

Sequence-Structure Relationship Guided Comparative Analysis of DEPDC1A and DEDC1B, Cancer Related Proteins

PALAK AHUJA

Department of Biotechnology, Faculty of Science, Jamia Hamdard, New Delhi-110062, India

palakahuja0314@gmail.com

Abstract: *Cancer is one of the most widely, studied multifaceted disease, involving a number of proteins which are differentially expressed in the tumor cells when compared to that of normal healthy cells. One such protein is DEP domain containing protein 1, which is expressed as two forms DEPDC1A and DEPDC1B. The two proteins have been found to be overexpressed in various type of cancers, making them a therapeutic target. However, to develop a more effective and disease specific drug, it is important to have a better understanding of the structural and functional characteristics of the target proteins, namely, DEPDC1A and DEPDC1B. The present study focuses on identifying the structural similarities/ dissimilarities between the two proteins and relating the observed results to their functional aspect.*

Keywords: *Oral Cancer; DEPDC1A; DEPDC1B; Potential Biomarkers*

Abbreviations: *DEPDC1: DEP domain containing protein 1; DEPDC1A: DEP domain containing protein 1A; DEPDC1B: DEP domain containing protein 1B; FA: Focal Adhesion; PTPRF: Focal Adhesion associated Protein Tyrosine Phosphatase Receptor Type F*

1. INTRODUCTION

Cancer, one of the most widely studied multifaceted disease, involves irregular growth and proliferation of the cells with a potential of metastasizing to other body parts [1], owing to an increased oncogene function, loss of function of several tumor suppressor genes resulting in unusual regulation of cell cycle [2]. Cancer is associated with alterations in a diverse number intracellular pathways, thereby leading to a transformed tumor cell metabolism and differential expression of a wide array of proteins, and enhancing the survival rate and growth of the tumor cell [3]. One such novel protein, often found to be differentially expressed in a number of cancers is, DEPDC1 i.e. DEP domain containing 1 protein. For instance, a study in the year 2007 reported over expression of DEPDC1 in the bladder cancer cells in comparison with 24 normal/ control human tissue and established via northern blot as well as immune-histochemical assays [4]. Another study in the year 2014, reported over expression of the protein DEPDC1B (DEP domain containing protein 1B), in oral cancer patients, where in it was observed that DEPDC1B mediates its function of cell migration as well as invasion, upon its interaction with another protein Rac1 [5]. Apart from stimulating Rac1 activation followed by cell proliferation [6], DEPDC1B has shown to play a crucial role in directing the de-adhesion events followed by cell cycle progression during mitosis. The study reported that DEPDC1B upon its accumulation in G2 phase of the cell cycle, competes with RhoA for its binding with ‘Focal Adhesion (FA) associated protein tyrosine phosphatase receptor type F (PTPRF)’ and induces the disassembling of Focal adhesions (FAs) leading to the morphological changes important for mitotic entry [7], indicating that its overexpression might enhance the cell cycle progression leading to carcinogenesis. Also, a study conducted in the year 2013, observed the role of DEPDC1B paralog, DEPDC1A in multiple myeloma, where in increased expression of the protein in malignant plasma cells lead to low survival rate in patients and knockdown of DEPDC1A protein hindered human myeloma cell line growth [8].

Structurally, DEPDC1B comprises of two conserved domains, namely DEP, a 90 amino acid long globular domain initially discovered in three proteins: *Drosophila* disheveled, EGL-10 of *Caenorhabditis elegans* and mammalian Pleckstrin, and RhoGAP [9-11]. The DEP domain is known

to facilitate the interaction between DEPDC1B and G-protein coupled receptor and the membrane phospholipids, essential for Wnt mediated signaling pathway and RhoGAP domain mediates the Rho GTPase signaling pathway [12-14]. However, not much is known about the structural as well as functional features of the DEPDC1B paralog, DEPDC1A, a poorly characterized protein with only a few published studies, for instance, it has been reported as a poor prognostic marker in breast, bladder, lung cancer and more recently in multiple myeloma [8, 15-17]

With both the proteins being involved directly or indirectly in tumorigenesis, it becomes important to have a better understand of the structural similarities and dissimilarities between DEPDC1A and DEPDC1B, and the influence of structural differences on their functional aspect. Also, the differential expression of the two proteins in the tumor cells when compared to the normal healthy cells, indicate the prominence of DEPDC1A and DEPDC1B as a potential protein based biomarker and an efficient diagnostic, prognostic and therapeutic purposes. The present study therefore, focuses on comparing the secondary structure of the two proteins and relating the observed differences/ similarities to their functional aspect.

2. METHODOLOGY

2.1 Retrieval of Protein Sequences

The protein sequences of the proteins DEPDC1A (UniProt ID: Q5TB30) and DEPDC1B (UniProt ID: Q8WUY9) was downloaded in FASTA format from UniProt (<http://www.uniprot.org/>) [18].

2.2 Pairwise Sequence Alignment of the Protein Sequences

The retrieval of the respective protein sequences was followed by the alignment of the two sequences using the software EMBOSS needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) [19], in order to identify the similarities/ dissimilarities between the sequences.

2.3 Prediction of Secondary Structure

The secondary structure of the two proteins were predicted using GOR IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) [20], a secondary structure prediction server, followed by their comparison in order to understand the structural dissimilarities that might have occurred due to differences in their protein sequences.

2.4 Identification of Domain/motif

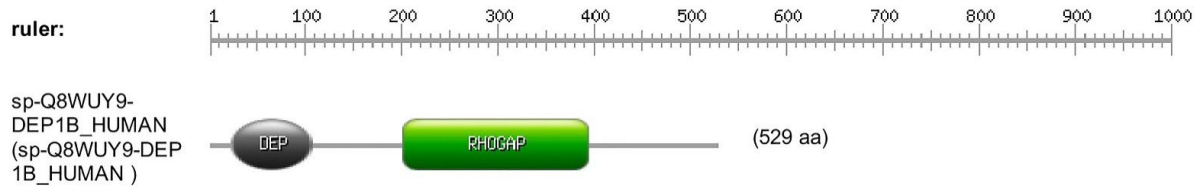
PROSITE (<http://prosite.expasy.org/>) [21], a server for analyzing and identifying the domain/motif in a protein, was used to identify the PROSITE domain in the proteins, DEPDC1A and DEPDC1B.

2.5 Comparative Analysis of the DEP Domain Protein Sequences of the Proteins, DEPDC1A and DEPDC1B

The protein sequence of the DEP domain in the two proteins- DEPDC1A and DEPDC1B, was compared using the server EMBOSS needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) [19], a server for the pairwise sequence alignment of the respective protein sequences.

3. RESULT

The protein sequences of DEPDC1A (UniProt ID: Q5TB30) and DEPDC1B (UniProt ID: Q8WUY9), obtained via UniProt, upon pairwise sequence alignment using EMBOSS needle showed that the two protein sequences were only 34.4% identical and 46.3% similar (Figure 1). Also, both the proteins contained DEP domain, from amino acid residue 24 to amino acid residue 108, identified using the server PROSITE (Figure 2). The sequence alignment of specifically the protein sequences of DEP domain, common in both the proteins was indicated that the two sequences are 71.8% identical, i.e. 61 amino acid residues out of 85 amino acid residues were identical, and 84.7% similar (Figure 3). Apart from the DEP domain, a second domain, RhoGAP domain, from amino acid residue 201 to amino acid residue 393 was identified using PROSITE only in DEPDC1B and but not in DEPDC1A (Figure 2). Pairwise sequence alignment of the protein sequences was followed by the prediction of the secondary structure of both the proteins using GOR IV. It was observed that DEPDC1A contained 35.76% of alpha helix, 14.06% of extended strand and 50.18% of random coil, whereas DEPDC1B consisted of 40.78% of alpha helix, 14.18% of extended strand and 44.05% of random coil (Figure 4).



b)

Figure2. Identification of the presence of the domain(s) in the proteins a) *DEPDC1A* and b) *DEPDC1B* using the server, PROSITE [21]

```
#=====
#
# Aligned_sequences: 2
# 1: DEP1A_HUMAN
# 2: DEP1B_HUMAN
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 85
# Identity:      61/85 (71.8%)
# Similarity:   72/85 (84.7%)
# Gaps:         0/85 ( 0.0%)
# Score: 342.0
#
#=====
```

DEP1A_HUMAN	1	FRAGMPLRKHRQHFKKYGNCFPAGEAVDWLYDLLRNNSNFGPEVTRQQTI	50
DEP1B_HUMAN	1	FRAKMPLRKHRCRFKSYEHCFATAEAVDWLHELLRCSQNFGEVTRKQTV	50
DEP1A_HUMAN	51	QLLRKFLKNHVIEDIKGRWGSENVDDNNQLFRFPA	85
DEP1B_HUMAN	51	QLLKKFLKNHVIEDIKGRWGEEDFEDNRHLYRFPP	85

Figure3. Pairwise Sequence Alignment of the DEP domain specific protein sequences in *DEPDC1A* and *DEPDC1B*, respectively using EMBOSS needle [19].

```

      10      20      30      40      50      60      70
MESQVPPGYPYRATKLNWVTTTSFRAGMPLRKHRQHFKKYGNCFPAGEAVDWLYDLLRNNSNFGPEVTRQ
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
QTIQLLRKFLKNHVIEDIKGRWGSENVDDNNQLFRFPATSPKLTLPRIYKLNKNIENFSKDKDSIFKFL
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
RNLRRTPKRHGLHLSQENGEKIKHEIINEDQENAINRELSQEDVEVRYVILYLQTLGVPVLEEVEE
hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
INPKQVPIQYIMNMANTSARGVILQNKSDLLPHWLSAMKCLANWPRSNMNNPITYVGFERDVFRTIA
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
DYFLDLPEPLTFEYELFVNILVVCGYITVSDRSAGIHKIQDDPQSSKFLHLNLSNLSFKSTECILLL
hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
HREKNKEESDSTERLQISNPGFERCAKMKQVLNLRNRVSANDIMGGSCHNLIGLSNMHDLSSNSKPRC
hhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
CSLEGIVDVPGNSSKEASSVFHQSFPIEGQNKLFLSKPKQFLLNLSHENIYKQPFSAQKRTSTLT
eeeeecccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
VQDQEELCNGCKSKQLCRSQQLLRSSSTRNRSYINTPVAEIMKPNVGGSTSVQTAMESELGESSATI
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
NKRLCKSTIELENSLIPASSMLTGTQSLQPLHRAIDAQLCCLLPPNRRKQLLMRMSRMSQNL
hhhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
VDMPLKLDAMGTRSLMIHFTSRVLCFAEVDLDELLAGRLVSLFMDHQEILQVPSYLQTAWEKHLDYL
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
KKGHIENPGDGLFAPLPTYSYCKQISAQEFDEQKIVSTSQAAAELENLNIKNRSLPLKEKRKLKQFQKE
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
YPLIQKRFPTTSEEAALFGDKPTIKQPMILIRKPKFRSLR
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

```

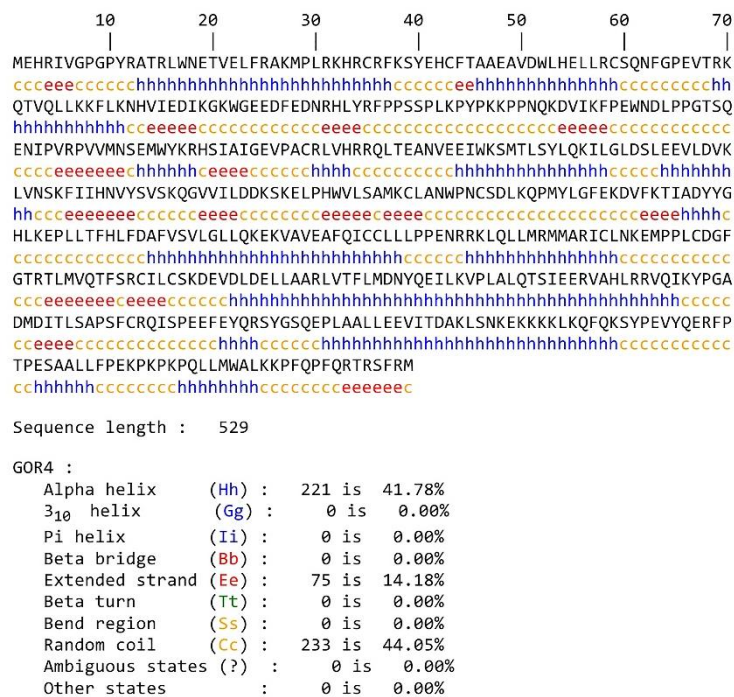
```

Sequence length : 811

GOR4 :
Alpha helix   (Hh) : 290 is 35.76%
310 helix    (Gg) :   0 is  0.00%
Pi helix      (Ii) :   0 is  0.00%
Beta bridge   (Bb) :   0 is  0.00%
Extended strand (Ee) : 114 is 14.06%
Beta turn     (Tt) :   0 is  0.00%
Bend region   (Ss) :   0 is  0.00%
Random coil   (Cc) : 407 is 50.18%
Ambiguous states (?) :   0 is  0.00%
Other states  :     0 is  0.00%

```

a)



b)

Figure4. Prediction of Secondary Structure of the protein DEPDC1A and DEPDC1B using GOR IV [20]

4. DISCUSSION

Both DEPDC1A and DEPDC1B, have been implicated in carcinogenesis as both the proteins have been reported to be differentially expressed in tumor cells when compared to the normal cells. However, the mechanism by which both proteins exert their pathological function might differ because of the differences in their protein sequences observed, as only 34.4% of the protein sequences were found to be identical and 46.7% similar, upon performing pairwise sequence alignment using a server, EMBOSS needle (Figure 1). Therefore, this huge difference in the protein sequence not only affects the structural characteristic of the protein but also its interaction with other proteins and hence its functionality. However, both the proteins have one common feature that is, the presence of DEP domain from amino acid residue 24 to amino acid residue 108 (a total of 84 amino acid residues) (Figure 2), as observed via the server, PROSITE, indicating that the both the protein might function similarly with respect to DEP domain mediated signaling, i.e. like DEPDC1B, DEPDC1A might also interact with G-protein coupled receptors and the negatively charged membrane phospholipids, resulting in the initiation of Wnt mediated signaling pathway [12, 13]. Out of the 84 amino acid residue DEP domain, 61 amino acid residues in the DEP domain of both the proteins, DEPDC1A and DEPDC1B, were found to be identical (71.8%) and 72 amino acid residues between the two protein being similar (84.7%) (Figure 3). The major difference in the protein sequences of the two proteins was only in the case of 13 amino acid residues, for instance, 16(K→S), 31(D→E) etc. (Figure 3), which may or may not affect the functionality of the domain and requires further research. Also, the presence of second domain, RhoGAP from the amino acid residue 201 to amino acid residue 393 in the protein DEPDC1B (Figure 2) and not in its paralog, DEPDC1A, indicates that only DEPDC1B is capable of mediating Rho GTPase signaling pathway [14], in addition to DEP domain stimulated signaling. The prediction of secondary structure of the two proteins was done using GOR IV. Due to the differences in the protein sequence of the two proteins, for instance the amino acid residue sequence from 100 to 103 was ‘PELR’ in case of DEPDC1A and ‘PNQK’ in case of DEPDC1B, resulting in the replacement of alpha helix in DEPDC1A with a random coil in case of DEPDC1B. The overall percentage variation observed in different secondary structures can be seen in Figure 4.

5. CONCLUSION

The two proteins under study, DEPDC1A and DEPDC1A play an important role in tumorigenesis as per the previously conducted studies, making them a potential biomarker, a therapeutic target and a diagnostic as well as prognostic marker. The present study focused on the better understanding of the

structural and functional characteristics of the two proteins, and therefore helping in the designing an effective and specific drug based on either of the two proteins, being implicated in various types of cancers. Even though the two proteins contains DEP domain, there exists structural difference which might result in functional differences, which must be investigated further.

REFERENCES

- [1] Seyfried TN and Shelton LM. Cancer as a Metabolic Disease. *Nutrition and Metabolism*. 7:7 (2010) [PMCID: PMC3941741]
- [2] Mendes RA. Oncogenic Pathways in the Development of Oral Cancer. *J Carcinogene Mutagene*. 3: 133 (2012). [doi: 10.4172/2157-2518.1000133]
- [3] Rob A. Cairns, et al. Regulation of Cancer Cell Metabolism. *Nature Reviews Cancer*. 11:85-95 (2011). [PMID: 21258394]
- [4] Kanehira M, et al. Involvement of upregulation of DEPDC1 (DEP domain containing 1) in bladder carcinogenesis. *Oncogene*. 26, 6448–6455 (2007). [PMID: 17452976]
- [5] Su et al. A putative novel protein, DEPDC1B, is overexpressed in oral cancer patients, and enhanced anchorage-independent growth in oral cancer cells that is mediated by Rac1 and ERK. *Journal of Biomedical Science*. 21:67 (2014). [PMID: 25091805]
- [6] Vial E, et al. ERK-MAPK signaling coordinately regulates activity of Rac1 and RhoA for tumor cell motility. *Cancer Cell*. 4:67–79 (2003). [PMID: 12892714]
- [7] Marchesi S, et al. DEPDC1B Coordinates De-adhesion Events and Cell Cycle Progression at Mitosis. *Developmental Cell*. 31: 420-433 (2014). [PMID:25458010]
- [8] Kassambara A, et al. Inhibition of DEPDC1A, a Bad Prognostic Marker in Multiple Myeloma, Delays Growth and Induces Mature Plasma Cell Markers in Malignant Plasma Cells. *PLoS ONE*. 8(4): e62752 (2013). [PMID:23646139]
- [9] Burchett SA. Regulators of G protein signaling. *J Neurochem*, 75:1335–1351 (2000). [Doi: 10.1046/j.1471-4159.2000.0751335.x]
- [10] Wharton KA Jr. Runnin' with the Dvl: proteins that associate with Dsh/Dvl and their significance to Wnt signal transduction. *Dev Biol*. 253:1–17 (2003). [PMID:12490194]
- [11] Wong HC, et al. Structural basis of the recognition of the dishevelled DEP domain in the Wnt signaling pathway. *Nat Struct Mol Biol*. 7:1178–1184 (2000). [PMID:11101902]
- [12] Sokol S. A role for Wnts in morpho-genesis and tissue polarity. *Nat Cell Biol*. 2: E124–E125 (2000). [PMID:10878822]
- [13] Ballon DR, et al. DEP-domain-mediated regulation of GPCR signaling responses. *Cell*. 126:1079–1093 (2006). [PMID:16990133]
- [14] Martemyanov KA, et al. The DEP domain determines subcellular targeting of the GTPase activating protein RGS9 in vivo. *J Neuro Sci*. 23:10175–10181 (2003). [PMID: 14614075]
- [15] Kretschmer C, et al. Identification of early molecular markers for breast cancer. *Mol Cancer*. 10: 15 (2011). [PMID:21314937]
- [16] Harada Y, et al. Cell permeable peptide DEPDC1-ZNF224 interferes with transcriptional repression and oncogenicity in bladder cancer cells. *Cancer Res*. 70: 5829–5839 (2010). [PMID:20587513]
- [17] Okayama H, et al. Identification of Genes Up-regulated in ALK-positive and EGFR/KRAS/ALK negative Lung Adenocarcinomas. *Cancer Res*. (2011). [PMID:22080568]
- [18] The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res*. 43: D204-D212 (2015). (<http://www.uniprot.org/>) [PMCID: PMC4384041]
- [19] Rice P, et al. EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*. 16(6): 276-7 (2000). (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) [PMID: 10827456]
- [20] Garnier J, et al. *Methods in Enzymology*. R.F. Doolittle Ed., Volume 266,540-553. (1996) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)
- [21] De Castro E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*. 34(Web Server issue):W362-5 (2006). (<http://prosite.expasy.org/>) [PMID:16845026]