Advances in Life Science and Technology
ISSN 2224-7181 (Paper) ISSN 2225-062X (Online)
Vol 6, 2012

Computational analysis of *M.tuberculosis* - CarD protein

V.G.Shanmuga Priya [1*] Dr.U.M.Muddapur[1] Megha Mehta [2]

1*,2 .Assistant Professor ,Department of Biotechnology,KLE'S College of Engineering and Technology, Belgaum-8, Karnataka, India

1. Associate Professor, Department of Biotechnology,KLE'S College of Engineering and Technology, Belgaum-8, Karnataka, India

*E-mail of the corresponding author: priyaabhi2000@yahoo.com, priyaabhirajan@gmail.com

Abstract

CarD protein which regulates rRNA transcription by interacting with RNA polymerase (RNAP), seems to be essential for the survival of *M.tuberculosis*. TRCF, a protein involved in DNA repair process has same type of interaction with RNAP but found not essential for the organism's survival.

In this study, computational analysis of CarD sequence retrieved from NCBI and comparative study of CarD and TRCF sequences are carried out using bioinformatics tools. ProtParam tool analysis of CarD sequence predicts the molecular weight to be 17 Kda and pI to be 5.49.The absorbance at the concentration of 0.1% in water at 280nm is 0.947.The instability index indicates that the protein will be stable in the test tube. Blastp tool search against database Human proteins supports a view that CarD can be used as a drug target as it has no similar proteins in humans. SMART tool predicted the presence of a single domain CarD_TRCF in CarD protein. STRING database is queried to get the possible interacting proteins for both CarD and TRCF. The comparative analysis of the results speculates that CarD's association with proteins of Isoprenoid pathway may be responsible for the indispensability of the protein to the organism.

Keywords: CarD, TRCF, ProtParam, SMART, STRING.

1. Introduction

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* is an ancient infection that has plagued humans throughout recorded and archeological history. Despite the availability of effective chemotherapy and a moderately protective vaccine, recently 8.8 million people fell ill with TB and 1.4 million died from TB (WHO, Fact sheet N 104, March 2012).This health crisis is exacerbated by the emergence of MDR, XDR and TDR strains and this has urged the necessity of finding new drug targets and new anti-tuberculosis agents to decrease the global incidence of tuberculosis disease.

M.tuberculosis in host faces a stress environment of host derived mutagens, starvation of nutrients and oxygen. To persist in this hostile environment *M.tuberculosis* enters a state of dormancy and rapidly down regulates ribosome biogenesis to match declining translational need, a response that requires coordinate transcriptional regulation of all ribosome components.(Betts et al.,2002).Bacteria accomplish this via the stringent response,a global regulatory mechanism in which transcription of stable RNAs is inhibited,in part by the production of the hyperphosphorylated guanine nucleotides ppGpp and pppGpp(Avarbock et al., 2000;Chatterji and Ojha.,2001;Magnusson et al.,2005).DksA protein potentiates the effect of (p)ppGpp by directly binding the RNAP(Paul et al.,2004). DksA homologs are absent in some divisions of bacteria, whereas (p)ppGpp sythetases are broadly distributed. Recent work has identified CarD protein which interacts with RNAP β subunit to control rRNA transcription in *Mycobacterium tuberculosis*.It was found that CarD is essential for viability, in that declining carD transcript levels directly correlated with cell death. CarD depletion leads to an upregulation of 16s rRNA and rpsH ribosomal protein transcripts in the organism.Mutant cells lacking CarD were found to be highly susceptible to antibiotics,when compared to wild strains and hence CarD is suggested to be a potential drug target .(Stallings et al.,2009).

TRCF (Transcription repair coupling factor) is necessary for DNA strand-specific repair during transcription. A lesion in the template strand blocks the RNA polymerase complex. The RNAP-DNA-RNA complex is specifically

recognized by TRCF which releases RNAP and the truncated transcript. The TRCF may replace RNAP at the lesion site and then recruit the uvrA/B/C repair system (Westblade et al.,2010;Chambers et al.,2003).

CarD proteins interacts with the N terminus of the RNAP β subunit, in a binding site shared by transcription repair coupling factor (TRCF).Though the degree of interaction seem to be the same for both, TRCF is found not essential in *M.tuberculosis*. Attempts to complement CarD depletion with TRCF were unsuccessful, indicating that RNAP binding is not sufficient for CarD activity. Other proteins which interact with CarD can give a clue for its essentiability. (Stallings et al., 2009).

In this study, CarD sequence is retrieved from database and is analysed with bioinformatics tools (Pradeep et al.,2012) and the comparative study between CarD and TRCF is carried out using computational tools and the possible reasons why CarD is essential for the survival of the organism is discussed.

2. Materials and Methods

2.1 Sequence retrieval

Sequence of CarD and TRCF of Mycobacterium tuberculosis were retrieved from NCBI protein database through entrez search engine (http://www.ncbi.nlm.nih.gov).

2.2 *Characterisation of physico-chemical properties*

ProtParam tool under Swiss Institute of Bioinformatics, computes various physico- chemical properties that can be detected from a protein sequence.The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).(Gasteiger et al., 2003). CarD sequence is submitted to this tool for characterisation.

2.3 *Finding homologs in Human*

BLAST tool, which is under NCBI, finds regions of local similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches (Altschul et al, 1990).Blastp program searches for similar sequences for the query protein sequence in protein databases. Search was made with CarD sequence of M.tuberculosis against Human proteins in databases and run with default parameters.

2.4 *Domain Analysis*

SMART (a Simple Modular Architecture Research Tool) allows the identification and annotation of genetically mobile domains and the analysis of domain architectures. More than 500 domain families found in signalling, extracellular and chromatin-associated proteins are detectable (Schultz et al., 1998). CarD and TRCF sequence is submitted to this tool for domain analysis.

2.5 *To predict interacting Proteins*

STRING currently holds 730 000 proteins in 180 fully sequenced organisms. It is a database of known and predicted protein interactions. The interaction include direct (physical) and indirect (functional) associations. They are derived from four sources: Genomic context, high throughput experiments, co expression and already known association as in literature etc. (Szklarczyk et al.,2011). To get the possible functionally associated proteins with both CarD and TRCF of *M.tuberculosis*, STRING database was queried. To increase the confidence level, CarD and TRCF sequences of other related mycobacteria- *Mycobacterium smegmatis, Mycobacterium bovis* and *Mycobacterium avium* were also submitted to STRING.

3. Results and Discussion

3.1 *Sequence from NCBI*

CarD Sequence retrieved from NCBI protein database has the accession number GAA43479 and GenBank Id 15610719 and was under the name transcription factor. Also in Uniprot KB database the protein is present as possible transcription protein with Id: 053568.It was confirmed as CarD from literature study. The CarD protein has 162 amino acids.

TRCF Sequence retrieved from NCBI protein database from the accession number CAB06859 and GenBank Id 3261715. It has 1234 amino acids.

3.2 *ProtParam tool*

Some important parameters known from the ProtParam tool are, the molecular weight is 17907.3 Dalton and Theoretical pI is 5.49.The Protein was found to contain more of alanine,Valine and glutamic acid residues. The absorbance of this protein at 280 nm, with the concentration of 0.1% in water was calculated to be 0.947. The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell, for this protein it is found to be less than 10 hours in Bacterial organism.The instability index of the protein was calculated to be 28.89, which indicates that the protein will be stable in the test tube. The average hydropathy value of the amino acids in the protein was found to be -0.313 indicating that the protein is hydrophilic in nature. These parameters will be applicable in wet lab experiments.

3.3 *Blastp*

Blastp search done against Human protein database has given 4 hits with very less score (Table-1).The aligned region between the query and hits is less and with high E-value. This confirms the absence of highly similar proteins in humans.The hits were also analysed and found that they do not possess the domain CarD_TRCF which is present in CarD protein.

3.4 *SMART Tool*

Smart tool analysis shows the presence of CarD_TRCF domain from 2-112 residues on the CarD protein with e-value 8.62e-33.The lower e-value increases prediction confidence. TRCF protein has found to have 4 domains- CarD_TRCF, Dead-like helicase superfamily, helicase superfamily C-terminal domain,TRCF domain. The domain CarD_TRCF present in both the proteins verify the similar mechanism of interaction with RNA polymerase.The other 3 domains present only in TRCF protein indicates the multiple functions carried out by the protein.

3.5 *STRING database*

The proteins interacting with CarD and TRCF of mycobacterium species got from STRING database were analyzed and their consensus taken and tabulated(Table 2 & 3).The important predictions got from this are- two proteins 2-C-methyl –D-erythritol 2,4 –cyclodiphosphate synthase and 2-c-methyl-D-erythritol 4-phosphate cytidylyltransferase both of which are involved in isoprenoid biosynthesis pathway and are already recognized as drug targets (HAMAP MF 00107 & HAMAP MF 00041 respectively) were found to be functionally associated with CarD (based on its genomic neighborhood) and not with TRCF.As Isoprenoids are essential for the survival of the organism (Heuston et al,2012;Campbell et al,2002;Beutow et al,2007)this gives a clue of possible function of CarD ,other than RNAP interaction ,which makes it essential for the survival of the organism, as when compared with TRCF.

DNA directed RNA polymerase subunit beta and 50 s ribosomal protein L25 are common proteins who have functional association with CarD and TRCF relates to their common function. Other proteins with which these two proteins are associated play a role in carrying out their specific function.

4. Conclusion

CarD protein is found to be essential for the survival of M.tuberculosis inside the host as it controls rRNA transcription level by interacting with RNA polymerase. Also, TRCF involved in DNA repair interacts similarly with RNA polymerase to carry out its function but is unable to control rRNA transcription and found not essential for survival. Taking the suspicion that CarD might have some other unique function and it also can be used as a drug target, a study is carried out to analyse these things computationally.

CarD protein sequence is retrieved from NCBI with accession number GAA43479 and length of 162 a.a. ProtParam tool which analyses the physico-chemical characters of a protein from its sequence predicted the molecular weight to be 17 KDa and pI to be 5.49. The absorbance of this protein at 280 nm, in the concentration of 0.1% in water was calculated to be 0.947. The instability index of the protein was calculated to be 28.89,which indicates that the protein will be stable in the test tube. The average hydropathy value of the amino acids in the protein was found to be -0.313 indicating that the protein is hydrophilic in nature, etc. These details will help to handle the protein in wetlab experiments.

Blastp search against Human proteins in databases gave 4 hits which has less sequence similarity to the CarD protein and has no similar domain in them as in CarD. This holds up the possibility of using CarD as a potential drug target.

Both CarD and TRCF have CarD_TRCF domain which explains their similar type of interaction with RNAP. The accession number of it is SM01058 in SMART database. TRCF has 3 more domains which notifies its role in DNA repair process.

Proteins predicted to have interaction with CarD and TRCF are known from STRING database, which predicts physically and functionally associated proteins of a query protein, based on genomic context association, expression and protein interaction in text available. Analysis of the results show that 2-C-methyl –D-erythritol 2,4 –cyclodiphosphate synthase and 2-c-methyl-D-erythritol 4-phosphate cytidylyltransferase both of which are involved in isoprenoid biosynthesis pathway are found to be associated with CarD based on gene position neighborhood in mycobacterial genomes. This suggests that this functional relationship between them makes CarD an essential protein.

CarD proteins specific mode of functional relation with isoprenoid biosynthesis pathway proteins has to be detected further and in future work, drug candidates targeting CarD needs to be screened or built insilico and has to be tested with wetlab experiments.

References

Altschul, S; Gish, W; Miller, W; Myers, E; Lipman, D (October 1990). "Basic local alignment search tool". *Journal of Molecular Biology* **215** (3): 403–410. DOI:10.1016/S0022-2836(05)80360-2. PMID 2231712

Avarbock, D., Avarbock, A., and Rubin, H. (2000).Differential regulation of opposing RelMtb activities by the aminoacylation state of a tRNA.ribosome. mRNA.RelMtb complex". Biochemistry 39, 11640–11648. http://dx.doi.org/10.1021/bi001256k, PMid:10995231

Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., and Duncan, K. (2002).Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol. Microbiol. 43,717–731.http://dx.doi.org/10.1046/j.1365-2958.2002.02779.x,PMid:11929527

Buetow L, Brown AC, Parish T, Hunter WN(2007), The structure of Mycobacteria 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase, an essential enzyme, provides a platform for drug discovery.BMC Struct biol. 2007;7:68 http://dx.doi.org/10.1186/1472-6807-7-68,PMid:17956607 PMCid:2151065

Chatterji, D., and Ojha, A.K. (2001). Revisiting the stringent response, ppGpp and starvation signaling. Curr. Opin. Microbiol. 4, 160–165.http://dx.doi.org/10.1016/S1369-5274(00)00182-X

Campbell TL, Brown ED. Characterization of the depletion of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase in Escherichia coli and Bacillus subtilis. J. Bacteriol. 2002;184:5609-5618. http://dx.doi.org/10.1128/JB.184.20.5609-5618.2002, PMid:12270818 PMCid:139617

Chambers, A.L., Smith, A.J., and Savery, N.J. (2003). A DNA translocation motif in the bacterial transcription-repair coupling factor, Mfd. Nucleic Acids Res. 31, 6409–6418
http://dx.doi.org/10.1093/nar/gkg868, PMid:14602898 PMCid:275562

Christina L. Stallings,Nicolas C. Stephanou,Linda Chu,Ann Hochschild,Bryce E. Nickels, and Michael S. Glickman(2009).CarD Is an Essential Regulator of rRNA Transcription Required for Mycobacterium tuberculosis Persistence.Cell.138,146-159.

http://dx.doi.org/10.1016/j.cell.2009.04.041, PMid:19596241 PMCid:2756155

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis.Nucleic Acids Res. 31:3784-3788(2003).
http://dx.doi.org/10.1093/nar/gkg563, PMid:12824418 PMCid:168970

Heuston S, Begley M, Gahan CG, Hill C.(2012),Isoprenoid biosynthesis in bacterial pathogens.
Microbiology. 2012 Jun; 158(Pt 6):1389-401. Epub 2012 Mar 30.PMid: 22466083

Jörg Schultz,Frank Milpetz,Peer Bork, and Chris P. Ponting(May 1998),SMART, a simple modular architecture research tool: Identification of signaling domains. PNAS May 26, 1998 vol. 95 no. 11, 5857-5864

Magnusson, L.U., Farewell, A., and Nystrom, T. (2005). ppGpp: a global regulator in Escherichia coli. Trends Microbiol. 13, 236–242.http://dx.doi.org/10.1016/j.tim.2005.03.008 ,PMid:15866041

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W., and Gourse, R.L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell 118, 311–322.,http://dx.doi.org/10.1016/j.cell.2004.07.009,PMid:15294157.

Pradeep N V, Anupama, Ankitha K, Pooja J(2012). Categorizing Phenomenal features of α-Amylase (Bacillus species) using Bioinformatic Tools.Advances in life science and Technology,Vol 4 ,27-34.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C.(Jan 2011),The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.Nucleic Acids Res. 39(Database issue):D561-8. Epub 2010 Nov 2.http://dx.doi.org/10.1093/nar/gkq973, PMid:21045058 PMCid:3013807

Westblade LF, Campbell EA, Pukhrambam C, Padovan JC, Nickels BE, Lamour V, Darst SA.(Dec 2010),Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. Nucleic Acids Res.;38(22):8357-69. Epub 2010 .http://dx.doi.org/10.1093/nar/gkq692,PMid:20702425 PMCid:3001067

V.G.Shanmuga Priya from Vellore,Tamil Nadu, India has done B.Sc (zoology) at Auxilium college, Vellore under Madras University and secured University First Rank(1995) and M.Sc(Bioinformatics) at Auxilium College under Thiruvalluvar University and secured University Third Rank(2005) and M.Phil(Bioinformatics) at Bharathidasan University, Tamil Nadu(2007). Is Presently working as Assistant Professor in Dept of Biotechnology, KLE'S College of Engineering and Technology, Belgaum, Karnataka, India.

Dr.U.M.Muddapur from Belgaum, Karnataka has done B.Sc, M.Sc(1999) and PhD(2004) –all in Biochemistry, from Karnataka University,Dharwad.,Karnataka. Is Presently working as Associate Professor in Dept of Biotechnology, KLE'S College of Engineering and Technology, Belgaum, Karnataka, India.Is a life member of ISTE and SBC societies.

Megha Mehta from Bilaspur, Chhatisgarh,India has done B.Sc(Microbiology) in the year 2000 & M.Sc(Biotechnolology) in the year 2003 from Guru Ghasidas university, Bilaspur Chhastisgarh. Is Presently working as Assistant Professor in Dept of Biotechnology, KLE'S College of Engineering and Technology, Belgaum, Karnataka, India.

Table 1-blastp hits for CarD sequence against Human proteins in databases.

S.No	Human proteins	Score	Query Coverage	E value	Maximum identity
1.	pantothenate kinase 2, mitochondrial isoform 1 preproprotein	29.6	35%	1.3	29%
2.	plexin domain-containing protein 2 precursor	28.9	26%	2.0	37%
3.	importin-9	27.7	33%	5.0	27%
4.	telomeric repeat-binding factor	27.7	33%	5.7	26%

Table 2 – STRING –predicted proteins interacting with CarD

Based on Criteria of	**CarD associated proteins** **(in the order of decreasing score)**
Neighborhood	**2-C-methyl –D-erythritol 2,4 –cyclodiphosphate synthase**
Neighborhood	**2-c-methyl-D-erythritol 4-phosphate cytidylyltransferase**
Neighborhood	Cysteinyl –tRNA synthetase
Neighborhood+ coexpression + experiments	***DNA-directed RNA polymerase subunit beta***
Neighborhood	tRNA/rRNA methyltranferase
Neighborhood + Cooccurrence	***50S ribosomal protein L25***
Neighborhood	DNA repair protein RadA
Neighborhood	phosphoglyceromutase
Neighborhood	Glutamyl-tRNA synthetase
Neighborhood + Cooccurrence	DNA integrity scanning protein DisA
Neighborhood	S-adenosylmethionine synthetase
Neighborhood	Putative DNA-binding/iron metalloprotein/AP endonuclease
Neighborhood	hydrolase
Neighborhood	Glycerophophoryl diester phosphodiesterase family protein

The essential proteins interacting with CarD are highlighted in Bold and proteins which also interact with TRCF are highlighted in Bold italics

Table 3- STRING –predicted proteins interacting with TRCF

Based on Criteria of	**TRCF associated proteins** (in the order of decreasing score)
Neighborhood	Nucleoside triphosphate pyrophosphohydrolase
Neighborhood+ Experiments+ textmining	Excinuclease ABC subunitA
Neighborhood+ Cooccurrence	Alanyl-tRNA synthetase
Neighborhood	Peptidyl –tRNA hydrolase
Neighborhood	UDP-N-acetylglucosamine pyrophosphorylase GlmU
Neighborhood+Experiments+text mining	***DNA- directed RNA polymerase subunit beta***
Neighborhood	Ribose-phosphate pyrophosphokinase
Neighborhood+ Cooccurrence	GTP-binding protein EngA
Neighborhood+ Cooccurrence	***50s ribosomal protein L25***/general stress protein Ctc.
Neighborhood	Putative phosphoserine phosphatase SerB
Neighborhood+Coexpression	GTP-dependent nucleic acid-binding protein EngD
Neighborhood	Translation-associated GTPase
Cooccurance + textmining	Excinuclease ABC subunit C
Neighborhood	folypolyglutamate synthase protein FolC
Cooccurance	Phosphopantetheine adenylyltransferase
Neighborhood	Transcriptional regulator,TetR family protein

Proteins which interact also with CarD are highlighted in Bold, italics.