ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 18(4)

Research Article

Received March 1, 2022; Revised April 30, 2022; Accepted April 30, 2022, Published April 30, 2022

DOI: 10.6026/97320630018411

Declaration on Publication Ethics:
The author's 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.

Declaration on official E-mail:
The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

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

Edited by P Kangueane
Citation: Bakhrebah *et al.* Bioinformation 18(4): 411-419 (2022)

Sequence to predicted structure and function for DHX8

Muhammed A Bakhrebah[1], Majed S Nassar[1*], Fahad M Alhoshani[1], Abdulkhaliq Munawir Alanazi[2] & Saeed A Asiri[3]

[1]Life Science and Environment Research Institute, King Abdulaziz City for Science and Technology (KACST), P.O. Box 6086, Riyadh 11442, Kingdom of Saudi Arabia; [2]Department of Respiratory Services, Ministry of National Guard Hospital and Health affairs (MNGHA), P.O. Box 22490, Riyadh, Kingdom of Saudi Arabia; [3]Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Najran University; *Corresponding author

Affiliation URL: https://www.kacst.edu.sa/
https://www.nu.edu.sa/

Author contacts:
Muhammed A Bakhrebah – E-mail: mbakhrbh@kacst.edu.sa
Majed S Nassar - E-mail: mnassar@kacst.edu.sa
Fahad M Alhoshani – E-mail:fhoshani@kacst.edu.sa
Abdulkhaliq Munawir Alanazi – E-mail:alanaziab33@ngha.med.sa
Saeed A Asiri – E-mail:saaasiri@nu.edu.sa

ISSN 0973-2063 (online) 0973-8894 (print)

Bioinformation 18(4): 411-419 (2022)

Abstract:
DEAH-box helicase 8 (DHX8) is a protein coded by the DHX8 gene in humans. Available information on the sequence to structure to function for this protein is limited in the literature. Therefore, it is of interest to document insights gleaned from the sequence to structure to functional analysis of the DHX8 protein using known software tools such as SMART, Prosite, Pfam, PANTHER, and InterProScan. We also present the predicted preliminary structure data for the protein to glean structure based functional insights.

Keywords: DHX8, Pfam, Inter Pro Scan beta turn, α-helix, SMART, Prosite, PANTHER, Ramachandran plot, hydrophobicity and antigenic region

Background:
DHX8 (DEAH box helicase 8) is a protein encoded by the DHX8 gene in humans). The conserved motif DEAH (single letter code for amino acids) consists of protein residues Asp-Glu-Ala-His **[1].** DHX8 is an ATP-dependent RNA helicase that is used in splicing and& regulation of spliced mRNA from the spliceosomes of the nucleus **[2].** The RNA binding domain (S1) contains DEAH/DEAD box, helicase domain (C-terminal), HA2 and an oligo-nucleotide/oligosaccharide binding connected by intrinsically disordered areas made of DHX8 **[3].** DHX8 performs various biological functions such as ATP binding, RNA binding and enzymatic activity (RNA helicase) **[4-6].** Moreover, its functional homolog in yeast (Prp22) shows splicing loyalty and proof-reading **[7,8].** This implies alternative splicing activation **[9, 10].** DHX8 mutations cause incomplete mRNA splicing and cell division abnormalities in zebra fish. siRNA suppression of DHX8 in human cells (HeLa), in splicing and thus, suggesting DHX8 is further more vital for mitotic exit **[11**]. The DEAH-box helicases of yeast spliceosomes have ATPase & RNA-unwinding functions at Prp22, Prp2, Prp16, & Prp43 **[12-13].** However, structural information for the CTD of the mammalian DEAH/RHA proteins is known **[14].** Nonetheless, comprehensive information on DHX8 is still limited in the literature. Therefore, it is of interest to document insights gleaned from the sequence to structure to functional analysis of this protein.

Methodology:
Sequence data:
The DHX8 gene (ID: 1659) sequence in humans is downloaded from URL https://www.ncbi.nlm.nih.gov/gene/1659

Hydrophobicity, amino acid charge and antigen prediction:
Amino acid charge and antigenic region were predicted using the emboss PLUGIN in Geneious prime. The signal cleavage region, and helicase domains have been assigned using the InterProScan plugin **[15].**

Domain Organization:
Pfam, scan Prosite, InterProScan, and SMART tools were utilized for finding domain organization **[16-18].** The domains in DHX8 are Ia, Ib, II, III, IV and VI in the RNA helicase as described elsewhere **[19].**

Structure data:
The helicase domain, including ATP binding domain and C-terminal helicase domain (362 amino acid) were developed using the SWISS-MODEL server (https://swissmodel.expasy.org/) with a suitable template (6HYU) **[16-20].** Ramachandran plots were drawn using the PROCHECK as described elsewhere **[21-23].** Motifs such as helix, strand, beta-turn, gama-turn, and beta-hairpin were also identified. The topology of the generated models was completed using PDBsum **[24].**

Structural helices:
The turns & nets accept a recovered helical value of 3.6 residues per turn **[23, 24].**

Beta turns:
A helicase domain (362 residues) is characterized by four contiguous residues represented by i, i+1, i+2, & i+3 as described elsewhere **[25, 26].** The classification VIa1 and VIa2 were used to recognize two subclasses of VIa turns by the phi, psi edges of buildup i+1 in the beta, & poly proline area of the Ramachandran plot. Types VIa1, VIa2, & VIb turns have supplementary conditions that residue i must be a cis-proline. Turns that do not follow the standards are ordered as type IV **[27-30].**

Table 1: Helices in the helicase domain of DHX8 (*Homo sapiens*)

S. No.	Start	End	Type	No. of residue	Length	Unit rise	Residues per turn	Pitch	Deviation form ideal	Sequence
1*	Gln2	Asp6	H	5	7.98	1.51	3.57	5.40	21.9	QAVHD
2*	Lys20	Ala31	H	12	17.77	1.48	3.80	5.61	14.5	KTTQITQYLAEA
3*	Arg46	Glu59	H	14	21.14	1.49	3.60	5.36	11.7	RVAAMSVAKRVSEE
4*	Met91	Cys96	H	6	9.32	1.50	3.40	5.10	2.8	MLLREC
5	Ala113	Glu115	G	3	-	-	-	-	-	AHE
6	Ile118	Val131	H	14	21.76	1.51	3.72	5.62	12.8	IHTDVLFGLLKKTV
7	Val149	Tyr154	H	6	9.50	1.51	3.52	5.32	4.6	VKFSQY
8	Tyr182	Thr196	H	15	22.10	1.45	3.70	5.38	10.3	YLDASLITVVMQIHLT
9	Gln209	Leu226	H	18	26.58	1.46	3.61	5.28	11.0	QEEIDTACEILYERMKSL
10	Ser244	Arg249	H	6	9.79	1.54	3.43	5.28	4.2	SEMQTR
11	Asn266	Ala268	G	3	-	-	-	-	-	NIA
12	Gln307	Gly315	H	9	14.14	1.52	3.55	5.41	7.4	QAQAKQRAG
13	Arg316	Gly318	G	3	-	-	-	-	-	RAG
14	Glu331	Asp348	H	6	9..06	1.49	3.58	5.33	8.8	ERAYRD
15	Glu345	Arg348	G	4	6.16	1.49	3.73	5.55	41.8	EIQR
16	Ala352	Met361	H	10	15.73	1.54	3.59	5.51	3.0	ASTVLSLKAM

*Number of helices in chain Y: 16; *Asterisked motifs correspond to those illustrated in the motif plots at the top of the page.*

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)

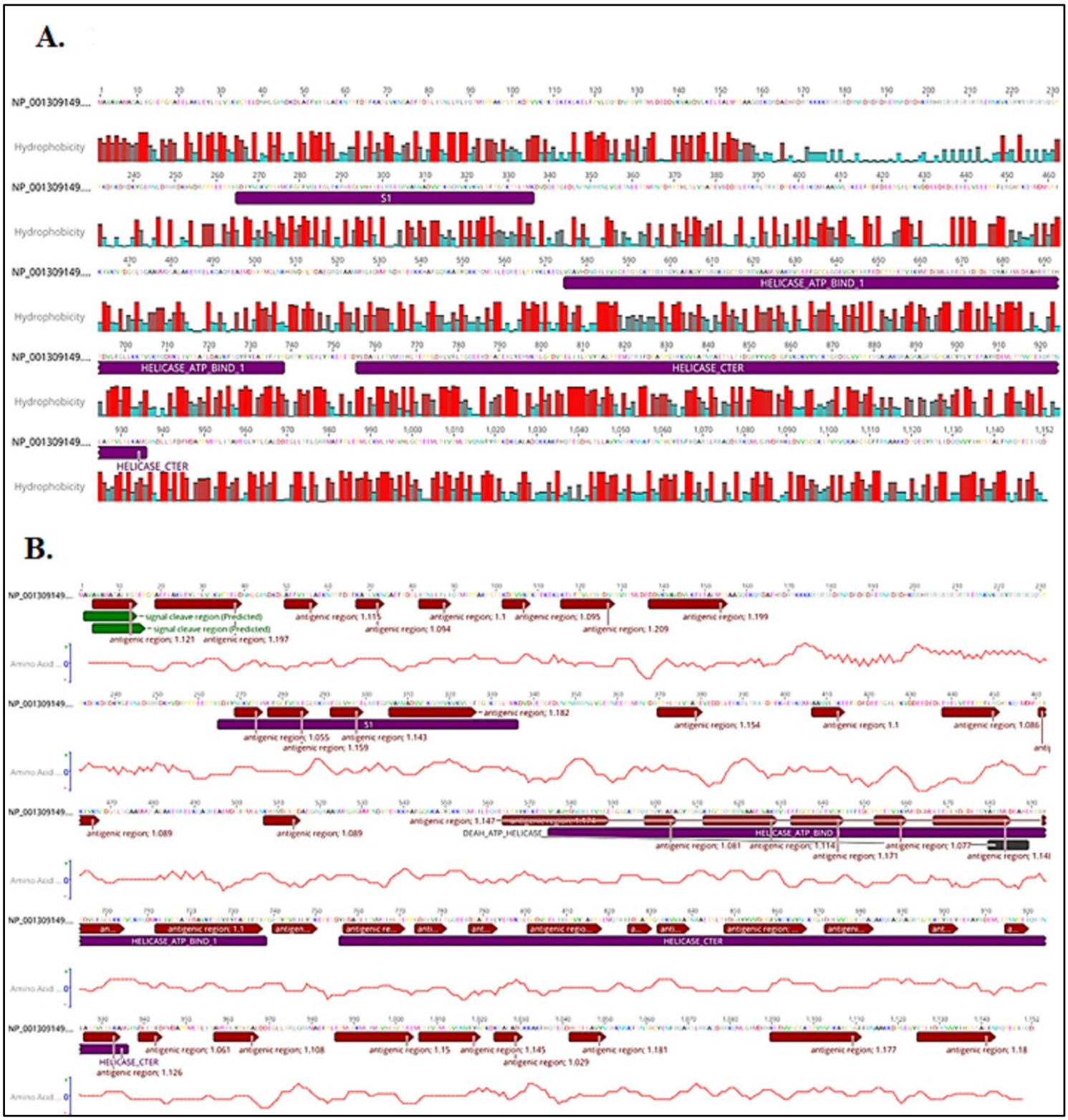


Figure 1: (A) Predicted hydrophobicity and **(B)** Predicted signal cleave region (green color), antigenic region (red color), and conserved domain (blue color)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)

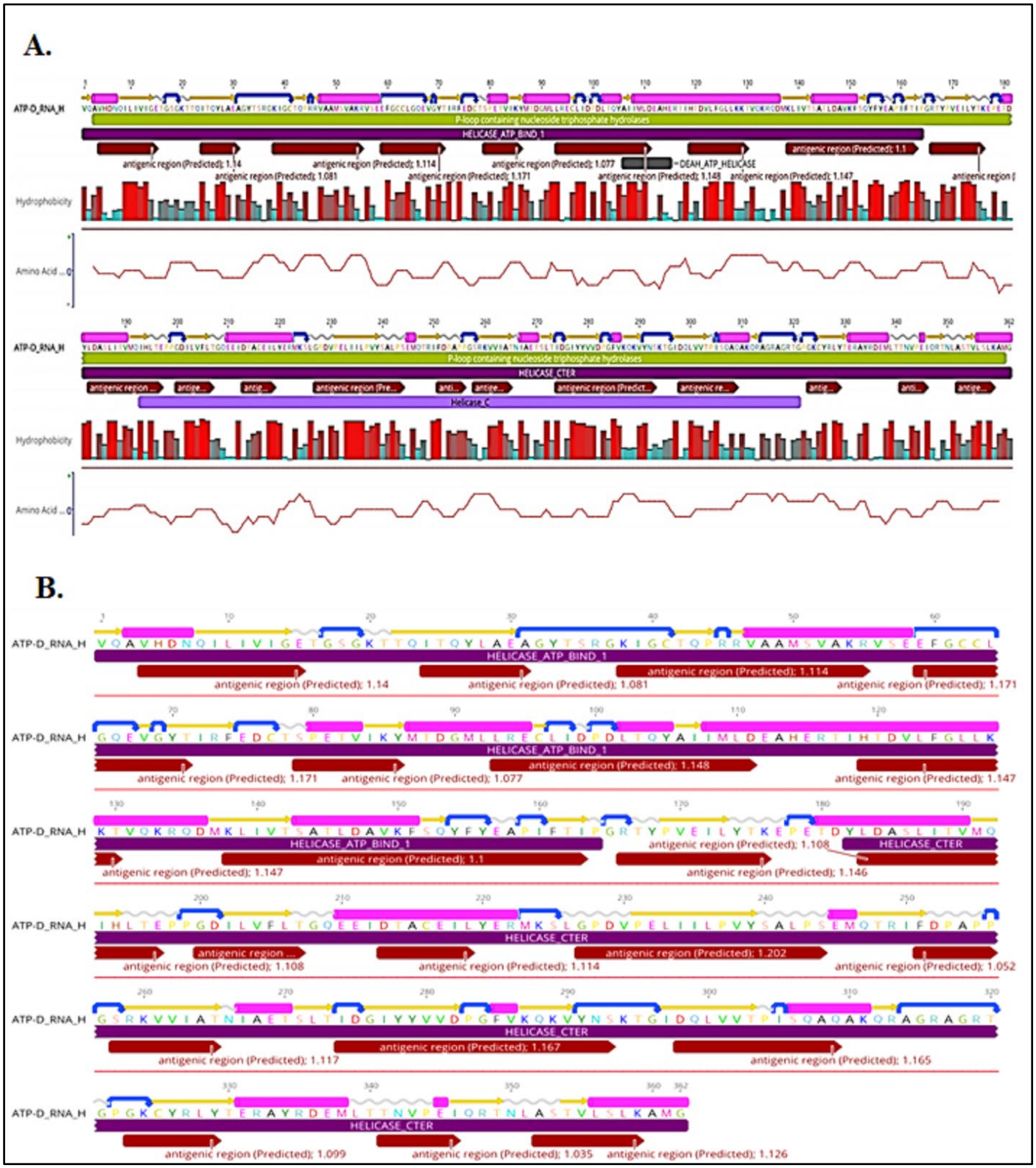


Figure 2: (A) Predicted hydrophobicity with amino acid charge and **(B)** Predicted secondary structure.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)

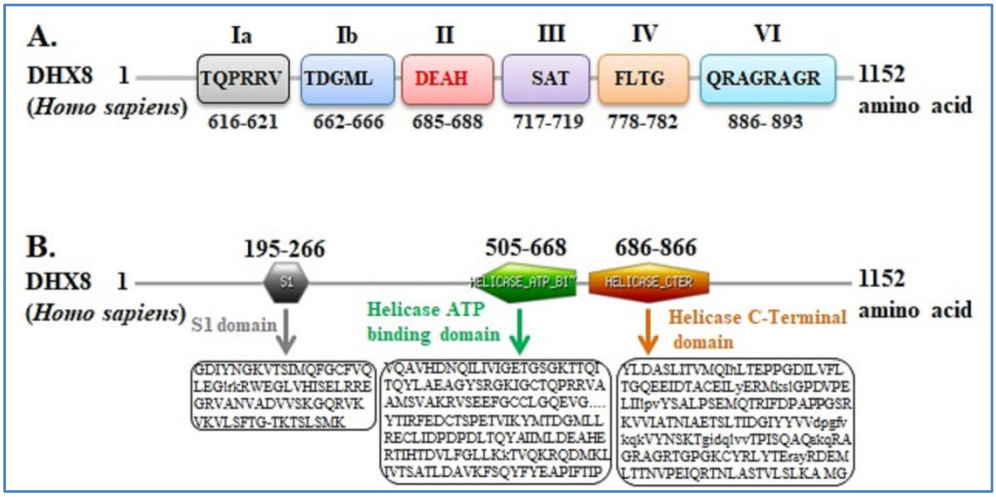


Figure 3: (A) Diagrammatic representation of the human ATP-dependent RNA helicase with conserved motifs and **(B)** ATP binding regions.

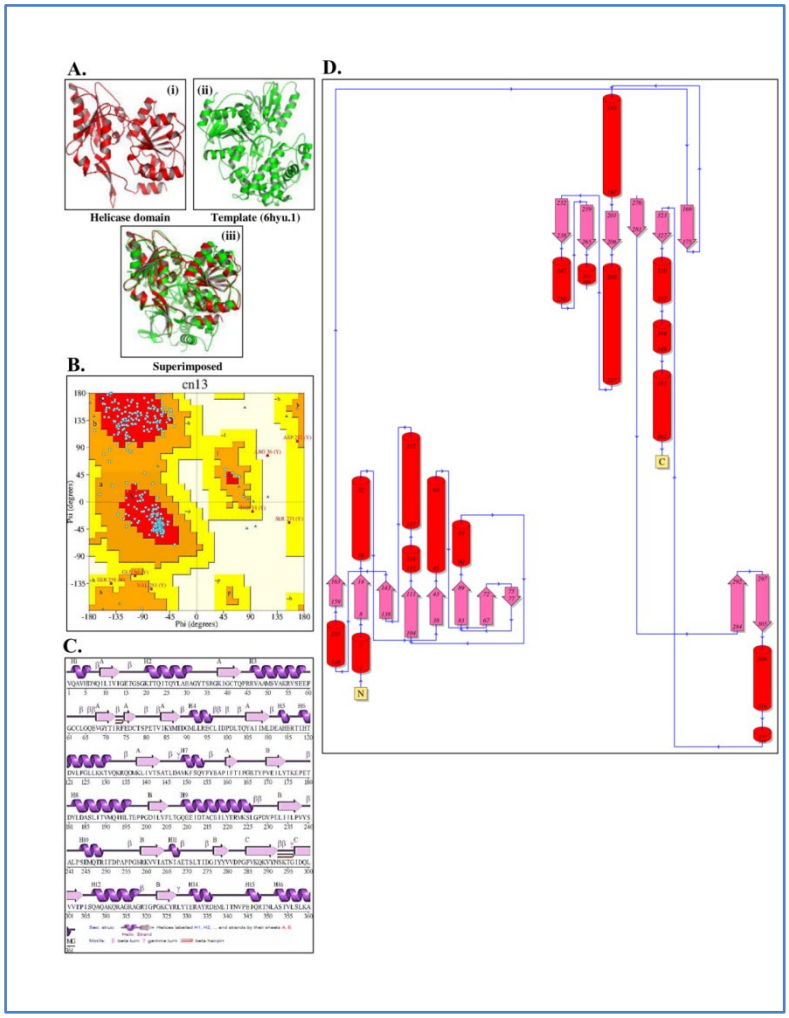
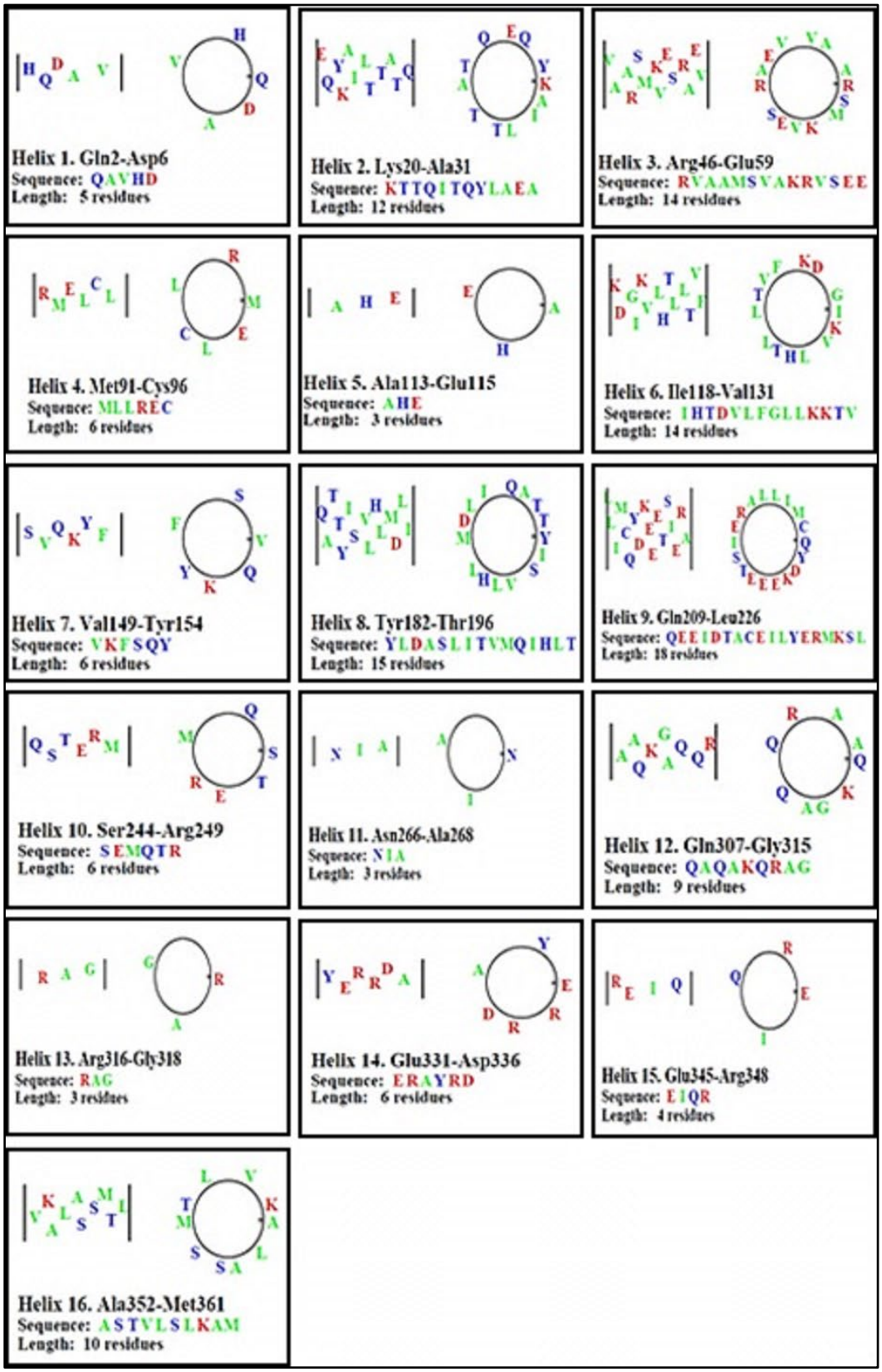


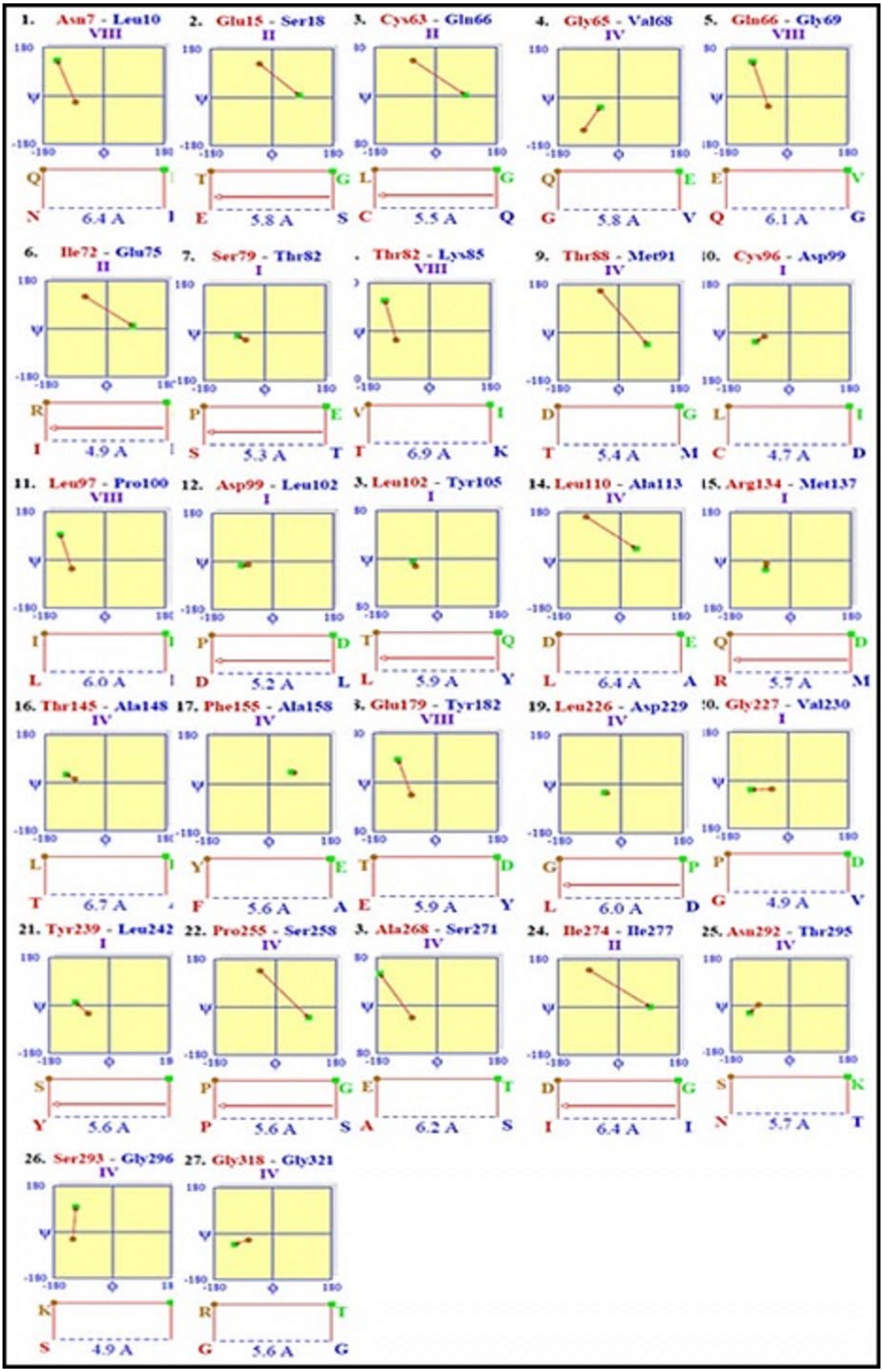
Figure 4: (A) Secondary structure analysis in the helicase domain; **(B)** Ramachandran plot for the helicase domain and**(C-D)** The observed topology of the helicase domain.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)



Figures 5: The helical diagram is shown

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)



Figures 6: The plots for turns using brown circle (residue i+1) & green square (residue i+2)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 18(4): 411-419 (2022)

Results and Discussion:

It is of interest to use Bioinformatics tools such as SMART, Prosite, Pfam, PANTHER, and InterProScan to analyze the DHX8 protein sequence to infer predicted function. It is also of interest present predicted structure data for the protein. We report such information to gain insights on the predicted function of DHX8. Hydrophobicity, antigenic area, and charge per amino acid residue have been predicted for in the DHX8 protein. We also plotted the charge per amino acid in the sequence. The family and domain were assigned using the InterProScan server (**Figure 1**). The antigenic region and secondary structure of the protein are shown (**Figure 2**). DHX8 (*Homo sapiens*) contains domain (ATP binding & C terminal) identified using the Bioinformatics tools (Pfam, scan Prosite, InterProScan, and SMART) as shown in **Figure 3**. The modeled structure was generated for the (i) Helicase domain; (ii) using the template (PDB ID: 6HYU.1) and (iii) superimposed structure (**Figure 4A**). Ramachandran plot showed residues in favored regions (**Figure 4B**). The observed topology and the predicted the secondary structure of the helicase domain is shown (**Figure 4C**). Structure (secondary) elements & their comparative 3D positions & rough alignments are shown (**Figure 4D**). The information given in **Table 1** shows the helix number (appointed consecutively beginning with one at the N-end of the protein), the buildup numbers compared to the beginning and end of the helices and the helix type (H (alpha helix), or G (3, 10) helix in DHX8. The 3.6 residues per turn helical value is shown (**Figure 5**). A beta-turn for DHX8 protein is shown (Figure 6). The corresponding data for beta turns in the DHX8 protein is given in Table 2 and Table 3. Thus, the use of computer aided tools in the understanding of biological sequences with less known information is illustrated.

Table 2: Beta strands in the helicase domain of DHX8

S.No.	Start	End	Sheet	No. of residue	Edge	Sequence
1	Ile9	Ile13	A	5	No	ILIVI
2	Lys38	Gln43	A	6	No	KIGCTQ
3	Val68	Ile72	A	5	No	VGYTI
4	Glu75	Cys77	A	3	Yes	EDC
5	Ile84	Thr88	A	5	No	IKTMT
6	Tyr105	Leu110	A	6	No	YAIIML
7	Lys138	Ser143	A	6	No	KLIVTS
8	Ile160	Thr162	A	3	Yes	IFT
9	Val170	Tyr174	B	5	Yes	VEILY
10	Asp201	Phe205	B	5	No	DILPV
11	Leu233	Val238	B	6	Yes	LIILPV
12	Arg259	Ala264	B	6	No	RKVVA
13	Ile277	Val280	B	4	No	IYYV
14	Phe285	Asn292	C	8	Yes	FVKQKVYN
15	Ile297	Pro304	C	8	Yes	IDQLVVTP
16	Gly323	Arg327	B	5	No	GKVYR

Number of beta strands in chain Y: 16

Table 3: Beta turns in the helicase domain of DHX8

S. No.	Turn	Sequence*	Turn Type	Residue i+1			Residue i+2			I to i+3 CA-dist	H bond
				Phi	Psi	Chi1	Phi	Psi	Chi1		
1.*	Asn7-Leu10	NQIL	VIII	-83.3	-23.9	-58.5	-138.1	132.2	-59.3	6.4	Yes
2.*	Glu15-Ser18	ETGS	II	-37.1	125.4	-58.6	80.4	7.9	-	5.8	No
3.*	Cys63-Gln66	CLGQ	II	-67.0	131.4	-174.4	86.1	4.0	-	5.5	No
4.*	Gly65-Val68	GQEV	IV	-103.0	-122.3	-60.0	-55.1	-37.6	-63.9	5.8	Yes
5.*	Gln66-Gly69	QEVG	VIII	-55.1	-37.6	-63.9	-101.9	129.0	174.1	6.1	Yes
6.	Ile72-Glu75	IRFE	II	-64.3	119.5	-171.7	78.5	9.4	-55.2	4.9	Yes
7.	Ser79-Thr82	SPET	I	-56.6	-27.8	-29.9	-80.1	-12.9	-61.4	5.3	No
8.	Thr82-Lys85	TVIK	VIII	-97.8	-34.5	172.8	-130.4	113.8	-49.4	6.9	No
9.	Thr88-Met91	TDGM	IV	-53.8	155.7	73.6	84.3	-44.2	-	5.4	Yes
10.	Cys96-Asp99	CLID	I	-73.5	-14.9	-56.6	-98.5	-34.3	-65.6	4.7	No
11.	Leu97-Pro100	LIDP	VIII	-98.5	-34.3	-65.6	-132.6	95.4	-180	6.0	Yes
12.	Asp99-Leu102	DPDL	I	-72.7	-12.9	21.1	-91.6	-16.1	-73.6	5.2	Yes
13.	Leu102-Tyr105	LTQY	I	-63.8	-29.0	68.4	-70.3	-12.1	-50.4	5.9	Yes
14.	Leu110-Ala113	LDEA	IV	-99.3	162.5	-153.4	47.5	44.3	-56.5	6.4	No
15.	Arg134-Met137	RQDM	I	-72.2	-12.4	-66.2	-73.8	-34.0	-58.6	5.7	Yes
16.	Thr145-Ala148	TLDA	IV	-90.7	12.4	-65.4	-117.0	31.4	-51.9	6.7	Yes
17.	Phe155-Ala158	FYEA	IV	71.4	41.6	-51.8	66.3	44.0	-58.3	5.6	Yes
18.	Glu179-Tyr182	ETDY	VIII	-70.4	-47.2	-60.4	-109.3	82.4	-165.4	5.9	Yes
19.	Leu226-Asp229	LGPD	IV	-45.9	-33.1	-	-48.2	-32.6	-31.0	6.0	No
20.	Gly227-Val230	GPDV	I	-48.2	-32.6	-31.0	-109.3	-34.6	-64.2	4.9	Yes
21.	Tyr239-Leu242	YSAL	I	-62.2	-31.9	51.5	-101.4	9.8	-	5.6	No
22.	Pro255-Ser258	PPGS	IV	-45.2	135.5	-30.5	98.7	-40.6	-	5.6	No
23.	Ala268-Ser271	AETS	IV	-74.9	-47.0	171.2	-171.0	117.4	-76.7	6.2	Yes
24.	Ile274-Ile277	IDGI	II	-85.7	137.6	-146.8	95.3	0.2	-	6.4	No
25.	Asn292-Thr295	NSKT	IV	-94.5	1.7	54.9	-121.9	-28.5	-66.6	5.7	Yes
26.	Ser293-Gly296	SKTG	IV	-121.9	-28.5	-66.6	-113.0	97.7	-59.1	4.9	Yes
27.	Gly318-Gly321	GRTG	IV	-75.0	-25.2	-176.6	-117.2	-41.1	47.6	5.6	Yes

Number of beta turns in chain Y: 27

Conclusion:

We document derived data to provide insights gleaned from the sequence to structure to predicted functional analysis of DHX8 protein using software tools in this report. Thus, the power of computer aided prediction tools in the understanding less known biological sequences is illustrated.

Conflict of Interests: There is no conflict of interests.

Acknowledgments:

The authors would like to thank King Abdulaziz City for Science and Technology (KACST) for the support.

References:

[1] www.genecards.org/
[2] https://pfam.xfam.org/
[3] https://www.ebi.ac.uk/
[4] https://www.ebi.ac.uk/GOA/news

[5] https://www.uniprot.org/uniprot/Q14562
[6] Mayas RM *et al. Nat. Struct. Mol. Biol.* 2006 **13**: 482. [PMID: 16680161]
[7] Semlow DR & Staley JP, *Trends Biochem. Sci.* 2012 **37:** 263. [PMID: 22564363]
[8] Semlow DR. *Cell*. 2016 **164:** 985. [PMID: 26919433]
[9] Park JW. *Proc. Natl Acad. Sci.*2004 **101:** 15974. [PMID: 15492211]
[10] English MA *et al. Dev. Dynam*. 2012 **241:**879. [PMID: 22411201]
[11] Martin A. *et al.* B. *J. Biol. Chem*. 2002 **277:** 17743. [PMID: 11886864]
[12] Tanaka N. & Schwer B. *Biochemistry*. 2005 **44:** 9795. [PMID: 16008364]
[13] Kudlinzki D *Proc. Natl Acad Sci.* 2012 **393:** 1131. [PMID: 23096351]
[14] T P Hopp & K R Woods. *Proc. Natl Acad Sci.* 1981 **78:** 3824. [PMID: 6167991]
[15] Tarique M *et al. Frontiers in microbiology.* 2017 **8:**130. [PMID: 28232818]
[16] Afaq S *et al. Bioinformation*. 2020 **16:**229. [PMID: 32308265]
[17] Bateman A *et al. Nucleic acids research.* 2002 **30:** 276.[PMID: 11752314]
[18] Tarique M *et al. Gene.* 2013 **522:**46. [PMID: 23562722]
[19] ARMOLD K: *Bioinformatics.* 2006 **22:** 195. [PMID: 16301204]
[20] Sheik S*Bioinformatics.* 2002 **18:** 1548. [PMID: 12424132]
[21] Felisberto-Rodrigues C *Biochem J.* 2019 **476:** 2521. [PMID: 31409651]
[22] Du C. *Breast Cancer Res Treat*. 2018 **172:** 561. [PMID: 30191368]
[23] Laskowski RA *Trends in biochemical sciences.* 1997 **22:** 488. [PMID: 9433130]
[24] May C & Appels R: *Theoretical and Applied Genetics*. 1987 **74:** 617. **[**PMID: 24240218]
[25] Wilmot C & Thornton J: *Protein Engineering Design and Selection.* 1990 **3:** 479. [PMID: 2371257]
[26] Ramachandran G & Venkatachalam C: *Original Research on Biomolecules.* 1968 **6:** 1255. [PMID: 5669465]
[27] Hutchinson EG & Thornton JM: *Protein Science.* 1994 **3:** 2207. [PMID: 7756980]
[28] Lewis PN *et al. Biochimica et Biophysica Acta.* 1973 **303:** 211. [PMID: 4351002]
[29] Richardson JS: *Advances in protein chemistry*. 1981 **34:**167. [PMID: 7020376]