ORIGINAL ARTICLE

An *In Silico* Strategy against Adam's Oliver Syndrome by Predicting RNAi Molecules against Dominant Genes by Suppressing Rho GTPases

S. F. Choragudi*, M. S. Ekklesia Sesham, Narasimha Vakkalagadda, Neeraj Krishna Vedantam, Devendranadh Reddy Janga, Dhanya Koneru, Angirekula Harisairam, Krishna Keerthika Oruganti

Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

Abstract

Aim: To analyse the In Silico Strategy against Adam's Oliver Syndrome by Predicting RNAi Molecules against Dominant Genes by Suppressing Rho GTPases. **Methods:** There are six genes known (ARHGAP31, DLL4, NOTCH1, RBPJ, DOCK6, and EOGT) reported so far, which can cause this syndrome. We retrieved CCDS and CDS sequences of each gene from the NCBI database and multiple sequence alignment was performed using ClustalΩ. After target identification, we chose ARHGAP31 as the target gene, as it is a major cause of cutis aplasia, also the regulator of Rho GTPase activity. siRNA and miRNA were designed using Invitrogen Block-iT. **Results:** The Proteins related to ARHGAP31 were obtained using Blastx. GC content analysis of ARHGAP31 using ENDMEMO was found to be 54.55%, and RNA-RNA interaction was interpreted using IntaRNA. Heat capacity, c = 17.88025 J/kg ºk, was analyzed using OligoCalc server and prediction of secondary structure was done using Modeller 9.22 and SAVES. The dope score, analyzed from the RC plot, for qseq4.B99990004 is –48,617.11328. The Z-score value was calculated using ProSA. The Z-score value for the modeled protein is 9.12. **Conclusion:** Study indicates that by suppressing Rho GTPase activity of ARHGAP31 gene, we can reduce the occurrence of AOS during early embryonic stages.

Key words: Aplasia cutis congenita, clustalΩ, inherited congenital disorder, ramachandran plot, RNA-RNA interactions, terminal transverse limb defects

INTRODUCTION

The Adams-Oliver syndrome (AOS) is a rare heterogeneous congenital incongruity interpreted by the presence of aplasia cutis congenita and with transverse terminal limb defects. Apart from these two, cutis marmorata and other anomalies include cardiovascular, respiratory, central nervous system, and orofacial defects were observed. These anomalies are caused due to mutations in six genes.[1] The genes identified were ARHGAP31, NOTCH1, EOGT, DOCK6, DLL4, and RBPJ.

ARHGAP31, located on chromosome 3, encodes a Rho GTPase-Activating Protein 31. Signaling by Rho GTPases and signaling by GPCR are among the associated pathways of this gene. This gene functions as an activating protein for RAC1 and CDC42. ARHGAP31 acts as a molecular switch that controls many aspects of cell activity through the mechanism of cycling between two conformational forms.

Address for correspondence:
S. F. Choragudi, Department of Biotechnology,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur - 522 502,
Andhra Pradesh, India.
E-mail: felicebt@kluniversity.in

Received: 30-04-2022
Revised: 08-08-2022
Accepted: 21:08:2022

Between the active and inactive cycles of gene are controlled by Cdc42/Rac1 and guanine nucleotide exchange factors pathways are crucial for development processes of limbs and scalp, and their signaling directly impacts on cell migration and proliferation in a cell-specific manner. Although CdGAPs function in vascular development is unknown, its mutation causes superficial vessel defects and subcutaneous edema.[2]

The single-pass transmembrane receptor protein is known as notch receptor. Signaling pathways stimulated by notch proteins play crucial role in the developmental events of many tissues all over the body, as well as the heart, bones, muscles, liver, and blood cells, amidst other anomalies. Mutations in Notch signaling have mainly depicted with bone and vascular disorders and cardiac malformations. Notch genes play a key role in determining cell fate. Notch pathway has four receptors, namely, NOTCH1, NOTCH2, NOTCH3, and NOTCH4. Mutations in NOTCH1 are the most common cause of AOS.

The EOGT gene translates to an enzyme, which acts in the lumen of the endoplasmic reticulum, by transferring a molecule called N-acetyl glucosamine to modify eukaryotic growth factor (EGF)-like domains, including the Notch receptors, in turn, regulating development signaling. Very few proteins are altered by EOGT protein which is known, but it is known that NOTCH proteins can be modified by EOGT.[3]

DOCK6 delivers instructions for a protein called guanine nucleotide exchange factor, which is known as activating proteins called GTPases. They play a significant role in cells by their chemical signaling. GTPases such as Cdc42 and Rac1 are activated through the presence of DOCK6 protein by the attached GDP which is exchanged from GTP. Once Cdc42 and Rac1 became active, the signals are transmitted which are crucial for various events of embryonic developments. This type of regulation of GTPases is due to DOCK6 protein in the developmental events of the skull, limbs, and heart. DOCK 6 conjointly performs a task within the development of fibers (axons) that stretch of nerve cells.[4]

MATERIALS AND METHODS

Retrieval of CCDS and CDS sequences

The fastA formats of CCDS and CDS sequences were retrieved from NCBI [https://www.ncbi.nlm.nih.gov/] with accession numbers – ARHGAP31 (NG_007665), DLL4 (NG_046974), NOTCH1 (NG_007458), RBPJ (NG_030343), DOCK6 (NG_031953), and EOGT (NG_042829).

Phylogenetic analysis

The fastA format CCDS and CDS sequences were retrieved and, with the help of ClustalΩ,[5] multiple sequence alignment was performed to find the evolutionary relationship.

Target identification

From the analysis of the phylogenetic tree obtained from ClustalΩ, ARHGPA31 gene was considered as the target gene due to its high functionality.

siRNA and miRNA designing

siRNA and miRNA were designed by taking CCDS sequences of the genes as input which is responsible for AOS with the help of Invitrogen BLOCK-iT by Thermo Fisher [https://rnaidesigner.thermofisher.com/rnaiexpress/].

Evaluation of result

Sequence retrieval

Protein BLAST was performed using BLASTx [https://blast.ncbi.nlm.nih.gov/Blast.cgi] and based on E values and percent identity hits were retrieved.

GC content

GC content of the nucleotide sequence of ARHGAP31 gene was calculated using the ENDMEMO [http://www.ENDMEMO.com/].

RNA-RNA interaction

RNA-RNA interaction was done with the help of IntaRNA[6] by taking miRNA sequences having less than 50% GC content against the target protein sequence.

Heat capacity

Using OligoCalc,[7] the heat capacity was determined for the sequence of interest. The formula used to calculate heat capacity

$$c=Q/m\,\Delta T$$

Where, Q= heat loss, Δ T= temperature, and m= mass

Prediction of secondary structure

3D structures of proteins were predicted using MODELLER 9.22 software[8] in which totally five structures were predicted based on dope scope; the structures in PDB[9] format were considered for further analysis.

Evaluation of PDB using RC plot

Energy was computed for the PDB files obtained from the Modeller 9.22 software. The energy values for the PDB files computed before and after energy minimization.

Based on energy value, the stable protein structure is finalized and analysis of the protein is done using SAVES server [HTTPS://SERVICESN.MBI.UCLA.EDU/SAVES/] and Ramachandran plot was obtained from the PROCHECK.[10]

RESULTS AND DISCUSSION

Retrieval of CCDS and CDS sequences

The FastA formats of CCDS and CDS of all the six genes involved in the Adams-Oliver syndrome are retrieved from NCBI and provided in Table 1.

CCDS of all genes

ARHGAP31

ATGAAGAACAAGGGTGCTAAGCAGAAGC
TGAAACGAAAGGGAGCCGCCAGC
GCGTTTGGCTGTGACCTGACGGAGTAT
CTGGAAAGCTCGGGACAGGATGTTCCATA
CG.GAGACCTCAACCAGCTGTTTTTACC
AGCCTCAGCGGAGATCAGTAATTCTGGATG
GAAGAAGTGGGAGGCAAATAGAATGA.

DDL4

ATGGCGGCAGCGTCCCGGAGCGCCTCTGG
CTGGGCGCTACTGCTGCTGGTGGCACT
TTGGCAGCAGCGCGCGGCCGGCTCCG
GCGTCTTCCAGCTGCCCAGTCTG
TGTGTTTGATATCAGAGGAGAGGAATGAA
TGTGTCATTGCCACGGAGGTATAA.

DOCK6

ATGGCTGCCTCCGAGCGCCGCGCCTTC
GCGCACAAGATCAACAGGACGGTGGCCG
CAGAGGTGCGGAAGCAGGTGTCC
CGGGAACGCAGTGGCTCCCCCCACTC
CAGCACCCCAGCTGATGGCACCCACC
CCACCCGGCCTCAGGAACTCC
TTGAACAGAGCAAGTTTCCGAA
AGGCAGACCTCTGA.

EOGT

ATGTTAATGTTGTTTGTCTTTGGAG
TCTTACTTCATGAAGTCTCACTGA
GTGGTCAGAATGAAGCTCCTCCT
AATACTCACAGCATTCCAGGCGAAC
CTCTGTATAACTATGCGACCAC
GTATTGCAACACCCAAAGTGGCCA
TTTAAGAAGAAACATGATGAGCTATAA.

NOTCH1

ATGCCGCCGCTCCTGGCGCCCCT
GCTCTGCCTGGCGCTGCTGCC
CGCGCTCGCCGCACGAGGCCC
GCGATCCGAGGGCGTCTCCAG
CCCTCCCACCAGCATGCAGTCCC
AGATCGCCCGCATTCCGGAGGCCTTC
AAGTAA.

RBPJ

ATGCGAAATTATTTAAAAGAGC
GAGGGGATCAAACAGTACTTA
TTCTTCATGCAAAAGTTGCACA
GAAGTCATATGGAAATGAAA
AAAGGTTTTTTTGCCCACCTCCT
ACAGCGAGGGAAGTTACACAAA
CGCCAGCACAATTCAACCA
GTGTCACATCATCTACAGCCA CAGTGGTATCCTAA.

Table 1: Genes and retrieved CDS accession no. with version

Gene name	ARHGAP31	DLL4	DOCK6	EOGT	NOTCH1	RBPJ
Accession and version	AB033030.1	AB036931.1	AB037816.2	AJ868234.1	AB209873.1	AK302230.1
	AK293726.1	AB043894.1	AK295664.1	AK091089.1	AF308602.1	AK303244.1
	BC112163.1	AF253468.1	AK316063.1	AK126187.1	AK000012.1	BC064976.1
	BC112165.1	AK313831.1	BC008335.1	AK290356.1	CR457221.1	BC020780.1
		AY358894.1	BC051330.1	AK294101.1	M73980.1	D14041.1
		BC106950.2	BC146786.1	AK304102.1		L07872.1
				BC028935.1		L07874.1
				BC060887.1		L07876.1
				BX640821.1		
				KC347596.1		

Phylogenetic tree analysis

Multiple sequence alignment was done to know the sequence homology between the genes, first, individual multiple sequence alignment was performed for the CDS sequence of each gene separately, and later, total multiple sequence alignment was done for all the genes, as shown in Figure 1.

Target identification

On phylogenetic analysis observation, ARHGPA31 is considered as the main target as mutations in ARHGAP31 which is the major cause of cutis aplasia (this condition was observed in almost all AOS cases) and this gene is also the regulator of Rho GTPase activity, especially during the forming of the heart skull and limbs in embryonic development stage.

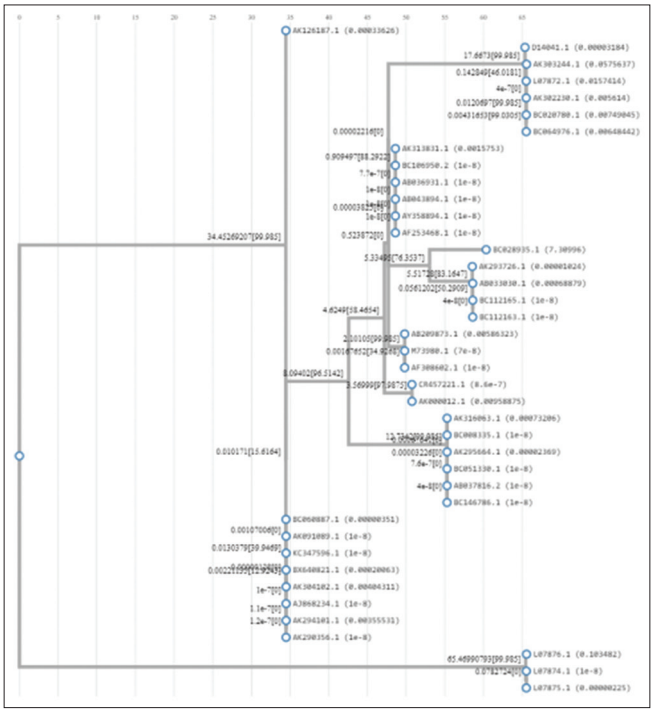

Figure 1: Complete phylogenetic tree analysis of all the six genes

siRNA and miRNA designing

Invitrogen BLOCK-iT by Thermo Fisher used for designing siRNA and miRNA as both regulate the gene expression, the sequences are tabulated in Tables 2 and 3.

Evaluation of result

Blastx

The sequence homology of the gene was detected by performing protein blast using Blastx. Protein homology had recognized that percent identity alone was an inferior method of differentiating homology among proteins compared, based on E values and percentage identity, maximum hits were retrieved.

Proteins obtained from blastx based on % identity

Gene: ARHGAP31.

Protein accession no

3IUG_A, 5C5S_A.

Percent identity

68.00%, 36.31.

E value

3e-90, 1e-20.

GC content

Using ENDMEMO, the GC content of ARHGAP31 gene was calculated P31 and the GC content of the ARHGAP31 gene is 54.55%.

Table 2: siRNA designed for ARHGAP31

S. No.	Start	Sequence	GC%	Tuschl's pattern match
1	534	CCTCAGGTCTAAAGAAATT	36.85	
2	882	GCGAAAGCTCTCCAGTAAA	47.37	B
3	1015	GCTACTATCCGACCAGCTA	52.64	1015
4	1778	CCTTGAGCTCTCAACATTT	42.11	1778
5	2036	CCAGCCCAATTCAGCCTAT	52.64	BD
6	2601	GGTTGAGATCGTCTCACAA	47.37	B
7	2652	GCCTTCAGACTGTGACGAA	52.64	B
8	2855	GCCATTCTCTAGATAGCAA	42.11	B
9	2971	GCACCCAGGAGAGAGATTA	52.64	
10	4030	CCTCAGAGCCTAATCTTAT	42.11	BD

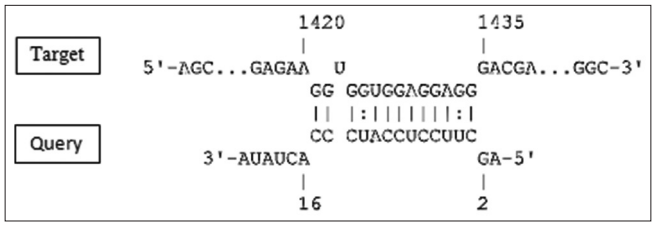

Figure 2: Selected interaction between target and query sequences of RNA1

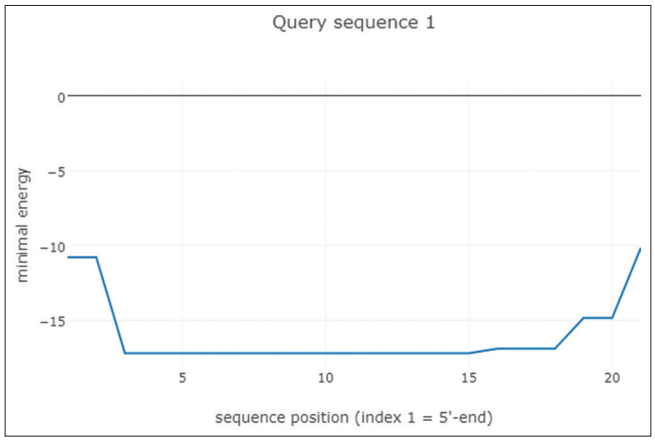

Figure 3: Position-wise minimal energy profile

Table 3: miRNA designed for ARHGAP31

S. No.	Start	Sequence	GC%
1	410	AGCTTCCTCCATCCCACTATA	47.62
2	880	AAGCGAAAGCTCTCCAGTAAA	42.86
3	938	CTGGATCAGACTCCAAATCAA	42.86
4	1079	CCAAGGGAAATTTCAATCGAA	38.1
5	2448	TCTCTACATAGACCAGCTGAA	42.86
6	2599	GAGGTTGAGATCGTCTCACAA	47.62
7	2845	CTTCGCCAGAGCCATTCTCTA	52.39
8	2943	GAGGAATTCTGACCCTCTTCA	47.62
9	2985	GATTACTGGATGGGATGAGAA	42.86
10	3727	ACTCAGAAACCTGCCAAAGAT	42.86

RNA-RNA interaction

RNA-RNA interactions analysis was done with the help of IntaRNA by taking miRNA against the target gene [Figure 2] sequence and the following results were obtained and tabulated in Table 4.

Values related to the selected interaction RNA1

Energy: –17.19620 kcal/mol [Figure 3]
Hybridization energy: 22.7 kcal/mol
Unfolding energy – Target: 5.53496
Unfolding energy – Query: 0.01086
Position – Target RNA: 1421–1434
Position – Query RNA: 3–15
Position seed – Target RNA: 1424–1430
Position seed – Query RNA: 7–13.

Heat capacity

Using OligoCalc, the heat capacity was determined for the sequence of interest using the following formula,

$$c=Q/m\ \Delta T$$

Q=37.526 cal; m=24123454.14g; ΔT=87 ºk.
c=17.88025 J/kg ºk

Prediction of secondary structure of protein

The secondary three-dimensional structure of proteins was speculated using MODELLER 9.22 software in which totally five structures were predicted. Based on dope scope, these structures in PDB format were considered for further analysis.

Evaluation of protein using RC plot

PDB flies of proteins were obtained using Swiss PDB software (five PDB files for each gene) that is previously obtained from the MODELLER 9.22 software; the total energy was computed before and after energy minimization.

Table 4: Identified interactions between the target and query sequences

Target	Position	Query	Position	Energy (kcal/mol)
ARHGAP31	1421–1434	RNA1	3–15	–17.19620
ARHGAP31	1597–1610	RNA8	7–20	–12.71700
ARHGAP31	1442–1457	RNA5	2–18	–11.45040
ARHGAP31	1253–1265	RNA9	2–15	–10.49060
ARHGAP31	520–535	RNA6	3–18	–9.20387
ARHGAP31	1468–1478	RNA2	9–19	–8.90014
ARHGAP31	1428–1441	RNA7	8–20	–7.88810
ARHGAP31	413–425	RNA4	4–15	–7.65874
ARHGAP31	178–194	RNA3	1–16	–7.33259

Table 5: The energy values for the PDB files computed before and after energy minimization

S. No.	PDB id	Total energy before energy minimization (k cal/mol)	Total energy after energy minimization (k cal/mol)	Difference in energy (k cal/mol)
1.	qseq1.B99990001	7,382,181	82,501	7,299,680
2.	qseq1.B99990002	11,336,477	102,977	11,233,500
3.	qseq1.B99990003	25,373,424	83,640.930	25,289,784
4.	qseq1.B99990004	43,927,516	89,997	43,837,519
5.	qseq1.B99990005	20,664,946	77,391	20,587,555

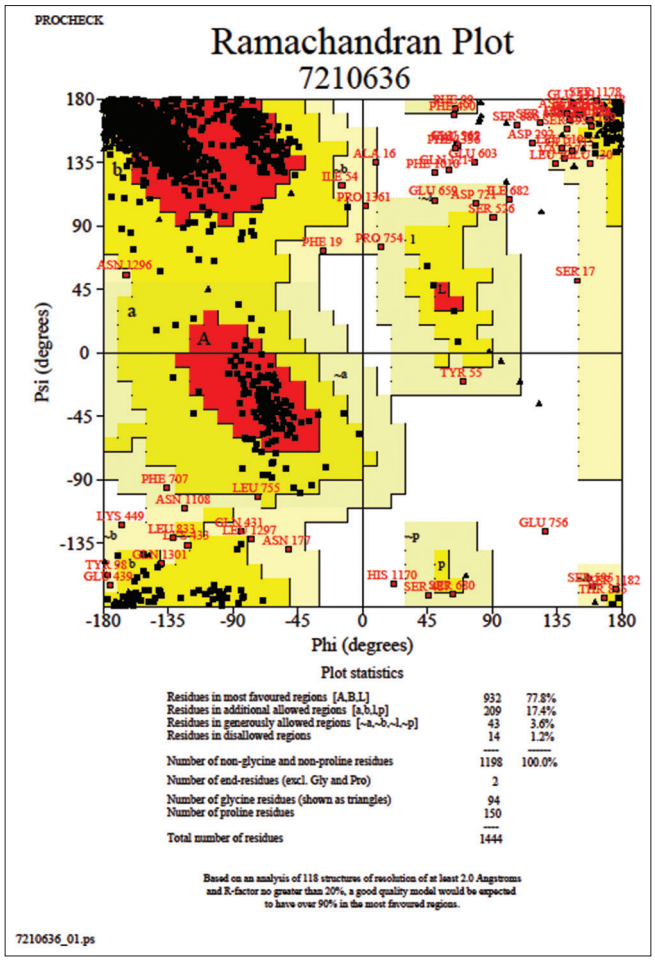

Figure 4: Ramachandran plot of ARHGAP31

Before and after energy minimization, the values were recorded for each protein using the same method, and finally, the PDB file which recorded greater energy difference value, before and after energy minimization, was selected. Ramachandran plot was obtained using SAVES-PROCHECK server as [Figure 4]and analysis was done using ProSA[11] and the data are tabulated in Table 5.

qseq4.B99990004 was considered as dope score was lower for this specific id than the other. The dope score for qseq4. B99990004 is –48,617.11328.

Ramachandran plot

The Z-score value was calculated using ProSA. The Z-score value for the modeled protein is 9.12, Figure 5.

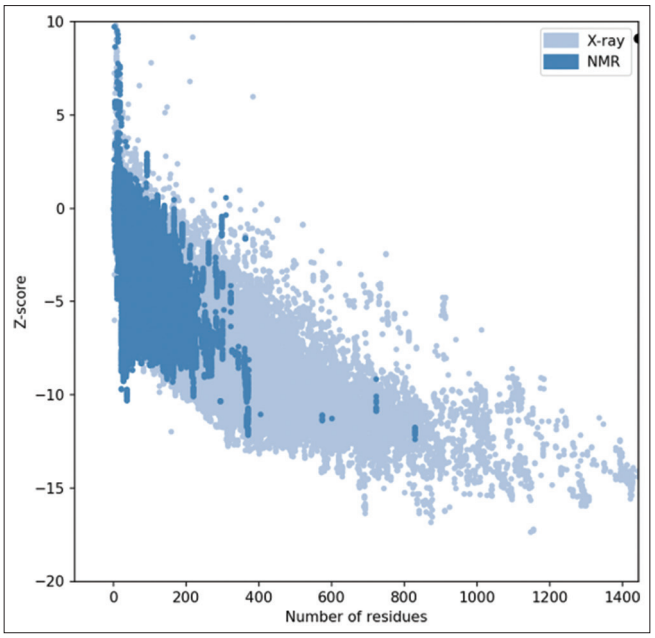

Figure 5: Number of residues targeted against Z-score with value 9.12

CONCLUSION

Adams-Oliver syndrome is an inherited congenital disorder, following either a dominant or recessive pattern. There are six genes known (ARHGAP31, DLL4, NOTCH1, RBPJ, DOCK6, and EOGT) reported so far, which can cause this syndrome. We retrieved CCDS and CDS sequences of each gene from the NCBI database and multiple sequence alignment was performed using ClustalΩ. After target identification, we chose ARHGAP31 as the target gene, as it is a major cause of cutis aplasia, also the regulator of Rho GTPase activity. siRNA and miRNA were designed using Invitrogen Block-iT. Proteins related to ARHGAP31 were obtained using Blastx. GC content analysis of ARHGAP31 using ENDMEMO was found to be 54.55%, and RNA-RNA interaction was interpreted using IntaRNA. Heat capacity, c = 17.88025 J/kg ºk, was analyzed using OligoCalc server and prediction of secondary structure was done using Modeller 9.22 and SAVES. The dope score, analyzed from the RC plot, for qseq4.B99990004 is –48,617.11328. The Z-score value was calculated using ProSA. The Z-score value for the modeled protein is 9.12. By suppressing Rho GTPase activity of ARHGAP31 gene, we can reduce the occurrence of AOS during early embryonic stages.

ACKNOWLEDGMENT

Authors thank the management of KL Educational Foundation, Vaddeswaram, for providing laboratory facilities for this work.

CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

AUTHORS' CONTRIBUTION

Dr. Choragudi S.F and Ekklesia Sesham M S have designed and directed through the project; Narasimha Vakkalagadda, Neeraj Krishna Vedantam, Devendranadh Reddy Janga, Dhanya Koneru, and Angirekula Harisairam and Krishna Keerthika Oruganti performed the experiment part and all together has analyzed the results obtained. Dr. Chorgudi S.F. and Ekklesia Sesham M S reverified the results and rectified the mistakes if any. Both authors read and approve the final version of the manuscript.

REFERENCES

1. Madan A, Sardana K, Garg VK. Adams oliver syndrome. Indian Pediatr 2015;52:633-4.
2. Caron C, DeGeer J, Fournier P, Duquette PM, Luangrath V, Ishii H, *et al*. CdGAP/ARHGAP31, a Cdc42/Rac1 GTPase regulator, is critical for vascular development and VEGF-mediated angiogenesis. Sci Rep 2016;6:27485.
3. Müller R, Jenny A, Stanley P. The EGF repeat-specific O-GlcNAc-transferase Eogt interacts with notch signaling and pyrimidine metabolism pathways in Drosophila. PloS one 2013;8:e62835.
4. Ekklesia Shesham MS, Sai Pushpa Sree V, Lakshmi V, Vidya Sree T, Sri Lalitha M. Adam's oliver syndrome. Int J Sci Tech Res 2019;8:2970-3.
5. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al*. Clustal W and clustal X version 2.0. Bioinformatics 2007;23:2947-8.
6. Mann M, Wright PR, Backofen R. IntaRNA 2.0: Enhanced and customizable prediction of RNA-RNA interactions. Nucleic Acid Res 2017;45:W435-9.
7. Kibbe WA. OligoCalc: An online oligonucleotide properties calculator. Nucleic Acid Res 2007;35:W43-6.
8. Webb B, Sali A. Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics 2016;54:5.6.1-37.
9. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The protein data bank. Nucleic Acid Res 2000;28:235-42.
10. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK-a program to check the stereochemical quality of protein structures. J App Crystal 1993;26:283-91.
11. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acid Res 2007;35:W407-10.

Source of Support: Nil. **Conflicts of Interest:** None declared.