International Journal of Biomedical Research

COMPARATIVE SEQUENCE AND STRUCTURAL CONFORMATION ANALYSIS OF RECA PROTEIN

Amajala Krishna Chaitanya*, DSVGK. Kaladhar, S. Suresh, I. Bhaskar Reddy

Department of Biochemistry/Bioinformatics, GITAM Institute of Science, GITAM University, Visakhapatnam-530045, Andhra Pradesh, India

***Corresponding Author:** chy2ak@gmail.com

This article is available online at www.ssjournals.com

ABSTRACT

Water molecules are very important for stabilizing the protein structures. The influence of hydration (or) dehydration on protein structures leads to the changes in the activity of proteins as well their conformation. Our current study deals in understanding the various organisms' RecA protein sequence similarity with respect to *E. coli* RecA protein (*EcRecA*) and comparative structural conformation studies between *M. smegmatis* RecA protein (*MsRecA*) (dehydrated state) and *E. coli* RecA + DNA bound structures (active state) as well ATP-complexed D. *radiodurans* RecA protein (*DrRecA*) structure. From the sequence similarity studies, it was identified that the *E.coli* RecA protein exhibits the highest sequence similarity (100%) with the *Shigella dysenteriae serotype 1 (strain Sd197), Shigella flexneri* and the lowest sequence similarity (46%) with the *Mycoplasma genitalium.* From the structural alignment & superposition studies, it can be inferred that the low hydrated MsRecA protein conformation is very closely similar to the active conformation of the EcRecA protein. This confers that the lower less water molecules favors the active conformational state of the RecA protein in both *M. smegmatis & E. coli* whereas the ATP-bound complex of RecA protein of *D. radiodurans* is exhibiting different conformation from the active state conformation.

Keywords: RecA Protein, Sequence Alignment, Structural superposition, Molegro Virtual Docker (MVD)

1. Introduction

1.1. RecA Protein: The RecA protein is a multifunctional protein that is essential to three distinct, but related biological processes: (a) genetic recombination; (b) SOS response; and (c) the error-prone replicative bypass of DNA lesions[1]. The central reaction in recombination is the exchange of strands between two homologous DNA molecules, catalyzed by the RecA family of ATPases[2]. Homologous genetic recombination plays an important role in the repair of DNA damage, generation of genetic diversity, maintenance of genomic integrity, and proper segregation of chromosomes[3]. Other than recombination process, Rec A protein is involved in SOS response and DNA lesion repairs during the replication process. For the active state of RecA protein, cofactor such as NTP is essential. The complex of ATP & RecA protein bound to the DNA specifically and catalyzes the RecA functionality. Hence the present study is to analyse the RecA protein sequence homology across various organisms with respect to *E. coli* and to understand the conformation of the protein in different states such as dehydrated state, ATP-complexed state and ss/dsDNA-bound state.

2. Methodology

2.1. RecA Protein Sequences Data Collection: RecA protein sequences were retrieved from the latest updates of the Swiss-Prot, which is manually annotated and reviewed. Swiss-Prot is one of the sections of UniProt Knowledgebase. The UniProt Knowledgebase (UniProtKB) is the central hub for collection of functional information on proteins, with accurate, consistent and rich annotation. By restricting the query RecA 'protein name' and only 'reviewed', 101 (UniProtKB/Swiss-Prot) protein sequences of various organisms are retrieved. An easy download option is available to retrieve the selected or required sequences in desired format. The sequences data in FASTA format are retrieved for further comparative sequence identity analysis with respect to *E. coli* RecA protein sequence.

2.2. Comparative Sequence Identity Analysis: The pairwise DNA-DNA or protein-protein sequence comparisons can be done using tool 'Align Sequences Protein BLAST'. This tool produces the alignment of two given sequences using BLAST algorithm for local alignment. A World Wide Web version of the program can be used interactively at the NCBI WWW site. The

tool provides various types of BLAST programs in which appropriate program (blastp) is selected. The default values are assigned for the parameters. The two sequences which are to be aligned must be in FASTA format. The results from the program provide information like alignment of Query and Subject, Score, Expect Value, Identities, Positives, and Gaps. The resulting alignments are presented in both graphical and text form.

2.3. Retrieval of PDB Structures of RecA Protein: The Protein Data Bank (PDB) is an archive of experimentally determined three-dimensional structures of biological macromolecules. The data contained in the archive include atomic coordinates, bibliographic citations, primary and secondary structure, information, and crystallographic structure factors and NMR experimental data. The required 3D structure can be retrieved by typing its PDB accession code or protein name in the search box. The atomic coordinates and structure factors which have been deposited in the RCSB PDB with the accession codes are retrieved **(Table 1).**

Table 1. Details of PDB Structures Retrieved

PDB ID	TITLE	SOURCE
2OFO	MSrecA-native	*Mycobacterium smegmatis*
2OE2	MSrecA-native-low humidity 95%	*Mycobacterium smegmatis*
1XP8	Deinococcus radiodurans RecA in complex with ATP-gamma-S	*Deinococcus radiodurans*
1N03	Model for Active RecA Filament	*Escherichia coli*
3CMU	Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures	*Escherichia coli*

2.4. Structure Alignment Using MVD: Molegro Virtual Docker (MVD) is an integrated environment for studying and predicting how ligands interact with macromolecules. The identification of ligand binding modes is done by iteratively evaluating a number of candidate solutions (ligand conformations) and estimating the energy of their interactions with the macromolecule. MVD requires a three-dimensional structure of both protein and ligand. MVD performs flexible ligand docking, so the optimal geometry of the ligand will be determined during the docking[11].

Identity Analysis: The details of comparative sequence identities of 100 RecA protein sequences with respect to *Escherichia coli* (strain K12) are given in the **table 2**. Among the one hundred sequences, the highest sequence similarity (100%) are exhibited by the *Shigella dysenteriae serotype 1 (strain Sd197), Shigella flexneri* and the lowest sequence similarity (46%) is exhibited by the *Mycoplasma genitalium.* For structure conformation analysis, we have chosen the *Mycobacterium smegmatis, Deinococcus radiodurans* respective RecA proteins as their crystal structures are solved which are exhibiting the sequence similarities of 65 & 61 percent identities respectively.

3. Results and Discussion

3.1. Comparative RecA Protein Sequence

Table 2. RecA protein sequence identities of various organisms with respect to *E.coli.*

S.no	Accession	Organism	Length	Identities (%)
1	P0A7G6	*Escherichia coli (strain K12)*	353	-
2	P47581	*Mycoplasma genitalium*	340	46
3	P29226	*Mycoplasma pulmonis*	339	53
4	P42443	*Deinococcus radiodurans*	363	61
5	Q5QJ16	*Mycobacterium smegmatis*	349	65
6	P65977	*Salmonella typhimurium*	353	97
7	P0A7G9	*Shigella flexneri*	353	100
8	Q32CM9	*Shigella dysenteriae serotype 1 (strain Sd197)*	353	100

3.2. Comparative RecA Protein Structure Alignment & Superposition Studies: By performing the structural alignment & superposition studies **(Figure 1)** of the PDB structures 2OF0, 3CMU, 3CMT, 3CMV, 1NO3, 1XP8 onto the 2OE2 (*Mycobacterium smegmatis-Low Hydrated structure*), the obtained Root Mean Square Deviation (RMSD) values & the number of amino acid residues matched in the structural alignment **(Figure 2a,**

b) are tabulated in the **table 3**. The lower the RMSD values, the higher the structural similarities / conformations of the structures[10]. In our study, it can be inferred that the low hydrated RecA protein conformation is very closely similar to the native/active conformation of the protein i.e., the RecA protein in complex with the ss/ds DNA structural conformation. This indicates that the lower hydration environment (less water molecules) favors the active conformational state of the RecA protein in both *M. smegmatis & E. coli* whereas the ATP-bound complex of RecA protein of *D. radiodurans* is exhibiting different conformation from the native/active state conformation.

Table 3. Root Mean Square Deviation values obtained in Structural Superposition Studies

PDB ID	SOURCE / TITLE	2OE2 MS LH	Matches
2OFO	*Mycobacterium smegmatis* RecA-native	2.35	326
3CMU A	RecA-ssDNA/dsDNA structures / *Escherichia coli*	5.61	24
3CMT A	RecA-ssDNA/dsDNA structures / *Escherichia coli*	5.65	24
3CMV A	RecA-ssDNA/dsDNA structures / *Escherichia coli*	5.69	24
1NO3	Model for Active RecA Filament / *Escherichia coli*	13.34	26
1XP8	*Deinococcus radiodurans* RecA-ATP-gamma-S complex	16.50	20

Figure 1. Executing the Secondary Structure alignment Application in MVD

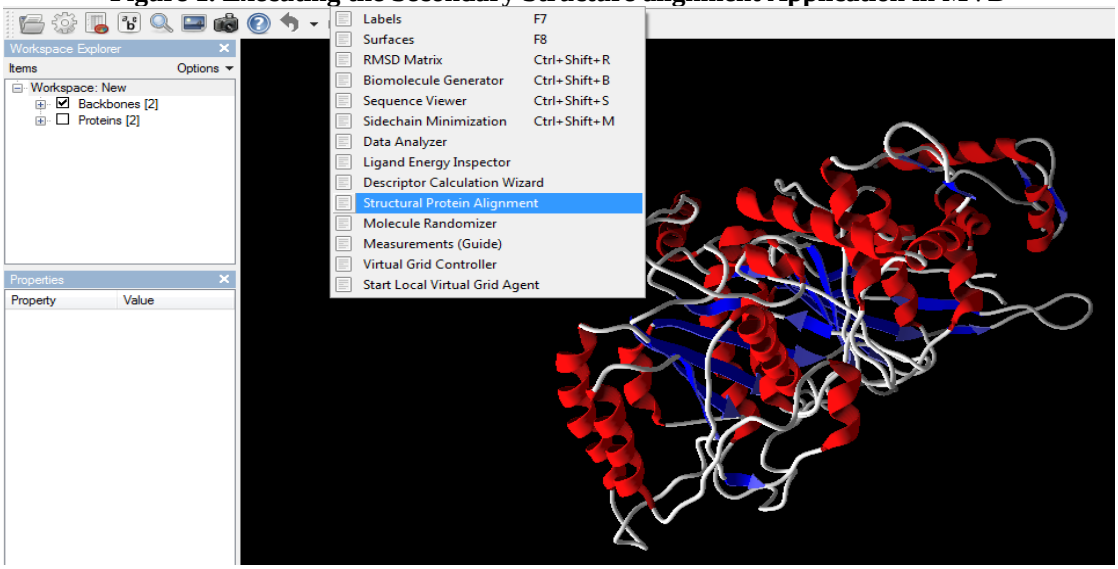


Figure 2a. Structure Alignment of the Proteins 1N03 with 20E2

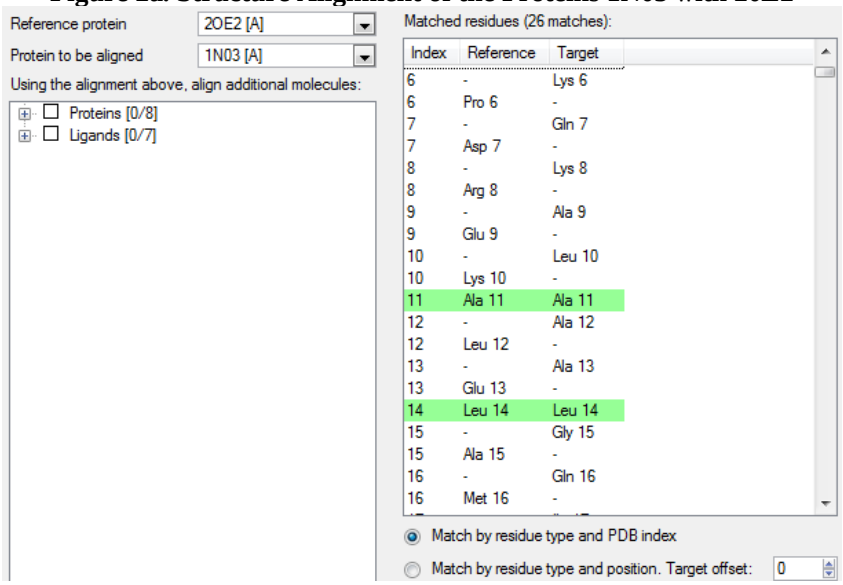
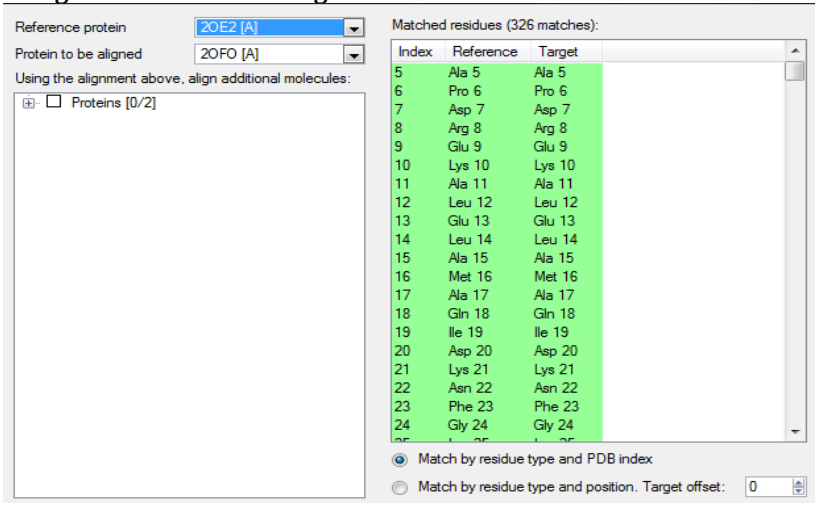


Figure 2b. Structure Alignment of the Proteins 2OFO with 20E2

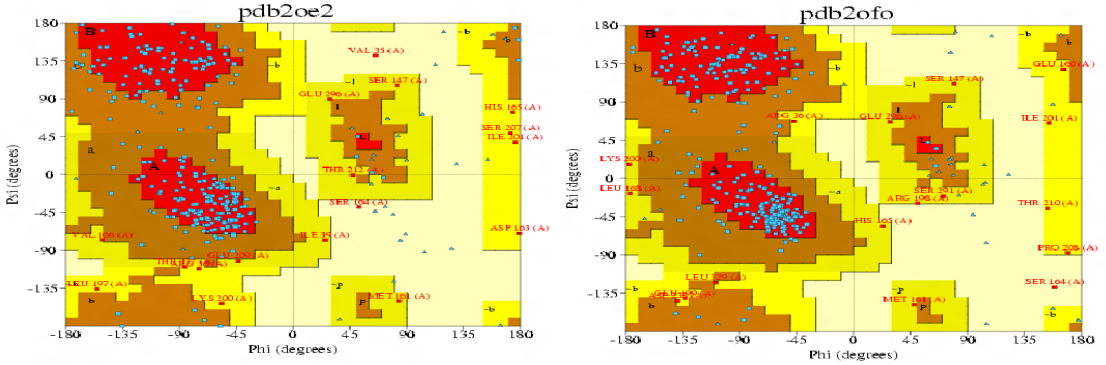


3.3. Ramachandran Plot Analysis: The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure (except those at the chain termini). Glycine residues are separately identified by triangles as these are not restricted to the regions of the plot appropriate to the other sidechain types[12]. The Ramachandran plots for the PDB structures: 2OE2 and 2OFO are generated by using PROCHECK. From the structural superposition studies, it can be inferred that the *Mycobacterium smegmatis* RecA-native (2OFO) structure conformation is very closely similar to the *Mycobacterium smegmatis* RecA-native-low humidity 95% (2OE2) structure conformation. From the Ramachandran Plot analysis it can be observed that the percentage of residues in most favoured regions are not varying in larger number and the existence of glycine & proline residues are similar in both the protein structures. The residues which are in disallowed regions are also very less in both the structures. The Procheck statistics are tabulated in **table 4**.

Table 4. Procheck Statistics

	No. of Residues		%-tage	
PDB IDs	**2OE2**	**2OFO**	**2OE2**	**2OFO**
Most favoured region [A,B,L]	193	218	69.4	78.1
Additional allowed regions [a,b,l,p]	68	45	24.5	16.1
Generously allowed regions [~a, ~b, ~l, ~p]	15	15	5.4	5.4
Disallowed regions [XX]	2	1	0.7	0.4
Non glycine and non proline residues	278	279	100	100
End residues (excl,gly & pro)	1	1	-	-
Glycine residues	36	37	-	-
Proline residues	11	11	-	-
Total number of residues	326	328	-	-

Fig 3. Ramachandran Plots of 2OE2 & 2OFO



4. Conclusion

The RecA protein of *E. coli* is compared with various organisms' RecA protein and comparative sequence identity analysis is performed. From the structural alignment & superposition studies, the RMSD values suggest that MsRecA dehydrated and EcRecA+DNA crystal structures are in active conformation. From the Ramachandran plot analysis, it can be generalized that the both native form and dehydrated form of RecA protein structures exhibits the similar conformations. Hence from the present study, it can be inferred that the dehydrated crystal structure mimics the native active state (action state) of the RecA protein.

References

1. Piero,R Bianco., Stephen C Kowalczykowski. RecA Protein. Advanced article, *Encyclopedia of LifeSciences* 2005
2. Zhucheng Chen, Haijuan Yang and Nikola P. Pavletich. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* 2008; (458) 489-484.
3. Datta,S., Krishna,R., Ganesh,N., Chandra,N.R., Muniyappa,K. and Vijayan,M. Crystal structures of *Mycobacterium smegmatis* RecA and its nucleotide complexes. *J. Bacteriol* 2003 (185), 4280–4284.
4. Krishna,R., Prabu, Manjunath,G.P., Kumar,P., N.R., Muniyappa,K. and Vijayan,M. Snapshots of RecA protein involving movement of the c-domain and different conformations of the DNA-binding loops : Crystallographic and comparative analysis of 11 structures of *Mycobacterium smegmatis* RecA. *Journal of molecular biology* 2007, vol. 367, 1130-1144.
5. Krishna,R., Prabu, Manjunath,G.P., Kumar,P., Surolia,A., N.R., Muniyappa,K. and Vijayan,M. Crystallographic identification of an ordered C-terminal domain and a second nucleotide-binding site in RecA: new insights into allostery. *Nucleic Acids Res.* 2006, 34, 2186–2195.
6. Datta,S., Prabu,M.M., Vaze,M.B., Ganesh,N., Chandra,N.R., Muniyappa,K. and Vijayan,M. Crystal structures of *Mycobacterium tuberculosis* RecA and its complex with ADP-AlF4: implications for decreased ATPase activity and molecular aggregation. *Nucleic Acids Res.* 2000, 28, 4964–4973.
7. Samuel Karlin and Luciano Brocchieri. Evolutionary conservation of RecA genes in relation to protein structure and function. *Journal of Bacteriology* 1996, Vol. 178, No.7, 1881-1894.
8. Samuel Karlin, George M. Weinstock, and Volker Brendel. Bacterial classifications derived from RecA protein sequence comparisons. *Journal of Bacteriology* 1995, Vol. 177, No.23, 6881-6893.
9. A. Krishna Chaitanya, I. Bhaskar Reddy and DSVGK. Kaladhar. Comparative docking studies on inhibitors of anti-apoptotic proteins. *IJPT* 2011, 3(1).1515-1519.
10. A. Krishna Chaitanya, I. Bhaskar Reddy, DSVGK. Kaladhar and GNV. Santhosh. Huntingtin protein modeling and structure alignment studies. *IJPBS* 2011, 2(1), 147-152.
11. http://www.molegro.com/mvd-product.php
12. http://www.ebi.ac.uk/pdbsum/