Malaysian Applied Biology (2024) 53(4): 185-198
https://doi.org/10.55230/mabjournal.v53i4.3078

Research

From Zebrafish To Humans: *In Silico* Comparative Study of RAD50 Sequences

Nahid Khalili[1], Ahmed Najm[1], Ibrahim Mahmood[2], Douglas Law[3], Chyan Leong Ng[4], Ahmad Azfaralariff[1], and Shazrul Fazry[1,5,6*]

1. Department of Food Science, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
2. Dentistry department, Al-Rafidain University College, Baghdad, Iraq
3. Faculty of Health and Life Sciences, Inti International University, 71800 Nilai, Negeri Sembilan
4. Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
5. Centre of Excellence, Innovation Centre for Confectionery Technology (MANIS), Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
6. Tasik Chini Research Center, The Centre for Natural and Physical Laboratory Management UKM, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia
*Corresponding author: shazrul@ukm.edu.my

ABSTRACT

DNA damage, particularly the occurrence of DNA double-strand breaks (DSBs), presents a significant hazard to the integrity and viability of cells. Improper repair of DSBs can result in chromosomal alterations, oncogenic changes, or cell demise. The MRE11-RAD50-NBS1 (MRN) complex plays a crucial role in DNA repair and signaling under the Ataxia Telangiectasia Mutated (ATM) kinase regulation. In this study, we employed comprehensive computational techniques to analyze the structure of RAD50 in *Danio rerio* (Zebrafish), utilized as a model organism. Additionally, we conducted in silico assessments of RAD50 from both Zebrafish and humans, comparing their characteristics. The substantial sequence resemblance between DrRAD50 and HsRAD50 suggests that DrRAD50 could potentially serve as a valuable model for HsRAD50. However, it is important to acknowledge that sequence similarity alone does not necessarily imply functional equivalence. Further functional studies are needed to confirm the extent of their functional similarities. By examining the secondary and tertiary protein structures of RAD50, we observed a notable likeness between Zebrafish and Human RAD50 proteins. In silico analysis demonstrated that the sequence of RAD50 in zebrafish shares 70% similarity with the human RAD50 protein.

Key words: DNA repair, *Danio rerio*, Human, *in-silico*, RAD50

Article History
Accepted: 26 August 2024
First version online: 27 October 2024

Cite This Article:
Khalili, N., Najm, A., Mahmood, I., Law, D., Ng, C.L., Azfaralariff, A., Fazry, S. 2024. From zebrafish to humans: *In silico* comparative study of RAD50 sequences. Malaysian Applied Biology, 53(4): 185-198. https://doi.org/10.55230/mabjournal.v53i4.3078

INTRODUCTION

When cells are exposed to DNA damage, they undergo various forms of damage, with DNA double-strand breaks (DSBs) posing the most significant risk to cell integrity and viability. DSBs, if not properly repaired, can lead to chromosomal rearrangements such as translocations and deletions, which may result in oncogenic transformation or cell death (Shibata *et al*., 2017). The repair mechanism involves inserting a broken DNA strand into a homologous DNA duplex and synthesizing copies (Paull & Lee, 2005). The MRE11-RAD50-NBS1 (MRN) protein complex is crucial in responding to DSBs, hairpins, and other abnormal terminal DNA structures, regulating DNA repair and signal transduction through the Ataxia Telangiectasia Mutated (ATM) kinase (Hopfner *et al*., 2000). RAD50 proteins, part of the structural maintenance of chromosome (SMC) superfamily of ATPases, including Condensins and cohesins in eukaryotes, play a significant role in this complex (Park *et al*., 2017). ATM phosphorylates RAD50 at a specific site (Ser-635), which is pivotal in signaling for cell cycle control and DNA repair. Mutations in this signaling site result in the inability to correct radiosensitivity, impair DNA double-strand

break repair, and cause an S-phase checkpoint defect (Gatei *et al*., 2011).

Hypomorphic mutations occurring in NBS1 and MRE11 lead to the autosomal recessive conditions known as Nijmegen breakage syndrome (NBS) and ataxia-telangiectasia-like disorder (ATLD). In a study by Waltes *et al*. (2009), they documented the case of a patient initially diagnosed with predominantly NBS, presenting with microcephaly, mental retardation, a "bird-like" facial appearance, and short stature. Upon further investigation, they identified RAD50 deficiency in this patient, resulting in a clinical phenotype resembling NBS, termed NBS-like disorder (NBSLD) (Waltes *et al*., 2009).

Previous research has explored the role of genes across diverse organisms. Eijpe *et al*. (2000) delved into RAD50's function concerning meiotic recombination in mouse testes, proposing its involvement in priming chromatin for recombination initiation during early meiotic prophase. In Genes and Development, Petrini *et al*. (2002) detailed the creation of hippomorphic RAD50 mutants in mice, resulting in partial embryonic lethality when RAD50 activity was impaired. Research by Attwooll *et al*. (2009) demonstrated the necessity of the MRN complex in mouse embryonic fibroblasts for responding to telomere dysfunction, suggesting shared mechanisms with the response to interstitial DNA damage. Rezaeejam *et al*. (2018) investigated the impact of melatonin as a radioprotective agent on RAD50 expression in rat peripheral blood, finding increased RAD50 levels following melatonin treatment before irradiation. Furthermore, the MRN complex's involvement in telomere capping during Drosophila embryogenesis has been established. Studies on RAD50 disruption in Drosophila revealed lethality in mutant flies, with morphological abnormalities evident in RAD50-deficient late pupae, including lack of abdominal segmentation, black spots, and epithelial tumors (Gorski *et al*., 2004; Ciapponi *et al*., 2004; Gao *et al*., 2009).

DrRAD50 refers to the RAD50 protein in zebrafish (*Danio rerio*), while HsRad50 refers to the RAD50 protein in humans (*Homo sapiens*). RAD50 is a crucial component of the MRN complex (MRE11-RAD50-NBS1), essential for DNA repair processes like homologous recombination and double-strand break repair (Park *et al*., 2017).

In this study, we infer the function and structural properties of DrRAD50 from HsRad50, despite the conventional approach of using model organisms to predict human protein behavior. First, RAD50 proteins are highly conserved across eukaryotic species, with critical functional domains and motifs preserved. Structural studies have demonstrated that the overall architecture of RAD50, including its zinc-hook and ATPase domains, is remarkably similar between humans and zebrafish. Additionally, functional assays have shown that human RAD50 can rescue RAD50-deficient phenotypes in zebrafish, indicating a strong functional conservation. These findings collectively support the feasibility of using HsRad50 as a model to infer the properties of DrRAD50.

Danio rerio (zebrafish) is an ideal model organism for studying vertebrate development due to its genetic and organ homology with humans, including bones. The optical clarity, small size, and rapid development of zebrafish embryos make them suitable for large-scale mutagenesis experiments to identify mutants with skeletal defects and for high-throughput screenings to discover compounds that can reverse pathological phenotypes (Carnovali *et al*., 2019). Adult zebrafish also serve as valuable models for studying adult human bone diseases. Unlike embryos, adults exhibit bone turnover, repair, and remodeling. Various pathological models, including bone injury, osteoporosis, and genetic disorders such as osteogenesis imperfecta, have been established in adult zebrafish. Recently, models for metabolic diseases like type 2 diabetes and obesity have been developed in adult zebrafish, revealing osteoporosis-like phenotypes associated with metabolic alterations (Carnovali *et al*., 2019).

Here, we employ an *in silico* algorithm to analyze the structural and functional similarity of the RAD50 protein between humans and zebrafish.

MATERIALS AND METHODS

In silico analysis

Source of sequences

The RAD50 genes of *Danio rerio* (Zebrafish) and *Homo sapiens* (human) with GenBank accession (ID: XP_005167995.1 and AAB07119.1) respectively, were obtained from GenBank of the National Centre of Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov).

Source of pathogenic mutation variants

The RAD50 pathogenic variants were assembled from (i) previous literature, (ii) a search using the ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and (iii) the single nucleotide polymorphism database (dbSNP) (https://www.ncbi.nlm.nih.gov/snp/).

Analysis of sequences

NCBI bioinformatics online tool was used to analyze the RAD50 gene of *Danio rerio* and *Homo sapiens*. The conserved domains were predicted with the CD search program (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). BLAST pairwise alignment tool was used for sequence alignment in the phylogenetic analysis of *Danio rerio* RAD50 and RAD50s from other species (Altschul *et al*., 1990). The phylogenetic tree was constructed using a Neighbour-Joining method (NJ) with the help of the MEGA7 software.

Prediction of secondary structure

The PSIPRED program (http://bioinf.cs.ucl.ac.uk/psipred/) was used for the secondary structure prediction of the RAD50 proteins. The SOPMA server (https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) (Geourjon & Deleage, 1995) was used to determine the conformational statistics about positional probabilities over the α-helices, β-strands, turns, random and coils inside the protein structure.

Prediction of tertiary structure

The I-TASSER (https://zhanggroup.org/I-TASSER/) was used to predict the 3D structure of the full-length model for the protein of RAD50 of *Danio rerio* and *Homo sapiens* (Yang & Zhang, 2015). This online framework developed 3D models by removing continuous fragments from threading alignments and then reassembling them using replica-exchanged Monte Carlo simulations.

Modification and validation of predicted structure model

The optimized 3D structure of the model was assessed by the software; Rampage (http://mordred.bioc. cam.ac.uk/~rapper/rampage.php) which generates a Ramachandran plot using data derived and determines the chemical system viewpoints that are next to major chain and edge chain (Lovell *et al*., 2003).

***In silico* analysis of mutated RAD50 protein**

To investigate the possible effect of edited DNA on protein structure, an in-silico analysis of the model structure obtained from I-Tasser was conducted. The predicted structures were viewed and edited using Chimera from the University of California San Francisco (UCSF). A single-point mutation on residue was generated and the side chain rotamer was optimized.

Statistical analysis

The study includes a rigorous statistical validation of the models used. This involves calculating confidence levels, *p*-values, and other relevant statistical measures to support the reliability of the *in silico* predictions. For each model, we performed [95% confidence intervals for the sequence similarity scores between DrRAD50 & HsRAD50 and *p*-value of [specific *p*-value], which is below the conventional threshold of 0.05], ensuring that our findings are not only based on sequence similarity but are also statistically robust. These statistical measures provide a quantifiable level of confidence in the potential functional equivalence of DrRAD50 and HsRAD50, further reinforcing the need for corroborative functional studies to confirm these predictions.

RESULTS AND DISCUSSION

In silico analysis of RAD50

In vitro mutagenesis and corresponding mutant protein expression, or functional tests and characterization, is a burdensome time, workload, and cost activity. For these factors, *in silico* research is an easy, quick, economical, and efficient way of improving our understanding of protein structure, protein-protein interactions, protein docking, and even how protein structure and function may be influenced by amino acid substitution. The 3D protein structure model helps the recognition of amino acid substitutions. In this section, we used extensive computational techniques to screen for structural analysis of RAD50 of *Danio rerio* (Zebrafish), which was used as a model organism in this study. Also, *in silico* analyses for Zebrafish and human RAD50 were carried out and compared. The high sequence similarity between *Danio rerio* RAD50 (DrRAD50) and *Homo sapiens* RAD50 (HsRAD50) suggests that DrRAD50 can serve as a useful model for HsRAD50 (Figure 5).

Clinical pathogen variant

A list of pathogen variants was compiled from the Clinvar databases at the National Center for

Biotechnology (NCBI). The predicted HsRAD50 secondary structure was used to determine the position of these clinvar mutant variants. As shown in Figure 1, it is possible to identify which part of the protein structure of the variants is located based on the mutant number. Based on the significance of each of these structural modes (α-helix, strand, and coiled), the significance of each mutant can be identified. To identify the significance of a mutant based on the location and context of the mutation within the protein sequence. Also determined of the structural context, whether the mutation occurs in an α-helix, strand (β-sheet), or coiled region. Analyze the impact on local structure mutations in α-helices or β-strands can disrupt hydrogen bonding patterns, leading to destabilization of these structures. Coiled regions are generally less ordered and more flexible, so mutations here might affect protein dynamics or interactions with other molecules. Identify any functional domains affected by the mutation. For example, if the mutation is within an active site, binding site, or a domain critical for protein-protein interactions, it is likely to have significant functional implications. (Table 1) shows the 11 ClinVar variants of Nijmegen Breakage Syndrome-like disease (NBSLD) previously reported and collected in the ClinVar database, including their positions, clinical significance, and sequences. In the recognition and repair of defective DNA, the MRE11, RAD50-NBS1 complex plays a key role. Through interaction with other key members of the DNA damage response, the MRN complex is involved in various DNA damage repair pathways. Mutations in any member of the complex can lead to hypersensitivity to genotoxic agents, which in turn increases an individual's susceptibility to cancer. Recent research pointed to the role of the MRN complex in tumorigenesis and cancer treatment and explored potential approaches for targeting this complex for cancer treatment (Katri Heikkinen *et al*., 2006; Petroni *et al*., 2023). The limited increased risk of cancer could be associated with a single RAD50 mutation (Wang *et al*., 2008; Walters *et al*., 2009; Lin *et al*., 2016; Thompson *et al*., 2016; Bian *et al*., 2019). In (2009), a patient with RAD50 deficiency was identified and described by Waltes *et al*. (2009) resulting in a first-time clinical phenotype that could be classified as NBS-like disease. (NBSLD) (Waltes *et al*., 2009). Germline mutations in RAD50 (R1093X & Y1313X) cause NBS-like disorder (Syed & Tainer, 2018). All observations demonstrate the relevance of genetic interaction research and functional analysis to further confirm findings from high-throughput screening mutation research.

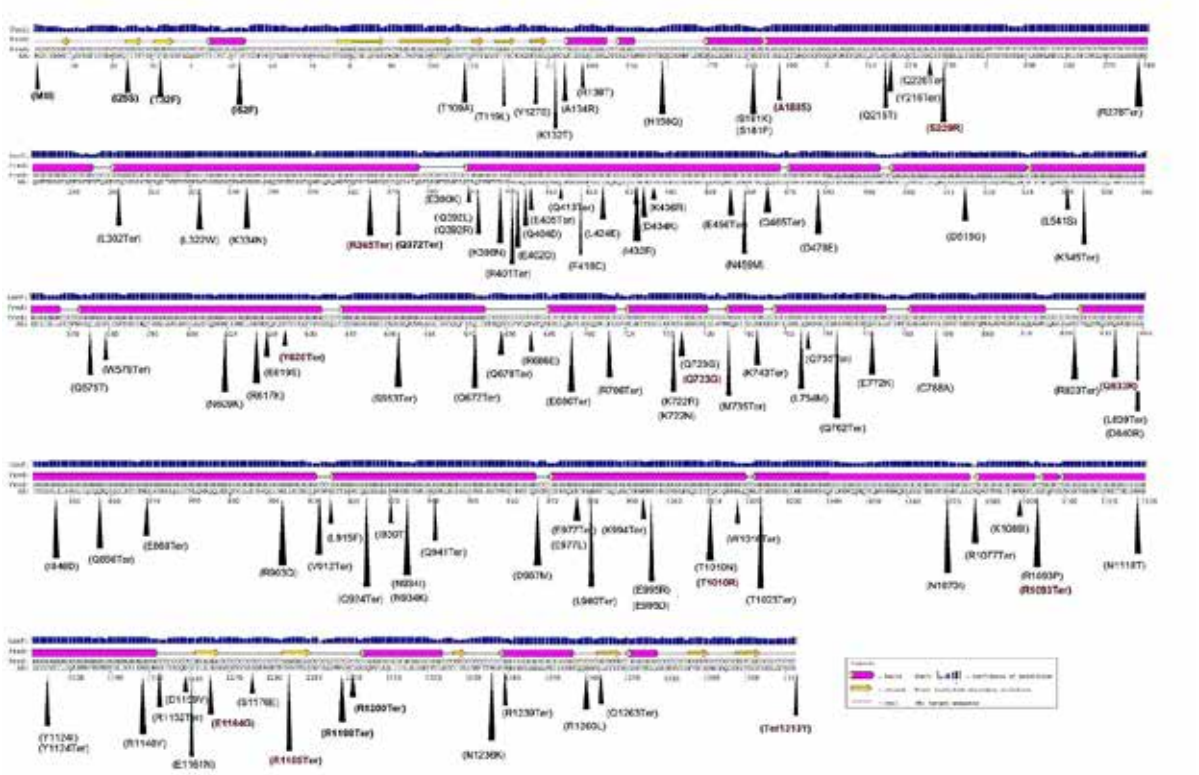

Fig. 1. Summary of pathogenic clinical pathogen variant on HsRAD50 protein secondary structure. The secondary structure was created and all 114 pathogenic variants of RAD50 were pinned to specific sites where mutation occurred.

The significance of human clinical pathogen variants to *Danio rerio* (zebrafish) lies in their shared genetic makeup with humans, making zebrafish a valuable model for studying human diseases. Variants in human clinical pathogens can be examined in zebrafish to understand their impact on analogous genes and pathways. Additionally, zebrafish are frequently used to model human diseases, including those stemming from genetic mutations. The transparency and rapid development of zebrafish embryos facilitate the observation of developmental processes. Investigating how human pathogen variants affect zebrafish development can shed light on the roles of these variants in human development and congenital diseases. Zebrafish can be genetically modified to express or suppress specific genes, enabling the functional analysis of human pathogen variants and clarifying their roles in cellular processes and disease pathways (Howe *et al*., 2013).

Primary structure analysis

The sequence of the RAD50 protein of Danio rerio was obtained from the NCBI database and used as a query sequence for homology modeling. The similarity search for the sequence was carried out using of BLAST tool. Then, the phylogenetic tree was provided to assess the phylogenetic position of zebrafish among other groups of organisms. The homologous, orthologous, and paralogous sequences of *Danio rerio* are revealed by the rooted structure as compared to others. As can be seen in Figure 2, *Danio rerio* belongs to the Cyprinidea family.

Table 1. Position, clinical significance, and SNP sequence of previously characterized RAD50 pathogenic (NBSLD) variants

No.	Variant	Exon	Domain	Clinical significance	SNP sequence
1	A188S	5	ABC-Rad5o	Likely pathogenic	ATTAAA[-/A]GCCTTA
2	S229R	5	AAA	Likely pathogenic	TACAAG[-/T]AAGGA
3	R365Ter	14	SMC	Likely pathogenic	-
4	Y625Ter	12	SMC	pathogenic	TCCAGTTA[C/G/T]GAAGACA
5	Q723G	13	Zink hook	pathogenic	AAAAAAA[-/A/T]GGAAAAG
6	Q833R	15	SMC	pathogenic	GAGAAAC[-/AA]GAGAAAC
7	T1010R	19	PRK	Likely pathogenic	ATATTGATA[-/CACA]GAAGGTAGG
8	R1093Ter	20	Mplasa	pathogenic	GAAAGAACTT[C/T]GAGAACCAC

A conserved domain search revealed that RAD50 contains three main domains (Figure 4). Two domains are ATP-binding cassette domain (ABC) which include 90 amino acid residues at the C-terminal (residues 1-317), and 102 amino acid residues at the N-terminal (residues 1222-11312). To explain the key role of RAD50 in DNA double-strand break repair (DSBR), Hopfner *et al*. have characterized the catalytic ATP-free RAD50 (RAD50cd) and domain ATP-bound RAD50 as biochemically and structurally essential for understanding the molecular mechanisms of DSBR. To predict the secondary structure of the RAD50 protein, we utilized both the PSIPRED and SOPMA programs. While each tool provides valuable insights individually, consensus-predicted features from multiple sources are generally more accurate. Therefore, we have integrated the predictions from both SOPMA and PSIPRED to enhance the reliability of our secondary structure predictions. They reported that RAD50cd crystal structures classify possible protein and DNA connectors, and show an ABC-ATPase loop connecting RAD50 molecular pathways to ABC transports, including P glycoprotein and transmembrane conductivity regulator for cystic fibrosis (Hopfner *et al*., 2000). RAD50 zinc hook is the other domain that contracts in 649-709 residues of sequence. For the recruitment of MRN to DSBs on chromatin, the RAD50 zinc hook domain is essential and it can also activate ATM protein in HR-mediated DSB repair in mammalian cells. (prosit. expasy.org & interpro< EMBL www.ebi.ac.uk) suggests that the RAD50 zinc hook domain in the RAD50 sequence of Zebrafish consists of 100 amino acids (635-734) (Barbi *et al*., 1991; Bally-Cuif *et al*., 2003; Barbelanne *et al*., 2014). Both Humans and Zebrafish share identical Zinc hook sequence of RAD50 and the zinc hook domain of RAD50 is essential for initiating responses of DNA damage and for mediating DSB repair (He *et al*. 2012).

Secondary structure prediction of RAD50 The secondary structure of the RAD50 protein was predicted by two systems: SOPMA (Self Configured Prediction System with Alignment) and PSIPRED. The findings of SOPMA are presented in Table 2. Compared to other secondary structure elements, these findings indicate a larger number of coils. (Alpha helix, Beta turn, & Random coil). The Pro-Motif analysis showed the modeled RAD50 structure, with 1312 amino acids, containing 7 beta-hairpins, 17 strands, 13 helices, 28 beta-turns and 1 gamma turn (Figure 5).

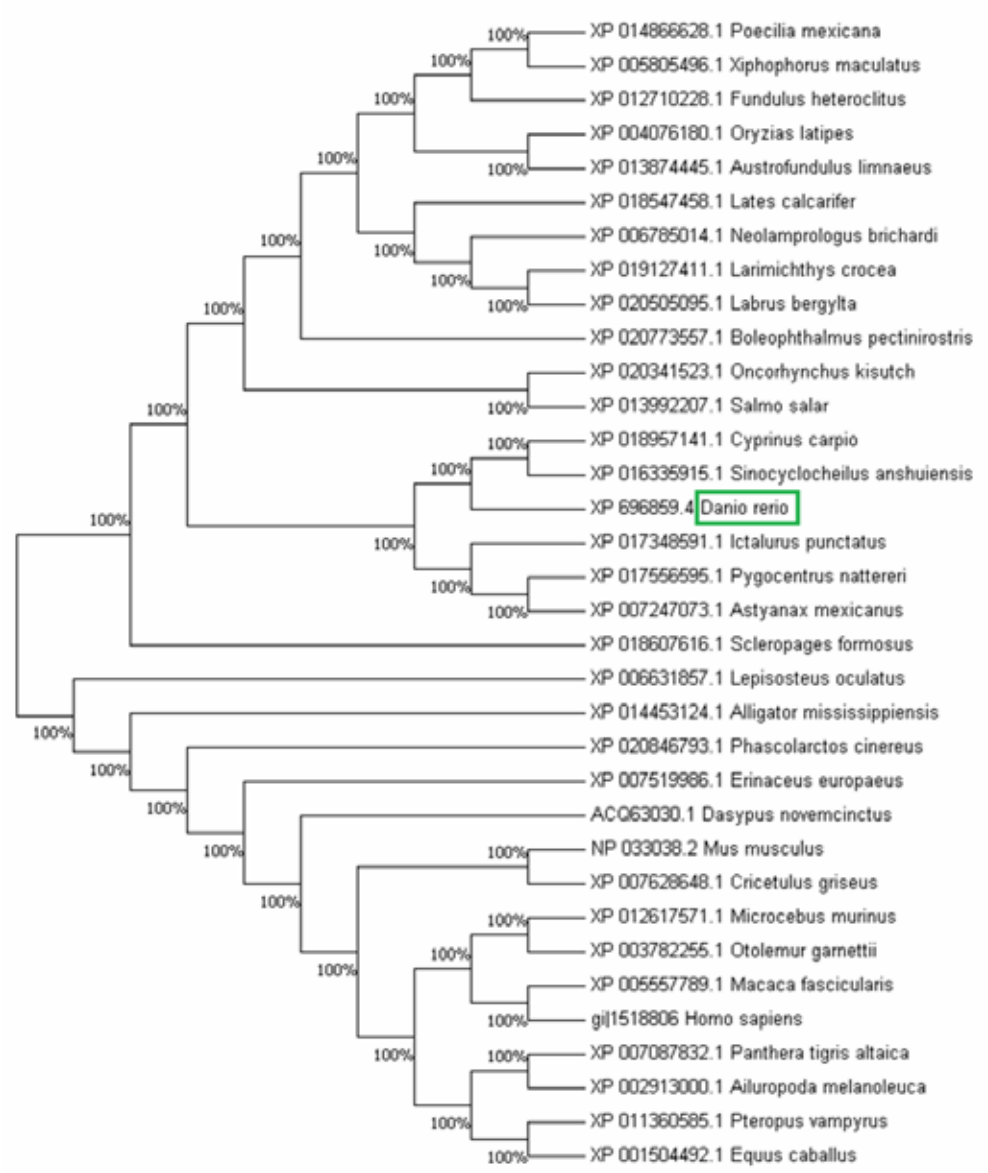

Fig. 2. Phylogenetic tree of 29 RAD50 genes from different groups of organisms. Bootstrap values, indicated at the nodes, were obtained from 1000 bootstrap replicates and are reported as percentages. The RAD50 of *Danio rerio* is marked in the green box.

Fig. 3. Pairwise sequence alignment of RAD50 protein sequence of human and zebrafish (1312 residues). 71.0% identity in 1312 residues overlap; Score: 4886.0; Gap frequency: 0.0%. Red residues show the residues related to Modification Sites in Orthologues. The phosphorylation Ser635 site is highlighted in yellow.

```
Seq.Human     721 KKEKRRDEMLGLVPMRQSIIDLKEKEIPELRNKLQNVNRDIQRLKNDIEEQETLLGTIMP
Seq.Danio     721 RKERRRDEMMTLKPIRQSLVELQEKELPELRNQLQRVNRDIEKLKGDIEEQETLLCTLVS
                   ** *****  * * ***   * *** ***** ** *****  ** ********* *

Seq.Human     781 EEESAKVCLTDVTIMERFQMELKDVERKIAQQAAKLQGIDLDRTVQQVNQEKQEKQHKLD
Seq.Danio     781 EEETAKACLQDISLMDRFQLDLKDVERKIAQHAAKLQGVDLSRTMQQVSQEKQETQHRLD
                  *** ** ** *   * ***  ********** ****** ** ** *** ***** ** **

Seq.Human     841 TVSSKIELNRKLIQDQQEQIQHLKSTTNELKSEKLQISTNLQRRQQLEEQTVELSTEVQS
Seq.Danio     841 TTSSKIELKRKLIQDQQEQIQALRSSVNEIRGEKLQISSNMQKRQQLEEQCVEFSTEIQT
                  * ****** ************ * *  **   ****** * * ******* ** *** *

Seq.Human     901 LYREIKDAKEQVSPLETTLEKFQQEKEELINKKNTSNKIAQDKLNDIKEKVKNIHGYMKD
Seq.Danio     901 LHRDIRDAKEQASPLAATLEKLQQEKQDLVERRRKKQEEGQEKINAIKEKMKNITLFEKE
                  * * * ***** ***  **** ****  *           * * * **** ***    *

Seq.Human     961 IENYIQDGKDDYKKQKETELNKVIAQLSECEKHKEKINEDMRLMRQDIDTQKIQERWLQD
Seq.Danio     961 ITKYIEEGKDSYKEQKETELQEMDKQLHEAEKQREKTNKDMGNIRQDIDTQKVQERWLQD
                  *  **  *** ** ******     ** * **  ** * **   ******** *******

Seq.Human    1021 NLTLRKRNEELKEVEEERKQHLKEMGQMQVLQMKSEHQKLEENIDNIKRNHNLALGRQKG
Seq.Danio    1021 NLTLRKRVEELKEVSRKREALIKEMGNMQVLQLRNERREVERKLEDLKKNRSVALGRQKG
                  ******* ******   *    **** *****   *    *      * *   *******

Seq.Human    1081 YEEEIIHFKKELREPQFRDAEEKYREMMIVMRTTELVNKDLDIYYKTLDQAIMKFHSMKM
Seq.Danio    1081 YEDEILRFRKELNEDQYCRAEDLYRDKMIVMRTTELANKDLDIYYKALDQTIMRFHSMKM
                  ** **  * *** * *   **  **  ********* ********* *** ** ******

Seq.Human    1141 EEINKIIRDLWRSTYRGQDIEYIEIRSDADENVSASDKRRNYNYRVVMLKGDTALDMRGR
Seq.Danio    1141 EEINKIIRDLWRSTYRGQDIEYVEIRSDVDENASAGVKRRTYNYRVVMVKGDAALDMRGR
                  ********************** ***** *** **  *** ******* *** *******

Seq.Human    1201 CSAGQKVLASLIIRLALAETFCLNCGIIALDEPTTNLDRENIESLAHALVEIIKSRSQQR
Seq.Danio    1201 CSAGQKVLASLIIRLALAETFCLNCGILALDEPTTNLDRENIESLAHALVEIIKSRSRQR
                  ************************** ***************************** **

Seq.Human    1261 NFQLLVITHDEDFVELLGRSEYVEKFYRIKKNIDQCSEIVKCSVSSLGFNVH
Seq.Danio    1261 NFQLLVITHDEDFVELLGRSNYVEHFYRIRKNQDQCSEISKCSINSLNSYLH
                  ******************** *** **** ** ****** ***  **    *
```

Fig. 3. (Continued). Pairwise sequence alignment of RAD50 protein sequence of human and zebrafish (1312 residues). 71.0% identity in 1312 residues overlap; Score: 4886.0; Gap frequency: 0.0%. Red residues show the residues related to Modification Sites in Orthologues. The phosphorylation Ser635 site is highlighted in yellow.

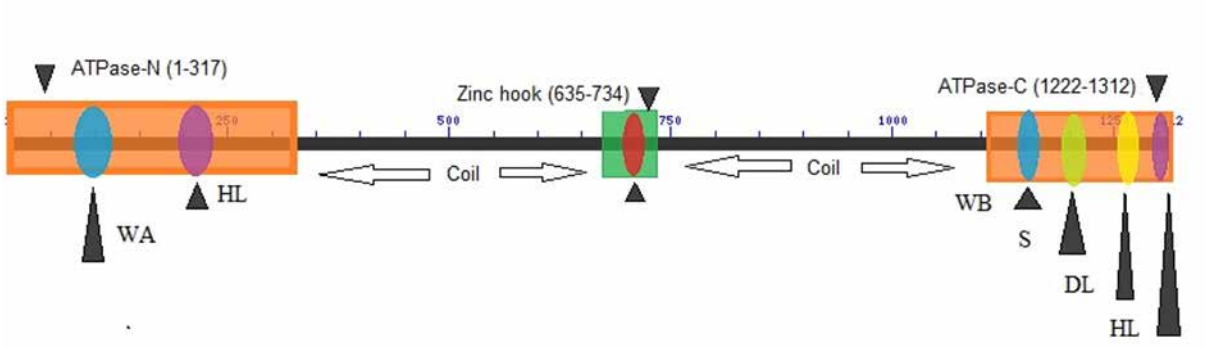

Fig. 4. Domain structure of DrRAD50. Sequence contains 1312 amino acids and domain analysis showed RAD50 of DrRAD50 and HsRAD50 contains three main domains which are ATP-binding cassette domain at each end of N and C terminal and RAD50 zinc hook domain at the middle of the sequences. They have seven motifs which are Walker A (WA) and Walker B (WB) motifs at either end of the protein, signature motif, Q loop (QL), D loop (DL), and H loop (HL).

Table 1. Calculated secondary structure elements by SOPMA

S.NO	parameters	Value (%) *Danio rerio*	Value (%) *Homo sapiens*
1.	Alpha helix	72.10%	72.56%
2.	310 helix	0.00%	0.00%
3.	Pi helix	0.00%	0.00%
4.	Beta bridge	0.00%	0.00%
5.	Extended strand	8.16%	8.38%
6.	Beta turn	4.57%	4.27%
7.	Bend region	0.00%	0.00%
8.	Random coil	15.17%	14.79%
9.	Ambiguous states	0.00%	0.00%
10.	Other states	0.00%	0.00%

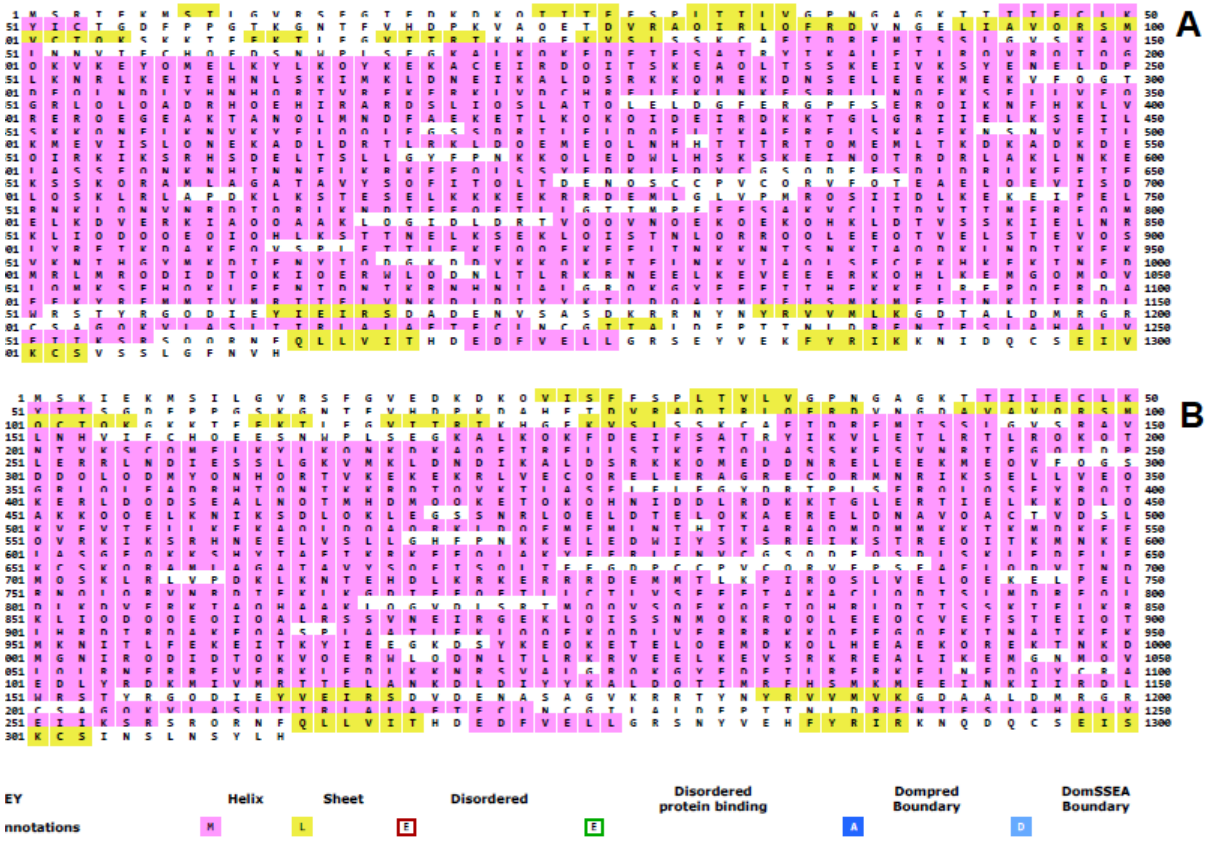

Fig. 5. Secondary Structure Map of RAD50, Feature predictions are color-coded onto the sequence according to the sequence feature key shown of RAD50 of A) *Homo sapiens* and B) *Danio rerio* by PSIPRED server.

3D structure modelling and structure quality

The tertiary structure of the whole RAD50 protein is not found available in the Protein Data Bank (PDB). Hence, homology modeling was performed to generate a 3D structure model of RAD50 protein; 3D models of RAD50 were generated for zebrafish and humans (Figure 6). The superimposed 3D protein structure was done for DrRAD50 and HsRAD50. UCSF Chimera prophesied superimposed model of HsRAD50 and DrRAD50, shows that both DrRAD50 and HsRad50 share high structure similarity.

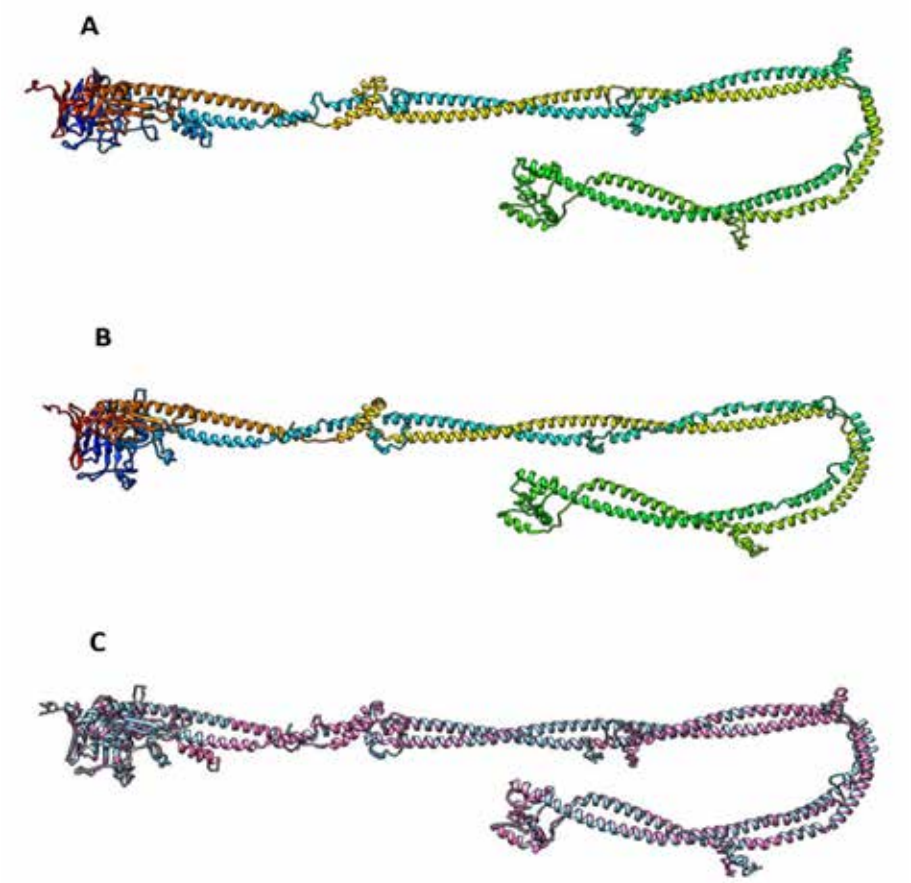

Fig. 6. 3D structure predicted by I-TASSER, based on the cluster size of structure along with its respective score > 1.5. A) *Homo sapiens* RAD50 and B) *Danio rerio* RAD50 C) Superimpose of created PDB of HsRAD50 and DrRAD50. HsRAD50 in pink and DrRAD50 in blue color have been specified. C-score > 1.5: This score range generally indicates a high confidence in the predicted structure, often correlating with a Root Mean Square Deviation (RMSD) of around 1-2 Ångströms (Å) from the native structure. C-score between -1.5 and 1.5: This range represents medium confidence. C-score < -1.5: Lower scores indicate lower confidence.

The Ramachandran plot is one of the local validation criteria, which is a standard part of structure analysis before structure deposition (Pražnikar *et al*. 2019). We used a rampage tool to determine the Psi/Phi Ramachandran plot to analyze the 3D model RAD50 protein quality, which was predicted by I-Tasser. The distributions of the Ramachandran plots of non-glycine and non-proline residues are summarized in Figure 7. Table 3 shows the summary results of Non-Proline and Non-Glycine residues for the model produced for DrRAD50 and HsRAD50. These results revealed that in 72.7% of *Danio rerio* and 74% of *Homo sapiens,* the residues were present in the most favored regions; 17.2% for *Danio rerio* and 16.9% of *Homo sapiens* in the additionally allowed regions; 10.2% for *Danio rerio* and 9.1% of *Homo sapiens* of the residues were seen in the disallowed regions in the constructed model of RAD50. However, the validation of the predicted 3D structure of RAD50 for *Danio rerio* and *Homo sapiens* by the Psi/Phi Ramachandran plot suggests that the structures are satisfactory, and comparing these results between *Danio rerio* and *Homo sapiens* were shown high similarity in RAD50 protein structure (Rojowska *et al*., 2014). The Ramachandran plot shows the statistical distribution of backbone dihedral angle combinations Phi and Psi (Azqueta *et al*., 2014 & Babb *et al*., 2004). The allowed regions of the Ramachandran plot indicate, in principle, which Phi / Psi angle values are appropriate for an amino acid, X. In fact, for structure validation, the distribution of Phi / Psi values found in a protein structure may be used (Ramakrishnan *et al*., 2007). For the dihedral angles, the Ramachandran plot visualizes energetically allowed and forbidden regions (Appel & Chitnis, 2002 & Álvarez *et al*., 2014). Many dihedral angles are found in the forbidden regions of the Ramachandran plot for low-quality homology models. These variations typically suggest structural issues (Wiltgen, 2018). The models need to be refined to reduce the number of residues in disallowed regions. This can be achieved by using advanced computational techniques and software that optimize the 3D structure based on stereochemical parameters. The current study's findings, with 10.2% and 9.1% residues in disallowed

regions for *Danio rerio* and *Homo sapiens* RAD50 proteins, highlight the need for further refinement and validation. Reducing the number of residues in disallowed regions is crucial for improving the accuracy and reliability of the predicted protein structures.

This section emphasizes the use of homology modeling to generate a 3D structure of the RAD50 protein. Models for both zebrafish (*Danio rerio*) and human RAD50 were created and compared, showing high structural similarity. The superimposed models of human and zebrafish RAD50 proteins confirm this similarity. Validation through Ramachandran plots indicates the structural models' quality, although further refinement is needed to reduce the number of residues in disallowed regions.

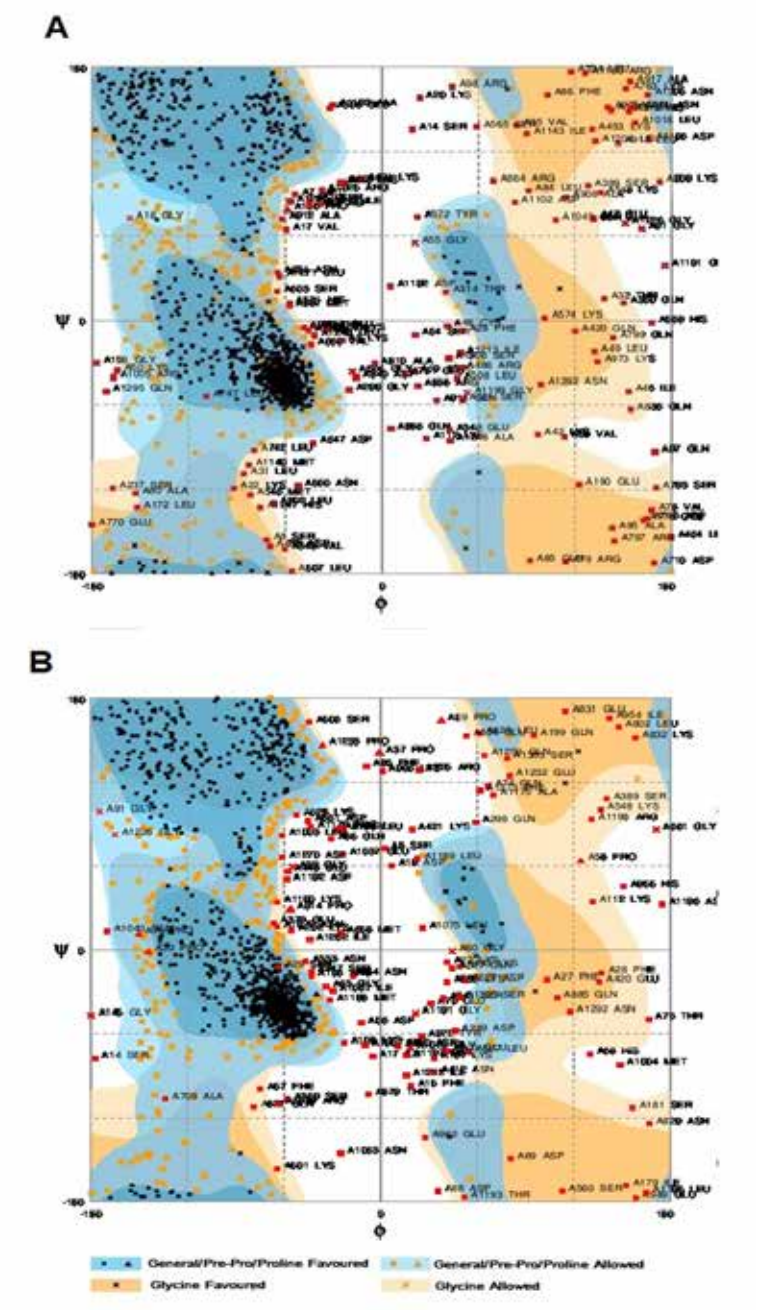

Fig. 7. Ramachandran plot showing 89.9% and 90.9% of the residues for (A) DrRAD50 and (B) HsRAD50 models present in the most favored or allowed region. 17.2% for *Danio rerio* and 16.9% of *Homo sapiens* in the additionally allowed regions; 10.2% for *Danio rerio* and 9.1% of *Homo sapiens* of the residues were seen in the disallowed regions in the constructed model of RAD50. A) *Danio rerio* B) *Homo sapiens*.

Table 2. Summary Results of Non-Proline and Non-Glycine Residues for Model Produced for RAD50 Protein in human and zebrafish

	Residues in most favoured regions		Residues in additional allowed regions		Residues in outlier-allowed regions	
	%	number	%	number	%	number
Danio rerio	72.7	952	17.2	225	10.2	133
Homo sapiens	74	970	16.9	221	9.1	119

Limitations of the study

1- Functional Divergence: While there is 70% sequence similarity between DrRAD50 and HsRAD50, the remaining 30% difference can lead to significant functional divergence. These differences may affect the protein's interactions with other molecules, its stability, and its activity within the cell. Such divergences could result in different functional outcomes despite the overall similarity.

2- Context-Specific Differences: Zebrafish and humans exhibit distinct physiological and cellular environments. These differences could influence how RAD50 functions in DNA repair and interacts with other components of the DNA damage response pathway. Therefore, the context-specific nuances might not be fully replicated in zebrafish.

3- Incomplete Representation of Human Disease: Zebrafish models might not entirely capture the complexity of human diseases, especially those influenced by genetic and environmental factors. This limitation can affect the model's ability to accurately predict human-specific disease outcomes and responses.

4- Evolutionary Distance: Zebrafish are more evolutionarily distant from humans compared to mammalian models like mice. This evolutionary gap may mean that certain aspects of RAD50's function and regulation are not conserved, potentially leading to misleading conclusions about its role in human biology.

CONCLUSION

In this research, computational analysis was conducted on the zebrafish RAD50 protein across three structural conformations. These analyses were juxtaposed with those of the human RAD50 protein. Upon scrutinizing the secondary and tertiary structures of the RAD50 protein and their analyses, a notable resemblance was observed between the zebrafish and human RAD50 proteins.

ACKNOWLEDGEMENTS

This work was supported by Ministry of Science, Technology and Innovation of Malaysia (MOSTI) (02-01-02-SF1279); Malaysian Ministry of Higher Education (MoHE) (FRGS/1/2023/STG01/UKM/02/5); Universiti Kebangsaan Malaysia (ST-2019-013; UKM-TR-018; GGPM-2014-006; and ST-2021-010); and Inti International University (INTI-FHLS-01-26-2023; and INTI-FHLS-01-03-2022).

ETHICAL STATEMENT

Not applicable

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abe, T., Ishiai, M., Hosono, Y., Yoshimura, A., Tada, S., Adachi, N., Koyama, H., Takata, M., Takeda, S., Enomoto, T. & Seki, M. 2008. KU70/80, DNA-PKcs, and Artemis are essential for the rapid induction of apoptosis after massive DSB formation. Cellular Signalling, 20(11): 1978-1985. https://doi.org/10.1016/j.cellsig.2008.07.006

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology, 215(3): 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Álvarez-Quilón, A., Serrano-Benítez, A., Lieberman, J.A., Quintero, C., Sánchez-Gutiérrez, D., Escudero,

L.M. & Cortés-Ledesma, F. 2014. ATM specifically mediates the repair of double-strand breaks with blocked DNA ends. Nature Communications, 5: 3347. https://doi.org/10.1038/ncomms4347

Appel, B. & Chitnis, A. 2002. Neurogenesis and specification of neuronal identity. Results and Problems in Cell Differentiation, 40: 237-251. https://doi.org/10.1007/978-3-540-46041-1_12

Attwooll, C.L., Akpinar, M. & Petrini, J.H. 2009. The mre11 complex and the response to dysfunctional telomeres. Molecular Cell Biology, 20: 5540-51. https://doi.org/10.1128/MCB.00479-09

Azqueta, A., Slyskova, J., Langie, S.A.S., Neill Gaiv, I. & Collins, A. 2014. Comet assay to measure DNA repair: approach and applications. Frontiers in Genetics, 5: 288. https://doi.org/10.3389/fgene.2014.00288

Babb, S.G. & Marrs, J.A. 2004. E-cadherin regulates cell movements and tissue formation in early zebrafish embryos. Developmental Dynamics, 230(2): 263-277. https://doi.org/10.1002/dvdy.20057

Bally-Cuif, L. & Hammerschmidt, M. 2003, February 1. Induction and patterning of neuronal development, and its connection to cell cycle control. Current Opinion in Neurobiology, 13(1): 16-25. https://doi.org/10.1016/S0959-4388(03)00015-1

Barbelanne, M. & Tsang, W.Y. 2014. Molecular and Cellular Basis of Autosomal Recessive Primary Microcephaly. BioMed Research International, 2014: 547986. https://doi.org/10.1155/2014/547986

Barbi, G., Scheres, J.M.J. C., Schindler, D., Taalman, R.D.F. M., Rodens, K., Mehnert, K. & Muller, M. 1991. Chromosome instability and X-ray hypersensitivity in a microcephalic and growth-retarded child. American Journal of Medical Genetics, 40(1): 44-50. https://doi.org/10.1002/ajmg.1320400109

Bian, L., Meng, Y., Zhang, M. & Li, D. 2019. MRE11-RAD50-NBS1 complex alterations and DNA damage response: Implications for cancer treatment. Molecular Cancer, 18: 169. https://doi.org/10.1186/s12943-019-1100-5

Carnovali, M., Banfi, G. & Mariotti, M. 2019. Zebrafish models of human skeletal disorders: Embryo and adult swimming together. Biomed Research International, 2019: 1253710. https://doi.org/10.1155/2019/1253710

Ciapponi, L., Cenci, G., Ducau, J., Flores, C., Johnson-Schlitz, D., Gorski, M.M. & Gatti, M. 2004. The Drosophila Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. Current Biology, 14(15): 1360-1366. https://doi.org/10.1016/j.cub.2004.07.019

Eisen, J.S. & Smith, J.C. 2008. Controlling morpholino experiments: Don't stop making antisense. Development, 135(10): 1735-1743. https://doi.org/10.1242/dev.001115

Gao, G., Bi, X., Chen, J., Srikanta, D. & Rong, Y.S. 2009. Mre11-Rad50-Nbs complex is required to cap telomeres during Drosophila embryogenesis. Proceedings of the National Academy of Sciences of the United States of America, 106(26): 10728-10733. https://doi.org/10.1073/pnas.0902707106

Gatei, M., Jakob, B., Chen, P., Kijas, A.W., Becherel, O.J., Gueven, N., Birrell, G., Lee, J.H., Paull, T.T., Lerenthal, Y., Fazry, S., Taucher-Scholz, G., Kalb, R., Schindler, D., Waltes, R. & Dörk, T. 2011. ATM protein-dependent phosphorylation of Rad50 protein Regulates DNA repair and cell cycle control. Journal of Biological Chemistry, 286(36): 31542-31556. https://doi.org/10.1074/jbc.M111.258152

Gatei, M., Jakob, B., Chen, P., Kijas, A.W., Becherel, O.J., Gueven, N., Birrell, G., Lee, J.H., Paull, T.T., Lerenthal, Y., Fazry, S., Taucher-Scholz, G., Kalb, R., Schindler, D., Waltes, R. & Dörk, T. 2011. ATM protein-dependent phosphorylation of Rad50 protein regulates DNA repair and cell cycle control. Journal of Biological Chemistry, 286(36): 31542-31556. https://doi.org/10.1074/jbc.M111.258152

Geourjon, C. & Deléage, G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics, 11(6):681-684. https://doi.org/10.1093/bioinformatics/11.6.681

Gorski, J.C., Huang, S.M., Pinto, A., Hamman, M.A., Hilligoss, J.K., Zaheer, N.A., Desai, M., Miller, M. & Hall, S.D. 2004. The effect of echinacea (*Echinacea purpurea* root) on cytochrome P450 activity in vivo. Clinical Pharmacology & Therapies, 75(1): 89-100. https://doi.org/10.1016/j.clpt.2003.09.013

He, J., Shi, L.Z., Truong, L.N., Lu, C.S., Razavian, N., Li, Y. & Negrete, A. 2012. RAD50 zinc hook is important for the Mre11 complex to bind chromosomal DNA double-stranded breaks and initiate various DNA damage responses. Journal of Biological Chemistry, 287(38): 31747-31756. https://doi.org/10.1074/jbc.M112.384750

Heikkinen, K., Rapakko, K., Karppinen, S.M., Erkko, H., Knuutila, S., Lundán, T., Mannermaa, A., Børresen-Dale, A.L., Borg, Å, Barkardottir, R.B., Petrini, J. & Winqvist, R. 2006. *RAD50* and *NBS1* are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis, 27(8): 1593-1599. https://doi.org/10.1093/carcin/bgi360

Hopfner, K.P., Karcher, A., Shin, D.S., Craig, L., Arthur, L.M., Carney, J.P. & Tainer, J.A. 2000. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell, 101(7): 789-800. https://doi.org/10.1016/S0092-8674(00)80890-9

Howe, K., Clark, M.D. & Torroja, C.F. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature, 496(7446): 498-503. https://doi.org/10.1038/nature12111

Kämäräinen, M., Heiskala, K., Knuutila, S., Heiskala, M., Winqvist, O. & Andersson, L.C. 2003. RELP is a novel human REG-like protein with up-regulated expression in inflammatory and metaplastic gastrointestinal mucosa. The American Journal of Pathology, 163(1): 11-20. https://doi.org/10.1016/S0002-9440(10)63625-5

Li, Y., Wang, J., Zhou, G., Lajeunesse, M., Le, N., Stawicki, B.N. & Lopez Corcino, Y. 2017. Nonhomologous end-joining with minimal sequence loss is promoted by the Mre11-Rad50-Nbs1-Ctp1 Complex in Schizosaccharomyces pombe. Genetics, 206: 481-496. https://doi.org/10.1534/genetics.117.200972

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G. & Richardson, J. 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics, 50(3): 437-450. https://doi.org/10.1002/prot.10286

Park, Y.B., Hohl, M., Padjasek, M., Jeong, E., Jin, K.S., Krężel, A., Petrini & J.H.J. 2017. Eukaryotic Rad50 functions as a rod-shaped dimer. Nature Structural & Molecular Biology, 24(3): 248-257. https://doi.org/10.1038/nsmb.3369

Paull, T.T. & Lee, J.-H. 2005. The Mre11/Rad50/Nbs1 complex and its role as a DNA double-strand break sensor for ATM. Cell Cycle, 4(6): 737-740. https://doi.org/10.4161/cc.4.6.1715

Petroni, M., La Monica, V., Fabretti, F., Augusto, M., Battaglini, D., Polonara, F., Di Giulio, S. & Giannini, G. 2023. The Multiple Faces of the MRN Complex: Roles in Medulloblastoma and Beyond. Cancers,15(14): 3599. https://doi.org/10.3390/cancers15143599

Petroni, M., Sardina, F., Infante, P., Bartolazzi, A., Locatelli, E., Fabretti, F. & Di Giulio, S. 2018. MRE11 inhibition highlights a replication stress-dependent vulnerability of MYCN-driven tumours. Cell Death & Disease, 9(9): 895. https://doi.org/10.1038/s41419-018-0924-z

Pražnikar, J., Tomić, M. & Turk, D. 2019. Validation and quality assessment of macromolecular structures using complex network analysis. Scientific Reports, 9: 1678. https://doi.org/10.1038/s41598-019-38658-9

Ramakrishnan, K., Sparks, R.A. & Berryhill, W.E. 2007. Diagnosis and treatment of otitis media. American Family Physician, 76(11): 1650-1658.

Rezaeejam, H., Shirazi, A., Izadi, P., Bazzaz, J.T., Ghazi-Khansari, M. & Valizadeh, M. 2018. Radioprotective effect of melatonin on expression of *Cdkn1a* and *Rad50* genes in rat peripheral blood.

Rezaie, A., Parker, R.D. & Abdollahi, M. 2007. Oxidative stress and pathogenesis of inflammatory bowel disease: an epiphenomenon or the cause?. Digestive Diseases and Sciences, 52: 2015-2021. https://doi.org/10.1007/s10620-006-9622-2

Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M.M. & Maity, R. 2014. DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. Molecular Cell, 53(1): 7-18. https://doi.org/10.1016/j.molcel.2013.11.003

Singh, N., Basnet, H., Wiltshire, T.D., Mohammad, D.H., Thompson, J.R., Héroux, A. & Botuyan, M.V. 2012. Dual recognition of phosphoserine and phosphotyrosine in histone variant H2A.X by DNA damage response protein MCPH1. Proceedings of the National Academy of Sciences of the United States of America, 109(36): 14381-14386. https://doi.org/10.1073/pnas.1212366109

Syed, A. & Tainer, J.A. 2018. The MRE11-RAD50-NBS1 complex conducts the orchestration of damage signaling and outcomes to stress in DNA replication and repair. Annual Review of Biochemistry, 87: 263-294. https://doi.org/10.1146/annurev-biochem-062917-012415

Waltes, R., Kalb, R., Gatei, M., Kijas, A.W., Stumm, M., Sobeck, A. & Wieland, B. 2009. Human RAD50 deficiency in a nijmegen breakage syndrome-like disorder. The American Journal of Human Genetics, 84(5): 605-616. https://doi.org/10.1016/j.ajhg.2009.04.010

Wang, Y., Xu, Z., Bach, S.J. & McAllister, T.A. 2008. Effects of phlorotannins from *Ascophyllum nodosum* (brown seaweed) on *in vitro* ruminal digestion of mixed forage or barley grain. Animal Feed Science and Technology, 145 (1-4): 375-395. https://doi.org/10.1016/j.anifeedsci.2007.03.013

Wiltgen, M. 2018. Algorithms for structure comparison and analysis: Homology modelling of proteins. Encyclopedia of Bioinformatics and Computational Biology, 1: 38-61. https://doi.org/10.1016/B978-0-12-809633-8.20484-6

Yang, J. & Zhang, Y. 2015. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Research, 43(W1): W174-W181. https://doi.org/10.1093/nar/gkv342