STRUCTURAL AND PHYLOGENETIC STUDY OF RecA, AlkB & ENDONUCLEASES AND THEIR ROLE IN MUTATIONAL DYNAMICS

[1]Akash Ajay, [2]Sneha Dhankar, [3]Gayatri Singh

[1]Designation of 1st author => PhD Student, SES, Jawaharlal Nehru University,
[2]Designation of 2nd Author => PhD Student, SES, Jawaharlal Nehru University,
[3]Designation of 3rd Author => PhD Student, SES, Jawaharlal Nehru University

[1]School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India 110067

Abstract : DNA repair proteins are important in influencing the fidelity of genomic apparatus of the cell and thereby affecting the mutational dynamics. Here we study the phylogeny and structural characteristics of DNA repair proteins in E.coli – RecA, AlkB & Endonucleases and understand their implications in mutational dynamics in E.coli.

IndexTerms – **DNA repair, RecA, AlkB, Endonuclease V, Mutations**

I. INTRODUCTION

At the heart of evolution lies the intricate dance of mutations, shaping the genetic landscape of living organisms over vast spans of time. (Rosenberg et al., 2001) Mutational dynamics refers to the patterns, frequencies, and mechanisms by which genetic mutations arise and propagate within populations (Drake et al., 1998). This fascinating field offers a window into the evolutionary tapestry, shedding light on the forces that drive genetic diversity, adaptation, and speciation (Rosenberg et al ., 2001). Mutations are the raw material of evolution, serving as the foundation upon which natural selection acts (Simmons et al., 2007). They arise from a variety of sources, including errors during DNA replication, exposure to mutagenic agents, and even the activity of transposable elements— genetic sequences that can "jump" around the genome, sometimes causing mutations in the process. Understanding the types of mutations that predominate in a population is also crucial (Dake et al., 1998). Point mutations involve single-base changes and are the most common type of mutation. Insertions and deletions, where small segments of DNA are added or removed, can have significant consequences for the resulting protein products. Larger structural mutations, such as chromosomal rearrangements, can lead to even greater genetic diversity.

Mutational dynamics also interact with natural selection, the process through which advantageous mutations become more prevalent in a population (Rosenberg et al., 2001). Beneficial mutations that enhance an organism's fitness have a higher chance of being passed on to subsequent generations, contributing to adaptation and survival. Conversely, detrimental mutations might be eliminated from the gene pool through negative selection. In this study we try to understand the role of three important proteins involved in the DNA replication and error correction mechanisms in E.coli – alkB, recA, EndonucleaseV (Sancar et al., 1988). We will use phylogeny and homology based structural modelling as tools to understand the mechanistic influence of these proteins on mutational dynamics (Kumar et al., 1994, Kreiger et al.,2003).

II.   RESEARCH METHODOLOGY

1.   Data and Sources of Data

For this study secondary data has been collected from NCBI and EBI websites. For phylogenetic alignment we have used standard sequences from EBI nucleotides. For homology-based modelling using Expasy, we used the sequence form NCBI protein. We have included in Prokaryotes and not included eukaryote proteins in our analysis.

2.1 Theoretical framework

We have chosen 3 important DNA repair enzymes to characterise
1.   RecA
In the intricate world of DNA maintenance and repair, RecA stands as a pivotal protein that plays a central role in orchestrating processes critical for maintaining the integrity and diversity of genetic material. RecA, first discovered in Escherichia coli (E. coli), has since been identified in various organisms and is known for its remarkable ability to catalyze DNA strand exchange—a fundamental step in DNA repair and genetic recombination. RecA's most well-known function is its involvement in homologous recombination, a process essential for the repair of DNA double-strand breaks and the exchange of genetic information between homologous DNA molecules. When a double-strand break occurs, RecA assembles onto the single-stranded DNA regions generated by the break. This RecA-coated single-stranded DNA then searches for a homologous region within another DNA molecule, often the sister chromatid or a homologous chromosome. Once a suitable region is found, RecA catalyzes the strand invasion, leading to the formation of a joint molecule called a Holliday junction. This junction can then be resolved to repair the break or facilitate genetic exchange. Beyond its role in recombination, RecA is also central to the SOS response, a complex and highly regulated

cellular reaction to severe DNA damage. In response to extensive DNA damage, the cell activates the SOS response, which involves the upregulation of a suite of genes involved in DNA repair, tolerance, and mutagenesis. RecA plays a key role in this response by mediating the autocatalytic cleavage of LexA, a repressor protein that normally prevents the expression of SOS genes. The cleavage of LexA allows the induction of these genes, providing the cell with tools to cope with the DNA damage.

2. AlkB Enzymes

Discovered in various organisms, from bacteria to humans, AlkB enzymes are part of the AlkB family of proteins. These enzymes utilize a unique oxidative demethylation mechanism to reverse the damage caused by alkylated DNA bases. They specifically target alkylated bases like 1-methyladenine (1meA) and 3-methylcytosine (3meC), catalyzing the removal of the alkyl group from the damaged base.

The AlkB repair process involves several steps: recognition of the damaged base, oxidation of the methyl group, and subsequent restoration of the DNA base to its unmodified state. This remarkable mechanism ensures that the genetic information encoded in DNA remains accurate, even in the face of challenges posed by alkylating agents, which can include environmental toxins and endogenous metabolites.

3. Endonucleases

Endonucleases play a pivotal role in DNA repair. The cellular machinery responsible for repairing damaged DNA relies on endonucleases to initiate the excision of DNA segments containing lesions or errors. The repair process can lead to the removal of damaged segments and the synthesis of new DNA strands, ultimately restoring the integrity of the genome.

2.2 Bioinformatics tools

2.2.1 MEGA

We used MEGA12 (Kumar et al., 1994) to plot the maximum likelihood trees. We aligned the DNA sequences using MUSCLE algorithm and then used neighbor joining method to cluster the sequences. The phylogenetic trees were constructed using Maximum likelihood method with underlying sequence model GTR +G+I which is the most realistic model to plot the phylogenetic trees. We performed 100 bootstrap iterations.

2.2.2 Swiss Expasy

We used Swiss Expassy suite of tools to find the most suitable template to model the protein sequence (Krieger et al., 2003). The protein sequeces were modelled using Homology modeling which works under the principle of similar sequences tend to fold to similar sequences. It infers similarity in the sequence to be the basis of homology and models the unknown sequence of protein using the template the quality of the homology model generated by Expasy depends on the similarity between your target protein and the chosen template. The more similar they are in terms of sequence and function, the more reliable the model is likely to be.

2.2.3 AlphaFold2

We then modelled the protein and viewed the protein using AlphaFold 2 and compared with our previously obtained results using homology modelling (Skolnick et al., 2021). AlphaFold2's predictions are known to be highly accurate rivalling the experimental methods such as X-ray crystallography.

2.2.4 Ramachandran plot

A Ramachandran plot is a graphical representation used in structural biology and biochemistry to analyze the dihedral angles of a protein's backbone. It helps researchers understand the allowed and disallowed regions of torsional angles (phi and psi) for amino acid residues in a protein's structure

Protein structures can be analyzed by plotting the phi and psi angles of each amino acid residue onto the Ramachandran plot. Data points representing individual residues are plotted as dots on the graph.
- Points within the allowed regions indicate that the protein's backbone conformation is in a stable and energetically favorable state.
- Points in the disallowed regions suggest that the protein may have structural problems or steric clashes, and these regions should be avoided during protein modeling or refinement.
- Points near the boundaries between allowed and disallowed regions may indicate regions of potential structural strain or flexibility that warrant further investigation.

III. RESULTS AND DISCUSSION

1.1 Phylogenetic trees of the DNA repair proteins

Plotting the DNA sequences given in Table 1 for recA, alkB and endonucleases gave us phylogenetic trees for the respective proteins. We included only prokaryote genes for plotting the proteins and excluded eukaryote genes. The phylogenetic tree for the three DNA repair proteins is given in Fig1 A,B&C respectively

Figure 1 – Phylogenetic tree for DNA repair proteins- (A.) RecA (B.) AlkB (C.) Endonuclease
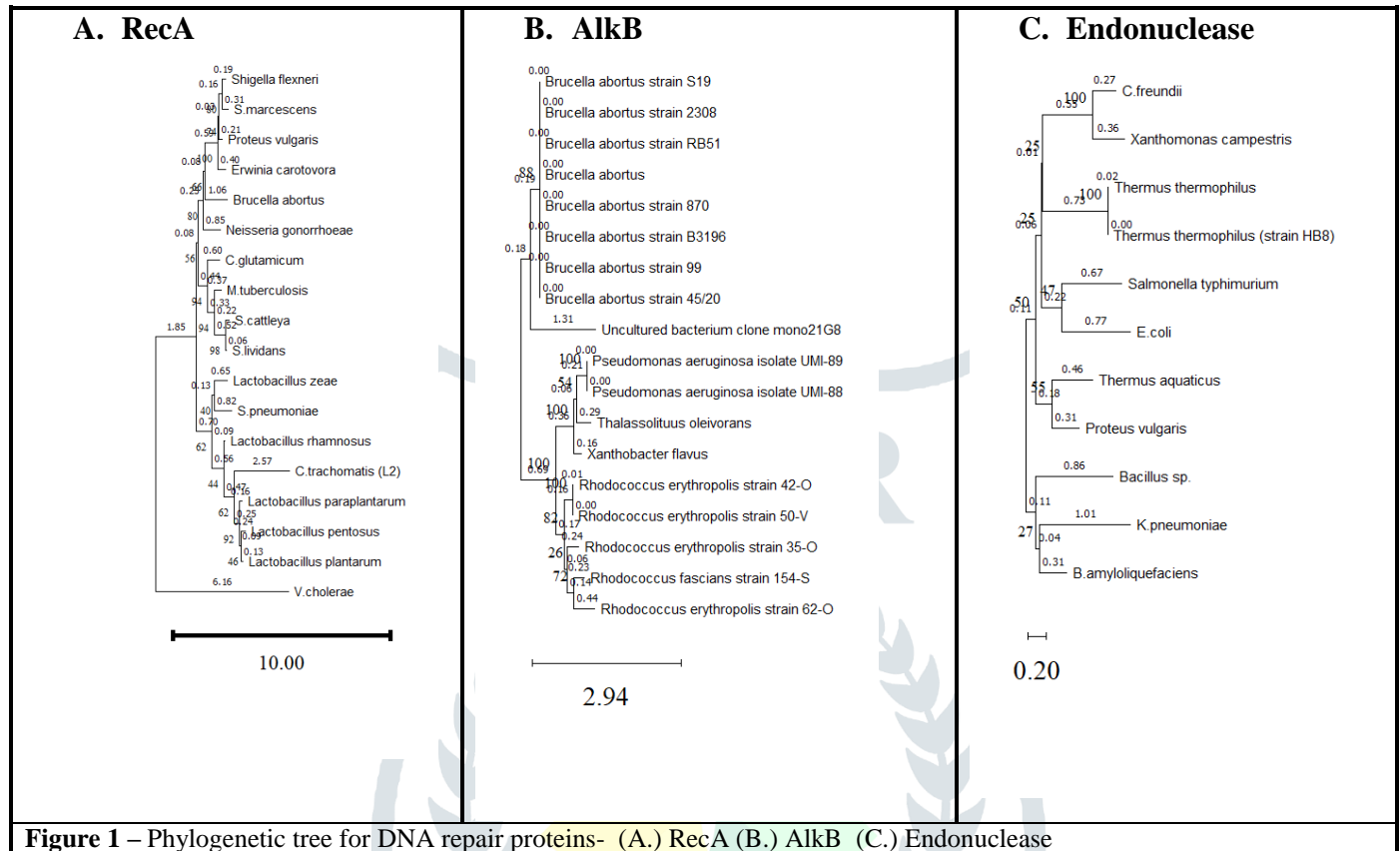
The phylogenetic tree helps us to understand the evolutionary relationships between the 3 DNA repair proteins – For recA we can see that V.cholarae diverged early from the common ancestor. The Lactobacillus recA proteins are forming a clade. Pathogens *Shigella flexneri, Proteus vulgaris, Erwina carotovora,* form a monophyletic clade indicating shared evolutionary history. *S.pneumonae* falls within Lactobacillus clades for recA. For AlkB we can see that *Brucella* strains forms a monophyletic clade and *Rhodococcus* clades form another monophyletic clade. *Pseudomonas aeruginosa,* a pathogen causing pneumonia and hospital infections is closer to *Rhodococcus* than *Brucella* genus. For Endonuclease we can see that *Proteus vulgaris* is closer to *Thermus aquaticus* which is thermophilic bacteria and is the source of the molecular biology enzyme Taq polymerase.

1.2 Homology modelling with Swiss Expasy

The homology modelling of recA, alkB and endonuclease was done with the help of Swiss Expasy. The amino -acid sequences of the 3 proteins were downloaded from NCBI website. The templates were searched by Swiss Expasy itself and then these templates were used to construct protein models for the unannotated sequences. These modelled proteins were viewed on AlphaFold2.
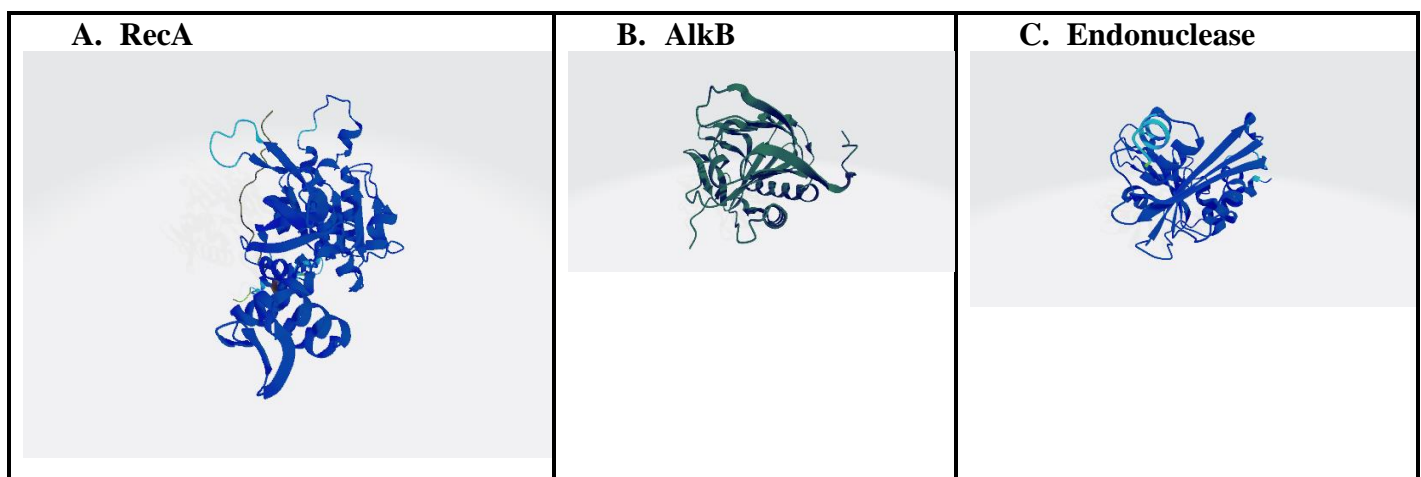
Figure 2 – Homology modeling for DNA repair proteins- (A.) RecA (B.) AlkB (C.) Endonuclease

Homology models are theoretical-computational approximations of the real protein structures, and therefore require validation and sometimes refinement and optimization. A very popular validation tool is the Ramachandran plot which analyzes the stereochemical quality of protein structures. We plot Ramachandran plots for the modelled proteins.
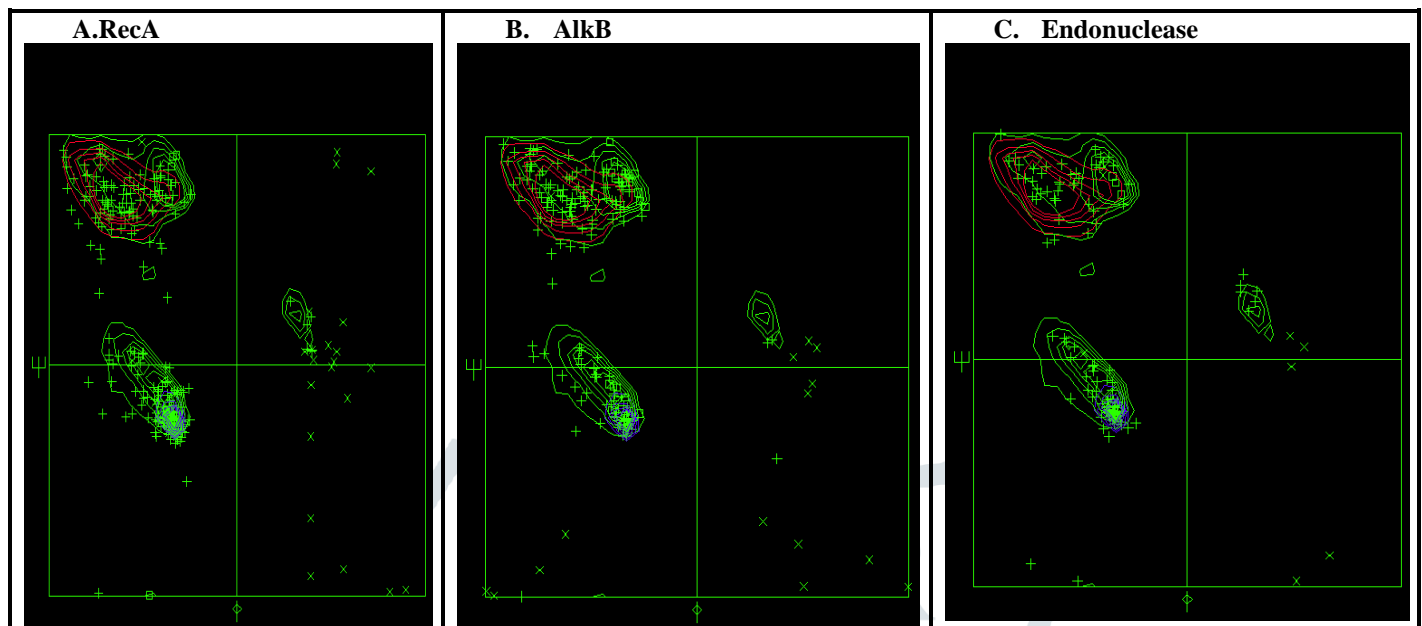
Figure 2 –Ramachandran plot for models plotted in Figure 2- (A.) RecA (B.) AlkB (C.) Endonuclease

We can see form our Ramachandran plots that most of the datapoints inferred form the homology models of recA, AlkB and Endonuclease are accepted in this region. Most of phi and psi angles are in energetically favourable regions, which are also called as allowed regions. Very few points are in unfavourable energy conditions which may indicate minor inaccuracies in modelling

REFERENCES

[1] Rosenberg, S. M. (2001). Evolving responsively: adaptive mutation. *Nature Reviews Genetics*, *2*(7), 504-515.

[2] Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), 1667-1686.

[3] Sancar, A., & Sancar, G. B. (1988). DNA repair enzymes. *Annual review of biochemistry*, *57*(1), 29-67.

[4] Roca, A. I., Cox, M. M., & Brenner, S. L. (1990). The RecA Protein: Structure and Functio. *Critical reviews in biochemistry and molecular biology*, *25*(6), 415-456.

[5] Fedeles, B. I., Singh, V., Delaney, J. C., Li, D., & Essigmann, J. M. (2015). The AlkB family of Fe (II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *Journal of Biological Chemistry*, *290*(34), 20734-20742.

[6] Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cellular and molecular life sciences*, *62*, 685-707.

[7] Krieger, E., Nabuurs, S. B., & Vriend, G. (2003). Homology modeling. *Structural bioinformatics*, *44*, 509-523.

[8] Kumar, S., Tamura, K., & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Bioinformatics*, *10*(2), 189-191.

[9] Skolnick, J., Gao, M., Zhou, H., & Singh, S. (2021). AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *Journal of chemical information and modeling*, *61*(10), 4827-4831.

[10] Simmons, M. J., & Crow, J. F. (1977). Mutations affecting fitness in Drosophila populations. *Annual review of genetics*, *11*(1), 49-78.