparent activation of a transposase, no significant mutations are seen in the repaired DNA [43]. But ESDSA also leaves certain questions unanswered. Why does the synthesis extend for so long? How processive is the process? How such long strands of single stranded DNA are maintained without any secondary structure and that there are no slippages during possibly several rounds of strand invasion? And what are the key players in this mechanism? Earlier studies have implicated roles for RecA, RecFOR, DNA Pol III and DNA Pol A, but their counterparts are present in other radiosensitive bacteria too. *Could novel genes be involved in this process*? Almost 50 % of the Deinococcal proteome, like proteomes of many species, comprises of hypothetical proteins, which have no homology to known proteins and it has been speculated earlier that these proteins may play important role in radiation reistance [44]. Studying the exact mechanism of ESDSA and the possible role of novel proteins in

this process is important in understanding the robust error free repair system in *D. radiodurans*.

Besides the machinery of proteins involved in repair another important aspect of repair process is its regulation. *D. radiodurans* employs well known and possibly novel pathways to repair a highly shattered genome with negligible errors which means that regulation of metabolic processes like nucleolytic degradation, synthesis of DNA etc becomes very important. It is known that DNA repair is linked to cell cycle as cell cycle progression takes place only after repair. Again as the plot below highlights (Fig 1.2.2.1) the pattern of expression at 3 kGy is different from that of 15 kGy and surprisingly, in a highly shattered genome, selectively more genes are over-expressed at 15 kGy than at 3 kGy.

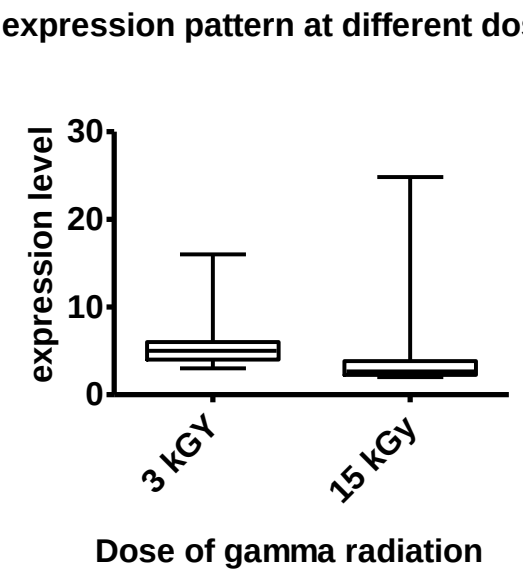


Fig 1.2.2.1. Altered expression profile of genes involved in DNA repair at different doses at 3 kGy and 15 kGy shows that expression of genes sharply increases at higher doses.[Das and Misra, Unpublished data,[45]]

It has previously been shown that upstream to several genes over-expressed in radiation stress in *D. radiodurans*, like DR0326, DRA0346, DR0423 and several DNA repair genes, a strong palindromic motif designated as radiation/desiccation response motif (RDRM) is present [46] and this is a conserved motif as it is present also in *D.geothermalis* and *D.deserti*. Thus, possibly *D. radiodurans* employs novel regulatory networks to repair the numerous DSBs after a dose of radiation.

1.3 Hypothetical Proteins and their Role in *D. radiodurans*

1.3.1 Introduction to Hypothetical Proteins

Algorithms to predict genes, predict stretches of DNA with proper start and stop codons and corresponding amino acid sequences and name them as 'Open Reading Frame'. If translations from these ORFs match with atleast 30 % identity with known proteins in the public databases like NCBI, EBI, DDBJ they are annotated with the names of already known genes, otherwise the polypetide sequences derived from the ORFs are named

'Hypothetical Proteins'. Fischer and Eisenberg coined the name 'ORFans' for the parent ORFs of these hypothetical proteins [47] .

It was widely believed that the number of hypothetical proteins would reduce as more and more genomes are sequenced, however as the plot below shows (Fig 1.3.1.1) that has not been the case. With every new genome sequence, a newer set of hypothetical proteins are added to the databases. A caveat has been that half of the ORFans encoding hypothetical proteins, are less than 150 bases and may be wrongly annotated as encoding for proteins, and in fact some reports also refer to these small ORFs as ELFs ("evil little fellows"), hinting to the fact that these might actually be noncoding regions or regulatory regions [48].

So what is the origin of hypothetical proteins? This has been an intriguing question and there are two predominant theories for the existence of hypothetical proteins.

1) These are remnants of once active genes, which are now nonfunctional and would in future gather more stop codons, frameshifts etc and morph as 'pseudogenes' and then in the future get eliminated. An example is GULO (L-gulono-γ-lactone oxidase). GULO aids in the biosynthesis of vitamin C in most of the mammals but it is disabled in humans and primates [49]. Another example is an inactive form of caspase 12 in humans [50]. But in this case, the inactive form is selected for and maintained as these individuals are more resistant to severe sepsis.

2) The other theory is that many of these hypothetical proteins are duplicate copies of functional genes (paralogs) and are slowly evolving newer functions (*de novo* evolution). For example, BSC4 gene in *S.cerevisiae* is expressed as non-coding RNA in species closely related to yeast but in yeast itself, it is a protein involved in DNA repair. CLLU1,

C22ORF45 and DNAH10OS in humans are functional proteins detected in proteomics data but these are not transcribed in primates [51].

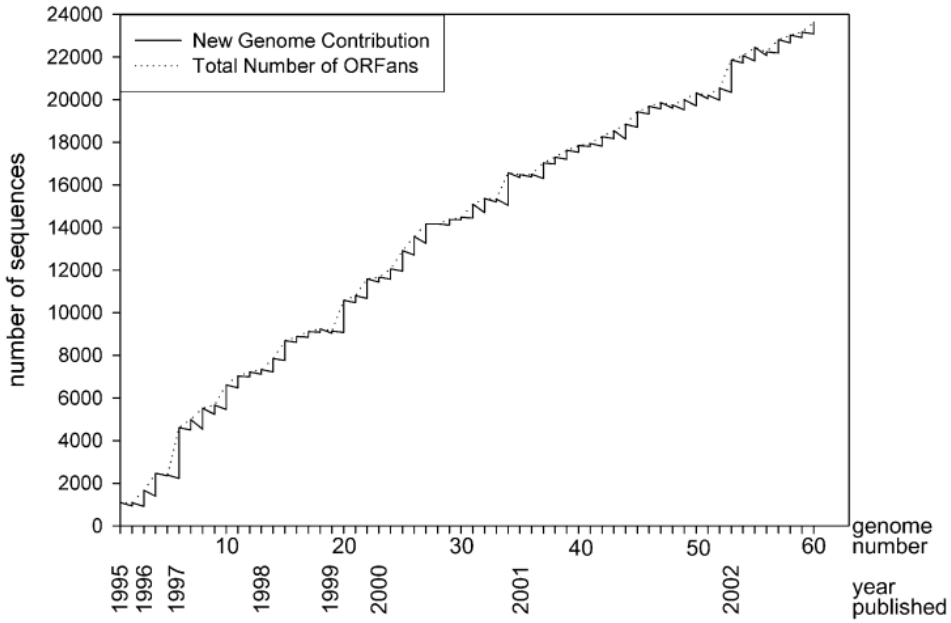


Fig 1.3.1.1. Plot of hypothetical proteins vs genomes published. Hypothetical proteins (ORFans) are increasing with the number of genomes sequenced (adapted from Siew *et al*.,2003) [48].

Many studies show that hypothetical proteins are actually expressed for e.g Bennet *et al*.,2001[39] have shown that around 3000 hypothetical proteins in *S.cerevisiae* are actively transcribed while similar studies in *E. coli* have shown that several hypothetical proteins are actively expressed [52]. The acknowledgement of the role hypothetical proteins in cell survival has been slow and only recently they are being biochemically characterized, as till 1998 only a single hypothetical protein from *B.licheniformis* was characterized, whereas in 2011 alone 44 such hypothetical proteins were characterized.

1.3.2 Hypothetical Proteins in Radiation Response in D. radiodurans

A study implicating hypothetical proteins in the radiation response of *D. radiodurans* was that of Liu *et al*.,2003 [45]. As the plot below shows (Fig 1.3.2.1) several hypothetical proteins were higly expressed in a pattern similar to the total cellular expression indicating that they are are necessary in the repair process of *D. radiodurans.* They concluded that a majority of induced genes were poorly characterized and needed further characterization. Suspected ELFs were also present, for e.g in the case of DRA0234 it was only 171 bp long and the theoretical transcript formed stable stem-loop structures, indicative of noncoding RNA. Nevertheless, it showed that hypothetical proteins like DRB0098-DRB0100, uncharacterized ABC transporters DR1356-DR1359, kinases with unknown specificity DR2467, DR0394, DR0609 and DR1564 etc were highly induced in response to radiation.

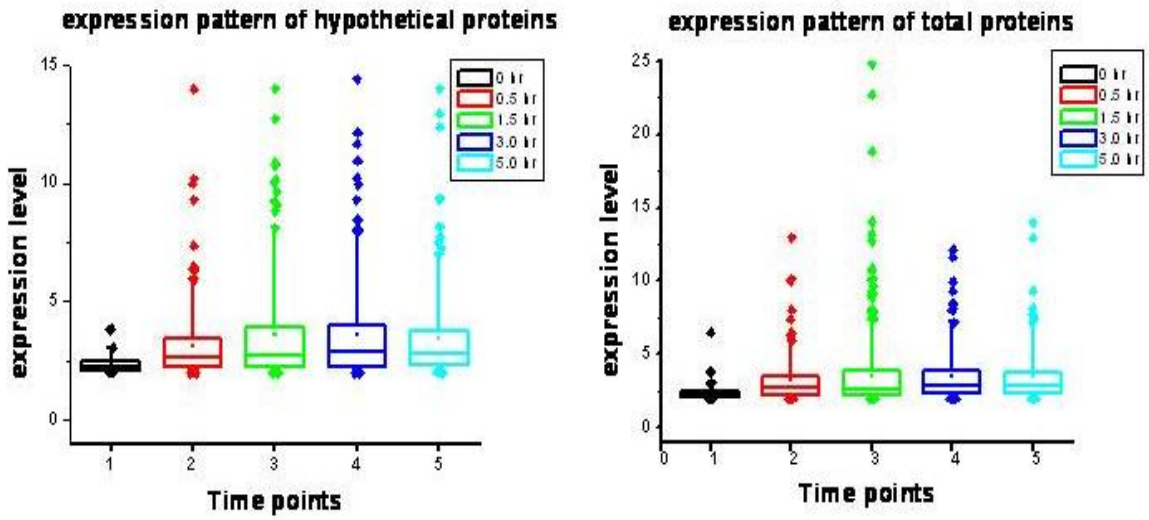


Fig 1.3.2.1. Expression profile of hypothetical proteins with respect to total proteins. Box Plot showing the similarity of the pattern of expression of hypothetical proteins to the expression pattern seen during the repair process for all genes [Das and Misra,unpublished data,[45]].

Another study by Tanaka *et al*., 1996 [53] deleted the genes of hypothetical proteins which were over-expressed and found out that in several cases, the deletion mutant was debilitated. Thus the role of hypothetical proteins like *ddrA* (DR0423), *ddrB* (DR0070), *pprA* (DRA0346) etc in the radiation response of was discovered. Purified DdrA protein was shown to bind to 3' ends of single stranded DNA and protect them from exonuclease degradation [54]. Likewise purified PprA was characterized *in vitro* as a DNA end binding protein which prevented DNA from Exo III attack and phenomenally increased ligase activity of ATP dependent T4 DNA ligase and NAD dependent E. coli DNA ligase [55]. As the table below shows (Table 1.3.2.1) hypothetical proteins play an important

role in the metabolism and repair process in *D. radiodurans* and although we distinguish hypothetical proteins as a separate entity, in *D. radiodurans* as in other organisms, they are integrated seamlessly into the metabolism.

Table 1.3.2.1 A representative list of ORFans encoding for hypothetical proteins characterized in *D. radiodurans*

Sr No	Gene	Locus tag	function	reference
1	*irr*E or *ppr*I	DR0617	A deletion study was done. IrrE is a regulatory protein which enhances RecA transcription	[56]
2	*mnt*E	DR1236	manganese efflux protein	[57]
3	*sig*1	DR0180	Deletion of the gene makes *D. radiodurans* sensitive to IR ECF derived heat shock Sigma Factor, controls groESL and dnaKJ operons.	[58]
4	*sig*2	DR0804	Same as above, but deletion of the gene has a lesser drastic effect	[59]
5	*ddr*A	DR0423	Deletion of the gene made *D. radiodurans* sensitive to radiation and the purified protein protected ssDNA *invitro*	[54]
6	*lex*A2	DRA0074	A disruption mutant showed a	[60]

			higher resistance to radiation	
7	*drarnl*	DRB0094	DraRnl seals RNA strand nicks in double stranded RNA or RNA:DNA hybrid	[61]
8	*ddr*B	DR0070	The deletion mutant was sensitive to γ radiation and purified protein stimulated single stranded DNA annealing	[62]
9	*nim*A	DR0842	A reductase involved in metronidazole antibiotic resistance	[63]
10	*dr*RRA	DR2418	The deletion mutant was sensitive to radiation and studies showed it to be a DNA binding response regulator	[64]
11	*ddr*P	DRB0100	The purified protein forms a complex with AMP but does not ligate DNA or RNA	[65]
12	-	DRB0098	The purified protein acts as a polynucleotide kinase. Only other example from bacteria is in clostridia	[65]
13	*drOxy*R	DR0615	The deletion mutant is sensitive to H_2O_2, but not to IR. It is a transcriptional activator involved in Mn/Fe transport	[66]

14	-	DR0505	the purified protein was a hairpin endonclease and a 3'-5' exonuclease	[67]
15	*cru*F	DR0091	the purified protein is a carotenoid 1-2 hydratase involved in synthesis of the pigment deinoxanthin, which is an important ROS scavenger	[68]

1.4 Role of Bioinformatics in Studying Hypothetical Proteins

As we have seen above, hypothetical proteins have an important role in repair, general metabolism and radiation resistance in *D. radiodurans*. Microarray studies have shown several hypothetical proteins being induced and it important to understand the function of these proteins

As shown in the figure below (Fig 1.4.1.1), there are several bioinformatics tools for searching for functions of hypothetical proteins:-

1) Sequence based – BLAST or FASTA search in nucleotide or protein databases for functional homologs

2) Subsequence or motif based search in PROSITE or PHI-BLAST for finding conserved functional sites

3) Structure based – by homology modeling or fold prediction

4) Feature based-where models are generated based on features derived from amino acid sequence of the query. Feature based methods have also been used to build classifiers for machine learning approaches like Neural Networks and Support Vector Machines which can predict specific functional groups in a high throughput way.

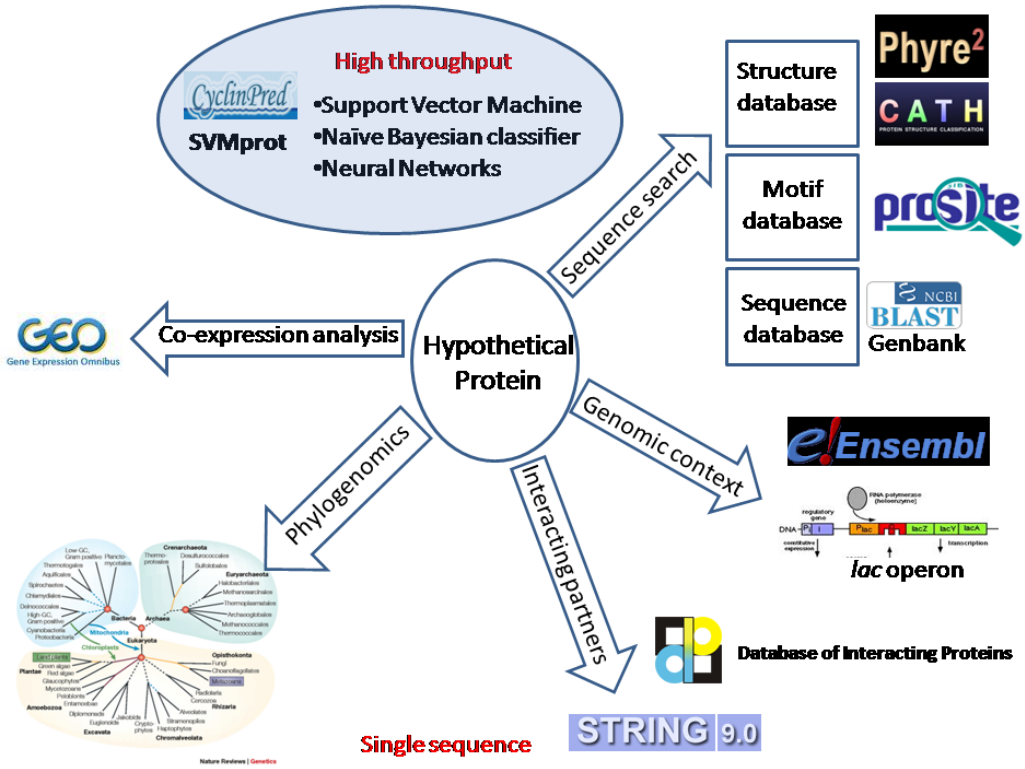


Fig 1.4.1.1. Schematic showing bioinformatics approaches for protein function determination

Because of the common evolutionary past, the key active site residues, residues involved in salt bridges and overall structure of proteins are conserved. This overall conservation is reflected in the amino acid sequence but due to poor sequence homology, BLAST, the standard search tool for comparing an unknown sequence to proteins in a database cannot identify any function for these proteins. An improved algorithm called Position Specific Iterated -BLAST (PSI-BLAST) was developed which would build a statistical probability distribution for the presence of amino acids in a alignement [69]. PSI-BLAST has since then been successfully used to detect remote homologs in several cases. Also, as the table below shows (Table 1.4.1.1) several specialized databases are available which improve the chances of getting relevant matches.

Table 1.4.1.1. List of databases of nucleic acids and proteins adapted from Washietl and Hofacker .,2010 [70]

Database	Description
EMBL, Genbank, DDBJ	central sequence repository
RefSeq	non-redundant and curated sequence
GEO (Gene Expression Omnibus)	gene expression experimental dataset
PDB (Protein Data Bank)	database of experimentally determined

	structure of proteins, nucleic acids and complex assemblies
Entrez Genome	database of complete and nearly complete genome sequences
Swissprot	manually curated protein database
DIP (Database of Interacting Proteins)	online database of experimentally determined protein-protein interaction
PROSITE	database of protein domains, families and functional sites

Secondary structures are the basis for tertiary structure formation and for computational reasons these can be done faster. Also topology and connectivity conflicts between residues are easily revealed by secondary structures. Today secondary structure prediction are fairly accurately done by Neural Networks [71].

Knowledge of the tertiary structure helps in understanding the mechanism of a protein and in evolution the structure that is conserved and not the sequence. Thus hypothetical proteins with remote sequence identity can be properly identified and a function can be predicted. The basic independent structural domain is a fold and as the plot below shows

(Fig 1.4.1.2) most sequence families can be molded onto few 3-D folds in a structure database.

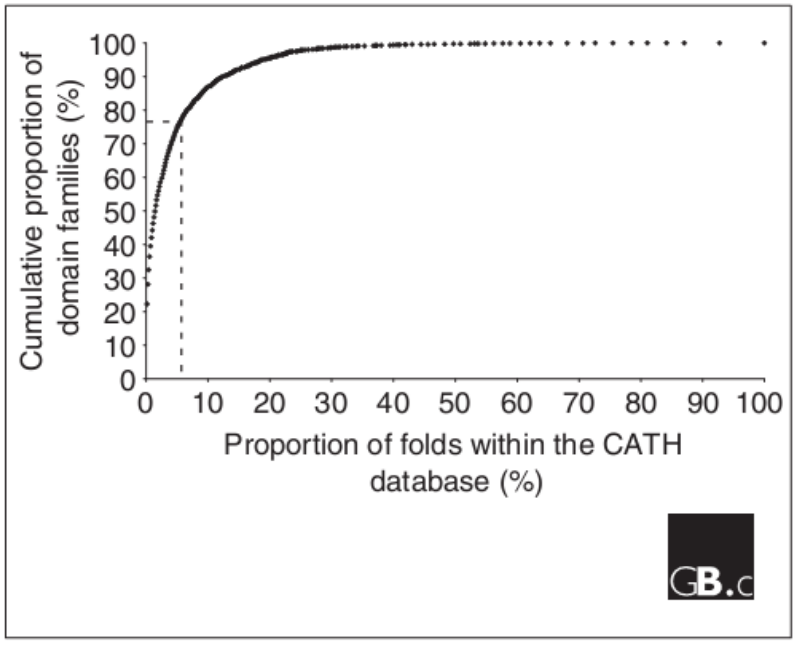


Fig 1.4.1.2. Proportion of the sequence domain families represented by CATH fold groups. A group of 54 CATH fold groups cover 76% of all domains.(Adapted from Pearl *et al*.,2005) [52].

As seen in the table below (Table 1.4.1.2) , there is a correlation between structure and function at the fold level [72]. Again, the protein structure also has a correlation with the ligands it interacts with, as there is a selection pressure on the shape of the clefts and grooves which fit the ligand [73]. Thus most sequences have common folds and most folds have a common function.

Table 1.4.1.2. Relationship between folds in a protein and their function.(Adapted from Marti-Renom *et al*.,2000) [72] .

Fold	Functional Annotation of homologous proteins	Number of members	Percent of the fold
Rossman Fold	Enzyme	131	57%
TIM Barrel	Enzyme	94	85%
IG Fold	Antibody	17	89%
Globins	Oxygen transporter	16	100%
Beta-propeller	Alpha-sialidase	6	66%

Protein structure prediction methods available today can be grouped into (1) homology modeling, (2) threading and (3) ab initio methods. The simplest and fastest method is homology modeling. If the sequence identity is >50%, then this is a reliable method. Here the query sequence is searched across sequences of available structures and a template is chosen. Then the query is modeled on the structure of the template. The 3-D model obtained usually has a resolution of 3-4°A RMSD or even better, and is good for deducing the overall folds and thus the possible function. One can superimpose any motifs found in the query sequence, independently from a Motif Database and try to map the active site in the query.

In bacteria genes are arranged in operons. Thus if an unknown gene in the species of interest, lies in an operon, then it could be predicted to act in that pathway. The plot below shows the co-relation between gene order conservation and evolutionary distance (Fig 1.4.1.3). Gene order conservation has been used as an important tool for finding protein function.

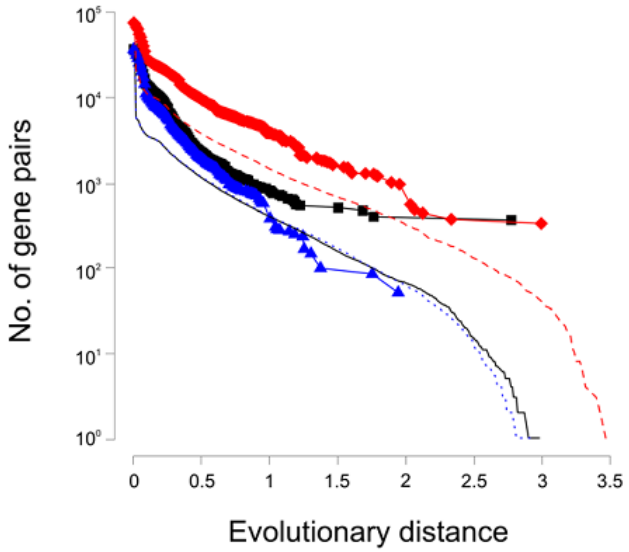


Fig 1.4.1.3. Evolutionary conservation and relative gene orientation (Adapted from Lopez *et al*.,2011) [74].

Besides these methods above, tools have also been developed to predict sub-cellular location, post translational modification and interacting partners. There are also tools available for docking possible substrates with the protein of interest. A very useful approach for gene prediction has been 'Phylogenomics'. Evolutionary processes indicate that all genes have evolved from a common ancestor and changes in genes are co-

terminus with speciation, thus proteins with similar phylogenetic profiles should have similar functions [75,76]

Thus today, with the help of bioinformatics and high throughput experiments like microarray based expression profiling the role of hypothetical proteins in the life of organisms are slowly being revealed.

1.5 Aims and Objectives of this Study

The genesis of this study began with an observation of *D. radiodurans* cells undergoing post irradiation recovery (PIR), after exposure to a 4kGy dose of γ radiation. At this dose of γ radiation the DNA inside the cell shatters into >100 DNA pieces and during repair these pieces are accurately reassembled back by a poorly known network of the proteins which execute ESDSA. In the initial phase a regulated nuclease activity converts the shattered double stranded DNA to single stranded DNA and hence nucleases are an important part of ESDSA. When the cell lysate of the 1hour PIR sample was fractionated by heparin column followed by ion exchange chromatography, a fraction was obtained, which degraded exogenoulsy added double strand DNA in the presence of 5 mM Mn^{2+}. This degradation of the added DNA was inhibited when 1mM ATP was added to the assay mixture. Nucleases need metal ions for their activity and ATP chelates out metal ions but here the ratio of Mn^{2+}/ATP was sufficiently high to rule out this effect, which meant that ATP was modulating the activity of this nuclease. Many proteins are known to be regulated by kinases and inhibition by ATP meant a possible regulation by kinases. Also there are certain reports of nucleases being inhibited by ATP. For all the above

reasons we got curious to find out the possible nucleases/kinases in this fraction. To identify the possible nucleases/kinases, this fraction was analyzed by a sensitive technique, Fourier Transform Ion Cyclotron Resonace-Mass Spectroscopy, which detected ~300 different proteins but no known nuclease or kinase was found. However 11 'hypothetical proteins' were also found and an initial BLAST search showed that 6 (DR0041, DR0390, DR0461, DR1654, DR2417m, DRA0282) of them had retrived matches with some proteins involved in DNA metabolism while the other five had no matches in the databases at an e-value cutoff of 1.0.

Thus there was a possibility for discovering a novel nuclease and its regulatory kinase involved in the unique ESDSA phase of DNA repair in *D. radiodurans*. Also, bioinformatics tools were available to detect the presence of nuclease or kinase functional/structural motifs. In this thesis we have attempted to predict the functions of these 6 hypothetical proteins for possible nuclease/kinase activity using bioinformatics and functionally characterized two of these proteins DR2417 and DRA0282 both *in vitro* and *in vivo*. Total work described in this thesis has been accomplished through following objectives:

1. Theoretical prediction of functional domains in the above ORFs with the help of bioinformatics tools and online databases.

2. *In vitro* characterization of recombinant DR2417 and DRA0282 for their predicted functions and their effect on the response of transgenic *E. coli* cells to DNA damaging agents

3. Molecular genetic studies to understand the role of DR2417 and DRA0282 in DNA repair and radiation resistance of *D. radiodurans.*

Chapter 2

Materials and Experimental Procedures

2.1. Materials

2.1.1 Plasticware and Glassware

Disposable polypropylene microcentrifuge tubes (1.5 ml and 2 ml) and micropipette tips were obtained from Tarsons, India or Axygen, USA. PCR tubes were procured from Axygen, USA. Polypropylene SS34 tubes and GSA bottles were from Tarsons, India. All the plasticwares and glasswares were sterilized by autoclaving. Glasswares were also sterilized by baking at higher temperature for 3 h. Cryovials (2 ml) and low temperature storage boxes were obtained from Laxbro, India and Axygen, USA. All the glasswares were from Corning, USA or Borosil, India.

2.1.2 Chemicals and Media Ingredients

Dehydrated components of culture media (Bacto-tryptone and Bacto-yeast extract) and agar-agar (Bacto-agar) powder were from Difco Laboratories, USA. Fine chemicals were from Sigma Chemical Company, or USB-Amersham, or E.Merck, or Roche Biochemicals GmbH or Pharmacia LKB, or S.R.L. Ltd., or Bangalore Genei (P) Ltd., India. Inorganic salts, urea for sequencing, organic solvents and acids of Excel-R or

Anal-R quality were from Qualigens (Glaxo) India Ltd. or E.Merck India Ltd. Molecular Biology Grade phenol was purchased from SRL, India.

2.1.3 Enzymes and Other Molecular Biology Reagents

T4 Polynucleotide Kinase, T4 DNA ligase, restriction endonucleases and DNA polymerase (Klenow fragment) were from New England Biolabs, USA or Bangalore Genei (P) Ltd., India or USB-Amersham, UK or Roche Biochemicals GmbH, Germany. T4 DNA polymerase was from New England Biolabs, USA or Roche Biochemicals GmbH, Germany. Exonuclease I, Shrimp Alkaline Phosphatase were from USB-Amersham, UK. Lysozyme (from chicken egg-white), DNase I (from bovine pancreas) and RNase (from bovine pancreas) were from Sigma Chemical Company, USA. *Taq* DNA Polymerase was obtained from Bangalore Genie Pvt Ltd, Bangalore. *Pwo* DNA polymerase and deoxynucleotidetriphosphates (dNTPs) were from Roche Biochemical Germany. DNA molecular weight standards were from Bangalore Genei (P) Ltd., India. PCR extraction kit, Gel extraction kit and RNA extraction kit were from QIAGEN, Germany. All the primers used in this study were synthesized commercially.

2.1.4 Radionucleotides and Photographic Materials

[32p] γ-ATP [32p] α-ATP were from Board of Radiation and Isotope Technology (BRIT), India. X-ray films of medical type, polyester based, double-sided were from Kodak India Ltd, India.

2.1.5 Preparation of Dialysis Tubing

Dialysis tubing was cut in 30 cm pieces and soaked in sterile 50mM Tris buffer, pH 7.5 for overnight at 4^{o}C. The buffer was replaced with similar buffer containing 2 % Glycerol and 10 mM EDTA and autoclaved before storage at 4^{o}C. (The tubes from the stock were handled aseptically to avoid microbial growth).

2.1.6 Composition of Stock Solutions

Preparation and storage of most of the stock solutions is given in Table 2.1.6.1. Unless specifically mentioned, the solutions were made in ultrapure water and sterilized without adjusting the pH. Wherever necessary, addition was done to agar media after prior cooling of the medium to ~50°C. Preparation of all the antibiotic stock solutions is given in Table 2.1.6.2. Addition to agar media was done just prior to pouring the plates and only after the media was cooled to ~50°C.

2.1.6.1 List of chemicals

Chemical	**Concentration**		**Storage**
	Stock	*Final*	
Adenosine triphosphate (ATP)	10 mM	1 mM	-20°C
Ammonium persulfate (APS)	100mg/ml	1 mg/ml	-20°C

BSA	10.0 %	As required	-20°C
$CaCl_2$	1.0 M	100mM	4°C
dNTPs	100 mM	As required	-20°C
DTT	1 M	1mM	-20°C
Ethidium bromide	5 mg/ml	0.5 μg/ml	RT.
EDTA (Na salt)	0.5 M	As required	RT.
D-glucose	20.0 %	0.4 %	4°C
Glycerol	98 %	As required	RT.
KCl	2M	As required	RT
IPTG	200 mg/ml	160 μg/ml	-20°C
$MgCl_2$	1.0 M	As required	4°C
$MgSO_4$	1.0 M	10 mM	4°C
PMSF	100mM	1mM	-20°C
SDS (Na salt)	10 %	As required	RT.
Sodium acetate	3.0 M	As required	RT.
Sodium chloride	5.0 M	As required	RT.
N,N,N',N'-Tetramethyl ethylenediamine (TEMED)		As required	4°C

DMSO		As required	RT
Tris.HCl	1.0 M	As required	RT.

2.1.6.2 List of antibiotics

Antibiotic	**Stock Solution**	**Working Concentration**
	(mg/ml)	(μg/ml)
Ampicillin	100 in sterile DW	100
Chloramphenicol	34 in Ethanol	3-5
Kanamycin	25 in sterile DW	5-30
Spectinomycin	100 in sterile DW	75

2.1.7 Composition of Commonly Used Analytical Reagents

All commonly used analytical reagents like 1 M Tris.Cl, 0.5 M DTT, 0.5 M EDTA(pH 8.0), Phenol, TE buffer, GTE(for alkaline miniprep), 5 M Potassium Acetate, 5X TBE, 10X TAE, 6X DNA loading dye, were prepared as per Sambrook et al.,2001 [77] .

Phosphate Buffered Saline

Dissolve 8 gm NaCl, 0.2 gm KCl, 1.44 gm Na_2HPO_4 and 0.24 gm KH_2PO_4 in 800 ml distilled water. The pH to 7.4 was adjusted with dilute HCl and volume was made up to 1000 ml. The solution was dispensed in 100 ml bottles and sterilized by autoclaving. The buffer was stored at RT.

PMSF

174 mg of the phenylmethylsulfonylfluoride powder was dissolved in 10 ml isopropanol and solution was vortered vigorously to dissolve the content for a stock concentration of 100 mM . One milliliter aliquots were taken in alumunium foil wrapped vials and preserved at –20^{o}C. The solution was kept on ice for 10- 20 min just before the use.

Mitomycin C

20 microgram mitomycin C was dissolved in 1 ml sterile double distilled water for a stock concentration of 20 µg/ml and preserved at –20^{o}C in dark.

Metal Affinity Purification Matrices

Immoblized metal affinity chromatogrpahy matrices were procured from GE healthcare (Sweden) as fast flow metal chelating materials and NINTA agarose from QIAgen (Germany).

Oligonucleotides

Normal oligonucleotides and RNA oligos were synthesised from Metabion, Switzerland/Sweden and MWG, Germany.

2.1.8 Composition of Media

All the media described here were sterilized by autoclaving at 15 p.s.i. for 20 min. and stored at RT., unless otherwise mentioned. Agar-agar was added at a final concentration of 1.5% (unless otherwise mentioned) to prepare solidified media.

Luria Bertani Broth (LB)

NaCl, 5 gm.; Bacto-yeast extract, 5 gm.; Bacto-tryptone, 10 gm.;Milli RO grade water to a final volume of 1 litre and the pH adjusted to 7.4 with NaOH.

TGY medium

Tryptone, 5g; Yeast Extract, 3g, Glucose, 1g were dissolved in 1000 ml of distilled water and pH=7.4 was adjusted with NaOH.

2.1.9 Composition of Other Reagents Used in Molecular Biology Studies

2X Refolding Buffer

0.2 M Tris-HCl pH 7.6 ;0.8 M L-Arginine ;20% glycerol ;0.5 mM GSSG* ;5 mM GSH*

Prepare in sterile glassware and with sterile milliQ water. Autoclave for 5 minutes.*Add GSSG and GSH after autoclaving.

Protein Storage Buffer

10 mM Tris-HCl pH 7.6 ;50 mM KCl ;0.1 mM EDTA ;50 % glycerol

DNase I

5 mg DNase I, 2.6 ml sterile (98 %) glycerol and 150 µl 5 M NaCl. were mixed to dissolve in 2.25 ml sterile Milli Q grade water by repeatedly tapping or vortexing, dispensed as 500 µl aliquots and stored at -20°C.

RNase A

50 mg RNase A was dissolved in 5 ml of 10mM sodium acetate, pH 5.2. The solution was heated at 115^{o}C for 15 min and heating was put off and temperature was allowed to decrease slowly to RT. To this, add 500µl of 1MTris.HCl , pH 7.5 for a final concentration of 10 mg/ml and dispensed in aliquots and stored at –20^{o}C.

Lysozyme

Lysozyme was prepared at a concentration of 50 mg/ml in 10 mM Tris.HCl (pH = 8.0) and used fresh.

Acrylamide Stock

29.2gm acrylamide and 0.8 gm of N, N-Methylene (bis) acrylamide was dissolved in 70 ml double distilled water at 37^{o}C. The mixture was stirred for 30 min to assure complete dissolution and then the volume was made up to 100 ml. The final stock concentration was 30%. The solution was filtered through Whatman 540 filter disc and stored at 4^{o}C.

Urea gel-loading dye

One milliliter of deionised formamide (100 %) was mixed with 20 µl of 0.5M EDTA, pH 8.0 and 40 µl of each of the bromophenol blue (2%) and xylene cyanol FF (2%). The 100 µl aliquots were made and preserved at –20^{o}C.

2X Laemmelli Cracking buffer

Dissolve 460 mg of SDS, 7.6 mg of EGTA, 20 mg of sodium azide, 2 ml of Glycerol, 150 µl of 100mM PMSF, 1.0 ml β-mercaptoethanol in 2.5 ml of 0.5M Tris-HCl, pH 6.8 and 40µl of 0.5 % bromophenol blue was added and volume was made up to 10 ml with distilled water.

10 x SDS PAGE electrophoresis buffer

Dissolve 30 gm Tris base, 144 gm Glycine and 10 gm SDS in 800 ml double distilled water. The volume was made up to 1000ml with distilled water.

2.2 Methods

2.2.1 Microbiological Methods

2.2.1.1 Maintenance of the Bacterial Stocks of *Escherichia coli*

Bacterial stocks were maintained under frozen conditions in presence of cryoprotectants. The logarithmically growing culture was spun and resuspended in minimal medium and

/or LB medium to cells density of 10^9 cells per ml. The cell suspension was mixed with 20% glycerol in autoclaved polypropylene screw capped tubes and stored in freezing temperature tolerant boxes, at –70^{o}C. The stocks were numbered and entered in the bacterial culture book with full details.

2.2.1.2 Growth of the Bacterial Strains

Different bacterial strains used in this study are summarized in Table 2.2.1. Bacterial culture was taken from long-term storage stocks directly on LB / TGY agar plates and or plates containing appropriate antibiotics if required. The plates were incubated at appropriate temperature overnight. Single isolated colonies were inoculated in broth with or without antibiotics and allowed to grow at 37^{o}C for *E. coli* and 30^{o}C for *D. radiodurans* at shaking speed of 150 RPM overnight. The overnight grown liquid culture was diluted to 1:100 in fresh broth with selection pressure if required, in the flasks having 10 times the air space to the volume of the medium. The culture was allowed to grow with shaking at 200 RPM at appropriate temperature. The aliquots were drawn at different time intervals and optical density was measured at 600 nm. In parallel an appropriate dilutions were plated on agar plates and the numbers of colonies were counted. The growth curve drawn under defined set of conditions was used in determining the different growth phases of the culture for different experiments. Total number of colonies appeared on agar plates after different treatments of *D. radiodurans* cells were taken for calculating the survival efficiency.

2.2.1.1 List of bacterial strains

Bacterial strains	**Genotype**	**Source**
Deinococcus radiodurans R1	Wild type	Schafer *et al* .,1998[78]
ΔdrA0282 *Deinococcus radiodurans mutant*	*drA_0282*	this study
Δdr2417 *Deinococcus radiodurans mutant*	*dr2417 partial mutant*	this study
E. coli DH5alpha	*F^- / endA1 hsdR17 glnV44 thi-1 recA1 gyrA relA Δ (lacIZYA-argF) U169 deoR (Φ80dlacΔ (lacZ)M15)*	NEB
E. coli BL21	*F^-ompT gal [dcm] [lon] $hsdS_B$ DE3::T7RNA polymerase gene*	NEB
E. coli AB1157	*thr-1, araC14, leuB6(Am), Δ(gpt-proA)62, lacY1, tsx-33, qsr'-0, glnV44(AS), galK2(Oc),* LAM-, *Rac-0, hisG4(Oc), rfbC1, mgl-51, rpoS396(Am), rpsL31(strR), kdgK51, xylA5, mtl-1, argE3(Oc), thi-1*	Dewitt *et al*.,1962

E. coli JC1553	*leuB6 fhuA2 lacY1 glnV44(AS) gal-6 hns-1 hisG1(Fs) rfbC1 galP63 argG6 malT1(λ^R) xyl-7 mtlA2 metB1 recA1 rpsL104*	Clark *et al*.,1964
E. coli DM49	*lexA3*	Mount *et al*.,1972[79]

2.2.1.3. Radiation Treatment of Cells

Radiation treatment of *D. radiodurans* cells

D. radiodurans cells were grown to early stationary phase and mid log phase as required, in TGY medium. The cells were collected and washed with sterile normal saline. The cells were resuspended in half volume of fresh TGY broth and irradiated at different doses of γ radiation in ice in γ irradiator (GC5000, BRIT) at the dose rate of 9.87kGy/h. The cells were diluted appropriately and plated on TGY agar plates. For UV treatment, the cells plated on agar plates were directly irradiated with different doses of UV radiation. Mitomycin C (20µg/ml) was used as described in (Khairnar *et al.,* 2008). The plates were incubated at 30^{o}C for at least 36 hours for recording total viable counts in response to specific treatments.

Radiation treatment of *E. coli* cells

For UV radiation 1% inoculum of recombinant *E. coli* was grown in LB + 100 µg/ml ampicillin for 2 hours at 37°C/180 RPM till midlog phase and 200µM IPTG was added

and and the culture was further grown for 2 hrs more and then exposed to various doses of UV at 0.861 J/m^2/s.For γ radiation induced cells were irradiated in a GC-220 (BRIT) γ irradiator at a dose rate of 7 gray/min.

2.2.2 Methods used in Molecular Studies

2.2.2.1 Isolation of Plasmid DNA (mini prep)

Plasmid DNA was prepared using alkaline lysis method as described in Sambrook *et al.*,2001[77].

2.2.2.2 Isolation of Genomic DNA

The 2 ml overnight grown culture of *D. radiodurans* was spun down, washed with 70% ethanol and resuspended 400 µL SDS-EB buffer (20% SDS,400 mM NaCl, 40 mM EDTA, 100 mM Tris-HCl, pH 8.0). After mixing well, an equal volume of phenol-chloroform, was added and then spun at 10K for 10 min at 4°C. The aqueous layer was extracted with equal volumes of 24:1 mixture of chloroform-isoamyl alcohol. The supernatent was mixed with 1/10^{th} volume of 3M sodium acetate pH 4.5 & 2 volumes of ethanol and incubated at -20^oC for ½ hour. DNA was collected at 15K for 15 min and dissolved in sterile water. DNA concentration was ascertained with OD_{260} and on agarose gel.

2.2.2.3 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used for the qualitative and quantitative analysis of DNA. 0.5µg/ml of ethidium bromide was added to molten 1 % low EEO agarose and the

solidified gel was used to fractionate DNA .The samples were mixed with gel loading dye solution and loaded onto the agarose gel submerged in electrophoresis buffer. Electrophoresis was done at 11 V/cm. After the electrophoresis has completed, the DNA bands were visualized on UV transilluminator (UVP Model 3UVTM Transilluminator) at 254nm or with hand UV lamp (Model UVGL-58) for cloning experiments at 320 nm.

2.2.2.4 Urea-PAGE gel Electrophoresis

Urea-PAGE gels were run in a DNA sequencing unit. A set of rigorously cleaned matched glass plate set were coated with Sigmacote® silane solution. The plates were sanwhiched with 0.1mm spacers and sealed. 8M urea-14% PAGE sequencing gel was poured in this assembly. The gel was stored overnight at 4 °C for polymerization to set and was then run the next day. The wells were flushed with 0.5 X TBE and the gel was pre-run at 1000 volts 18 mA for 30 min. Then samples were added and the gel was run till dye front reached 2/3 of the gel. All radioactive buffer was dispensed carefully and DNA Sequencing Unit was carefully washed. The gel was blotted onto Whatman Blotting sheet and wrapped with Saranwrap. Then this assembly was sandwiched with an X-ray film in a X-ray film cassette and the autoradiogram was left to develop in -20 °C. The X-ray films were developed in an automated X-ray film processor (Velopex Extra).

8M Urea-14% PAGE Sequencing Gel

Urea	25 gms
40 % acrylamide stock	20 ml

5X TBE	6 ml
10% Ammonium persulfate	1 ml
Formamide	6 ml

Mix everything and then make up volume to 60 ml. Filter through whatman 541 paper. Then degas for 10 minutes, add 30 µl TEMED and pour the gel.

2.2.2.5 Electrophorectic Mobility Shift Assay

Electrophoretic Mobility Shift Assay was done to study the DNA binding characteristics of DRA0282. A 5' 32P end labelled 200 bp double stranded DNA was used as a probe. 100 ng of DRA0282 was incubated with 10 pmol of the probe for 30 min on ice. In separate experiments this protein-substrate complex was supplemented with 100 fold excess concentration of different forms of DNA, incubated for 30 min at 37 °C and these complexes were mixed with 50% glycerol and loaded on a 5 % native polyacrylamide gel. The gel was run at 150 V for 2h at room temperature, then dried and exposed for 4-16 h to X-ray film at –80 C and the X-ray films were developed in a automated X-ray film processor (Velopex Extra).

5% Native Polyacrylamide Gel

milliQ-water	9 ml
5X TBE	3 ml
30% Acrylamide	3 ml
10% Ammonium persulfate	10 µL
50% glycerol	2.5 ml

TEMED 10 µL

Combine the ingredients for a final volume of 15 ml and pour between a 1.5 mm hollow 13 cm X 14 cm glass plate sandwhich and allow the gel to set.

2.2.2.6 PCR Amplification

Primers were designed manually using the corresponding nucleotide sequences from *D. radiodurans* genomes database. FASTPCR, a freely available software http://primerdigital.com/fastpcr.html was used to design the primers. Relevant factors like Restriction Endonuclease sites, GC content, primer dimer formation etc, were optimized with the help of menus in FASTPCR. HPSF® purified primers were obtained commercially from Metabion. PCR amplification was mostly carried out in 50µl volume. A typical PCR reaction contains template (50ng), primers 10 pmol, dNTPs 200µM, Taq DNA polymerase buffer 1X, 3.0 mM $MgCl_2$ and Taq DNA polymerase 2.5U. Dimethyl Sulfoxide (DMSO) was mixed as a denaturant in the PCR mix while amplifying with the high GC *D. radiodurans* genomic DNA. The PCR mix was amplified on Eppendorf Mastercycler® thermocycler for 30 cycles. The results were visualized on 1 % agarose gel with ethidium bromide. After PCR, when required, the unused components were removed by Roche PCR purification kit as per manufacturer's instruction.

2.2.2.7 Restriction Digestion

The protocols followed for restriction digestion with a particular enzyme were largely as described by manufacturers. The ratio of enzyme to DNA was kept less than 10 and the concentration of glycerol in reaction mixture was maintained to less than 10%. The double digestion of DNA with two different restriction endonucleases requiring same

buffer was carried out by adding both the enzymes and incubating for 3-5 hour. Double digestion with restriction endonucleases, which have incompatibility in reaction condition, was carried out one after another. An aliquot digests with individual enzymes along with the double digest was taken and analysed on agarose gels.

2.2.2.8 Ligation

Ligations were done with commercially available Quick Ligase Kits either from Roche or Bangalore Genei. A typical ligation reaction mixture consist of 100ng of gel purified insert and vector to give 3:1 molar ratio with 1X ligation buffer and 1 Weiss unit T4 DNA ligase in 10 µl reaction and incubated for 30 minutes.

2.2.2.9 DNA End Labelling

Synthetic oligonucleotides were radiolabelled as per following protocol:

5' End Labeling	vol	**3' End Labelling**	vol(µL)
DNA 10 pmol/µL	5	DNA 10 pmol/µL	5
PNK buffer	5	TdT buffer	5
milli Q water	28	milli Q water	23
[32p] γ-ATP ,3.3 µCi/pmol ATP	2	2.5 M $CoCl_2$	5
Poly Nucleotide Kinase	2	[32p] α-ATP, 3.3 µCi/pmol	2
		Terminal deoxynucleotidyl transferase	2

After mixing the above ingredients the mixture was incubated at 37°C for 1 hour, then heated at 70°C for 10 min and the radiolabelled probe was purified by illustra™ Microspin™ G-25 columns.

2.2.2.10 RNA Preparation

Effect of DR2417 was evaluated using total RNA purified from *E. coli*. The glassware and water was DEPC treated. Total RNA was extracted from *E. coli* by TRI –Reagent and then extracted twice with water saturated phenol:Chloroform. This RNA was then dephosphorylated by treating with calf intestinal alkaline phosphatase and again extracted twice with water saturated phenol:Chloroform. The RNA concentration was estimated to be 124 nM by UV spectroscopy. This RNA was then radio-labelled with $\gamma\ ^{32}$P ATP

2.2.2.11 ATPase Assay

We checked possible ATPase activity of DRA0282 by adopting the protocol of Buxbaum *et al*.,1990 [80]. The reaction was setup by incubating 200 ng of DRA0282 with 6.6 pmol [32P] γ ATP in its buffer, either in the presence or absence of 500 ng dsDNA in a total volume of 100 µL and incubating these at 37°C for 30 minutes. A control without any DRA0282 was setup in parallel. After incubation 50 µL ice cold Stop Solution (5 % TCA, 0.14% ammonium heptamolydate, 3% sulphuric acid, 1mM orthophosphoric) was added. The 100 µL of icecold 1:1 mixture of isobutanol/ethylacetate was added and mixed by vortexing. Then this was centrifuged at 10K for 2 minutes and 70 µL was

withdrawn into 10 ml scintillation cocktail. Then radioactivity was measured in a Packard Tri-Carb 2000 TR Liquid Scintillation Counter.

Different vectors used in this study

Recombinant vectors for expression of recombinant proteins in *E. coli* and for generation of knockouts in *D. radiodurans* R1 are provided below (Table 2.2.2.1):

2.2.2.1 List of plamids used

Sr No.	**Name of the plasmid**	**Characteristics**	**Source**
1	pBluescript SK+	Amplicilin, 2.930kb, colE1	Stratagene
2	pET28a+	Kanamycin, 5.320kb, colE1	Novagene
3	pRADgro	The *groESL* promoter cloned in pRAD1 , 6.885kb	The *groESL* promoter cloned in pRAD1 , 6.885kb
4	pNOKOUT	900bp nptII in pBluescript SK+, 3.9kb	This study
5	pTRC99a	Ampicillin,4.19 kb, colE1	Amersham Pharmacia
6	pTRC282	drA0282 cloned in pTRC99a	This study
7	pTRC2417	dr2417 cloned in pTRC99a	This study
8	pNOK282	nptII flanked by 1 kb upstream and downstream	This study

		sequence of drA0282	
9	pNOK2417	nptII flanked by 1 kb upstream and downstream sequence of dr2417	This study
10	pET2147	dr2417 cloned in pET28	This study
11	pET282	drA0282 cloned in pET28	This study

2.2.3. Bacterial Transformation

2.2.3.1 Transformation in Escherichia coli

In *E. coli*, competent cells were prepared using the standard protocol. In brief, the overnight grown culture of desired *E. coli* strain was diluted to 100 fold in fresh LB medium. The culture was allowed to grow at 37^{o}C till OD_{600} reaches to 0.3 to 0.4 and thereafter the cells were chilled on ice for 30 min. The culture was transferred to pre-chilled SS34 tubes and centrifuged at 6k for 5 min at 4^{o}C. The pellet was gently suspended in ½ culture volume of 100mM $CaCl_2$ and suspension was incubated on ice for 30 min. Thereafter it was centrifuged at 6000 rpm for 10 min and competent cells were gently suspended in 0.1 culture volume of 100mM $CaCl_2$. 100µl of the competent cells were aliquoted in 1.5 ml pre-chilled tubes and to it approximately 100ng of the plasmid DNA was added and mixed by tapping. The mixture was incubated on ice for 30 to 45 min and heat shock treated at 42^{o}C for 2 min followed by 5 min on ice incubation. The

transformation mixture was diluted with 800µl of 1x LB broth. For the expression of antibiotic resistance genes, the mixture was incubated at 37^{o}C for 20 min. The transformation mixture was plated on LB agar plates containing appropriate antibiotics. The plates were incubated at 37^{o}C for overnight and transformants were scored and characterized.

2.2.3.2 Transformation in Deinococcus radiodurans

20 ml of 2X TYG in 150 ml flask was inoculated with 1/50 dilution of overnight grown *D. radiodurans* culture. This was then incubated at 32 °C for 2 hours. Then $CaCl_2$ was added at a final concentration of 50 mM and the culture was further grown for 30 min. Then the culture was spun down and resuspended in 1/4th volume 2X TYG with 50 mM $CaCl_2$. Then DNA was added to 1ml aliquot of cell suspension and this was further incubated on ice for 30 min. then the culture was shifted at 32 C and incubated under rotation for 1 hr. Then 5 ml TYG was added and this was grown overnight at 32 °C and then was plated on TYG plates with appropriate antibiotics

2.2.4 Methods used in Protein Purification

2.2.4.1 Inducible Expression of Genes in *Escherichia coli*

E. coli strain BL21 DE(3) pLys*S* harboring recombinant pET vectors were grown overnight in the presence of kanamycin (25µg/ml). The cells were diluted in fresh LB

medium containing antiboitics and allowed to grow for 3-4 h to get density OD600 0.3 and then IPTG (0.5mM) was added and growth was continued for a further period of 5h. A small aliquot (200µl) was drawn for checking for the inducible synthesis of protein and remaining cells were harvested, washed and preserved at –70^{o}C for downstream processing. The samples were analyzed for induction of proteins along with vector controls by SDS-PAGE.

2.2.4.2 SDS-PAGE Analysis of Proteins

Polyacrylamide gel provides a matrix for electrophorectic seperation of polypeptides. SDS binds fairly specifically in a mass ratio of 1.4:1 and confers a negative charge to the polypeptide. In denaturing SDS-PAGE therefore, migration is determined by the molecular weight.

2.2.4.3.1 Preparation of SDS-PAGE gel

Ordinary SDS-PAGE gel with uniform 10 % Acrylamide concentration was made as below.

Composition of 10 % separating polyacrylamide gel

Acrylamide (30%)	5.3 ml
1.5M Tris-HCl, pH-8.8	5.0 ml
10 %SDS	200 µl

10 % ammonium persulfate	200 µl
TEMED	20 µl
Water	3.4 ml
Total volume	20 ml

Composition of 5% stacking gel

Acrylamide (30%)	0.83ml
1.0 M Tris-HCl, pH 6.8	0.63ml
10 % SDS	50 µl
10 % ammonium persulfate	50 µl
TEMED	15 µl
Water	3.4 ml
Total volume	5.0 ml

2.2.4.3.2 Procedure of pouring and running the gel

Glass plates were cleaned with liquid soap, rinsed with water and wiped with 70% ethanol. The plate sandwich was assembled with 1.0 mm spacers on the stand. Gaps between plate and stand were sealed with agar and SDS-PAGE mixture for separating gel was poured. It was over layered with water-saturated n-butanol. After 20-30 min when gel is polymerised, the butanol was removed and washed with water. Stacking mixture was poured and comb was fitted in the gel. Once the stacking gel is polymerised, the

comb is removed and wells were cleaned with water 3-4 times and then samples were loaded. The samples were prepared by heating equal volume of sample and 2X SDS gel loading dye at 95^{o}C. The denatured samples were spun at 12,000 RPM for 10 min and clear supernatant was loaded in the gel. The electrophoresis tank was filled with electrophoresis buffer (10% SDS-Tris-Glycine, pH 8.8) and gel was run for 1 hour at 50 volts and then at 100 volts for 2 hour. The gel was then processed as needed.

2.2.4.3.3 Colloidal Coomassie Brilliant blue staining

Coomassie blue (CBB) dye binds with the basic amino acids of the polypeptide. This is an irreversible method for protein staining. The detection limit of this method is 100 ng. We hence adopted a more sensitive Collodial Coomassie Blue Staining with a lower detection limit of 30 ng. Also, this method dispenses with frequent solution change. After running the gel, it is briefly rinsed with water for 1 minute and then immersed in the dye O/N. Next day the gel is briefly rinsed in water for 1 min and visualized.

Composition of Collodial coomassie blue stain

Component	Final Concentration	Amount
D/W		650 ml
Ammonium Sulphate	10 %	100 gm
Coomassie G250	0.1 %	20 ml of 5 % solution in water

Ortho-phosphoric Acid	3 %	30 ml
Ethanol	20 %	200 ml
	final volume	1000 ml

Add the ingredients in the order listed. Make sure the ammonium sulphate is dissolved before adding the coomassie G-250 solution to a final of 0.1 %. Bring up the final volume upto 1 Ltr with water. Store tightly sealed at RT.

2.2.4.4.1 Preparation of cell free extract of E. coli

The cell pellet was thawed on ice and 1 gm pellet was suspended in 2 ml lysis buffer A (20mM Tris-HCl, pH 7.5; 0.1mM EDTA and 1mM DTT). Lysozyme was added to a final concentration 200μg/ml and incubated on ice for 20 min. To it, Triton X-100 was added to a final concentration of 1% and sonicated at 50% duty cycle, for 3 min with the 1 min intermittent cooling on ice. Suspension was diluted to two folds and centrifuged at 12k (SS34, Sorvall) for 15 min.

2.2.4.4.2 Preparation of cell free extract of D. radiodurans

Cell pellet was resuspended in 2 ml of suspension buffer (20mM Tris-Cl pH-7.6, 50mM NaCl, 1mM EDTA, 0.5mM DTT, and 0.5mM PMSF). To this lysozyme was added to a final concentration of 2 mg/ml and incubated at 37^{o}C for 30 min. The cell lysate was subjected to freeze-thaw cycles in liquid nitrogen. After this, cells were sonicated using

30 sec. pulse cycles with intermittent cooling for 1 min each, at 40% duty cycle for total time of 10 min. Cell lysate was centrifuged at 12,000g for 20 min. and clear cell free extract was used as soluble protein fraction.

2.2.4.5 Protein Estimation

Proteins were routinely assayed by the Bradford protein estimation method [81] . In solution the reduced form of Coomassie G-250 binds to proteins in a quantitative way and this can be assayed in a spectrophotometer at 595 nm. In brief, 200µl of the 5X dye solution was mixed with 800µl of total solution containing proteins and buffer. The content was mixed by vortexing and left at RT for 15 min before color development was monitored spectrophotometrically at 595nm. The OD_{595} was compared with a calibrated curve that was made using standard concentration of BSA.

2.2.4.6 Metal Affinity Purification of Recombinant Proteins

Immobilized Metal ion affinity Chromatography exploits molecules affinity for cheated metal ions. The amino acid histidine present in the his tagged protein forms complexes with transition metal ions such as Ni_2+ ,Cu_2+, Zn_2+ and Fe_3+. Initial efforts were made to purify soluble DRA0282 and DR2417 from recombinant *E. coli* cell extracts. However, since suitable yeilds were not obtained we purified both the proteins under denaturing conditions.

2.2.4.6.1 Purification in denaturing conditions

The recombinant protein was purified using a standard kit protocol. In brief, the cell contents were solubilized in buffer containing 100mM NaH_2PO_4, 10mM Tris-Cl and 8M urea, pH8.0 (binding buffer). The lysate was centrifuged at 10,000 RPM for 30 min at

room temperature to pellet the cellular debris. 1 ml of the 50% Ni-NTA was loaded onto a column. The cell lysate was diluted 10 fold with binding buffer and loaded onto the column and allowed to flow slowly. The flow through obtained was again passed through the column. An aliquot was collected and stored for SDS PAGE analysis. The column was washed twice with 4 ml of Buffer C (100mM NaH_2PO_4, 10mM Tris-Cl, 8M urea, pH 6.3). The protein fractions each of 500µl were eluted with buffer D (100 mM NaH_2PO_4, 10 mM Tris-Cl, 8 M Urea, pH-5.9) followed by buffer E (100mM NaH_2PO_4, 10 mM Tris-Cl, 8 M urea, pH 4.5). The collected fractions as well as flow through were analyzed by SDS PAGE. The elution fraction which showed pure protein were refolded by serially dialyzing the protein in Refolding Buffer with decreasing concentration of urea. The protein samples then were stored in Protein Storage Buffer.

2.2.4.7 Western Blot

Western Blot Transfer Buffer

25 mM Tris-HCl

192 mM Glycine

20 % Methanol pH 8.3

10X Tris Buffered Saline (TBS)

To prepare 1Ltr 10X TBS: 24.2 g Tris base, 80 g NaCl; adjust pH to 7.6 with HCl (use at 1X)

TBS-T for Western Blot

To 1 Ltr of TBS add 1 ml Tween 20

Blocking solution for Western Blot

Nonfat dried milk 3% ; sodium azide 0.02%

Add the above in 1X TBS.

Western blots were done to detect the production of recombinant proteins. The samples were run on a SDS-PAGE gel along with a pre-stained marker. A PVDF membrane of the same size as the SDS-PAGE was equilibrated with methanol for 15 minutes and then rinsed with water. After running the gel, it was briefly washed and the gel, the membrane, and the extra-thick blotting paper were equilibrated with the transfer buffer for 15 minutes. The PVDF membrane was marked with pencil for orientation and then the gel was placed on top of the PVDF membrane and this was sandwhiched between the blotting sheets. Then this assembly was placed inside a Biorad Trans Blot Semi-Dry Transfer Cell Apparatus. Then blotting was done for 45 minutes at 100 mA and progress was monitored by the rise in voltage. After the blot, the membrane was immersed in 3% skim milk blocking buffer for 30 minutes and then rinsed once with TBS-T. Then the membrane was incubated with the primary antibody in TBS-T in 1:10,000 dilution with gentle shaking for 2 hrs at RT. Then the membrane was rinsed thrice with TBS-T and the AP conjugated secondary antibody in TBS-T was added. This was again incubated similarly for 1 hour. The blots were developed with NBT/BCIP to detect the bands.

Chapter 3

RESULTS

3.1 Bioinformatics analysis of selected hypothetical proteins identified from the pool of DNA binding proteins exhibiting ATP sensitive nucleolytic function

Heparin binding proteins isolated from *D. radiodurans* cells collected at 1 hour Post Irradiation Recovery (PIR) phase showed an ATP inhibited nulceolytic degradation on double stranded DNA substrate [82]. Since the cells were in the Extended Synthesis Dependent Strand Annealing (ESDSA) phase of repair, which is a unique repair process discovered in *D. radiodurans* and nucleases play an important role in ESDSA, we wanted to find the identity any putative nuclease and the reasons for its inhibition by ATP. When this fraction of heparin binding proteins was analyzed by Fourier Transform Ion Cyclotron Mass Spectroscopy (FTICR-MS), it detected over 300 proteins in this fraction [83]. In this mass spectroscopy analysis no annotated nuclease or kinase was found but a set of 11 hypothetical proteins in the fraction were also found. Hypothetical or putative proteins make up roughly 50% of ORFs in *D. radiodurans* and have been implicated in the extraordinary resistance of *D. radiodurans*. We worked on the premise that this novel activity could be amongst those 6 hypothetical proteins. Although there are broad guidelines, there is no defined bioinformatics protocol for finding the function of hypothetical proteins and hardly any work of this nature was documented for *D. radiodurans* before. Most nucleases differ in sequence but they do contain conserved minimal motifs, which usually consist of acidic and basic residues forming the active site

and they can be neatly arranged in 8 fold families [84]. We hence tried to use the known bioinformatics tools to find possible nuclease functions for these hypothetical proteins based on the above information, both from the practical standpoint of finding a novel nuclease as well as test whether these tools also can also successfully predict function of these hitherto uncharacterized proteins from *D. radiodurans.* We have used a phylogenomic approach to study these proteins. In phylogenomics, the protein of interest is clustered with its homologs from known species based on a distance tree (phylogenetic tree). Since function diverges with speciation, usually the protein of interest clusters closely with its functional homologs. This approach has been sucessfully used to attribute functions to unknown genes [76] before and hence we decided to follow this approach. This chapter describes our efforts predicting the possible functions for the above hypothetical proteins.

3.1.1 Methods

The subcellular location of the protein was checked in pSORTb server [85]. Additionally the structural features of the protein sequence were highlighted by Hydrophobic Cluster Analysis [86]. This analysis is based on the fact that hydrophobic residues in a protein are clustered together and when the sequence is mapped over a helix these clusters are highlighted. Thus unusual amino acid distributions and hydrophobic patches can be easily identified especially for pseudogenes and frameshifts which have unusual hydrophobic cluster plots. PSI-BLAST (Position Specific Iteration BLAST) was done with the RefSeq

database at NCBI for each of the sequences [69]. Entrez keywords used were “DNA repair”, “nuclease”, “kinase” and sequences with e-value higher than E-value cutoff of 10^{-3}were visually screened and subjected to another round of iteration. This was done till 20-25 sequences were obtained or till convergence, whichever came earlier. The sequences obtained were aligned in ClustalX [87] along with the seed alignments obtained from PFAM database [88] and conserved sites usually corresponding to metal binding sites or ligand binding sites were seen. Also to improve the alignment, the alignments were superimposed with theoretically calculated secondary structure data. This was done to ensure that gaps would be forced in the alignment building process where only loops were present. A phylogenetic distance tree using Neighbor Joining algorithm was built from this alignment. For statistical accuracy this tree was bootstrapped 1000 times. All phylogenomic inferences were made from this bootstrapped tree. Additionally, the fasta formatted sequence for each individual sequence under study was submitted to I-TASSER website [89] for obtaining the homology model of the protein. Where I-TASSER failed to provide a structural model of resolution 4° A or less, the protein sequence was aligned to a sequence whose structural data was available and this alignment was submitted to SWISSMODEL [90] in the alignment mode. All structural models obtained were checked for accuracy by seeing that maximum residues were present in the allowed region of the Ramchandran Plot. Since our aim was limited to studying conservation at fold level, we did not energy minimize the structures.

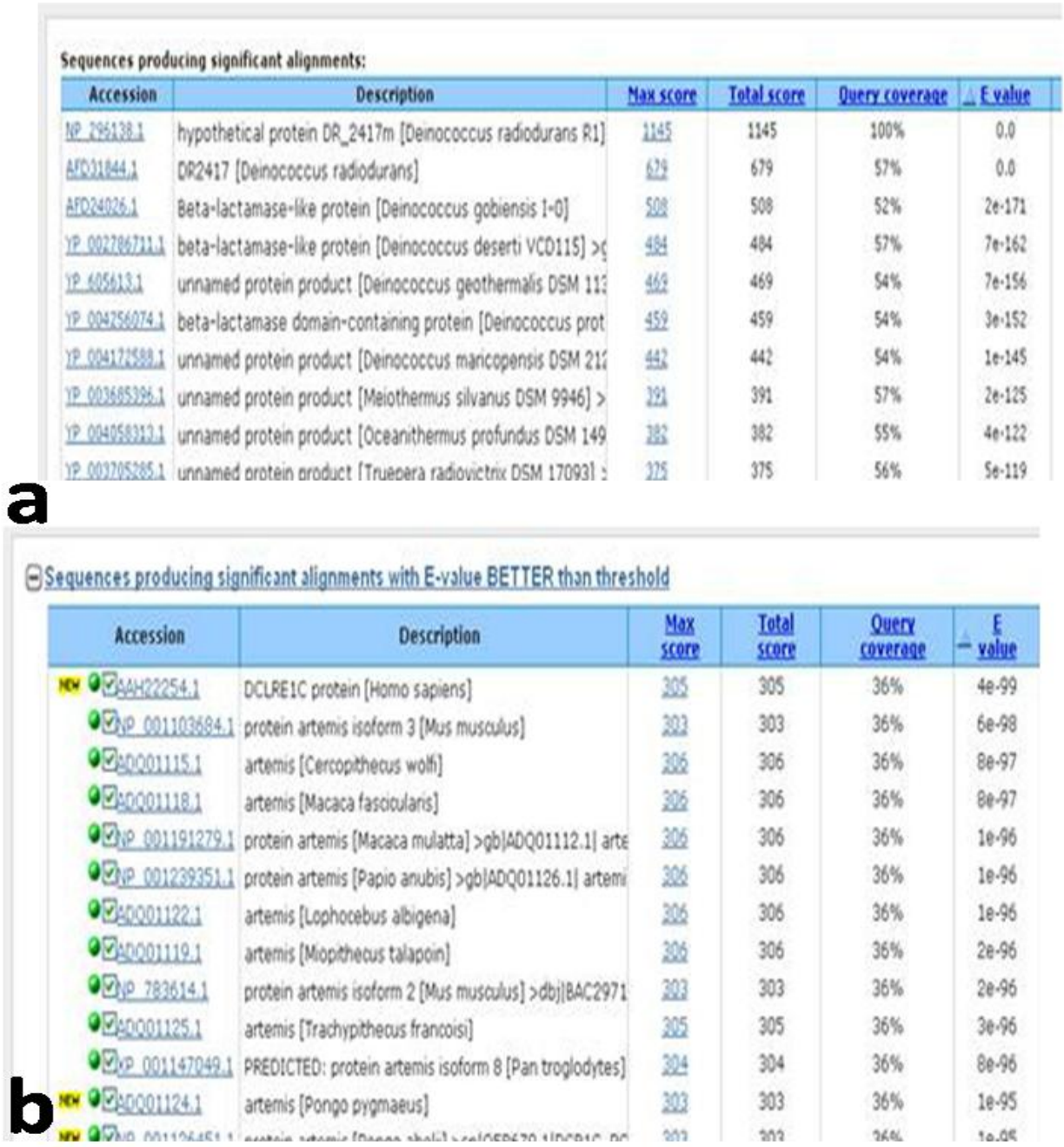
3.1.2 Results

3.1.2.1 BLAST Search

By doing a default BLAST search for the 11 hypothetical proteins, we mostly retrieved other hypothetical proteins. Since these proteins were found in a PIR fraction showing nuclease activity, we did successive BLAST searches with the Entrez key words “DNA repair”, “nuclease”, “exonuclease”, “ATP”, “ATP inhibited”, “kinase” and found that 6 of those proteins had matches with DNA metabolic proteins. FTICR-MS data of this 1 hour PIR cell lysate fraction detected 84 peptides of DRA0282, 45 peptides of DR0390 , 36 peptides of DR2417 , 39 peptides of DR1654 while it detected only 5 peptides of DR0041 and 4 peptides of DR0461. Since the number of peptides for DR0041 and DR0461 were very low, we focussed on initially trying to predict the function of DRA0282, DR2417, DR0390 and DR1654.

As BLASTP (BLAST for proteins) did not retrieve functionally characterized proteins, we used the Position Specific Iterated BLAST (PSI-BLAST) algorithm. In PSI-BLAST, we queried the RefSeq database at a cutoff e-value of 0.001 with the Entrez keywords used earlier and examined the matches retrieved. We chose alignments of the query and subject obtained from the first search as ‘significant matches’ usually based on clustering of histidine or acidic or basic residues, an e-value lower than our cutoff and alignments with few gaps. These alignments were iterated for several rounds. We found functional homologs of DR2417 like Artemis, Pso2, Apollo which are known to be involved in DSB repair etc by doing these iterations. Thus by using PSI-BLAST, the sensitivity of our

search improved and as shown in the figure below (Fig 3.1.2.1a & 3.1.2.1b), this was not found in BLAST. This improvement was also seen for the other ORFs as shown below (Table 3.1.2.1). A Multiple Sequence Alignment (MSA) was built with the protein sequences obtained from the BLAST search in ClustalX and we tried to find conserved motifs amongst the proteins (Table 3.1.2.1). After obtaining this data, the phylogenetic relationship of proteins in the MSA are obtained. The hypothetical protein is assigned the function of the proteins of the sub branch in which the hypothetical protein clusters. Then the conserved residues and mutational changes can be crosschecked with the computed structural model of the hypothetical protein. The best results were obtained when the conserved sites in the MSA mapped at known active sites or ligand binding site or protein-protein interaction sites in the structure.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value
NP_296138.1	hypothetical protein DR_2417m [Deinococcus radiodurans R1]	1145	1145	100%	0.0
AFD31844.1	DR2417 [Deinococcus radiodurans]	679	679	57%	0.0
AFD24026.1	Beta-lactamase-like protein [Deinococcus gobiensis I-0]	508	508	52%	2e-171
YP_002786711.1	beta-lactamase-like protein [Deinococcus deserti VCD115] >g	484	484	57%	7e-162
YP_605613.1	unnamed protein product [Deinococcus geothermalis DSM 11	469	469	54%	7e-156
YP_004256074.1	beta-lactamase domain-containing protein [Deinococcus prot	459	459	54%	3e-152
YP_004172588.1	unnamed protein product [Deinococcus maricopensis DSM 21	442	442	54%	1e-145
YP_003685396.1	unnamed protein product [Meiothermus silvanus DSM 9946] >	391	391	57%	2e-125
YP_004058313.1	unnamed protein product [Oceanithermus profundus DSM 149	382	382	55%	4e-122
YP_003705285.1	unnamed protein product [Truepera radiovictrix DSM 17093] >	375	375	56%	5e-119

a

Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E value
AAH22254.1	DCLRE1C protein [Homo sapiens]	305	305	36%	4e-99
NP_001103684.1	protein artemis isoform 3 [Mus musculus]	303	303	36%	6e-98
ADQ01115.1	artemis [Cercopithecus wolfi]	306	306	36%	8e-97
ADQ01118.1	artemis [Macaca fascicularis]	306	306	36%	8e-97
NP_001191279.1	protein artemis [Macaca mulatta] >gb|ADQ01112.1| arte	306	306	36%	1e-96
NP_001239351.1	protein artemis [Papio anubis] >gb|ADQ01126.1| artemi	306	306	36%	1e-96
ADQ01122.1	artemis [Lophocebus albigena]	306	306	36%	1e-96
ADQ01119.1	artemis [Miopithecus talapoin]	306	306	36%	2e-96
NP_783614.1	protein artemis isoform 2 [Mus musculus] >dbj|BAC2971	303	303	36%	2e-96
ADQ01125.1	artemis [Trachypithecus francoisi]	305	305	36%	3e-96
XP_001147049.1	PREDICTED: protein artemis isoform 8 [Pan troglodytes]	304	304	36%	8e-96
ADQ01124.1	artemis [Pongo pygmaeus]	303	303	36%	1e-95

b

Fig 3.1.2.1. BLAST search output. Here we can see the improvement in retrieving well characterized protein matches to a query hypothetical protein DR2417. **(a)** Results from BLASTP search showing mostly other hypothetical proteins **(b)** Results from PSI-BLAST search showing eukaryotic DNA repair nuclease Artemis.

Table 3.1.2.1 Difference between output of BLAST and PSI-BLAST result for the ORFs

ORF	BLAST	PSI-BLAST	Conserved motifs found after sequence alignment with results of PSI-BLAST
DR 2417m (64 kD)	Hypothetical protein	Artemis, PSO2, Apollo	D, H-X-H-X-DH (nuclease), H, VL-X-[IS]-GD (lactamase) E, SGH, [HV]-GE (CASP)
DRA 0282 (55 kD)	Hypothetical protein	Ku80 and PRP19	HD-X-DL-X-P (ATP binding) AYRL-X-DP (DNA Ligase) fibronectin type III domain
DR 0390 (59 kD)	Hypothetical protein	DAK kinase	DAK 2 domain/DegV domain
DR 1654 (19 kD)	Hypothetical protein	helicase	nil

3.1.2.2 Analysis for DR2417

Initial PSI-BLAST analysis showed DR2417m has a match with a β-CASP family nuclease- Artemis. The β-CASP family of nucleases are involved in diverse functions like 3' end processing of RNA precursors, Non Homologous End Joining (NHEJ) repair pathway, V(D)J recombination etc. These proteins have several key conserved motifs in

their polypeptide sequence, where motifs 1, 2, 3 and 4 belong to the catalytic metallo-β lactamase domain and motifs A, B and C belong to the substrate specificity determining CASP domain. Artemis has been shown to play a role in DNA repair in Mammals. Mice having a null mutation in *snm1* gene which codes for Artemis develop Severe Combined Immunodeficiency (SCID) because lymphocytes in these mice cannot undergo V-D-J recombination during their maturation [91] . V-D-J recombination involves rejoining of hairpin loop DNA substrate by the NHEJ apparatus. In V-D-J recombination, asymmetric double strand breaks are created by the RAG1 and RAG2 proteins in the linking region between the V, D and J exons and this creates a hairpin loop. In *Snm1-/-* mice, the V-D-J exons cannot be religated as there is no Artemis to cleave the hairpin loop at the ends for joining by ligase(s). Artemis has a 5'-3' single stranded DNA exonuclease activity which gets modified to a hairpin endonuclease after it is phosphorylated by DNA-PK protein complex and this endonuclease activity is indispensible in V-D-J recombination. In *D. radiodurans*, it is known that double strand breaks are formed after irradiation and repaired during its post irradiation recovery. Thus the retrieval of Artemis as match for DR2417 was relevant in this bacterium. . The *dr2417m* ORF was annotated as having a 'genuine frameshift' at position 996 of the ORF where an additional 'T' residue was present as compared to other deinococcus species. Sequence alignments showed that DR2417m had all the conserved motifs of the β-CASP family of nucleases (Fig 3.1.2.2 a) and the corresponding Hydrophobic Cluster Analysis plot showed an unusually arginine and proline rich C-terminus as per annotated sequence (Fig 3.1.2.2b). Importantly it had valine at motif C, the same as Artemis, which is a DNase while all other RNase members had histidine (Fig 3.1.2.2a) [92].

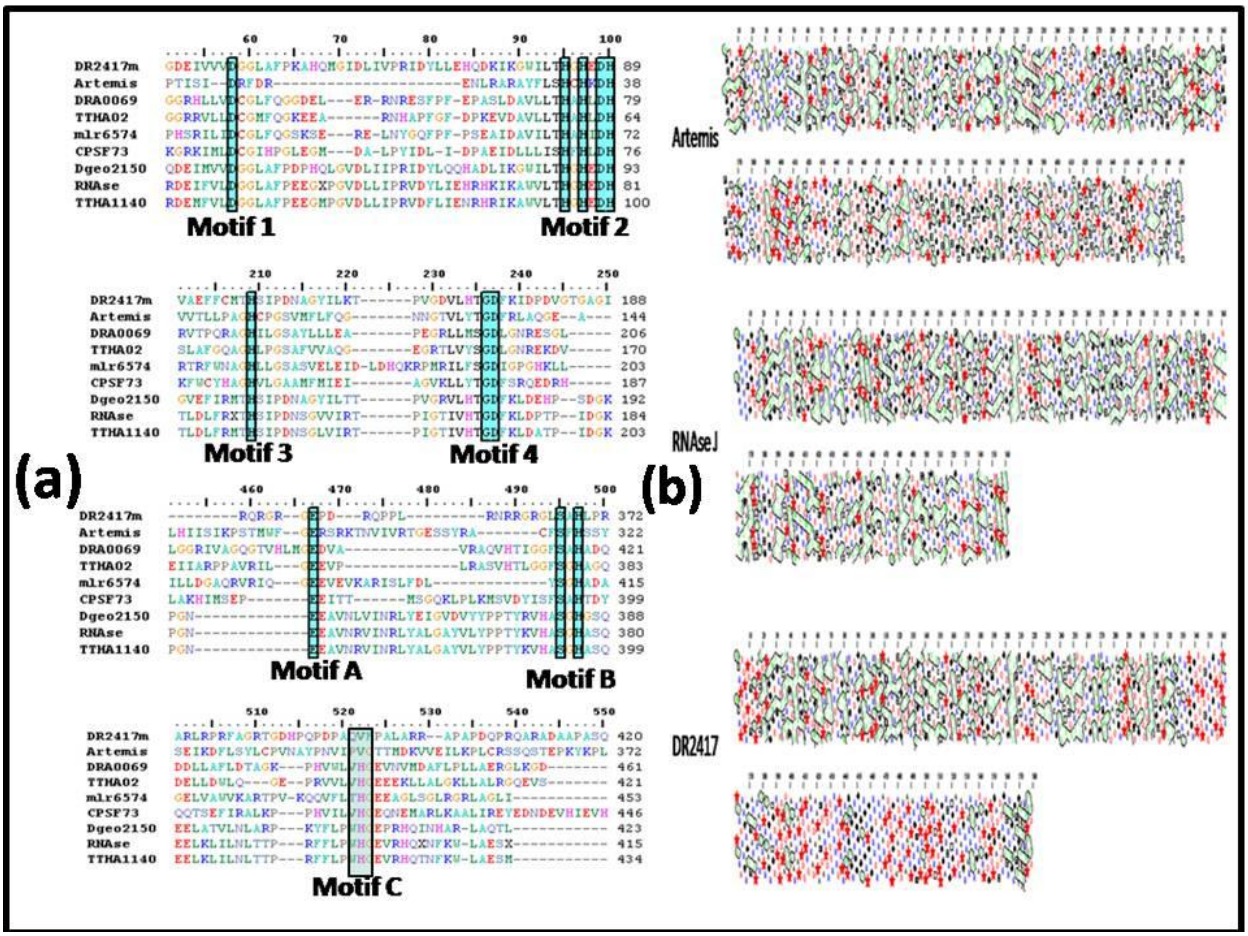


Fig 3.1.2.2. **(a)** multiple sequence alignment of DR2417m with functional homologs obtained by PSI-BLAST. The boxed regions are the conserved motifs. Motif 1,2,3 and 4 belong to metallo-β-lactamase domain while Motif A,B,C belong to the CASP domain. DR2417m has a valine at motif C while all RNases have histidine. **(b)** A Hydrophobic Cluster Analysis (HCA) plot of sequence of DR2417 and its closest homologs RNase J and Artemis shows that DR2417m has an unusual sequence profile rich in proline (red stars) and arginine (red R label). This profile is absent in both RNase J, an RNase and Artemis, a DNase

DR2417m had 56% identities with RNase J, an RNase from *Thermus thermophilus* while we speculated its role as a DNase. To better understand the possible function of DR2417 we built a homology model of DR2417m based on the crystal structure of RNase J (PDB ID 3T3N). The 3-D structure showed that DR2417m had a couple of β-sheets absent in the β-lactamase as well as the CASP domain, which might make it more flexible (Fig 3.1.2.3 a). The model indicated that DR2417 should exist at least as a dimer to act upon DNA, as the active site of the monomer could accommodate only a single strand (Fig 3.1.2.3b). The model also showed that possibly the DNA ends need to be frayed for the protein to accommodate them. The phylogenetic distance tree of DR2417m with its homologs showed that DR2417m is closer to Dgeo2150 from *D.geothermalis* but not its paralog DRA0069 and it was evolutionarily closer to RNase J from *Thermus* rather than the DNase Artemis (Fig 3.1.2.3c). However, since RNase J had a histidine at motif C, while DR2417m had a valine at that site and after all since Artemis is a eukaryotic protein which would have collected mutations at other sites other than motif C, we concluded that structural similarity apart, DR2417m should act as a DNase.

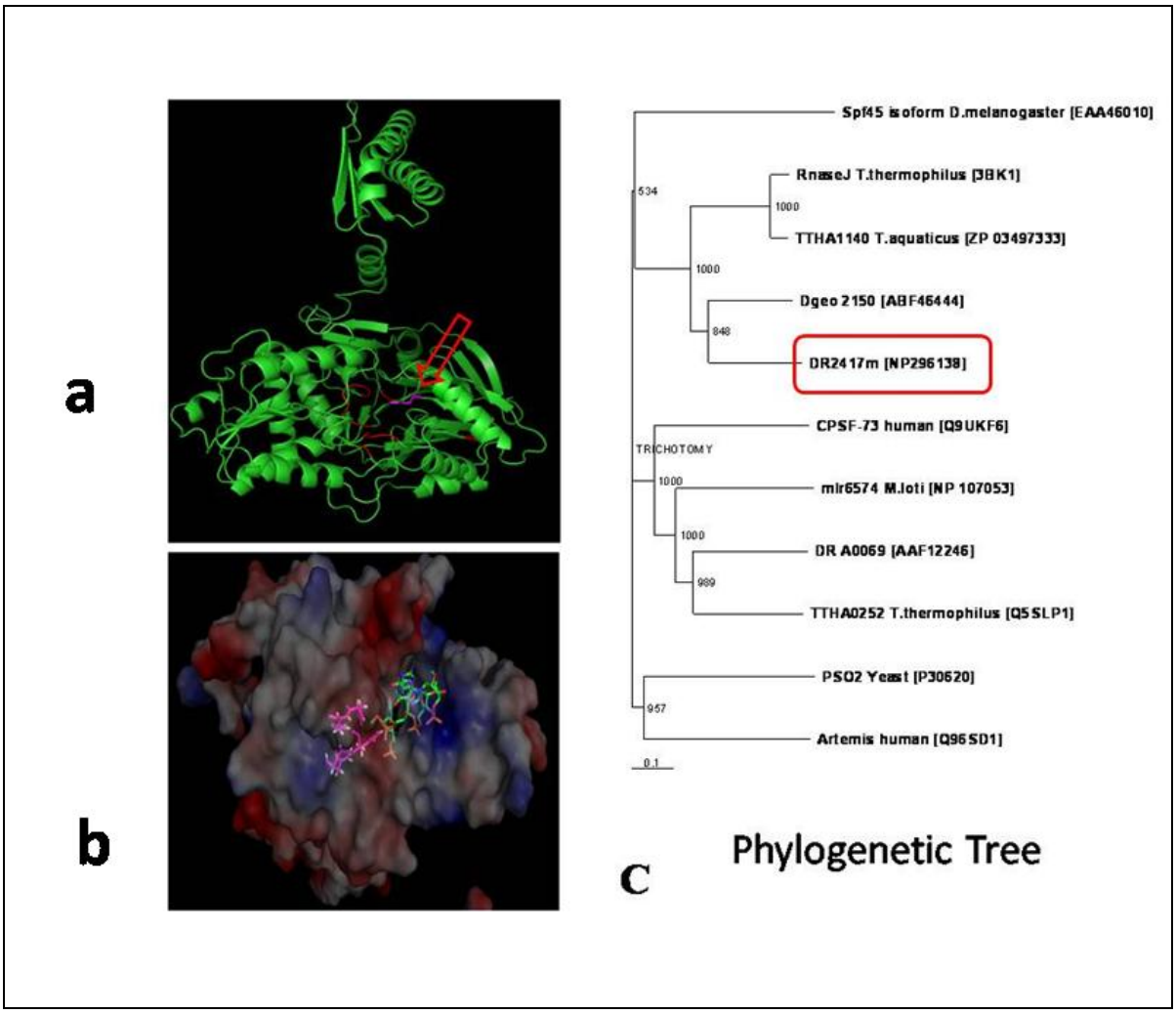


Fig 3.1.2.3. **(a)** The cartoon of a 3-D homology model of DR2417m based on RNase J template (3T3O). The arrow marks the valine residue close to the active site and the regions in red highlight regions where DR2417m differs from RNaseJ **(b)** A space filling plot of the active site shows the closeup of the active site. The single strand of nucleic acid (green colored stick model) sits on a positively charged channel colored in blue. The active site residues are colored pink. This shows that the active site has space only for one strand. **(c)** the bootstrapped phylogenetic tree shows that DR2417m clusters with RNase J

3.1.2.3 Analysis of DRA0282

We were interested in DRA0282 as it showed the maximum number of peptides (84 peptides) obtained from FTICR-MS, in the DNA binding proteins fraction exhibiting ATP sensitive nuclease activity. To find the function of DRA0282 we did a PSI-BLAST with DRA0282 in the RefSeq database with the keywords like 'DNA repair', 'nuclease' etc, as DRA0282 was found in a PIR fraction. Mostly Ku homologues were obtained as the nearest matches after 3-4 rounds of iteration. We aligned DRA0282 with these sequences for getting further insights in the possible functions of this protein. Ku protein is involved in repairing DSBs by the NHEJ pathway. The important feature of NHEJ is that a break is resealed without the need of extensive homology and Ku is a key player in this pathway and one of the earliest proteins to be recruited to DSBs in eukaryotic system [93]. In mammals, where it was first identified, Ku is a functional heteromer of two polypetides, Ku70 and Ku80, which have similar structure but share only 14% sequence identity.

They combine as Ku70/80 to form a ring structure which slides into the broken DNA ends and it holds the two broken ends like a molecular tape. Ku70/80 has a modular structure having an N-terminal VonWillebrand (vWA) domain which binds to nucleotides and is postulated to be involved in protein-protein interaction, a core domain which binds to broken DNA ends and C-terminal SAP domain which mediates protein-DNA interaction. Ku recognizes the broken ends and recruits DNA dependent protein kinase C (catalytic subunit) DNA-PKcs, which complexes with Ku and this complex, now termed as DNA-PK, then recruits the downstream end processing nucleases and ligase complex of XRCC4/DNA ligase IV to reseal the two ends.

This recruitment is done by site specific phosphorylation of the downstream factors by the DNA-PK complex [94]. Also, in most cases of double strand breaks, the broken ends are not exactly blunt and hence incompatible for ligase activity. This end processing is accomplished by Artemis (our match for DR2417). Recently Ku homologs have also been discovered in bacteria like YkoV from *Bacillus subtilis* and *Mt*-Ku from *Mycobacterium tuberculosis* [95]. These are smaller proteins as compared to the eukaryotic Ku and are not as extensively regulated as their eukaryotic counterparts. Besides, bacteriophage Mu Gam protein has also been shown to be homologous to Ku protein and the discovery of Ku in bacteria indicates that Ku is an ancient protein which has undergone species specific modification. Thus there was a possibility of a Ku type protein having overall structural similarity yet having a poor sequence similarity as structure is more conserved than sequence. Our Multiple Sequence Alignment showed that DRA0282 had an overall homology with Ku proteins (Fig 3.1.2.4).

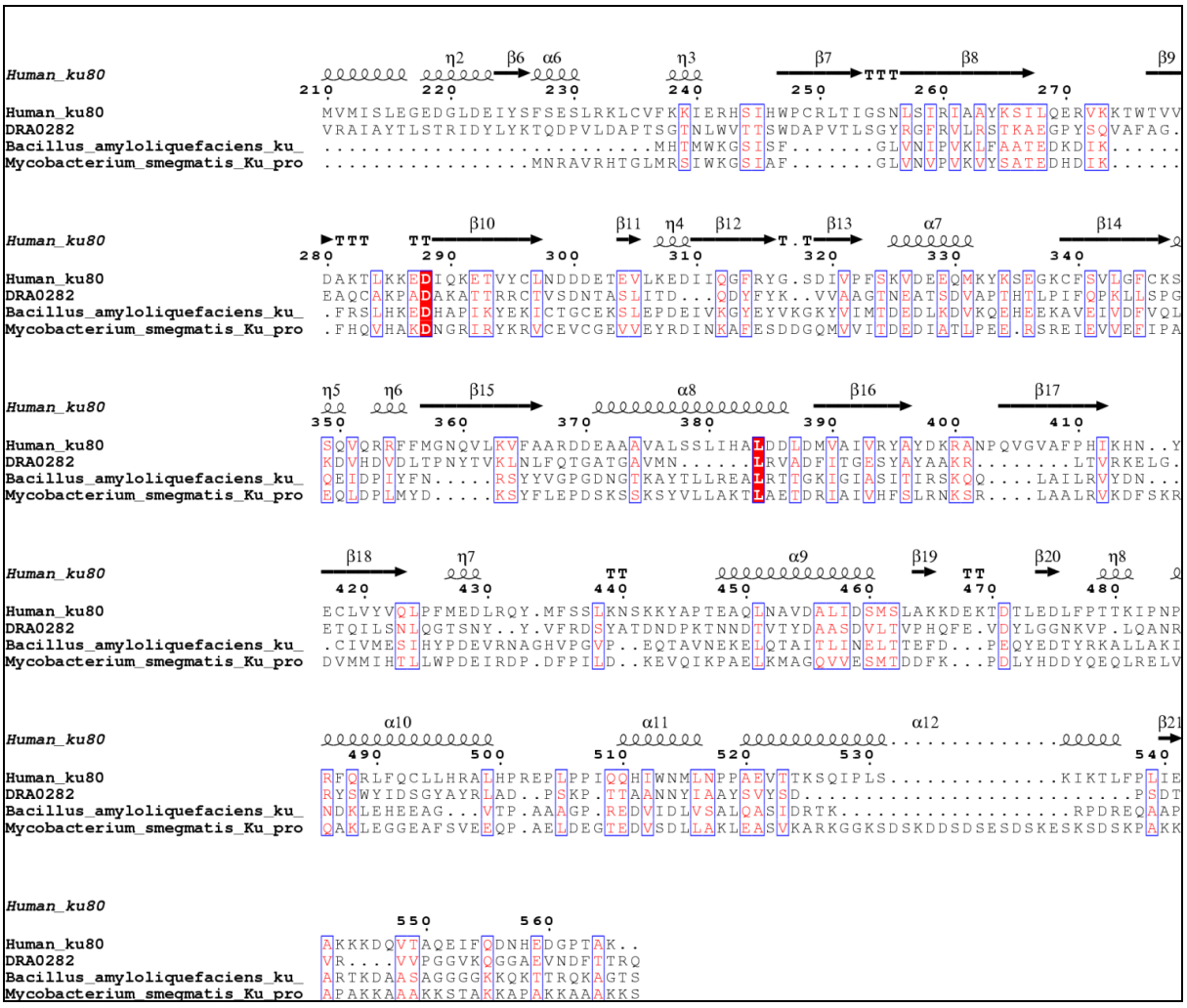


Fig 3.1.2.4. Multiple Sequence Alignment of DRA0282 with its homologs shows the conserved regions in the core domain. Structural studies have shown that the core domain binds to DNA

Since DNA binding property of any protein would depend upon its structure, we built a 3-D model of DRA0282 with the structure of Ku80 from Protein Data Bank (PDB code-1JEY). A model with RMSD of 4.0 was built. We observed that DRA0282 had a remarkable homology with Ku80 (Fig 3.1.2.5a). The N terminal Von Willebrand A

domain, which binds nucleotides and the DNA binding β-barrel domain were seen. Interestingly, in the phylogenetic tree DRA0282 showed a closer match with eukaryotic Ku, than with prokaryotic Ku proteins (Fig3.1.2.5b).

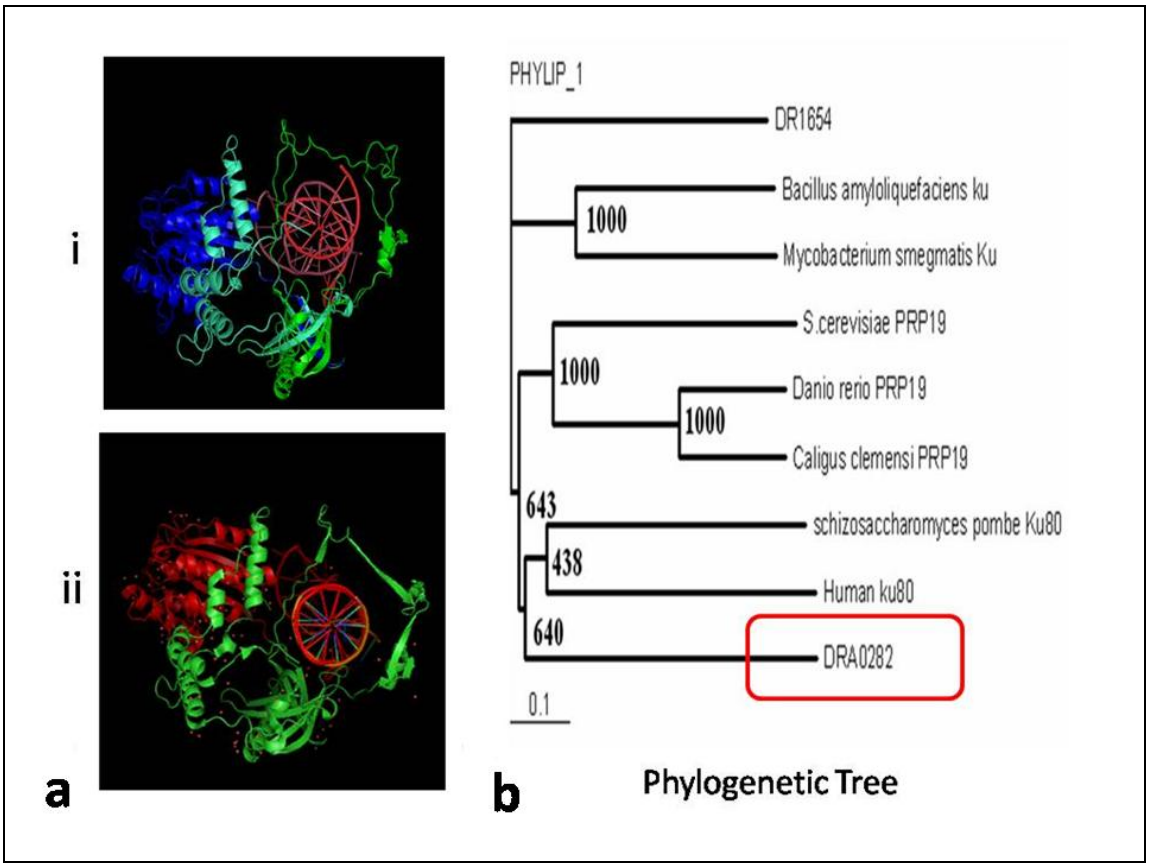


Fig 3.1.2.5. **(a)** The 3-D model of DRA0282 with its homolog human Ku80.The DNA is modelled as a red spiral. i) The homology model of DRA0282; note the vWA domain in blue and the DNA binding core domain in green. ii) The template human ku80 molecule (1JEY). **(b)** The bootstrapped phylogenetic tree of DRA0282 showing its close homology to eukaryotic type ku proteins.

Based on bioinformatic analysis the structural similarity of DRA0282 with human Ku80 and therefore DRA0282 functioning as a DNA end binding protein could be predicted.

3.1.2.4 Analysis of DR0390

DR0390 with 45 peptides was the 3rd most abundant protein and the preliminary BLAST search showed matches with kinases. In NHEJ repair pathway of mammals, the activity of Ku (our match with DRA0282) and other factors of the pathway are regulated by a kinase DNA-PKcs. DNA-PKcs is a serine-threonine kinase encoded by the *PRKDC* gene in humans and needs to be bound to DNA-protein complex to be catalytically active. Ku recruits DNA-PKcs and this forms a holoenzyme DNA-PK, which then recruits Artemis, an end processing nuclease and then a ligase complex. Mechanistically, DNA-PKcs bound to Ku and DNA undergoes autophosphorylation and by which it gets activated. Activated kinase phopshorylates Ku, and then this holoenzyme complex recruits Artemis and phosphorylates it. Since we had already found homologs of Ku and Artemis we wanted to check whether indeed DR0390 is homologous to DNA-PKcs [96]. By querying the sequence databases and building the multiple sequence alignment we found that DR0390 is a multi-domain protein (Fig 3.1.2.7a & 3.1.2.7b). The N terminal had a Dihydroxy Acetone Kinase (DAK2) domain while the C terminal has DegV family domain. The DAK kinase domain is present in all enzymes involved in glycerol metabolism [97]. The unique feature of DAK kinase in *E. coli* is accepting phosphate from a phosphoprotein rather than a NTP, while in other cases like in *C.freundii*, this kinase uses ATP. The N-terminal of DR0390 matched structurally with the kinase from

C.freundii (PDB ID- 1UN9) with 1°A RSMD, indicating a good homology (Fig 3.1.2.7bI). However, all the residues at the active site are not conserved indicating that it may have an altered substrate preference (Fig 3.1.2.6).

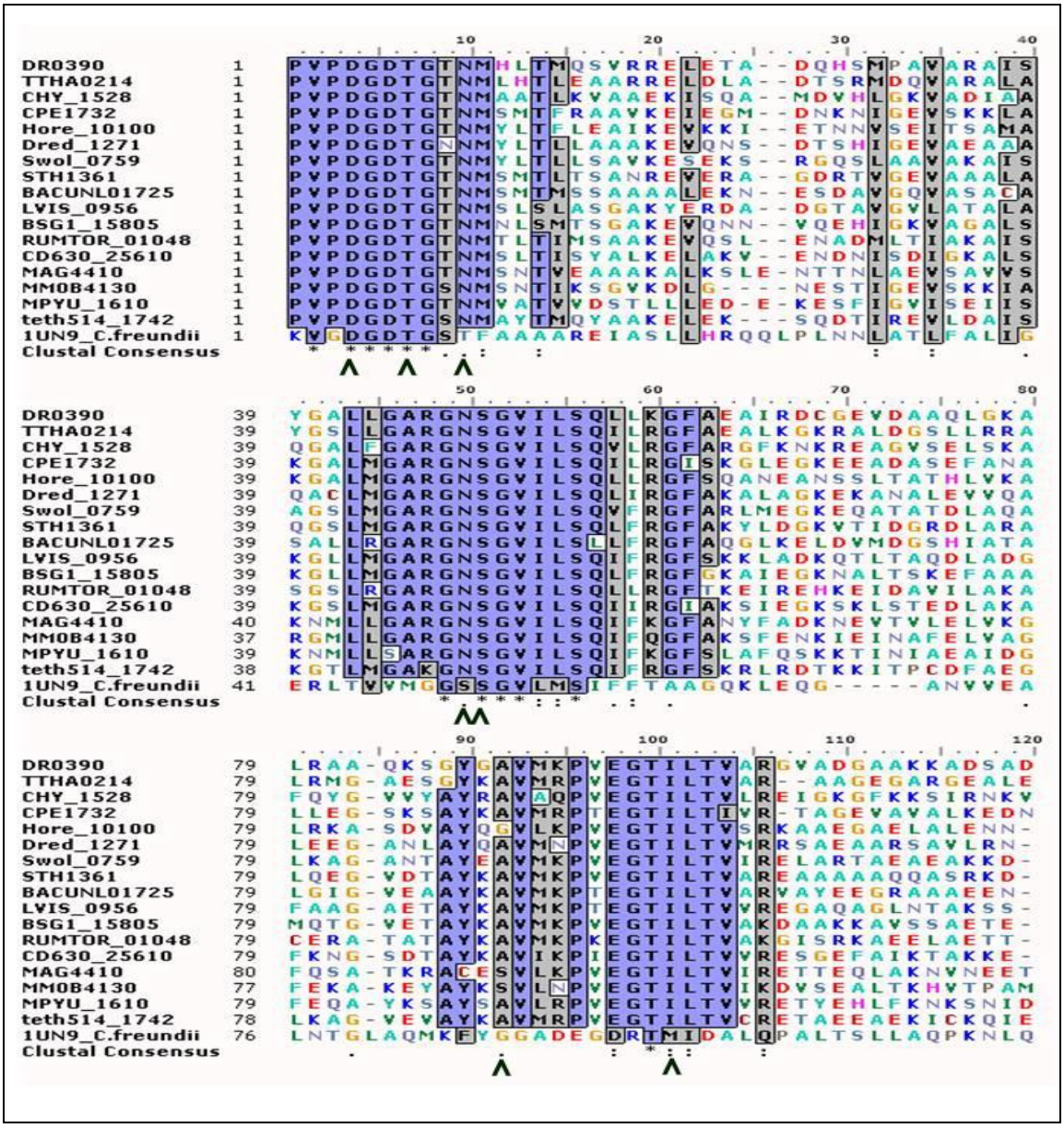


Fig 3.1.2.6. A partial Multiple Sequence Alignment of the N terminal kinase domain of DR0390 was done with the homologs of this protein obtained from pfam (PF02734). We

also included the sequence of Dihydroxy Acetone Kinase from *C.freundii* (1UN9_C.freundii) whose structure is available. The arrows indicate residues at the active site.

DegV family was first described as a conserved domain in *Bacillus spp*. A hypothetical protein from *Thermotoga maritima* (PDB ID-1MGP) which belonged to this family was crystallized and was found to bind to fatty acids and hence it is presumed that proteins with this domain might act in the lipid metabolism pathway [98]. The C-terminal of DR0390 could be modelled on 1MGP with 1° RMSD and the model showed that DR0390 has a wide shallow groove which has a capacity accept a variety of lipid/non lipid substrates (Fig 3.1.2.7bII). Due to the broad nature of the active sites on the DAK kinase and DegV domains, we can say that DR0390 is a kinase. Although, not to greater extent but the possibility of this protein involved in protein phosphorylation cannot be ruled out. Since, as seen in the figure below (Fig 3.1.2.7a and 3.1.2.7c), the domain organization of DR0390 and DNA-PKcs is quite different and we could not find any DNA binding domain therefore there was a little possibility of DR0390 having DNA-PKcs like function.

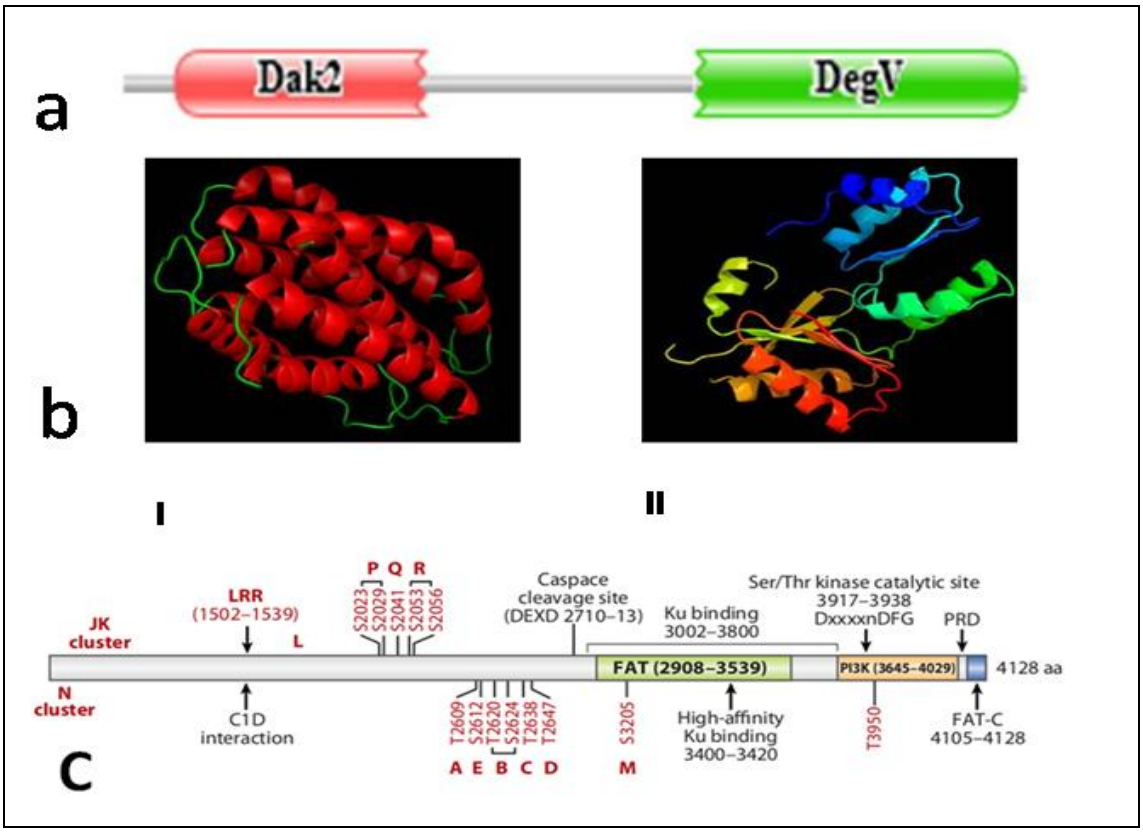


Fig 3.1.2.7. **(a)** A cartoon representation of the 2 domains present in DR0390. **(b)** The 3-D structure model of (I) the kinase domain (DAK2) in the N-terminal and (II) the lipid binding domain (DegV) in the C-terminal. In both cases the overall fold is conserved, but individual residues at the active site do not always match. Also, the inter-domain region was not modelled, possibly due to low complexity.**(c)** A schematics of DNA-PKcs. LRR is the Lecine rich region. PI3K region is the kinase site. At the center is the FAT domain found in PIK family of kinases which act as intracellular sensors. This domain is necessary for protein-protein interactions. A-E,P-R and M in red mark the phosphorylation sites.

3.1.2.5 Analysis of DR1654

With 39 peptides DR1654 was one of the most abundant proteins in the 1 hour PIR fraction. Analysis at PROSITE showed that the first 20 residues of DR1654 has a signal peptide for attachment to the membrane (MRGMKKILLASVAGAGLLAS**C**) (Fig 3.1.2.8a). The lipoprotein signal peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached. Proteins with similar membrane bound signals include major outer membrane lipoprotein (murein-lipoproteins) (*lpp*), *E. coli* lipoprotein-34 (*nlpB*), *E. coli* osmotically inducible lipoprotein B (*osmB*), *E. coli* peptidoglycan-associated lipoprotein (*pal*), a number of *Bacillus* β-lactamases, *Chlamydia trachomatis* outer membrane protein 3 (*omp3*), *Haemophilus influenzae* proteins Pal and Pcp etc [99].

PSI-BLAST search retrieved matches with DEAD box helicases and helix-turn-helix domain bearing proteins, but the e-values were 0.1, much higher than the threshold 0.001 and the alignments had intermittent gaps. Also the secondary structure obtained from JPRED server did not show any helix-turn-helix topology (Fig 3.1.2.8c).Thus the BLAST search and sequence alignments did not give any clue to the function of DR1654. We next attempted to obtain the 3-D model of this protein to gain insight into the structure and function of this protein by submitting the sequence at the PHYRE threading server. The server could model only 36 of the 182 residues. But significantly it modelled a transmembrane domain for the residues just next to the signal sequence based on COMB10, a transmembrane protein in *H.pylori* (PDB ID 2BHV) [100]. We aligned the DR1654 sequence with members of the conserved sequence family to which COMB10 belongs (pfam family of bacterial conjugation Trb-I like protein), but found that key

conserved residues are missing. A Kyte and Doolittle hydropathy plot showed that a majority of the residues had a hydropathy index >1 indicating that DR1654 is a largely hydrophobic protein and this is corroborated by the secondary structure showing stretches of β- sheet which would be hydrophobic (Fig 3.1.2.8b). A possibility exists for DR1654 that it may be involved in binding of proteins/membranes with these hydrophobic patches. We had to conclude that DR1654 is a membrane bound protein whose function could not be properly deciphered.

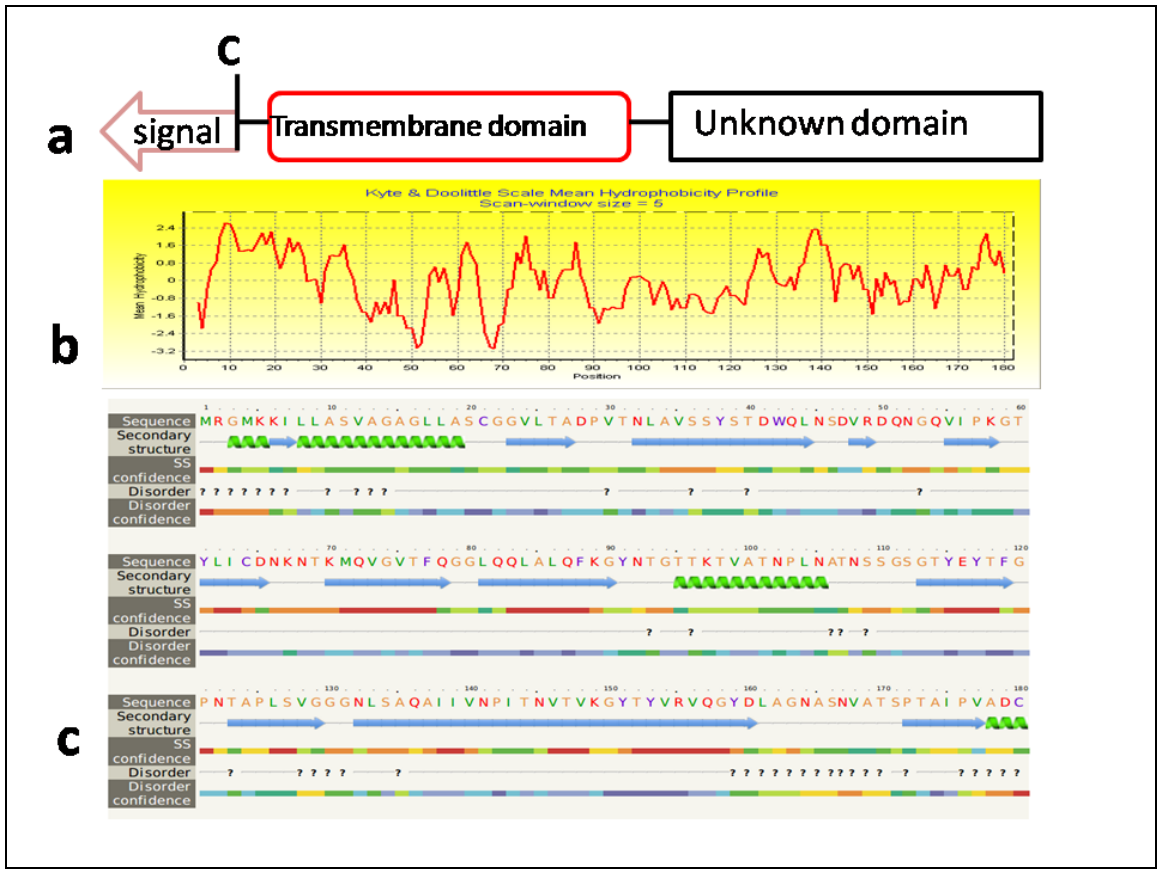


Fig 3.1.2.8. **(a)** A schematic of the domain organization of DR1654; at the N terminal is the signal sequence for homing onto periplasm and this signal is cleaved by signal peptidase II at the cysteine residue marked in the figure. Next to this signal is a transmembrane domian beyond which is the region which could not be predicted. **(b)** The

hydropathy plot shows the predominantly hydrophobic nature of this protein **(c)** the predicted secondary structure shows a major segment of β sheets which would impart hydrophobicity.

Thus, the bioinformatic analysis of these four proteins predicted the putative functions of DRA0282 as Ku like protein, DR2417m as β-CASP family protein, DR0390 as a kinase and DR1654 as a membrane bound protein. The functions predicted for DRA0282 and DR2417 proteins also fall within the DNA metabolic functions that are integral to efficient DSB repair. Surprisingly, the other two proteins do not come under the category of DNA binding proteins as they do not have known DNA binding domains but they may be secondary proteins binding to DNA binding proteins. DR0390 is a kinase having nucleotide interacting domain which could retained it to the heparin column. The binding of DR1654, a cell wall binding protein to heparin column is still intriguing. However, the hydrophobic domain in DR1654 might make this protein a good candidate for stabilizing the protein –protein interaction and that might have helped this protein to get trapped along with other DNA binding proteins during purification.

The 1 hour PIR fraction showed a novel nuclease activity and DR2417 was predicted to be a nuclease. DR2417 matched with Artemis which functions in the NHEJ pathway. The activity of Artemis is modulated by the DNA-PK complex comprising Ku and DNA-PKcs. Our analysis showed DRA0282 as a Ku homlog which also functions in the NHEJ pathway. We assumed that probably a similar interaction may occur with DR2417-DRA0282 and hence further studies were concentrated on the characterization of DRA0282 and DR2417m both *in vivo* and *in vitro*.

3.2 Functional characterization of DRA0282 for its role in radiation resistance in bacteria

Bioinformatic analysis suggested that DRA0282 is homologous to human Ku80 and could act to bring dsDNA strands together. This protein was found from the 1 hour Post Irradiation Repair fraction where there was a possibility of broken double strand DNA ends in the need for proteins which could hold them together like a 'molecular tape' and although DRA0282 did not have nuclease/kinase domains, its homology to Ku indicated that it could interact with a nuclease. Also Ku type bacterial proteins were known to be present in the phylogenetically close *Mycobacteria* spp [101]. Ku was discovered originally as an autoantigen in human patients with poly-myositis-scleroderma overlap syndrome and is one of the most abundant DNA end-binding proteins in the human cell [94]. It is a heterodimeric protein made up of two subunits, Ku70 and Ku80 and functions in NHEJ repair pathway in eukaryotes. Recent studies have demonstrated the presence of Ku and NHEJ pathway even in prokaryotes [25]. It is a highly versatile regulatory protein that has been implicated in multiple nuclear processes, e.g., DNA repair, telomere maintenance and apoptosis. Accordingly, Ku is thought to play a crucial role in maintenance of chromosomal integrity and cell survival. Specifically, prior studies suggest that a delicate balance exists in Ku expression, as overexpression of Ku proteins promotes oncogenic phenotypes, including hyperproliferation and resistance to apoptosis; whereas deficient or low expression of Ku leads to genomic instability and tumorigenesis [102]. **S**tructurally both Ku 70 and 86 are similar although their sequence-based

structural alignment is only 14% identical. It is made of three distinct domains with a β -barrel domain at the core which binds non-specifically to the sugar-phosphate backbone of DNA. Ku is now known to be one of the early proteins to be recruited at a DSB and bind to a variety of discontinuities in double-stranded DNA, including single-stranded 'gaps' and 'bubbles' of non-complementarity, but with highest affinity to blunt, 5'or 3'overhanging, and hairpin DNA double strand ends.After binding to the DNA end, Ku recruits the 470 kDa DNA-dependent protein kinase DNA-PK catalytic subunit (DNA-PKcs) and triggers its catalytic activity. In contrast to DSB repair in vertebrates, which usually occurs by NHEJ, most DSB repairs in yeast are repaired by homologous recombination. Nevertheless, NHEJ appears to be important in haploid yeast cells in the G1 phase of the cell cycle and *D. radiodurans* curiously shares a similarity of repair process with yeast. As stated earlier, *in silico* sequence searches of microbial genomes provided convincing evidence for the presence of putative Ku genes in a variety of phylogenetically diverse prokaryotic genomes. The bacterial Ku homologues (~30–40 kDa) are considerably smaller than their eukaryote counterparts (~70–85 kDa) and comprise the central core dimerization and DNA binding region of Ku70 and Ku80 [25]. However, bacterial Ku homologues typically lack both the divergent von Willebrand factor A (vWA) domain present as an amino-terminal extension in Ku70/Ku80 and the SAP domain present as a carboxy-terminal extension of Ku70. The vWA and SAP domains have been implicated in mediating protein-protein and protein-DNA interactions, respectively [95]. Our bioinformatics studies revealed that DRA0282 has the N-terminal vWA domain and the core DNA binding domain and phylogenetic studies show it to be closer to eukaryotic Ku type proteins. Also, DRA0282 has a paralog

DRB0068 which at 90 kD is much larger in size and has a signal peptide for transport to the membrane. Interestingly, in humans Ku has a membrane bound form which serves as a 'docking molecule' for Matrix Metalloproteinse 9 which is involved in cell-cell interaction [103]. Thus a study of DRA0282 would reveal an interesting aspect of DNA repair in *D. radiodurans*.

3.2.1 Methods

3.2.1.1 Construction of DRA0282 expression plasmid and recombinant protein purification

Genomic DNA of *D. radiodurans* was prepared as previously established[40]. The *drA0282* coding sequences were PCR amplified from genomic DNA of *D. radiodurans* using gene specific forward primer 282F (5'GAATTCATATGTTCATGAAGAGCAAGG 3') and reverse primer 282R (5'ACGGGATCCTAACTGACGGGAGGTGA3') having the required restriction sites incorporated at the 5' end of individual primers. The PCR product was ligated at *Nde*I and *Bam*HI in pET28a+ to yield pET282 and at *NcoI* and *BamHI* sites in pTrc99 to yield pTrc282. The pTrc282 plasmid was used for studying the effect of DRA0282 expression on radiation response of wild type and *recA* mutant of *E. coli* k-12. The pET282 plasmid was transferred to *E. coli* BL21 (DE3) pLysS and transgenic cells were induced with 200µM IPTG. The recombinant DRA0282 protein expressing (his)6 tag at N-terminal was purified under denaturing conditions using metal affinity chromatography as

described earlier [104]. The protein was refolded by serial dilution of urea in 0.8 M arginine buffer with increasing concentration of dithiothreitol as described earlier [105]. The purified and refolded protein was dialyzed in 10mM Tris-HCl, pH 7.5, 15mM KCl, 1mM DTT and 50% glycerol and stored in -20^{o}C for further use.

3.2.1.1 Isolation of a Deletion Mutant of *drA0282* in *D. radiodurans*

The Δ*drA0282* deletion mutant was generated using protocols described previously [106]. In brief, the 1 kb upstream and downstream fragments of the coding sequences of *drA0282* were PCR-amplified using sequence-specific primers. The upstream fragment was amplified using primers 282upF (5' ATGGGCCCGCTTTGTACTCCA-GAGCA3') and 282upR (5'-CGGAATTCTTCATGAACACCAGGT-3'), and the downstream fragment was amplified using 282dnF (5'- TGGATCCACCACCCGTCAGTAAGCT-3') and 282dnR (5'-CGT-CTAGAAGCGCCCTGGTTGACCGT-3'). The products were cloned independently and sequentially in pNOKOUT to yield pNOK282. This plasmid was linearized with ScaI and *D. radiodurans* was transformed with this plasmid to generate the Δ*drA0282* mutant. PCR amplification was carried out using sequence-specific primers and complete replacement of *drA0282* with the nptII gene was confirmed. These clones were considered as Δ*drA0282* mutants, and a single such clone was used for further studies. To generate the double mutant of *pprA* and *drA0282*, linearized pNOK282 was transferred into the pprA :: cat mutant of *D. radiodurans* [55]. Recombinant cells were sub-cultured for many generations in TGY broth containing both

chloramphenicol (5 mg/ml) and kanamycin (10 mgml). Genomic DNA was prepared, and the absence of *drA0282* in the pprA mutant background was ascertained by PCR amplification using gene-specific primers.

3.2.1.2 Cell Survival studies

Cell survival studies were carried out as described previously [107]. In brief, mutant and wild-type *D. radiodurans* cells were grown in TGY broth to the late-exponential phase at 32 °C. Cells were suspended in sterile phosphate buffered saline and exposed to various doses of γ radiation at a dose rate of 5.89 kGy/hr (γ 5000, ^{60}Co, Board of Radiation and Isotopes Technology, Department of Atomic Energy, India) as described previously [107]. Appropriate dilutions were plated on TGY agar plates and incubated at 32 °C. *E. coli* cells expressing DRA0282 on pET282 and pTrc282 were treated with different doses of UV and γ radiation as described previously [67]. In brief, exponentially growing *E. coli* cells induced with 200 µM IPTG were washed with phosphate buffered saline and resuspended in phosphate buffered saline, and treated with various doses of γ radiation at a dose rate of 356 Gy/hr in a Gammacell 220 irradiator, MDS Nordion) and UV at 254 nm. UV-treated cells were plated on LB agar plates containing kanamycin (25 mg/ml) and incubated in the dark. The plates were incubated at 37°C overnight for *E. coli*. Similarly deletion mutants of DRA0282 and *wild type D. radiodurans* cells were exposed to various doses of UV and these cells were grown for 36 hrs at 32°C.

3.2.1.3 DNA binding and ExoIII protection studies

The DNA-binding assay was carried out as described previously [40]. In brief, 100 ng purified DRA0282 was incubated with 200 ng PCR-amplified deinococcal DNA/super-helical form of pBluescript SK+/single-stranded linear DNA in buffer containing 10 mM Tris/HCl, pH 7.5, 5 mM $MnCl_2$, 15 mM KCl and 2 % (v/v) glycerol for 20 min at 37°C. To determine the effect of ATP, the proteins were pre-incubated with 2 mM ATP for 20 min at 37 °C, followed by incubation with 200 ng DNA substrate for 30 min. The ExoIII protection assay was performed as described elsewhere [108]. In brief, 200 ng DNA substrate was pre-incubated with different concentrations of purified DRA0282 for 10 min in Exonuclease III Buffer (New England Biolabs). The reaction mixture was further incubated for 20 min in the presence of an increasing concentration of ExoIII. When required, the nucleoprotein complex was dissociated by heating at 65 °C in the presence of 95 % formamide and 25 mM EDTA. Products were analysed on 1 % agarose gels and DNA bands were visualized with ethidium bromide. For determination of the affinity constant of DRA0282 for different DNA substrates, 200 ng DNA was incubated with an increasing concentration of recombinant DRA0282 (4 nM to 1.42 mM) for 10 min, and products were analysed on 0.8 % agarose gels or 5% native PAGE gels. Ethidium bromide-stained DNA fluorescence intensity was measured using ImageJ software [109]. The integrated intensity for bound and unbound fractions per unit area was measured separately, and the fraction bound to the protein was plotted as a function of the protein concentration using GraphPad Prism 5. The Kd for curve fitting of individual plots was determined by the software, working on the principle of the least-squares method and

applying the formula $Y = (Bmax \times [X]) \div (Kd + [X])$ where B_{max} is the maximum binding capacity,Y is the bound fraction and [X] is the protein concentration.

3.2.2 Results

3.2.2.1 Recombinant DRA0282 of *D.radidodurans* was purified from *E. coli*

The *drA0282* ORF was cloned and expressed in *E. coli* BL21 (DE3) and recombinant DRA0282 was purified to homogeneity (Fig 3.2.2.1). Attempts were made to purify the soluble form of the protein, by inducing at lower temperatures (16°C) or by adding 2% ethanol or 2 µg/ml chloramphenicol after the culture attained mid-log phase, however we found that at 1 mM IPTG majority of the protein would be found in the inclusion bodies and hence we purified the protein under denaturing conditions and then refolded it in an 0.8 M arginine containing buffer.

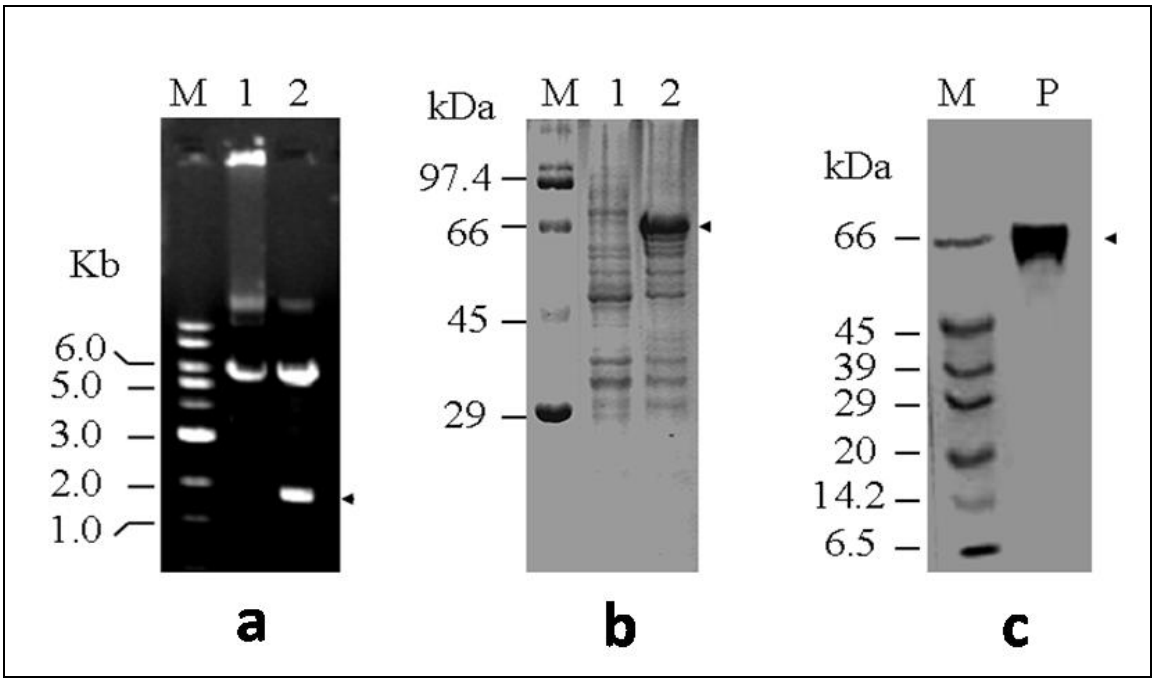


Fig 3.2.2.1. Expression of DRA0282 in *E. coli* and purification of recombinant protein. **(a)** The coding sequences of DRA0282 were PCR-amplified and cloned into pET28a+ to yield pET282. pET28a+ (1) and pET282 (2) were digested with NdeI and BamHI, and the size of the insert released was compared with the size marker (M). **(b)** pET282 was transformed into *E. coli* BL21, and cells were grown and induced with IPTG. Total proteins from the uninduced control (1) and induced (2) cells were analysed by SDS-PAGE. The size of the recombinant protein was confirmed with respect to a protein molecular mass marker (M). **(c)** Transgenic *E. coli* expressing DRA0282 was used for the purification of recombinant DRA0282 protein (P). This purified protein was used for activity characterization.

3.2.2.2 DRA0282 is a DNA-binding protein with an ability to protect double stranded DNA in solution

The purified recombinant DRA0282 protein was incubated with single stranded DNA (ssDNA) and linear double stranded DNA (dsDNA) in the presence of different metal ions. The protein showed strong binding to both double stranded DNA and single stranded DNA in the presence of Mn^{2+}, while this activity was not observed in the presence of Ca^{2+} and Mg^{2+} (Fig. 3.2.2.2a). DRA0282 showed stronger binding to single stranded DNA than to double stranded DNA in the presence of Mn^{2+} (Fig. 3.2.2.2a). It did not bind to double stranded DNA in the presence of Zn^{2+}, while its binding to single stranded DNA in the presence of Zn^{2+} was minor compared with its binding in the presence of Mn^{2+}. Recombinant DRA0282 showed a strong binding to Covalently Closed Circular DNA (CCC DNA), without affecting the topology or size of the plasmid DNA. This was confirmed by the complete recovery of the substrate in its original form upon heating with 95 % formamide and 25 mM EDTA at 65 °C (Fig. 3.2.2.2b), which breaks the nucleoprotein interaction in vitro [110]. This indicated that DRA0282 was a Mn^{2+}-dependent DNA-binding protein. A critical feature of Ku type proteins is their preference for blunt ends over resected ends [111] and hence we wanted to study the DNA end preference of DRA0282. The effects of DNA with different types of, single stranded DNA and CCC DNA on double stranded DNA binding were monitored by a competition assay. A 200 bp ^{32}P-labelled double stranded DNA probe was used to study the binding preferences of DRA0282. The double stranded DNA with different types of ends generated by BamHI, EcoRV or PstI did not disrupt the binding of DRA0282 to the

probe, which, however, was readily disrupted by single stranded DNA (Fig 3.2.2.2c,lane 6) and supercoiled DNA (Fig. 3.2.2.2c, lane 7). These activities were Mn^{2+}-dependent and sensitive to ATP. The inhibition of DRA0282 by ATP was unlike Ku protein, which shows ATP dependent weak helicase activity in solution (Fig.3.2.2.2c) [94].

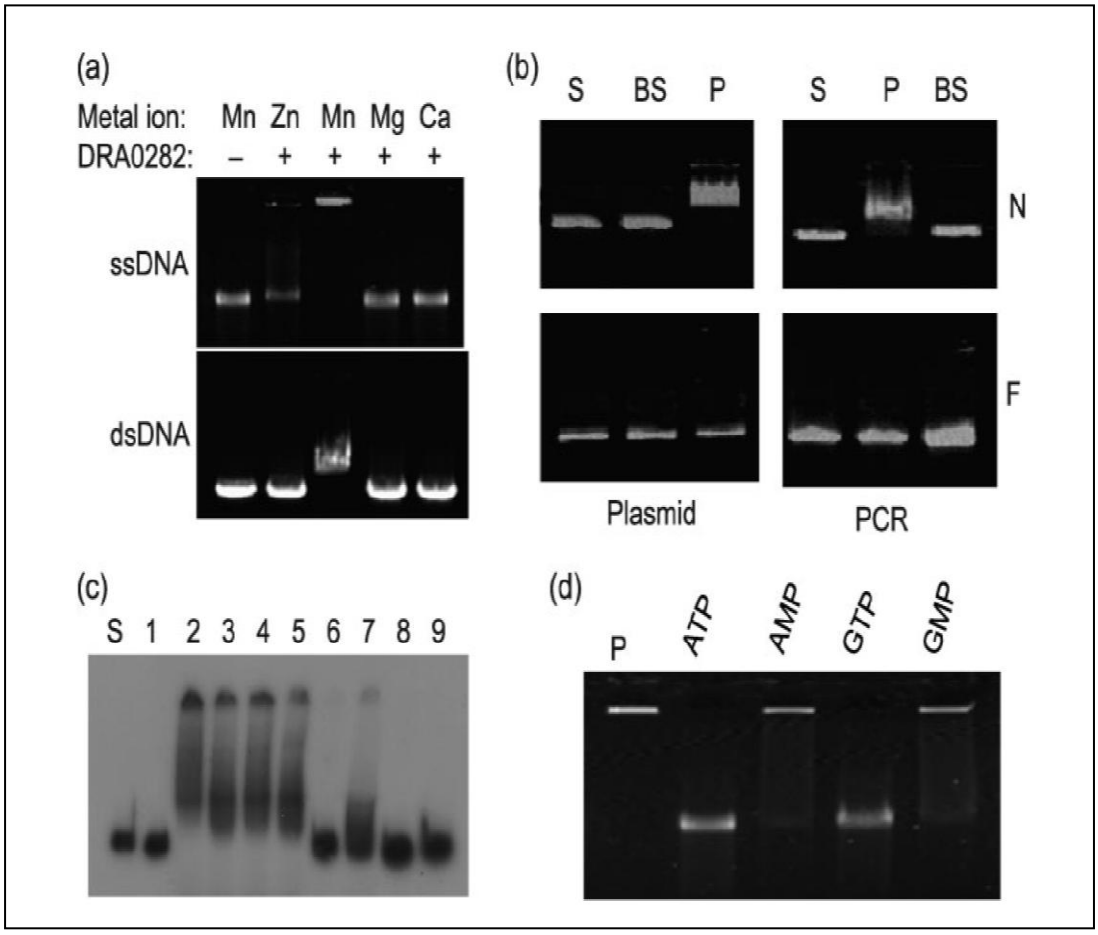


Fig 3.2.2.2. DNA binding activity characterization of DRA0282. **(a)** The 100ng purified recombinant DRA0282 was incubated with 200ng M13mp18 ssDNA and 200ng dsDNA substrates in buffer containing 5mM of each $ZnCl_2$, $MnCl_2$, $MgCl_2$ and $CaCl_2$ and products were analyzed on 1% agarose. **(b)** The effect of protein on DNA was analyzed by incubating both linear (PCR) and superhelical (plasmid) dsDNA (S) with DRA0282

(P) and BSA (BS). One set of reaction mixture was treated with 95 % formamide 25mM EDTA (F) and other set was treated as control (N). Both sets were analyzed on 1% agarose gel. **(c)** Autoradiogram showing the 200bp dsDNA labeled with [^{32}P] at 5'end (S) was incubated either with BSA (1) or DRA0282 (2-9). For competition assay, the 100ng protein was pre-incubated with 50ng of each *Bam*HI (3), *Eco*RV, (4), *Pst*I (5) digested dsDNA, and M13 mp18 ssDNA (6) and supercoiled plasmid DNA (7) separately, before adding radiolabeled substrate. The DNA binding activity (2) was also checked in absence (8) and presence of 5mM Mn^{2+} + 2mMATP complex (9). Products were separated on 6% native-PAGE, dried and autoradiogram was developed. **(d)** Effect of 2mM of each ATP, AMP, GTP and GMP on ssDNA binding activity of DRA0282 was monitored on agarose gel.

Further, the purified protein showed the lowest binding constant (K_d) of 2.93 ± 0.5 nM for CCC plasmid DNA followed by 60.22 ± 9.1 nM for linear single stranded DNA and 196.9 ± 30.8 nM for linear double stranded DNA (Fig. 3.2.2.3) indicating that this protein had highest affinity for supercoiled DNA than single stranded DNA and least to linear double stranded DNA. Since this protein contains a Ku80-type DNA-binding pocket and nucleotide interaction domain, the effect of nucleoside phosphates on DNA binding and the protection of DNA ends from ExoIII degradation were examined. The DNA–DRA0282 interaction was inhibited by ATP and GTP, while AMP and GMP did not show any measurable effect on this activity (Fig. 3.2.2.2d). These results suggested that DRA0282 was a DNA-binding protein with a relatively higher affinity for supercoiled DNA. The inhibition of the DNA– protein interaction by nucleoside

triphosphates and not by monophosphates suggested the involvement of terminal phosphates rather than the adenyl group in the inhibition of the DNA-binding activity of this protein.

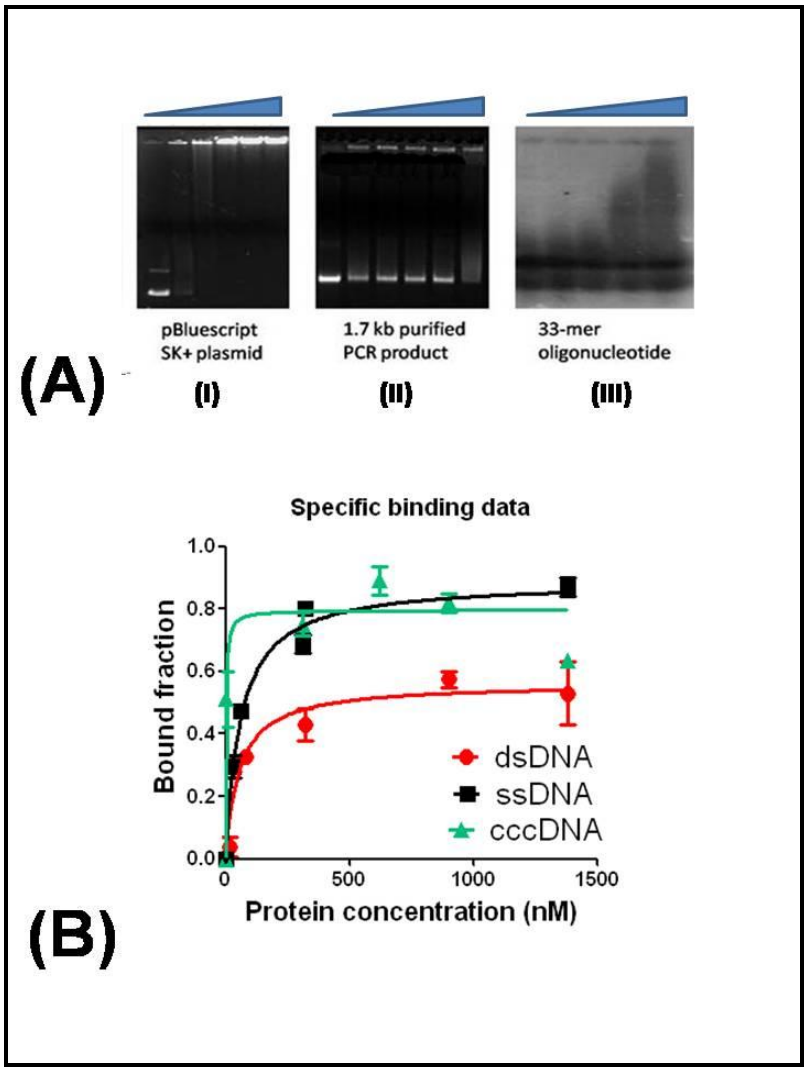


Fig. 3.2.2.3. Determination of the DNA-binding constant of DRA0282 after incubating it with linear dsDNA(●), supercoiled DNA (▲) and linear ssDNA (■) with increasing concentrations of recombinant DRA0282 for 10 min at 37 °C. (A) (I) different concentration of DRA0282 (in pmol) as indicated above the gel was incubated with 200

ng supercoiled plasmid DNA and analysed on a 1 % agarose gel (II) as in (I) but here the substrate was 200 ng of linear dsDNA (III) 20 pmol of a 32P radiolablled 33mer oligonucleotide was incubated with the indicated amount of DRA0282 (in pmol) and the product was analyzed on a 5% PAGE gel (B) The DNA–protein bound fraction was plotted against protein concentration, and affinity constants of the protein for different DNA substrates were computed using GraphPad Prism 5 graphics, as described in Methods.

Exonuclease III protection assay is a reliable tool to detect the presence of DNA binding activity in a protein. Recombinant DRA0282 protected linear double stranded DNA from ExoIII degradation, whereby as much as 50U of nuclease could not degrade the double stranded DNA substrate completely after pre-incubation with DRA0282 (Fig. 3.2.2.4a). An increasing concentration of DRA0282 afforded an increase in protection of DNA when challenged with 10 U ExoIII, which was able to degrade the double stranded DNA substrate completely in the absence of DRA0282 (Fig. 3.2.2.4b). Since DNA binding by DRA0282 was inhibited by ATP, the effect of ATP on the double stranded DNA protection of DRA0282 was examined. In the presence of ATP, DRA0282 failed to protect DNA from ExoIII degradation, and eventually more than 90 % of the DNA was degraded, in spite of the presence of 72 nM DRA0282 (Fig. 3.2.2.4c), which had shown complete DNA protection in an earlier assay (Fig. 3.2.2.4b). Since ATP alone showed no effect on ExoIII activity (Fig. 3.2.2.4c, lane 7), and the protection of double stranded DNA from the action of ExoIII was not seen in the presence of BSA as a control (Fig. 3.2.2.4 c, lane 6), it is possible that the binding of ATP to DRA0282 prevents its binding to double stranded DNA, eventually leading to DNA degradation by ExoIII. These results

suggest that DRA0282 is able to protect double stranded DNA from nucleolytic degradation by direct binding to the DNA substrate. Its higher affinity for supercoiled DNA indicated a strong possibility of this protein interacting directly with the toroids formed in *D. radiodurans* genome. Thus DRA0282 does protect DNA ends from exonuclease activity except in the presence of ATP. In context of DNA repair this makes sense as at 1 hour PIR phase, the ATP levels actually rise inside *D. radiodurans* and this also correlates with maximal nuclease activty [82]. It is possible that DRA0282 protects DNA ends till specific repair phase exonucleases are induced. The possibility of other proteins functioning like DRA0282 cannot be ruled out.

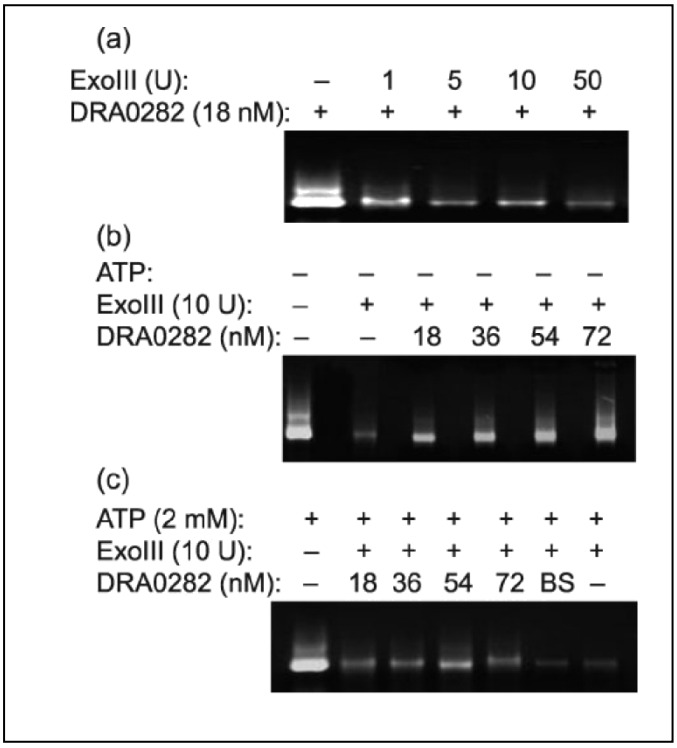


Fig.3.2.2.4. DRA0282 protection of dsDNA from ExoIII degradation. The ~150 nM 1 kb PCR-amplified dsDNA was pre-incubated with ~18 nM DRA0282 for 5 min before increasing concentrations of ExoIII were added. **(a)** Similarly, the DNA substrate was pre-incubated with increasing concentrations (18–72 nM) DRA0282 in the presence of 2 mM $MnCl_2$ **(b)** and 2 mM ATP along with 5 mM Mn^{2+} (**c**) before 10 U ExoIII was added. Reaction mixtures containing both BSA (BS) and ExoIII, and ExoIII alone, were used as controls. All reactions were carried out at 37 6C for 20 min, and products were analysed on a 1 % agarose gel.

The damage response of DRA0282 in bacteria was further investigated by expressing this protein in *E. coli* and by generating a knockout of this gene in *D. radiodurans*.

3.2.2.3 DRA0282 Enhances γ and UVC Tolerance in *E. coli*

The effect of DRA0282 on DNA damage tolerance was monitored in wild-type and mutant strains of *E. coli*. Wild-type cells expressing DRA0282 showed a nearly 10-fold higher tolerance to γ radiation at a dose of 200 Gy (Fig. 3.2.2.5a), and a nearly three-log increase in UVC tolerance at a dose of 1.2 kJ m^2, as compared with untransformed cells (Fig. 3.2.2.5c). Interestingly, the *recA* mutant (JC1553) expressing DRA0282 did not show any increase in γ radiation (Fig. 3.2.2.5b) and UVC tolerance (Fig. 3.2.2.5d). This indicated that DRA0282 functions in *E. coli* through RecA. Since RecA is involved in both homologous recombination and SOS response mechanisms in *E. coli*, the

contribution of RecA in supporting DRA0282 function in *E. coli* via these pathways was also examined.

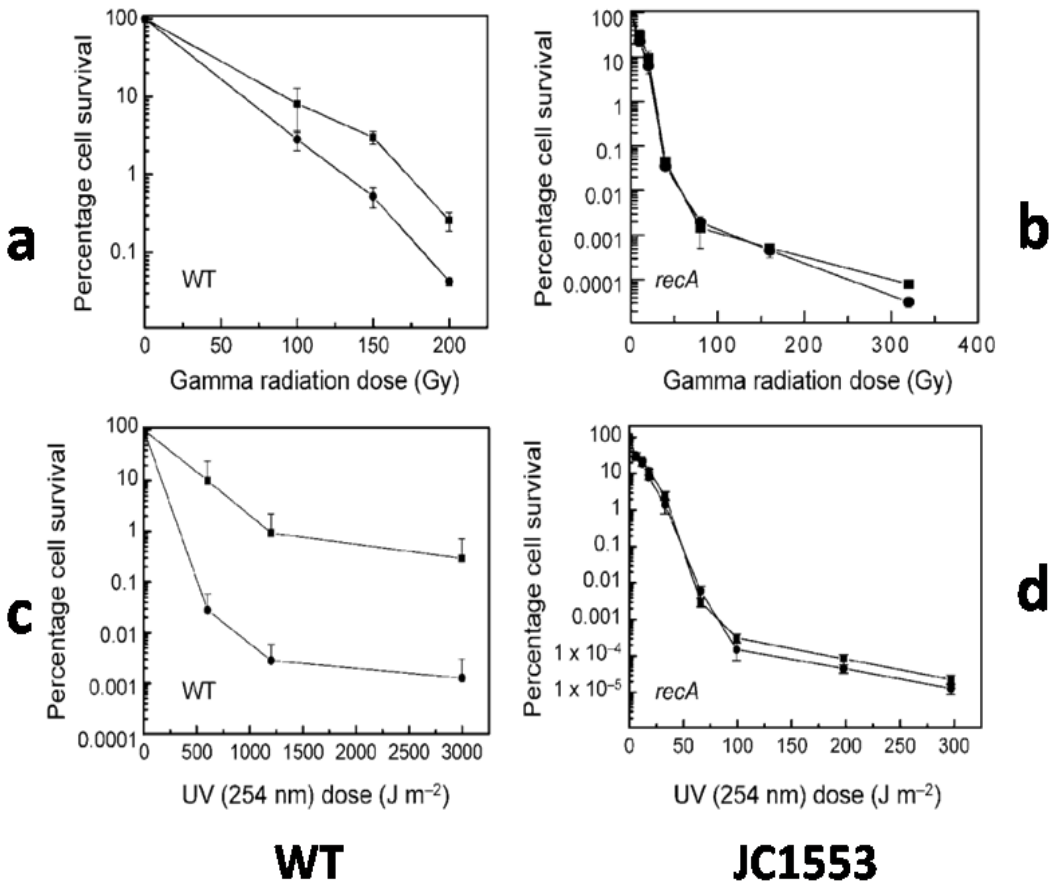


Fig 3.2.2.5. Effect of DRA0282 overexpression on the DNA damage response of *E. coli*. *E. coli* AB1157 (wild-type; WT) and JC1553 (*recA*$^-$) cells expressing DRA0282 (■) and corresponding controls (●) were treated with different doses of γ and UV (254 nm) radiation. c.f.u. were monitored after overnight incubation at 37 °C. Cell survival of 100 % corresponds to $5.423X10^7$ and $3.124X10^7$ cells/ ml of the wild-type and *recA* mutant, respectively.

DRA0282 was expressed in *E. coli* strain DM49 (*E. coli* K-12, *recA*$^{+}$ *lexA3*) [79], and the effect of both UVC and γ radiation was monitored. These cells showed an enhanced tolerance to both UV and γ radiation (Fig. 3.2.2.6), albeit slightly lower than that observed in the wild-type background. The DM49 strain has a wild-type copy of RecA but a point mutation in the RecA co-protease cleavage site of LexA (*lexA3*), making it SOS-deficient but still recombination-proficient. These results suggested that DRA0282 contributed to DNA damage tolerance of *E. coli* through mechanisms that involve RecA, but still we could not dissect the exact pathway in which DRA0282 interacts with RecA. Since DRA0282 has a strong binding affinity to single stranded DNA along with supercoiled double stranded DNA, the possibility of this protein helping in bridging the double stranded DNA with single stranded DNA before RecA performs its functions could be speculated.

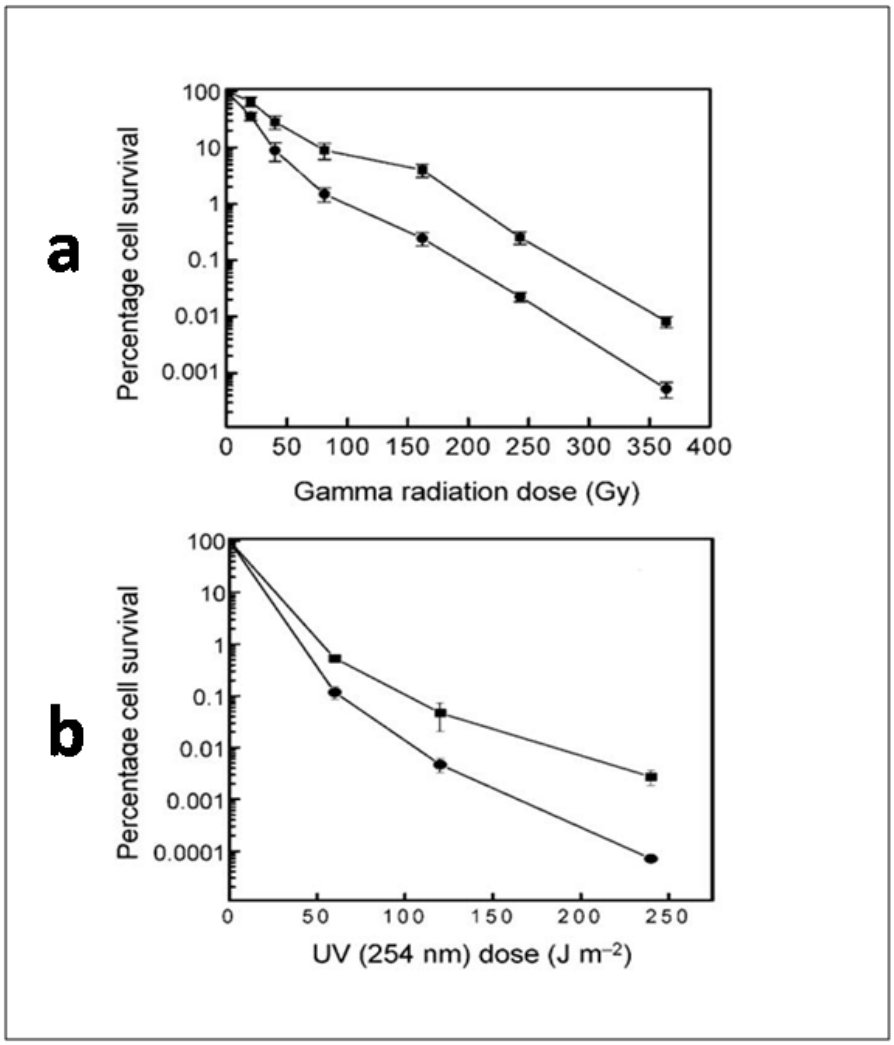


Fig3.2.2.6. Effect of DRA0282 on DNA damage tolerance of SOS-deficient and recombination-proficient *E. coli* . *E. coli* DM49 (*lexA3* mutant) cells expressing DRA0282 (■) or harboring the expression vector (●) were treated with different doses of γ **(a)** and UV (254 nm) **(b)** radiation, and cell survival was monitored at 37°C. Cell survival of 100 % corresponds to $7.21X10^7$ cells ml.

3.2.2.4 DRA0282 Deletion Mutant Does Not Alter the *pprA* Mutant Phenotype in *D. radiodurans*

The Δ*drA0282* mutant showed a dose-dependent response to γ radiation-induced DNA damage. It showed nearly a 10-fold reduction with respect to wild-type tolerance at 14 kGy γ radiation (Fig. 3.2.2.7a), while no difference in the γ radiation response was noted in either a pprA single or a *pprA* : : cat Δ*drA0282* double mutant. Both these mutants showed a nearly two-log decrease with respect to the wild-type tolerance of γ radiation (Fig. 3.2.2.7a). This suggests that although both these proteins contribute to the radiation resistance of this bacterium, PprA plays the dominant role. The levels of expression of the *drA0282* gene were increased nearly 20-fold within 1 hour of post-irradiation growth (Fig.3.2.2.8), further supporting the possible involvement of DRA0282 in the radiation response of *D. radiodurans*. Curiously, the Δ*drA0282* mutant is 4 fold more resistant as compared to the wild type at 2500 J/m^2 UVC (Fig 3.2.2.7b). Some parallel can be drawn with its *in vitro* studies where it prevent exonuclease III digestion of the DNA, indicating that it prevents that access of DNA metabolic proteins to DNA. Also, DRA0282 has been found to be a major component of *D. radiodurans* DNA toroid [112]. Thus DRA0282 could act as a 'gatekeeper' for access of proteins to DNA.

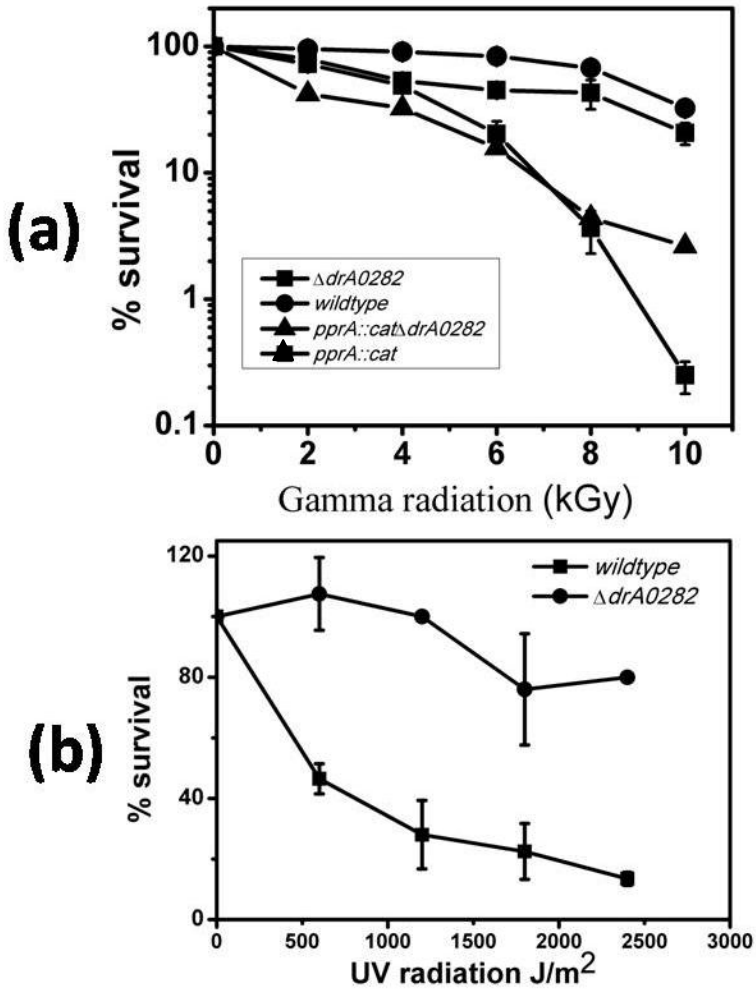


Fig 3.2.2.7. **(a)**Effect of *drA0282* deletion on the γ radiation tolerance of *D. radiodurans*. Cells of *D. radiodurans* R1 (wild-type; ●) and its Δ*drA0282* single mutant (■), *pprA:: catΔdrA0282 double mutant* (▲) and *pprA:: cat* single mutant (▼) were treated with different doses of γ radiation, and cell survival was monitored. Cell survival of 100 % corresponds to $1.45X10^{7}$, $4.12\ X10^{6}$, $2.54\ X10^{7}$ and $1.72\ X10^{6}$ cells/ml for the wild-type, *ΔdrA0282* , *pprA:: cat* and *pprA:: catΔdrA0282* strains, respectively.**(b)** Effect of *drA0282* deletion on UV radiation tolerance of wild type *D. radiodurans* R1(■) and its Δ*drA0282* single mutant (●) were treated with different doses of UVC radiation, and cell

survival was monitered. Cell survival of 100% corresponds to $3X10^8$ cells/ml and 5.2 X 10^8 cells/ml for the wild type R1 and Δ*drA0282* single mutant respectively.

Previously it has been reported that γ-irradiated *D. radiodurans* R1 cells show rapid DNA degradation, which is arrested within 90 min of PIR [34,105]. During this period the cells show a higher rate of DNA synthesis, leading to a net increase in DNA content and DSB repair. This period coincides with the expression of DNA-protecting proteins such as PprA and DdrA [54,55], and many uncharacterized ORFs, including *drA0282* (Fig.3.2.2.8). The protection of linear double stranded DNA from ExoIII activity by PprA, DdrA and now the DRA0282 protein (Fig. 3.2.2.4), which is also found to have a role in radiation resistance of *D. radiodurans* (Fig. 3.2.2.7) suggested that DNA protection from nucleolytic degradation plays an important role in the radiation resistance of this bacterium. The fact that DRA0282 enhanced the UVC and γ radiation resistance of wild-type *E. coli* and a *lexA3* mutant, although to a lesser extent than the wild-type, while not doing so in a *recA* mutant, suggests a role for this protein in recombination-dependent mechanism(s) that contribute to DNA repair in *E. coli*. On the other hand, the molecular mechanisms that support the role of this protein in the radiation resistance of *D. radiodurans* are not yet clear. However, DRA0282 showed structural homology to Ku homologues, and the different roles of Ku-type proteins, including non-homologous end-joining (NHEJ) and DSB repair, have been shown in higher organisms [113].

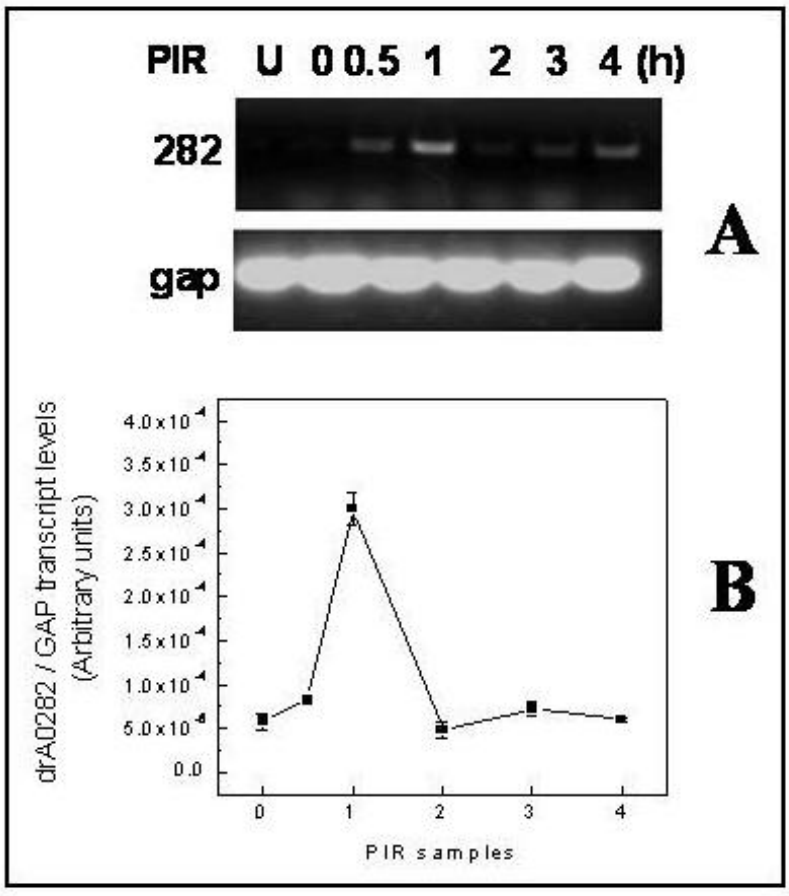


Fig. 3.2.2.8. Effect of γ radiation treatment on the expression of the *drA0282* gene in *D. radiodurans*. Both unirradiated (U) and γ-irradiated cells (0) were allowed to recover for different time intervals, 0.5, 1.0, 2.0, 3.0 and 4.0 h, and total RNA was isolated. RNAs were converted to cDNA using random primers. RT-PCR was carried out using *drA0282* -specific (282) and glyceraldehyde-3-phosphate dehydrogenase-specific (gap) primers. **(a)** products were analysed on a 1% agarose gel and **(b)** the intensities of the ethidium bromide-stained DNA bands were quantified by density scanning; the ratios of *drA0282* to glyceraldehyde-3-phosphate dehydrogenase (GAP) (drA0282/GAP) transcript levels are presented herein.

For example, human hPso4, a protein having similarity to DRA0282 , has been shown to bring Metnase, a human DSB repair protein with SET and transposase/nuclease domains [114], to the sites of DSB repair [115,116,117]. The hPso4 is known to bind double stranded DNA non-specifically, and has pleiotropic functions in DNA recombination and error-prone repair. The hPso4 protein is induced 15- to 30-fold in cells exposed to γ radiation and chemical mutagens, but not UV radiation [118]. These findings indicate the possible role of Ku homologues in DSB repair by both NHEJ and recombination functions. Although the existence of NHEJ-type activity has not been demonstrated in *D. radiodurans*, its genome encodes several other hypothetical proteins which could constitute an NHEJ pathway, and therefore the occurrence of such a mechanism cannot be ruled out in this bacterium. The existence of an NHEJ-type mechanism has been reported in other bacteria, including *Bacillus subtilis* [25] and *Mycobacterium* [95,101]. Recently, an alternative mechanism of DNA ends joining has been demonstrated in *E. coli* [119]. Our study has confirmed that DRA0282 is a DNA-binding protein with a higher affinity for supercoiled DNA and with a role in double stranded DNA protection, as well as its contribution to bacterial resistance to DNA damage by a mechanism that requires a functional RecA, at least in *E. coli*.

3.3. Characterization of DR2417 for its predicted functions and its role in bacterial resistance to γ radiation

D. radiodurans has an amazingly high tolerance to γ radiation with a D_{10} value of 10 kGy. A rough calculation shows that this is equivalent to a single cell tolerating a temperature equivalent of a billion degrees. Such high doses of radiation shatter the DNA into >200 pieces, which is then accurately reassembled back within 6 hours when grown in a rich medium. In the early phase of DNA repair in *D. radiodurans* the damaged DNA is formatted for subsequent RecA mediated homologous repair process. Not much is known of this phase other than the fact that initially the double stranded DNA is converted to single stranded DNA by exonuclease activity and then long stretches of single stranded DNA are synthesized [5]. As mentioned before, in an earlier study, [82] a cell lysate fraction which contained an ATP inhibited nuclease activity, was obtained from cells undergoing repair in this early phase. DR2417m was detected as one of the several hypothetical proteins from this fraction. Our bioinformatics studies showed that DR2417m was homologous to Artemis. Artemis is a key protein in the NHEJ repair process in eukaryotes and certain bacteria and one of the few DNases belonging to the versatile β-CASP family of nucleases. Since DR2417m was found in a cell lysate fraction exhibiting nuclease activity where no known nuclease could be detected, it was imperative to study the properties of this protein. Adding to the relevance of studying this

protein was the fact that the *dr2417m* ORF deposited in the public genome database was annotated as having a genuine frameshift at position 996 of the ORF due to which its C-terminal half had no regular secondary structure and had a high proportion of proline and arginine residues (please refer to Chapter 3.1).

Our studies showed that a hypothetical protein DR2417m in this fraction was a member of β-CASP family of nucleases. β-CASP family proteins are members of the broad metallo-β-lactamase family of proteins which act on substrates having ester bonds and a negative charge. Callebaut *et al*., 2002 first identified the β-CASP family as metallo-β-lactamase family protein with a CASP domain inserted in the lactamase domain which makes these members specific for nucleic acids [120]. All these nucleases are known to be inhibited by ATP and the nuclease activity in the fraction too was ATP inhibited. Also *dr2417m* ORF had a frameshift which disrupted the C-terminal secondary structure. Such unusual structures are usually seen in pseudogenes, but DR2417m was a functionally expressed protein as detected in fraction exhibiting nuclease activity and a multiprotein complex expressisng similar DNA processing function. Thus we decided to recheck the sequenece of DR2417m and to study the biochemical and molecular genetic aspects of DR2417m in context of radiation resistance in *D. radiodurans*.

3.3.1 Methods

3.3.1.1 Cloning, Purification and Expression of DR2417

Genomic DNA was prepared from actively growing cultures of *D. radiodurans* R1 as described earlier [105]. The standard recombinant DNA techniques including plasmid DNA isolation from *E. coli* were used as described in [104]. The *dr2417m* coding region was PCR amplified from total genomic DNA of *D. radiodurans* R1 using gene specific primers (Forward primer, 5'-GAATTCCATATGACCAGACCAGAACAA-3' and reverse primer, 5'-CGGGATCCTCAATCAAAGAAAGGGAA -3') and having required restriction endonuclease site incorporated at the 5' ends in respective primers. The PCR product was sequenced to confirm the identity of *dr2417* and the absence of mutation in the coding region of the gene. It was cloned at *Nde*I- *Bam*HI sites in pET28a+. The recombinant plasmid containing *dr2417* gene under IPTG inducible T7 promoter, was named as pET2417. Transgenic *E. coli* BL21 cells harboring pET2417 were induced with 1mM IPTG and the cells were sonicated on ice with 50% duty cycle with pulses of 30s. Lysed cells were centrifuged at 5000x g for 5 min followed by 20000xg for 30 min at 4^{o}C. The pellet containing recombinant protein in inclusion was dissolved in buffer B (20mM TrisHCl, pH8.0, 8M urea and 5mM $MgCl_2$) as described in manufacturer's protocols (QIAGEN, Germany). After incubating at 37^{o}C for 30 min followed by centrifugation at 20000x g for 30 min the clear supernatant was mixed with required

quantity of 50% Ni-NTA agarose slurry and passed through column in a buffer system containing 10% glycerol and 2% ethanol. The column was washed with buffer C (buffer B at pH 6.3). The (his)6-tagged recombinant protein was eluted from matrix with elution buffer D (buffer B, pH 5.3) followed by buffer E (buffer B, pH 4.6). The fractions containing more than 95% pure protein were pooled and refolded in the presence of 0.8 M arginine by employing the serial dilution of urea with increasing concentration of dithiothreitol (0-5mM). This preparation of protein was further purified by both Q-sepharose and SP-sepharose ion-exchange and Sephacryl S-100 HR gel filtration chromatography. Purity of recombinant DR2417m was checked on SDS-PAGE. The fractions showing more than 99% purity were pooled and dialysed overnight in buffer H (50mM NaH_2PO_4, 50mM NaCl, 1mM PMSF, 1mM DTT, 5mM $MgCl_2$ and 50% glycerol, pH 8.0). Protein concentration was determined using Bradford's dye binding method.

3.3.1.2 Generation of *dr2417* Knockout and Copy Number Studies

For generating *dr2417* gene deletion in *D. radiodurans* genome, the required construct was made using strategy as described in [105].In brief, 1 kb upstream to start codon and 1 kb downstream to stop codon of *dr2417* were PCR amplified using gene specific primers. The 2417upF (5- ′ATGGTACCGCCGCAGCCTCGCCA-3′) and 2417upR (5′-ATGGGC- CCTATGCCTCCGCTTCCCT -3′) primers for upstream fragment and 2417dnF (5′ATGGATCCGTGGGAGCGGGCAGG 3′) and 2417dnR (5′-

CGTCTAGAACGTTG- CTGCCTTCG -3′) primers were used for downstream fragment. These fragments were cloned at KpnI-ApaI and BamHI-XbaI, respectively in pNOKOUT [105] to yield pNOK2417. The pNOK2417 is a suicidal vector in *D. radiodurans*, containing an expressing cassette of nptIIflanked by DNA sequences homologous to the regions flanking the *dr2417* locus in the genome of this bacterium. Integration of nptII cassette in genome of *D. radiodurans* provides stable expression of kanamycin resistance to the cells and eventually marked the replacement of *dr2417* with nptII through homologous recombination. The recombinant plasmid was linearized and transfer to *D. radiodurans* . Transformants were grown in TGY containing kanamycin (8μ g/ml) for several generations. The presence of *dr2417* and nptII genes was monitored in surviving fraction at every subculture using both internal primers dr2417F3 (P3) and dr2417R3 (P4) as well as flanking primers P1 (5′-ATGACCAGACCAGAACAA-3′) and P2 (5′-TCAATCAAAGAAAGGGAA-3′) for *dr2417* and NptF (5 ′-AGGCCACGTTGTGTCTCA-3 ′) and NptR (5′-TGCTCTGCCAGTGTTACA-3′). The yield of PCR product using internal primers of *dr2417* was compared with the yield of PCR product of *dr1343* gene amplified using dr1343F and dr1343R primers as described above, as an internal reference control. For copy number determination, DNA was extracted from R1 and the clones 4,9,10 corresponding to the growth curve as per SDS-EB protocol. 50 μg of DNA was treated with 66 ng of RNase A for 1 hr at 37 °C and then this sample was extracted twice with phenol:$CHCl_3$ and $CHCl_3$:Isoamyl alchohol. $O.D_{260/280}$ was checked in this sample to ensure its purity. Real-time PCR was carried out using the Rotor Gene 3000 (Corbett Life Science) machine. The PCRs were setup by mixing 10X SYBR green PCR mix with 100 ng of template DNA 10 pmol each of

forward and reverse primers in 20 µl reaction system. The above reaction mixtures were amplified in the following steps: step 1—denaturation at 95 °C for 5 min; step 2—denaturation at 95 ° C for 30 s; step 3—annealing at 55 °C for 30 s; step 4—extension at 72 °C for 45 sec; step 5—melting curve analysis. Steps 2–4 were repeated for 35 cycles. The specificity of respective amplicons was confirmed from the melting curve analysis. The amplification of *dr2417* ORF was carried out on wild type *D. radiodurans* and 3 partial deletion mutants with 3 technical replicates for each group. The threshold cycle (CT) values obtained from above runs were used for calculating the amplification levels of genes by REST-384 version 2 software [121]. The expressions of genes were normalized against that of a housekeeping gene, GAPDH (*dr1343*), and plotted as relative change in yield with respect to wild type *D. radiodurans* .

Table 3.2 list of primers used in copy number study

Name of the gene	locus	primer name	primer sequence	size of amplicon
DR2417	dr2417	dr2417-F3	5’-GCACGTCGTCGTTCACGT	511 bp
		dr2417-R3	5’-GTCGGCGAACTGCAAGACCA	
GAPDH glyceraldehyde-3 phosphate dehydrogenase	dr1343	dr1343-F	5’-CTTCACCAGCCGCGAAGGGGCCTCCAAGC	103 bp
		dr1343-R	5’-GCCCAGCACGATGGAGAAGTCCTCGCC	

3.3.1.3 γ Irradiation and Cell Survival Studies

γ radiation treatment to bacterial cells was carried out as described earlier [73]. In brief, the exponentially growing cells of wild type and partial knockout mutants of *D. radiodurans* were washed and resuspended in 1/10th volume of normal saline. These cells were irradiated with different doses of γ radiation on ice, at a dose rate 6.47 kGy/h at 4 °C in γ chamber (GC 5000, ^{60}Co, Board of Radiation and Isotopes Technology, DAE, India) as described earlier [82] and the colony forming units were determined as described earlier [104] . For the effect of DR2417 in *E. coli* , the *dr2417* gene was cloned in pTRC99 (Pharmacia) and recombinant protein was expressed in *E. coli* AB1157. The expression of protein was confirmed by SDS-PAGE. Transgenic *E. coli* AB1157 expressing DR2417 were checked for their survival under various doses of γ radiation and UV radiation as described in [122]. Treated cells were plated on LB agar with appropriate antibiotics and 50 μM IPTG and the cell forming units at 37 °C, were determined.

3.3.1.4 Nucleotide Sequencing Studies

dr2417m was PCR amplified using sequence specific primers (Forward primer, (5′GAAT TCCATATGACCAGACCAGAACAA3′ and reverse primer, 5 ′CGGGATCCTCAATCAA AGAAAGGGAA3′) from genomic DNA of *D. radiodurans* . PCR product was purified and full-length gene was sequenced using both terminal

primers used for PCR amplification and internal primers (dr2417F3 — 5′GCACGTCGTCGTTCACGT 3’ and dr2417R3 — 5′GTCGGCGAACTGCAAGACCA 3 ′) on automated DNA sequencer (ABI PRISM 3100 Avant Genetic Analyzer) using dideoxy chain termination chemistry. The nucleotide sequence of DR2417m, annotated as a hypothetical protein, was aligned with the nucleotide sequence obtained by sequencing. The nucleotide sequence obtained by sequencing was also translated into protein sequence in BioEdit sequence alignment editor. Functional motif search and structure prediction studies were carried out as described in earlier analysis in [83].

3.3.1.5 RNase activity

For RNase assays the glassware and water was DEPC treated. Total RNA was extracted from *E. coli* by TRI –Reagent and then extracted twice with water saturated phenol:Chloroform. This RNA was then dephosphorylated by treating with calf intestinal alkaline phosphatase and again extracted twice with water saturated phenol:chloroform. The RNA concentration was estimated to be 124 nM by UV spectroscopy. This RNA was then radio-labelled with γ 32P ATP. An assay was set up in 50 µL with 7 µL radiolabelled RNA and 300 ng DR2417, 20 U of Exo I and 600 ng of RNAse A respectively and incubated at 37 C for 1 hour. then equal volume of stop solution (7% TCA+ 100 ug/ml BSA) was added and the mixture was incubated on ice for 10 minutes and then the mixture was spun down at 12K for 10 min at 4 C. then 40 µL of the supernatant was mixed with 7ml scintillation cocktail and activity was measured in a scintillation counter. RNase A was assumed to be 100 %, and relative % activity has been

reported. Further, the 20mer synthetic RNA (5′ UGGUGGUGGAU CCCGGG AUC3 ′) as used in earlier studies [123] , was commercially synthesized and it was labeled at 5′end with polynucleotide kinase.

The 5 pmol of labeled RNA substrate was incubated with 100 ng of DR2417 in the assay buffer (10 mM Tris –HCl pH 8.0, 50 m M NaCl and 1 mM D TT supplemented with 2 mM $ZnCl_2$) and aliquots were taken at different time of incubation. The reaction mixture was separated on 14% denaturing Urea-PAGE gel, and products were visualized by autoradiography.

3.3.2 Results

3.3.2.1 The *dr2417* ORF was Reconfirmed; Cloned, Expressed and Recombinant Protein was Purified.

The strain of *D. radiodurans* which was sequenced and made available in public databases was ATCC BAA-816 and not *D. radiodurans* R1 and that strain had collected several mutations [60]. To cite a case, mis-annotation was seen in the case of RecJ (*dr1126*), where 2 ‘G’ residues were missing in the published sequence, leading to a erroneous annotation [32]. Since the C-terminal had an unusual secondary structure, and only *D. radiodurans* had an additional’ T’ in the ORF amongst all *Deinococcus spp* , we resequenced the gene and found that indeed the sequence in the genome had erroneously reported an additional ‘T’ base at position 996 of the ORF (Fig 3.3.2.1a &3.3.2.1b). After resequencing, the size of the ORF reduced from 1739 bp to 1680 bp and the size of the

protein corrected to 62 kD. Also the C-terminal half had proper secondary structures devoid of excess proline and arginine residues. To confirm the sequencing data the purified protein was analyzed by MALDI-TOF and a MASCOT search of the peptides identified the sample to be DR2417 (Fig 3.3.2.2). The mass of peptides obtained experimentally and tryptic fragments generated *in silico* after correcting for the frameshift for DR2417,showed a complete match and these peptides could be mapped on DR2417 protein (Fig.3.3.2.1c). And thus sequencing and MALDI analysis confirmed that *dr2417m* ORF, did not have frame shift mutation as reported earlier [19]. We henceforth referred to the ORF as *dr2417* and the protein as DR2417.

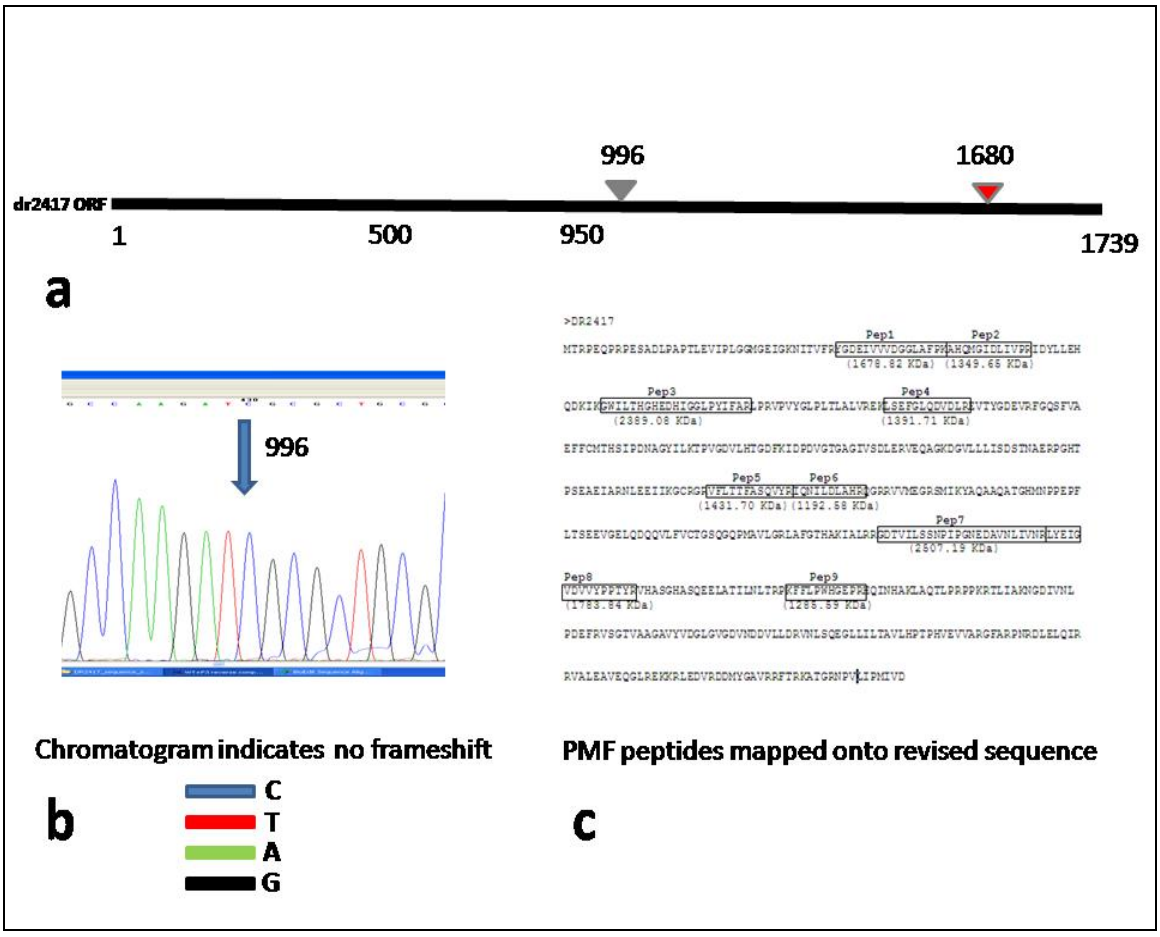


Fig. 3.3.2.1. Reannotation of *dr2417* gene. **(a)** The bar is a schematic of the *dr2417* ORF and the genome annotated length of this ORF is 1739 bp. The grey tab at position 996 marks the site of reported frameshift and the red tab marks the revised end for the ORF at position 1680 based on our study.**(b)** The genome sequence at NCBI reported a 'genuine frameshift' at position 996 of the gene. Our chromatogram from sequencing this region showed that *dr2417* in *D. radiodurans* has no frameshift. This reduced the length of the gene by 59 bases. The bars represent the colors of the bases in the chromatogram **(c)** The sequence data was confirmed by mapping peptide mass fragments obtained by mass spectroscopy to the theoretically translated polypeptide.

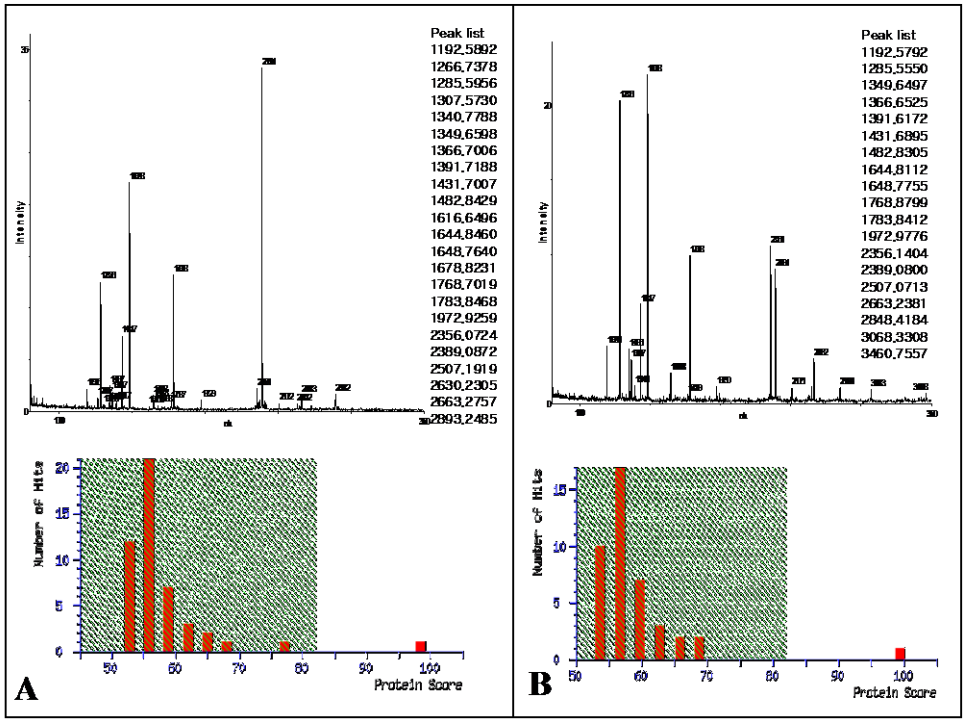


Fig 3.3.2.2. MALDI TOF MS analysis of pure DR2417 . Two samples of pure protein (A and B) was analysed for peptide mass fingerprints (upper panel). Peptide profiles of each sample were searched in Mascot library (lower panel) and identity of protein was defined.

Recombinant DR2417 (~ 60 kDa) was purified to near homogeneity from transgenic *E. coli* as described in materials and methods (Fig 3.3.2.3a). Besides purifying by three different stages of purification, a sham control of *E. coli* BL21(DE3) cell lysate was also kept in the first round of NiNTA purification. The elution fractions obtained from the cell lysate of the sham control was pooled and checked for presence of single stranded DNA nuclease activity. No significant nuclease activity was seen confirming the purity of the DR2417 protein purified by NiNTA column (Fig 3.3.2.3 b). However as an additional precaution, the purified protein was further purified by Ion Exchange (Fig 3.3.2.3a).

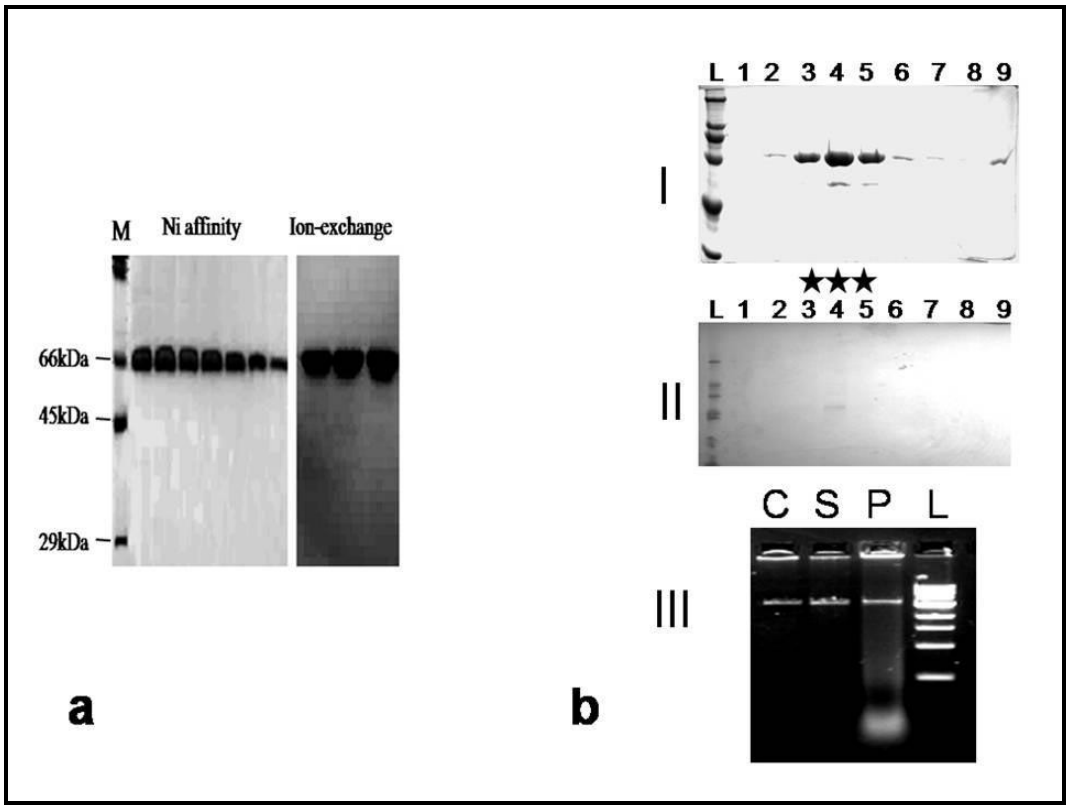


Fig 3.3.2.3.**(a)** SDS-PAGE analysis of recombinant DR2417 purified through different steps as described in materials and methods. Recombinant DR2417 was purified twice by metal affinity column chromatography using Ni-NTA (QIAgen, Germany) matrix (Ni affinity) in presence of 2% ethanol and 10% glycerol. Fractions showing best purification were pooled and further purified by passing protein through coupled SP sepharose - Q sepharose columns and eluted from Q-sepharose (Ion-exchnage). All the gels were stained with silver nitrate and checked for purity.**(b)** A sham control was maintained from the cell lysate of IPTG induced *E. coli* cells containing the vector pET28(a) only. (I) elution fractions from NiNTA column of DR2417 purified from cells having recombinant pET2417 vector.lane numbers indicate elution fractions(II) corresponding elution fractions from cells having sham control.(III) nuclease activity on

agarose gel . Fraction 3,4 and 5 marked by asterix, were pooled for DR2417 and sham control respectively and assayed for nuclease activity. C-control DNA, S- sham control, P-DR2417, L- 1kb ladder

3.3.2.2 DR2417 was characterized as a Mn^{2+} dependant nuclease

Therefore, the recombinant protein DR2417 is the correct protein encoded from *dr2417* gene on the genome of this bacterium. Nucleases have metal ion preference for their activity e.g. most RNases prefer Zn^{2+} while DNases prefer Mg^{2+} or Mn^{2+} [124] which was also evident in functional homologs of DR2417 from both bacteria and higher organisms [40,125]. Therefore the activity of DR2417 was checked in the presence of Zn^{2+}, Mn^{2+} and Mg^{2+} on both ssDNA and dsDNA. It showed a strong binding with both single stranded DNA and double stranded DNA in presence of Mn^{2+} ion (Fig. 3.3.2.4A). When DR2417 was incubated with DNA in presence of Mn^{2+} , it showed a better nuclease activity followed by Zn^{2+} and nearly no nuclease activity was observed with Mg^{2+} (Fig. 3.3.2.4B). Interestingly, the DR2417 interaction with DNA was inhibited by ATP and GTP but not with either AMP or GMP (Fig. 3.3.2.4C).

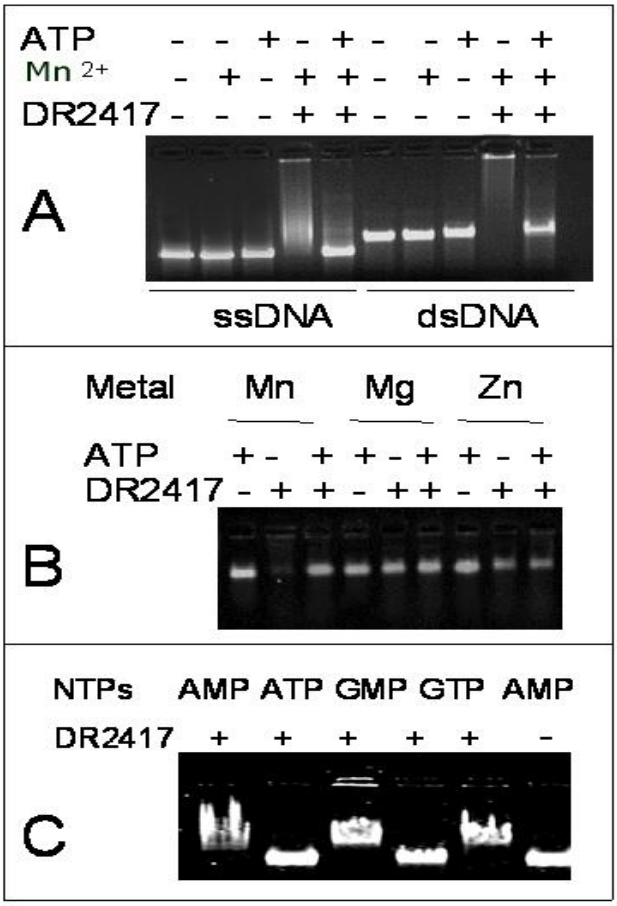


Fig 3.3.2.4. Nucleolytic activity assay of DR2417 on DNA substrates. The 100 ng pure recombinant DR2417 was incubated with M13mp18 virion DNA (ssDNA) and PCR amplified DNA (dsDNA) in buffer containing 10 mM $MgCl_2$, in presence and absence of 1 mM ATP. **(A)** The plasmid DNA digested with HindIII was incubated for 30 min with 100 ng purified DR2417 in buffer (10 mM Tris–HCl pH8.0, 50 mM NaCl and 1 mM DTT) supplemented with 5 mM $MnCl_2$ (Mn), 10 mM $MgCl_2$ (Mg) and 5 mM $ZnCl_2$ (Zn) as required, and nuclease activity was checked in presence and absence of 1 mM ATP. **(B)** Reaction mixture was heated at 85 °C for 15 min in presence of 65% formamide to break nucleoprotein complex. The effect of ATP, GTP, AMP and GMP

was also checked on its ssDNA binding activity. **(C)** Products were analyzed on 1% agarose gel and change if any, was compared with respective substrates (S) control. Experiments were repeated at least three times and results of a typical experiment shown.

Homology modelling of DR2417 showed that it is structurally similar with RNase J from *Thermus thermophilus* HB27 (PDB id 3T30 and gene id TTC0775). RNase J can hydrolyze 5′P RNA substrate but not 5′PPP RNA. Crystal structure of RNase J showed that it has a nucleotide binding pocket just upstream to the active site, hence, if there is an NTP or if the first residue of an oligonucleotide is pppNMP then subsequent phosphodiester bond is not aligned with the active site which prevents the nuclease activity. Unlike other β-CASP family nucleases, which are ambivalent [126], DR2417 was a Mn^{2+} dependent nuclease with a high specificity for DNA. This is interesting because Mn^{2+} inhibits the activity of the homologous RNase TTHA0252 from *T. thermophilus* [127] and there are no known cases of other β-CASP family nucleases that prefer Mn^{2+}. Recently however, there are reports suggesting that the different species of *Deinococcus* including *D. radiodurans* accumulates higher levels of Mn^{2+}, which helps these bacteria in higher oxidative stress tolerance and protection of biomolecules from oxidative damage [15,29,128]. Several DNA repair enzymes like X family DNA repair polymerase [129,130], SbcCD enzyme [40] and DNA processing nuclease DR0505 [67] of *D. radiodurans* have been shown to prefer Mn^{2+} for their catalytic functions. A multiprotein DNA processing complex which was isolated from this bacterium, and contains DR2417, also prefers Mn^{2+} for the DNA processing activity [131]. These findings independently suggested that since, this bacterium accumulates

higher concentration of manganese, its DNA metabolic proteins seem to have evolved for the better use of Mn^{2+} for their functions. Since DR2417 also shows a limited activity with Zn^{2+}, we presume that Mn^{2+} dependent activity is an evolutionary modification of this protein in this bacterium.

The corrected polypeptide sequence of DR2417 showed histidine at motif C of the β-CASP domain. Thus if the bioinformatics analysis be true, it should act as an RNase. Hence, we assayed the nucleolytic activity of DR2417 on RNA substrate using total RNA extracted from *D. radiodurans* and from *E. coli*. The recombinant DR2417 visually showed poor activity on RNA when assayed in presence of either Mn^{2+} or Zn^{2+}, separately (Fig. 3.3.2.5A) and analyzed on agarose. When DR2417 was incubated with radiolabeled total RNA substrate it showed a significant release of TCA soluble radiolabeled products from RNA substrate (Fig. 3.3.2.5B), but it was nearly 6 and 1.5 folds lower than RNase A and Exonuclease I controls, respectively. Incubation of DR2417 with 5 ′ kinased synthetic RNA for different time intervals showed the release of [32P]-NMP from 30 min onward. However, complete degradation of total RNA substrate was not achieved even in 2 h of incubation (Fig. 3.3.2.5C), indicating that DR2417 had a lower preference for the RNA substrate. These findings suggested that DR2417 was an efficient Mn^{2+} dependent DNase with a poor RNase activity, at least *in vitro.*

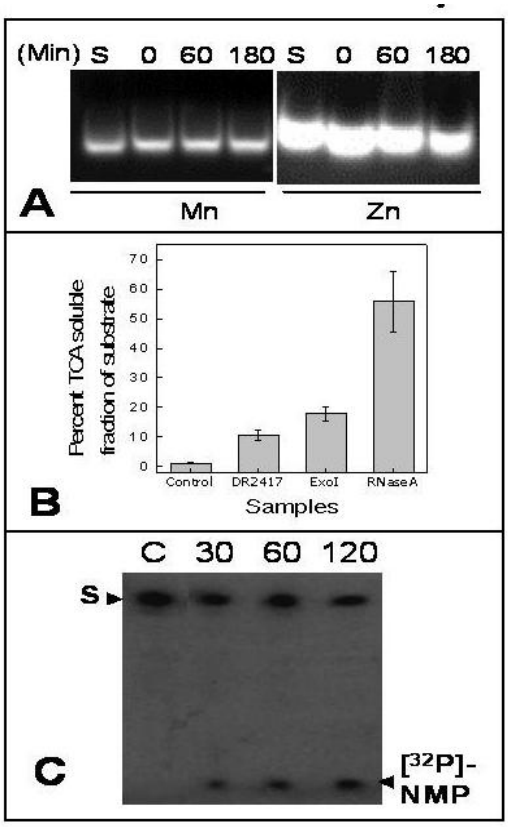


Fig 3.3.2.5. **(A)** Nucleolytic activity of DR2417 on RNA substrate was evaluated using total RNA extracted from *D. radiodurans* and from *E. coli*. The recombinant DR2417 showed undetectable levels of nuclease activity on RNA substrate assayed in presence of either Mn^{2+} or Zn^{2+}, separately and analyzed on agarose. **(B)** When radiolabeled RNA was incubated with DR2417, it showed some release of TCA soluble radiolabeled products from RNA substrate, but this was nearly 6 and 1.5 folds lower than RNaseA and exonuclease I controls, respectively. **(C)** DR2417 was incubated with 5′ labelled RNA oligonucleotide and the reaction was sampled at 30, 60 and 120 min and the products were visualized on a 16% Urea-PAGE gel

Our results showed that DR2417 acts preferentially on DNA, which is different both from the known characteristics of RNaseJ with which this protein had close structural similarities and almost all other β-CASP family nucleases. Most of the earlier studies have shown that these molecules show a considerable movement to accommodate the substrate and show substrate specificity by forming a narrow cleft suitable only for RNA molecules. Molecular basis underlying the differential utilization of RNA substrate by DR2417 and RNase J is not clear. However, DR2417 lacks few β sheet regions as compared to RNase J and also, there are several local regions, especially loops where energy constraints exist in the model (Fig 3.3.2.6). These changes are away from the active site and thus substrate preference may not just be determined by the active site residues.

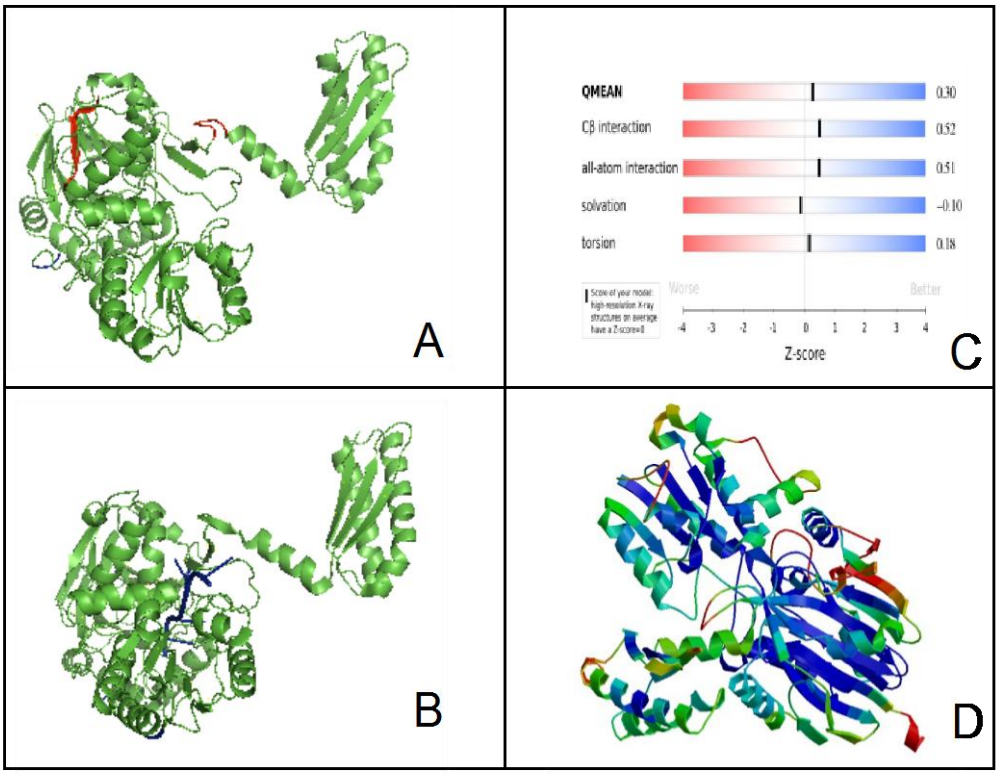


Fig 3.3.2.6. Homology model structure of DR2417. The 3-D structure of **(A)** DR2417 was generated by swissmodel using **(B)** RNase J as a template. The red colored regions highlight the differences between the template and the target. RNase J has well defined beta sheets, which lacks in DR2417. **(C)**The Q-mean plot showing that the overall quality of the model is satisfactory and **(D)** the estimated per residue inaccuracy is shown in the model, with local inaccuracies shown in red color.

3.3.2.3 DR2417 Showed 3′→ 5′ Exonuclease and ssDNA/dsDNA Junction Endonuclease

The proposed model of ESDSA based Double Strand Break (DSB) repair has suggested that the processing of shattered DNA in both polarities is required for efficient DSB repair [32,34]. However, the enzymes that process DNA during ESDSA are not precisely known but suggested to be amongst the recombination nucleases. Recent reports implicate RecJ as an important nuclease involved in recombination repair [32,132]. Also, certain proteins having DNA ends processing and processive degradation of DNA strand in $5' \rightarrow 3'$ or $3' \rightarrow 5'$ polarity have been reported [40,67,129]. The possible role of DR2417 in processing of DNA substrate for ESDSA cannot be ruled out. We have also seen that the expression of *dr2417* was highest at 1 hour PIR, the time during which extensive nucleolytic degradation is seen (Fig.3.3.2.7).

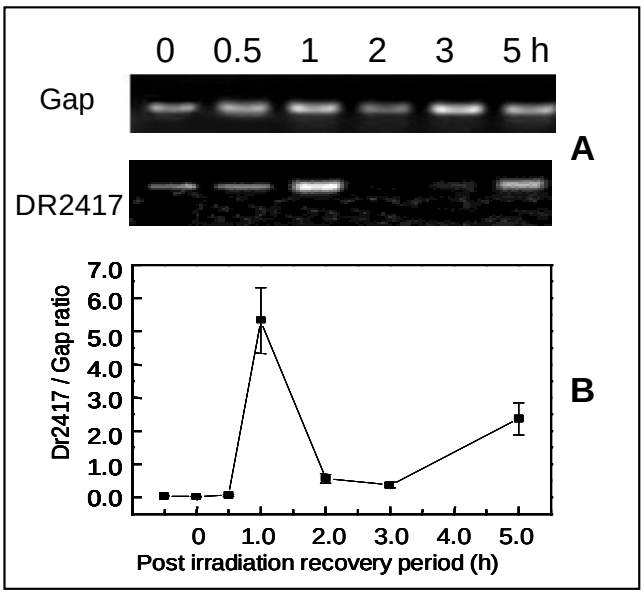


Fig 3.3.2.7. Expression kinetics of *dr2417* during post irradiation recovery of *D. radiodurans*. Total RNA was prepared from bacterial cells collected at different time interval (0, 0.5, 1, 2, 3 and 5 hour) of 4 kGy post irradiation and cDNA was synthesized using standard protocol.**(A)** Reverse transcriptase - PCR was carried out using *dr2417* (DR2417) and *dr1343* (GAP) specific primers. Products were analysed on 1% agarose gene. **(B)** The intensity of PCR product was quantified by densitometry and data obtained from three such independent experiments were computed for ratios of *dr2417* to *dr1343* genes expression.

We therefore, checked the DNA processing activity of DR2417 on DNA substrates having various types of DNA ends like single stranded DNA, linear double stranded DNA, stem– loop DNA, 5′ overhang and 3′ overhang DNA substrates. DR2417 showed very poor activity with a 5′ labeled 33mer single strand oligonucleotide [99F] substrate (Fig. 3.3.2.8A). When 5′ labeled 99F was annealed with equal size complementary

oligonucleotide to make a 33mer dimer, the enzyme showed relatively faster degradation than single stranded DNA substrate, but the release of terminal radio labeled nucleotide was seen only after 2 hour of incubation at 37 °C (Fig. 3.3.2.8B). These results strongly argued that although DR2417 possibly has 3′ → 5′ exonuclease activity on single strand DNA and double strand DNA substrates, it acted slowly on synthetic substrates lacking 5′ phosphate on opposite strand. Hence, the above 33mer dimer substrate was phosphorylated at 5′ with cold ATP and incubated with DR2417. Now the enzyme showed faster degradation with the release for terminal nucleotide in less than 30 min incubation (Fig. 3.3.2.8C). Although the results indicated that DR2417 had 3′ → 5′ exonuclease activity, the possibility of the products release due to exonucleolytic degradation in opposite polarity could not be ruled out from this result. To check this possibility, DR2417 was incubated with DNA substrate having 37-nucleotide long stem and 4- nucleotide long loop structure (HP78) labeled on both ends separately. DR2417 incubated with HP78 radiolabeled at 5′ end showed a series of typical products that are normally observed during exonucleolytic degradation of DNA (Fig. 3.3.2.8D). This enzyme also produced the islands of products from 0.5 hour onward incubation. This suggested that DR2417 has 3′ → 5′ exonuclease activities and a strong possibility of an endonuclease function at the single strand/double strand DNA junction, that could produce islands of products of 37 nucleotides and 41 nucleotides on 5′ labeled HP78. DR2417 however, showed very poor activity on 3′ radiolabeled HP78, which had a hydroxyl group at 5′end (Fig. 3.3.2.8E) and the 3′ radiolabeled [^{32}P] dNMP could be detected only after 3 hour of incubation. This further indicated that DR2417 possibly requires both 5′ phosphate and 3′ hydroxyl group in double stranded

DNA for optimum activity. This hypothesis was tested with 3′ radiolabeled HP78 also phosphorylated at 5′ end with cold ATP. The enzyme showed a faster release of 3′ terminal [32P] dNMP in less than 30 min incubation under similar condition (Fig. 3.3.2.8F). This result along with the results obtained when 3′ radiolabeled 33mer dimer substrate was kinased at 5′ position (Fig. 3.3.2.8C) suggested the requirement of 5′ phosphate for the 3′ → 5′ exonuclease activity of DR2417.

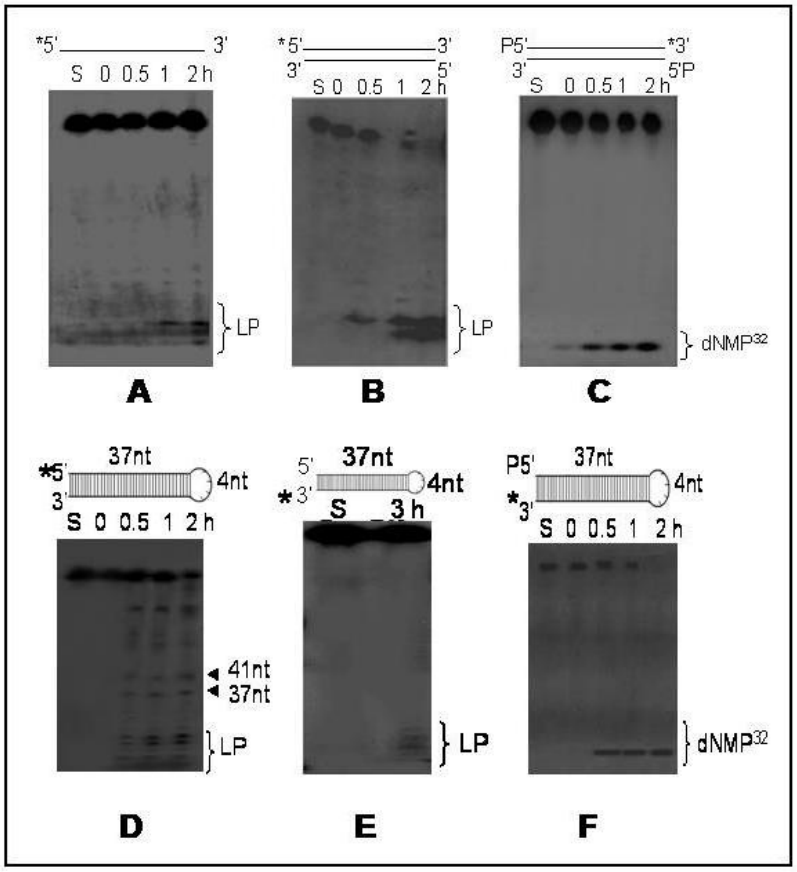


Fig 3.3.2.8. DNA processing activity characterization of DR2417 on different types of DNA substrates. DNA substrates like **(A)** 5′ labeled linear 33 bases ssDNA **(B)** 33 bp dsDNA **(C)** 3 ′ labeled 33 bp dsDNA having 5 ′ phosphorylation in opposite strand, **(D)**

78mer (HP78) stem loop DNA , **(E)** HP78 with and **(F)** without 5′ phosphorylation, were prepared. These substrates (S) were incubated with DR2417 at different time interval at 37 °C. Reaction was stopped with DNA sequencing dye (50% formamide, 25 mM EDTA, pH8.0) and mixtures were heated at 95 °C for 15 min. Products were analyzed on 16% Urea-PAGE and autoradiograms were developed. Arrow marked showing limit products (LP), terminal mononucleotide ($dNMP^{32}$) and ss/dsDNA junction endonucleolytic products. Experiments repeated two times and data of a reproducible experiment is shown.

The 1 hour PIR fraction readily degraded exogenously added PCR DNA. However when DR2417 was incubated with PCR DNA, the activity was sluggish (Fig 3.3.2.9). From our studies on bioichemical characterization (Fig 3.3.2.8E & 3.3.2.8F), we realized that DR2417 may not act on PCR amplified DNA as both ends are hydroxylated. On incubating DR2417 with EcoRI cut plasmid DNA which has 5'-P, it readily degraded it (Fig 3.3.2.9). We then treated PCR dsDNA with Polynucleotide Kinase to phosphorylate the 5'ends and when DR2417 was incubated with this modified substrate, it readily degraded this DNA (Fig 3.3.2.9). Thus there was a possibility of a DNA binding polynucleotide kinase in the cell lysate fraction from which DR2417 was isolated.

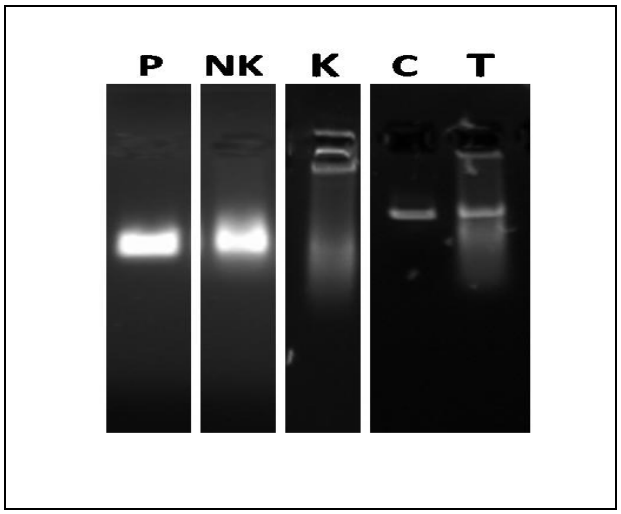


Fig 3.3.2.9. 100 ng of DR2417 was incubated either with 1 µg of PCR product (lane P, NK & K) or 200 ng of EcoRI cut pBSK+ plasmid DNA (lane C &T) at 37°C for 30 min. lane P-control PCR DNA, lane NK- PCR DNA with DR2417, lane K- PCR DNA with ends phosphorylated by PNK and then incubated with DR2417, lane C- EcoRI cut pBSK+, lane T- EcoRI cut pBSK+ treated with DR2417. Representative images of independent assays have been provided

The requirement of 5′ phosphate for exonuclease activity has been shown as a characteristic function of β-CASP family nuclease [133]. Interestingly, these results suggested that (i) double stranded DNA substrates that contain both 5′ phosphate and 3′ hydroxyl group at the same end, which would eventually be present in native DNA would be used as preferred substrate for the inherent 3′ → 5′ exonuclease activity of DR2417, and (ii) DR2417 was a ss/ds junction endonuclease.

3.3.2.4. Unlike in *E. coli*, DR2417 Showed a Dominant Phenotype in *D. radiodurans*

Several deinococcal proteins expressed in *E. coli* and vice versa have shown remarkable effects on DNA damage tolerance of transgenic bacteria [124,134]. On the other hand the deletion of certain DNA repair genes from *D. radiodurans* genome does not affect the normal growth of this bacterium [44,135]. While creating deletion mutation of *dr2417* gene in *D. radiodurans* genome, we observed that the cells approaching complete replacement of *dr2417* with nptII marker gene (Fig. 3.3.2.10), did not survive under normal growth conditions and mutants having lower copy numbers of this gene showed poor growth under normal conditions (Fig. 3.3.2.11). These cells, survived less when subjected to γ radiation stress as compared to the wild type *D. radiodurans* (Fig. 3.3.2.11). This observation was very much similar to those favoring the essentiality of a gene in bacterial system. Recently, the indispensability of RecJ in growth and DSB repair of *D. radiodurans* has been demonstrated [28]. These results might allow us to speculate that indispensability of DR2417 as an essential gene for survival and eventual γ radiation resistance in *D. radiodurans*.

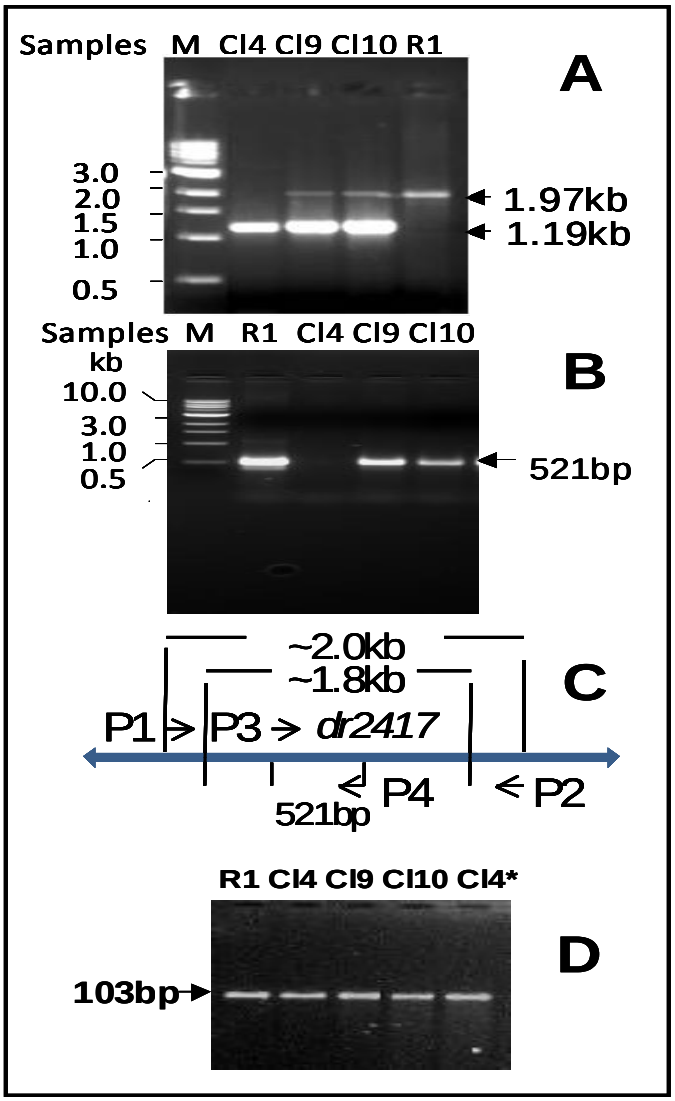


Fig 3.3.2.10. Characterization of *D. radiodurans* cells for *dr2417* deletion. Deinococcus cells transformed with linear pNOK2417 were subcultured in TGY broth supplemented with Kanamycin (8 μg/ml) and grown several generations. Certain clones were randomly picked up at specific time of subculturing and tentatively assigned as clone 9 (Cl9) clone 10 (Cl10) and clone 4 (Cl4). The genomic DNA was made and the status of *dr2417* gene was checked by PCR ampli fication using**(A)** flanking primers P1 and P2 and **(B)** internal primers P3 and P4 and the amplification of the expected size products as marked in **(C)** was ascertained on agarose gel. **(D)** almost equal levels of amplification of *dr1343* gene

taken as an internal reference control; in all the samples equal amount of template DNA was taken in each PCR reaction .

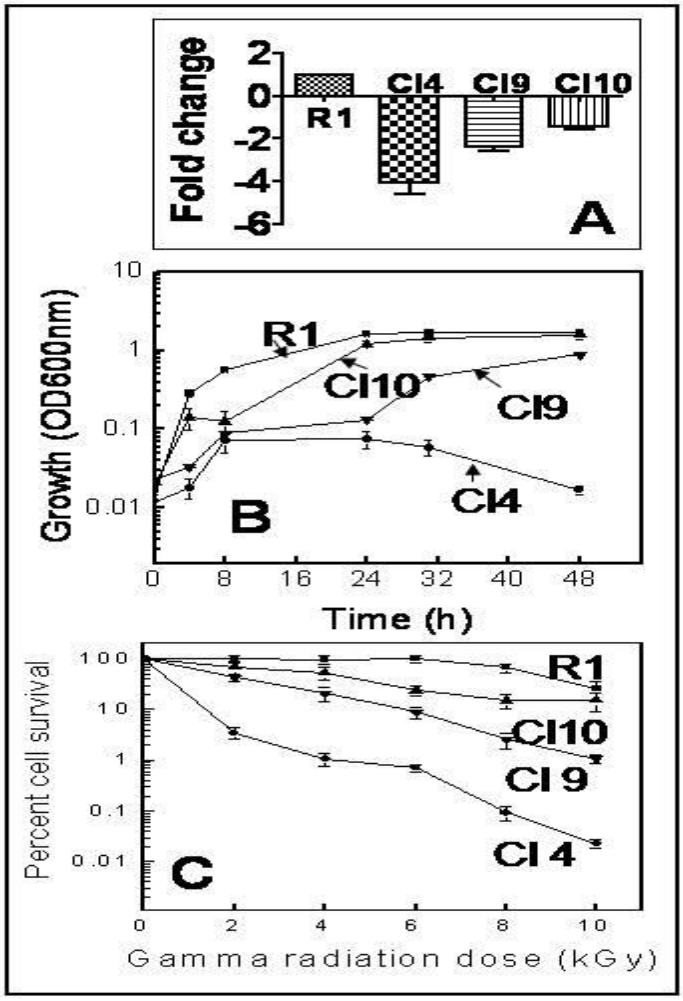


Fig 3.3.2.11. Effect of deletion of *d2417* on growth and γ radiation response of *D. radiodurans*. Genomic DNA was prepared from clone 4 (Cl4), clone 9 (Cl9), clone 10 (Cl10) as described in Fig. 3.3.2.10 , and wild type (R1) cells. Copy number of *dr2417* was determined by real time PCR using internal primers of *dr2417* and *dr1343* as a reference control, as described in methods. Levels of *dr2417* sequence amplification in

each sample were normalized by taking levels of *dr1343* amplification in respective sample as reference control. **(A)** Fold change in *dr2417* copy number was calculated by dividing its PCR yields in respective clones with the same in wild type control .**(B)**These clones were checked for growth characteristics and **(C)** responses to different doses of γ radiation. Different clones having different fold reduction in *dr2417* copy number showed proportional growth retardation and sensitivity to γ radiation

Effect of DR2417 on DNA damage response of *E. coli* to γ-rays, UVC rays and to interstrand crosslinks caused by mitomycin C was studied. The wild type *E. coli* expressing this protein on a multicopy plasmid did not show any significant effect (Fig3.3.2.12). This could be due to different levels of ATP in *E. coli* and *D. radiodurans* or some other factor like functional redundancy by other proteins or some inhibitory activity in *E. coli*. A greater possibility of *E. coli* not having the pathway (s) where DR2417 works and that contribute to extreme resistance in *D. radiodurans* may be suggested. It is pertinent to note that although homologs of DR2417 abound in gram positive bacteria where they are involved in RNA metabolism they are absent in *E. coli* and other enterobacteria [123]. Thus it is possible that the pathways in which β-CASP nucleases operate upon are missing in *E. coli* and hence we could not see any significant effect. Also, the threshold of DSB DNA needed for DR2417 to act upon is simply not generated in the radiation sensitive *E. coli* cells.

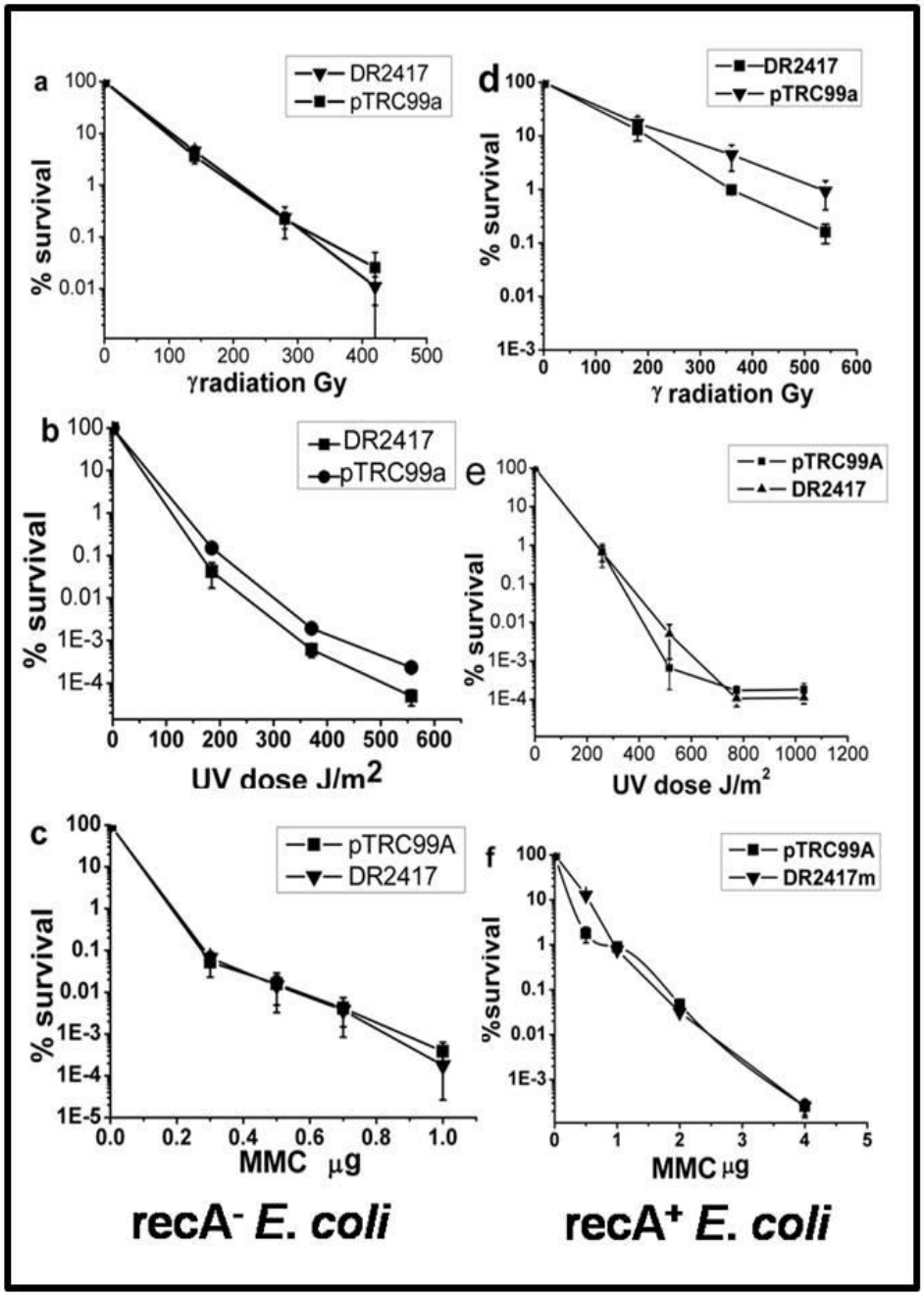


Fig. 3.3.2.12. Response of *E. coli* cells expressing DR2417 to DNA damaging agents. *recA E. coli* DH5α cells were exposed to **(a)** γ radiation **(b)** UVC radiation and **(c)** mitomycin C at various doses. Similarly wildtype *E. coli* strain AB11557(recA+) and

were exposed to **(d)** γ radiation **(e)** UVC radiation and **(f)** mitomycin C at various doses. Results from three independent experiments are shown

Further studies at cellular and molecular levels would be required for providing the direct evidence on indispensability of DR2417 in this bacterium. The reported mistake in form of a frameshift at 996th position in *dr2417*m ORF has been corrected and submitted to GenBank. Recombinant protein showed both exo- and endo-nuclease and needs a 5′ phosphate for its activity akin to Artemis, and is inhibited by ATP as observed in hSNM1, a homolog of Artemis. However, unique to DR2417 was its preference for double stranded DNA for its 3′ → 5′ exonuclease activity and a significantly low activity on single stranded DNA and RNA substrates. Although, the molecular basis for low RNase activity in DR2417 is not clear, the possibilities of other amino acids and motifs involvement in determining the substrate specificity by Motif C of β-CASP family nucleases could be speculated. All the prokaryotic proteins of this family like ykqC from *Bacillus subtilis*, TTHA0252 from *T. thermophilus*, Pab-RNase J from *P. abyssi*, have a conserved histidine in motif C similar to DR2417 but function as RNases [92,126,136]. However, the DNA specific nucleases of this family like Artemis and hSNM1 in eukaryotes have a conserved Valine in motif C and express a considerably low activity on RNA substrates [120]. Low RNase activity in DR2417 could be attributed to certain notable differences in the amino acids contributed to 3-D modeled structure of this protein (Fig 3.3.2.6). The strong endonuclease activity on HP78, a stem–loop substrate (Fig. 3.3.2.8D) also confirmed that DR2417 confers ssDNA/dsDNA endonuclease function. These results suggested that DR2417 was a DNA processing enzyme. Although, the functional significance of this protein is not clear in

D. radiodurans, our results indicated that DR2417 plays an essential role in growth and eventually γ radiation resistance in this bacterium.

We noted that at higher temperatures, DNA is degraded in the presence of Mn^{2+} but DR2417 prevents this degradation (Fig 3.3.2.13a). When DR2417 is incubated for more than 30 minutes with with double stranded DNA, DR2417 is resistant to digestion by proteinase K proteolysis (Fig 3.3.2.13b). Thus DNA is protected by the protein and DNA protects the protein. A possible explanation of this phenomenon can be if an inaccessible supramolecular complex is formed and this phenomenon needs further investigation. The inability to generate deletion mutation, a strong affinity of this protein to DNA, the protection of DNA from higher temperature together might allow to speculate the crucial role of this protein in structure maintenance and normal DNA metabolism. Inhibition of its nuclease activity by ATP is a feature which could be useful in regulation of its functions under different physiological conditions. Recently, it has been shown that the levels of ATP changes during post irradiation recovery in *D. radiodurans* [82]. With the present data it is difficult to decipher the indispensable role of this protein.

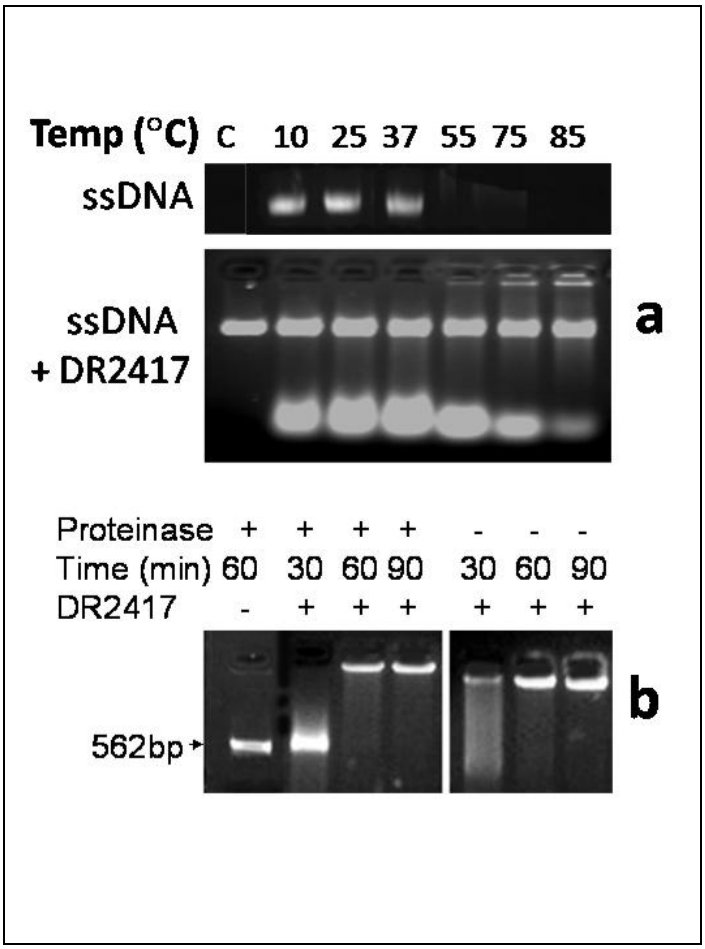


Fig 3.3.2.13. **(a)** 100 ng of DR2417 was incubated with 1 µg of ØX 174 virion DNA in a buffer containing 10 mM Tris Cl pH 8.0, 50 mM NaCl, 1 mM DTT and 2mM $MnCl_2$ at the indicated temparatures.**(b)** Effect of proteinase K on DNA-protein complex stability. The 500ng purified DR2417 was incubated with 200ng of 562bp PCR amplified dsDNA for different time interval. One set was treated proteinase K for 30 min and other was kept as control. Reaction mixture was analysed on 1% agarose gel.

A wide occurrence of similar type of proteins in *Bacilli* and other bacteria might add further to the importance of this family protein for the growth and maintenance of living organisms. Although, further work will be needed to understand the exact

mechanism of DR2417 function, the provided evidences suggest that DR2417 was a genuine DNA damage inducible novel nuclease of β- CASP family as against the reported frameshift in genome sequence, and its DNA processing activity is regulated by levels of high energy phosphates and may be essential for survival and radiation resistance of *D. radiodurans*.

Chapter 4

Discussion

As mentioned earlier, *D. radiodurans* is unique amongst living organisms in having an efficient repair process which can reseal a large number of Double Strand Breaks (DSBs) [135]. It has adapted a time consuming modification of the homologous repair pathway rather than the quicker Non Homologous End Joining (NHEJ) pathway, presumably to maintain the fidelity of information. This is an efficient pathway as more than 200 fragments are efficiently resealed even when evidence exists for transposon activity [43]. The initial phase of the DNA repair pathway, named as ESDSA pathway has a lot of resemblance with Synthesis Dependent Strand Annealing (SDSA) pathway as seen in yeasts which is essentially a non reciprocal crossover mechanism [34]. In yeasts SDSA is mediated by the RAD52 epistasis group comprising of RAD50, RAD51 (RecA), RAD52, RAD54, RAD55, RAD57, Xrs2 and MRE11[35]. RAD50/MRE11 form an ATP regulated dsDNA specific exonuclease activity and it has been shown in *D. radiodurans* that this activity is contributed by SbcCD [40]. It has been shown that DdrA could be a functional homolog of RAD52 [54] and DR1259 has been identified as a RAD54 homolog. Thus *D. radiodurans* has almost all the functional homologs of the yeast DNA repair apparatus. It is interesting to note that several fungi too share the property of radioresistance as seen in *D. radiodurans* [38] and it may be speculated that the non reciprocal homologous crossing over mechanism (SDSA/ESDSA) coupled with control of cell cycle is the reason for radio-resistance. A single and seminal study by Kota *et al*.,2008 to purify a DNA repair complex during the crucial early phase of repair however does not list these above deinococcal proteins together but listed several other proteins including DR2417 which were associated with putative DNA repair roles [131]. However, transcriptome data show that DdrA, RecA, DR1259, SbcD all are

overexpressed during the repair process. RT-PCR analysis in our study also showed the inducible expression of DRA0282 and DR2417 genes during the early phase of Post Irradiation Recovery (PIR). Thus there could be a possibility of DdrA, RecA, DR1259, SbcD being active in ESDSA but not directly by interacting with each other but through a more elaborate mechanism which might include DRA0282 and DR2417.

Most authors suggest that homologous recombination is the major repair pathway in *D. radiodurans* but that does not preclude NHEJ or similar pathways of DSB repair, at least in the initial few hours. A look at the protein profile of the first hour of PIR shows several proteins which are novel [131,137]. For example, many of these proteins are membrane or cell wall bound and have putative nucleases/kinases/protease domains and hence they could be involved in DNA repair or cell cycle arrest.

In this thesis, we had started with an observation that a cellular lysate fraction of *Deinococcus* cells in the early phase of repair showed a nuclease activity which was inhibited by ATP. In prokaryotic systems ATP stimulated nucleases are more prevalent, a classic case being the RecBCD system and hence the observed ATP inhibited nuclease activity was unusual [138]. It is also well known that extensive but regulated nuclease activity occurs in the early phase of repair in *D. radiodurans* [34] and hence finding the identity of this possible nuclease and/or kinase was of significance. Analysis of mass spectroscopy data by MASCOT identified a large dataset >300 proteins, but none of them was annotated as a nuclease or kinase in the database. Here it is pertinent to remember that automatic annotation of protein function is not always accurate. As outlined by Alfonso Valencia *et al*.,2005 [139], the general model is that rate of evolution decreases from catalytic site, to protein core, to substrate specificity site and to surface

regulation.The degree of divergence of functional annotation is therefore, dependent on the degree of divergence of sequence similarity. For an unknown sequence which has at least 15% sequence identity with a known protein, the first Enzyme Commissin digit (EC digit) can be accurately assigned which meant that whether a protein is an oxidoreductase or a ligase can be fairly and correctly predicted. However, in proteins with even at 70% sequence identity, the fourth EC digit which specifies substrate specificity can be and has been wrongly assigned and this fact has been confirmed by other studies too [140] . There have been cases of mis-annotataion before [141] and thus the presence of nucleases/kinases amongst the annotated proteins cannot be ruled out and would be worth studying. Analyzing all these proteins for nuclease or kinase activty experimentally would have required high throughput protein purification and assay systems, which would be time consuming and purifying a protein from a complex collection is equally demanding. Hence we relied on other alternative i.e. *insilico* analysis. Most nuclease and kinase domains are well characterized and are cataloged in databases. Since MASCOT could not find any nuclease/kinase in pool of proteins exhibiting ATP sensitive nuclease functions in our earlier studies, we focussed on the structure-function prediction of the 11 hypothetical proteins also found in this fraction.

The set of criteria we used for analyzing the data was (1) protein sequence analysis (2) genomic context (3) subcellular location (4) peptide abundance in FTICR-MS data (5) theoretically modeled structure and (6) phylogenetic relationships. In this study, we employed both bioinformatics as well as experimental approaches for functional characterization of both these proteins.

Our experience showed that the output obtained with each individual analysis was not seamless. For example, the PSI-BLAST results indicated DRA0282 as Ku type protein but the structure prediction servers did not predict DRA0282 as Ku type protein but as human fibronectin (fibronectin type III domains bind DNA) and receptor protein tyrosine phosphatase type protein. The best analysis came when we combined results from the sequence search with the structure data and phylogenetic tree information. Based on initial BLAST searches, we had presumed that we are looking into components of NHEJ pathway and accordingly our intent was to see whether indeed these initial searches were true. For both DR2417 and DRA0282 this method proved effective and provided us with testable predictions but for DR0390, we could only predict its role as a kinase with possible role in binding to lipid substrates and it failed to predict any functional role for DR1654 save it being possibly a membrane bound protein. Both DRA0282 and DR2417 were found in a cell lysate fraction at 1 hour PIR phase and both these proteins were unannotated and hence their study assumed importance. Our bioinformatics results showed that these proteins may have a role in DNA processing. DR2417 showed homology to Artemis, a nuclease involved in processing noncompatible ends for NHEJ and DRA0282 was homologous with Ku80 which too was involved in the same pathway. Artemis is an important DNA repair enzyme which modifies its activity from a 5' ssDNA exonuclease to a hairpin endonuclease. We would have expected DR2417 to show a similar kinase dependent switch from 5' single stranded DNA exonuclease to a hairpin endonuclease in the presence of DRA0282 and DR0390 (presuming it had DNA-PKcs activity, which it may not have). Our assays showed that DR2417 was a double stranded DNA exonulease as well as a hairpin endonuclease but it did both these activities without

being modulated by any other protein and our preliminary investigation showed that DRA0282 hindered this nuclease activity rather than modulating it. At the same time DR2417 was inhibited by ATP which could be easily explained by the 3-D model obtained. However, the model also suggested that DR2417 should act on single stranded DNA or RNA, but our biochemical assays show that it does not degrade single stranded DNA or RNA. More importantly, earlier bioinformatics studies on β-CASP family of nucleases concluded that valine at motif C contributed to the discrimination between RNase and DNase. Our study has adequate proof that motif C is inconsequential to RNase/DNase discrimination and the reason for this lies elsewhere. Indirectly we assume that the biochemical activity could be explained if DR2417 acts exclusively as a dimer. Also, it may be well worth noting that the closest homolog of DR2417 is RNAse J from *Thermus thermophilus* which functions an RNAse in the closely related Thermus species. DR2417 is the first reported DNase of the β-CASP family amongst prokaryotes and this assumes importance as members of this family are present in more than 200 species of gram positive bacteria. Our studies indicate that possibly in all other species β-CASP nucleases act as RNases, while in *Deinococcus* spp. they have evolved into DNases. A crucial part of the evolution would be a dimer interface which would not allow single stranded DNA but would become relaxed enough for double stranded DNA in its presence. Artemis null mutants of mice do not survive due to a compromised immune system as NHEJ pathway is required for T and B cell maturation, but this mutation is not lethal *per se*. We found that Δ*dr2417* is possibly a lethal mutation in *D. radiodurans* and thus DR2417 could have an essential role in the lifecycle of *D. radiodurans*. It was difficult to study the biological effects as cells lacking the gene would die off. But

nevertheless, its nuclease activities are consistent with its role during early phase repair where extensive double stranded DNase activity converts the double stranded DNA to extensive tracts of single stranded DNA which is then paired with homologous strands in a *recA* dependent pathway. We also overlooked a crucial piece of evidence from our own study. Hydrophobic Cluster Analysis plots (HCA plots) highlight regions of irregular secondary structure. Our HCA plot showed that the C-terminal of DR2417 is devoid of regular secondary structure with an excess of proline and arginine residues as seen in pseudogenes. This fact should have alerted us , instead we found a valine at a suitable position in the multiple sequence alignment. Since DR2417 was a DNase, we assumed that this valine is the site in motif C. The earlier studies had provided circumstantial evidence for valine at motif C being discriminatory for DNase/RNase without providing any structural/biochemical evidence. We took this evidence at face value, which now we find was erroneous. Nevertheless, resequencing the *dr2417* ORF and Peptide Mass Fingerprinting of the protein corrected this oversight.

We identified DRA0282 as a Ku type protein. The reputable automated structure prediction server ITASSER predicted DRA0282 as a functional homolog of a sialidase from *Micromonospora viridifaciens* (PDB id 1EUT) which is quite different from a DNA binding protein. But we combined our PSI-BLAST search and multiple sequence alignment with Ku homologs and built 3-D model with Ku80 (PDB id1JEY). A close inspection of both models of this proteins reveal a central β barrel domain, a N-terminal β sheet roll, thus although the structure was similar but function was different. Our biochemical studies show that DRA0282 did indeed bind to DNA and also protected ends of double stranded DNA from Exo III nuclease activity. But, crucially, it did not enhance

intermolecular or intramolecular ligation and neither did it promote endjoining of noncompatible ends *invitro*. Also DRA0282 had no ATPase activity like Ku rather it was inhibited by ATP and although 3-D modeling showed that it had a clearly defined ATP binding domain it was not clear how ATP binding would inhibit DNA binding. Curiously DRA0282 shows highest affinity for single stranded DNA followed by supercoiled DNA and shows a much lower affinity for linear double stranded DNA. This is in contrast to Ku70/80 dimer in mammals or Ku type proteins in bacteria which show preference to linear double stranded DNA [142]. Also phylogenetic studies show that DRA0282, matches more closely with eukaryotic Ku80 and has only 7 orthologs in bacteria. This would mean that its Ku like structure and function is a case of convergent evolution. Our molecular genetic studies on DRA0282 have revealed apparently contrasting results in the response of two different bacterial species to DNA damage. While transgenic *E. coli* expressing this protein show enhanced protection to UV, the *D. radiodurans* deletion mutants show a marginally improved survival. Deleting this gene causes no apparent loss in survival of *D. radiodurans* or its radiation resistance. We have some evidence from *E. coli* that DRA0282 functionally interacts with RecA in its physiological milieu *in vivo*. The mechanism of *E. coli* RecA and *D. radiodurans* RecA is different as *D. radiodurans* RecA has preference for double stranded DNA while *E. coli* has preference for single stranded DNA [143]. Also DRA0282 has higher affinity for single stranded DNA compared to double stranded DNA [144]. Although it is mere speculation, possibly this affinity for single stranded DNA allows DRA0282 to interact better with *E. coli* RecA and enhance the homologous recombination pathway in *E. coli*, while in *D. radiodurans* its role may be to prevent DNA from premature attack by exonucleases prior to being

coated by RecA, akin to SSB. The 3-D model suggests that like Ku, DRA0282 encloses DNA in a ring structure which can only be released by a cleaving the protein. Interestingly there is a serine protease DRA0283 just upstream to DRA0282. It has been earlier shown that this protease undergoes phosphorylation *in vivo* and is under the control of PprI a global switch for radioresistance [145]. Although the null mutant of DRA0282 was not debilitated earlier in survival or in response to γ radiation, this protein has been found in many proteomics studies on *D. radiodurans* [137,145,146]. DRA0282 may play a redundant role, or may have an important role under environmental conditions which cannot be mimicked under laboratory conditions.

Hence, from our studies we would like to conclude that both these proteins should be involved in DNA metabolism especially in the repair process. Recently, it has been shown that both DR2417 and DRA0282 are associated with the *D. radiodurans* nucleoid, the condensed structures which are formed during stationary phase and especially after extensive DNA damage [112]. Toroidal structure of *D. radiodurans* genome may be an added advantage to extreme phenotypes of this

bacterium has not been ruled out. Thus it corroborates our results that DRA0282 and DR2417 are DNA metabolic genes involved in maintenance of genome integrity. It would indeed be interesting to further unravel the real time functions of these two proteins in the metabolism of the cell.

Chapter 5

Summary, Conclusion and Future Perspectives

Summary

D. radiodurans was initially isolated in 1956 from canned meat treated to 4 kGy radiation dose and classified as *Micrococcus spp*.[2]. This dose is 250 times more than the tolerance level of the gut bacterium *E. coli*. Over the next 60 odd years, the molecular basis of its unusual radiation resistance has been investigated and several hypotheses have been propounded. These include the role of antioxidants , the compactness and toroidal structure of genome, protection of double strand breaks (DSBs) from nucleolytic degradation and an efficient DSB repair pathway which now has been shown to be mainly facilitated by ESDSA followed by RecF pathway of recombination [34]. While all these have independently proven in contributing to extreme radiation resistance of this bacterium, the concerted action of these molecular events in radioresistance has not been sufficiently detailed. In this thesis, a fraction of DNA binding proteins showing ATP sensitive nuclease activity and autophosphorylation of several proteins was subjected to mass spectrometric identification of its protein components. Interestingly, none of the known nucleases and kinases were found in a pool of 300 polypeptides detected by FTICR-MS. Out of the 11 hypothetical proteins present in this fraction, 4 were analysed using various tools of bioinformatics for structure and function predictions. Two of these DRA0282 and DR2417 showed well defined functional motifs relating to DNA metabolism while another DR0390 was having putative motifs for a kinase. DRA0282 and DR2417 proteins were found structurally close to Ku80 and Artemis type nuclease, respectively, while DR0390 was closer to DAK/DegV kinase. DRA0282 and DR2417 were further characterized *in vivo* and *in vitro* for the known functions related to their

closer homologues. DRA0282 was a DNA binding protein, a function that was found to be sensitive to ATP addition. Quite unlike Ku type proteins, DRA0282 showed higher affinity to supercoiled double stranded DNA followed by a lower affinity to single stranded DNA and the lowest affinity to linear dsDNA amongst these topologically different substrates. This protein also showed DNA protection from exonuclease degradation as observed with Ku type homologues [113]. The role of this protein in γ radiation resistance of *D. radiodurans* was redundant as its deletion mutant did not show change in DNA damage response of wild type while absence of this protein made *D. radiodurans* more tolerant to UVC as compared to wild type. This result is intriguing. Interestingly, deletion of *drA0282* from the genome of PprA (another deinococcal protein with nearly similar function *in vitro*) mutant of *D. radiodurans*, did not add further to pprA mutant's radiation sensitivity. Further, over expression of this protein increased both γ and UVC resistance of transgenic *E. coli*. This phenotype in *E. coli* required a recombinationally functional RecA and was independent of SOS response. These results indicated that DRA0282 functions in tandem with RecA in recombination repair of DNA damage at least in *E. coli* while it has functional homologues in *D. radiodurans*. Its function in tandem with RecA in *D. radiodurans* will be investigated independently.

DR2417m was wrongly annotated with a so called genuine frameshift questioning the functional authenticity of this protein in *D. radiodurans*. However, this protein was consistently detected in the functional pools of protein in independent studies. DR2417m was sequenced and the mistakes of initial annotation were corrected. DR2417 showed perfect match with proteins of β-CASP family of which Artemis is a member. Recombinant protein was made and *in vitro* activities were checked. DR2417 showed

both exonculease and ss/dsDNA junction endonuclease on artificially synthesized heteropolymeric DNA substrates. DR2417 exhibited DNA processing activity albeit 3'-5' exonucelase, which are integral to efficient DSB repair as observed in *D. radiodurans*. Quite surprisingly, inspite of having very similar structure with conserved motifs as that of RNase J from *Thermus spp.* and *Bacillus spp.*, and other β-CASP family nuclease, this protein showed poor RNase activity. This result is further interesting but intriguing. Molecular genetic studies confirmed that *dr2417* is an essential gene in *D. radiodurans* as dose dependent reduction in copy number correlated very well with loss of cell survival and γ radiation resistance.

In this study we have used integrated approach of bioinformatics tools for prediction of structure and functions of four proteins selected from the pool of 11 hypothetical proteins and further validated the predicted functions of two of these proteins experimentally. The evidences gathered on the various activities of these proteins have helped us in assigning new functions to two of the hypothetical proteins from nearly 50% of uncharacterized proteins of *D. radiodurans*. Apparently, these functions are found to be integral to the DNA metabolism reported during recombination repair of DNA double strand breaks, for which *D. radiodurans* is best characterized and is being studied.

Conclusion

Our study has accentuated the role of bioinformatics in finding relevant proteins in a complex mixture of proteins undertaking a certain biochemical activity in this case e.g. ATP sensitive nucleolytic function. Extrapolation is avoided in science and bioinformatics is no exception to this. While learning these facts we could assign novel

functions to two hypothetical proteins in *D. radiodurans* genome and corrected one erroneous annotation during the course of this study. Another key finding has been that one cannot rely heavily on a single bioinformatic tool to decipher the function of novel proteins rather an integrated approaches and authentication by wet biology is must. Thus out of 300 proteins, we could home onto a single nuclease DR2417 and another DNA binding protein DRA0282 for its similarity to Ku80. The findings of this study led to conclude the following:

- On the basis of multiple sequence alignments, phylogenetic trees and 3-D homology model, we predicted that DR2417 and DRA0282 are functional counterparts of Artemis and Ku 70/80, respectively
- DR2417 is a Mn^{2+} dependent DNase and DRA0282 is a DNA end protecting protein.
- DRA0282 is a Mn^{2+} dependent DNA end protecting protein having higher affinity to supercoiled DNA and single stranded DNA as compared to linear double stranded DNA.
- Both these proteins display unique functional properties like 3'-5' double stranded exonuclease activity and endonuclease with propensity to single stranded-double stranded DNA junctions for DR2417 and protection of linear double stranded DNA from exonucelase III digestion and higher affinity for supercoiled DNA by DRA0282.
- Molecular genetic studies show that DR2417 appears to be an essential gene in *D. radiodurans* while much effect was not seen in *E. coli*.DRA0282 helps

significantly in UV and γ radiation tolerance of *E. coli* but has nearly no effect on γ radiation response while it seems to have a role in UVC resistance in *D. radiodurans*

Future Perspectives

As we conc*lude th*is thesis, it would not be inappropriate to raise the questions that we have shied away from- was it possible for the already annotated protein to have an additional nuclease/kinase function and not the hypothetical proteins? Yes. That possibility existed and a way to analyze this would be a study to take all the sequences and analyze for known nuclease/kinase motifs by writing user defined programs. What we have undertaken here is the approach of using available bioinformatics tools with a biologists 'common sense' approach. This had its limitations in that no clear prediction was possible for DR0390 and DR1654. But all the same, our results allow us to peer a little deeper into the enigmatic life of *D. radiodurans*. While this study has answered some of the questions and hypotheses raised both from the experimental data of earlier work and bioinformatic analyses of certain proteins, it has left several unanswered questions.

1. DR2417 almost is unique amongst all β-CASP family of nucleases in its weak exonuclease activity on single stranded DNA or RNA although at the same time it is able to act upon gross single stranded Ø X174 DNA to form island products. From the structure- function point of view this aspect is certainly an exciting area of research.

During the course of this study we gained evidence that *dr2417* might be an essential gene. Earlier RecJ, a 5'-3' exonuclease was shown to be essential for *D. radiodurans*, showing that crossover mechanisms involving single stranded DNA are a part of the lifecycle in *D. radiodurans*. Rec J has been postulated to work in the RecFOR pathway but cells expressing SbcB which degrades 3' ssDNA and eventually arrests RecFOR, grow normally like the wild type [100] and hence *D. radiodurans* may have another single stranded DNA dependent mechanism involving Rec J and possibly DR2417 working in tandem. Thus the exact physiological role played by DR2417 is well worth investigating. Recombinant DR2417 protects DNA from thermal and pH denaturation. The mechanism underlying this characteristic would also be worth elucidating.

2. DRA0282 is omnipresent and has been shown to be a part of the nucleoid [110]. As described earlier, absence of this gene product does not hamper the survival of *D. radiodurans* to radiation stress in rich medium. The survival and DNA repair of the Δ*drA0282* deletion mutant under physiological stresses like dessication would be interesting as the DdrA deletion mutant shows a debilitated phenotype only under dessication [62]. DRA0282 shows preference to supercoiled DNA, followed by single stranded DNA and protection of linear double stranded DNA from exonuclease degradation and this protein also function in tandem with recombination function of RecA in *E. coli*. It is possible that this protein in brings the interaction of supercoiled DNA with single stranded DNA. The functional significance of this interaction by DRA0282, in the context of efficient DSB repair would be interesting to ascertain.

3.DR0390 is predicted to be a kinase. The activity characterization of this protein and significance if any, of this protein being together with DRA0282 and DR2417 would be

interesting to study. Last but not the least, further studies on functional characterization of DR0390 keeping in mind its role as kinase and its involvement in radiation resistance may be investigated.

Reference List

1. Mora C, Tittensor DP, Adl S, Simpson AG, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biol 9: e1001127.

2. Duggan DE, Anderson AW, Elliker PR (1963) Inactivation of the Radiation-Resistant Spoilage Bacterium *Micrococcus radiodurans*. Ii. Radiation Inactivation Rates as Influenced by Menstruum Temperature, Preirradiation Heat Treatment, and Certain Reducing Agents. Appl Microbiol 11: 413-417.

3. Yuan M, Zhang W, Dai S, Wu J, Wang Y, et al. (2009) *Deinococcus gobiensis* sp. nov., an extremely radiation-resistant bacterium. Int J Syst Evol Microbiol 59: 1513-1517.

4. Yang Y, Itoh T, Yokobori S, Shimada H, Itahashi S, et al. (2010) *Deinococcus aetherius* sp. nov., isolated from the stratosphere. Int J Syst Evol Microbiol 60: 776-779.

5. Slade D, Radman M (2011) Oxidative stress resistance in *Deinococcus radiodurans*. Microbiol Mol Biol Rev 75: 133-191.

6. Sedelnikova OA, Redon CE, Dickey JS, Nakamura AJ, Georgakilas AG, et al. (2010) Role of oxidatively induced DNA lesions in human pathogenesis. Mutat Res 704: 152-159.

7. Battista JR (2000) Radiation resistance: the fragments that remain. Curr Biol 10: R204-205.

8. Griffiths E, Gupta RS (2004) Distinctive protein signatures provide molecular markers and evidence for the monophyletic nature of the deinococcus-thermus phylum. J Bacteriol 186: 3097-3107.

9. Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, et al. (2005) Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. BMC Evol Biol 5: 57.

10. Daly MJ (2009) A new perspective on radiation resistance based on *Deinococcus radiodurans*. Nat Rev Microbiol 7: 237-245.

11. Cantu D, Schaack J, Patel M (2009) Oxidative inactivation of mitochondrial aconitase results in iron and H2O2-mediated neurotoxicity in rat primary mesencephalic cultures. PLoS One 4: e7095.

12. Misra HS, Khairnar NP, Barik A, Indira Priyadarsini K, Mohan H, et al. (2004) Pyrroloquinoline-quinone: a reactive oxygen species scavenger in bacteria. FEBS Lett 578: 26-30.

13. Khairnar NP, Misra HS, Apte SK (2003) Pyrroloquinoline-quinone synthesized in Escherichia coli by pyrroloquinoline-quinone synthase of *Deinococcus radiodurans* plays a role beyond mineral phosphate solubilization. Biochem Biophys Res Commun 312: 303-308.

14. Lavy A, Neeman Y, Fuhrman B (2005) The antioxidative effect of the bacteria *Dienococcus radiophilus* against LDL lipid peroxidation. Eur J Nutr 44: 281-284.

15. Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, et al. (2004) Accumulation of Mn(II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance. Science 306: 1025-1028.

16. Krisko A, Radman M (2010) Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. Proc Natl Acad Sci U S A 107: 14373-14377.

17. Joshi B, Schmid R, Altendorf K, Apte SK (2004) Protein recycling is a major component of post-irradiation recovery in *Deinococcus radiodurans* strain R1. Biochem Biophys Res Commun 320: 1112-1117.

18. Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, et al. (1998) A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature 391: 43-50.

19. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, et al. (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286: 1571-1577.

20. Martin SA, Lord CJ, Ashworth A (2008) DNA repair deficiency as a therapeutic target in cancer. Curr Opin Genet Dev 18: 80-86.

21. Shuck SC, Short EA, Turchi JJ (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. Cell Res 18: 64-72.

22. Pham P, Rangarajan S, Woodgate R, Goodman MF (2001) Roles of DNA polymerases V and II in SOS-induced error-prone and error-free repair in *Escherichia coli*. Proc Natl Acad Sci U S A 98: 8350-8354.

23. Aguilera A, Boulton SJ (2007) How to exchange your partner. Workshop on recombination mechanisms and the maintenance of genomic stability. EMBO Rep 8: 28-33.

24. Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev 72: 642-671, Table of Contents.

25. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, et al. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297: 1686-1689.

26. Englander J, Klein E, Brumfeld V, Sharma AK, Doherty AJ, et al. (2004) DNA toroids: framework for DNA repair in *Deinococcus radiodurans* and in germinating bacterial spores. J Bacteriol 186: 5973-5977.

27. Zimmerman JM, Battista JR (2005) A ring-like nucleoid is not necessary for radioresistance in the Deinococcaceae. BMC Microbiol 5: 17.

28. Daly MJ, Gaidamakova EK, Matrosova VY, Kiang JG, Fukumoto R, et al. (2010) Small-molecule antioxidant proteome-shields in *Deinococcus radiodurans*. PLoS One 5: e12570.

29. Daly MJ, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, et al. (2007) Protein oxidation implicated as the primary determinant of bacterial radioresistance. PLoS Biol 5: e92.

30. Gutman PD, Fuchs P, Minton KW (1994) Restoration of the DNA damage resistance of *Deinococcus radiodurans* DNA polymerase mutants by *Escherichia coli* DNA polymerase I and Klenow fragment. Mutat Res 314: 87-97.

31. Varghese AJ, Day RS, 3rd (1970) Excision of cytosine-thymine adduct from the DNA of ultraviolet-irradiated *Micrococcus radiodurans*. Photochem Photobiol 11: 511-517.

32. Bentchikou E, Servant P, Coste G, Sommer S (2010) A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans*. PLoS Genet 6: e1000774.

33. Zahradka K, Slade D, Bailone A, Sommer S, Averbeck D, et al. (2006) Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature 443: 569-573.

34. Slade D, Lindner AB, Paul G, Radman M (2009) Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. Cell 136: 1044-1055.

35. McVey M, Adams M, Staeva-Vieira E, Sekelsky JJ (2004) Evidence for multiple cycles of strand invasion during repair of double-strand gaps in Drosophila. Genetics 167: 699-705.

36. Paques F, Leung WY, Haber JE (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol Cell Biol 18: 2045-2054.

37. Mueller JE, Clyman J, Huang YJ, Parker MM, Belfort M (1996) Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. Genes Dev 10: 351-364.

38. Holloman WK, Schirawski J, Holliday R (2007) Towards understanding the extreme radiation resistance of *Ustilago maydis*. Trends Microbiol 15: 525-529.

39. Bennett CB, Lewis LK, Karthikeyan G, Lobachev KS, Jin YH, et al. (2001) Genes required for ionizing radiation resistance in yeast. Nat Genet 29: 426-434.

40. Kamble VA, Misra HS (2010) The SbcCD complex of *Deinococcus radiodurans* contributes to radioresistance and DNA strand break repair in vivo and exhibits Mre11-Rad50 type activity in vitro. DNA Repair (Amst) 9: 488-494.

41. Aylon Y, Kupiec M (2004) DSB repair: the yeast paradigm. DNA Repair (Amst) 3: 797-815.

42. Zou L, Elledge SJ (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science 300: 1542-1548.

43. Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, et al. (2010) Irradiation-induced Deinococcus radiodurans genome fragmentation triggers transposition of a single resident insertion sequence. PLoS Genet 6: e1000799.

44. Cox MM, Battista JR (2005) *Deinococcus radiodurans* - the consummate survivor. Nat Rev Microbiol 3: 882-892.

45. Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, et al. (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Proc Natl Acad Sci U S A 100: 4191-4196.

46. de Groot A, Dulermo R, Ortet P, Blanchard L, Guerin P, et al. (2009) Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. PLoS Genet 5: e1000434.

47. Davids W, Fuxelius HH, Andersson SG (2003) The Journey to smORFland. Comp Funct Genomics 4: 537-541.

48. Siew N, Fischer D (2003) Unravelling the ORFan Puzzle. Comp Funct Genomics 4: 432-441.

49. Nishikimi M, Fukuyama R, Minoshima S, Shimizu N, Yagi K (1994) Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. J Biol Chem 269: 13685-13688.

50. Xue Y, Daly A, Yngvadottir B, Liu M, Coop G, et al. (2006) Spread of an inactive form of caspase-12 in humans is due to recent positive selection. Am J Hum Genet 78: 659-670.

51. Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12: 692-702.

52. Pearl F, Todd A, Sillitoe I, Dibley M, Redfern O, et al. (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res 33: D247-251.

53. Tanaka A, Hirano H, Kikuchi M, Kitayama S, Watanabe H (1996) Changes in cellular proteins of *Deinococcus radiodurans* following gamma-irradiation. Radiat Environ Biophys 35: 95-99.

54. Harris DR, Tanaka M, Saveliev SV, Jolivet E, Earl AM, et al. (2004) Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1. PLoS Biol 2: e304.

55. Narumi I, Satoh K, Cui S, Funayama T, Kitayama S, et al. (2004) PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. Mol Microbiol 54: 278-285.

56. Earl AM, Mohundro MM, Mian IS, Battista JR (2002) The IrrE protein of *Deinococcus radiodurans* R1 is a novel regulator of recA expression. J Bacteriol 184: 6216-6224.

57. Sun H, Xu G, Zhan H, Chen H, Sun Z, et al. (2010) Identification and evaluation of the role of the manganese efflux protein in *Deinococcus radiodurans*. BMC Microbiol 10: 319.

58. Schmid AK, Howell HA, Battista JR, Peterson SN, Lidstrom ME (2005) Global transcriptional and proteomic analysis of the Sig1 heat shock regulon of *Deinococcus radiodurans*. J Bacteriol 187: 3339-3351.

59. Schmid AK, Lidstrom ME (2002) Involvement of two putative alternative sigma factors in stress response of the radioresistant bacterium *Deinococcus radiodurans*. J Bacteriol 184: 6182-6189.

60. Satoh K, Ohba H, Sghaier H, Narumi I (2006) Down-regulation of radioresistance by LexA2 in *Deinococcus radiodurans*. Microbiology 152: 3217-3226.

61. Martins A, Shuman S (2004) An RNA ligase from *Deinococcus radiodurans*. J Biol Chem 279: 50654-50661.

62. Xu G, Lu H, Wang L, Chen H, Xu Z, et al. (2010) DdrB stimulates single-stranded DNA annealing and facilitates RecA-independent DNA repair in *Deinococcus radiodurans*. DNA Repair (Amst) 9: 805-812.

63. Leiros HK, Kozielski-Stuhrmann S, Kapp U, Terradot L, Leonard GA, et al. (2004) Structural basis of 5-nitroimidazole antibiotic resistance: the crystal structure of NimA from *Deinococcus radiodurans*. J Biol Chem 279: 55840-55849.

64. Wang L, Xu G, Chen H, Zhao Y, Xu N, et al. (2008) DrRRA: a novel response regulator essential for the extreme radioresistance of *Deinococcus radiodurans*. Mol Microbiol 67: 1211-1222.

65. Blasius M, Buob R, Shevelev IV, Hubscher U (2007) Enzymes involved in DNA ligation and end-healing in the radioresistant bacterium *Deinococcus radiodurans*. BMC Mol Biol 8: 69.

66. Chen H, Xu G, Zhao Y, Tian B, Lu H, et al. (2008) A novel OxyR sensor and regulator of hydrogen peroxide stress with one cysteine residue in *Deinococcus radiodurans*. PLoS One 3: e1602.

67. Kota S, Kumar CV, Misra HS (2010) Characterization of an ATP-regulated DNA-processing enzyme and thermotolerant phosphoesterase in the radioresistant bacterium *Deinococcus radiodurans*. Biochem J 431: 149-157.

68. Sun Z, Shen S, Wang C, Wang H, Hu Y, et al. (2009) A novel carotenoid 1,2-hydratase (CruF) from two species of the non-photosynthetic bacterium *Deinococcus*. Microbiology 155: 2775-2783.

69. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

70. Washietl S, Hofacker I (2010) Data Miniing Techniques for the Life Sciences; Carugo O, Eisenhaber F, editors. New York: Humana Press.

71. Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232: 584-599.

72. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291-325.

73. Grant A, Lee D, Orengo C (2004) Progress towards mapping the universe of protein folds. Genome Biol 5: 107.

74. Lopez MD, Samuelsson T (2011) eGOB: eukaryotic Gene Order Browser. Bioinformatics 27: 1150-1151.

75. Eisen JA, Kaiser D, Myers RM (1997) Gastrogenomic delights: a movable feast. Nat Med 3: 1076-1078.

76. Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet 6: 361-375.

77. Sambrook J, Russell D (2001) Molecular Cloning : A Laboratory manual. New York: Cold Spring Harbor Laboratory Press.

78. Schafer M, Schmitz C, Horneck G (1998) High sensitivity of *Deinococcus radiodurans* to photodynamically-produced singlet oxygen. Int J Radiat Biol 74: 249-253.

79. Mount DW, Low KB, Edmiston SJ (1972) Dominant mutations (lex) in *Escherichia coli* K-12 which affect radiation sensitivity and frequency of ultraviolet lght-induced mutations. J Bacteriol 112: 886-893.

80. Buxbaum E, Schoner W (1990) Blocking of Na+/K+ transport by the MgPO4 complex analogue Co(NH3)4PO4 leaves the Na+/Na(+)-exchange reaction of the sodium pump unaltered and shifts its high-affinity ATP-binding site to a Na(+)-like form. Eur J Biochem 193: 355-360.

81. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254.

82. Kamble VA, Rajpurohit YS, Srivastava AK, Misra HS (2010) Increased synthesis of signaling molecules coincides with reversible inhibition of nucleolytic activity during postirradiation recovery of *Deinococcus radiodurans*. FEMS Microbiol Lett 303: 18-25.

83. Das AD, Misra HS (2011) Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage. Microbiology 157: 2196-2205.

84. Nishino T, Morikawa K (2002) Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. Oncogene 21: 9022-9032.

85. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26: 1608-1615.

86. Gaboriaud C, Uze G, Lutfalla G, Mogensen K (1990) Hydrophobic cluster analysis reveals duplication in the external structure of human alpha-interferon receptor and homology with gamma-interferon receptor external domain. FEBS Lett 269: 1-3.

87. Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244.

88. Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28: 405-420.

89. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738.

90. 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.

91. Rooney S, Sekiguchi J, Zhu C, Cheng HL, Manis J, et al. (2002) Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice. Mol Cell 10: 1379-1390.

92. de Villartay JP, Shimazaki N, Charbonnier JB, Fischer A, Mornon JP, et al. (2009) A histidine in the beta-CASP domain of Artemis is critical for its full in vitro and in vivo functions. DNA Repair (Amst) 8: 202-208.

93. Lieber MR, Ma Y, Pannicke U, Schwarz K (2003) Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Biol 4: 712-720.

94. Featherstone C, Jackson SP (1999) Ku, a DNA repair protein with multiple cellular functions? Mutat Res 434: 3-15.

95. Pitcher RS, Brissett NC, Doherty AJ (2007) Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol 61: 259-282.

96. Goodarzi AA, Yu Y, Riballo E, Douglas P, Walker SA, et al. (2006) DNA-PK autophosphorylation facilitates Artemis endonuclease activity. Embo J 25: 3880-3889.

97. Siebold C, Arnold I, Garcia-Alles LF, Baumann U, Erni B (2003) Crystal structure of the Citrobacter freundii dihydroxyacetone kinase reveals an eight-stranded alpha-helical barrel ATP-binding domain. J Biol Chem 278: 48236-48244.

98. Schulze-Gahmen U, Pelaschier J, Yokota H, Kim R, Kim SH (2003) Crystal structure of a hypothetical protein, TM841 of Thermotoga maritima, reveals its function as a fatty acid-binding protein. Proteins 50: 526-530.

99. Klein P, Somorjai RL, Lau PC (1988) Distinctive properties of signal sequences from bacterial lipoproteins. Protein Eng 2: 15-20.

100. Terradot L, Bayliss R, Oomen C, Leonard GA, Baron C, et al. (2005) Structures of two core subunits of the bacterial type IV secretion system, VirB8 from *Brucella suis* and ComB10 from *Helicobacter pylori*. Proc Natl Acad Sci U S A 102: 4596-4601.

101. Gong C, Bongiorno P, Martins A, Stephanou NC, Zhu H, et al. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. Nat Struct Mol Biol 12: 304-312.

102. Gullo C, Au M, Feng G, Teoh G (2006) The biology of Ku and its potential oncogenic role in cancer. Biochim Biophys Acta 1765: 223-234.

103. Monferran S, Paupert J, Dauvillier S, Salles B, Muller C (2004) The membrane form of the DNA repair protein Ku interacts at the cell surface with metalloproteinase 9. Embo J 23: 3758-3768.

104. Misra HS, Khairnar NP, Kota S, Shrivastava S, Joshi VP, et al. (2006) An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance. Mol Microbiol 59: 1308-1316.

105. Khairnar NP, Kamble VA, Misra HS (2008) RecBC enzyme overproduction affects UV and gamma radiation survival of *Deinococcus radiodurans*. DNA Repair (Amst) 7: 40-47.

106. Khairnar NP, Kamble VA, Mangoli SH, Apte SK, Misra HS (2007) Involvement of a periplasmic protein kinase in DNA strand break repair and homologous recombination in *Escherichia coli*. Mol Microbiol 65: 294-304.

107. Rajpurohit YS, Gopalakrishnan R, Misra HS (2008) Involvement of a protein kinase activity inducer in DNA double strand break repair and radioresistance of *Deinococcus radiodurans*. J Bacteriol 190: 3948-3954.

108. Wang J, Julin DA (2004) DNA helicase activity of the RecD protein from *Deinococcus radiodurans*. J Biol Chem 279: 52024-52032.

109. Abramoff MD, Viergever MA (2002) Computation and visualization of three-dimensional soft tissue motion in the orbit. IEEE Trans Med Imaging 21: 296-304.

110. O'Donovan A, Scherly D, Clarkson SG, Wood RD (1994) Isolation of active recombinant XPG protein, a human DNA repair endonuclease. J Biol Chem 269: 15965-15968.

111. Hiom K (1999) Dna repair: Rad52 - the means to an end. Curr Biol 9: R446-448.

112. Toueille M, Mirabella B, Guerin P, Bouthier de la Tour C, Boisnard S, et al. (2012) A comparative proteomic approach to better define *Deinococcus* nucleoid specificities. J Proteomics 75: 2588-2600.

113. Downs JA, Jackson SP (2004) A means to a DNA end: the many roles of Ku. Nat Rev Mol Cell Biol 5: 367-378.

114. Lee SH, Oshige M, Durant ST, Rasila KK, Williamson EA, et al. (2005) The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc Natl Acad Sci U S A 102: 18075-18080.

115. Zhang N, Kaur R, Lu X, Shen X, Li L, et al. (2005) The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. J Biol Chem 280: 40559-40567.

116. Lu X, Legerski RJ (2007) The Prp19/Pso4 core complex undergoes ubiquitylation and structural alterations in response to DNA damage. Biochem Biophys Res Commun 354: 968-974.

117. Beck BD, Park SJ, Lee YJ, Roman Y, Hromas RA, et al. (2008) Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. J Biol Chem 283: 9023-9030.

118. Mahajan KN, Mitchell BS (2003) Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. Proc Natl Acad Sci U S A 100: 10746-10751.

119. Chayot R, Montagne B, Mazel D, Ricchetti M (2010) An end-joining repair mechanism in *Escherichia coli*. Proc Natl Acad Sci U S A 107: 2141-2146.

120. Callebaut I, Moshous D, Mornon JP, de Villartay JP (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res 30: 3592-3601.

121. Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30: e36.

122. Kota S, Misra HS (2006) PprA: A protein implicated in radioresistance of *Deinococcus radiodurans* stimulates catalase activity in *Escherichia coli*. Appl Microbiol Biotechnol 72: 790-796.

123. Mathy N, Hebert A, Mervelet P, Benard L, Dorleans A, et al. (2010) *Bacillus subtilis* ribonucleases J1 and J2 form a complex with altered enzyme behaviour. Mol Microbiol 75: 489-498.

124. Connelly JC, de Leau ES, Leach DR (2003) Nucleolytic processing of a protein-bound DNA end by the *E. coli* SbcCD (MR) complex. DNA Repair (Amst) 2: 795-807.

125. D'Amours D, Jackson SP (2002) The Mre11 complex: at the crossroads of DNA repair and checkpoint signalling. Nat Rev Mol Cell Biol 3: 317-327.

126. Clouet-d'Orval B, Rinaldi D, Quentin Y, Carpousis AJ (2010) Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. J Biol Chem 285: 17574-17583.

127. Ghosal D, Omelchenko MV, Gaidamakova EK, Matrosova VY, Vasilenko A, et al. (2005) How radiation kills cells: survival of *Deinococcus radiodurans* and *Shewanella oneidensis* under oxidative stress. FEMS Microbiol Rev 29: 361-375.

128. Shashidhar R, Kumar SA, Misra HS, Bandekar JR (2010) Evaluation of the role of enzymatic and nonenzymatic antioxidant systems in the radiation resistance of *Deinococcus*. Can J Microbiol 56: 195-201.

129. Bentchikou E, Servant P, Coste G, Sommer S (2007) Additive effects of SbcCD and PolX deficiencies in the in vivo repair of DNA double-strand breaks in *Deinococcus radiodurans*. J Bacteriol 189: 4784-4790.

130. Cao Z, Julin DA (2009) Characterization in vitro and in vivo of the DNA helicase encoded by *Deinococcus radiodurans* locus DR1572. DNA Repair (Amst) 8: 612-619.

131. Kota S, Misra HS (2008) Identification of a DNA processing complex from *Deinococcus radiodurans*. Biochem Cell Biol 86: 448-458.

132. Cao Z, Mueller CW, Julin DA (2010) Analysis of the recJ gene and protein from *Deinococcus radiodurans*. DNA Repair (Amst) 9: 66-75.

133. Dorleans A, Li de la Sierra-Gallay I, Piton J, Zig L, Gilet L, et al. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. Structure 19: 1252-1261.

134. Schafer M, Schmitz C, Facius R, Horneck G, Milow B, et al. (2000) Systematic study of parameters influencing the action of Rose Bengal with visible light on bacterial cells: comparison between the biological effect and singlet-oxygen production. Photochem Photobiol 71: 514-523.

135. Battista JR, Earl AM, Park MJ (1999) Why is *Deinococcus radiodurans* so resistant to ionizing radiation? Trends Microbiol 7: 362-365.

136. Li de la Sierra-Gallay I, Zig L, Jamalli A, Putzer H (2008) Structural insights into the dual activity of RNase J. Nat Struct Mol Biol 15: 206-212.

137. Basu B, Apte SK (2012) Gamma radiation-induced proteome of *Deinococcus radiodurans* primarily targets DNA repair and oxidative stress alleviation. Mol Cell Proteomics 11: M111 011734.

138. Chen HW, Randle DE, Gabbidon M, Julin DA (1998) Functions of the ATP hydrolysis subunits (RecB and RecD) in the nuclease reactions catalyzed by the RecBCD enzyme from *Escherichia coli*. J Mol Biol 278: 89-104.

139. Valencia A (2005) Automatic annotation of protein function. Curr Opin Struct Biol 15: 267-274.

140. Lukk T, Sakai A, Kalyanaraman C, Brown SD, Imker HJ, et al. (2012) Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. Proc Natl Acad Sci U S A 109: 4122-4127.

141. Iyer LM, Aravind L, Bork P, Hofmann K, Mushegian AR, et al. (2001) *Quod erat demonstrandum*? The mystery of experimental validation of apparently erroneous computational analyses of protein sequences. Genome Biol 2: RESEARCH0051.

142. Doherty AJ, Jackson SP (2001) DNA repair: how Ku makes ends meet. Curr Biol 11: R920-924.

143. Kim JI, Sharma AK, Abbott SN, Wood EA, Dwyer DW, et al. (2002) RecA Protein from the extremely radioresistant bacterium *Deinococcus radiodurans*: expression, purification, and characterization. J Bacteriol 184: 1649-1660.

144. Kim JI, Cox MM (2002) The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways. Proc Natl Acad Sci U S A 99: 7917-7921.

145. Lu H, Gao G, Xu G, Fan L, Yin L, et al. (2009) *Deinococcus radiodurans* PprI switches on DNA damage response and cellular survival networks after radiation damage. Mol Cell Proteomics 8: 481-494.

146. Lipton MS, Pasa-Tolic L, Anderson GA, Anderson DJ, Auberry DL, et al. (2002) Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. Proc Natl Acad Sci U S A 99: 11049-11054.

Microbiology (2011), **157**, 2196–2205 **DOI** 10.1099/mic.0.040436-0

Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage

Anubrata D. Das and Hari S. Misra

Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400 085, India

Correspondence
Hari S. Misra
hsmisra@barc.gov.in

Received 12 April 2010
Revised 20 January 2011
Accepted 18 April 2011

DRA0282, a hypothetical protein, was found in a pool of nucleotide-binding proteins in *Deinococcus radiodurans* cells recovering from gamma radiation stress. This pool exhibited an unusual inhibition of nuclease activity by ATP. The N terminus of DRA0282 showed similarity to human Ku80 homologues, while the C terminus showed no similarities to known proteins. The recombinant protein required Mn^{2+} for its interaction with DNA and protected dsDNA from exonuclease III degradation. The binding of the protein to supercoiled DNA with a K_d of ~2.93 nM was nearly 20-fold stronger than its binding to ssDNA and nearly 67-fold stronger than its binding to linear dsDNA. *Escherichia coli* cells expressing DRA0282 showed a RecA-dependent enhancement of UV and gamma radiation tolerance. The Δ*drA0282* mutant of *D. radiodurans* showed a dose-dependent response to gamma radiation. At 14 kGy, the Δ*drA0282* mutant showed nearly 10-fold less survival, while at this dose both *pprA*::*cat*Δ*drA0282* and *pprA*::*cat* mutants were nearly 100-fold more sensitive than the wild-type. These results suggested that DRA0282 is a DNA-binding protein with a preference for superhelical DNA, and that it plays a role in bacterial resistance to DNA damage through a pathway in which PprA perhaps plays a dominant role in *D. radiodurans*.

INTRODUCTION

Deinococcus radiodurans R1 shows extraordinary tolerance to several abiotic stresses, including ionizing and non-ionizing radiation (Makarova *et al.*, 2001; Cox & Battista 2005; Blasius *et al.*, 2008). This bacterium possesses an efficient DNA double-strand break (DSB) repair mechanism (Minton, 1994; Daly & Minton, 1996) and strong oxidative stress tolerance mechanisms (Markillie *et al.*, 1999; Daly *et al.*, 2004, 2007). DSB repair in this organism displays RecA-dependent biphasic kinetics (Daly *et al.*, 1994; Slade *et al.*, 2009). Phase I involves extended synthesis-dependent strand annealing (ESDSA) processes (Zahradka *et al.*, 2006), followed by the second phase of the maturation of individual chromosomes, which presumably involves homologous recombination (Daly & Minton, 1996). The *D. radiodurans* R1 genome (White *et al.*, 1999) shows the absence of *recBC* and *sbcB*/*sbcA* genes, while the *recD2* protein (Montague *et al.*, 2009) and components of the complete RecF pathway are present (Blasius *et al.*, 2008). Expression of *sbcB* (Misra *et al.*, 2006) and *recBC* (Khairnar *et al.*, 2008) genes of *Escherichia coli* in *D. radiodurans* makes this bacterium sensitive to gamma radiation by delaying its DSB repair, further supporting the contention that the RecF recombination pathway contributes to DSB repair and radiation resistance in this bacterium. The role of the RecFOR pathway in DSB repair through an ESDSA mechanism has been suggested in *D. radiodurans* (Bentchikou *et al.*, 2010). However, the significance of DNA protection in radiation resistance and DSB repair during the early phase of post-irradiation recovery (PIR) has also been shown. Two proteins, namely DdrA (Harris *et al.*, 2004) and PprA (Narumi *et al.*, 2004), have been characterized for their roles in DNA protection and radiation resistance in *D. radiodurans*. Also, recently, the involvement of reversible protein phosphorylation in the regulation of nucleolytic degradation during PIR has been suggested (Kamble *et al.*, 2010).

Here, we report the characterization of DRA0282 from *D. radiodurans* for its role in DNA protection from exonuclease degradation *in vitro* and in radiation resistance *in vivo*. The recombinant protein showed stronger binding to superhelical DNA and protected dsDNA from exonuclease III (ExoIII) degradation *in vitro*. The DR_A0282 deletion mutant (Δ*drA0282*) of *D. radiodurans* showed a dose-dependent effect, and these cells displayed nearly 10-fold less resistance as compared with the wild-type at a 14 kGy dose of gamma radiation. Deletion of *drA0282* in a *pprA*-minus background (*pprA*::*cat* Δ*drA0282*), however,

Abbreviations: DSB, double-strand break; ExoIII, exonuclease III; FTICR-MS, Fourier transform ion cyclotron resonance-MS; PIR, post-irradiation recovery.

Five supplementary figures and two supplementary tables are available with the online version of this paper.

showed no additive effect to the *pprA* mutant phenotype, although a double mutant showed nearly a 10-fold higher sensitivity than the Δ*drA0282* mutant. *E. coli* cells expressing DRA0282 showed a RecA-dependent enhancement of UV (254 nm) and gamma radiation resistance. Although SOS-deficient but recombination-proficient *E. coli* cells expressing DRA0282 showed an increase in DNA damage tolerance, the increase was significantly less than that of wild-type cells expressing the same protein. This indicated that both recombination and SOS mechanisms are important for DRA0282 function *in vivo*. Thus, we characterized DRA0282 as a DNA-binding protein with a preference for superhelical DNA, and demonstrated that it contributes to bacterial resistance to DNA damage by mechanism(s) that require RecA, at least in *E. coli*.

METHODS

Bacterial strains and materials. *D. radiodurans* ATCC13929 was a generous gift from Dr M. Schäfer, Germany (Schäfer *et al.*, 2000). The wild-type and its derivatives were grown aerobically in TGY (0.5 % Bacto tryptone, 0.3 % Bacto yeast extract, 0.1 % glucose) with or without kanamycin (8 μg ml^{-1}). TGY agar plates were incubated at 32 °C for 48 h and colonies were counted. *E. coli* and its derivatives were grown in Luria–Bertani (LB) medium supplemented with appropriate antibiotics. Other recombinant techniques were as described in Sambrook & Russell (2001). Molecular biology grade chemicals were purchased from Sigma, Roche Biochemicals, New England Biolabs and Bangalore Genei.

DNA-binding protein fractionation. Exponentially growing cells were irradiated with 6 kGy of gamma radiation. Aliquots were collected at 1 h intervals and cell-free extract was prepared in lysis buffer (20 mM Tris/HCl, pH 7.6, 1 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF, 200 μg lysozyme ml^{-1}), as described previously (Kota & Misra, 2008). Soluble proteins were diluted with 10 volumes of buffer A (10 mM sodium phosphate, pH 7.0, 50 mM NaCl, 1 mM EDTA, pH 8.0, 1 mM DTT, 1 mM PMSF) and passed through a heparin–Sepharose affinity column as described previously (Kamble *et al.*, 2010). Proteins were eluted with 500 mM NaCl in buffer A. The protein concentration was determined, and the heparin-purified fraction showing ATP inhibition of nucleolytic function was treated with alkaline phosphatase and refractionated on a MonoQ column by FPLC. Proteins were eluted with a salt gradient (50–500 mM NaCl) in 50 mM Tris/HCl buffer, pH 7.5, containing 10 % (v/v) glycerol and 1 mM DTT. Fractions were assayed for nuclease activity in the presence and absence of 2 mM ATP. Fraction 12, which showed ATP inhibition of nucleolytic function, was analysed by Fourier transform ion cyclotron resonance-MS (FTICR-MS).

Functional domain analysis. Functional domain analysis of DRA0282 was carried out using standard online bioinformatics tools. In brief, a PSI-BLAST search was done with the Swiss-Prot database, with default parameters and the keyword 'DNA repair'. Initial iterations showed Ku80 and PRP19/PSO4 from yeast, and these were selected for further iterations. After five iterations, several Ku-like proteins were found. We selected these sequences, and multiple sequence alignment was carried out, including prokaryotic Ku proteins, in CLUSTAL_X. This procedure was repeated further to align DRA0282 with Ku proteins alone. DRA0282 again clustered with eukaryotic Ku. Next we built a neighbour-joining tree and bootstrapped it 1000 times. Secondary structure was inferred from PSIPRED, JNET and Professor at the Quick 2D server at the Max Planck Institute for Developmental Biology. For 3D structure determination, we initially submitted the protein sequence to online servers such as I-TASSER and PHYRE, and the templates with closest matches were obtained from these servers. We subsequently built a 3D model in SWISS-MODEL, in the alignment mode, using 1JEY, the human Ku heterodimer structure, as the template (Arnold *et al.*, 2006), and residues 1–486 were modelled at minimum free energy stabilization.

Construction of a DR_A0282 expression plasmid and recombinant protein purification. Genomic DNA of *D. radiodurans* was prepared as described in Battista *et al.* (2001) and used for PCR amplification of *drA0282* coding sequences using a gene-specific forward primer 282F (5′-GAATTCATATGTTCATGAAGAGCAAGG-3′) and reverse primer 282R (5′-ACGGGATCCTAACTGACGGG-AGGTGA-3′), which incorporated restriction enzyme sites at the 5′ ends of the primers. The 1.515 kb PCR product was ligated at *Nde*I and *Bam*HI in pET28a+ to yield pET282, and at *Nco*I and *Bam*HI sites in pTrc99 to yield pTrc282. The pTrc282 plasmid was used for studying the effect of DRA0282 expression on the radiation response of the wild-type (*E. coli* AB1157) and *recA* mutant (*E. coli* JC1553) of *E. coli* K-12 (Supplementary Table S1). The pET282 plasmid was transferred to *E. coli* BL21(DE3) pLysS, and cells were induced with 200 μM IPTG for purification of recombinant DRA0282 using metal affinity chromatography, as described previously (Misra *et al.*, 1998). DRA0282 purified under denaturing conditions was refolded by serial dilution of urea in buffer with an increasing concentration of DTT, as described previously (Khairnar & Misra, 2009). The refolded protein was dialysed in 10 mM Tris/HCl, pH 7.5, 15 mM KCl, 1 mM DTT and 50 % (v/v) glycerol, and stored at −20 °C for further use.

Isolation of a deletion mutant of *D. radiodurans*. The *drA0282* deletion mutant was generated using protocols described previously (Khairnar *et al.*, 2008). In brief, the 1 kb upstream and downstream fragments of the coding sequences of *drA0282* were PCR-amplified using sequence-specific primers. The upstream fragment was amplified using primers 282upF (5′-ATGGGCCCGCTTTGTACTCCA-GAGCA-3′) and 282upR (5′-CGGAATTCTTCATGAACACCAGGT-3′), and the downstream fragment was amplified using 282dnF (5′-ATGGATCCACCACCCGTCAGTAAGCT-3′) and 282dnR (5′-CGT-CTAGAAGCGCCCTGGTTGACCGT-3′). The products were cloned independently and sequentially in pNOKOUT (Khairnar *et al.*, 2008) to yield pNOK282. This plasmid was linearized with *Sca*I and transformed into *D. radiodurans* to generate the Δ*drA0282* mutant. PCR amplification was carried out using sequence-specific primers, and complete replacement of *drA0282* with the *nptII* gene was confirmed. These clones were considered as Δ*drA0282* mutants, and a single such clone was used for further studies.

To generate the double mutant of *pprA* and *drA0282*, linearized pNOK282 was transformed into the *pprA*::*cat* mutant of *D. radiodurans* (Narumi *et al.*, 2004). Recombinant cells were sub-cultured for many generations in TGY broth containing both chloramphenicol (5 μg ml^{-1}) and kanamycin (10 μg ml^{-1}). Genomic DNA was prepared, and the absence of *drA0282* in the *pprA* mutant background was ascertained by PCR amplification using gene-specific primers.

Cell survival studies. Cell survival studies were carried out as described previously (Rajpurohit *et al.*, 2008). In brief, mutant and wild-type *D. radiodurans* cells were grown in TGY broth to the late-exponential phase at 32 °C. Cells were suspended in sterile PBS and exposed to various doses of gamma radiation at a dose rate of 5.89 kGy h^{-1} (Gamma 5000, ^{60}Co, Board of Radiation and Isotopes Technology, Department of Atomic Energy, India). Appropriate dilutions were plated on TGY agar plates and incubated at 32 °C. *E. coli* cells expressing DRA0282 on pET282 and pTrc282 were treated with different doses of UV (254 nm) and gamma radiation as

described previously (Kota & Misra, 2006). In brief, exponentially growing *E. coli* cells induced with 200 μM IPTG were washed with and resuspended in PBS, and treated with various doses of gamma radiation at a dose rate of 356 Gy h^{-1} (Gammacell 220 irradiator, MDS Nordion) and UV at 254 nm. UV-treated cells were plated on LB agar plates containing kanamycin (25 μg ml^{-1}) and incubated in the dark. The plates were incubated at 37 °C overnight for *E. coli* and for 36 h at 32 °C for *D. radiodurans* before c.f.u. were counted.

DNA binding and ExoIII protection studies. The DNA-binding assay was carried out as described previously (Kamble *et al.*, 2010). In brief, 100 ng purified DRA0282 and 500 ng heparin-purified proteins were incubated with 200 ng PCR-amplified deinococcal DNA/super-helical form of pBluescript SK+/single-stranded linear DNA in buffer containing 10 mM Tris/HCl, pH 7.5, 5 mM $MnCl_2$, 15 mM KCl and 2 % (v/v) glycerol for 20 min at 37 °C. To determine the effect of ATP, the proteins were pre-incubated with 2 mM ATP for 20 min at 37 °C, followed by incubation with 200 ng DNA substrate for 30 min. The ExoIII protection assay was performed as described elsewhere (Wang *et al.*, 2005). In brief, 200 ng DNA substrate was pre-incubated with different concentrations of purified DRA0282 for 10 min in Exonuclease III Buffer (New England Biolabs). The reaction mixture was further incubated for 20 min in the presence of an increasing concentration of ExoIII. When required, the nucleoprotein complex was dissociated by heating at 65 °C in the presence of 95 % formamide and 25 mM EDTA. Products were analysed on 1 % agarose gels and DNA bands were visualized with ethidium bromide.

For determination of the affinity constant of DRA0282 for different DNA substrates, 200 ng DNA was incubated with an increasing concentration of recombinant DRA0282 (4 nM to 1.42 μM) for 10 min, and products were analysed on 0.8 % agarose gels. Ethidium bromide-stained DNA fluorescence intensity was measured using ImageJ software (Abramoff *et al.*, 2004). The integrated intensity for bound and unbound fractions per unit area was measured separately, and the fraction bound to the protein was plotted as a function of the protein concentration using GraphPad Prism 5. The K_d for curve fitting of individual plots was determined by the software, working on the principle of the least-squares method and applying the formula $Y = B_{max} \times [X]/K_d + [X]$, where B_{max} is the maximum binding capacity, Y is the bound fraction and [X] is the protein concentration.

RESULTS AND DISCUSSION

Identification of DRA0282 from a protein mixture exhibiting ATP-stimulated DNA protection

Heparin-binding proteins purified from cells undergoing PIR show kinetically the appearance of nucleolytic activity (Kamble *et al.*, 2010). The 1 h PIR proteins did not show any inherent nucleolytic activity, which, however, was expressed upon dephosphorylation with alkaline phosphatase. Alkaline phosphatase-treated 1 h PIR proteins were refractionated on a Mono-Q column (Fig. 1a). The fractions were checked for nuclease activity towards PCR-amplified DNA with and without ATP. It was found that the nucleolytic activity was fractionated (Fig. 1b), and only fraction 12 showed ATP inhibition of nucleolytic function (Fig. 1c). Fractions 6, 7 and 8 showed strong DNA-binding activity. Other fractions, such as 9, 10, 13–18 and 20, showed neither nucleolytic activity nor an effect of ATP on the quality and quantity of the DNA substrate as compared with the control. This suggested that the nucleolytic function of fraction 12 was sensitive to ATP. This fraction was analysed by FTICR-MS, and matched a number of proteins represented by at least five or more peptides (Supplementary Table S2). These included some known proteins as well as a large number of uncharacterized ORFs, including DR_0129, DR_A0282, DR_1124, DR_1768, DR_2417 m and DR_2563. However, no proteins were annotated as nucleases or kinases in this fraction. Among these, the ORFs DR_0129, DR_1124, DR_1768, DR_2417 m and DR_2563 have also been found in a multiprotein DNA-processing complex identified independently in the same bacterium that also exhibits an uncharacterized ATP-sensitive exonucleolytic function (Kota & Misra, 2008). Fraction 12 also contained DRA0282, a hypothetical protein, with the largest number of FTICR-MS peptides (84), which was characterized here for its *in vitro* activity and *in vivo* role, if any, in radiation resistance.

DRA0282 shows structural similarities to Ku-type proteins

Multiple sequence alignments of DRA0282 with human Ku80 showed a close similarity to the β-barrel of the Ku

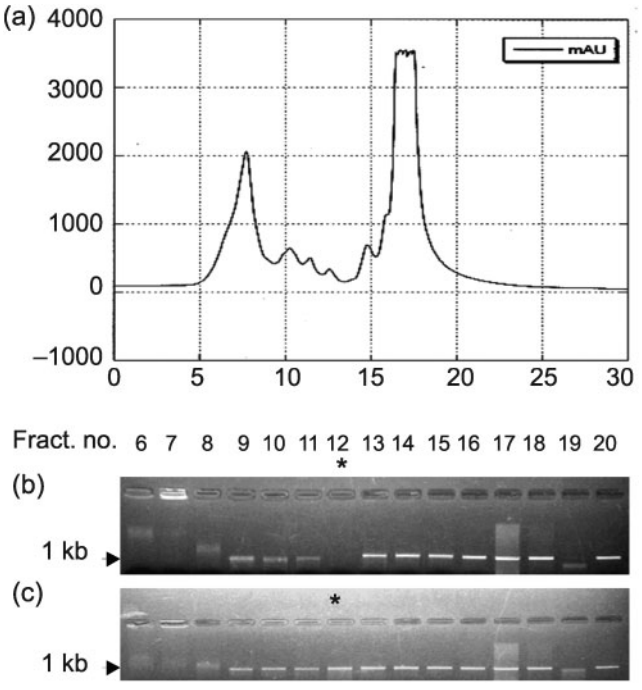


Fig. 1. Fractionation and nucleolytic activity assay of heparin-binding proteins from *D. radiodurans*. Heparin-binding proteins were isolated from cells allowed to recover for 1 h post-irradiation, and were fractionated on MonoQ columns by FPLC (a). Nearly 500 ng of proteins of each fraction was tested for nucleolytic activity with 100 ng dsDNA substrate in buffer containing 5 mM $MnCl_2$, both in the absence (b) and the presence (c) of 2 mM ATP pre-incubated with the proteins for 20 min. (b, c) Asterisks mark fraction 12, which contained the hypothetical protein DRA0282.

core domain (Supplementary Fig. S1). Phylogenetic analysis showed DRA0282 clustering with eukaryotic Ku proteins (Supplementary Fig. S2). This was further verified by the alignment of DRA0282 with other Ku-like proteins from both prokaryotes and eukaryotes. Results showed that ~150 amino acid residues from the N terminus of DRA0282 were homologous to the N terminus of human and yeast Ku80. This forms the α/β domain, also known as the Rossman fold, which signifies the binding of nucleotides by this protein. Further, the 150 amino acid residues from 250–400 of DRA0282 showed the highest homology to the central domain known as the Ku core domain in Ku proteins, which is responsible for their binding to free DNA ends (Aravind & Koonin, 2001). The secondary structure predicted by PyMOL showed extensive β-sheets at regular intervals, interspersed by coils. Although the DRA0282 sequence showed the overall α/β domain structure of Ku80m, it lacked certain segments, such as $\alpha 2$ and $\beta 17$ (Supplementary Fig. S1 and Fig. 2). The overall similarities of this protein to Ku-like proteins prompted us to build a 3D model using the human Ku heterodimer as template (Arnold *et al.*, 2006). Regions with gaps had high root mean square deviation (RMSD), while regions with sequence overlap had a maximum RSMD of 4°, and the modal RMSD value was 0.8. Unlike human Ku, which is activated by ATP, DRA0282 was inhibited by it. This difference could be attributed to the absence of the $\alpha 2$ loop in DRA0282 and the resulting changes in its structure. The core domain of this protein resembled the Ku80 core domain. The $\alpha 7$–$\beta 18$ loop, which is responsible for dsDNA binding in Ku80, showed a significant match with DRA0282, while $\beta 10$–$\beta 13$ showed a poor structural alignment. The homology of DRA0282 to human Ku80 and its homologues was therefore through its N terminus, while its C-terminal domain was different. It could be speculated that this protein has evolved through genomic rearrangement under the pressure of natural selection. The results from online servers were intriguing. PHYRE returned a receptor-type tyrosine protein phosphatase as the best match [Protein Data Bank (PDB) ID: 2V5Y], while I-TASSER returned a sialidase from *Micromonospora* spp. (PDB ID: 1EUT), and FUGUE returned measles virus haemagglutinin (PDB ID: 2ZB6). These proteins have in common extensive β-sheet structures. These results indicated the structural resemblance of DRA0282 to Ku80 protein in having DNA-binding domains.

DRA0282 is a DNA-binding protein with an ability to protect dsDNA in solution

The purified recombinant DRA0282 protein (Supplementary Fig. S3) was incubated with ssDNA and linear dsDNA in the presence of different metal ions. The protein showed strong binding to both dsDNA and ssDNA in the presence of Mn^{2+}, while this activity was not observed in the presence of Ca^{2+} and Mg^{2+} (Fig. 3a). DRA0282 showed stronger binding to ssDNA than to dsDNA in the presence of Mn^{2+} (Fig. 3b). It did not bind to dsDNA in the presence of Zn^{2+}, while its binding to ssDNA in the presence of Zn^{2+} was minor compared with its binding in the presence of Mn^{2+}. Recombinant DRA0282 showed strong binding to superhelical plasmid DNA, without affecting the topology or size of the plasmid DNA. This was confirmed by the complete recovery of the substrate in its original form upon heating with 95 % formamide and 25 mM EDTA at 65 °C (Fig. 3b), which breaks the nucleoprotein interaction *in vitro* (O'Donovan *et al.*, 1994). This indicated that DRA0282 was a Mn^{2+}-dependent

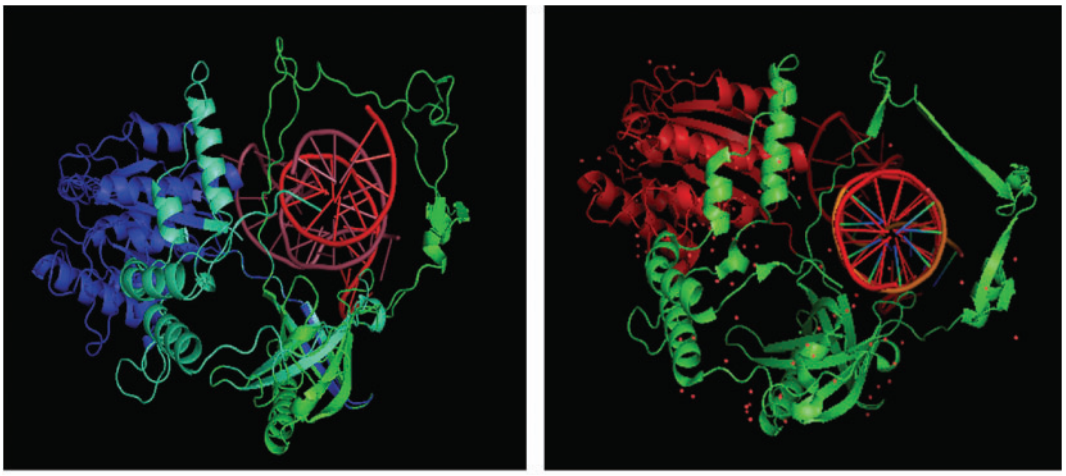


Fig. 2. Structural similarities of DRA0282 and human Ku80. The 3D models of DRA0282 and human Ku80 were compared to find similarities in the dsDNA-binding motifs. The N-terminal Von Willebrand domain is coloured blue in DRA0282 and red in human Ku80, while the Ku core domain is coloured green in both molecules.

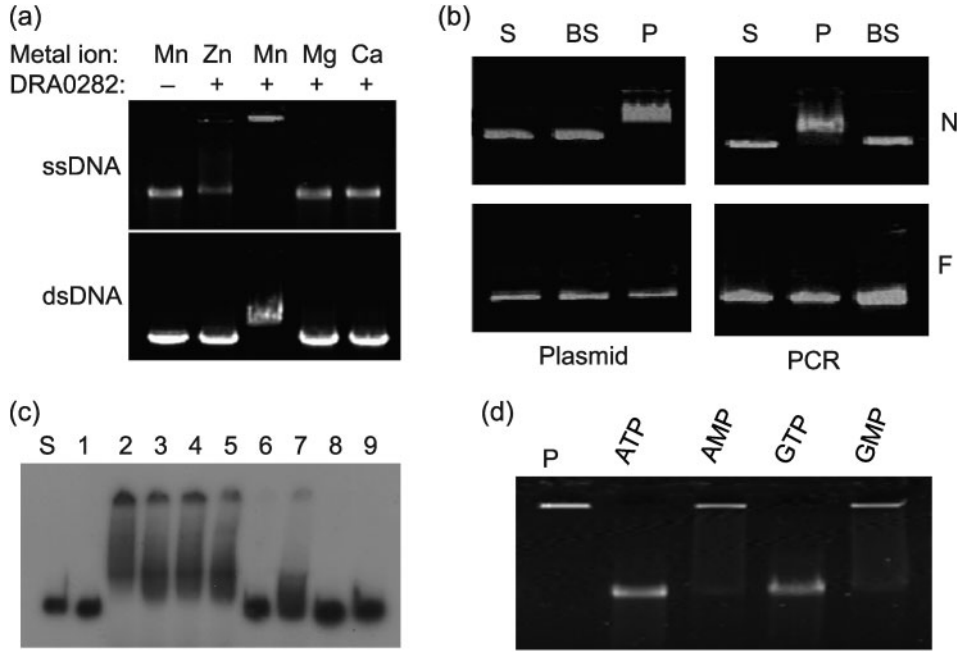


Fig. 3. Characterization of the DNA-binding activity of DRA0282. (a) One hundred nanograms of purified recombinant DRA0282 was incubated with 200 ng M13mp18 ssDNA and 200 ng dsDNA in buffer supplemented with 5 mM $ZnCl_2$, $MnCl_2$, $MgCl_2$ or $CaCl_2$, and products were analysed on 1 % agarose. (b) DRA0282 (P) was incubated with both linear (PCR) and superhelical (plasmid) dsDNA (S), with BSA (BS) as a control. One set of reaction mixtures was treated with 95 % formamide, 25 mM EDTA (F), while another set was kept as a control (N), and both sets were analysed on 1 % agarose gels. (c) Autoradiogram showing the 200 bp dsDNA labelled with ^{32}P at the 5′ end (lane S) incubated with either 100 ng BSA (lane 1) or 100 ng DRA0282 (lanes 2–9). For the competition assay, the protein was pre-incubated separately with 50 ng *Bam*HI- (lane 3), *Eco*RV- (lane 4) or *Pst*I-digested dsDNA (lane 5), M13mp18 ssDNA (lane 6) or supercoiled plasmid DNA (lane 7), before the addition of radiolabelled probe. DNA-binding activity (lane 2) was also checked in the absence (lane 8) and presence (lane 9) of 5 mM Mn^{2+} + 2 mM ATP. Products were separated by 6 % native PAGE and dried, and the autoradiogram was developed. (d) The effect of 2 mM ATP, AMP, GTP or GMP on the ssDNA-binding activity of DRA0282 was monitored on an agarose gel.

DNA-binding protein. The effects of the types of DNA ends and ssDNA on dsDNA binding were monitored by a competition assay. A 200 bp ^{32}P-labelled dsDNA probe was used to study the binding preferences of DRA0282. dsDNA with different types of ends generated by *Bam*HI, *Eco*RV or *Pst*I did not disrupt the binding of DRA0282 to the probe, which, however, was readily disrupted by ssDNA (Fig. 3c). These activities were Mn^{2+}-dependent and sensitive to ATP (Fig. 3c). Further, the purified protein showed the highest binding constant (K_d) of 2.93 nM for superhelical plasmid, followed by 60.22 nM for linear ssDNA and 196.9 nM for linear dsDNA (Fig. 4).

Since this protein contains a Ku80-type DNA-binding pocket and nucleotide interaction domain, the effect of nucleoside phosphates on DNA binding and the protection of DNA ends from ExoIII degradation were examined. The DNA–DRA0282 interaction was inhibited by ATP and GTP, while AMP and GMP did not show any measurable effect on this activity (Fig. 3d). These results suggested that DRA0282 was a DNA-binding protein with a relatively higher affinity for superhelical DNA and linear ssDNA as compared with linear dsDNA. The inhibition of the DNA–protein interaction by nucleoside triphosphates and not by monophosphates suggested the involvement of terminal phosphates rather than the adenyl group in the inhibition of the DNA-binding activity of this protein. Similarly, recombinant DRA0282 protected linear dsDNA from ExoIII degradation, whereby as much as 50 U nuclease could not degrade the dsDNA substrate completely after pre-incubation with DRA0282 (Fig. 5a). An increasing concentration of DRA0282 afforded an increase in protection of DNA when challenged with 10 U ExoIII, which was able to degrade the dsDNA substrate completely in the absence of DRA0282 (Fig. 5b). Since DNA binding by DRA0282 was inhibited by ATP, the effect of ATP on the dsDNA protection of DRA0282 was examined. In the presence of ATP, DRA0282 failed to protect DNA from ExoIII degradation, and eventually more than 90 % of the DNA was degraded, in spite of the presence of 72 nM DRA0282 (Fig. 5c), which had shown complete DNA protection in an earlier assay (Fig. 5b). Since ATP alone showed no effect on ExoIII activity (Fig. 5c, lane 6), and the

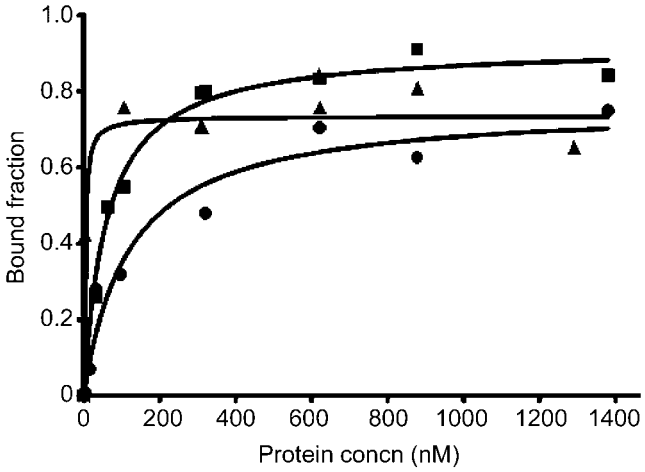


Fig. 4. Determination of the DNA-binding constant of DRA0282. Two hundred nanograms of linear dsDNA (●), superhelical DNA (▲) and linear ssDNA (■) substrates were incubated with increasing concentrations of recombinant DRA0282 for 10 min at 37 °C. Products were analysed on a 1 % agarose gel. The DNA–protein bound fraction was plotted against protein concentration, and affinity constants of the protein for different DNA substrates were computed using GraphPad Prism 5 graphics, as described in Methods.

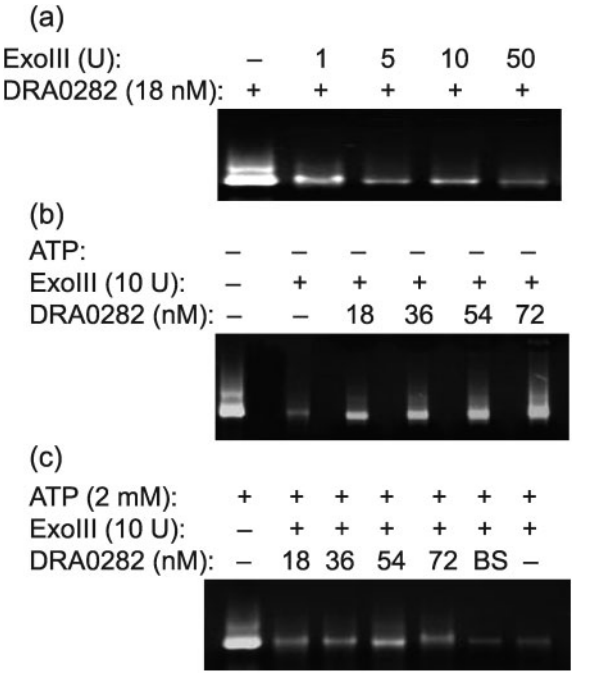


Fig. 5. DRA0282 protection of dsDNA from ExoIII degradation. The ~150 nM 1 kb PCR-amplified dsDNA was pre-incubated with ~18 nM DRA0282 for 5 min before increasing concentrations of ExoIII were added (a). Similarly, the DNA substrate was pre-incubated with increasing concentrations (18–72 nM) DRA0282 in the presence of 2 mM $MnCl_2$ (b) and 2 mM ATP along with 5 mM Mn^{2+} (c) before 10 U ExoIII was added. Reaction mixtures containing both BSA (BS) and ExoIII, and ExoIII alone, were used as controls. All reactions were carried out at 37 °C for 20 min, and products were analysed on a 1 % agarose gel.

protection of dsDNA from the action of ExoIII was not seen in the presence of BSA as a control (Fig. 5c, lane 7), it is possible that the binding of ATP to DRA0282 prevents its binding to DNA, eventually leading to DNA degradation by ExoIII. These results suggest that DRA0282 is able to protect dsDNA from nucleolytic degradation by direct binding to the DNA substrate. Its higher affinity for superhelical DNA indicated a strong possibility of this protein interacting directly with the bacterial genome. The possibility that DRA0282 has a role in the DNA damage response of bacteria was further investigated by expressing this protein in *E. coli* and by generating a knockout of this gene in *D. radiodurans*.

DRA0282 enhances the UVC and gamma radiation tolerance of *E. coli*

The effect of DRA0282 on DNA damage tolerance was monitored in wild-type and mutant strains of *E. coli* (Supplementary Table S1). Wild-type cells expressing DRA0282 showed a nearly 10-fold higher tolerance to gamma radiation at a dose of 200 Gy (Fig. 6a), and a nearly three-log increase in UVC tolerance at a dose of 1.2 kJ m^{-2}, as compared with untransformed cells (Fig. 6c). Interestingly, the *recA* mutant (JC1553) expressing DRA0282 did not show any increase in gamma radiation (Fig. 6b) and UVC tolerance (Fig. 6d). This indicated that DRA0282 functions in *E. coli* through RecA. Since RecA is involved in both homologous recombination and SOS response mechanisms in *E. coli*, the contribution of RecA in supporting DRA0282 function in *E. coli* via these pathways was examined. DRA0282 was expressed in *E. coli* strain DM49 (*E. coli* K-12, *recA*$^+$ *lexA3*) (Mount *et al.*, 1972), and the effect of both UVC and gamma radiation was monitored. These cells showed an enhanced tolerance to both UV and gamma radiation (Fig. 7), albeit slightly lower than that observed in the wild-type background. The DM49 strain has a wild-type copy of RecA but a point mutation in the RecA co-protease cleavage site of LexA (*lexA3*), making it SOS-deficient but still recombination-proficient. These results suggested that DRA0282 plays an important role in the DNA damage tolerance of *E. coli* and works through mechanisms that involve RecA.

DRA0282 deletion does not alter the *pprA* mutant phenotype in *D. radiodurans*

The Δ*drA0282* mutant showed a dose-dependent response to gamma radiation-induced DNA damage. It showed nearly a 10-fold reduction with respect to wild-type tolerance at 14 kGy gamma radiation (Fig. 8), while no difference in the gamma radiation response was noted in either a *pprA* single or a *pprA*::*cat* Δ*drA0282* double mutant. Both these mutants showed a nearly two-log

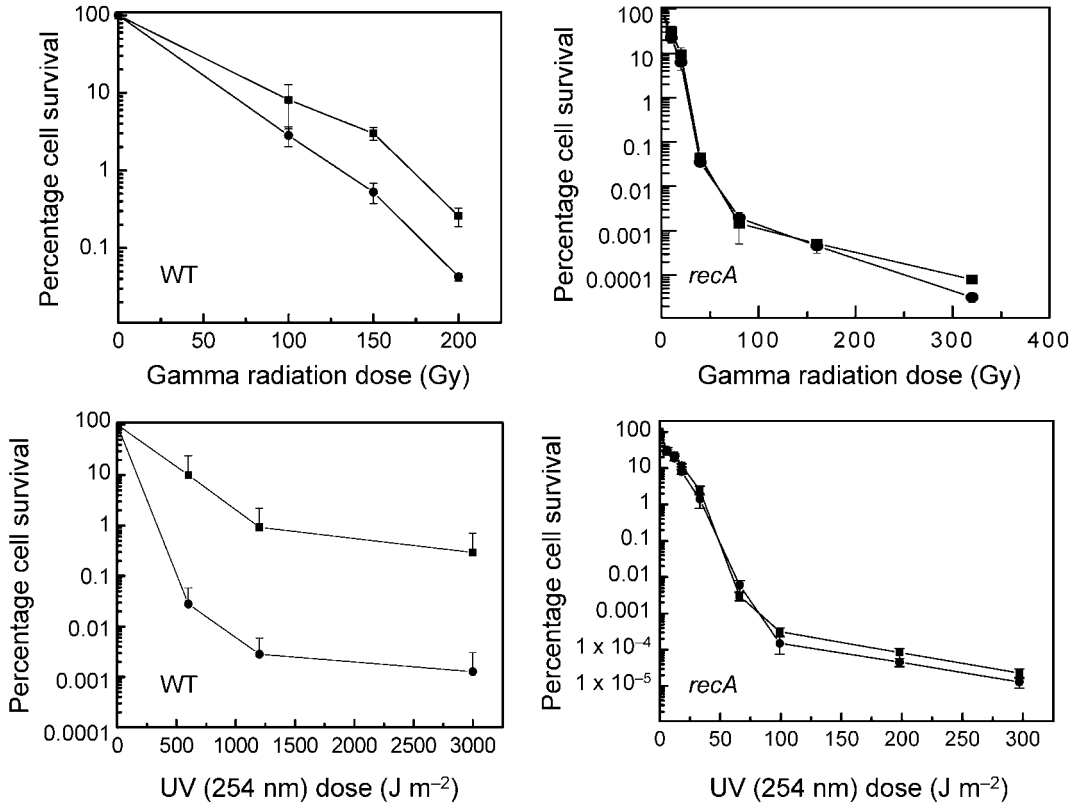


Fig. 6. Effect of DRA0282 overexpression on the DNA damage response of *E. coli*. *E. coli* AB1157 (wild-type; WT) and JC1553 (*recA*) cells expressing DRA0282 (■) and corresponding controls (●) were treated with different doses of gamma and UV (254 nm) radiation. c.f.u. were monitored after overnight incubation at 37 °C. Cell survival of 100 % corresponds to 5.423×10^7 and 3.124×10^7 cells ml^{-1} of the wild-type and *recA* mutant, respectively.

decrease with respect to the wild-type tolerance of gamma radiation (Fig. 8). This suggests that although both these proteins contribute to the radiation resistance of this bacterium, PprA plays the dominant role. The levels of expression of the *drA0282* gene were increased nearly 20-fold within 1 h of post-irradiation growth (Supplementary Fig. S4), further supporting the possible involvement of DRA0282 in the radiation response of *D. radiodurans*.

Previously it has been reported that gamma-irradiated *D. radiodurans* R1 cells show rapid DNA degradation, which is arrested within 90 min of PIR (Khairnar *et al.*, 2008; Slade *et al.*, 2009). During this period the cells show a higher rate of DNA synthesis, leading to a net increase in DNA content and DSB repair. This period coincides with the expression of DNA-protecting proteins such as PprA (Narumi *et al.*, 2004) and DdrA (Harris *et al.*, 2004), and many uncharacterized ORFs, including DR_A0282 (Liu *et al.*, 2003; Tanaka *et al.*, 2004; Supplementary Fig. S4). The protection of linear dsDNA from ExoIII activity by PprA, DdrA and now the DRA0282 protein (Fig. 5), which is also found to have a role in radiation resistance of *D. radiodurans* (Fig. 8), suggested that DNA protection from nucleolytic degradation plays an important role in the radiation resistance of this bacterium. The fact that DRA0282 enhanced the UVC and gamma radiation resistance of wild-type *E. coli* and a *lexA3* mutant, although to a lesser extent than the wild-type, while not doing so in a *recA* mutant, suggests a role for this protein in recombination-dependent mechanism(s) that contribute to DNA repair in *E. coli*. On the other hand, the molecular mechanisms that support the role of this protein in the radiation resistance of *D. radiodurans* are not yet clear. However, DRA0282 showed structural homology to Ku homologues, and the different roles of Ku-type proteins, including non-homologous end-joining (NHEJ) and DSB repair, have been shown in higher organisms (Downs & Jackson, 2004). For example, human hPso4, a protein having similarity to DRA0282 (Supplementary Fig. S5), has been shown to bring Metnase, a human DSB repair protein with SET and transposase/nuclease domains (Lee *et al.*, 2005), to the sites of DSB repair (Zhang *et al.*, 2005; Lu & Legerski, 2007; Beck *et al.*, 2008). hPso4 is known to bind dsDNA non-specifically, and has pleiotropic functions in DNA recombination and error-prone repair. The hPso4 protein is induced 15- to 30-fold in cells exposed to gamma radiation and chemical mutagens, but not UV radiation (Mahajan & Mitchell, 2003). These findings indicate the

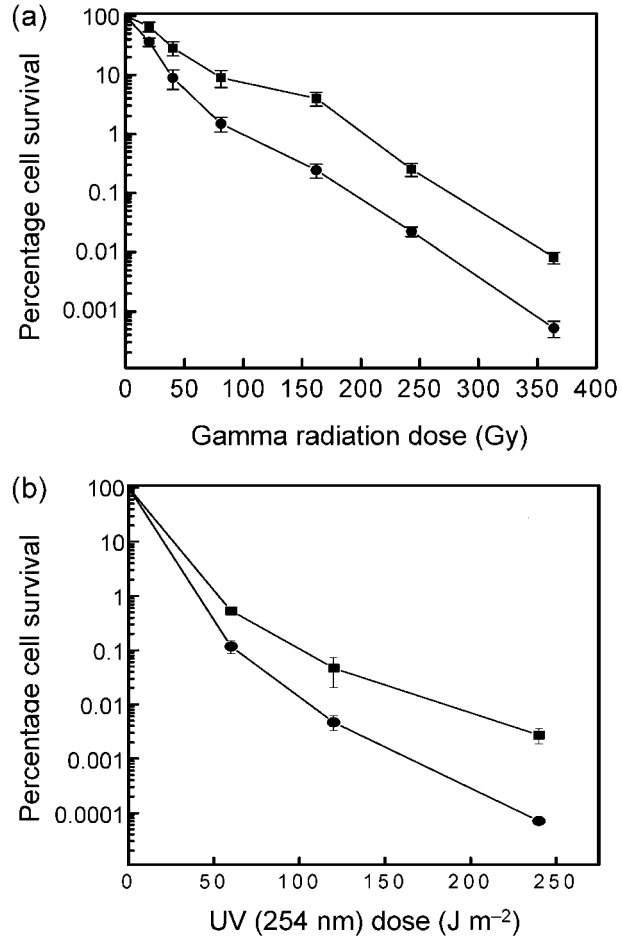


Fig. 7. Effect of DRA0282 on DNA damage tolerance of SOS-deficient and recombination-proficient *E. coli*. *E. coli* DM49 (*lexA3* mutant) cells expressing DRA0282 (■) or harbouring the expression vector (●) were treated with different doses of gamma (a) and UV (254 nm) (b) radiation, and cell survival was monitored at 37 °C. Cell survival of 100 % corresponds to 7.21×10^7 cells ml^{-1}.

possible role of Ku homologues in DSB repair by both NHEJ and recombination functions. Although the existence of NHEJ-type activity has not been demonstrated in *D. radiodurans*, its genome encodes several other hypothetical proteins which could constitute an NHEJ pathway, and therefore the occurrence of such a mechanism cannot be ruled out in this bacterium. The existence of an NHEJ-type mechanism has been reported in other bacteria, including *Bacillus subtilis* (Weller *et al.*, 2002) and *Mycobacterium* (Gong *et al.*, 2005; Pitcher *et al.*, 2007). Recently, an alternative mechanism of DNA ends joining has been demonstrated in *E. coli* (Chayot *et al.*, 2010). Although further studies will be required to understand the molecular mechanism underlying the roles of DRA0282 in the radioresistance of *D. radiodurans*, this study reports the characterization of the hypothetical protein DRA0282 as a DNA-binding protein with a higher affinity for superhelical DNA and with a role in dsDNA protection, as well as its

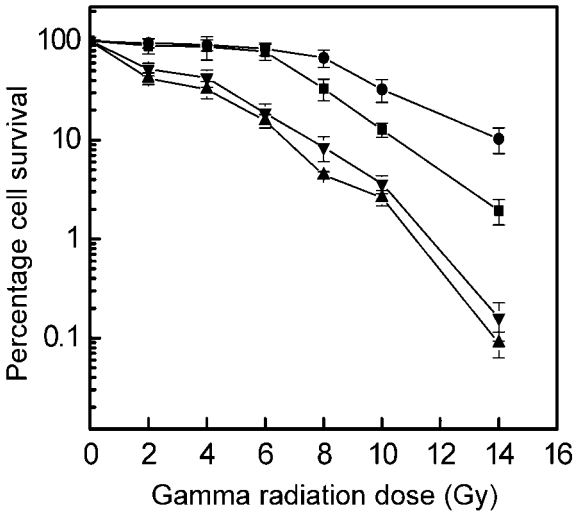


Fig. 8. Effect of *drA0282* deletion on the gamma radiation tolerance of *D. radiodurans*. Cells of *D. radiodurans* R1 (wild-type; ●) and its Δ*drA0282* single mutant (■), *pprA*::*cat*Δ*drA0282* double mutant (▼) and *pprA*::*cat* single mutant (▲) were treated with different doses of gamma radiation, and cell survival was monitored. Cell survival of 100 % corresponds to 1.45×10^7, 4.12×10^6, 2.54×10^7 and 1.72×10^6 cells ml^{-1} for the wild-type, Δ*drA0282*, *pprA*::*cat* and *pprA*::*cat*Δ*drA0282* strains, respectively.

contribution to bacterial resistance to DNA damage by a mechanism that requires a functional RecA, at least in *E. coli*.

ACKNOWLEDGEMENTS

The authors thank Dr S. K. Apte for his criticism and constructive comments during the preparation of the manuscript, Professor John R. Battista, Louisiana State University Baton Rouge, for providing laboratory space to H. S. M., and Professor Mary Lipton for proteome analysis. We gratefully acknowledge Dr Knut Ivan Kristiansen, Oslo University Hospital, Oslo, Norway, and Professor I. Narumi, Japan Atomic Energy Research Institute, Takasaki, Japan, for providing *E. coli* DM49 and the *pprA* mutant, respectively. H. S. M. thanks the Department of Biotechnology, Government of India, for an overseas senior research associateship.

REFERENCES

Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. (2004). Image processing with ImageJ. *Biophoton Int* **11**, 36–42.

Aravind, L. & Koonin, E. V. (2001). Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* **11**, 1365–1374.

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.

Battista, J. R., Park, M. J. & McLemore, A. E. (2001). Inactivation of two homologues of proteins presumed to be involved in the

desiccation tolerance of plants sensitizes *Deinococcus radiodurans* R1 to desiccation. *Cryobiology* **43**, 133–139.

Beck, B. D., Park, S. J., Lee, Y. J., Roman, Y., Hromas, R. A. & Lee, S. H. (2008). Human Pso4 is a metnase (SETMAR)-binding partner that regulates metnase function in DNA repair. *J Biol Chem* **283**, 9023–9030.

Bentchikou, E., Servant, P., Coste, G. & Sommer, S. (2010). A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans*. *PLoS Genet* **6**, e1000774.

Blasius, M., Hübscher, U. & Sommer, S. (2008). *Deinococcus radiodurans*: what belongs to the survival kit? *Crit Rev Biochem Mol Biol* **43**, 221–238.

Chayot, R., Montagne, B., Mazel, D. & Ricchetti, M. (2010). An end-joining repair mechanism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **107**, 2141–2146.

Clark, A. J. & Margulies, A. D. (1965). Isolation and characterization of recombination deficient mutants of *E. coli* K12. *Proc Natl Acad Sci U S A* **53**, 451–459.

Cox, M. M. & Battista, J. R. (2005). *Deinococcus radiodurans* – the consummate survivor. *Nat Rev Microbiol* **3**, 882–892.

Daly, M. J. & Minton, K. W. (1996). An alternative pathway of recombination of chromosomal fragments precedes *recA*-dependent recombination in the radioresistant bacterium *Deinococcus radiodurans*. *J Bacteriol* **178**, 4461–4471.

Daly, M. J., Ouyang, L., Fuchs, P. & Minton, K. W. (1994). In vivo damage and *recA*-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*. *J Bacteriol* **176**, 3508–3517.

Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Venkateswaran, A., Hess, M., Omelchenko, M. V., Kostandarithes, H. M. & other authors (2004). Accumulation of Mn(II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance. *Science* **306**, 1025–1028.

Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., Lai, B., Ravel, B., Li, S. M. & other authors (2007). Protein oxidation implicated as the primary determinant of bacterial radioresistance. *PLoS Biol* **5**, e92.

Downs, J. A. & Jackson, S. P. (2004). A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* **5**, 367–378.

Gong, C., Bongiorno, P., Martins, A., Stephanou, N. C., Zhu, H., Shuman, S. & Glickman, M. S. (2005). Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol* **12**, 304–312.

Harris, D. R., Tanaka, M., Saveliev, S. V., Jolivet, E., Earl, A. M., Cox, M. M. & Battista, J. R. (2004). Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1. *PLoS Biol* **2**, e304.

Kamble, V. A., Rajpurohit, Y. S., Srivastava, A. K. & Misra, H. S. (2010). Increased synthesis of signaling molecules coincides with reversible inhibition of nucleolytic activity during postirradiation recovery of *Deinococcus radiodurans*. *FEMS Microbiol Lett* **303**, 18–25.

Khairnar, N. P. & Misra, H. S. (2009). DNA polymerase X from *Deinococcus radiodurans* implicated in bacterial tolerance to DNA damage is characterized as a short patch base excision repair polymerase. *Microbiology* **155**, 3005–3014.

Khairnar, N. P., Kamble, V. A. & Misra, H. S. (2008). RecBC enzyme overproduction affects UV and gamma radiation survival of *Deinococcus radiodurans*. *DNA Repair (Amst)* **7**, 40–47.

Kota, S. & Misra, H. S. (2006). PprA: a protein implicated in radioresistance of *Deinococcus radiodurans* stimulates catalase activity in *Escherichia coli*. *Appl Microbiol Biotechnol* **72**, 790–796.

Kota, S. & Misra, H. S. (2008). Identification of a DNA processing complex from *Deinococcus radiodurans*. *Biochem Cell Biol* **86**, 448–458.

Lee, S.-H., Oshige, M., Durant, S. T., Rasila, K. K., Williamson, E. A., Ramsey, H., Kwan, L., Nickoloff, J. A. & Hromas, R. (2005). The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. *Proc Natl Acad Sci U S A* **102**, 18075–18080.

Liu, Y., Zhou, J., Omelchenko, M. V., Beliaev, A. S., Venkateswaran, A., Stair, J., Wu, L., Thompson, D. K., Xu, D. & other authors (2003). Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci U S A* **100**, 4191–4196.

Lu, X. & Legerski, R. J. (2007). The Prp19/Pso4 core complex undergoes ubiquitylation and structural alterations in response to DNA damage. *Biochem Biophys Res Commun* **354**, 968–974.

Mahajan, K. N. & Mitchell, B. S. (2003). Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. *Proc Natl Acad Sci U S A* **100**, 10746–10751.

Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J. (2001). Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol Mol Biol Rev* **65**, 44–79.

Markillie, L. M., Varnum, S. M., Hradecky, P. & Wong, K. K. (1999). Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (*katA*) and superoxide dismutase (*sodA*) mutants. *J Bacteriol* **181**, 666–669.

Minton, K. W. (1994). DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Mol Microbiol* **13**, 9–15.

Misra, H. S., Pandey, P. K., Modak, M. J., Vinayak, R. & Pandey, V. N. (1998). Polyamide nucleic acid–DNA chimera lacking the phosphate backbone are novel primers for polymerase reaction catalyzed by DNA polymerases. *Biochemistry* **37**, 1917–1925.

Misra, H. S., Khairnar, N. P., Kota, S., Shrivastava, S., Joshi, V. P. & Apte, S. K. (2006). An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance. *Mol Microbiol* **59**, 1308–1316.

Montague, M., Barnes, C., Smith, H. O., Chuang, R. Y. & Vashee, S. (2009). The evolution of RecD outside of the RecBCD complex. *J Mol Evol* **69**, 360–371.

Mount, D. W., Low, K. B. & Edmiston, S. J. (1972). Dominant mutations (*lex*) in *Escherichia coli* K-12 which affect radiation sensitivity and frequency of ultraviolet light-induced mutations. *J Bacteriol* **112**, 886–893.

Narumi, I., Satoh, K., Cui, S., Funayama, T., Kitayama, S. & Watanabe, H. (2004). PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. *Mol Microbiol* **54**, 278–285.

O'Donovan, A., Scherly, D., Clarkson, S. G. & Wood, R. D. (1994). Isolation of active recombinant XPG protein, a human DNA repair endonuclease. *J Biol Chem* **269**, 15965–15968.

Pitcher, R. S., Brissett, N. C. & Doherty, A. J. (2007). Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* **61**, 259–282.

Rajpurohit, Y. S., Gopalakrishnan, R. & Misra, H. S. (2008). Involvement of a protein kinase activity inducer in DNA double strand break repair and radioresistance of *Deinococcus radiodurans*. *J Bacteriol* **190**, 3948–3954.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schäfer, M., Schmitz, C., Facius, R., Horneck, G., Milow, B., Funken, K. H. & Ortner, J. (2000). Systematic study of parameters influencing the action of Rose Bengal with visible light on bacterial cells: comparison between the biological effect and singlet-oxygen production. *Photochem Photobiol* **71**, 514–523.

Slade, D., Lindner, A. B., Paul, G. & Radman, M. (2009). Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. *Cell* **136**, 1044–1055.

Tanaka, M., Earl, A. M., Howell, H. A., Park, M. J., Eisen, J. A., Peterson, S. N. & Battista, J. R. (2004). Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. *Genetics* **168**, 21–33.

Wang, J., Li, T., Guo, X. & Lu, Z. (2005). Exonuclease III protection assay with FRET probe for detecting DNA-binding proteins. *Nucleic Acids Res* **33**, e23.

Weller, G. R., Kysela, B., Roy, R., Tonkin, L. M., Scanlan, E., Della, M., Devine, S. K., Day, J. P., Wilkinson, A. & other authors (2002). Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* **297**, 1686–1689.

White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C. & other authors (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**, 1571–1577.

Zahradka, K., Slade, D., Bailone, A., Sommer, S., Averbeck, D., Petranovic, M., Lindner, A. B. & Radman, M. (2006). Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* **443**, 569–573.

Zhang, N., Kaur, R., Lu, X., Shen, X., Li, L. & Legerski, R. J. (2005). The Pso4 mRNA splicing and DNA repair complex interacts with WRN for processing of DNA interstrand cross-links. *J Biol Chem* **280**, 40559–40567.

Edited by: T. Msadek

Biochimica et Biophysica Acta 1820 (2012) 1052–1061

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

DR2417, a hypothetical protein characterized as a novel β-CASP family nuclease in radiation resistant bacterium, *Deinococcus radiodurans*

Anubrata D. Das, Hari S. Misra *

Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai-400 085, India

ARTICLE INFO

Article history:
Received 9 July 2011
Received in revised form 1 March 2012
Accepted 20 March 2012
Available online 30 March 2012

Keywords:
β-CASP family nuclease
Deinococcus
DNA processing
DSB repair
Junction endonuclease
Radioresistance

ABSTRACT

Background: Deinococcus radiodurans survives extreme doses of radiations contributed by efficient DNA repair pathways. DR2417 (DncA) was detected separately both in a pool of nucleotide binding proteins and multiprotein complex isolated from cells undergoing DNA repair.
Scope of review: DR_2417m ORF was sequenced and amino acid sequence of DncA was search for structural similarities with other proteins and functional motifs. Recombinant DncA was characterized for its DNA metabolic functions *in vitro* and its role in radiation resistance.
Major conclusions: Sequencing of DR_2417m did not show the reported frame shift at 996th nucleotide position of this gene. DncA showed similarities with β-CASP family nucleases. Recombinant protein acted efficiently on dsDNA and showed an Mn^{2+} dependent 3′→5′ exonuclease and ssDNA/dsDNA junction endonuclease activities while a very low level activity on RNA. The DNase activity of this protein was inhibited in presence of ATP. Its transcription was induced upon γ radiation exposure and a reduction in its copy number resulted in reduced growth rate and loss of γ radiation resistance in *Deinococcus*. *Conclusion* — our results suggest that DncA was a novel nuclease of β CASP family having a strong dsDNA end processing activity and it seems to be an essential gene required for both growth and γ radiation resistance of this bacterium.
General significance: Traditionally DncA should have shown both DNase and RNase functions as other members of β CASP family nucleases. A strong DNase and poor RNase activity possibly made it functionally significant in the radioresistance of *D. radiodurans*, which would be worth investigating independently.

1. Introduction

Deinococcus radiodurans R1 has a remarkable ability of surviving intense doses of γ radiation. It tolerates up to 6 kGy γ radiation with remarkable endurance, a dose, which would be lethal for almost all forms of life. During this exposure, the genome breaks down to nearly 200 fragments but within 3–4 h, all this is accurately reassembled without a measurable loss of cell viability [1,2]. These phenotypes are attributed to an efficient DNA strand break repair mechanism [3,4] and a strong oxidative stress tolerance [5] due to antioxidant enzymes [6] and by the non-enzymatic components like pyrroloquinoline quinine (PQQ) [7], carotenoids [8] and accumulation of Mn (II) [9,10]. DNA strand break repair in this bacterium is biphasic [11,12] with an initial ESDSA phase [13] and a later homologous recombination phase. Interestingly, the involvement of RecA has been demonstrated in both these phases [14]. The genome sequence of *D. radiodurans* revealed that it had the usual complement of DNA repair machinery such as nucleotide excision repair, base excision repair, recombination repair [3], as known in other radiation sensitive bacteria [15]. However, there are some unique features like — (i) presence of proteins involved in extended synthesis dependent strand annealing (ESDSA) [13] that supports phase I of DSB repair, (ii) PprA [16] and DdrA [17] involved in protection of DNA and (iii) modulation of nucleolytic functions during post irradiation recovery (PIR), through protein phosphorylation/dephosphorylation [18]. The involvement of a novel DNA damage responsive membrane protein kinase in extreme radiation resistance and DSB repair [19] has been recently demonstrated. These processes help in error free reassembling of shattered genome without loss of genetic information. Some of these proteins characterized recently, suggest the importance of DNA processing and its functional significance in radiation resistance of this bacterium [20–22]. Also, another unique source for contributing to the extraordinary resistance of *D. radiodurans* may lie with the remaining ~30% uncharacterized proteins. Recently, we observed that DNA binding proteins isolated from cells collected at different PIR period show a kinetic change in nucleolytic activity on double stranded DNA substrate [18]. Mass spectrometric identification of the proteins from this pool showed the presence of 11 “hypothetical proteins”. Among them, DR2417m was found to be relatively abundant with 36 peptides [23]. Further, DR2417 was also detected in a multiprotein DNA processing complex isolated

* Corresponding author. Tel.: +91 22 25595417; fax: +91 22 25505151.
E-mail address: hsmisra@barc.gov.in (H.S. Misra).

0304-4165/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagen.2012.03.014

from this bacterium [24]. The DR_2417m ORF encoding this protein was annotated as having a genuine frameshift at its 996th position in *D. radiodurans* [25]. These intriguing features and detection of this protein from a functionally important pool of proteins prodded us to characterize this protein's functions both *in vitro* and *in vivo*.

Here, we report the characterization of DR2417 as a *D*NA processing *n*uclease found in a multiprotein DNA processing *c*omplex (hereafter referred as DncA) and for its role in growth and radiation resistance in *D. radiodurans*. Interestingly, nucleotide sequencing of PCR amplified DR_2417m ORF did not show the reported frameshift in coding region of this protein. Recombinant DncA showed an efficient nuclease activity on double stranded DNA substrates but a significantly low activity on both single stranded DNA and RNA substrates. It was found to be an endonuclease also having 3′→5′ exonuclease activity with a propensity for single/double strand junctions *in vitro*. The activity of this enzyme was inhibited by ATP but not by AMP. Molecular genetic studies showed that DncA had an indispensable role in growth and eventually in γ radiation tolerance in this bacterium, as the cells approaching the complete replacement of *dncA* with *nptII* marker gene did not survive.

2. Materials and methods

2.1. Bacterial strains and materials

D. radiodurans strain R1, ATCC13939, was a generous gift from Dr. M. Schaefer [26] and was grown in TGY medium (0.5% Bacto tryptone, 0.3% Yeast Extract, 0.1 % Glucose). The *Escherichia coli* strain DH5α was used for maintaining the cloned gene while *E. coli* BL21 (DE3) pLysS was used for expression of the recombinant protein. *E. coli* cells harboring pET28a+, an expression vector, and its derivative were grown at 37 °C in LB broth or LB agar plates supplemented with kanamycin (25 μg/ml). All the molecular biology grade chemicals were purchased from Sigma Chemical Company, USA, Bethesda Research Laboratory, USA, and Sisco Research Laboratory, India. The restriction enzymes and DNA modifying enzymes were obtained from Roche Biochemicals, Germany, New England Biolabs, USA, and Bangalore Genei, India.

2.2. Gamma irradiation and cell survival studies

Gamma radiation treatment to bacterial cells was carried out as described earlier [18]. In brief, the exponentially growing cells of wild type and partial knockout mutants of *D. radiodurans* were washed and resuspended in 1/10th volume of normal saline. These cells were irradiated with different doses of γ radiation on ice, at a dose rate 6.47 kGy/h at 4 °C in γ chamber (GC 5000, ^{60}Co, Board of Radiation and Isotopes Technology, DAE, India) as described earlier [27] and the colony forming units were determined as described earlier [28]. For the effect of DR2417 (DncA) in *E. coli*, the *dncA* was cloned in pTRC99 (Pharmacia) and recombinant protein was expressed in *E. coli* AB1157. The expression of protein was confirmed by SDS-PAGE. Transgenic *E. coli* AB1157 expressing DncA were checked for their survival under various doses of γ radiation and UV radiation as described in [29]. Treated cells were plated on LB agar with appropriate antibiotics and 50 μM IPTG and the cell forming units at 37 °C, were determined.

2.3. Gene expression studies

Total RNA was extracted from γ radiation treated cells at different time interval of PIR and cDNA was prepared using modified protocol used in [30] and described in [19]. PCR amplification of internal fragment of *dncA* was carried out using gene specific primers. The equal amount of template used in different samples was ascertained with amplification of internal fragment of *dr1343* encoding glyceraldehyde 3-phosphate dehydrogenase (Gap) as an internal control using dr1343F (5′ CTTCACCAGCCGCGAAGGGGCCTCCAAGC 3′) and dr1343R (5′GCCCAGCACGATGGAGAAGTCCTCGCC 3′) primers. Products were analyzed on 1% agarose gels. Amount of PCR amplified product would reflect the levels of corresponding gene transcripts at different PIR period. DNA band intensity was quantified by densitometric scanning and plotted as the ratio of *dncA*/*dr1343* against PIR time.

2.4. Nucleotide sequencing and bioinformatic studies

DR_2417m was PCR amplified using sequence specific primers (Forward primer, (5′GAAT TCCATATGACCAGACCAGAACAA3′ and reverse primer, 5′CGGGATCCTCAATCAA AGAAAGGGAA3′) from genomic DNA of *D. radiodurans* R1. PCR product was purified and full-length gene was sequenced using both terminal primers used for PCR amplification and internal primers (dr2417F3 — 5′GCACGTCGTCGTTCACGT 3′ and dr2417R3 — 5′GTCGGCGAACTGCAAGACCA 3′) on automated DNA sequencer (ABI PRISM 3100 Avant Genetic Analyzer) using dideoxy chain termination chemistry. Nucleotide sequence was translated into protein sequence. Functional motif search and structure prediction studies were carried out as described in earlier analysis in [23]. In brief, the amino acid sequence of DncA, annotated as a hypothetical protein, was subjected to a BLAST search at the NCBI portal. The close homologs were aligned with protein sequence in ClustalX, and conserved motifs were checked and further a bootstrapped phylogenetic tree was made (Fig S1). For making structure model of complete DR2417 molecule, its amino acid sequence was submitted to the Swiss model server in automated mode and the model was visualized in PyMOL as outlined in [31].

2.5. Construction of DncA expression plasmid

Genomic DNA was prepared from exponentially growing cultures of *D. radiodurans* as described earlier [32]. The standard recombinant DNA techniques including plasmid DNA isolation from *E. coli* were used as described in [33]. The *dncA* [25] coding region was PCR amplified from total genomic DNA of *D. radiodurans* using gene specific primers (Forward primer, 5′GAATTCCATATGACCAGACCAGAACAA3′ and reverse primer, 5′CGGGATCCTCAATCAAAGAAAGGGAA3′) and having required restriction endonuclease site incorporated at the 5′ ends in respective primers. The PCR product was sequenced to confirm the identity of *dncA* and the absence of mutation in the coding region of the gene. It was cloned at *Nde*I–*Bam*HI sites in pET28a+. The recombinant plasmid containing *dncA* gene under IPTG inducible T7 promoter, was named as pET2417.

2.6. Purification of recombinant protein

Transgenic *E. coli* BL21 cells harboring pET2417 were induced with 100 μM IPTG and cells expressing recombinant protein were disrupted by mild sonication at 4 °C. Protein was purified under denaturing conditions using nickel-affinity chromatography as described in manufacturer's protocols (QIAGEN, Germany). In brief, the cell pellet was suspended in lysis buffer A (10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM PMSF, 10 mM $MgCl_2$, lysozyme (1 mg/ml) and incubated at 4 °C for 30 min. Cells were sonicated on ice with 50% duty cycle at 4 μ tip with the sonication pulse of 30 s. Cell suspension was centrifuged at 5000 ×*g* for 5 min followed by 20,000 ×*g* for 30 min at 4 °C. The pellet presumably containing recombinant protein in inclusion was dissolved in buffer B (20 mM Tris–HCl, pH 8.0, 8 M urea and 5 mM $MgCl_2$), incubated at 37 °C for 30 min followed by centrifugation at 20,000 ×*g* for 30 min. The clear supernatant was mixed with required quantity of 50% Ni-NTA agarose slurry and passed through column in a buffer system containing 10% glycerol and 2% ethanol. The column was washed with buffer C (buffer B at pH 6.3). The (his)6-tagged recombinant protein was eluted from matrix

with elution buffer D (buffer B, pH 5.3) followed by buffer E (buffer B, pH 4.6). The fractions containing more than 95% pure protein were pooled and refolded by employing the serial dilution of urea in a refolding buffer (100 mM Tris–HCl, 400 mM arginine, 0.5 mM oxidized glutathione, 5 mM reduced glutathione and 10% glycerol), with increasing concentration of dithiothreitol (0–5 mM). This preparation of protein was further purified by both Q-sepharose and SP-sepharose ion-exchange and Sephacryl S-100 HR gel filtration chromatography. Purity of recombinant DncA was checked on SDS-PAGE. The fractions showing more than 99% purity were pooled and dialyzed overnight in buffer H (50 mM Tris–HCl, pH7.6, 50 mM NaCl, 1 mM PMSF, 1 mM DTT, 5 mM $MgCl_2$ and 50% glycerol, pH 8.0). Protein concentration was determined using Bradford's dye binding method.

2.7. Mass spectrometric analysis

Matrix assisted laser desorption ionization (MALDI) mass spectrometric analysis of this protein was carried out using standard procedure for in-gel trypsin digestion [34]. In brief, the gel pieces were dehydrated with 100 μl of acetonitrile and dried in a vacuum centrifuge. The dried gel pieces were rehydrated with 20 μl of 50 mM ammonium bicarbonate containing 0.2 μg modified trypsin (Promega) for 45 min on ice. After removal of solution, 70 μl of 50 mM ammonium bicarbonate was added. The digestion was performed overnight at 37 °C. The peptide solution was desalted using C18 nano column. Custom-made chromatographic columns were used for desalting and concentration of the peptide mixture prior to mass spectrometric analysis. A column consisting of 100–300 nL of poros reverse phase R2 material (20–30 μm bead size, PerSeptive Biosystems) was packed in a constricted GE Loader tip (Eppendorf, Hamburg, Germany). A 10 ml syringe was used to force liquid through the column by applying a gentle air pressure. Thirty microliters of the peptide mixture from the digestion supernatant was diluted 30 μl in 0.1% TFA, loaded onto the column, and washed with 30 μl of 0.1% TFA. For analyses by MS, peptides were eluted with matrix solution (15–20 g/L 4HCCA in 60% acetonitrile/0.1% TFA). Aliquots (30 μl) of peptide sample solution were loaded onto the MALDI-TOF MS sample plate together with 0.5 μl of the matrix solution (15–20 g/L 4HCCA in 60% acetonitrile/0.1% TFA). The MALDI analysis was carried out using a delayed-extraction reflectron time-of-flight mass spectrometer (Model Voyager DE-STR MALDI-TOF; ABI, USA) by outsourcing it at AB Sciex Instruments, USA. The mass of peptides was compared with the masses of *in silico* generated tryptic fragment of DncA protein sequence.

2.8. DNA binding and nuclease activity assay

The DNA binding and nuclease activities of this protein were determined as described earlier [22,24]. In brief, the 1 μg of purified recombinant protein was incubated with 1 μg dsDNA (PCR product) and 1 μg of ØX174/M13 mp18 virion DNA (ssDNA) separately at 37 °C for 10 min, in buffer containing 10 mM Tris–HCl pH8.0, 50 mM NaCl and 1 mM DTT supplemented with 5 mM $MnCl_2$ as required. Reaction was stopped with 5 mM EDTA in 95% formamide and heated at 65 °C for 15 min when required to disrupt nucleoprotein complex for monitoring nucleolytic products. The reaction products were analyzed on 1% agarose gel. For studying the effect of nucleotides, the 2 mM each of ATP, AMP, GTP and GMP was pre-incubated with protein before reaction was initiated with the addition of DNA substrate.

2.9. DNA processing activity assay

For checking the DNA processing activity, the double stranded DNA substrates with varying structure like 5′ overhang, 3′ overhang, recessed end, stem–loop, forked end and ssDNA substrates were prepared. One of the strands was labeled at either 5′ or 3′ end as required, and purified. Nearly equimolar ratio of both labeled strand and its complementary unlabeled strand was mixed in annealing Buffer (50 mM Tris–HCl pH 7.6, 75 mM NaCl, 1 mM DTT), and incubated at 95 °C for 3 min followed a gradual cooling to 25 °C on heating block. The double stranded DNA substrate was made by annealing FDs2417 (*5′* AGCACATCGGCTTCGCCGGGTGACACAACCGCT3′) and RDs2417 (5′ AGCGGTTGT GTCACCCGGCGAAGCCGATGTGCT3′) oligos, and 3′ and 5′ overhang DNA were generated by annealing 99F (5′GGAATTCCAAT-GAACCGCAAAACCGCAAAAACC GTACCGA3′) with 5OV99F (5′TC GGTA CGGTTTTTGCGGTTCATA 3′) and 3OV99F (5′ TTTTTGCGGTTCATATG-GAATTCC3′) oligos, respectively. The HP78 (5′GTTTCTA TTCAGCCCT-TTGACGTAATCCAGCCCCGGGTTTTCCCGGGGCTGGATTACGTCAAAGGGC-TGAATAGAAAC3′) as a stem–loop DNA [35], and 99F and FDs2417 oligonucleotides separately as ssDNA substrates were used. DNA processing activity was checked as described earlier [22]. In brief, 5 pmol of labeled DNA substrate was incubated with 100 ng of DncA in the assay buffer and aliquots were taken at different time of incubation. The reaction mixture was separated on 14% denaturing Urea-PAGE gel, and products were visualized by autoradiography.

2.10. RNase activity assay

Effect of DncA on RNA was evaluated using total RNAs purified from *D. radiodurans* using standard protocol as described in [33], and a commercially synthesized RNA substrate as described in [36]. Nearly 1 μg total RNA of *D. radiodurans* was incubated with 100 ng DncA at 37 °C for 10 min, in a buffer containing 10 mM Tris–HCl pH 8.0, 50 mM NaCl and 1 mM DTT supplemented with 5 mM $MnCl_2$/$ZnCl_2$, as required. Reaction was stopped with 5 mM EDTA in 95% formamide and heated at 65 °C for 15 min when required to disrupt nucleoprotein complex for monitoring nucleolytic products. The reaction products were analyzed on 1% agarose gel. Total RNA purified from exponentially growing *E. coli* was treated with DNase free of RNase (GE, Healthcare). Purified RNA was labeled with polynucleotide kinase in standard reaction buffer supplemented with [^{32}P] γ-ATP. Radio labeled substrate was purified by ethanol precipitation at −20 °C, and free nucleotide was removed. Approximately 500 K cpm equivalent labeled RNA substrate was incubated with 100 ng purified DncA/1 unit exonuclease I/200 ng RNaseA at 37 °C for 1 h in a buffer containing 10 mM Tris–HCl pH 8.0, 50 mM NaCl and 1 mM DTT supplemented with 5 mM $ZnCl_2$. Reactions were stopped with 5 mM EDTA in 95% formamide and heated at 65 °C for 15 min. TCA soluble counts were monitored as described earlier [27]. Further, the 20mer synthetic RNA (5′ UGGUGGUGGAUCCCGGGAUC3′) as used in earlier studies [36], was commercially synthesized and it was labeled at 5′end with polynucleotide kinase. The 5 pmol of labeled RNA substrate was incubated with 100 ng of DncA in the assay buffer (10 mM Tris–HCl pH 8.0, 50 mM NaCl and 1 mM DTT supplemented with 2 mM $ZnCl_2$) and aliquots were taken at different time of incubation. The reaction mixture was separated on 14% denaturing Urea-PAGE gel, and products were visualized by autoradiography.

2.11. Generation of dr2417 knockout and copy number studies

For generating *dncA* gene deletion in *D. radiodurans* R1 genome, the required construct was made using strategy as described in [27]. In brief, 1 kb upstream to start codon and 1 kb downstream to stop codon of DR2417 were PCR amplified using gene specific primers. The 2417upF (5′ATGGTACCGCCGCAGCCTCGCCA 3′) and 2417upR (5′ ATGGGCCCTATGCCTCCGCTTCCCT 3′) primers for upstream fragment and 2417dnF (5′ATGGATCCGTGGGAGCGGGCAGG 3′) and 2417dnR (5′ CGTCTAGAACGTTG CTGCCTTCG 3′) primers were used for downstream fragment. These fragments were cloned at *KpnI-ApaI* and *BamHI-XbaI*, respectively in pNOKOUT [27] to yield pNOK2417. The pNOK2417 is a suicidal vector in *D. radiodurans*, containing an expressing cassette of *nptII* flanked by DNA sequences homologous to

the regions flanking the *dncA* locus in the genome of this bacterium. Integration of *nptII* cassette in genome of *D. radiodurans* provides stable expression of kanamycin resistance to the cells and eventually marked the replacement of *dncA* with *nptII* through homologous recombination. The recombinant plasmid was linearized and transfer to *D. radiodurans*. Transformants were grown in TGY containing kanamycin (8 μg/ml) for several generations. The presence of *dr2417* and *nptII* genes was monitored in surviving fraction at every subculture using both internal primers dr2417F3 (P3) and dr2417R3 (P4) as well as flanking primers P1 (5′ATGACCAGACCAGAACAA3′) and P2 (5′TCAATCAAAGAAAGGGAA3′) for *dncA* and NptF (5′ AGGCCACGTTGTGTCTCA3′) and NptR (5′TGCTCTGCCAGTGTTACA3′). The yield of PCR product using internal primers of *dncA* was compared with the yield of PCR product of *dr1343* gene amplified using dr1343F and dr1343R primers as described above, as an internal reference control. For copy number determination, the genomic DNA was prepared from cells collected at different stages of sub-culturing and treated with RNase free from DNase before number of copies of *dncA* was determined by real time PCR (RT-PCR) using dr2417F3 and dr2417R3 internal primers as described above. Data was normalized with *dr1343* as an internal reference. The clones having different status of *dncA* in genome were checked for growth and γ radiation resistance.

3. Results and discussion

3.1. DncA is structurally similar with β-CASP family of nucleases

DR2417 (DncA) was identified from a pool of nucleotide binding proteins of *D. radiodurans*, exhibiting the ATP sensitive nucleolytic degradation of dsDNA [18] and independently in a multiprotein DNA processing complex also showing the ATP inhibition to its nuclease activity [24]. DncA was annotated a hypothetical protein encoded from DR_2417m ORF reported to have a genuine frameshift at 996th in this ORF [25]. In order to understand the functional significance of this protein, as it exists in the multiprotein DNA processing complex, and to ascertain the frameshift in annotated ORF, the PCR product of DR_2417m was completely sequenced. Data showed the absence of reported frameshift at 996th position (Fig S2), which may be referred as DncA. The correct nucleotide sequence of erstwhile DR_2417m was submitted to GenBank (Accession No. JQ432552). Amino acid sequences derived from *dncA* gene showed homology with various members of β CASP family proteins including a well-characterized NHEJ repair protein "Artemis", RNase J and hSNM1. Earlier, Callebaut and colleagues [37] had also identified DR2417m as a β-CASP family protein on the basis of its N-terminal similarities with other proteins. However, they evaluated it to be an inactive paralog to an active gene DRA0069 [37] perhaps due to its unusual C-terminal structure. Due to wrongly annotated frameshift at 996th position, the amino acid sequences beyond 332nd position were not correct and thus motif B and motif C could not be assigned in the earlier analysis. Now, the C-terminal part of DncA has a regular structure, 19 amino acids shorter and both motif B and motif C are very apparent, as compared to erstwhile DR2417m. Nevertheless, it showed all the metallo β-lactamase motifs M1-M4, motif A, motif B and motif C of the β-CASP fold including a key motif H-X-H-X-DH, present in nucleases (Fig. 1). Interestingly, DncA shows a histidine at motif C, which is similar to motif C of other β CASP nucleases also exhibiting RNase activity. Phylogenetic analysis revealed that DncA was closest to Dgeo_2150 (http://www.jgi.doe.gov) and not its paralog DR_A0069 and high bootstrap values indicate that this relationship is true. Since DncA was found in a fraction processing dsDNA, we surmised, it could have an Artemis type function. However, the critical motif C, which distinguishes DNases from RNases within this family of proteins,

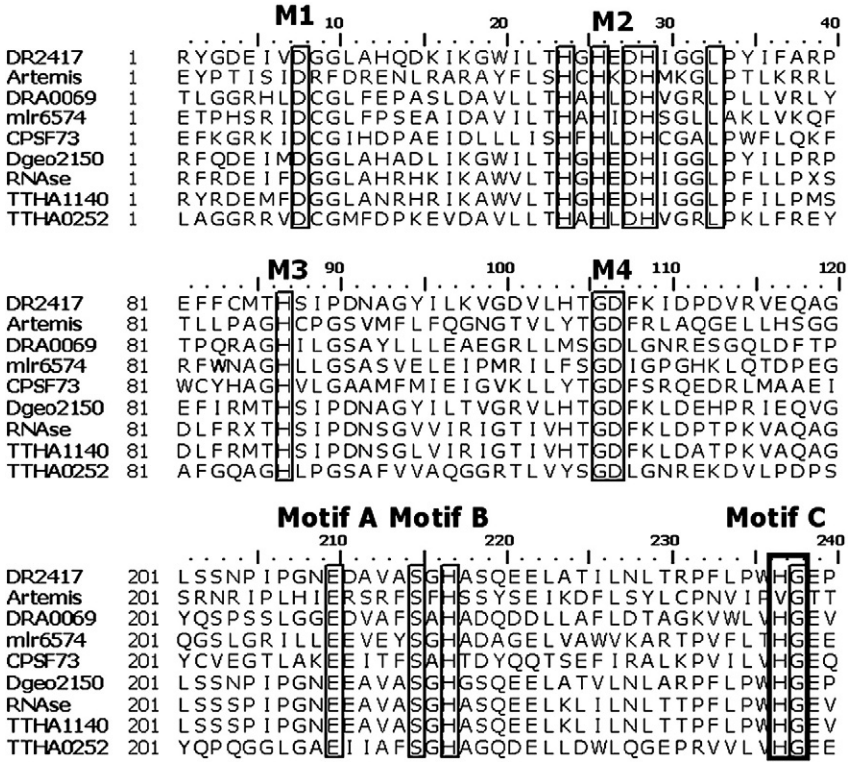


Fig. 1. Bioinformatic analysis of DR2417 (DncA). Amino acid sequence of DncA was searched for similarities with existing proteins in database. DncA was subsequently aligned with members of the beta CASP family proteins in ClustalX. Multiple sequence alignment of DncA with its homologs showed the presence of conserved β lactamase motif 1 (M1), 2 (M2), 3 (M3) and 4 (M4) and Motif A, Motif B and Motif C in β CASP domain of the proteins belonging to this family. For clarity of representation, the regions of alignment having gaps were removed during analysis.

showed the presence of histidine (H) in DncA while this amino acid is valine (V) in Artemis. These analyses indicated that DncA is a member of β-CASP family nuclease and is much closer to RNaseJ.

3.2. DncA was characterized as a manganese dependent DNase

Recombinant DncA (~63 kDa) was purified to near homogeneity from transgenic *E. coli* as described in methods (Fig S3). Mass spectrometric analysis confirmed the identity of recombinant protein as DncA (Fig S4). The mass of peptides obtained experimentally and tryptic fragments generated *in silico* for DncA, showed complete match (Fig. 2A) and these peptides could be mapped on DncA protein (Fig. 2B). The peptides detected from purified recombinant protein were similar to *in silico* generated tryptic fragments of C-terminal of DncA. Since, the C-terminal of DncA is likely to different from original peptide due to a wrong annotation of frame shift in reading frame, MALDI analysis further confirmed that PCR cloned erstwhile DR_2417m ORF, did not have frame shift mutation as reported earlier [25]. Therefore, the recombinant protein DncA is correct protein encoded from *dncA* gene on the genome of this bacterium.

Nucleases have metal ion preference for their activity *e.g.* most RNases prefer Zn^{2+}, DNases prefer Mg^{2+} or Mn^{2+} [38] which was also evident in functional homologs of DcnA from both bacteria and higher organisms [21,39]. Therefore the activity of DncA was checked in the presence of Zn^{2+}, Mn^{2+} and Mg^{2+} on both ssDNA and dsDNA. It showed a strong binding with both ssDNA and dsDNA in presence of Mg^{2+} ion (Fig. 3A). DncA incubated with DNA in presence of Mn^{2+} showed the better nuclease activity followed by Zn^{2+} and nearly no nuclease activity was observed with Mg^{2+} (Fig. 3B). Interestingly, the DncA interaction with DNA was inhibited by ATP and GTP but not with either AMP or GMP (Fig. 3C). A reliable model of DncA was obtained from the SWISSMODEL server, with RNase J from *Thermus thermophilus* HB27 (PDB id 3T30 and gene id TTC0775) (Q-Mean score of 0.3 and R.M.S deviation of 0.063). RNase J can hydrolyze 5′P RNA substrate but not 5′PPP RNA. This could be accounted to the presence of its nucleotide binding pocket just upstream to the active site, which might prevents entry of triphosphate substrates. DncA model also showed placement of nucleotide-binding pocket with respect to active site (data not shown) and that could be attributed to on the inhibition of DncA nuclease activity by triphosphates and not monophosphates. Unlike other β-CASP family nucleases, which are ambivalent [40], DncA was a Mn^{2+} dependent nuclease with a high specificity for DNA. This is interesting because Mn^{2+} inhibits the activity of the homologous RNase TTHA0252 from *T. thermophilus* [41] and there are no known cases of other β CASP family nucleases that prefer Mn^{2+}. Recently however, there are reports suggesting that the

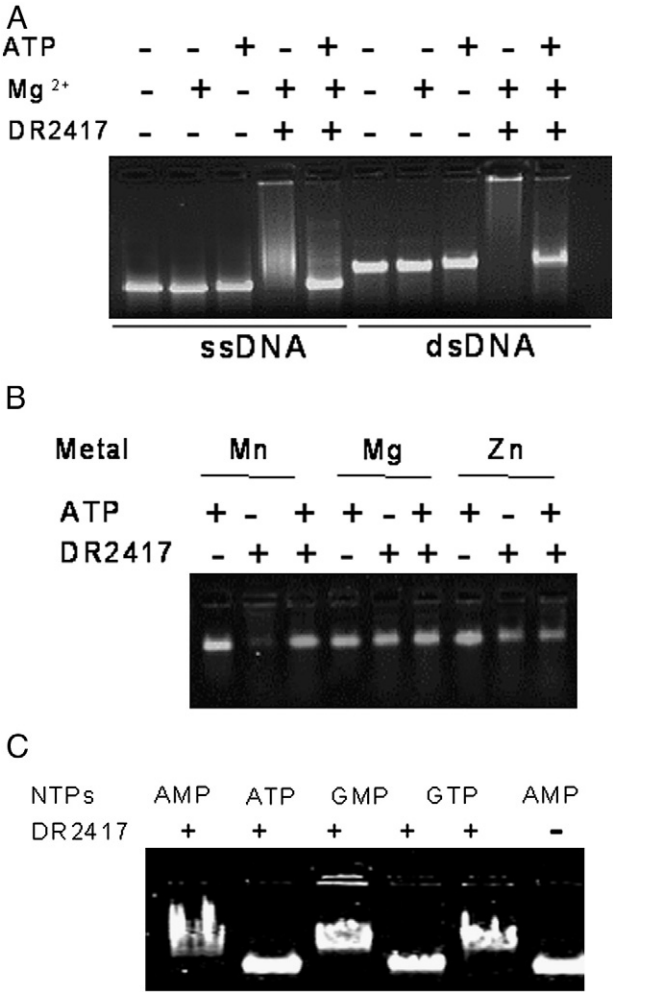


Fig. 3. Nucleolytic activity assay of DncA on DNA substrates. The 100 ng pure recombinant DncA was incubated with M13mp18 virion DNA (ssDNA) and PCR amplified DNA (dsDNA) in buffer containing 10 mM $MgCl_2$, in presence and absence of 1 mM ATP (A). The plasmid DNA digested with *Hind*III was incubated for 30 min with 100 ng purified DncA in buffer (10 mM Tris–HCl pH8.0, 50 mM NaCl and 1 mM DTT) supplemented with 5 mM $MnCl_2$ (Mn), 10 mM $MgCl_2$ (Mg) and 5 mM $ZnCl_2$ (Zn) as required, and nuclease activity was checked in presence and absence of 1 mM ATP (B). Reaction mixture was heated at 85 °C for 15 min in presence of 65% formamide to break nucleoprotein complex. The effect of ATP, GTP, AMP and GMP was also checked on its ssDNA binding activity (C). Products were analyzed on 1% agarose gel and change if any, was compared with respective substrates (S) control. Experiments were repeated at least three times and results of a typical experiment shown.

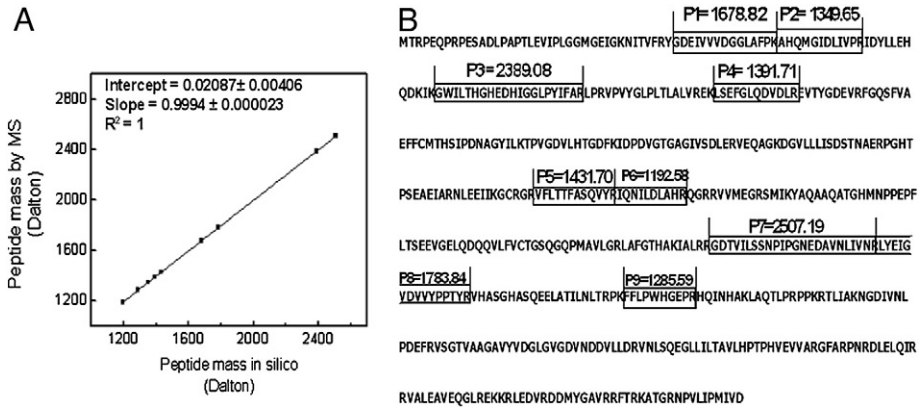


Fig. 2. Mass spectrometric analysis for confirmation of protein structure derived from the corrected nucleotide sequence of *dncA*. Recombinant DncA was purified to near homogeneity and subjected to peptide mass fingerprinting. The masses of the peptides were searched in Mascot database and identity of protein was confirmed as a polypeptide encoded from DncA (Fig S2). Nine of these peptides showed absolute match with *in silico* generated tryptic fragment of DncA as per revised protein sequence (A). All these peptides (P1–P9) were mapping on primary structure of DncA, which showed the detection of peptides (P5–P9) corresponding to C-terminal of DncA (B).

different species of *Deinococcus* including *D. radiodurans* accumulates higher levels of Mn^{2+}, which helps these bacteria in higher oxidative stress tolerance and protection of biomolecules from oxidative damage [10,42,43]. Several DNA repair enzymes like X family DNA repair polymerase [20,44], SbcCD enzyme [21] and DNA processing nuclease DR0505 [22] of *D. radiodurans* have been shown to prefer Mn^{2+} for their catalytic functions. A multiprotein DNA processing complex isolated from this bacterium, and contains DR2417, also prefers Mn^{2+} for the DNA processing activity [24]. These findings independently suggested that since, this bacterium accumulates higher concentration of manganese, its DNA metabolic proteins seem to have evolved for the better use of Mn^{2+} for their functions. Since DncA also shows a limited activity with Zn^{2+}, we presume that Mn^{2+} dependent activity is an evolutionary modification of this protein in this bacterium.

Nucleolytic activity of DncA on RNA substrate was evaluated using total RNA extracted from *D. radiodurans* and from *E. coli*. The recombinant DncA showed undetectable levels of nuclease activity on RNA substrate assayed in presence of either Mn^{2+} or Zn^{2+}, separately (Fig. 4A) and analyzed on agarose. Radio labeled RNA substrate when incubated with DncA showed a significant release of TCA soluble radiolabeled products from RNA substrate (Fig. 4B), but it was nearly 6 and 1.5 folds lower than RNaseA and exonuclease I controls, respectively. Incubation of DncA with 5′ kinased synthetic RNA for different time intervals showed the release of [^{32}P]-NMP from 30 min onward. However, complete degradation of total RNA substrate was not achieved in 2 h of incubation (Fig. 4C), indicating that DncA had a lower preference for the RNA substrate. These findings suggested that DncA was an efficient Mn^{2+} dependent DNase with a poor RNase activity, at least *in vitro*.

Our results showed that DncA acts preferentially on dsDNA, which is different from the known characteristics of RNaseJ, with which this protein had close structural similarities. Most of the earlier studies have shown that these molecules show a considerable movement to accommodate the substrate and show substrate specificity by forming a narrow cleft suitable only for RNA molecules. Molecular basis underlying the differential utilization of RNA substrate by DncA and RNase J is not clear. However, DncA lacks several β sheet regions as compared to RNase J and also, there are several local regions, especially loops where energy constraints exist in the model (Fig S5).

3.3. DncA showed 3′→5′ exonuclease and ssDNA/dsDNA junction endonuclease

The proposed model of ESDSA based DSB repair has suggested that the processing of shattered DNA in both polarities is required for efficient DSB repair [13,14]. However, the enzymes that could do DNA processing during ESDSA have not been precisely known but suggested to be among the recombination nucleases. A recent report implicates RecJ as an important nuclease involved in recombination repair [45]. Also, certain proteins having DNA ends processing and processive degradation of DNA strand in 5′→3′ or 3′→5′ polarity have been reported [20–22]. The possible role of DncA in processing of DNA substrate for ESDSA cannot be ruled out. We have also seen that the expression of *dncA* was highest at 1 h PIR (Fig S6), the time during which extensive nucleolytic degradation is seen. We therefore, checked the DNA processing activity of DncA on DNA substrates having various types of DNA ends like ssDNA, linear dsDNA, stem–loop DNA, 5′ overhang and 3′ overhang DNA substrates.

DncA showed very poor activity with a 5′ labeled 33mer single strand oligonucleotide [99F] substrate (Fig. 5A). When 5′ labeled 99F was annealed with equal size complementary oligonucleotide to make recessed ds33mer, the enzyme showed relatively faster degradation than ssDNA substrate, but the release of terminal radio labeled nucleotide was still seen after 2 h of incubation at 37 °C (Fig. 5B). These results strongly argued that although DncA possibly has 3′→5′ exonuclease activity on ssDNA and dsDNA substrates, it

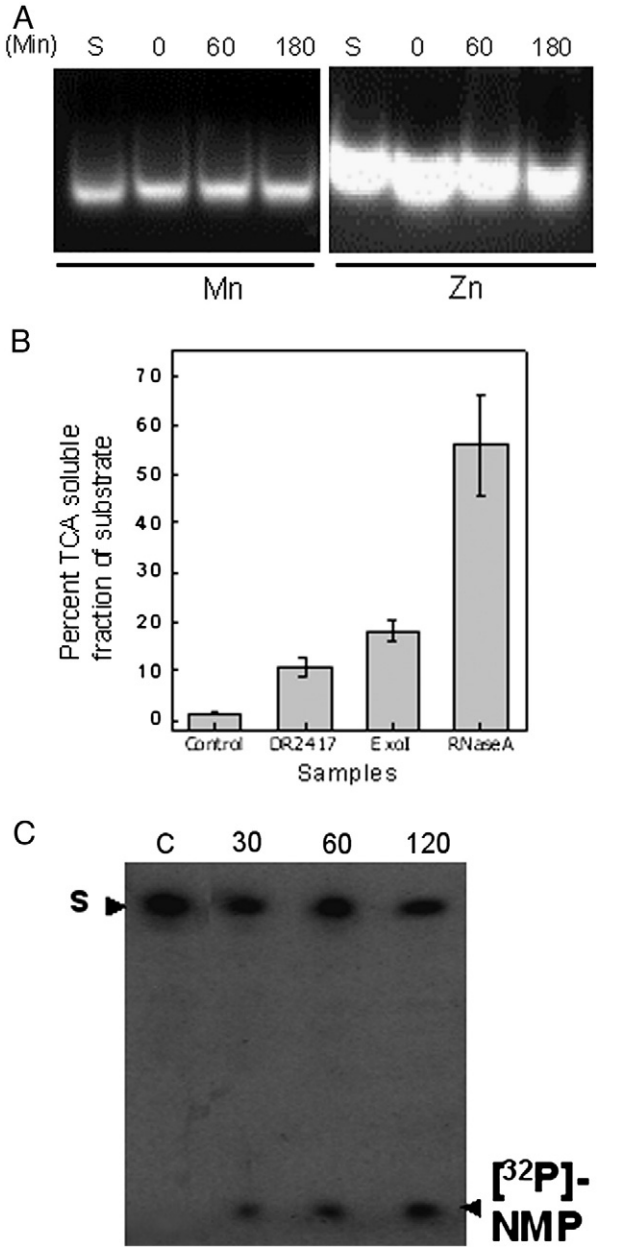


Fig. 4. Determination of RNase activity in recombinant DncA. Total RNA was isolated from *D. radiodurans* and incubated with DncA for different time interval in presence and absence of 5 mM $MnCl_2$ (Mn) and $ZnCl_2$ (Zn), separately and products were analyzed on agarose (A). Total RNA purified from *E. coli* was labeled and RNA degrading ability of recombinant DncA, and commercial preparation of exonuclease I (ExoI) and RNaseA (RNaseA) was measured as TCA soluble counts (B). RNase potential of DncA was further checked on labeled synthetic RNA substrate (C), incubated with DncA at 37 °C under standard assay conditions. Aliquots were taken at 30 min, 60 min and 120 min post incubation and products were analyzed on Urea-PAGE and detected by autoradiography (C).

acted slowly on synthetic substrate lacking 5′ phosphate on opposite strand. Hence, the above ds33mer substrate was phosphorylated at 5′ with cold ATP and incubated with DncA. Now the enzyme showed faster degradation with the release for terminal nucleotide in less than 30 min incubation (Fig. 5C). Although the results indicated that DncA had 3′→5′ exonuclease activity, the possibility of the products release due to exonucleolytic degradation in opposite polarity could not be ruled from this result. To check this possibility, DncA was incubated with DNA substrate having 37-nucleotide long stem and 4-nucleotide long loop structure (HP78) labeled on both ends separately. DncA incubated with HP78 radiolabeled at 5′ end showed a series of typical products that are normally observed during

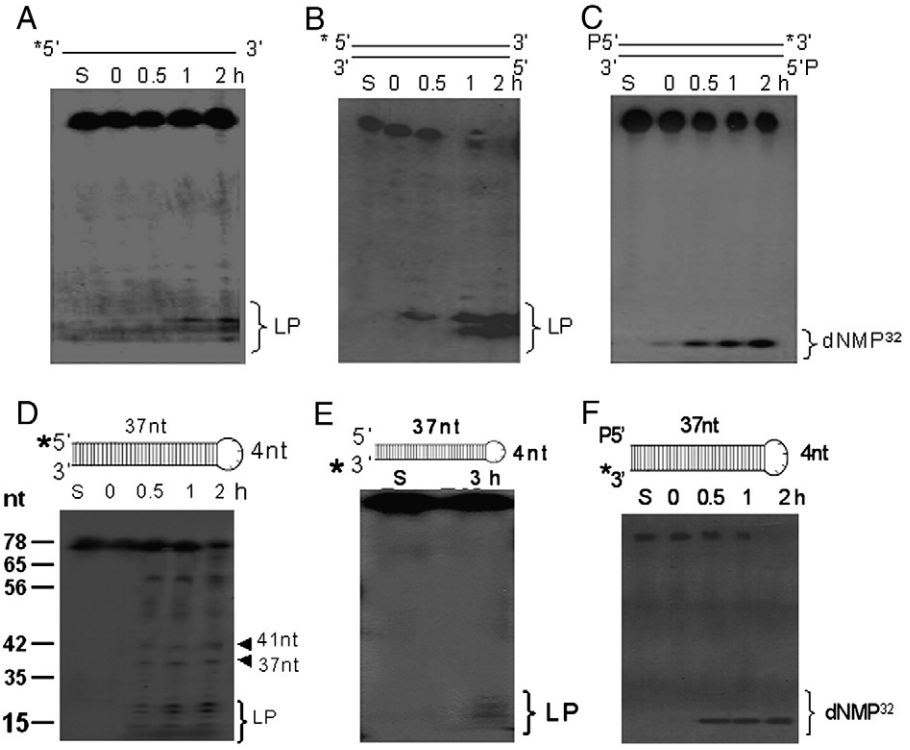


Fig. 5. DNA processing activity characterization of DncA on different types of DNA substrates. DNA substrates like 5′ labeled linear 33 bases ssDNA (A), recessed 33 bp dsDNA (B) and 78mer (HP78) stem loop DNA (D), and 3′ labeled recessed 33 bp dsDNA having 5′ phosphorylation in opposite strand (C), HP78 with (F) and without (E) 5′ phosphorylation, were prepared. These substrates (S) were incubated with DncA at different time interval at 37 °C. Reaction was stopped with DNA sequencing dye (50% formamide, 25 mM EDTA, pH8.0) and mixtures were heated at 95 °C for 15 min. Products were analyzed on 16% Urea-PAGE and autoradiograms were developed. Arrow marked showing limit products (LP), terminal mononucleotide ($dNMP^{32}$) and ss/dsDNA junction endonucleolytic products. Experiments repeated two times and data of a reproducible experiment is shown.

exonucleolytic degradation of DNA (Fig. 5D). This enzyme also produced the islands of products from 0.5 h onward incubation. This suggested that DncA has 3′→5′ exonuclease activities and a strong possibility of an endonuclease function at the ss/ds DNA junction, that could produce islands of products of 37nt and 41nt on 5′ labeled HP78. DncA however, showed very poor activity on 3′ radiolabeled HP78, which had a hydroxyl group at 5′end (Fig. 5E) and the 3′ radiolabeled [^{32}P] dNMP could be detected only after 3 h incubation. This further indicated that DncA possibly requires both 5′ phosphate and 3′ hydroxyl group in dsDNA for optimum activity. This hypothesis was tested with 3′ radio-labeled HP78 also phosphorylated at 5′ end with cold ATP. Enzyme showed a faster release of 3′ terminal [^{32}P] dNMP in less than 30 min incubation under similar condition (Fig. 5F). This result along with the results obtained when 3′ radiolabeled ds33mer substrate was kinased at 5′ position (Fig. 5C) suggested the requirement of 5′ phosphate for the 3′→5′ exonuclease activity of DncA. The requirement of 5′ phosphate and 3′ hydroxyl group for efficient function of DncA was further supported with the higher activity of DncA on kinased PCR amplified dsDNA as compare to non-phosphorylated PCR product, and the plasmid DNA linearized with restriction enzyme (data not shown). The requirement of 5′ phosphate for exonuclease activity has been shown as a characteristic function of β CASP family nuclease [46]. Interestingly, these results suggested that (i) dsDNA substrates that contain both 5′ phosphate and 3′ hydroxyl group, which would eventually be present in natural substrates were preferred substrate for the inherent 3′→5′ exonuclease activity of DncA and (ii) DncA was a ss/ds junction endonuclease.

3.4. Unlike in E. coli, DncA showed a dominant phenotype in D. radiodurans

Several deinococcal proteins expressed in *E. coli* and *vice versa* have shown remarkable effects on DNA damage tolerance of transgenic bacteria [26,35,36,47]. On the other hand the deletion of certain DNA repair genes from *D. radiodurans* genome does not affect normal growth of this bacterium [2,4]. While creating deletion mutation of *dncA* gene in *D. radiodurans* genome, we observed that the cells approaching complete replacement of *dncA* with *nptII* marker gene (Fig. 6), did not survive under normal growth conditions. Mutants having different copy numbers of this gene showed differential growth rate under normal conditions (Fig. 7). These cells, survived less when subjected to γ radiation stress as compared to the wild type *D. radiodurans* (Fig. 7). On the contrary when these cells were grown in absence of selection pressure, the *dncA* copy numbers increased several folds, which was reflected in growth recovery of these cells (data not given). This observation was very much similar to those favoring the essentiality of a gene in bacterial system. Recently, the indispensability of RecJ in growth and DSB repair of *D. radiodurans* has been demonstrated [45]. These results might allow us to speculate that indispensability of *dncA* was an essential gene for survival and eventual gamma radiation resistance in *D. radiodurans*. On the other hand the overexpression of this protein in wild type *E. coli* showed no significant effect (Fig S7). Further studies at cellular and molecular levels would be required for providing the direct evidence on indispensability of DncA in this bacterium.

This study has brought forth some interesting findings to suggest that *dncA* gene encodes a functional and stable protein in *D. radiodurans* that is a β CASP family nuclease. The reported mistake in form of a frameshift at 996th position in DR_2417m ORF has been corrected and submitted to database. Amino acid sequences deduced from the correct gene *dncA* showed best structural similarities with other members of β CASP family nucleases. Recombinant protein showed both exo- and endo-nuclease and needs a 5′ phosphate for its activity akin to Artemis, and inhibited by ATP as observed in hSNM1, a homolog of Artemis. However, uniqueness in DncA was its preference

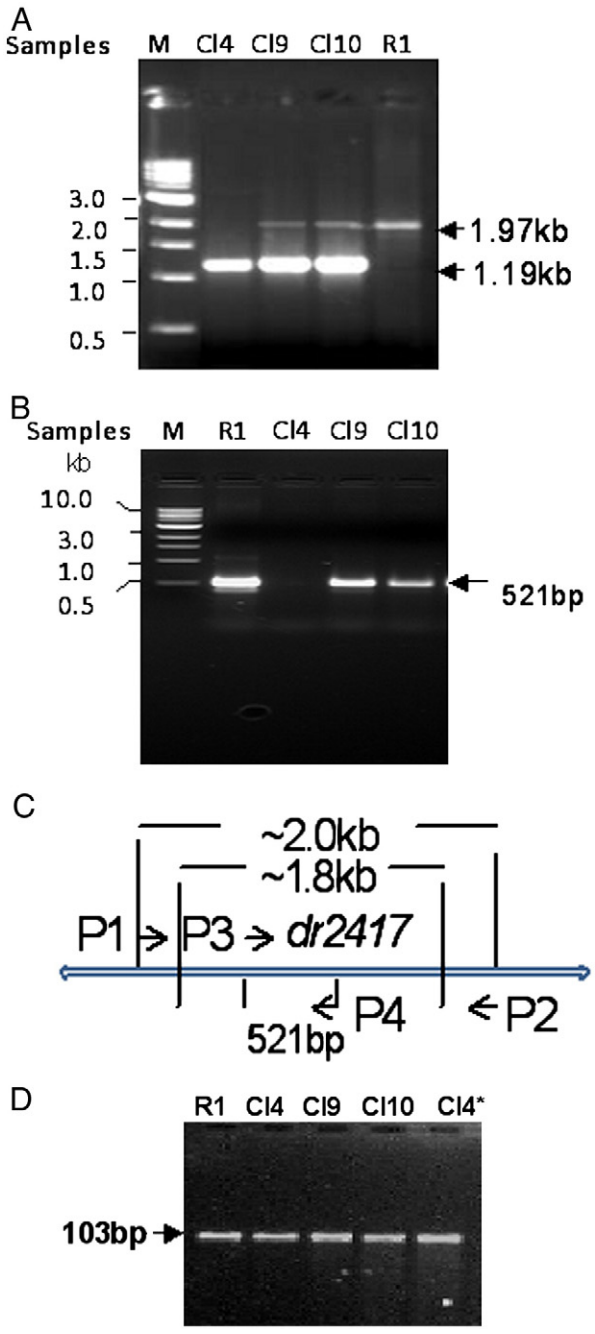


Fig. 6. Characterization of *D. radiodurans* cells for *dr2417 (dncA)* deletion. *Deinococcus* cells transformed with linear pNOK2417 were sub-cultured in TGY broth supplemented with Kan (8 μg/ml) and grown several generations. Certain clones were randomly picked up at specific time of sub-culturing and tentatively assigned as clone 9 (Cl9) clone 10 (Cl10) and clone 4 (Cl4). The genomic DNA was made and the status of *dncA* gene was checked by PCR amplification using flaking primers P1 and P2 (A) and internal primers P3 and P4 (B) and the amplification of the expected size products as marked in (C) was ascertained on agarose gel. Almost equal levels of amplification of *dr1343* gene taken as an internal reference control, in all the samples ascertained the equal amount of template DNA in each PCR reaction (D).

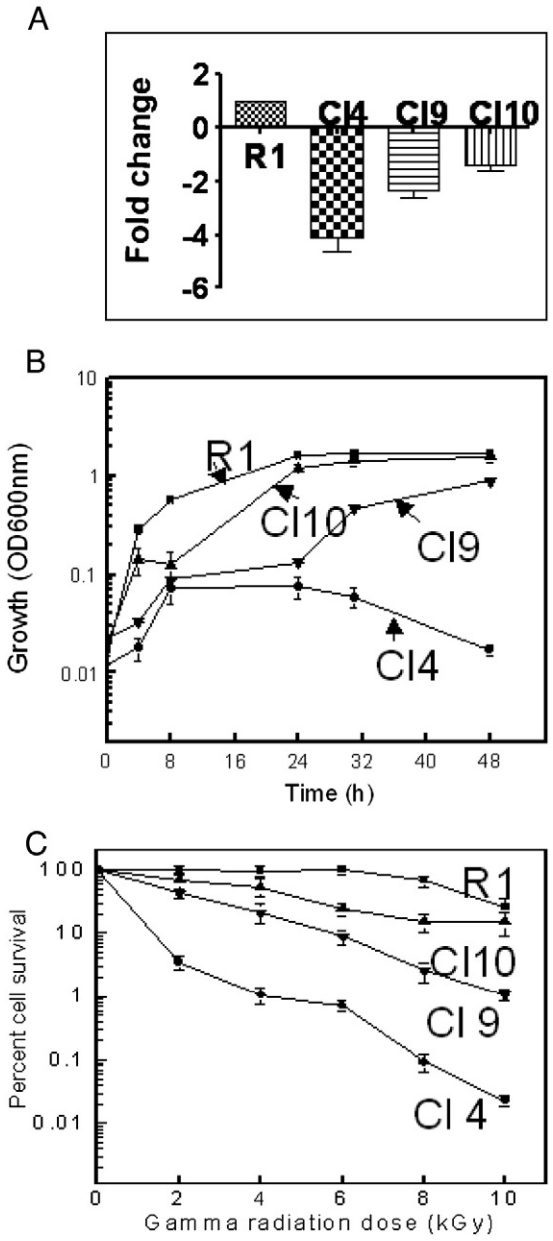


Fig. 7. Effect of *dncA* deletion on growth and γ radiation response of *D. radiodurans*. Genomic DNA was prepared from clone 4 (Cl4), clone 9 (Cl9), clone 10 (Cl10) as described in Fig. 6, and wild type (R1) cells. Copy number of *dncA* (*dr2417*) was determined by real time PCR using internal primers of *dncA* and *dr1343* as a reference control, as described in methods. Levels of *dncA* sequence amplification in each sample were normalized by taking levels of *dr1343* amplification in respective sample as reference control. Fold change in *dncA* copy number was calculated by dividing its PCR yields in respective clones with the same in wild type control (A). These clones were checked for growth characteristics (B) and responses to different doses of γ radiation (C). Different clones having different fold reduction in *dncA* copy number showed proportional growth retardation and sensitivity to γ radiation.

for dsDNA for its 3′→5′ exonuclease activity and a significantly low activity on ssDNA and RNA substrates. Although, the molecular basis for low RNase activity in DncA DNase is not clear, the possibilities of other amino acids and motifs involvement in determining the substrate specificity by Motif C of β CASP family nucleases could be speculated. Because, all the prokaryotic proteins of this family like ykqC from *Bacillus subtilis*, TTHA0252 from *T. thermophilus*, Pab-RNase J from *P. abyssi*, have a conserved histidine in motif C similar to DncA, but function as RNases [40,48,49]. However, the DNA specific nucleases of this family like Artemis and hSNM1 in eukaryotes have a conserved Valine in motif C and express a considerably low activity on RNA substrates [37]. Low RNase activity in DncA could be attributed to certain notable differences in the amino acids contributed to 3-D modeled structure of this protein (Fig S5). The strong endonuclease activity on HP78, a stem–loop substrate (Fig. 5D) also confirmed that DncA confers ssDNA/dsDNA endonuclease function. These results suggested that DncA was DNA processing enzyme.

Although, the functional significance of this protein is not clear in *D. radiodurans*, our results indicated that DncA plays an essential role

in growth and eventually γ radiation resistance in this bacterium. Inability to generate deletion mutation, a strong affinity of this protein to DNA, the protection of DNA from higher temperature and increasing the mobility of circular DNA (data not shown) together might allow to speculate the crucial role of this protein in structure maintenance and normal DNA metabolism. Inhibition of its nuclease activity by ATP is another novel feature of this protein, which could be useful in regulation of its functions under different physiological conditions. Recently, it has been shown that the levels of ATP changes during post irradiation recovery in *D. radiodurans* [18]. With the present data it is difficult to decipher the indispensable role of this protein. Since, the longer incubation of DncA with dsDNA protected it from proteinase K proteolysis (Fig S8) and it increased the mobility of circular DNA on agarose gel without cleavage (data not shown), the role of this protein in DNA scaffolding and thereby maintaining integrity of genome structure could be argued as one of the possible mechanisms of its action. The absence of such protein could make these processes defective and that might account to its indispensability in this organism. A wide occurrence of similar type of proteins in *Bacilli* and other bacteria (data not shown) might add further to the importance of this family protein for the growth and maintenance of living organisms. Although, further work will be needed to understand the exact mechanism of DncA function, the provided evidences suggest that DncA was a genuine DNA damage inducible novel nuclease of β CASP family as against the reported frameshift in genome sequence, and its DNA processing activity is regulated by levels of high energy phosphate by yet unknown mechanism, which might provide a plausible basis for its requirement in growth and eventually DNA damage tolerance in *D. radiodurans*.

Acknowledgements

Authors thank Dr S. K. Apte for his constant support and encouragement while pursuing this work and Mrs Swathi Kota, and Ms Kruti P. Mehta for critical readings and comments while preparation of this manuscript. Assistance of Dr (Mrs) Rita Mukhopadhyaya in nucleotide sequencing of this gene and Mr S. Jaykumar for real time PCR, is gracefully acknowledged. We gratefully acknowledge Dr (Mrs) Vidya A Kamble, for her initial work that led to detailed characterization of this protein.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.bbagen.2012.03.014.

References

[1] J.R. Battista, Radiation resistance: the fragments that remain, Curr. Biol. 10 (2000) R204–R205.
[2] M.M. Cox, J.R. Battista, *Deinococcus radiodurans* - the consummate survivor, Nat. Rev. Microbiol. 3 (2005) 882–892.
[3] K.S. Makarova, L. Aravind, Y.I. Wolf, R.L. Tatusov, K.W. Minton, E.V. Koonin, M.J. Daly, Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics, Microbiol. Mol. Biol. Rev. 65 (2001) 44–79.
[4] M. Blasius, S. Sommer, U. Hubscher, *Deinococcus radiodurans*: what belongs to the survival kit? Crit. Rev. Biochem. Mol. Biol. 43 (2008) 221–238.
[5] D. Slade, M. Radman, Oxidative stress resistance in *Deinococcus radiodurans*, Microbiol. Mol. Biol. Rev. 75 (2011) 133–191.
[6] L.M. Markillie, S.M. Varnum, P. Hradecky, K.K. Wong, Targeted mutagenesis by duplication insertion in the radioresistant bacterium *Deinococcus radiodurans*: radiation sensitivities of catalase (katA) and superoxide dismutase (sodA) mutants, J. Bacteriol. 181 (1999) 666–669.
[7] N.P. Khairnar, H.S. Misra, S.K. Apte, Pyrroloquinoline-quinone synthesized in *Escherichia coli* by pyrroloquinoline-quinone synthase of *Deinococcus radiodurans* plays a role beyond mineral phosphate solubilization, Biochem. Biophys. Res. Commun. 312 (2003) 303–308.
[8] B. Tian, Z. Sun, S. Shen, H. Wang, J. Jiao, L. Wang, Y. Hu, Y. Hua, Effects of carotenoids from *Deinococcus radiodurans* on protein oxidation, Lett. Appl. Microbiol. 49 (2009) 689–694.
[9] M.J. Daly, E.K. Gaidamakova, V.Y. Matrosova, A. Vasilenko, M. Zhai, A. Venkateswaran, et al., Accumulation of Mn(II) in *Deinococcus radiodurans* facilitates gamma-radiation resistance, Science 306 (2004) 1025–1028.
[10] M.J. Daly, E.K. Gaidamakova, V.Y. Matrosova, A. Vasilenko, M. Zhai, R.D. Leapman, et al., Protein oxidation implicated as the primary determinant of bacterial radioresistance, PLoS Biol. 5 (2007) e92.
[11] M.J. Daly, L. Ouyang, P. Fuchs, K.W. Minton, In vivo damage and recA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*, J. Bacteriol. 176 (1994) 3508–3517.
[12] M.J. Daly, K.W. Minton, An alternative pathway of recombination of chromosomal fragments precedes recA-dependent recombination in the radioresistant bacterium *Deinococcus radiodurans*, J. Bacteriol. 178 (1996) 4461–4471.
[13] K. Zahradka, D. Slade, A. Bailone, S. Sommer, D. Averbeck, M. Petranovic, A.B. Lindner, M. Radman, Reassembly of shattered chromosomes in *Deinococcus radiodurans*, Nature 443 (2006) 569–573.
[14] D. Slade, A.B. Lindner, G. Paul, M. Radman, Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*, Cell 136 (2009) 1044–1055.
[15] A. Kuzminov, Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda, Microbiol. Mol. Biol. Rev. 63 (1999) 751–813.
[16] I. Narumi, K. Satoh, S. Cui, T. Funayama, S. Kitayama, H. Watanabe, PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation, Mol. Microbiol. 54 (2004) 278–285.
[17] D.R. Harris, M. Tanaka, S.V. Saveliev, E. Jolivet, A.M. Earl, M.M. Cox, J.R. Battista, Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1, PLoS Biol. 2 (2004) e304.
[18] V.A. Kamble, Y.S. Rajpurohit, A.K. Srivastava, H.S. Misra, Increased synthesis of signaling molecules coincides with reversible inhibition of nucleolytic activity during postirradiation recovery of *Deinococcus radiodurans*, FEMS Microbiol. Lett. 303 (2010) 18–25.
[19] Y.S. Rajpurohit, H.S. Misra, Characterization of a DNA damage-inducible membrane protein kinase from *Deinococcus radiodurans* and its role in bacterial radioresistance and DNA strand break repair, Mol. Microbiol. 77 (2010) 1470–1482.
[20] E. Bentchikou, P. Servant, G. Coste, S. Sommer, Additive effects of SbcCD and PolX deficiencies in the in vivo repair of DNA double-strand breaks in *Deinococcus radiodurans*, J. Bacteriol. 189 (2007) 4784–4790.
[21] V.A. Kamble, H.S. Misra, The SbcCD complex of *Deinococcus radiodurans* contributes to radioresistance and DNA strand break repair in vivo and exhibits Mre11-Rad50 type activity in vitro, DNA Repair (Amst) 9 (2010) 488–494.
[22] S. Kota, C.V. Kumar, H.S. Misra, Characterization of an ATP-regulated DNA-processing enzyme and thermotolerant phosphoesterase in the radioresistant bacterium *Deinococcus radiodurans*, Biochem. J. 431 (2010) 149–157.
[23] A.D. Das, H.S. Misra, Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage, Microbiology (SGM) 157 (2011) 2196–2205.
[24] S. Kota, H.S. Misra, Identification of a DNA processing complex from *Deinococcus radiodurans*, Biochem. Cell Biol. 86 (2008) 448–458.
[25] O. White, J.A. Eisen, J.F. Heidelberg, E.K. Hickey, J.D. Peterson, R.J. Dodson, D.H. Haft, M.L. Gwinn, W.C. Nelson, D.L. Richardson, et al., Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1, Science 286 (1999) 1571–1577.
[26] M. Schaefer, C. Schmitz, R. Facius, G. Horneck, B. Milow, K.H. Funken, J. Ortner, Systematic study of parameters influencing the action of Rose Bengal with visible light on bacterial cells: comparison between the biological effect and singlet-oxygen production, Photochem. Photobiol. 71 (2000) 514–523.
[27] N.P. Khairnar, V.A. Kamble, H.S. Misra, RecBC enzyme overproduction affects UV and gamma radiation survival of *Deinococcus radiodurans*, DNA Repair (Amst) 7 (2008) 40–47.
[28] H.S. Misra, N.P. Khairnar, S. Kota, S. Shrivastava, V.P. Joshi, S.K. Apte, An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance, Mol. Microbiol. 59 (2006) 1308–1316.
[29] S. Kota, H.S. Misra, PprA: A protein implicated in radioresistance of *Deinococcus radiodurans* stimulates catalase activity in *Escherichia coli*, Appl. Microbiol. Biotechnol. 72 (2006) 790–796.
[30] H. Chen, G. Xu, Y. Zhao, B. Tian, H. Lu, X. Yu, et al., A novel OxyR sensor and regulator of hydrogen peroxide stress with one cysteine residue in *Deinococcus radiodurans*, PLoS One 3 (2008) e1602.
[31] A. Biegert, C. Mayer, M. Remmert, J. Soding, A.N. Lupas, The MPI Bioinformatics Toolkit for protein sequence analysis, Nucleic Acids Res. 34 (2006) W335–W339.
[32] J.R. Battista, M.J. Park, A.E. McLemore, Inactivation of two homologues of proteins presumed to be involved in the desiccation tolerance of plants sensitizes *Deinococcus radiodurans* R1 to desiccation, Cryobiology 43 (2001) 133–139.
[33] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, third ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.
[34] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels, Anal. Chem. 68 (1996) 850–858.
[35] J.C. Connelly, E.S. de Leau, D.R. Leach, DNA cleavage and degradation by the SbcCD protein complex from *Escherichia coli*, Nucleic Acids Res. 27 (1999) 1039–1046.
[36] N. Mathy, A. Hebert, P. Mervelet, L. Benard, A. Dorleans, Ines Li de la Sierra-Gallay, P. Noirot, H. Putzer, C. Condon, *Bacillus subtilis* ribonucleases J1 and J2 form a complex with altered behaviour, Mol. Microbiol. 75 (2010) 489–498.
[37] I. Callebaut, D. Moshous, J.P. Mornon, J.P. de Villartay, Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family, Nucleic Acids Res. 30 (2002) 3592–3601.
[38] J.C. Connelly, E.S. de Leau, D.R. Leach, Nucleolytic processing of a protein-bound DNA end by the *E. coli* SbcCD (MR) complex, DNA Repair (Amst) 2 (2003) 795–807.

[39] D. D'Amours, S.P. Jackson, The Mre11 complex: at the crossroads of DNA repair and checkpoint signaling, Nat. Rev. Mol. Cell Biol. 3 (2002) 317–327.
[40] B. Clouet-d'Orval, D. Rinaldi, Y. Quentin, A.J. Carpousis, Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity, J. Biol. Chem. 285 (2010) 17574–17583.
[41] H. Ishikawa, N. Nakagawa, S. Kuramitsu, R. Masui, Crystal structure of TTHA0252 from Thermus thermophilus HB8, a RNA degradation protein of the metallo-beta-lactamase superfamily, J. Biochem. 140 (2006) 535–542.
[42] D. Ghosal, M.V. Omelchenko, E.K. Gaidamakova, V.Y. Matrosova, A. Vasilenko, A. Venkateswaran, et al., How radiation kills cells: survival of *Deinococcus radiodurans* and Shewanella oneidensis under oxidative stress, FEMS Microbiol. Rev. 29 (2005) 361–375.
[43] R. Shashidhar, S.A. Kumar, H.S. Misra, J.R. Bandekar, Evaluation of the role of enzymatic and nonenzymatic antioxidant systems in the radiation resistance of Deinococcus, Can. J. Microbiol. 56 (2010) 195–201.
[44] N.P. Khairnar, H.S. Misra, DNA polymerase X from *Deinococcus radiodurans* implicated in bacterial tolerance to DNA damage is characterized as a short patch base excision repair polymerase, Microbiology (SGM) 155 (2009) 3005–3014.
[45] Z. Cao, C.W. Mueller, D.A. Julin, Analysis of the recJ gene and protein from *Deinococcus radiodurans*, DNA Repair (Amst) 9 (2010) 66–75.
[46] Y. Ma, H. Lu, B. Tippin, M.F. Goodman, N. Shimazaki, O. Koiwai, C.L. Hsieh, K. Schwarz, M.R. Lieber, A biochemically defined system for mammalian nonhomologous DNA end joining, Mol. Cell 16 (2004) 701–713.
[47] G. Gao, B. Tian, L. Liu, D. Sheng, B. Shen, Y. Hua, Expression of *Deinococcus radiodurans* PprI enhances the radioresistance of *Escherichia coli*, DNA Repair (Amst) 2 (2003) 1419–1427.
[48] J.-P. de Villartay, N. Shimazaki, J.-B. Carbonnier, A. Fischer, J.-P. Mormon, M.R. Leiber, I. Callebaut, A histidine in the β-CASP domain of artemis is critical for its full *in vitro* and *in vivo* functions, DNA Repair (Amst) 8 (2009) 202–208.
[49] I.L. de la Sierra-Gallay, L. Zig, A. Jamalli, H. Putzer, Structural insights into the dual activity of RNase J, Nat. Struct. Mol. Biol. 15 (2008) 206–212.

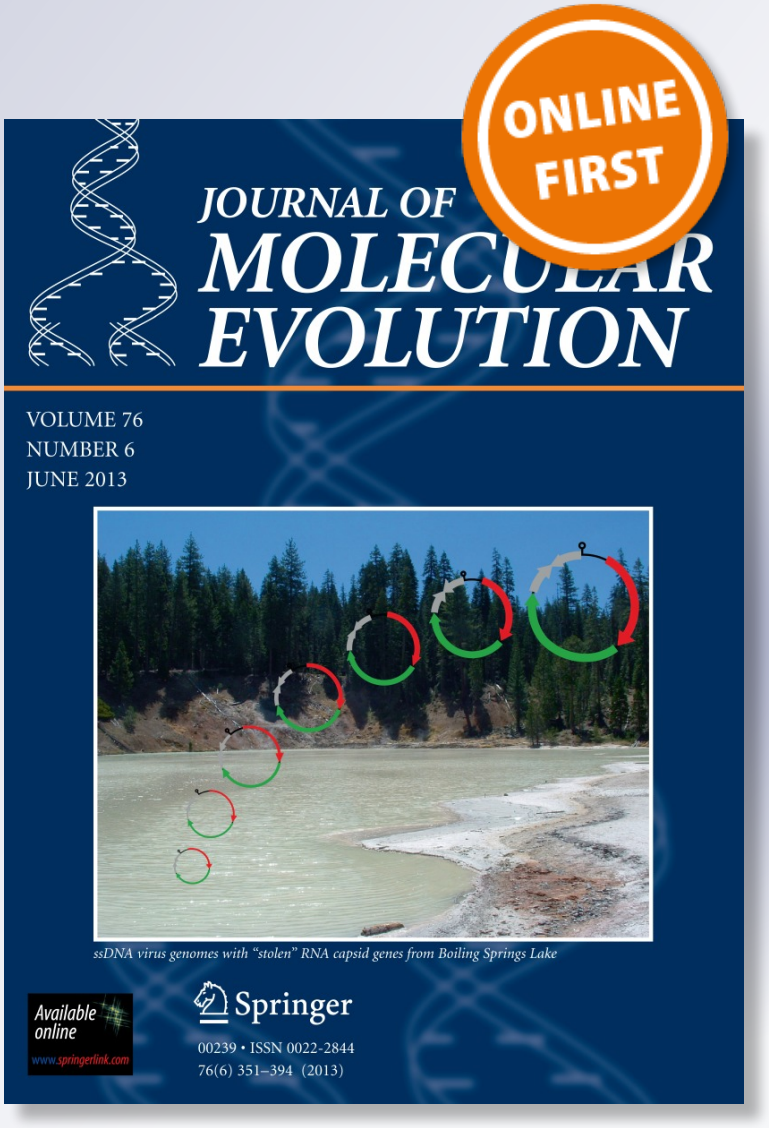
Hypothetical Proteins Present During Recovery Phase of Radiation Resistant Bacterium Deinococcus radiodurans *are Under Purifying Selection*

Anubrata D. Das & Hari S. Misra

Journal of Molecular Evolution

ISSN 0022-2844

J Mol Evol
DOI 10.1007/s00239-013-9577-9



Springer

J Mol Evol
DOI 10.1007/s00239-013-9577-9

ORIGINAL ARTICLE

Hypothetical Proteins Present During Recovery Phase of Radiation Resistant Bacterium *Deinococcus radiodurans* are Under Purifying Selection

Anubrata D. Das · Hari S. Misra

Received: 18 April 2013 / Accepted: 26 July 2013

Abstract *Deinococcus radiodurans* has an unusual capacity to recover from intense doses of ionizing radiation. The DNA repair proteins of this organism play an important role in repairing the heavily damaged DNA by employing a novel mechanism of DNA double-strand break repair. An earlier report stated that genes of many of these repair proteins are under positive selection implying that these genes have a tendency to mutate, which in turn provides selective advantage to this bacterium. Several "hypothetical proteins" are also present during the recovery phase and some of them have also been shown for their roles in radiation resistance. Therefore, we tested the selection pressure on the genes encoding these poorly characterized proteins. Our results show that a number of "hypothetical proteins" present during the repair phase have structural adaptations compared to their orthologs and the genes encoding them as well as those for the DNA repair proteins present during this phase are under purifying selection. Evidence of purifying selection in these hypothetical proteins suggests that certain novel characteristics among these proteins are conserved and seem to be under functional constraints to perform important functions during recovery process after gamma radiation damage.

Keywords *Deinococcus* · Hypothetical proteins · Selection pressure

Introduction

Bacteria belonging to the *Deinococcaceae* family have been isolated from the diverse environments. *Deinococcus radiodurans* a member of this family is characterized for its unusually higher resistance to several DNA damaging agents including radiations and desiccation. While the γ radiation D_{10} for human is 0.005 kGy, it is 0.25 kGy for *Escherichia coli* and as high as 10 kGy for *D. radiodurans*. Ionizing radiation kills by primarily causing DNA double-strand breaks (DSBs) in the DNA and the resistance to ionizing radiation is seen in but a handful of organisms notably among the Deinococci, the cyanobacteria like *Chroococcidiopsis* spp. and various fungi like *Filobasidium* (Slade and Radman 2011). Genome sequencing showed that *D. radiodurans* has a normal complement of repair proteins (White et al. 1999) and hence the question naturally arose as to how a repair complement common with other sensitive bacteria could repair the large number of DSBs created by exposing the cells to radiation. Molecular genetic experiments later showed that DSBs were found to be repaired initially by a novel pathway, the extended synthesis-dependent strand annealing (ESDSA) which involves non-reciprocal crossovers mediated primarily by RecJ, Rec A, DNA pol I, and DNA pol III followed by Rec FOR pathway of homologous recombination (Misra et al. 2006; Slade et al. 2009). ESDSA is similar to the synthesis-dependent strand annealing (SDSA) pathway seen in yeast, which itself can tolerate up to 0.8 kGy of γ radiation (Bennett et al. 2001). Again, *D. radiodurans* RecA shares 61 % identity with *E. coli* RecA but in contrast to the

Electronic supplementary material The online version of this article (doi:10.1007/s00239-013-9577-9) contains supplementary material, which is available to authorized users.

A. D. Das · H. S. Misra (✉)
Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400085, India
e-mail: hsmisra@barc.gov.in

A. D. Das
e-mail: anubrata@barc.gov.in

Published online: 10 August 2013

E. coli RecA, which prefers to bind to single-strand DNA, Deinococcal RecA preferentially binds to double-strand DNA (Kim and Cox 2002). Thus, possibly *D. radiodurans* has evolved to overcome damage due to radiation by improving on the SDSA pathway and by evolving minor but functional changes in known repair proteins. In this context, Sghaier et al. (2008) reported that the basal DNA repair machinery including DNA polymerases and DNA glycosylases are under positive selection. The presence of positive selection implies that there is a tendency to gain mutations in a protein which confers newer functions and subsequently allows the organism to adapt better to a environment, whereas the presence of negative selection or purifying selection means that the proteins are under functional constraints and there is a tendency to conserve the amino acids which are involved in that particular function or involved in the maintenance of structure (Nielsen et al. 2005).

Recovery from radiation also involves many other aspects like membrane regeneration, protein recycling, a regulated nucleolytic activity, signaling mechanisms etc., which are poorly understood. Transcriptomics (Liu et al. 2003), proteomics (Tanaka et al. 1996), and biochemical studies have shown that many hypothetical proteins are present during the recovery from radiation damage. Almost 40 % of the annotated ORFs in *Deinococcus* codes for "hypothetical proteins" or "putative proteins," which by definition do not have any functionally, characterized homologs (Siew and Fischer 2003). Since studies done earlier in other species have shown that the genes which code for hypothetical proteins (hereafter referred as ORFans) are the major component of positively selected genes in those species, implying that these are important for the fitness of these species under the respective living conditions (Soyer et al. 2009; Tai et al. 2011). We thus explored the subset of ORFans existing in the recovery phase in *D. radiodurans* for the evidence of positive selection as a simple test for their importance to the fitness and survival of *D. radiodurans.* Our results show that ORFans encoding for hypothetical proteins present during the radiation recovery phase are rather under purifying selection and this tendency of conservation indicates their essentiality in the recovery process.

Materials and Methods

Identification of Hypothetical Proteins and Orthologs

We shortlisted all hypothetical proteins obtained from the proteomics papers on *Deinococcus,* which have reported for recovery from radiation except for (Lipton et al. 2002), which did not. Transcriptome data were obtained from (Liu et al. 2003; Tanaka et al. 1996) and all genes with induction greater than twofold was considered significant. We analyzed the hypothetical proteins in the pSORTb server to find the sub cellular location of these proteins. Usually orthologs are obtained from closely related genomes in toto but due to lack of clear phylogenetic relationship between radiation resistant species, we decided to obtain orthologs from the curated eggNOG database (Powell et al. 2012). Since the accuracy of prediction improves by increasing the number of species involved in the study, in most cases 10 orthologs were taken for each "ORFan" under study. We carried out the similar exercise for the DNA repair proteins in this study.

Determination of Positive Selection and Homology Model of Structures

After obtaining the pre-aligned orthologs for each query, we obtained the corresponding set of coding sequences (CDS) of this set of proteins from EMBLCDS database and codon aligned them in PAL2NAL (Suyama et al. 2006). In a few cases especially for those proteins where adequate homologs were not available in the database, a reciprocal BLAST search with an E-value cutoff of 0.0001 was used to find additional homologs. Then CLUSTALX was used to align these additional homologs to the pre-aligned dataset obtained from the eggNOG database. Subsequently, for each query protein and its corresponding orthologs we built a maximum likelihood (ML) tree by DNAML in the PHYLIP package, with gamma distributed rates and a randomized input order. We obtained the 3-D homology models of the proteins from Phyre2 fold prediction server (Soding 2005) and I-Tasser server (Roy et al. 2010). These models were visualized in PyMOL. For positive selection test, we used the codeml program in PAML4.6 package (Yang 2007). The F3x4 codon substitution model was used to calculate likelihoods. The likelihood ratio test (LRT) was calculated from the likelihoods obtained from M1a and M2a models. Subsequently, the positively selected sites were seen by Bayesian empirical Bayesian (BEB) analysis using PAML.

Results and Discussion

Hypothetical Proteins were Upregulated During Recovery Phase of γ Irradiation

Deinococcus radiodurans cells exposed to γ radiation showed upregulation of ~832 genes at various time intervals of the recovery phase (Liu et al. 2003; Tanaka et al. 2004). Out of these, around 375–500 are ORFans (Fig. 1a). Through independent studies, the functional

significance of a number of hypothetical proteins in radiation resistance has been demonstrated. For example, Tanaka et al. (2004) generated deletion mutants of some of the hypothetical proteins, which were upregulated, and showed that several of these mutants were unviable. Thus, the roles of hypothetical proteins like *ddrA* (DR0423), *ddrB* (DR0070), *pprA* (DRA0346), DRB0100 etc. in the radiation resistance were discovered (Narumi et al. 2004; Harris et al. 2004; Kota et al. 2010). On comparing the gene expression levels at the two different doses i.e., 15 and 3 kGy, it was noticed that at 15 kGy dose, there are 473 ORFans showing greater than two-fold-induced expression and this number is well above the 26 ORFans induced at 3 kGy γ dose (Fig. 1b). This showed that the extent of cellular damage has a role in expression of ORFans. Interestingly, some of the highly induced ORFans at 3 kGy like DR0326, DR0491, DR0533, DR1439, DR1440 were not reported at 15 kGy, while reverse is true for some of the other ORFans like DR1358, DR1141, DR0697, DR1359. This qualitative shift hints to a possible dose-dependent or DNA damage-dependent gene regulation. We surveyed the data from independent proteomics studies on *D. radiodurans* and it showed that around 60 different hypothetical proteins were present during the post irradiation recovery (PIR) phase (Table S1) (Das and Misra 2011; Lu et al. 2009; Zhang et al. 2005; Kota and Misra 2008; Basu and Apte 2012). This number included both proteins induced in response to γ radiation as well as constitutively expressed proteins and what is interesting for this small subset is that most of these proteins were reported from first 2 h of recovery period when the novel ESDSA phase is highly active. An upregulation in the transcripts did not necessarily lead to the detectable levels of corresponding proteins. For example, the upregulation of DR0422 (20 fold), DR1141 (tenfold), DR2574 (sixfold), and DR1440 (fourfold) ORFans of *D. radiodurans* were reported in different transcriptomic studies, but the proteins corresponding of these ORFs are not yet reported. Another very interesting feature observed was that of the 23 paralogous genes belonging to 9 gene families, which have expanded in *D. radiodurans,* only DR2179 was found to be present during this phase (Omelchenko et al. 2005). The predicted functions of these proteins indicated that these constitute the group of proteins involved in lipid transport, exhibiting protease activity and DNA binding proteins including transcriptional activators. Almost one-third of all the hypothetical proteins detected were membrane bound or extracellular proteins (Fig. S1). These proteomic studies have also reported several small molecular weight proteins with an average molecular weight of 18 kDa having very few orthologs, including some like DRA0281 and DR1977, which were present only in *D. radiodurans*. Nevertheless, the upregulation of several ORFans and the presence of hypothetical proteins during the recovery phase indicated that these proteins also contribute to the recovery of *D. radiodurans* from γ radiation effects.

Many Hypothetical Proteins Have Evolved Altered Features on Common Structural Templates

Many of the proteins listed in Table S1 were previously annotated as "hypothetical," but subsequently have been assigned functions based on their sequence and structural features. However, in several cases we saw structural alterations different from what has been annotated, which could lead to the altered capabilities for substrate and/or protein interaction. Since these proteins were found in the recovery phase, we tried to understand their functions in the context of recovery from radiation damage and we have highlighted here a few such examples with greater details.

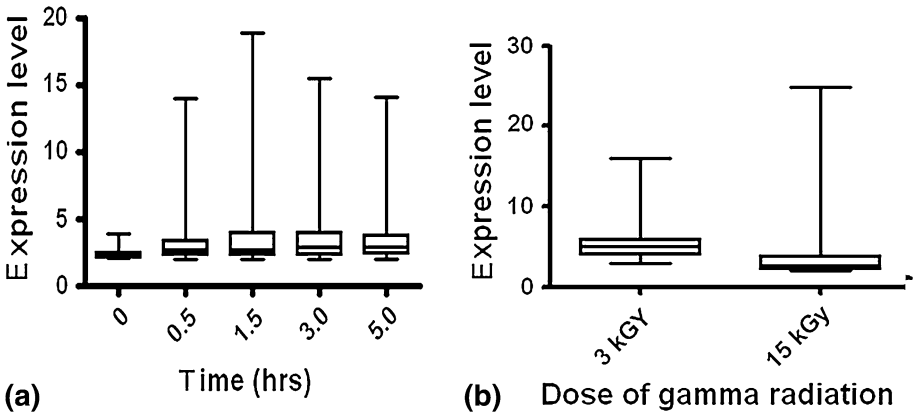


Fig. 1 Gene expression profile of ORFans encoding for hypothetical proteins in response to γ-radiation. **a** *Box plot* showing the pattern of expression of ORFans at various time intervals during the PIR phase as reported in (Tanaka et al. 1996). **b** Altered expression profile of ORFans involved in radiation recovery at different doses at 3 and 15 kGy shows that expression of select genes sharply increases at higher doses as reported in Liu et al. (2003)

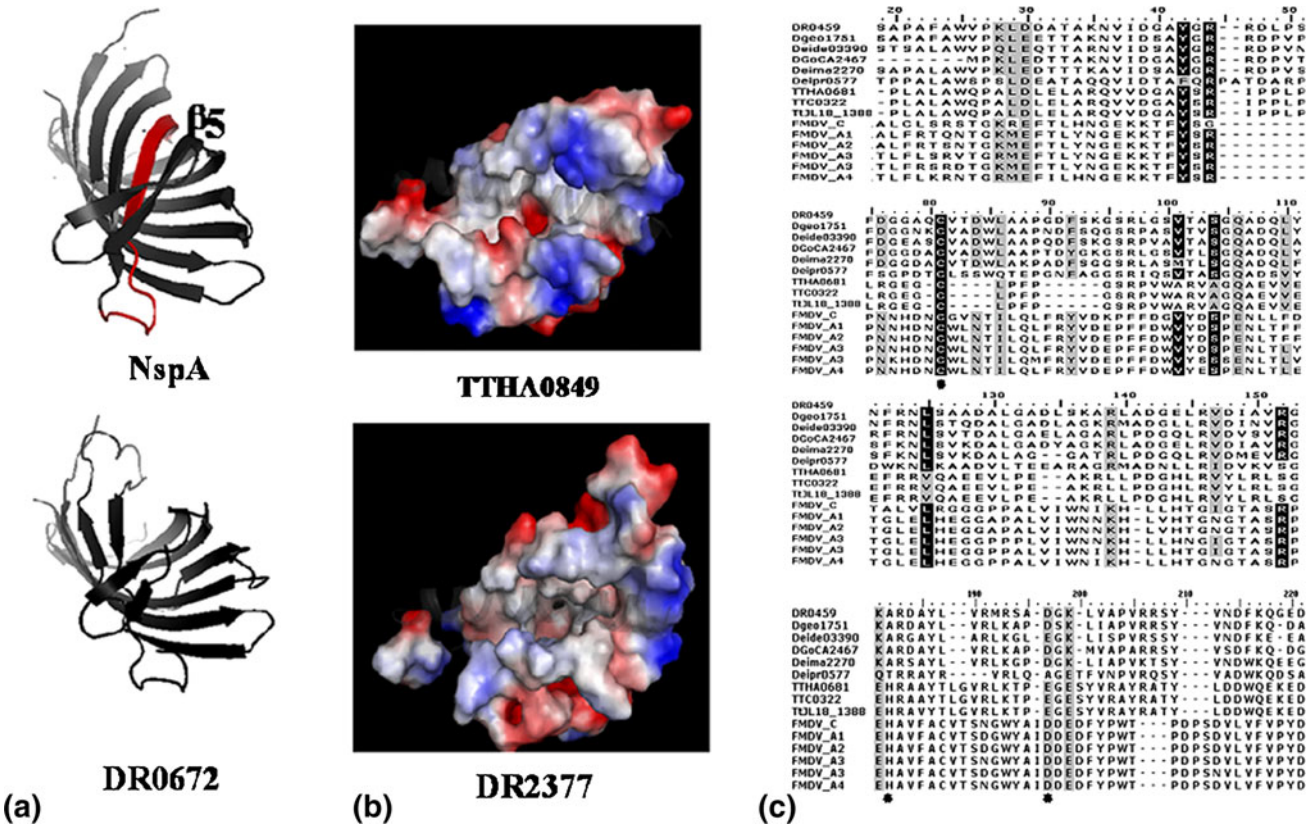


Fig. 2 The 3D model of DR0672 and DR2377 and the multiple sequence alignment of DR0459 with its homologs. **a** A cartoon representation of the 3D homology model of DR0672 (DR0672), which was obtained from the structure of NspA (1P4T) from *N. meningitides* with an RMSD of 5.0. The beta sheet in NspA, which is absent in DR0672 is highlighted in *red color*. **b** A space filling model of DR2377 shows the deeper substrate binding pocket of DR2377 compared to its homolog TTHA0849 from the closely related *T. thermophilus* spp. **c** Part of a multiple sequence alignment of DR0459 and its homologs from foot and mouth disease virus (FMDV) sequences. The conserved sites are marked with an *asterisk*. DR0459, Dgeo1751, Deide03390, DGoCA2467, Deima2270 are *Deinococcus* orthologs, while TTHA0681 and TTC0322 are *Thermus* orthologs (Color figure online)

DR0672 is a typical case where commonly available structural folds have been modified. It has remote sequence homology even within *Deinococci* but showed structural homology with Neisserial surface protein A (NspA) (Fig. 2a). A beta sheet β-5, was absent in DR0672 and this implies that DR0672 possibly forms a much more compact barrel. Unlike the hydrophobic residues present in NspA, the domain of DR0672 oriented toward the outer side is lined with charged residues in the large loops, as seen in extracellular lipid binding domains of OmpA proteins (LaLonde et al. 1994). This protein is highly divergent from the canonical OmpA proteins and may be involved in binding lipids or in maintaining membrane integrity as seen in *Porphyromonas gingivalis* (Iwami et al. 2007).

DR2377 is a homolog of TTHA0849 from *Thermus* spp., a member of the steroidogenic acute regulatory related lipid-transfer domain (START) superfamily (Iyer et al. 2001). Members of this family are involved in transport of various lipids. It has been reported that the cavity of active site of TTHA0849 is small so that it can only accommodate lipids smaller than cholesterol. A 3D model of DR2377 based on the template of TTHA0849 shows that its cavity is much larger and thus the nature of lipids transported should be different in *Deinococcus* and *Thermus* spp. (Fig. 2b). This protein is induced twofold higher during recovery from radiation and may be involved in the metabolism of lipids associated with radiation resistance.

DR0459 is a membrane bound protein with an N-terminal signal peptide. In DR0459, the structure could not be completely modeled but N-terminal of the protein shows homology to foot and mouth disease virus (FMDV) leader protease. The multiple sequence alignment (Fig. 2c) shows an interesting aspect. Homologs of DR0459 are present in both *Deinococci* and *Thermus* and in both of them have the conserved cysteine (Cys51) and aspartate (Asp163) residue of FMDV leader protease. However, the conserved histidine (His148), which completes the catalytic network, is replaced by alanine in *Deinococcus* spp. A similar mutation was observed in case of plant storage protein narbonin, which is an inactive form of chitinase (Hennig et al. 1992).

The C-terminal has an adhesin domain, thus earmarking DR0459 to the cell wall. DR0459 is induced threefold higher after a 1 kGy dose of γ radiation thus highlighting its role in the recovery of radiation damaged cells (Lu et al. 2009). It would be worth examining if the presence of Ala148 in place of mostly conserved His148 makes this protein inactive in *D. radiodurans*.

DR2623 was found induced when *D. radiodurans* cells were irradiated at 1 kGy dose and then allowed to undergo repair for 1 h (Zhang et al. 2005). DR2623 is structurally homologous to thioredoxin reductase (Fig. 3a, b). The thioredoxin system is critical in *Deinococcus* as this system works with thioredoxin-dependent thiol peroxidases, which scavenge the harmful reactive oxygen species (ROS) generated during γ radiation. The active sites of this class of proteins have conserved thiol rich active site motif "CXXC" and during catalysis the electrons flow from nicotinamide adenine dinucleotide (NADH) to the active site disulfide via flavin adenine dinucleotide (FAD) and then to thioredoxin (Fig. 3c) (Yamamura et al. 2009). In DR2623 the conserved cysteine residues are replaced by isoleucine and threonine and no contacts with FAD could be detected at least in the generated 3-D model, ruling it out as a reductase (Fig. 3d). Also missing is the C-terminal dimerization domain and thus probably DR2623 functions as a monomer. But homologs of DR2623 are widely distributed among the bacteria as seen in the 16S rDNA phylogenetic tree (Fig. 3e). These homologs do not have conserved residues seen in their characterized counterparts like glutathione reductase and lipoamide dehydrogenase and hence it would be interesting to know the exact function of this protein other than binding with FAD and NADH.

Among the other interesting proteins, the 3D structure of 16 kDa protein DR2179, which is induced during this phase, showed domain structure similar to 4VR domain initially identified as a novel small molecule binding domain (SMBD) in proteins (Fig. 4a, b) (Anantharaman

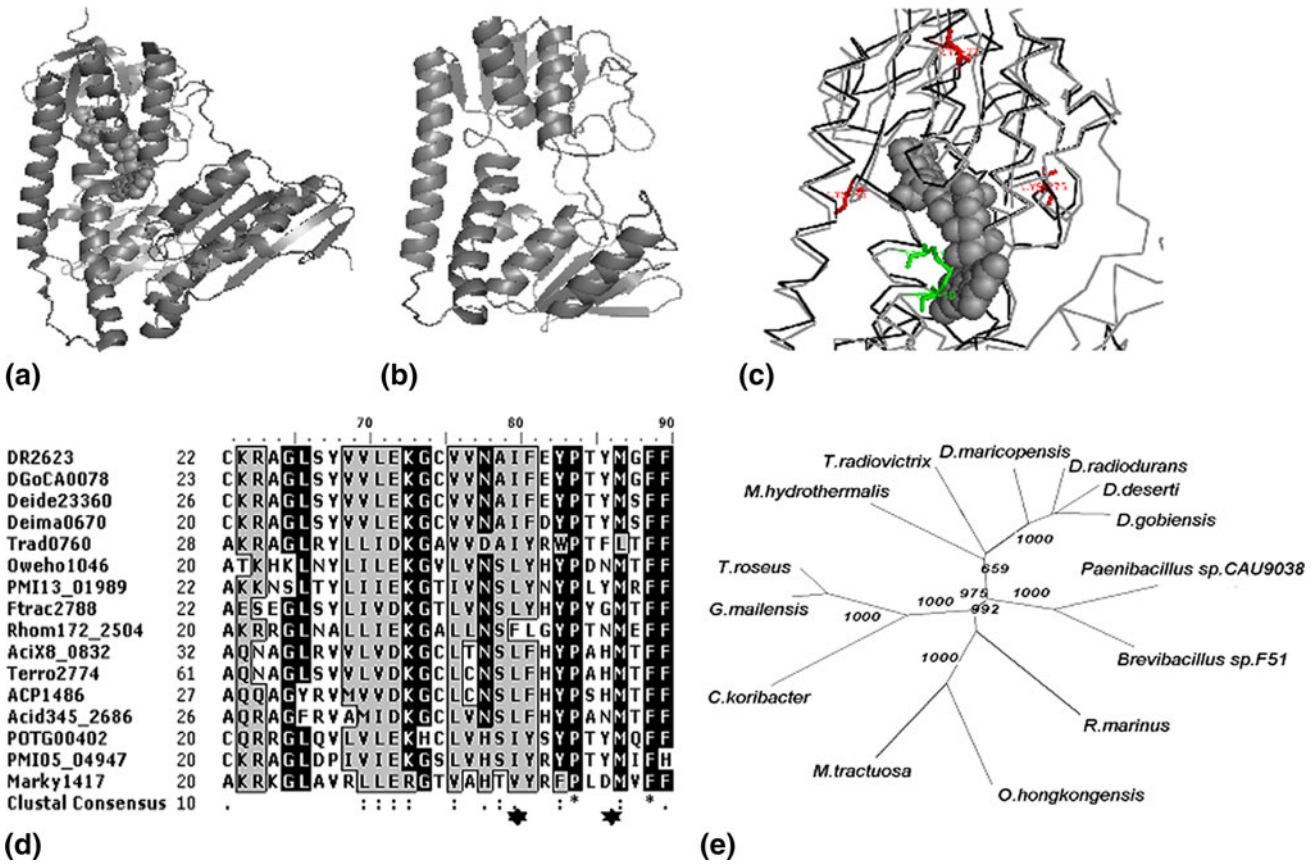


Fig. 3 Bioinformatic analysis of DR2623 with closely related proteins. **a** A cartoon representation of Glutathione reductase from *B. henselae* (PDB id 3T30). **b** A cartoon representation of the 3-D model of DR2623 modeled on GSR of *B. henselae*. **c** A close up view of DR2623 superimposed on 3T3O showing active site superimposition. The flavin adenine dinucleotide (FAD) molecule, which interacts with DR2623 is shown as a collection of *gray spheres* and the distant cysteine residues of DR2623 are marked in are positively selected are marked in *red* , while the catalytic cysteine residues of 3T3O are marked in *green*. **d** A multiple sequence alignment of homologs of DR2623. The site where cysteine is replaced is marked by *asterisk*. **e** A 16S rDNA phylogenetic tree of the homologs of DR2623. The bootstrap values are indicated at the nodes (Color figure online)

et al. 2001). A distinguishing feature of this domain is the presence of two conserved cysteine residues (Fig. 4c). A multiple sequence alignment (Fig. 4d) shows that DR2179 and its homologs have three conserved cysteine residues. However, the relative position of these residues is different with respect to the canonical standalone 4VR domain as identified in 23 kDa 4-vinyl reductase. DR2179 has been annotated as a heme NO binding protein and is structurally homologous to the H-NOX protein SO2144 from *Shewanella oneidensis.* SO2144 has been shown to sense the intracellular NO and regulate a cognate histidine kinase SO2145 by inhibiting its autophosphorylation (Price et al. 2007). DR2179 has the conserved S109 and R111 which interact with the heme moiety. But it lacks the conserved Y-X-S-X-R motif reported in H-NOX proteins. Thus, it is difficult to envisage its role as a typical H-NOX protein. The comparison of 3D model of DR2179 with the structure of a known pyrroloquinoline quinine (PQQ) binding protein like PqqB from *Klebsiella pneumoniae* did rule out the possibility of PQQ an antioxidant and essential cofactor for quinoprotein, interaction with this protein (Puehringer et al. 2008).

DR0390 is a 2-domain protein with its N-terminal matching the C-terminal nucleotide binding domain of dihydroxyacetone (Dha) kinase of *Citrobacter freundii* and the C-terminal matching with DegV family lipid binding proteins (Fig. 5a). The homolog of this protein is found mainly in Gram-positive bacteria like *Bacillus* spp. and *Clostridium* spp., while it is absent in Gram-negative (Fig. 5b). In Bacilli and Clostridia, it is adjacent to RecG but we could not ascertain any functional significance for this linkage. DR0390 in *D. radiodurans* and a homolog in *Bacillus halodurans* is constitutively expressed (Wallace et al. 2012). The Dha kinases utilize either ATP or PEP to phosphorylate dihydroxyacetone and other small aldoses or ketoses. The N-terminal of these proteins forms barrel shape structure comprised of 8 α-helices. In DR0390 the catalytic site is formed of the amino acids like D57, D59, T60, T150, S102, which are characterized as essential for catalytic function in Dha kinase of *C. freundii* (Fig. 5a) (Siebold et al. 2003). Another conserved residue D380 of Dha kinase is replaced by a functional analog N51 in this protein. DR0390 has a conserved threonine in the first loop instead of a histidine and this is usually a hallmark of two

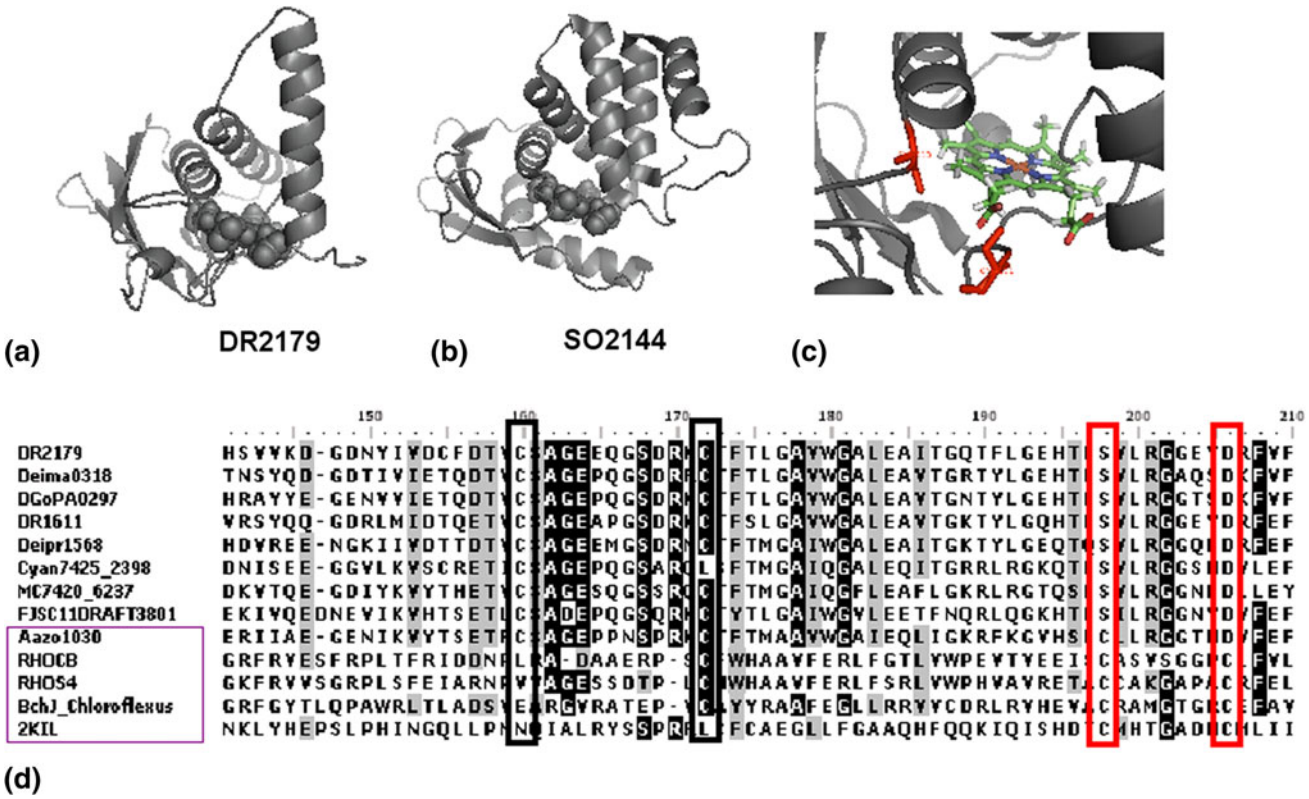


Fig. 4 Bioinformatic analysis of DR2179. The homology model of DR2179 (**a**) based on the crystal-structure cartoon of SO1244 (PDB id 2KIL) (**b**). The porphyrin ring of the heme molecule is shown as a collection of *gray spheres* for clarity. **c** A close up view of the putative active site of DR2179 showing the conserved cysteine residues (C101 and C113) forming bonds with the porphyrin ring. **d** A multiple sequence alignment of DR2179, its orthologs and the orthologs with SO2144, which are clustered within a *purple lined box*. The conserved cysteine residues of DR2179 are marked by *black outlines* while the conserved cysteine residues of the H-NOX homologs are marked by *red outlines* (Color figure online)

domain Dha kinases. Unusual in DR0390 is that, unlike a well formed capping loop as seen in PEP-dependent kinases (Siebold et al. 2003) or an unstructured region as in ATP-dependent kinases, it extends the helix H1 to encircle the active site. Structural studies show that DegV domain of both independent proteins and proteins with multi domain are composed of the conserved serine, threonine, arginine, and histidine residue, which interact with phospholipids/lipids substrates. In the DegV domain of DR0390, the conserved threonine is replaced by a lysine (K40) and the histidine is absent. The pattern of conservation on the 3D homology model shows that the C-terminal domain of DR0390 has different distribution of charged residues (Fig. 5a) compared to DegV proteins and that the substrate binding pocket is fairly large compared to the more compact Deg V family proteins. Thus, the nature of substrates could bind to DR0390 appears to be quite different as compared to DegV family proteins. *D. radiodurans* has at least three orthologs of DegV family proteins and none of them match to each other at their C-terminal regions.

DR2577 (SlpA) and its homolog DR1124 are annotated as S-layer proteins in *D. radiodurans* and have been shown to maintain the cell envelope structure (Rothfuss et al. 2006). But both these proteins have a C-terminal porin domain and N-terminal phenylalanine characteristic of outer-membrane proteins like OmpM1 from *Mitsuokella multacida* (Kalmokoff et al. 2009). Similarly, DRA0009 is an inactive homolog of sensor histidine kinase from *Thermotoga maritima* (Marina et al. 2005). All the catalytic residues of this protein are conserved with histidine kinase except asparagine the site for phosphorylation that is replaced by alanine DRA0009. The subtle and pervasive alterations seen in these proteins raised a pertinent question whether these observed changes were random or were selected during evolution.

Hypothetical Proteins Present During the Recovery Phase are Under Purifying Selection

Since many of these hypothetical proteins showed novel modifications of their structure, we wanted to see whether the corresponding genes are under positive selection and contribute to the ability of *D. radiodurans* to adapt to harsh environments like doses of high γ radiation. Selection pressure is determined by the ratio of non-synonymous substitutions to synonymous substitutions (dN/dS) for a given site in a gene (for convenience of readers dN/dS is

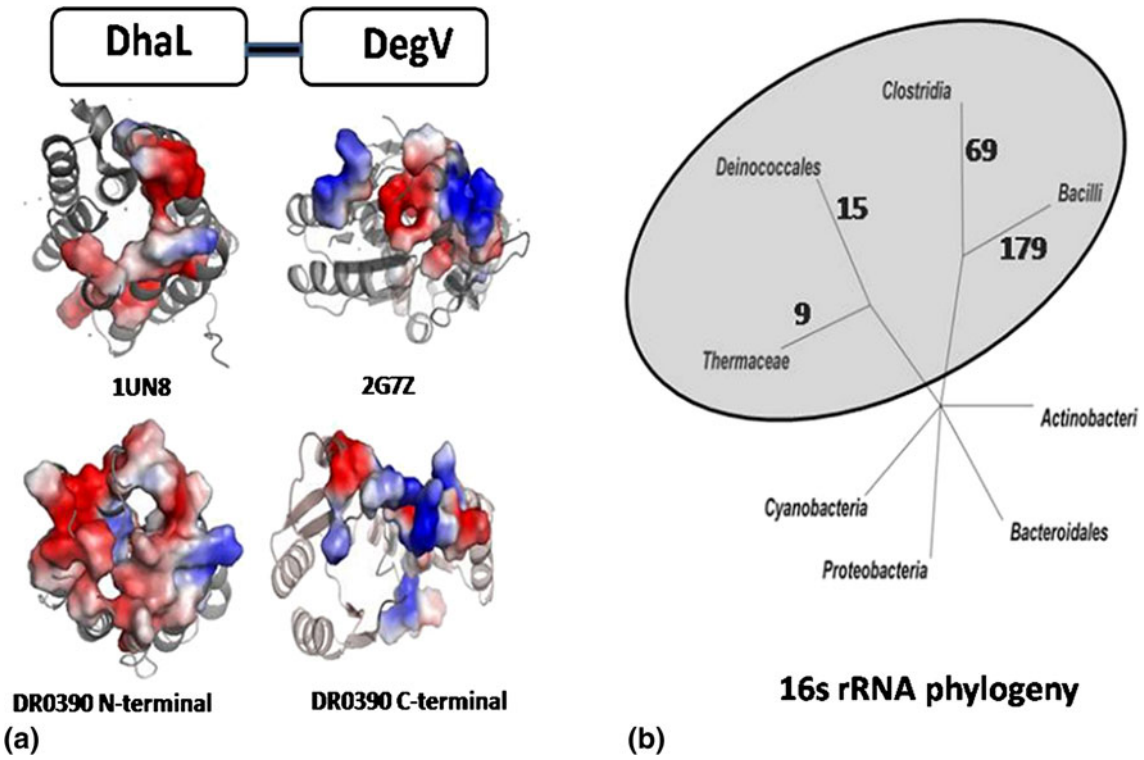


Fig. 5 Structure and functional domain distribution in DR0390. **a** DR0390 a 2-domain protein is comprised of N-terminal kinase domain as modeled on *C. freundii* Dha kinase (PDB id 1UN8) while C-terminal domain is modeled on DegV of *S. pyogenes*. The electrostatic surface potential of conserved residues of both the templates and DR0390 are mapped onto the respective structures to compare the similarities in the N-terminal and differences in the C-terminal. **b** The phylogenetic distribution of the homologs of DR0390 in bacteria. The *numbers* in the branches are the number of hits to DR0390 in BLAST search

represented as ω). If ω is <1, it means that mutations are deleterious and are preferentially removed and it was the case of purifying selection. On the other hand if ω is >1 then mutations are advantageous and these mutations are retained and this is a case of positive selection (Bielawski and Yang 2004). We calculated the dN/dS ratio for the hypothetical proteins and their homologs based on (a) a single ratio model (M0) for overall selection pressure and (b) a nested set of models (M1a vs M2a) for site-specific selection. A LRT based on the nested pair of models compares the probability of sequences having positively selected sites. We used BEB analysis is a complementary statistical tool to the above LRT and probability of ω in a given data set was estimates. Results showed that the actual sites are under positive selection. When LRT and BEB are in agreement, it means that the selection pressure is robust. Several multilocus-sequence typing (MLST) studies as well as whole genome studies have shown that housekeeping genes in bacteria are under purifying selection because the corresponding proteins have to conserve their structures/active sites for participating in similar biochemical functions across the living systems (Lan and Reeves 2001; Dingle et al. 2001). Likewise it has been seen that genes encoding most hypothetical proteins have recently evolved and are species-specific and usually they are under positive selection (Ge et al. 2008).

Orthologs selection-based studies for the hypothetical proteins were difficult in this study because though *Deinococcus* and *Thermus* are in the same phylum, there have been a lineage-specific gene gain and loss in both these species (Omelchenko et al. 2005). For example, *Deinococcus* has homologs of proteins including DR1252 a saccharopine dehydrogenase with N-terminal Rossman fold, which are well conserved in other species but are absent in *Thermus* spp. Also in several cases, the orthologs in *Thermus* and *Deinococcus* differ in length, e.g., DRA0009 is half the size of a homologous sensor histidine kinase from *Thermus*, indicating that these proteins are under different evolutionary constraints. For these reasons we obtained the set of orthologs from a curated database like eggNOG. There is an exponential decrease in the number of hypothetical proteins having more than 100 orthologs as compared to the DNA repair genes, which had ~900 orthologs each, indicating that these have evolved recently (Fig. 6a).

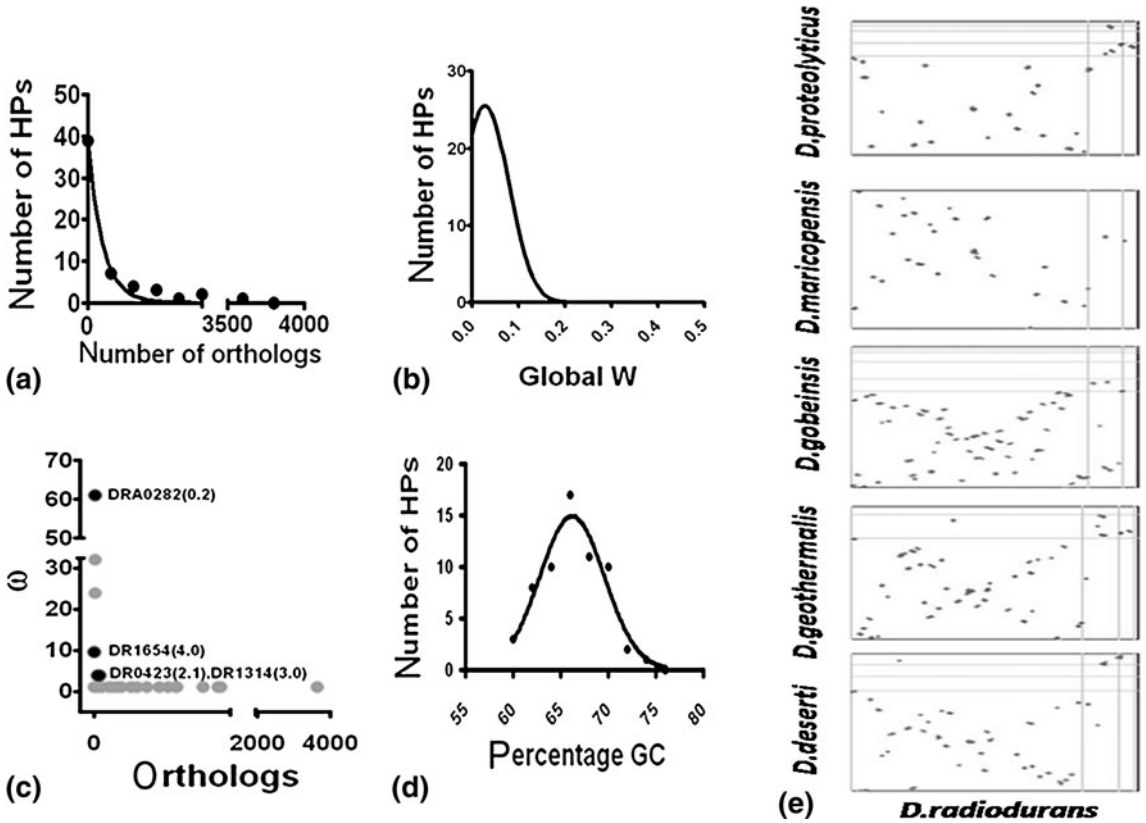


Fig. 6 Selection pressure and synteny of ORFans in *D. radiodurans*. **a** The relationship of hypothetical proteins and the number of orthologs. **b** The distribution of hypothetical proteins as a function of global ω values. **c** Relationship of site-specific '*w*' values and the number of orthologs. The proteins, which are positively selected are marked in *dark circles* and labeled with LRT values in brackets. **d** The distribution of %GC content in hypothetical proteins. **e** Lack of synteny in *Deinococcus* species

In some cases like DR0672 where suitable orthologs were absent we could not obtain data for selection pressure. When we applied the single ratio model (M0) for all the other cases, we found all the ORFans encoding for the hypothetical proteins were under purifying selection with $\omega < 1$ (Fig. 6b). Thus, the whole gene per se was under purifying selection but it was still possible that a few select sites would be under positive selection (Yang and dos Reis 2011). Next, even the LRT for site-specific selection showed no statistically significant evidence of site-specific positive selection in these genes (Table S1). In the case of DR1654, DR1314, DR0423 (DdrA), DR1940, and DRA0282, the LRT values were positive but below the levels of significance. In the given set of studied ORFans, the fractions of ORFans with positive LRT values were few (Fig. 6c), suggesting that these are newly evolved genes. In the case of extracellular proteins DR1654 and DR1940, analysis also showed that these proteins have positively selected sites with a confidence >90 %. DR1654 was found in heparin-binding fraction in cell lysate of *D. radiodurans* recovering from radiation injury (Das and Misra 2011). A low-resolution model obtained from I-Tasser (RMSD value of 10) showed it to be a homolog of the NC4 domain of collagen (Fig. 7a, b), which is known to bind with various macromolecules such as proteoglycans and heparin (Leppanen et al. 2007). Thus, homology prediction corroborated the observation that DR1654 is a heparin-binding protein (Das and Misra 2011). It is a poorly distributed protein and the homologs of these proteins are seen almost exclusively in radiation resistant bacteria (Fig. 7c). Like wise, the N-terminal of DR1940 is homologous to HslJ, a chaperone in *E. coli* and the C-terminal is homologous to Ecotin, a trypsin inhibitor (Fig. 7d, e). Also the homology model shows that positively selected residues Ser13, Leu15, Ala 185 are exposed on the same face of the chaperone domain suggesting that they may be involved in substrate binding. This protein is fairly well distributed among bacteria (Fig. 7e) and knowing its function can lead to insights on survival under stress. Pertinent to note is that BEB analysis showed positively selected sites only for the extracellular proteins, a class which is well-documented for positive selection, as they primarily interact with a changing environment (Nielsen et al. 2005; Petersen et al. 2007). Both DdrA and DRA0282 are intracellular and interact with

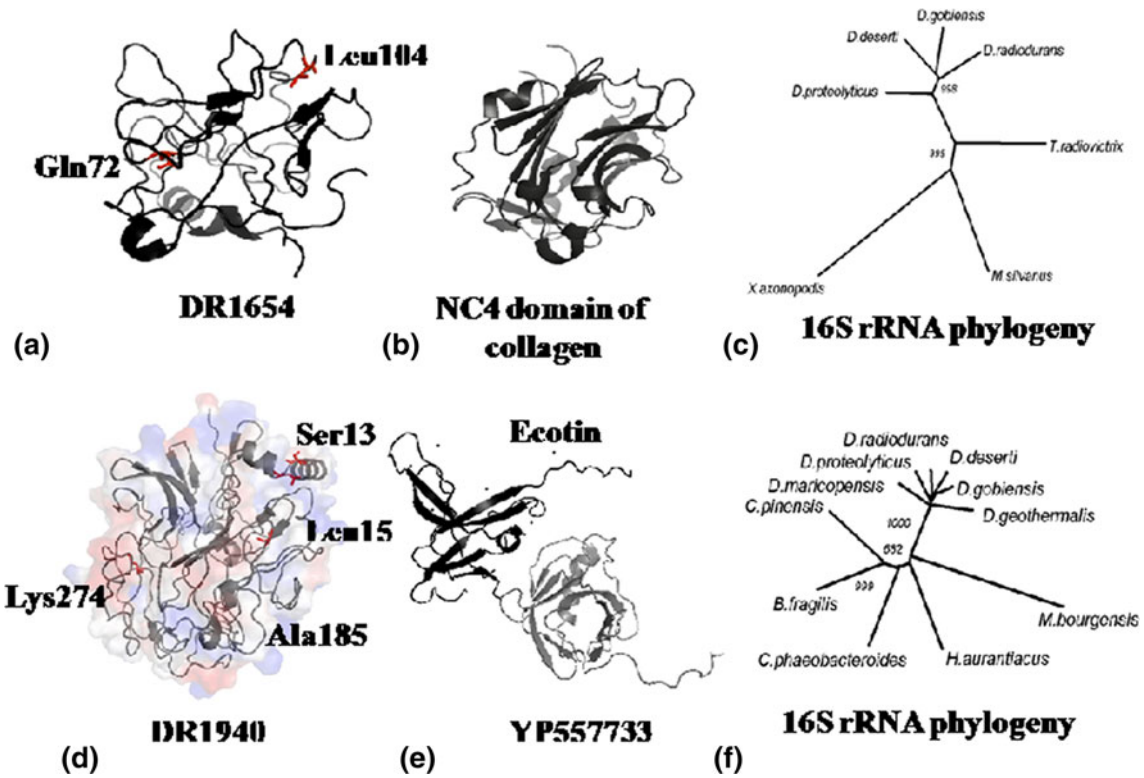


Fig. 7 Hypothetical proteins with positively selected sites in BEB analysis. **a** The 3D homology model of DR1654 showing the positively selected residues Gln72 and Leu104 in *red color*. **b** The structural homolog of DR1654, the NC4 domain of collagen. **c** The 16S rRNA phylogeny tree showing the sparse distribution of bacteria in which this protein is found. **d** The 3D homology model of DR1940 showing the positively selected residues Lys 274, Ser13, Leu15, and Ala185 in *red color*. **e** The structural homolog of DR1940, YP557733 is a hypothetical protein from *Burkholderia xenovorans,* which matches with the N-terminal of DR1940 while Ecotin, a trypsin inhibitor matches with the C-terminal of DR1940. **f** The 16S rRNA phylogeny tree showing the distribution of this protein in several bacteria (Color figure online)

DNA in vitro. These are upregulated during radiation recovery (Liu et al. 2003) and their deletion mutants show no effect when grown in rich medium (Harris et al. 2004; Das and Misra 2011). DR1314 is homologous to the PRC-H barrel domain of the photosynthetic reaction center of *Rhodopseudomonas viridis*. This domain is reported to be a key regulator in electron transfer between quinones in photosynthetic reaction center (Anantharaman and Aravind 2002). In *D. radiodurans,* the quinone like PQQ plays an important role to counter oxidative stress indicating some roles of DR1314 in electron transport processes during radiation recovery.

Since horizontal gene transfer has been a prevalent phenomenon in this genus, we tried to find out whether any of the ORFans studied here were recently transferred. *Deinococcus* genome is GC rich and a variation in the GC content of a gene could be an evidence for horizontal gene transfer. Our studies show that the hypothetical proteins under study have the same mean GC content as that of *Deinococci* and *prima facie* are not the case of any recent horizontal gene transfer (Fig. 6d). Another highlight is that, usually genes which have recently evolved like many of the ORFans are poorly expressed (Tautz and Domazet-Loso 2011) but in our case most of these ORFans are either constitutively expressed or are highly induced during the recovery process. Thus, we have realized that most of the hypothetical genes present in this phase have evolved novel features through purifying selection, which shows the tendency to conserve these features. Since purifying selection occurs due to functional constraints on a protein, it means that these proteins perform a key role in the recovery process.

Recombination/Repair Genes are also Under Purifying Selection

Since we found that most of the ORFans in the recovery phase are under purifying selection we decided to check the selection pressure on some of the recombination repair genes like *recFOR* pathway genes and a few other ubiquitous DNA repair proteins present in this phase (Table 1). A couple of earlier studies had shown that *recFOR* is important in the DSB repair of this bacterium (Misra et al. 2006; Bentchikou et al. 2010). Among the dozen repair genes studied by us with the rigorous model M1a versus M2a we found that they too were under purifying selection. This result remains true even when very divergent orthologs were selected and also has experimental support because when the DNA polymerase of *E. coli* is expressed in *Deinococcus,* it is able to participate in the DNA repair process and restore DSBs, which is possible only with a certain degree of conservation. Although, the sample size of housekeeping genes and recombination/repair genes analyzed under this study is not very large, this finding is in agreement with numerous reports, which show that housekeeping genes are under purifying selection (Petersen et al. 2007). Global genome studies have shown that the lack of gene order (synteny) is usually observed with organisms when they belong to distant phyla and that correlates with increasing dN/dS values in the corresponding orthologs (Novichkov et al. 2009). Thus the mechanism, which tend to conserve the gene order also conserve the sequence. But *Deinococci* lack synteny within themselves (Fig. 6e) and although protein sequences are conserved, the gene order is not and this implies that

Table 1 List of DNA repair genes checked for selection pressure in this study

Repair genes (locus name)	Gene name	LRT	Global ω	Site-specific Ω in M2a
DR0856	DNA pol III ε subunit	0.10	0.018	26.0
DR0906	DNA gyrase subunit B	0	0.031	1.0
DR1532	Transcription repair coupling factor	0	0.022	1.0
DR1707	DNA pol I	0	0.066	1.0
DR1913	DNA gyrase subunit A	0	0.006	1.0
DR1916	Rec G	0	0.026	1.0
DR2074	Methyladenine DNA glycosylase	0	0.004	1.0
DR0198	Rec R	0	0.030	1.0
DR1089	Rec F	0	0.117	1.0
DR1126	Rec J	0	0.035	1.0
DR1289	Rec Q	0	0.002	1.0
DR2340	Rec A	0	0.033	49.0

Gene sequences of the DNA recombination and repair genes were analyzed using LRT and the ratio of non-synonymous substitutions to synonymous substitutions (ω) were looked using nested set of model type M2a for site-specific selection (Ω)

operons would also not have been conserved. The DNA transposition activity reported in *D. radiodurans* may possibly be responsible for such a phenomenon (Mennecier et al. 2006). Thus in this study we have found that these hypothetical proteins in *Deinococcus* spp. are subjected to unique evolutionary process where the genes have been shuffled around during speciation but their phenotypes are conserved and they are robustly expressed.

Since positive selection is an indication of adaptive changes to the environmental conditions, we set out to find examples of positively selected hypothetical proteins present during the recovery phase of *D. radiodurans* from γ radiation. We have found a wide diversity in the phylogenetic distribution of these hypothetical proteins. For example there were conserved proteins like DR2623 with a wide phylogenetic distribution to proteins like DRA0281, which was present only in *D. radiodurans.* Homology models for a number of these hypothetical proteins showed distinct adaptations of the active site or binding surfaces or rearrangement of domains, which should lead to newer capabilities. All the ORFans encoding for these hypothetical proteins and the key DNA repair genes present in this phase were found to be under purifying selection. We suspect that these proteins have evolved adaptations for performing novel functions, which are necessary to overcome the effects of ionizing radiation, and hence these adaptations are conserved. A functional study of these proteins and their novel biochemical properties could throw a new light on the phenomenal radiation recovery of this organism. Moreover earlier studies have shown that a significantly low dN/dS value is an indication of a functional exon in eukaryotes (Nekrutenko et al. 2002) and we have seen evidence of purifying selection in many of these functional hypothetical proteins. Thus a possibility exists in future for finding of functional hypothetical proteins in *Deinococcus-Thermus* phylum by detecting for purifying selection.

Acknowledgments The authors thank Dr. S. K. Apte, for his constant support and encouragement while pursuing this work.

References

Anantharaman V, Aravind L (2002) The PRC-barrel: a widespread, conserved domain shared by photosynthetic reaction center subunits and proteins of RNA metabolism. Genome Biol 3(11):RESEARCH 0061

Anantharaman V, Koonin EV, Aravind L (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307(5):1271–1292

Basu B, Apte SK (2012) Gamma radiation-induced proteome of *Deinococcus radiodurans* primarily targets DNA repair and oxidative stress alleviation. Mol Cell Proteomics 11(1):M111 011734

Bennett CB, Lewis LK, Karthikeyan G, Lobachev KS, Jin YH, Sterling JF, Snipe JR, Resnick MA (2001) Genes required for ionizing radiation resistance in yeast. Nat Genet 29(4):426–434

Bentchikou E, Servant P, Coste G, Sommer S (2010) A major role of the RecFOR pathway in DNA double-strand-break repair through ESDSA in *Deinococcus radiodurans*. PLoS Genet 6(1):e1000774

Bielawski JP, Yang Z (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol 59(1):121–132

Das AD, Misra HS (2011) Characterization of DRA0282 from *Deinococcus radiodurans* for its role in bacterial resistance to DNA damage. Microbiology 157(Pt 8):2196–2205

Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, Bootsma HJ, Willems RJ, Urwin R, Maiden MC (2001) Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol 39(1):14–23

Ge G, Cowen L, Feng X, Widmer G (2008) Protein coding gene nucleotide substitution pattern in the apicomplexan protozoa *Cryptosporidium parvum* and *Cryptosporidium hominis*. Comp Funct Genomics. doi:10.1155/2008/879023

Harris DR, Tanaka M, Saveliev SV, Jolivet E, Earl AM, Cox MM, Battista JR (2004) Preserving genome integrity: the DdrA protein of *Deinococcus radiodurans* R1. PLoS Biol 2(10):e304

Hennig M, Schlesier B, Dauter Z, Pfeffer S, Betzel C, Hohne WE, Wilson KS (1992) A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8 A resolution. FEBS Lett 306(1):80–84

Iwami J, Murakami Y, Nagano K, Nakamura H, Yoshimura F (2007) Further evidence that major outer membrane proteins homologous to OmpA in *Porphyromonas gingivalis* stabilize bacterial cells. Oral Microbiol Immunol 22(5):356–360

Iyer LM, Koonin EV, Aravind L (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43(2):134–144

Kalmokoff ML, Austin JW, Cyr TD, Hefford MA, Teather RM, Selinger LB (2009) Physical and genetic characterization of an outer-membrane protein (OmpM1) containing an N-terminal S-layer-like homology domain from the phylogenetically Gram-positive gut anaerobe *Mitsuokella multacida*. Anaerobe 15(3): 74–81

Kim JI, Cox MM (2002) The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways. Proc Natl Acad Sci USA 99(12): 7917–7921

Kota S, Misra HS (2008) Identification of a DNA processing complex from *Deinococcus radiodurans*. Biochem Cell Biol 86(5): 448–458

Kota S, Kamble VA, Rajpurohit YS, Misra HS (2010) ATP-type DNA ligase requires other proteins for its activity in vitro and its operon components for radiation resistance in *Deinococcus radiodurans* in vivo. Biochem Cell Biol 88(5):783–790

LaLonde JM, Bernlohr DA, Banaszak LJ (1994) The up-and-down beta-barrel proteins. FASEB J 8(15):1240–1247

Lan R, Reeves PR (2001) When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol 9(9):419–424

Leppanen VM, Tossavainen H, Permi P, Lehtio L, Ronnholm G, Goldman A, Kilpelainen I, Pihlajamaa T (2007) Crystal structure of the N-terminal NC4 domain of collagen IX, a zinc binding member of the laminin-neurexin-sex hormone binding globulin (LNS) domain family. J Biol Chem 282(32):23219–23230

Lipton MS, Pasa-Tolic L, Anderson GA, Anderson DJ, Auberry DL, Battista JR, Daly MJ, Fredrickson J, Hixson KK, Kostandarithes H, Masselon C, Markillie LM, Moore RJ, Romine MF, Shen Y, Stritmatter E, Tolic N, Udseth HR, Venkateswaran A, Wong

KK, Zhao R, Smith RD (2002) Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. Proc Natl Acad Sci USA 99(17):11049–11054

Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, Stair J, Wu L, Thompson DK, Xu D, Rogozin IB, Gaidamakova EK, Zhai M, Makarova KS, Koonin EV, Daly MJ (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. Proc Natl Acad Sci USA 100(7):4191–4196

Lu H, Gao G, Xu G, Fan L, Yin L, Shen B, Hua Y (2009) *Deinococcus radiodurans* PprI switches on DNA damage response and cellular survival networks after radiation damage. Mol Cell Proteomics 8(3):481–494

Marina A, Waldburger CD, Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. EMBO J 24(24):4247–4259

Mennecier S, Servant P, Coste G, Bailone A, Sommer S (2006) Mutagenesis via IS transposition in *Deinococcus radiodurans*. Mol Microbiol 59(1):317–325

Misra HS, Khairnar NP, Kota S, Shrivastava S, Joshi VP, Apte SK (2006) An exonuclease I-sensitive DNA repair pathway in *Deinococcus radiodurans*: a major determinant of radiation resistance. Mol Microbiol 59(4):1308–1316

Narumi I, Satoh K, Cui S, Funayama T, Kitayama S, Watanabe H (2004) PprA: a novel protein from *Deinococcus radiodurans* that stimulates DNA ligation. Mol Microbiol 54(1):278–285

Nekrutenko A, Makova KD, Li WH (2002) The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. Genome Res 12(1):198–202

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Adams MD JJS, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3(6):e170

Novichkov PS, Wolf YI, Dubchak I, Koonin EV (2009) Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. J Bacteriol 191(1):65–73

Omelchenko MV, Wolf YI, Gaidamakova EK, Matrosova VY, Vasilenko A, Zhai M, Daly MJ, Koonin EV, Makarova KS (2005) Comparative genomics of *Thermus thermophilus* and *Deinococcus radiodurans*: divergent routes of adaptation to thermophily and radiation resistance. BMC Evol Biol 5:57

Petersen L, Bollback JP, Dimmic M, Hubisz M, Nielsen R (2007) Genes under positive selection in *Escherichia coli*. Genome Res 17(9):1336–1343

Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P (2012) eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res 40(Database issue):D284-289

Price MS, Chao LY, Marletta MA (2007) *Shewanella oneidensis* MR-1 H-NOX regulation of a histidine kinase by nitric oxide. Biochemistry 46(48):13677–13683

Puehringer S, RoseFigura J, Metlitzky M, Toyama H, Klinman JP, Schwarzenbacher R (2008) Structural studies of mutant forms of the PQQ-forming enzyme PqqC in the presence of product and substrate. Proteins 78(11):2554–2562

Rothfuss H, Lara JC, Schmid AK, Lidstrom ME (2006) Involvement of the S-layer proteins Hpi and SlpA in the maintenance of cell envelope integrity in *Deinococcus radiodurans* R1. Microbiology 152(Pt 9):2779–2787

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738

Sghaier H, Ghedira K, Benkahla A, Barkallah I (2008) Basal DNA repair machinery is subject to positive selection in ionizing-radiation-resistant bacteria. BMC Genomics 9:297

Siebold C, Arnold I, Garcia-Alles LF, Baumann U, Erni B (2003) Crystal structure of the *Citrobacter freundii* dihydroxyacetone kinase reveals an eight-stranded alpha-helical barrel ATP-binding domain. J Biol Chem 278(48):48236–48244

Siew N, Fischer D (2003) Unravelling the ORFan puzzle. Comp Funct Genomics 4(4):432–441

Slade D, Radman M (2011) Oxidative stress resistance in *Deinococcus radiodurans*. Microbiol Mol Biol Rev 75(1):133–191

Slade D, Lindner AB, Paul G, Radman M (2009) Recombination and replication in DNA repair of heavily irradiated *Deinococcus radiodurans*. Cell 136(6):1044–1055

Soding J (2005) Protein homology detection by HMM–HMM comparison. Bioinformatics 21(7):951–960

Soyer Y, Orsi RH, Rodriguez-Rivera LD, Sun Q, Wiedmann M (2009) Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected Salmonella serotypes. BMC Evol Biol 9:264

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34(Web Server issue):W609-612

Tai V, Poon AF, Paulsen IT, Palenik B (2011) Selection in coastal *Synechococcus* (cyanobacteria) populations evaluated from environmental metagenomes. PLoS ONE 6(9):e24249

Tanaka A, Hirano H, Kikuchi M, Kitayama S, Watanabe H (1996) Changes in cellular proteins of *Deinococcus radiodurans* following gamma-irradiation. Radiat Environ Biophys 35(2): 95–99

Tanaka M, Earl AM, Howell HA, Park M-J, Eisen JA, Peterson SN, Battista JR (2004) Analysis of *Deinococcus radiodurans'* transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. Genetics 168:21–33

Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12(10):692–702

Wallace JG, Zhou Z, Breaker RR (2012) OLE RNA protects extremophilic bacteria from alcohol toxicity. Nucleic Acids Res 40(14):6898–6907

White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin H, Jiang L, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Minton KW, Fleischmann RD, Ketchum KA, Nelson KE, Salzberg S, Smith HO, Venter JC, Fraser CM (1999) Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science 286(5444):1571–1577

Yamamura A, Okada A, Kameda Y, Ohtsuka J, Nakagawa N, Ebihara A, Nagata K, Tanokura M (2009) Structure of TTHA1623, a novel metallo-beta-lactamase superfamily protein from *Thermus thermophilus* HB8. Acta Crystallogr Sect F Struct Biol Cryst Commun 65(Pt 5):455–459

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586–1591

Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. Mol Biol Evol 28(3):1217–1228

Zhang C, Wei J, Zheng Z, Ying N, Sheng D, Hua Y (2005) Proteomic analysis of *Deinococcus radiodurans* recovering from gamma-irradiation. Proteomics 5(1):138–143