ISSN: 2249 - 7862 (Print)
Aureole, Vol. XII, December 2020 PP 88-94
www.aureoleonline.in

HOMOLOGY MODELLING AND STRUCTURE EVALUATION OF CDC42 PROTIEN

* **Sarath Sasi**

** **Biju Thomas**

Abstract

Human immune deficiency virus infection/acquired immunodeficiency syndrome (HIV/AIDS) is a disease of human system caused by infection with human immunity. HIV targets the T-cells in the human cell, which enables the membranes to fuse and allow the virus to enter the cell deficiency virus. The HIV generated DNA using reverse transcriptase integrates with human DNA by the enzyme integrase that is present in the viral core. Interferon induced GTP binding protein CDC42 is a protein that in humans is encoded by the CDC42 gene. Only recently, the importance of CDC42 in the inhibition of HIV propagation within the cell got acceptance in the scientific community. Not much work has been done in this field. CDC42 suppress infection by cell HIV-1 strains tested has equivalent or reduced effects on divergent simian immunodeficiency viruses. The primary step towards better understanding of the protein function is to determine the structure of the protein. The model was generated using CPH3 server using the template sequence 1 grn.

Keywords **:** *Homology modeling, CDC42 protein, NCBI, GenBank, FASTA, Query sequence, CPH model 3.2 server, Stride server, Ramachandran plot, Errat server.*

* *Student, St. Stephen's College, Uzhavoor*

** *Assistant Professor, Dept. of Zoology, St. Stephen's College, Uzhavoor*

Introduction

AIDS Human immunodeficiency virus infection/acquired immunodeficiency syndrome (HIV/AIDS) is a disease of the human immune system caused by infection with human immunodeficiency virus (HIV).
AIDS was first recognized by the Centers for Disease Control and Prevention (CDC) in 1981 and its cause-HIV infection-was identified in the early part of the decade. Since its discovery, AIDS has caused an estimated 36 million death (as of 2012.As of 2012 approximately 35.3 million people are living with HIV globally. HIV/AIDS is considered a pandemic- a disease outbreaks which is present over a large area and is actively spreading.

Human CDC42 Protein

Human CDC42 is a small GTPase of the Rho family, which regulates signaling pathways that control diverse cellular functions including cell morphology, migration, endocytosis, and cell cycle progression. This protein is highly similar to Saccharomyces cerevisiae CDC42, and is able to complement the yeast CDC42-1 mutant. The product of oncogene Dbl was reported to specifically catalyze the dissociation of GDP from this protein. This protein could regulate actin polymerization through its direct binding to Neural Wiskott- Aldrich syndrome protein (N-WASP), which subsequently activates Arp2/3 complex. Alternative splicing of this gene results in multiple transcript variants.

Methodology

The particular segment is taken from HIV Human interaction database. It has clear relation with the HIV protein.

Model building using CPH models-3.0 web server

The FASTA sequence of CXCR4 is submitted to the server by pasting the sequences in the corresponding field of CPH models 3.2 server [45]. The following methodologies are followed in the modeling program.

Identification of template protein

A position-specific scoring matrix (PSSM) is generated for a query sequence by searching for up to five iterations with default settings, against a local version of the uniprot database using PsiBlast. The PSSM generated by Blast is saved and used to search for a template in PDB. The blast procedure identified CDC42 protein with PDB id [1 grn] as the most related protein for our query sequence. Next, the query is aligned to the template.

Modeling

Once the best template has been found, C?-atom coordinates are extracted according to the sequence alignment and used as a starting point for the homology-modeling process. Missing atoms are added using the segmod program [30] and the structure was refined using the encad program, both [31]. From the Gene Mine package (www.bioinformatics.ucla.edu/genemine/).

Structure Evaluation

The correctness of the predicted structures is evaluated using verify 3d server and stride server. Ramachandran plot has been generated to check the validity of the structure.

Results and Discussions

The structure of the protein CDC42 of humans was modeled using CPHmodels-3.0 web server. The protein FASTA sequence of the protein was taken from GenBank (http://www.ncbi.nlm.nih.gov/genbank/). The overall protein sequence identity between them was around 94.7% and an E-value of 1e-109.This indicates that the template structure is best selection for modeling our target protein sequence and that the alignment has not occurred by random chance alone. Prior to the modeling, the server made an alignment of both the target and template protein sequences for identification of structurally conserved and variable regions between them. The alignments are given in the Figure 1

```
Entry: 1grn
Chain: A

Making profile-profile alignment ...

Score: 500.5 bits
Identity: 100.0 %

Query: 1   MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAG 60
           MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAG
Templ: 1   MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAG 60

Query: 61  QEDYDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLR 120
           QEDYDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLR
Templ: 61  QEDYDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLR 120

Query: 121 DDPSTIEKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALEPP 180
           DDPSTIEKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALEPP
Templ: 121 DDPSTIEKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALEPP 180

Query: 181 EPKKSRRCVLL 191
           EPKKSRRCVLL
Templ: 181 EPKKSRRCVLL 191
```

Figure 1: The alignment between the sequence of CDC42 sequence (query) and the sequence of template.

Since the identity between the two proteins is above 94.7%, homology modeling of our target protein is not an issue here.
The strand structure of CDC42 protein thus generated through homology modeling is shown in figure 2.

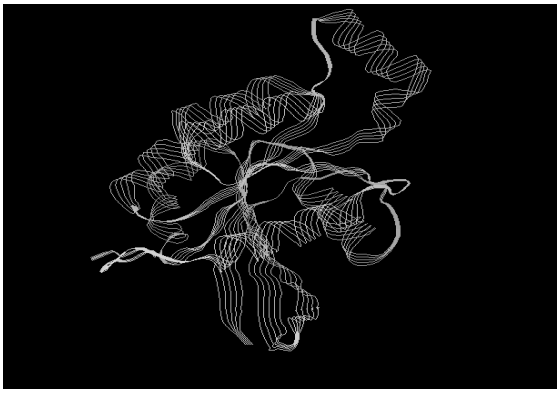

Figure 2: Figure showing the modeled structure of human cdc42 protein.

Secondary structural analysis of the protein was done at the stride server. The server is a very efficient tool for the assignment of secondary structure from the atomic coordinates of protein. It produces a visual output of various secondary structural features occupied by all the residues in the proteins in a diagrammatic view (figure 3)

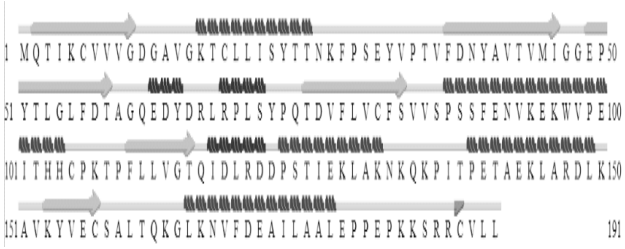

Figure 3: Secondary structure analysis of modeled humanCDC42 protein

The structure shows that there are 5 alpha helix,6 beta helix and 16 coiled (loop) regions.The 3-dimensinal arrangement of secondary structures of human CDC42 protein is given in figure 4.

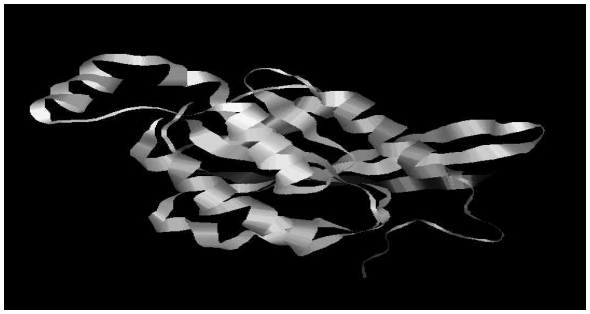

Figure 4: Three dimensional arrangement of different secondary structures in the modeled human CDC42proteins.

Stereochemical quality of the modeled human CDC42 protein is calculated using the verify 3D [47] and Ramachandran Plot analysis [48]. Ramachandran plot analysis was employed to evaluate the stereo chemical quality of the protein.94.7% residues were in most favored regions. 4.8% residues in generously allowed regions. Only 0.5%are in the disallowed regions. The structure also shows trans-membrane and nuclear regions very clearly.

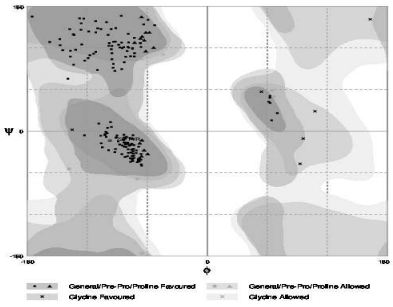

Figure 5: Ramachandran plot analysis data of CDC42 protein.

Evaluation of residues

```
Residue [   12 :GLY] ( -43.99, 110.57) in Allowed region
Residue [   36 :VAL] ( -69.91, -68.81) in Allowed region
Residue [   54 :GLY] ( -91.64,  96.03) in Allowed region
Residue [   59 :ALA] ( -65.34,-176.12) in Allowed region
Residue [   96 :LYS] (-138.17, -54.54) in Allowed region
Residue [  114 :GLY] ( -90.03,  99.45) in Allowed region
Residue [  116 :GLN] (  64.67,  58.72) in Allowed region
Residue [  150 :LYS] (  66.56,  49.81) in Allowed region
Residue [  188 :CYS] ( -66.21, 104.56) in Allowed region
Residue [   72 :TYR] ( -97.29, -13.65) in Outlier region
Number of residues in favoured region    (~98.0% expected)  :  179 ( 94.7%)
Number of residues in allowed region     ( ~2.0% expected)  :    9 (  4.8%)
Number of residues in outlier region                        :    1 (  0.5%)
```

Verify 3D analysis determines the compatibility of an atomic model (3D) with its own amino acid sequence (ID) by assigned a structural class based on its location and environment (alpha, beta, loop, polar, non-polar etc.) and comparing the results to good structures. The result indicates that the modeled CDC42 protein confines to allowed regions of the evaluation criteria (figure 6).

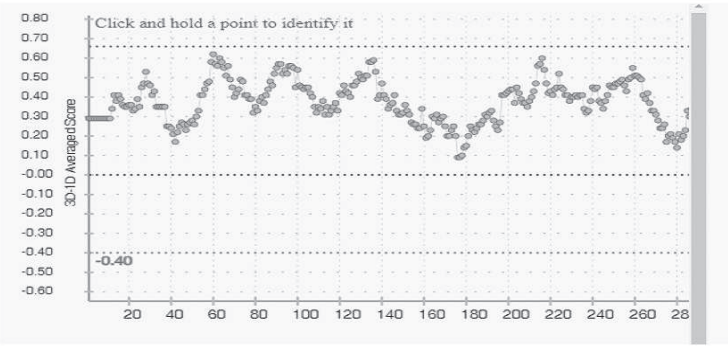

Figure 6: Verify 3D analysis data of humanCDC42 protein

The correctness of a proteins structure depends on satisfying the basic bio-physical parameters associated with the protein folding process such as correct bond lengths, bond angles, tetrahedral angles, phi angles, psi angles etc. Stereo chemical analysis can provide the extend of correctness of our modeled protein structure, thereby enabling us to the selection of the model as optimal structure for further analysis. Stereo chemical analysis by verify 3D followed by Ramachandran Plot analysis are in tune with the features of a correct macromolecular structural features.

Conclusion

Human CDC42 protein was modeled by CPH3 server using 1 grn as template. The overall protein sequence identity between the protein model and template was around 94.7% and an E-value of 1e-109. Secondary structure prediction using stride server identified 5 alpha helices, 6 beta strands and 16 coiled (loop) regions in the modeled human CDC42 protein. Ramachandran plot analysis was employed to evaluate the stereo chemical quality of the protein. 94.7% residues were in most favored regions, 4.8% residues in generously allowed regions, and only 0.5% are in the disallowed regions. The result indicates that the modeled CDC42 protein confines to allowed regions of the evaluation criteria. The generated structure can be used for further studies in future.

References

1. Sepkowitz KA (June 2001). "AIDS- the first 20 years". N. Engl. J. Med.344 (23): 1764-72
2. "Fact sheet".UNAIDS.org.2013.Retrieved December 4, 2013.
3. Kallings LO (2008). "The first postmodern pandemic: 25 years of HIV/AIDS". Journal of Internal Medicine 263 (3):218-43
4. Markowitz, edited by William N. Rom; associate editor, Steven B. (2007). Environmental and occupational medicine (4th ed.).Philadelphia: Kluwer/ Lippincott Williams& Wilkins.
5. "HIV and its transmission". Centers for Disease Control and Prevention.2003. Archived from the original on February 4, 2005. Retrieved may 23, 2006.
6. Sharp, PM; Hahn, BH (September 2011)."Origins of HIV and the AIDS pandemic". Cold spring Harbor perspectives in medicine a006841.doi:10.1101/ cshperspect.a006841
7. Montagnier, Luc. (1999) Human Immunodeficiency Viruses (Retroviridae). Encyclopedia of virology 763-774.
8. Zheng YH, Lovsin N, Peterlin BM (2005). "Newly identified host factors modulate HIV replication". Immunol.Lett.97 (2):225-34.
9. Hu WS, Temin HM. (1990). Retroviral recombination and reverse transcription.Science 250(4985):1227-1233.
10. Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ. (2006). Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. J Virol 80 (5):2472-2482.
11. Gelderblom HR (1997). " Fine structure of HIV and SIV ". In Los Alamos National Laboratory. HIV sequence compendium. Los Alamos National Laboratory. pp. 31-44.
12. Cocchi, F., A. L. De Vico, A. Garzino-Demo, S.K. Arya, R.C. Gallo, and P. Lusso. 1995. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV suppressive factors produced by CD8+ T cells. Science 270:1811-1815. Human MX2 is an interferon- induced post- entry inhibitor of HIV-1 Infection.