

Decoding the vital segments in human ATP-dependent RNA helicase

Vandana Kamjula, Ananya Kanneganti, Rohan Metla, Kusuma Nidamanuri, Sudarshan Idupulapati, Ashish Runthala*

Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India; Email: ashish.runthala@gmail.com.

*Corresponding Author

Vandana Kamjula - kamjulavandana7@gmail.com, Ananya Kanneganti - ananyakanneganti1998@gmail.com, Rohan Metla - metla.rohan420@gmail.com, Kusuma Nidamanuri - nidamanurikusuma@gmail.com, Sudarshan Idupulapati - sudarshanidupulapati1998@gmail.com, Ashish Runthala - ashish.runthala@gmail.com

Received January 22, 2020; Accepted February 20, 2020; Published February 29, 2020

DOI: 10.6026/97320630016160

Declaration on official E-mail:

The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:

The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Abstract:

An analysis of the ATP-dependent RNA helicase using known functionally close analogs help disclose the structural and functional information of the enzyme. The enzyme plays several interlinked biological functions and there is an urgent need to interpret its key active-site residues to infer function and establish role. The human protein q96c10.1 is annotated using tools such as interpro, go and cdd. The physicochemical properties are estimated using the tool protparam. We describe the enzyme protein model developed using modeller to identify active site residues. We used consurf to estimate the structural conservation and its evolutionary relationship is inferred using known close sequence homologs. The active site is predicted using castp and its topological flexibility is estimated through cabs-flex. The protein is annotated as a hydrolase using available data and ddx58 is found as its top-ranked interacting protein partner. We show that about 124 residues are found to be highly conserved among 259 homologs, clustered in 7 clades with the active-site showing low sequence conservation. It is further shown that only 9 loci among the 42 active-site residues are conserved with limited structural fluctuation from the wild type structure. Thus, we document various useful information linked to function, sequence similarity and phylogeny of the enzyme for annotation as potential helicase as designated by uniprot. Data shows limited degree of conserved sequence segments with topological flexibility unlike in other subfamily members of the protein.

Keywords: RNA helicase, innate immunity, motif, MODELLER, flexibility.

Availability: The constructed files/datasets analyzed in this study are available from the corresponding author on reasonable request.

Background:

RNA helicase is ubiquitously present in viruses, bacteria, archaea and eukaryotes, and is the largest cluster of enzymes linked with RNA metabolism [1]. Being a highly conserved enzyme, it plays a phenomenal role in the unwinding of the RNA duplexes [2] and requires the hydrolysis of nucleoside triphosphates [3]. The DEAD-box protein (DHX) family members are usually located in the nucleus region. The laboratory protein of genetics and physiology2 (LGP2) is a member of the DEAD-box protein family and belongs to the ATP-dependent RNA helicase family [4, 5], known to be involved in various steps of RNA metabolism [6] with several pleiotropic functions [7]. The catalytic core of these proteins encodes 12 highly conserved motifs [8]. LGP2 is a key regulator of interferon-induced with helicase-C domain1 (IFIH1)/ melanoma differentiation associated protein5 (MDA5) and DExD/H-Box helicase58 (DDX58)/retinoic acid-inducible gene (RIG-I)-mediated antiviral [9,10]. When the antiviral pathway gets perturbed, RIG-I usually initiates a cascade of deregulated events, which further causes the immunological disorders [11]. It shows a significant response against several viruses including newcastle disease, rhabdovirus, sendai, lassa, orthomyxoviruses (influenza), ebola and flaviviruses (hepatitis). While it acts both against the single or double-stranded RNA, MDA-5 is active against the long double-stranded RNA and recognizes picornaviruses and vaccinia viruses. Both these proteins are shown to have an active response against dengue, West Nile and Japanese encephalitis viruses [12]. Thus, helicases play key roles in regulating the innate immune responses [13]. Active research is going on RNA helicase, and enormous articles have been published to date (September 14th, 2019) [14, 15, 16, 17, 18, 19, 20, 21, 22]. The two databases national center for biotechnology information proteins (NCBI) and universal protein resource knowledgebase (UniProtKB) orderly contain 1335 and 422 sequences in contrast to 154 structures listed in the protein data bank (PDB). The ever-increasing sequence-structure gap for this protein makes its sequence, structure, conservation or phylogeny analysis quite elusive for the evolutionarily distinct human sequence variants. For its key role behind the regulation and control of gene regulation and RNA metabolism, there are growing implications for DHX subfamily in human diseases and their treatment [23, 24]. It is of interest to report an analysis of the ATP-dependent RNA helicase using known functionally close analogs to help disclose the structural and functional information of the enzyme.

Materials and Methods:

For functionally characterizing the un-annotated human protein sequence, the following strategy is developed, as depicted through a flowchart in (Figure 1).

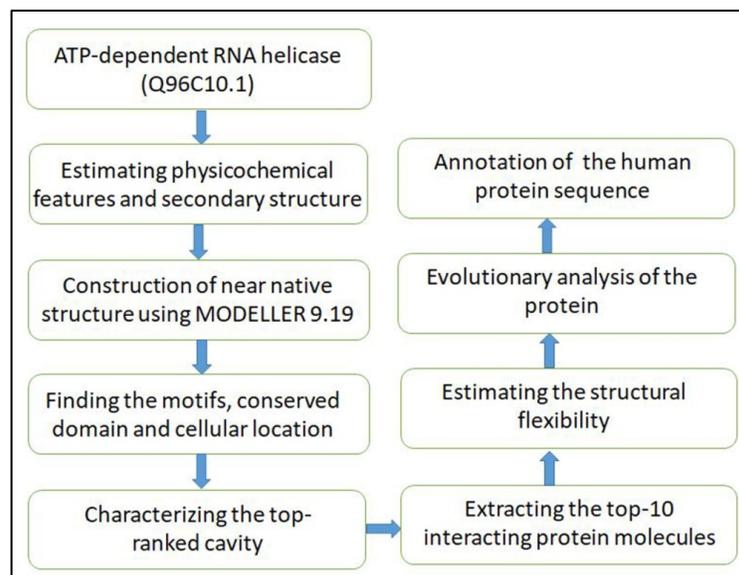


Figure 1: Flowchart showing the robust annotation algorithm for the human protein sequence. To ascertain the predictions, the methodology deploys the key sequence, structural and evolutionary measures.

Sequence retrieval:

The amino acid sequence of ATP-dependent RNA helicase (Q96C10.1) is retrieved from the UniProtKB/SwissProt database.

Prediction of physicochemical properties:

Several features, viz. residue composition, molecular weight, theoretical PI, instability index, extinction coefficient, atomic composition, aliphatic index, and grand average of hydropathicity (GRAVY) score are essential to define the physicochemical properties and to estimate the structural features of a protein sequence. The parameters are estimated through the ExPasy-Protparam tool (<https://web.expasy.org/protparam/>) [25].

Secondary data prediction:

PSIPRED algorithm is deployed to predict the three-state secondary structure (Helix, strand, and coil). It provides credible information corresponding to α -helices, beta-sheets, coils, transmembrane helices, signal peptides, membrane interactions, re-entrant helix, and putative domain boundaries [26].

Molecular modelling:

To construct a near-native structure of the RNA helicase sequence, HHPred [27] is used to screen the top-ranked functionally similar protein structure(s) (templates) from the PDB database by extending the sequence profile on the basis of 5 iterative rounds [28, 29]. The template 5F9FE, sharing the highest sequence similarity of 53%, is selected and the protein model is built using MODELLER9.19 [30]. The unaligned 1-residue N-terminal and 12-residue C-terminal (667-678) segments are truncated to curate the alignment file and to construct the 2-666 residue model structure. As the predicted decoy is found to encode several atomic clashes, it is energetically relaxed/refined through 3Drefine [31], and the best model is selected on the basis of qualitative model energy analysis (QMEAN) and ERRAT scores. The model is assessed through the discrete optimized potential energy (DOPE) and GA341 scores of MODELLER. By using the QMEAN server, the predicted top-ranked model is assessed through the Molprobity score on the basis of rotamer outliers and the atomic clash score. Ramachandran map is subsequently plotted through the PROCHECK server to assess the topological accuracy of the predicted structure on the basis of phi and psi angles.

Functional scrutiny:

The sequence is fed to InterPro server to retrieve the information regarding the superfamily, domains, repeats and gene ontology [32]. Conserved domain database (CDD) is subsequently screened to affirm the credibility of the screened domains for purging the spurious hits/superfamilies and selecting the credible ones [33]. To estimate the interaction of the selected protein with the closely-related sequences, the STRING database is used [34]. For a robustly accurate analysis, its algorithm deploys several parameters including gene fusion, gene neighborhood, gene co-occurrence, text mining, and co-expression to estimate a confidence score. The score ranges between 0 and 1 and, for all the considered features, it is expected to remarkably score the closely interacting protein pairs. To localize the three most-conserved motifs, the MEME suite is used [35]. On the basis of gapless local alignment of multiple sequences (GLAM2) protocol, it even covers the gapped motifs [36]. The algorithm helps to identify DNA and protein sequence motifs.

The default motif length range of 6-50 is used for the analysis. PROFUNC (<http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/>) is further used to estimate the biochemical functions through the sequence homology against the PDB database [37]. To reliably affirm the intracellular/cytosolic locus of the human helicase protein, the hidden markov models-dependent server (TMHMM2.0 www.cbs.dtu.dk/services/TMHMM) is used [38]. Peptide cutter, a web-based tool, (https://web.expasy.org/peptide_cutter/) is subsequently used to predict the location of probable cleavage sites of chemicals/proteolytic enzymes.

Conservation and flexibility analysis:

To reliably affirm the sequence conservation profile of the sequence, UniProtKB/SwissPROT database is screened through HMMER for the selected protein [39]. With a very strict E-value inclusion cutoff of 0.00001, the sequence profile is expanded through five iterative rounds. From a total of 728 ATP-dependent RNA helicases, 259 sequences are selected. As the sequence length of experimentally solved protein structures is found to be within 600-800, the sequence length filter (580-820 residues) is conservatively used along with the removal of bifunctional proteins. Sequences are retrieved using Batch-Entrez and aligned by ClustalW module of HHpred. ConSurf is subsequently used to track the degree of conservation across the chain [40]. Deploying the constructed sequence profile, the conservation scores are statistically estimated with the Bayesian probability across the chain on a scale of 1-9. To define the functional conservation across the chain, it takes input from sequence alignment and draws phylogeny connections among the sequences to plot it over the deployed/predicted reference structure through color gradations. Surface topography is further analyzed by computed atlas of surface topography of proteins (CASTp) to locate the active-site within the modelled protein structure [41]. It locates pockets, internal cavities, and the cross channels along with their surface area and volume, and reveals the functionally important sites within a protein structure. To study flexibility across the active-site and derive the root mean square fluctuation (RMSF) fluctuations across the cavity, the CABS-flex2.0 server is used [42]. It estimates flexibility/rigidity of the secondary structures and key residues of the constructed model. For localizing these flexible sites in correlation with the topology of the predicted model, Polyview-2D (<http://polyview.cchmc.org/>) server is used [43].

Phylogeny analysis:

To draw a credible evolutionary analysis, Gblocks is used for eliminating the evolutionary divergent regions and poorly aligned segments from the constructed alignment. It removes the ambiguous regions and takes into consideration only the conserved

regions to construct a phylogenetic tree. The resultant output is fed to a phylogeny server (http://phylogeny.lirmm.fr/phylo_cgi/index.cgi) to construct an evolutionary tree [44, 45]. Using the minimum value of SH-like statistically assesses the evolutionary relationship of the sequence dataset and Chi-2 based tests. The evolutionary distances are further computed using the Jonathan Taylor Thomas (JTT) matrix method.

Results:

Physicochemical properties:

The physicochemical properties of the ATP-dependent RNA helicase are estimated through ProtParam. For the 678-residue sequence, the molecular weight is estimated to be 76.6KDa. The sequence encodes 75 negatively and 73 positively charged residues, and it indicates that the protein is somewhat negatively charged. Theoretical pI is estimated to be 6.98 and it exhibits a slightly acidic nature. The extinction coefficient value and in-vitro half-life of the protein are respectively estimated to be 66,350 and 30 hours. The molecular formula is shown to be $C_{3365}H_{5391}N_{983}O_{994}S_{34}$ and it shows the GRAVY score of -0.294.

Secondary structure prediction:

The secondary structure elements define a protein structure and their encoded fractions play a key role in designing various bioanalytical experiments. Using PSIPRED, the fraction of α -helix, coil and β -strands are orderly estimated to be 46.6, 37.4 and 15.9 (Figure 2A), and it indicates a substantial predominance of α -helix than the remaining elements. The estimated secondary structures are marked across the chain, along with their statistical confidence (Figure 2B).

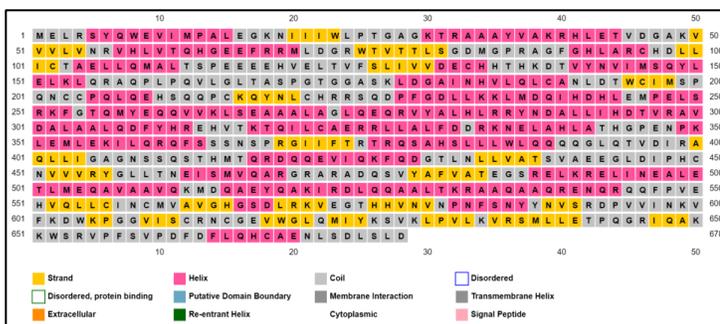


Figure 2A: PSIPRED result indicating the secondary structure and cellular location of the ATP-dependent RNA helicase

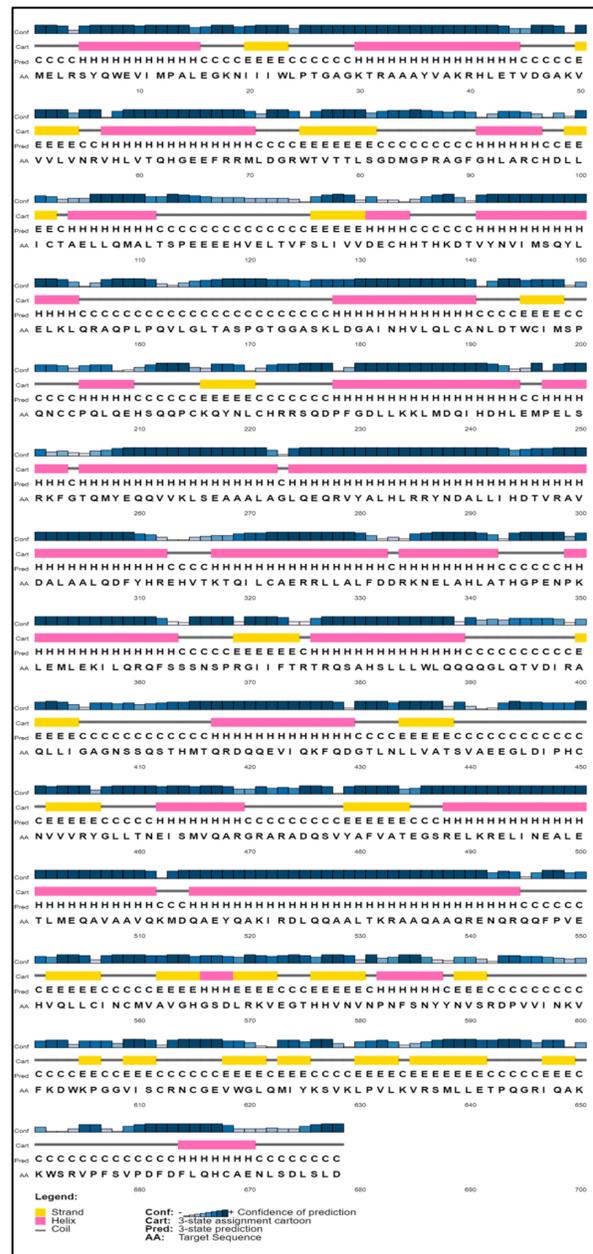


Figure 2B: Predicted three-state (helix/sheet/coil) secondary structure of ATP-dependent RNA helicase residues (AA) by PSIPRED at the confidence level (Conf); strands, helices, and coils are respectively represented as E, H, and C.

atomic clashes, the predicted structure is iteratively refined through 3Drefine to extensively sample its conformational space. The refined structure orderly shows a credible TM-Score and C α -RMSD of 0.96442 and 1.03 against 5F9FE. The model shows an ERRAT score of 94.1807, and it affirms the non-bonded interaction network in the model. The constructed decoy shows a DOPE and GA341 score of -77915.687 and 1.00 respectively. While the latter score indicates the structural compactness, the former energetic measure confirms the near-native credibility of the predicted model.

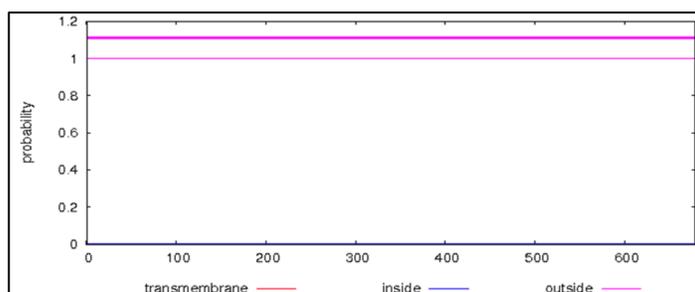


Figure 6: Graphical representation of the cellular location estimated by TMHMM for the ATP-dependent RNA helicase enzyme.

As shown in (Figure 7), a set of 90.70% and 7.50% residues are found to be localized within the most-favored and additionally allowed regions in the Ramachandran map, plotted through PROCHECK, as detailed in the following (Table 1) [48] Assessing through the QMEAN server, the model shows the clash score, rotamer outlier percentage and Molprobrity score of 2.65, 0.52% and 1.48 respectively. It affirms the local and global accuracy and suggests that the topological accuracy of the predicted decoy is comparable to a medium-resolution crystallographic structure.

Table 1: PROCHECK estimated Ramachandran plot for the predicted helicase structure. It shows 99.2% residues in the favorable regions and confirms the modelling credibility.

Plot statistics	Number	Percentage
Residues in most favored regions	547	90.7%
Residues in additional allowed regions	45	7.5%
Residues in generously allowed regions	6	1.00%
Residues in disallowed regions	5	0.80%
Number of non-glycine and non-proline residues	603	100%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues (shown as triangles)	36	
Number of proline residues	24	
Total number of residues	666	

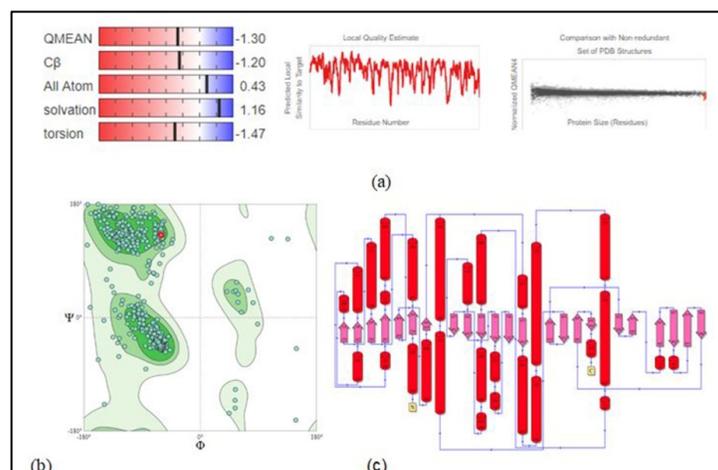


Figure 7: Structural assessment of the predicted protein model through (a) Qmean-score and Z score (b) PROCHECK-derived Ramachandran plot showing the 99.20% residues localized within the topologically allowed regions (c) Protein topology map, constructed using Profunc, with red color representing the helices and blue color representing the direction.

Conservation and flexibility analysis:

The conservation level, indicating the color gradation with maroon, white and turquoise to orderly represent the higher, medium and lower order of sequence conservation, is mapped onto the surface of the constructed protein model (Figure 8). The analysis reveals an average pairwise distance score of 1.49264, within the range of $1.01758e-07$ to 3.2305, across the entire sequence length. While only 68 residues are found to be completely conserved, 124, 308 and 178 loci are orderly found to be highly, moderately and poorly conserved.

The molecular surface of the helicase protein structure (Figure 9A) is analyzed through the CASTp server for identifying the pockets, cavities and cross channels. The biggest cavity shows the surface area and volume of 6077.513\AA^2 and 10513.441\AA^3 , and it signifies that the structure encodes a substantially broad cavity (Figure 9B). It is interesting to observe that only 42 residues (I22, L24, P25, A28, K30, V54, R56, V57, T103, E105, L106, M109, K138-D-T140, T167, Q256, M257, E259, Q260, R285, R375-T-R377, I404-G-A406, T438, S439, G444, L459, N461, R492, H576, F601, P606, L621, V632, K634, K650, W652, and S653) define the active site (Figure 9C). To estimate the atomic fluctuations across the cavity, the structural flexibility is estimated with the CABS-flex algorithm. On the basis of Ca α , C β , and side-chain representation, it quickly simulates

a protein structure and overcomes the size limitation of the classical molecular dynamics strategy.

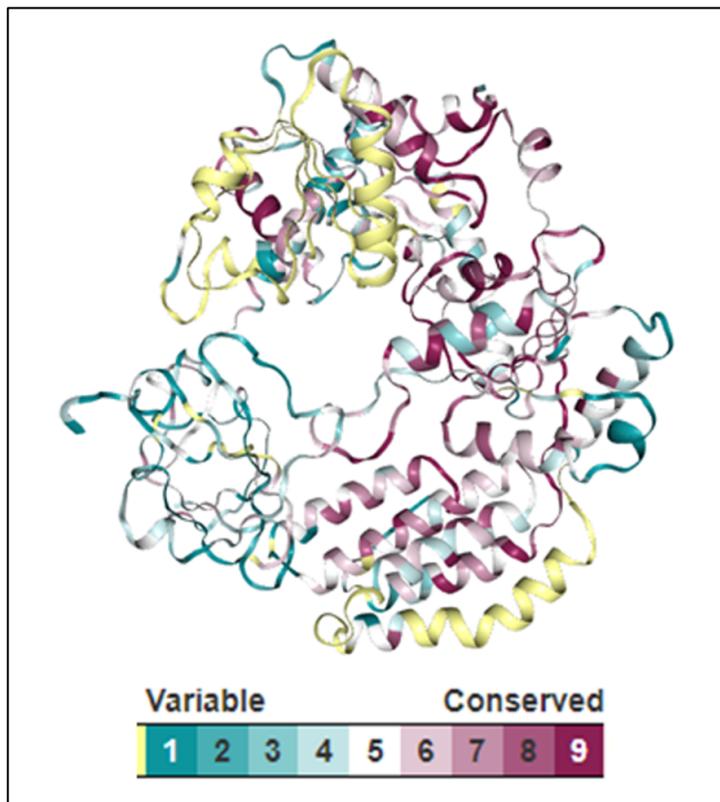


Figure 8: Consurf-derived conservation analysis of the human ATP-dependent RNA helicase. Color-coding is used to mark the evolutionary rate of residues over the predicted model. Low, mean and high evolutionary variability is orderly marked as maroon, white and turquoise.

For an input protein structure, the output ensembles the atomic-resolution profile representing the flexibility of the input structure. As the functionality of a protein is dependent on its topological flexibility, it is mandatory to map such vital sites across the protein sequence. Overlapping the sequence conservation map of this protein with the active-site cavity, it further illustrates that the core cavity is not highly conserved. Structural mapping through POLY VIEW-2D further shows that some flexible loci are significantly conserved and it delineates that these residues are essential for protein function. To analyze the residue flexibility score in correlation with the secondary structure and sequence conservation

of the residues, the results are overlapped and the average structural fluctuations are marked with a red line (**Figure 10**). The average and standard deviation of the RMSF scores for these loci are orderly found to be 0.781 and 0.495, unlike the respective scores of 0.804 and 0.593 for the complete structure, and it indicates that the active site is a bit more structurally stabilized. However, only 9 residues (L24, A28, K30, Q256, R285, T438, S439, G444, and L459) are found to be conserved, and it shows that the flexibility is natively vital for only a few residues.

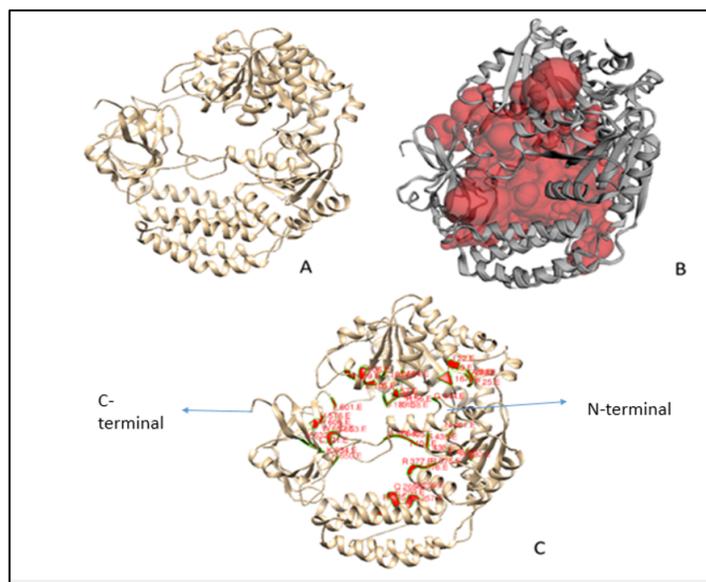


Figure 9: (A) Near-native model (B) Active-site zone estimated by CastP (C) Key active-site residues of the ATP-dependent RNA helicase (Q96C10.1).

Phylogeny analysis:

The 259-sequence dataset is aligned through ClustalW and is curated by eliminating the poorly aligned positions and divergent regions. Gblocks server is used to select the informative positions of the sequences. The dataset shows a mutual sequence identity within the range of 13.32-99.87, and the lower limit indicates a distant evolutionary linkage. Excavating it further, it shows 7 major evolutionary clades, and orderly defines the subsets of 25, 70, 23, 27, 17, 26 and 62 sequences. The sequence-identity range for these clades lies within the range of 22.55-96.27, 27.1-99.87, 29.31-99.84, 23.49-99.84, 36.19-98.24, 33.33-95.03 and 32.36-99.83 respectively. The mean sequence identity for each of these clades is orderly found to be 40.37±16.62, 40.44±12.53, 45.77±16.29, 39.07±16.07,

58.50±16.33, 45.75±10.88 and 54.78±12.38. It thus indicates that the clade5 members are evolutionary too close and 27 clade4 members share a distant relationship. However, the species are uniformly present across most of the clades (**Figure 11**).

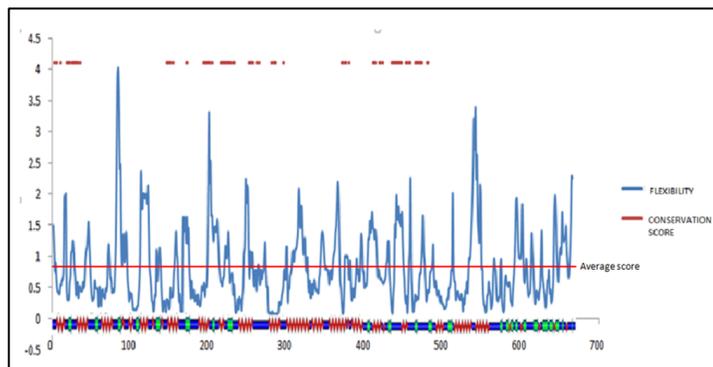


Figure 10: CABS-flex-estimated RMSF scores defined in correlation with their secondary-structure and the sequence conservation (marked with red superficial bars) for the ATP-dependent RNA helicase. Several flexible residues are found to be evolutionarily conserved across the topologically important secondary structure elements.

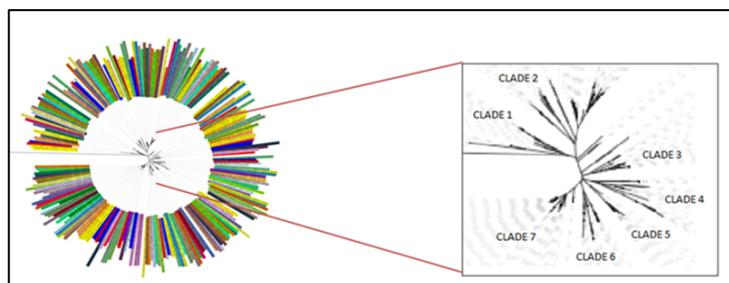


Figure 11: Phylogenetic tree of 259 sequences comprising 7 clades in which species in a clade are closely related to each other when compared to species in another clade.

Discussion:

The RNA helicase subfamily harbors several multi-functional enzymes. For its key role in various aspects of RNA metabolism, the ATP-dependent RNA helicase has been extensively studied [49]. However, the structural conservation and evolutionary divergence of several key sub familial members are still not functionally excavated through a functionally similar dataset [50]. While the sequence analysis of DHX58 protein reveals many characteristic features [51, 52], the STRING database has revealed

DDX58 as the top-ranked partner, as also shown in the recently published reports [53]. DHX58 is further found to interact with IFIH1, ISG15, RSAD2, IRF7, MX1, MAVS, DICER1, USP18 and OASL. However, IFIH1 has been recently shown to have an affinity for RNA [54], and the helicase motif of DICER (DICER1) has been shown crucial for processing the siRNA [55]. Further, as shown by recent microarray analysis, 1.9 fold upregulation of DHX58 is orderly found associated with a 2.1, 2.2, 2.2, 2.4, 2.8 and 4.0 fold upregulation of MX1, IFIH1, USP18, DDX58, OASL, and RSAD2 proteins [56, 57]. Hence, in accordance with the earlier studies, our estimated network of the top-ranked proteins (**Figure 4**) strongly indicates a potential role of the interacting partners in the immune signaling mechanism of DHX58. Motif segments have only been shown to be highly conserved in contrast to a significant variability across the N- and C-terminal domains, majorly responsible to interact with a diverse set of proteins. However, our ConSurf analysis shows statistically higher conservation for 168 residues through the HMM-profile of the constructed 259-sequence dataset, as mapped on the predicted near-native structure of DHX58 (**Figure 8**). Further, as estimated through CABS-flex (Figure 10), the structural fluctuations are found highest for some terminal residues, although the model shows a significant structural fluctuation across the chain. Although this is in agreement with the earlier results [58,59], it shows that the fluctuation of the key residues could possibly have a vital functional role. The evolutionary study shows a 7-clade evolutionary distribution of the constructed 259-sequence dataset, and each clade is found to span the sequences from all the available species. However, the structural superimposition based study of DEAD domains of DDX2A, DDX2B, DDX5, DDX10, DDX18, DDX20, DDX47, DDX52, and DDX53, and the helicase domains of DDX25 and DDX41 shows a $C\alpha$ -RMSD within 0.6-1.9Å over the diverse sequence identity range of 27%-86 [60]. The study thus adds on to the details reported earlier and it implicates that these protein structures are robustly conserved over the sequence alterations. Besides the interaction with a few molecules like β -catenin, a protein involved in the gene transcription [61], the active-sites of helicases have not been extensively excavated [62]. To extend it further, it is observed that the enzyme encodes a set of 42 active-site residues, of which 9 residues are found to be conserved. The active-site shows a lower topological fluctuation than the overall structure. Thus, the presented analysis provides a reliable framework for a more detailed evolutionary and structural analysis of ATP-dependent RNA helicase.

Conclusion:

It is of interest to annotate the human protein Q96C10.1 using known data to model structure and infer function with potential

role in the pathway. We document that ten proteins, including DDX58 and OASL are potential interacting protein partners with Q96C10.1. A dataset of 259 functionally similar homologs shows an evolutionary clustering within seven clades and shows conservation of only 9 active-site residues. It is inferred that active site residues are not highly conserved to link with the corresponding low structural similarity in these enzyme proteins.

Authors' contributions: AR planned and supervised the experimental methodology. VK and AK have carried out the work, and took the lead in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

The authors acknowledge the University and the department for providing the required resources/support. This work is not funded any agency.

Conflict of interests:

The authors declare no conflict of interest.

Funding: It is not a funded research.

References:

- [1] Anantharaman V *et al.* *Nucleic Acids Research*. 2002 **30**:1427 [PMID: 11917006].
- [2] Jing-Wen S & Yan-Hwa WL. *Clinica Chimica Acta*.2014 **436**:45 [PMID: 24835919].
- [3] Cordin O *et al.* *Gene* 2006 **367**:17 [PMID: 16337753].
- [4] Soto-Rifo R & Ohlmann T. *Wiley Interdisciplinary Reviews RNA*.2013 **4**:369 [PMID: 23606618].
- [5] Sharma D & Jankowsky E. *Critical Reviews in Biochemistry and Molecular Biology* 2014 **49**:343 [PMID: 25039764].
- [6] George WO. *Methods in Enzymology* 2012 **511**:385 [PMID: 22713330].
- [7] Linder P & Jankowsky E. *Nature Reviews Molecular Cell Biology* 2011 **12**:505 [PMID: 21779027].
- [8] Linder P & Fuller-Pace FV. *Biochimica Biophysica Acta* 2013 **1829**:750 [PMID: 23542735].
- [9] Satoh T *et al.* *Proceedings of the National Academy of Sciences USA* 2010 **107**:1512 [PMID: 20080593].
- [10] Murali A *et al.* *The Journal of Biological Chemistry* 2008 **283**:15825 [PMID: 18411269].
- [11] Shahanshah K *et al.* *International Review of Cell and Molecular Biology* 2019 **344**:215 [PMID: 30798989].
- [12] Yoneyama M *et al.* *The Journal of Immunology* 2005 **175**:2851 [PMID: 16116171].
- [13] Jankowsky E & Margaret F. *Current Opinion in Structural Biology* 2007 **17**:316 [PMID: 17574830].
- [14] Campagnoni AT. *Journal of Neurochemistry* 1988 **51**:1 [PMID: 2454292].
- [15] Fuller-Pace FV. *Trends in Cell Biology* 1994 **4**:271 [PMID: 14731588].
- [16] Jesus de la C *et al.* *Trends in Biochemical Sciences* 1999 **24**:192 [PMID: 10322435].
- [17] Venkat SRKY *et al.* *Frontiers in genetics* 2004 **119**:381 [PMID: 20225157].
- [18] Barhoumi M *et al.* *FEBS Journal* 2006 **273**:5086 [PMID: 17087726].
- [19] Matthias WH *et al.* *Nature Reviews Molecular Cell Biology* 2018 **19**:327 [PMID: 29339797].
- [20] Samantha B *et al.* *Journal of Biological Chemistry* 2019 **294**:11473 [PMID: 31175158].
- [21] Sowmya P *et al.* *Journal of Interferon & Cytokine Research* 2019 **39**:331 [PMID: 31090472].
- [22] Felisberto-Rodrigues C *et al.* *Biochemical Journal* 2019 **476**:2521. [PMID: 31409651].
- [23] Rahman MM *et al.* *Scientific Reports* 2017 **7**:15710 [PMID: 29146961].
- [24] Zhang H *et al.* *Hepatology* 2016 **64**:1033 [PMID: 27338082].
- [25] Gasteiger E *et al.* *The Proteomics Protocols Handbook* pp. 571-607.
- [26] Jones DT. *Journal of Molecular Biology* 1999 **292**:195 [PMID: 10493868].
- [27] Zimmermann L *et al.* *Journal of Molecular Biology* 2018 **430**:2237 [PMID: 29258817].
- [28] Runthala A. *Journal of Biomolecular Structure and Dynamics* 2012 **30**:607 [PMID: 22731875].
- [29] Runthala A & Chowdhury R. *Journal of Bioinformatics and Computational Biology* 2019 **17**:1950006 [PMID: 31057073].
- [30] Webb B & Sali A. *Current Protocols in Bioinformatics* 2016 **54**:5.6.1 [PMID: 27322406].
- [31] Bhattacharya D *et al.* *Nucleic Acids Research* 2016 **44**:W406 [PMID: 4987902].
- [32] Mitchell AL *et al.* *Nucleic Acids Research* 2019 **47**:D351 [PMID: 30398656].
- [33] Marchler-Bauer A *et al.* *Nucleic Acids Research* 2017 **45**:D200 [PMID: 27899674].
- [34] Szklarczyk D *et al.* *Nucleic Acids Research* 2019 **47**:D607 [PMID: 30476243].
- [35] Bailey TL & Elkan C. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 1994 **2**:28 [PMID: 7584402].
- [36] Frith MC *et al.* *PLOS Computational Biology* 2008 **4**:1 [PMID: 18437229].
- [37] Roman AL *et al.* *Nucleic Acids Research* 2005 **33**:W89 [PMID: 1160175].

- [38] Krogh A *et al.* *Journal of Molecular Biology* 2001 **305**:567 [PMID: 11152613].
- [39] Finn RD *et al.* *Nucleic Acids Research* 2011 **39**:W29 [PMID: 21593126].
- [40] Ashkenazy H *et al.* *Nucleic Acids Research* 2016 **44**:W344 [PMID:21766375].
- [41] Wei T *et al.* *Nucleic Acids Research* 2018 **46**:W363 [PMID: 29860391].
- [42] Jamroz *et al.* *Nucleic Acids Research* 2013 **41**:W427 [PMID: 23658222].
- [43] Porollo A *et al.* *Bioinformatics* 2014 **20**:2460 [PMID: 15073203].
- [44] Guindon S *et al.* *Systematic Biology* 2010 **59**:307 [PMID: 20525638].
- [45] Castresana J. *Molecular Biology and Evolution* 2000 **17**:540 [PMID: 10742046].
- [46] Runthala A & Chowdhury R. Protein Structure Prediction: Are We There Yet? In: Pham T, Jain LC (ed) Knowledge-Based Systems in Biomedicine and Computational Life Science, 2013 450thedn. Springer, Germany, pp 79-115
- [47] Runthala A & Chowdhury R. Unsolved problems of ambient computationally intelligent TBM algorithms. In: Bhattacharyya S, Dutta P, Chakraborty S (ed) Hybrid Soft Computing Approaches. Studies in Computational Intelligence, 2016, 611thedn. Springer, Germany, pp 75-105
- [48] Laskowski RA *et al.* *Journal of applied Crystallography* 1993 **26**:283 [PMID: 9008363].
- [49] Jankowsky E. *Trends in Biochemical Sciences* 2011 **36**:19 [PMID: 20813532].
- [50] Sharma D & Jankowsky E. *Critical reviews in Biochemistry and Molecular Biology* 2014 **49**:343 [PMID: 25039764].
- [51] Hilbert M *et al.* *Biological Chemistry* 2009 **390**:1237 [PMID: 19747077].
- [52] Takeuchi O & Akira S. *Current Opinion in Immunology* 2008 **20**:17 [PMID: 18272355].
- [53] Quicke KM *et al.* *Journal of interferon and cytokine research* 2019 **39**:669 [PMID: 31237466].
- [54] Zhang R *et al.* *Scientific reports* 2018 **8**:14189 [PMID: 30242207].
- [55] Deleris A *et al.* *Science* 2006 **313**:68 [PMID: 16741077].
- [56] MacKay CR *et al.* *Current Opinion in Immunology* 2014 **26**:49 [PMID: 24556400].
- [57] Aka JA *et al.* *Molecular and Cellular Endocrinology* 2017 **439**:175 [PMID: 27544780].
- [58] Fuller-Pace FV. *Nucleic Acids Research* 2006 **34**:4206 [PMID: 16935882].
- [59] Linder P & Jankowsky E. *Nature Reviews Molecular Cell Biology* 2011 **12**:505 [PMID: 21779027].
- [60] Schütz P *et al.* *PLoS ONE* 2010 **5**:e12791 [PMID: 20941364].
- [61] Waqar A *et al.* *Progress in Biophysics and Molecular Biology* 2018 **140**:79 [PMID: 29729328].
- [62] Linder P. *Nucleic Acids Research* 2006 **34**:4168 [PMID: 16936318].

Edited by P Kanguane

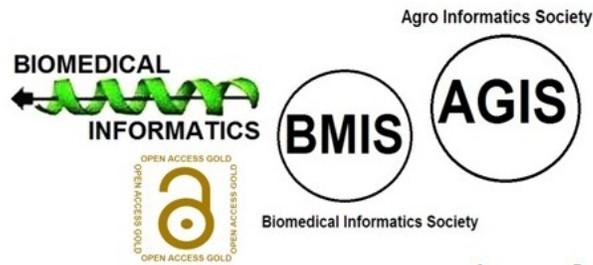
Citation: Kamjula *et al.* *Bioinformation* 16(2): 160-170 (2020)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

BIOINFORMATION

Discovery at the interface of physical and biological sciences



since 2005

BIOINFORMATION

Discovery at the interface of physical and biological sciences

indexed in

