Computational Biology and Chemistry 66 (2017) 11–20

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

Insights into structure and function of *30S Ribosomal Protein S2* (*30S2*) in *Chlamydophila pneumoniae:* A potent target of pneumonia

G. Koteswara Reddy[a,*], K. Nagamalleswara Rao[b], Kiran Yarrakula[a]

[a] Centre for Disaster Mitigation and Management, VIT University, Vellore-632014, India
[b] School of Civil and Chemical Engineering, VIT University, Vellore-632014, India

ARTICLE INFO

Article history:
Received 14 August 2016
Received in revised form 4 October 2016
Accepted 29 October 2016
Available online 9 November 2016

Keywords:
Chlamydophila pneumoniae
30S ribosomal protein S2
Pneumonia
Drug target
Structure
Function
Ramachandran plot
Profunc server

ABSTRACT

The gene *30S ribosomal protein S2* (*30S2*) is identified as a potential drug and vaccine target for *Pneumonia.* Its structural characterization is an important to understand the mechanism of action for identifying its receptor and/or other binding partners. The comparative genomics and proteomics studies are useful for structural characterization of *30S2* in *C. Pneumoniae* using different bioinformatics tools and web servers. In this study, the protein *30S2* structure was modelled and validated by Ramachandran plot. It is found that the modelled protein under most favoured "core" region was 88.7% and overall G-factor statistics with average score was −0.20. However, seven sequential motifs have been identified for *30S2* with reference codes (*PR0095, PF0038, TIGR01012, PTHR11489, SSF52313* and *PTHR11489).* In addition, seven structural highly conserved residues have been identified in the large cleft are *Lys160, Gly161* and *Arg162* with volume 1288.83 Å^3 and average depth of the cleft was 10.75 Å. Moreover, biological functions, biochemical process and structural constituents of ribosome are also explored. The study will be helped us to understand the sequential, structural, functional and evolutionary clues of unknown proteins available in *C. Pneumoniae.*

1. Introduction

The comparative genomics and proteomics studies are useful for understanding the sequential, structural, functional and evolutionary clues for pathogenic proteins using different bioinformatics tools and web servers (Piplani et al., 2016; Thakare et al., 2016). The gene *30S2* is responsible for translation and post medication in the cellular synthesis of all *C. Pneumoniae* strains. It is an essential to *Chlamydophila pneumonia* and non-human homologous among four strains. This gene is identified as a potential drug and vaccine target for *Pneumonia disease* Reddy et al., 2012). *C. Pneumoniae* is a widespread obligate intracellular gram negative bacterium, it causes upper and lower respiratory infections worldwide. In addition to acute infections, several chronic inflammatory diseases have been associated with *C. Pneumoniae* infection. Increasing evidence implicates that a persistent lung infection caused by *C. Pneumoniae* may contribute to the initiation, exacerbation and promotion of asthma symptoms (Black, 2007; Brandén et al., 2005).

Experimentally and structural genomics projects derived many protein structures that still have unknown functions (Lee et al., 2007). The protein structural characterization is an important to understand the mechanism of action in identifying its receptor and/or other binding partners (Laurent et al., 2016; Siddiqa et al., 2016; Tan et al., 2016). However, knowing the 3D structure of a protein opens up the possibility of ascertaining its function from an analysis of that structure (Berman and Westbrook, 2004; Berman et al., 2000). This process will be useful to predict the structure from sequence and function from its structure (Watson et al., 2005).

In the case of an experimental structure of a target protein is not available, then it may be possible to create a homology model of the target based on the experimental structure of a related protein (Jhoti and Leach, 2007; Sitbon and Pietrokovski, 2007). Sequence conservation in protein families, sequence motifs or blocks are the basic secondary structure elements namely alpha helices, beta strands, structured turns, and loops. Relations between protein sequence and structure can be analysed by either determining the sequence features of predefined structures. Proteins with similar sequences adopt similar structures (Bystroff et al., 1996; Chothia and Lesk, 1986; Cygler et al., 1993; Doolittle, 1981; Han and Baker, 1996; Mizuguchi and Blundell, 2000; Shenoy and Jayaram, 2010).

Inspiration of this study is to understand the sequence, structure and function at a molecular level besides making drug

* Corresponding author.
E-mail address: kotireddy.nit@gmail.com (G. Koteswara Reddy).

http://dx.doi.org/10.1016/j.compbiolchem.2016.10.014
1476-9271/

Nomenclature

3D	3 Dimensional
30S2	30S Ribosomal protein S2
Å	Angstrom (m)
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
C. Pneumoniae	Chlamydophila pneumoniae
DNA	Deoxyribonucleic acid
E-value	Expectation value
GO	Gene ontology
HMMs	Hidden Markov oodels
HTH	The helix-turn-helix
PDB	Protein Data Bank
PDFs	Probability density functions
Pfam	Protein Families Database
PIR	Protein information resource
R.M.S.D	Root-mean-square deviation (Å)
RLR	Right handed–left handed Ramachandran plot
SAVES	Structure analysis and verification server
SSM	Secondary structure matching
Uniprot	Universal protein resource

discovery faster and more efficient by replacing slow and expensive structural biology experiments with fast and less expensive bioinformatics tools (Ashburner and Bergman, 2005; Baker and Šali, 2001; Brenner et al., 1996; Dill et al., 2007; Li et al., 2002; Lo Conte et al., 2000; Omholt, 2013; Rost, 1997; Zhang, 2009). The Profunc server has been developed to help identify the likely biochemical function of a protein from its 3D structure and sequence. It combines a number of sequence-based and structure-based methods, utilising in-house software as well as external services through the use of web services to analyse an uploaded PDB structures (Laskowski et al., 2005).

The purpose of the present study is to investigate the sequential, structural, functional and evolutionary clues for *30S2* protein in *C. Pneumoniae*. The study is divided into three sections. First section carried out that the prediction of *30S2* protein 3D structure from its sequence using homology modelling, validation by Ramachandran plot and G-factor statistics. Second and third sections carried out that the prediction of protein *30S2* sequential, structural, functional and evolutionary motifs using Profunc web server.

2. Materials and methods

The protein *30S2* amino acid sequence was retrieved from Uniprot database as a query sequence with a total length of 277 amino acids and molecular weight of 31286. The database accession number is *Q9Z7K9* (http://www.uniprot.org/uniprot/ *Q9Z7K9*).

2.1. UCSF chimera

UCSF is a molecular modelling tool. It is a highly extensible program for interactive visualization and analysis of molecular structures and related data. High-quality 3D images and molecular dynamics can be produced (Pettersen et al., 2004).

2.2. Sequence to structure prediction and validation

The 3D structure of target protein can be built from its known target sequence by homology modelling method. This method comprises four main steps: *identification of structural, template*(*s*), *alignment of target sequence* and *template structure*(*s*), *model building*, and *model quality* evaluation. These steps can be repeated until a satisfying modelling result is achieved (Kuznetsov et al., 2016; Verma et al., 2016). The *Modeller9v7* is used for homology or comparative modelling of protein 3D structure prediction (Eswar et al., 2006).

2.2.1. Identification of structural template(s)

The query sequence *30S2* was searched to find out the related protein sequences as a template by the BLAST program against PDB. Five templates were identified with PDB entry IDs *1VS5*: *B*, *1P87*: *B*, *2GY9*: *B*, *1I94*: *B* and *1FJG*: *B*. The PDB entry ID *1VS5: B* The sequence that showed maximum identity 43% with highest bits score 205, better resolution 3.46 Å, better crystallographic R-factor 0.29 and less E-value.

2.2.2. Alignment of target sequence and template structure

Structure similarity searching was performed by standalone BLASTP against PDB for finding the similar structures. The results of BLASTP showed 43% identity and 63% similarity with protein *30S2* query from which is a crystal structure of the bacterial ribosome from *Escherichia Coli* with a PDB entry ID is *1VS5: B* (Schuwirth et al., 2006) of template which is X-ray crystallized structure at 3.46 Å and selected for backbone alignment.

2.2.3. Model building

The model was constructed by using the program *Modeller9v7* under Windows. Modeller is a comparative protein structure modelling software. It is based on spatial restraints derived from the alignment and probability density functions (PDFs) (Sali and Blundell, 1993).The 3D model of a protein is obtained by optimization of the molecular PDFs such that the model violates the input restraints as little as possible.

2.2.4. Model quality evaluation

It is suggested to assess the overall stereo chemical quality of the modelled structure via Ramachandran plot analysis using the program PROCHECK (Burley et al., 1994; Fernandez-Fuentes and Fiser, 2010; Laskowski et al., 1993) and G-Factor statistics. Further evaluation of modelled structure was done by VERIFY3D (Ramachandran et al., 1963), ERRAT is a protein structure verification algorithm that is especially well-suited for evaluating the progress of crystallographic model building and refinement (Lüthy et al., 1992) through structure analysis and verification server (https://services.mbi.ucla.edu/SAVES/). The methodology for protein structure prediction and validation is described in Fig. 1

2.3. Sequence based function prediction

2.3.1. Sequence search

Similar or relative sequences with annotated functions are identified by searching BLAST against UniProt database. BLAST results are provided that the aligned sequences, residue conservation with calculated scores and sequence alignment colour. The results will be displayed as a multiple sequence alignment coloured by residue conservation score (Altschul et al., 1990; UniProt, 2010).

2.3.2. Gene neighbours

It is the process of identification of neighbouring genes on genome. Neighbouring genes, particularly on bacterial genomes, are often functionally related. The homologous relative genome provides the functional information of query proteins role.