**SRI VENKATESWARA INTERNSHIP PROGRAM**
**FOR RESEARCH IN ACADEMICS**
**(SRI-VIPRA)**

SRI-VIPRA

# Project Report of 2023: SVP-2312

**"Analysing the structural properties of *Escherichia coli* encoded YqgF"**



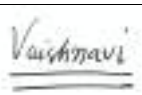
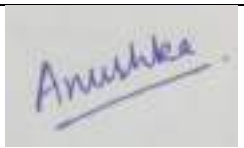
**IQAC**
**Sri Venkateswara College**
**University of Delhi**
**Benito Juarez Road, Dhaula Kuan, New Delhi**
**New Delhi -110021**

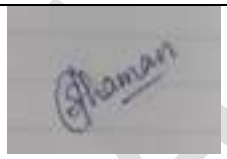
# SRIVIPRA PROJECT 2023

**Title : Analysing the structural properties of *Escherichia coli* encoded YqgF**

| | |
|---|---|
| **Name of Mentor:** Dr. Manoj Thakur<br>**Name of Department:** Botany<br>**Designation:** Assistant Professor |  |
| **Name of Mentor:** Dr. Shweta Sharma<br>**Name of Department:** Botany<br>**Designation:** Assistant Professor |  |

***List of students under the SRIVIPRA Project***

| S.No | Photo | Name of the student | Roll number | Course | Signature |
|---|---|---|---|---|---|
| 1 |  | Vaishnavi | 1821048 | B. Sc. (Hons.) Physics (Semester-V) |  |
| 2 |  | Parul Yadav | 1221031 | B. Sc. (Hons.) Biochemistry (Semester-V) |  |
| 3 |  | Anushka | 1421013 | B. Sc. (Hons.) Botany (Semester-V) |  |

| | | | | | |
|---|---|---|---|---|---|
| 4 |  | Tanisha Babbal | 1221044 | B. Sc. (Hons.) Biochemistry (Semester-V) |  |
| 5 |  | Sonal Thaman | 1421044 | B.Sc. (Hons.) Botany (Semester- V) |  |
| 6 |  | Ayush Ayushman | 1421017 | B.Sc. (Hons.) Botany (Semester- V) |  |

**Signature of Mentor**
Dr. Manoj Thakur

**Signature of Mentor**
Dr. Shweta Sharma

## Certificate of Originality

This is to certify that the aforementioned students from Sri Venkateswara College have participated in the summer project SVP-2312 titled "**Analysing the structural properties of *Escherichia coli* encoded YqgF"**. The participants have carried out the research project work under my guidance and supervision from 15 June, 2023 to 15th September 2023. The work carried out is original and carried out in an online/offline/hybrid mode.

**Signature of Mentor** **Signature of Mentor**

**Dr. Manoj Thakur.** **Dr. Shweta Sharma**

## Acknowledgements

We would like to sincerely thank each and every member of the team who joined this project with their brains together for understanding the basic structural aspects of the *Escherichia coli* encoded YqgF protein. A special thanks to **Dr. Shashank Auroli** (IISc, Bangalore) and **Ms. Kshitiza Mohan Dhyani** (IISc, Bangalore), without their assistance many turns or folds of this proteins would have not been unfolded. We appreciate that both of them have taken their precious time for teaching these students the basic principles associated with understanding the biochemistry of DNA and protein. We also extend our gratitude to our **TIC, Prof. Shukla Saluja** whose constant moral support and interest in our work was inspiring for us to proceed further in the research.

**Vaishnavi**: I wish to express my profound gratitude to Lord Venkateswara, Dr. K.C. Chandramani (Principal), Dr. Sharda Pasricha and Dr. S Krishnakumar (Coordinators, IQAC, SRIVIPRA 2023) and Dept. of Botany whose unwavering support and guidance made my SRIVIPRA project from 15th June, 2023 to 15th Sep, 2023 , possible. This endeavour has been a pivotal chapter in my academic journey, and I am deeply appreciative of the invaluable contributions from all involved. Foremost, I extend my heartfelt appreciation to my esteemed mentor, Dr. Manoj Thakur and Dr. Shweta Sharma. Their expertise, commitment, and patient guidance have been instrumental in shaping the trajectory of my research. I am profoundly indebted for their wisdom and mentorship. I would also like to acknowledge the exceptional collaboration and dedication of my colleagues. Their collective efforts, diverse perspectives, and collaborative spirit enriched our research. Furthermore, I extend my sincere gratitude to my friends and family for their unwavering encouragement and unwavering belief in my capabilities. This summer research undertaking has been transformative and illuminating, enhancing both my academic and personal growth. I am excited to carry forward the knowledge and experiences gained during this project into my future academic and professional endeavours. Once again, I extend my heartfelt thanks to everyone who contributed to the realization of this summer research project. Your unwavering support has been pivotal, and I eagerly anticipate the opportunity to share the findings and insights from this endeavour with the wider community.

**Parul Yadav**: This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in completion of this paper. I am grateful to my institute **Sri Venkateswara College** for this internship program that provided me an opportunity to gain practical exposure. I would like to express my gratitude to my internship supervisors **Dr. Manoj Thakur** and **Dr. Shweta Sharma** for their invaluable guidance and support throughout my internship. From the moment I started, they provided my clear directions

and were always available to answer my questions and provided valuable feedback. I am thankful to **Dr. Shashank** from Indian institute of Science who took out time from his busy schedule to give us insights about the topic and teach us how the work has to be done. I would like to extend my thanks to **Ms. Kshitiza Mohan Dhyani** for an informative session about various bioinformatics tools. Lastly, I would like to appreciate and thank my fellow team members for their full co-operation and support during internship.

**Ayush Ayushman:** I would like to extend my heartfelt gratitude to Dr. Manoj Thakur, Dr. Shweta Sharma and the entire staff at the "Sri Venkateshwara College" for providing me with the invaluable opportunity to complete my internship with Sri Vipra Internship Programme. Your guidance, support, and mentorship throughout this internship have been instrumental in my professional development. I am truly grateful for the knowledge and experience that I have gained during these three months, which I will carry with me as I continue my academic and career journey. Dr. Thakur's dedication to fostering a conducive learning environment and commitment to the success of his students have left a lasting impression on me. I would also like to express my appreciation to the staff and colleagues at Sri Vipra for their warm welcome, support, and collaboration during my internship. It has been a privilege to work with such a talented and dedicated team. Once again, thank you for this enriching experience, and I look forward to applying the skills and knowledge I have acquired to my future endeavours.

**Sonal Thaman**: I would like to express my heartfelt gratitude to my mentors, Dr. Manoj Thakur and Dr. Shweta Sharma, for their invaluable guidance, support, and expertise throughout my internship under the Sri Vipra Program. Their mentorship has been instrumental in shaping my professional growth and enhancing my skills in the research field. I am truly grateful for their unwavering dedication and the opportunities they have provided me. I would also like to extend my appreciation to my colleagues, Anushka and Ayushman, for their collaboration, teamwork, and support during this internship. Their contributions and camaraderie have made this experience even more enriching and enjoyable. Furthermore, I want to acknowledge the support and resources provided by Sri Venkateswara College, University of Delhi. The institution's commitment to academic excellence has created a conducive learning environment and has greatly contributed to my overall development. This experience has been truly transformative and will undoubtedly shape my future endeavours. Thank you all from the bottom of my heart!"

**Anushka**: I would like to take this opportunity to express my heartfelt gratitude to the individuals and organizations who have contributed to the successful completion of this research project. First and foremost, I am deeply thankful to Dr. Manoj Thakur and Dr. Shweta Sharma, my research advisor, whose unwavering support, expert guidance, and

insightful feedback have been instrumental throughout every phase of this project. Your mentorship has enriched my knowledge and skills in immeasurable ways. I am indebted to the members of my research team, Ayush Ayushman and Sonal, for their dedication, collaboration, and tireless efforts: Their diverse perspectives and expertise enriched the project, leading to more comprehensive results Your collective effort has been pivotal in shaping the outcomes of this research. This research project would not have been possible without your collective efforts and dedication.

**Tanisha Babbal:** I would like to express my profound gratitude to **Dr. Manoj Thakur** and **Dr. Shweta Sharma**, of the Botany Department of Sri Venkateswara College for their contributions to the completion of my Sri Vipra internship project. I would like to express my special thanks to our mentor **Dr. Manoj Thakur** for the time and efforts he provided throughout the course of the internship. His useful advice and suggestions were really helpful to me during the project's completion. Moreover, I am also extremely grateful to my peers who helped me by giving valuable critique and help. I would like to acknowledge that this project was completed entirely by me and not by someone else.

## TABLE OF CONTENTS

## Introduction

Studies from the past five decades have uncovered various DNA repair systems and many of these pathways have already being recognised with Noble prize in chemistry 2015 to A. Sancar, T. Lindal and P. Modrich [1]. One of the major molecular machineries that controls the chromosomal stability is represented by homologous Recombination (HR) pathway, that is operated and controlled by multiple protein-protein and protein-DNA network of interactions [2]. From 1960-1995, large wealth of information had been gathered regarding how this pathway works to achieve the recombination/recombinant products, but great deal of the molecular mechanism has been learnt recently based on biochemical, structural and single molecule based studies.

HR is conserved across all three domains of life and is associated with a number of key biological processes [3,4]. HR involves the exchange of information between two identical or nearly identical DNA molecules. In bacteria, HR was discovered in *Escherichia coli* by J. Lederberg and his colleagues (1947), and he named this process of genetic recombination as conjugation. The isolation of DNA damage sensitive recombination mutants indicated a role of HR in repair of damaged DNA [5-7]. Over the years, numerous genetic, biochemical and structural studies have uncovered important mechanistic details and established a role for HR in DNA damage repair, control of DNA replication fidelity and suppression of various types of cancer [8,9]. Additionally, HR contributes to the generation of genetic diversity in the population [10]. In bacteria, diversification of genomes can occur through a number of mechanisms. HR mediates integration of homologous foreign DNA during conjugation or transformation [11,12]. It assorts beneficial and deleterious mutations by allowing allelic recombination between closely related strains [13,14]. Generation of genomic diversity is implicated in the emergence of new strains of pathogenic bacteria, where intrachromosomal recombination is usually adaptive. In eukaryotes, HR is crucial during meiosis for gene-reassortment and proper chromosome segregation [15]. It also maintains genome integrity by preventing chromosome rearrangements and changes in chromosome number. Genomic instability is associated with an increased cancer risk and some examples of such mutations include BRCA1 and BRCA2 leading to breast cancer; the FA genes, leading to Fanconi anaemia; BLM, leading to Bloom syndrome and WRN, leading to Werner syndrome [16-20]. Therefore, our understanding of the role of HR components in the maintenance of genome stability has implications for therapeutic treatments. A vast amount of literature is available describing studies on different aspects of HR in bacteria, however, the general outline of the pathway appeared as described in the next section.

Based on genetic and biochemical analyses, the entire HR pathway can be viewed as four distinct steps: initiation, homologous pairing, extension of heteroduplex DNA and resolution. The initiation step involves nucleolytic processing of duplex DNA to generate ssDNA which is suitable substrate for RecA protein function. This is achieved by the concerted action of nucleases and helicases. In *E. coli*, RecBCD enzyme unwinds and degrades dsDNA but switches directionality of its nuclease activity on encountering a Chi site as described above. This leads to the formation of 3' tailed ssDNA which is the preferred polarity for RecA protein dependent invasion of recipient DNA [21-22]. *In vitro* RecBCD enzyme has many characteristic activities: DNA dependent ATPase, ss- and ds-DNA exonuclease, ssDNA endonuclease and ATP dependent DNA helicase. In addition to its non-specific nuclease activities, RecBCD also exhibit sequence specific nuclease activity that is specific for non-palindromic Chi-site. At first interaction with the dsDNA ends, the enzyme degrades the Chi-containing DNA strand of linear dsDNA processively with a 3′→5′ polarity until it encounters Chi-sequence which acts as a regulatory cis-element that attenuates the 3′-5′ exonuclease activity of RecBCD. While leaving the helicase function of the enzyme unaffected, interaction with the Chi-sequence also changes the polarity of the nuclease activity from 3′→5′ to 5′→3′ direction. Chi-modified RecBCD thus produces a long 3′ ss-DNA overhang which becomes a substrate for the RecA protein: these ssDNA:RecA filaments then undergo strand exchange with homologous pieces of

dsDNA. The next key step in HR is the alignment and pairing of homologous DNA molecules. In both RecBCD or RecFOR pathways, the first step entails the binding of RecA family of proteins to ssDNA, forming a helical nucleoprotein filament, formation of which requires ATP and $Mg^{2+}$ ion [21-26]. *E. coli* RecA can pair homologous ssDNA with duplex DNA with as few as eight bases of homology to form a synaptic complex [27-29]. The deproteinized synaptic complex has been shown to be a DNA triple helix80. Historically, the precise mechanism of homologous pairing remained uncertain; however, two alternate mechanisms have been considered. First, RecA nucleoprotein filament is aligned with homologous dsDNA via non-Watson-Crick hydrogen bonding in its major groove, allowing a transient formation of a novel DNA triplex. However, direct evidence for the formation of triplex structure in vitro, or to visualize it in the EM, have had no substantial success. Using multiple techniques, it was first showed and later substantiated by others that homologous alignment is based simply on Watson-Crick pairing interactions formed through the minor groove [30,31].

After homologous pairing, strand exchange is initiated between two recombining homologous DNA molecules and a process called branch migration occurs. In this process, the unpaired region of ssDNA displaces the complementary DNA strand, moving the branch point in a unidirectional manner. In contrast, spontaneous branch migration proceeds equally in both directions and is unlikely to complete recombination efficiently. Enzyme catalysed branch migration occurs between two DNA molecules provided that the following criteria are met: (a) one of the DNA molecules should have a region of ssDNA at the site which is homologous to the dsDNA and (b) one of the DNA molecules should have an end, a topological necessity, for the rotation of bases for the base pair switch. Branch migration catalysed by RecA family of proteins is directional, requires the presence of ATP; therefore, it is possible to complete recombination generating heteroduplex DNA that can be thousands of base pairs long 87-89. Biochemical and genetic studies indicate that extension of heteroduplex DNA driven by RecA family of proteins proceeds in the 3′→5′ direction with respect to the invading strand [32]. In the case of branch migration involving a pair of dsDNA molecules, as migration proceeds through the duplex-duplex region, HJ is generated. Resolution of this junction by RuvC protein yields products with two complementary heteroduplex structures [33]. Also, branch migration is driven by the junction-specific helicases RuvAB and RecG [34].

In this study we have characterised the structure of Yqgf protein and have concluded that the protein harbours the catalytic triad similar to that of the RuvC family of the proteins. Furthermore, YqgF contains both types of secondary protein elements such as α-helices, and β-sheets, and loops. Based on the already provided X-ray determined crystals structure in the PDB database, we observed that these secondary structure elements arranged 3-dimensionally to form globular quaternary structure that contains two protomers of YqgF subunits. The residues were found to be conserved in *Mycobacterium tuberculosis* and *Deinococcus radiodurans*. Based on our in silico analysis the functional role of three critical amino acids have become apparent. Determination of the active sites and its cognate substrate analysis could provide knowledge about its therapeutic potential in various pathogenic organisms such as mycobacterial species.

## Materials and Methods

### Bioinformatics Analysis & Multiple sequence alignment.

The strategy applied in generation of the figure for multiple sequence alignments was adapted from the previous studies [35, 36] The amino acid sequence of *M. tuberculosis* YqgF and RuvC was acquired from the TubercuList web server (http://www.Pasteur.fr/Bio/TubercuList). The amino acid sequences of other species of bacteria were retrieved from UniProt database and analysed for domain architecture, aligned using a multiple-alignment algorithm with the ClustalOmega series of programs, and visualized using

Jalview. The YqgF amino acid sequence was also analyzed for domain architecture using the Conserved Domain Database in NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

**Model construction**

The strategy applied was adapted from the previous studies as described briefly [37-39]. A three-dimensional models of MtYqgF, EcYqgF, and DrYqgF were constructed with the available crystal structures with the **PDB ID code: 7ESS, 1NMN** and **7W89** respectively, using PyMol. The obtained structures were labelled in the PowerPoint and the designated α-helices, and β-sheets, and loops were differentially coloured with appropriate format in the PyMoL. The appropriate options such as for the charge, hydrophobicity, polarity and for highlighting the individual amino acids as ball and sticks were selected for various models presented in the figure. Based on the literature survey, multiple sequence alignments and conserved domain analysis key residues were highlighted and presented as shown in figures.

**Results and discussion**

Holliday junction resolvases (HJRs) are key enzymes of DNA recombination that leads to the resolution of Holliday junction (HJ) which ultimately produces cross and non-crossovers recombinant products. A detailed bioinformatic analysis of the structural and evolutionary relationships of HJRs and other endonucleases suggests that the HJRs evolution has occurred independently from four distinct structural folds, namely RNase H, endonuclease, endonuclease VII-colicin E and RusA [39]. Within the RNase H fold, a new family of predicted HJRs was discovered in addition to the previously characterized RuvC family whose protein is typified by *E. coli* YqgF (EcYqgF), is ubiquitously present in bacteria. YqgF family members/proteins are likely to function as an alternative to RuvC or principal HJR in most bacteria and in low-GC Gram-positive bacteria and AQUIFEX respectively. Although, a remarkable replacement of RuvC with YqgF has been seen in many organisms across the eubacterial domain, however, *E. coli* genome contains the genes for both the proteins [40]. Despite this the important question remains lingering around the HR pathway that which of the two genes are mainly responsible for the HJ resolution. Weather or not both the gene products are equally responsible for the HR in *E. coli*. If one of the gene product is critical for the HR then what are the signals for their contextual involvement in the pathway? All these questions prompted us to structurally characterise Yqgf whose information is mostly lacking or underrepresented in the literature.

Taking advantage of the diverse set of RuvC sequences that are available in Uniprot as a result of bacterial genome sequencing projects, we investigated the sequence relationships of the YqgF proteins with RuvC proteins. A BLAST search initiated with the *E. coli* RuvC sequence detected other species of RuvC proteins or the orthologous bacterial proteins. In an analogous manner, the sequences of YqgF were retrieved as described in "materials and methods". Examination of a multiple alignment using the ClustalOmega program and Jalview revealed that both of the polypeptides exhibit 20.3 % identity. Intriguingly, the E67 of RuvC was previously shown to be involve in the catalysis of HJ resolution. By aligning EcRuvC with the EcYqgF, we found that Gly at 60$^{th}$ position is present in the alignment against glutamic acid, however we could see that one glutamate residue at 56$^{th}$ position is present in the E. coli but found to be invariable when compared YqgF from other species of bacteria (Figure 1). Furthermore, other than E67, two more aspartate's (D139 and D142) were previously shown to be involved in the reaction catalysed by RuvC. Based on our multiple sequence alignment, D122 of YqgF was found to be in nearby regions of D139 and D142 of RuvC. Since the exact alignment of D122 of YqgF was not observed with the two conserved aspartates of RuvC, the conclusion that D122 of YqgF might be preforming the similar roles as of D139 and D142 in RuvC could be drawn cautiously. The effect of inconsistency in the perfect

alignment could be negated because the flexibility and the molecular beathing could bring the precise juxtapositions of the amino acid residues being analysed.

Previous studies have shown that the signature catalytic residues of the RuvC protein encompasses the acidic residue, typically an aspartate towards the end of strand 1, a glutamate near the end of conserved strand 4 and two acidic residues (DxxD) associated with the C-terminal α-helix [39]. These residues mark the active site of RuvC and shown to be critical for the for the resolvase activity where all of these three residues form a spatially juxtaposed acidic triad that could coordinate divalent cations. Based on this knowledge, we developed a hypothesis with a naive question that; is there a possibility for the existence of such catalytic triad in YqgF? In line with this, model of the EcYqgF was generated using X-ray crystal based structures as described in "materials and method". After removing the water molecules, the obtained structure was analysed in PyMoL and differentially coloured to highlight the secondary structure elements present in the protein. As shown in Figure 2, panel A, EcYqgF shows the composition of α-helixes, and β-

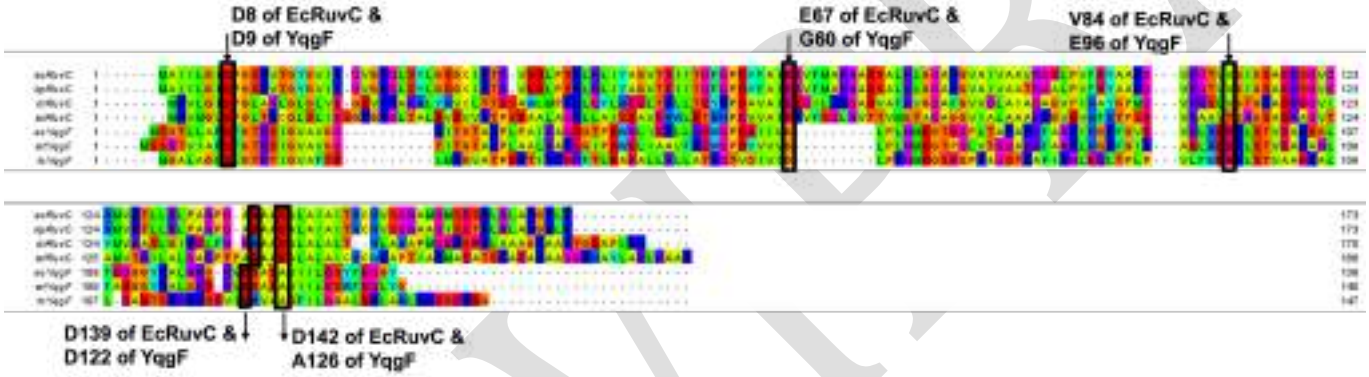



**Figure 1: Sequence alignment of RuvC and YqgF.** The sequences of all the mentioned proteins were retrieved from the UniProt database and the multiple alignment of all the amino acid sequences were performed using ClustalOmega. The image was generated using Jalview. The sequences used were of *Escherichia coli* RuvC, *Klebsiella pneumoniae* RuvC, Deinococcus radiodurans RuvC, *Mycobacterium tuberculosis* RuvC, *E. coli* YqgF, *M. tuberculosis* YqgF, and *Trichonephila clavipes* YqgF. The D9 of YqgF was found to be conserved with D8 of EcRuvC. D8, E67, D139 and D142 are the active site residues of the EcRuvC protein.

sheets, and loops. Precisely, EcYqgF contains three α-helixes, five β-sheets, and seven loops. The protomer of YqgF appears globular in shape and consisted of 138 amino acids. The crystal structure of YqgF have clearly revealed that it exists as dimers (Figure 2B). The secondary structures comes in the following hierarchical manner: N-terminus-β-sheet 1→ β-sheet 2→ β-sheet 3→α-helix 1 → β-sheet 4→α-helix 2 → β-sheet 5→α-helix 3- C-terminus (Figure 2C). The multiple alignment-based secondary structure prediction for the EcYqgF proteins revealed that such succession of elements are typical for the members of the RNase H fold. Furthermore, multiple alignments of the YqgF and RuvC members along with RuvC and YqgF's 3-dimensional structure indicates that the proximal and distal aspartates of the RuvC catalytic triad are conserved in the YqgF whereas the glutamate at the end of strand 4 in RuvC family, present as a conserved glutamate at the end of strand 5 in YqgF (Figure 2D-F). Altogether, the spatial proximity of the end of strand 5 to the two other conserved acidic residues strongly suggests that the YqgF is a nucleases with a catalytic mechanism similar to that of RuvC.

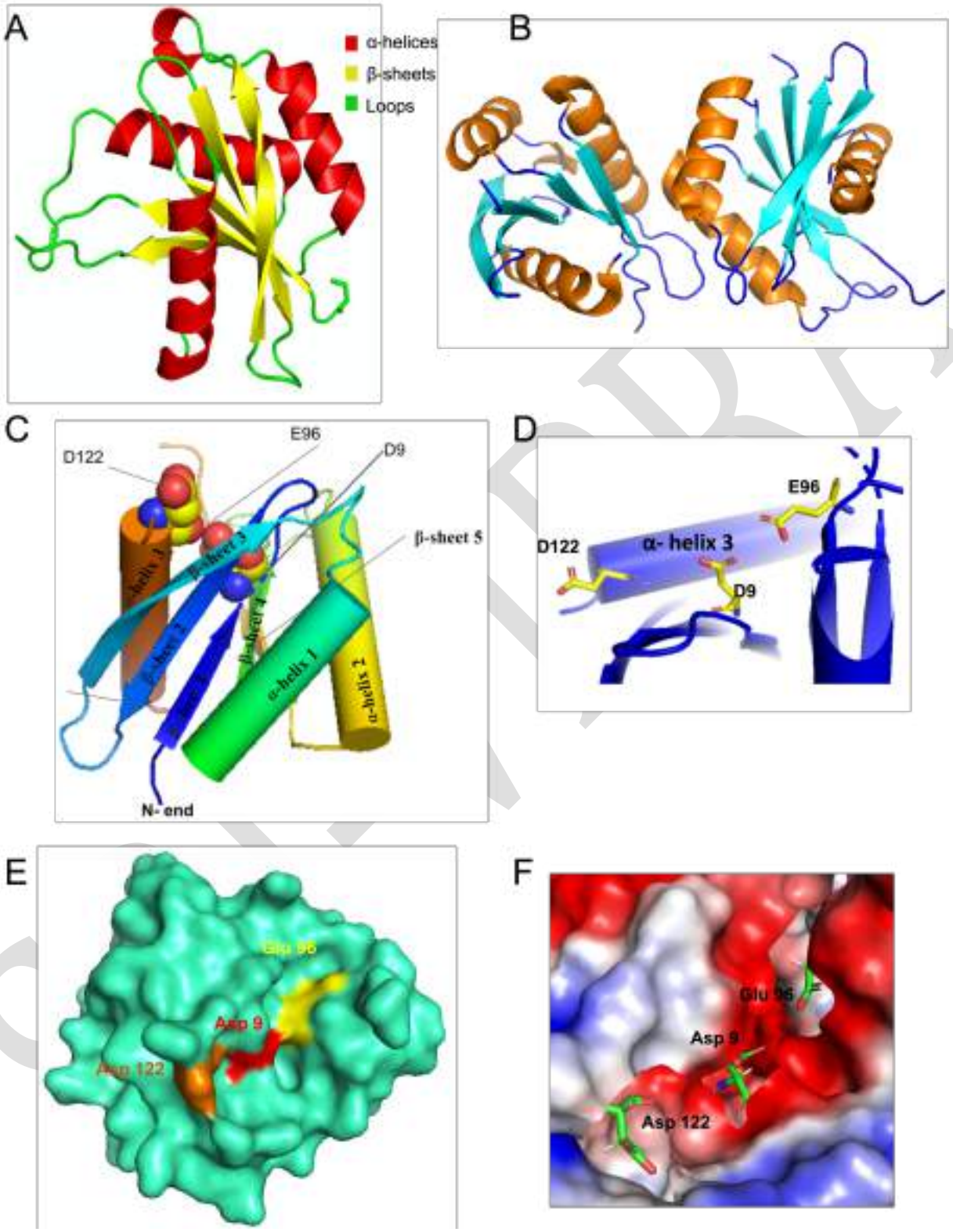



**Figure 2: EcYqgF harbours the catalytic triad analogous to the members of RuvC superfamily.** (A), Structure of YqgF monomer based on PDB ID: 1NMN; (B), Dimeric structure of YqgF; (C), Monomeric form of the protein showing the three amino acids juxtaposed to each other for probable function; (D) Zoomed in image of panel C to highlight the critical residues of catalytic triad; (E), Surface view along with three catalytic residues; and (F), Zoomed in image of panel E with critical residues of catalytic triad.

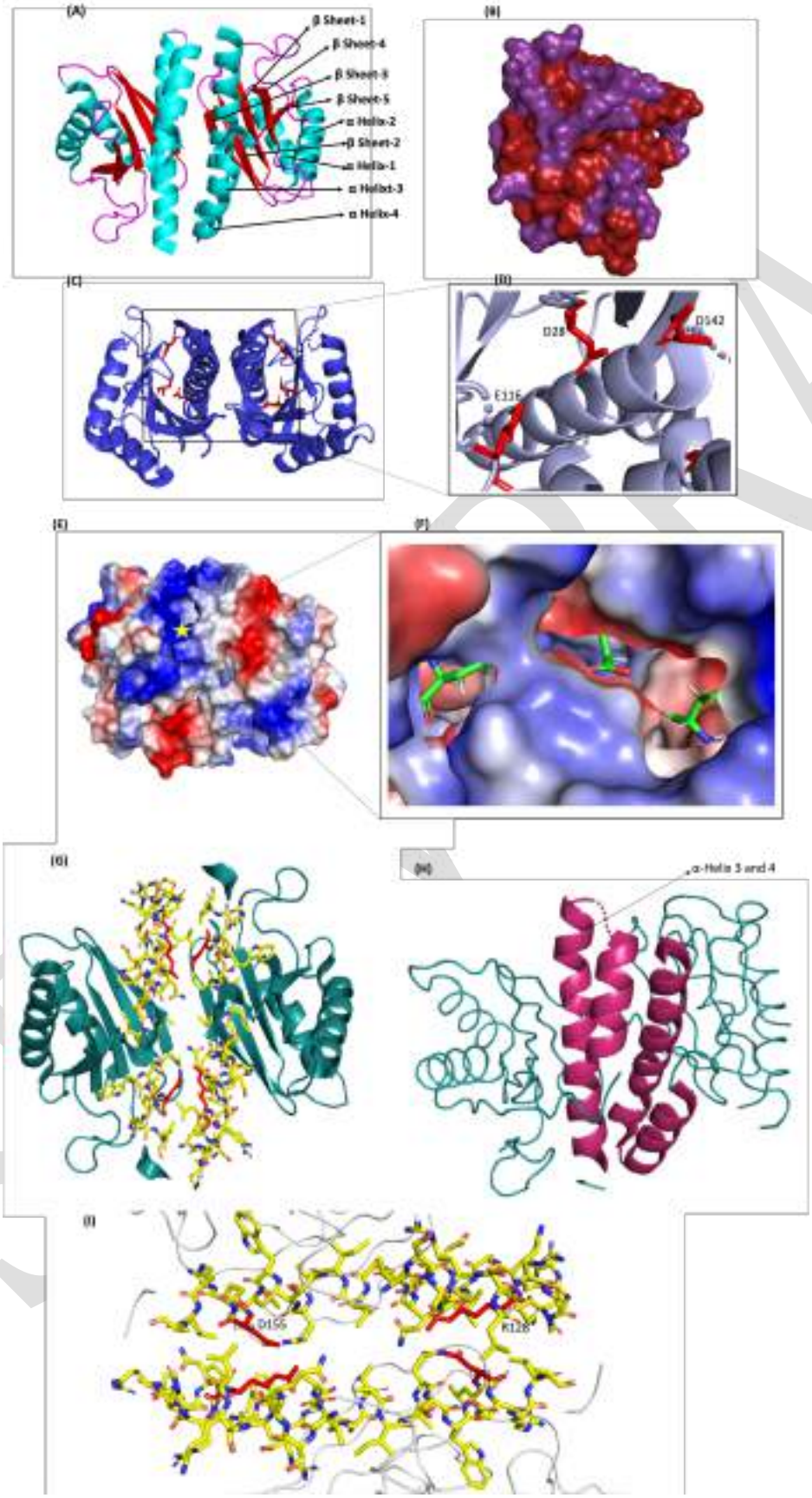

(A)
β Sheet-1
β Sheet-4
β Sheet-3
β Sheet-5
α Helix-2
β Sheet-2
α Helix-1
α Helix-3
α Helix-4
(B)
(C)
(D)
D28
D142
E126
(E)
(F)
(G)
(H)
α-Helix 3 and 4
(I)
D155
K128

**Figure 3: Structural characterization of *M. tuberculosis* YqgF**. (A-B) Overall architecture of crystal structure (PDB ID: 7ESS) based on ribbon diagram (A) and surface view (B) based on polar (purple) and hydrophobic (red) regions . (C) Crystal structure of MtYqgF highlighting the putative catalytic active site residues. (D) Zoomed in image of Panel C displaying the orientation of D28, E116 and D142 amino acid residues. Colour on the structure of MtYqgF is according the electrostatic potential with red extreme corresponds to the negative charge whereas blue regions corresponds to the positive charge on the protein. The electrostatic potential on the MtYqgF structure was determined using the adaptive Poisson Boltzmann Solver excluded surface by PyMOL. (G) Crsytal structure highlight the dimeric regions of YqgF. The residues involved are depicted in yellow sticks. The conserved resides in the dimeric regions are highlighted in the red sticks (panel I) (R128 and D155). (H) cartoon diagram highlighting the role of α-helices 3 and 4 in dimeric interface interactions.

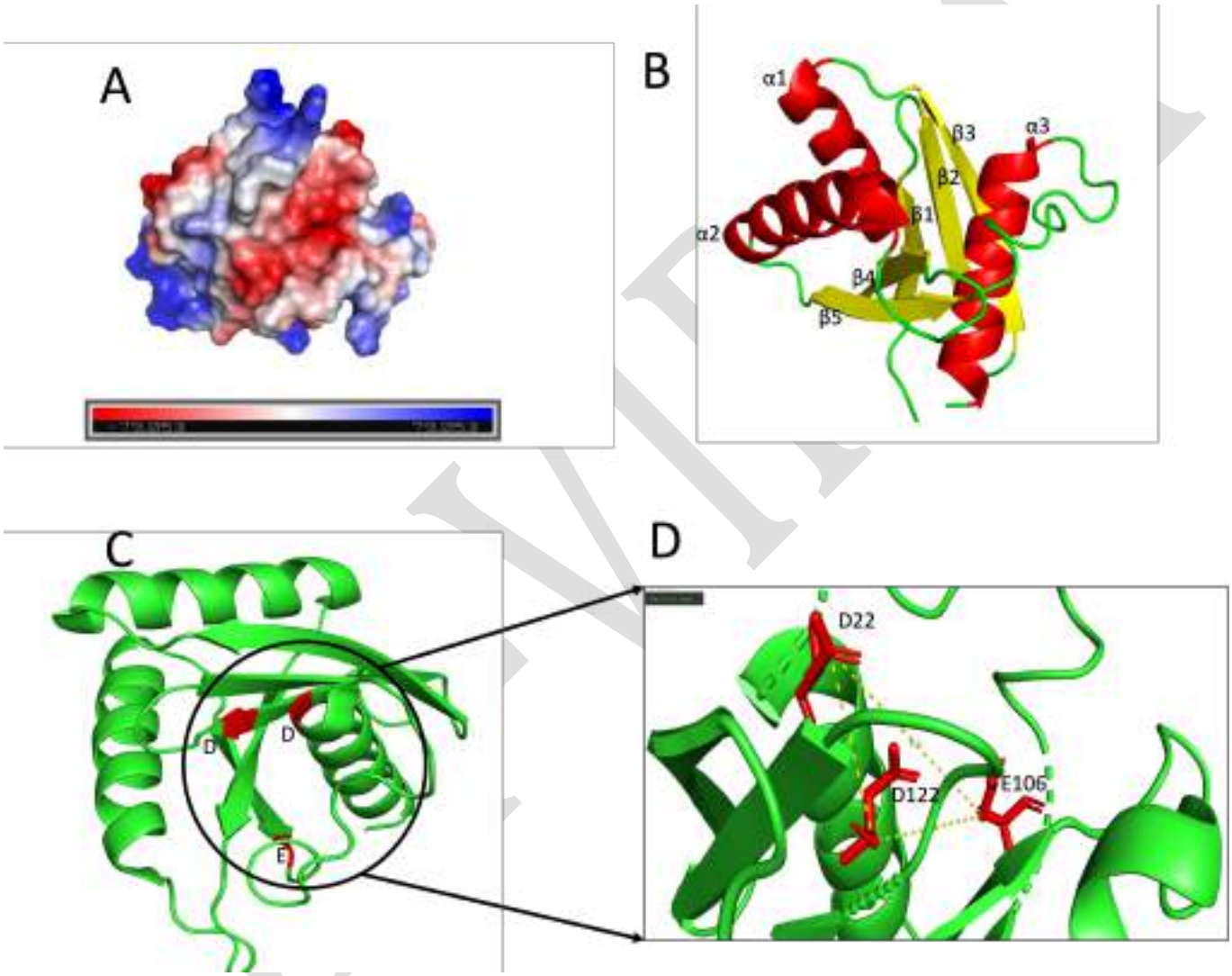



**Figure 4: Structural analysis of *Deinococcus radiodurans* YqgF.** Surface view (A), and structure based on crystal structure (B) with PDB ID: 7W89; (C), catalytic triad of DrYqgF consisted of DDE; (D) DDE corresponds to D22, D122 and E106.

Having a leverage of the availability of the crystals structures in the PDB database, we extended our study with orthologs of YqgF in *Mycobacterium tuberculosis* and *Deinococcus radiodurans.* Structural analysis of EcYqgF revealed that MtYqgF also exist as homodimers in solutions and appears globular in shape (Figure 3A-B). Consistent with the previous studies, PyMol analysis of the crystal structure of MtYqgF highlighted the catalytic triad which in this case consisted of residues D28, E116, and D143 (Figure 3C-F).

In contrast to EcYqgF, there existed a enough of the electron density corresponding to the dimeric interface with the involvement of conserved and many crucial residues in α-helix 3 and 4 (Figure 3G-I). Recently the crystal structure of the DrYqgF has been solved and thus we subjected this subunit to the same analysis. We observed that the in contrast to the EcYqgF and MtYqgF, DrYqgF exists as monomers in the crystals and in solutions based techniques [38]. It harbours DED as a catalytic triad where the two aspartates are present at 22 and 122 position and glutamate at 106$^{th}$ position in the polypeptide. Such startling differences could be due to the species-specific variation in the proteins. Overall, our study and its results highlight the fact that the YqgF indeed is a novel endonuclease which harbours the catalytic residues of RuvC. Although, the strong conclusions need further validations based on biochemical and genetic studies, such analysis can dictate the screening of biochemical inhibitors and future drug design of bacterial based pathogens.

**References**


1]. Kunkel TA. Celebrating DNA's Repair Crew. Cell. 2015 Dec 3;163(6):1301-3. doi: 10.1016/j.cell.2015.11.028. PMID: 26638062.

2]. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994) Biochemistry of homologous recombination in Escherichia coli, Microbiol Rev 58, 401-465.

3]. Lusetti, S. L. and Cox, M. M. (2002). The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. Annu. Rev. Biochem.71, 71-100.

4]. Zickler D. and Kleckner N. (2015) Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb. Perspect. Biol. 7(6) pii: a016626.

5]. Clark, A. J., and Margulies, A. D. (1965) Isolation and characterization of recombination-deficient mutants of Escherichia coli K12. Proc. Natl. Acad. Sci. U. S. A. 53, 451-459.

6]. Clark, A. J., Chamberlin, M., Boyce, R. P., and Howard-Flanders, P. (1966) Abnormal metabolic response to ultraviolet light of a recombination deficient mutant of Escherichia coli K12. J. Mol. Biol. 19, 442-454.

7]. Howard-Flanders, P. (1973). DNA repair and recombination. Br. Med. Bull. 29, 226-235.

8]. Cromie, G. A., Connelly, J. C., and Leach, D. R. (2001) Recombination at double-strand breaks and DNA ends: conserved mechanisms from phage to humans. Mol. Cell 8, 1163-1174.

9]. Krajewska, M., Fehrmann, R. S., de Vries, E. G. and van Vugt, M. A. (2015) Regulators of homologous recombination repair as novel targets for cancer treatment. Front. Genet. 6, 96. doi: 10.3389/fgene.2015.00096. eCollection.

10]. Biek, R., Pybus, O. G., Lloyd-Smith, J. O., and Didelot, X. (2015) Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. 30, 306-313.

11]. Smith, G. R. (1991) Conjugational recombination in E. coli: myths and mechanisms. Cell 64, 19-27.

12]. Lorenz, M.G., and Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. 58, 563-602.

13]. Otto, S. P. and Y. Michalakis. (1998) The evolution of recombination in changing environments. Trends Ecol. Evol. 13, 145-151.

14]. Feil, E. J. (2004). Small change: keeping pace with microevolution. Nat. Rev. Microbiol. 2, 483-495.

15]. Kleckner N. (1996) Meiosis: how could it work? Proc. Natl. Acad. Sci. U. S. A. 93, 8167-8174.

16]. Jasin, M (2002) Homologous repair of DNA damage and tumorigenesis: the BRCA connection. Oncogene 21, 8981–8993.

17]. Moynahan, M. E. (2002). The cancer connection: BRCA1 and BRCA2 tumor suppression in mice and humans. Oncogene 21, 8994-9007.

18]. Hickson, I.D., Robson, C.N., Atkinson, K.E., Hutton, L., and Emmerson, P.T. (1985) Reconstitution of RecBC DNase activity from purified Escherichia coli RecB and RecC proteins. J. Biol. Chem. 260, 1224-1229.
19]. Surralles, J., Jackson, S. P., Jasin, M., Kastan, M. B., West, S. C., and Joenje, H. (2004) Molecular cross-talk among chromosome fragility syndromes. Genes Dev. 18, 1359-1370.
20]. Kennedy, R. D. & D'Andrea, A. D. (2005) The Fanconi anemia/ BRCA pathway: new faces in the crowd. Genes Dev. 19, 2925–2940.
21].Dixon, D.A., and Kowalczykowski, S.C. (1991) Homologous pairing in vitro stimulated by the recombination hotspot, Chi. Cell 66, 361-371.
[22]. Konforti, B. B., and Davis, R. W. (1987) 3' homologous free ends are required for stable joint molecule formation by the RecA and single-stranded binding proteins of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 84, 690-694.
[23]. Konforti, B. B., and Davis, R. W. (1990) The preference for a 3' homologous end is intrinsic to RecA-promoted strand exchange. J. Biol. Chem. 265, 6916-6920.
[24]. Konforti, B. B., and Davis, R. W. (1991) DNA substrate requirements for stable joint molecule formation by the RecA and single-stranded DNA-binding proteins of Escherichia coli. J. Biol. Chem. 266, 10112-10121.
[25]. Sakai, A., and Cox, M. M. (2009) RecFOR and RecOR as distinct RecA loading pathways. J. Biol. Chem. 284, 3264-3272.
[26]. Leahy, M. C., and Radding, C. M. (1986) Topography of the interaction of recA protein with single-stranded deoxyoligonucleotides. J .Biol. Chem. 261, 6954-6960.
[27]. Hsieh, P., Camerini-Otero, C. S., and Camerini-Otero, R. D. (1992) The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. Proc. Natl. Acad. Sci. U. S. A. 89, 6492-6496.
[28]. De Vlaminck, I., van Loenhout, M. T., Zweifel, L., den Blanken, J., Hooning, K., Hage S., Kerssemakers, J., Dekker, C. (2012) Mechanism of homology recognition in DNA recombination from dual-molecule experiments. Mol. Cell 46, 616-624.
[29]. Rao, B. J., Dutreix, M., and Radding, C. M. (1991) Stable three-stranded DNA made by RecA protein. Proc. Natl. Acad. Sci. U. S. A. 88, 2984-9988.
[30]. Kumar, K. A., Muniyappa, K. (1992) Use of structure-directed DNA ligands to probe the binding of recA protein to narrow and wide grooves of DNA and on its ability to promote homologous pairing. J. Biol. Chem. 267, 24824-24832.
[31]. Baliga, R., Singleton, J. W., and Dervan, P. B. (1995) RecA oligonucleotide filaments bind in the minor groove of double-stranded DNA. Proc. Natl. Acad. Sci. U. S. A. 92, 10393-10397.
[32]. San Filippo, J., Sung, P., and Klein, H. (2008) Mechanism of eukaryotic homologous recombination. Annu. Rev. Biochem. 77, 229–257.
[33]. Renkawitz, J., Lademann, C. A and Jentsch, S. (2014) Mechanisms and principles of homology search during recombination. Nat. Rev. Mol. Cell Biol.15, 369-383.
34]. Whitby, M.C., Ryder, L., and Lloyd, R.G. (1993) Reverse branch migration of Holliday junction by RecG protein: a new mechanism for resolution of intermediates in recombination and DNA repair. Cell, 75, 341-350.
35]. Thakur M, Kumar MB, Muniyappa K. Mycobacterium tuberculosis UvrB Is a Robust DNA-Stimulated ATPase That Also Possesses Structure-Specific ATP-Dependent DNA Helicase Activity. Biochemistry. 2016 Oct 18;55(41):5865-5883. doi: 10.1021/acs.biochem.6b00558.
36]. Thakur M, Muniyappa K. Deciphering the essentiality and function of SxSx motif in Mycobacterium tuberculosis UvrB. Biochimie. 2020 Mar;170:94-105. doi: 10.1016/j.biochi.2020.01.003.

37]. Thakur M, Mohan D, Singh AK, Agarwal A, Gopal B, Muniyappa K. Novel insights into ATP-Stimulated Cleavage of branched DNA and RNA Substrates through Structure-Guided Studies of the Holliday Junction Resolvase RuvX. J Mol Biol. 2021 Jun 25;433(13):167014. doi: 10.1016/j.jmb.2021.167014.
38]. Sun Y, Yang J, Xu G, Cheng K. Biochemical and Structural Study of RuvC and YqgF from Deinococcus radiodurans. mBio. 2022 Oct 26;13(5):e0183422. doi: 10.1128/mbio.01834-22.
39]. Aravind L, Makarova KS, Koonin EV. SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. Nucleic Acids Res. 2000 Sep 15;28(18):3417-32. doi: 10.1093/nar/28.18.3417.
40]. Iwamoto A, Osawa A, Kawai M, Honda H, Yoshida S, Furuya N, Kato J. Mutations in the essential Escherichia coli gene, yqgF, and their effects on transcription. J Mol Microbiol Biotechnol. 2012;22(1):17-23. doi: 10.1159/000336517.