Molecular Phylogenetics and Evolution 109 (2017) 1–10

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Slr0006-like proteins: A TsaC/TsaC2/YciO subfamily exclusive to cyanobacteria

Leonor L. Carvalho, Tiina A. Salminen, Käthe M. Dahlström *

Structural Bioinformatics Laboratory, Biochemistry, Faculty of Science and Engineering, Åbo Akademi University, Tykistökatu 6A, 20520 Turku, Finland

ARTICLE INFO

Article history:
Received 6 September 2016
Revised 15 December 2016
Accepted 28 December 2016
Available online 30 December 2016

Keywords:
Protein evolution
Homology modeling
tRNA modification
Sll0216
Sll1866

ABSTRACT

The universally conserved TsaC/TsaC2/YciO family of proteins is essential for the N^6-threonylcarbamoyladenosine modification present in almost all ANN-decoding tRNAs. Previously, the family has been grouped into the TsaC/TsaC2 and YciO subfamilies. We used sequence analysis, phylogenetic methods and homology modeling to show that a third subfamily, the Slr0006-like subfamily, exists exclusively in some cyanobacteria. The Slr0006-like proteins are solely found together with both TsaC and YciO homologs, indicating a distinct function for the Slr0006-like subfamily. Accordingly, the homology models show that the amino acids in their putative binding clefts differ significantly. Hence, we introduce a new cyanobacterial subfamily of proteins with the TsaC-domain fold, along with the generated classification rules to assign new members to the correct cyanobacterial subfamily.

1. Introduction

The TsaC/TsaC2/YciO (previously YrdC/Sua5/YciO) family of proteins is universally conserved and can be split into the TsaC/TsaC2 and YciO subfamilies. Despite low sequence identity (<∼30%), the TsaC/TsaC2/YciO family members contain a conserved 3D structure with ∼185 residues (Fu et al., 2010; Jia et al., 2002). This TsaC-domain folds into an α/β twisted open-sheet structure with parallel and antiparallel β-strands, forming a central cleft with positive charge in all currently known structures (Agari et al., 2008; Fu et al., 2010; Jia et al., 2002; Kuratani et al., 2011; Parthier et al., 2012; Petkun et al., 2011; Teplova et al., 2000). Furthermore, this domain can be an independent protein, such as *Escherichia coli* TsaC (PDB ID: 1HRU (Teplova et al., 2000)), *E. coli* YciO (PDB ID: 1KK9 (Jia et al., 2002)) and *Streptococcus mutans* smu.1377c (PDB ID: 3L7V (Fu et al., 2010)), or part of a multidomain protein as in *E. coli* HypF (PDB ID: 3TSP (Petkun et al., 2011)), *Streptoalloteichus tenebrarius* TobZ (PDB ID: 3VEN (Parthier et al., 2012)) and *Sulfolobus tokodaii* TsaC2 (PDB ID: 2EQA (Agari et al., 2008), 3AJE (Kuratani et al., 2011), 4E1B (Parthier et al., 2012) [re-evaluation of PDB ID: 2EQA]). Nowadays, *E. coli* TsaC, *S. tokodaii* TsaC2 and *S. tenebrarius* TobZ are known to be important for the N^6-threonylcarbamoyladenosine (t^6A) modification pathway (El Yacoubi et al., 2009; Kuratani et al., 2011; Parthier et al., 2012), which alters position 37 in the anticodon loop of tRNA (t^6A_{37}) and maintains an accurate translation of the ANN codon (Auffinger and Westhof, 1998; Deutsch et al., 2012; El Yacoubi et al., 2012; Grosjean et al., 1995; Jühling et al., 2009). The central clefts of *E. coli* TsaC (PDB ID: 2MX1 (Harris et al., 2015)), *S. tokodaii* TsaC2 (PDB ID: 4E1B (Parthier et al., 2012)) and *S. tenebrarius* TobZ (PDB ID: 3VER (Parthier et al., 2012)) contribute to the first step of the modification, where *L*-threonine and ATP bind to the conserved signature motif (KxR. . .SxN (Harris et al., 2015)) and react to form the intermediate threonylcarbamoyladenylate (TCA) (Parthier et al., 2012). Thereafter, the threonylcarbamoyl moiety of TCA is transferred to A_{37} of tRNA by the bacterial threonylcarbamoyltransferase complex (Deutsch et al., 2012; Lauhon, 2012; Perrochia et al., 2013; Wan et al., 2013). Despite the advances in TsaC/TsaC2 characterization, the function of the homologous YciO is unknown (Montero et al., 2009). It has been linked to glycogen metabolism but, unlike TsaC and TsaC2, YciO is not essential for survival and it cannot compensate for the lack of TsaC or TsaC2 (Deutsch et al., 2012; Gerdes et al., 2011; Kaczanowska and Rydén-Aulin, 2005). Furthermore, the KxR. . .SxN signature motif observed in the TsaC/TsaC2 subfamily is replaced by a KxL. . .SxM motif in *E. coli* YciO (El Yacoubi et al., 2011; Gerdes et al., 2011; Harris et al., 2015).

Most organisms contain only one member of the TsaC/TsaC2/YciO subfamily, but some plants and several bacteria have two homologs (El Yacoubi et al., 2009). For example, 9% of 9176 bacterial genomes contain both TsaC and TsaC2, while 54% also encode

* Corresponding author.
E-mail address: kathe.dahlstrom@abo.fi (K.M. Dahlström).

http://dx.doi.org/10.1016/j.ympev.2016.12.033
1055-7903/

for YciO in addition to TsaC or TsaC2 or both of them (Thiaville et al., 2015b). Similarly, the cyanobacterium *Synechocystis* PCC 6083 encodes proteins homologous to TsaC (Sll1866; UniProtKB ID: P74144) and YciO (Sll0216; UniProtKB ID: P72724) (Carmel et al., 2013; Thiaville et al., 2015a). Additionally, *Synechocystis* encodes Slr0006 (UniProtKB ID: Q55667), which also has the TsaC-domain fold (Carmel et al., 2013). Limited access to inorganic carbon up-regulates the levels of Slr0006, but it is not important for cell survival (Battchikova et al., 2010; Carmel et al., 2012). We previously suggested that Slr0006 belongs to the YciO subfamily due to their similar signature motifs (Carmel et al., 2013). However, unlike YciO, glycogen levels were unaffected in a *slr0006* knock-out strain, leaving the function of Slr0006 unknown. Nevertheless, it is highly interesting that *Synechocystis* encodes a third, distinct protein with the TsaC-domain fold and it is intriguing to know its relationship to the other proteins in the TsaC/TsaC2/YciO family. Therefore, we analyzed the family in multiple cyanobacterial species with phylogenetic methods and structural modeling and analysis. Our study shows that several cyanobacterial genomes encode three different proteins, which can be classified as TsaC-, YciO- and Slr0006-like based on the general guidelines derived in this work. Interestingly, Slr0006-like proteins are unique to the cyanobacterial phylum and the species encoding Slr0006-like proteins always contain both TsaC- and YciO-like proteins, indicating that the Slr0006-like proteins indeed have a different function than the TsaC- and YciO-like proteins.

2. Methods

2.1. Sequence data

The *Synechocystis* sp. PCC 6803 Sll0216, Sll1866 and Slr0006 sequences were obtained from UniProtKB (Jain et al., 2009) and used as queries in NCBI (Altschul et al., 1990) and CyanoBase (http://genome.microbedb.jp/cyanobase) BLAST searches against non-redundant cyanobacterial genomes. Sequences with significant E-value (<0.001) and sequence identity from 22% to 71% were collected.

2.2. Sequence alignments, phylogenetic analysis and homology modeling

Multiple sequence alignments were generated with MALIGN (Johnson et al., 1996) in the BODIL modeling environment (Lehtonen et al., 2004). Sll0216, Sll1866 and Slr006 were aligned with their respective homologs and then combined into one multiple sequence alignment. *E. coli* TsaC and YciO were thereafter aligned to the previously aligned multiple sequence alignment. Phylogenetic reconstruction was done with Neighbor Joining (NJ) (Nei, 1987) and Maximum Likelihood (ML) in MEGA6 (Tamura et al., 2013). For ML analysis, the Le Gascuel (LG) 2008 (Le and Gascuel, 2008) substitution matrix, discrete gamma distributions and complete deletion of gaps and missing data were implemented, while NJ analysis was performed with the Jones-Taylor-Thornton (JTT) model (Jones et al., 1992). Branch support was assessed by bootstrapping (500 replications). A multiple structure-based alignment was generated by superimposing *E. coli* TsaC (PDB ID: 1HRU (Teplova et al., 2000)), *E. coli* YciO (PDB ID: 1KK9 (Jia et al., 2002)) and *S. tokodaii* TsaC2 (PDB ID: 4E1B (Parthier et al., 2012)) with VERTAA (Johnson and Lehtonen, 2000) in BODIL. Sll1866, Sll0216 and Slr0006 were aligned individually to the fixed structure-based alignment. All sequences except the target protein and the sequence for the crystal structure used as template were deleted prior to modeling. Homology models were generated with MODELLER (Sali and Blundell, 1993) and the model with the lowest value of the MODELLER objective function was chosen for further studies. Model evaluation was estimated with PROCHECK (Laskowski et al., 1993), ProQ (Wallner and Elofsson, 2003) and ProSA-web (Wiederstein and Sippl, 2007), as well as through visual inspection of superimpositions with the template structure. PyMOL (DeLano, 2002) was used for generating electrostatic surfaces (the APBS tool) and pictures.

2.3. Molecular dynamics simulations

Molecular dynamics (MD) simulations were calculated with Gromacs (Berendsen et al., 1995). A topology file was created with the OPLS-AA/L all atom force field (Jorgensen and Tirado-Rives, 1988) and the system was defined. A simulation box with cubic dodecahedron shape and water as solvent was created. The system energy was minimized for 500 steps to assure a reasonable geometry and solvent orientation of the starting structure. The surrounding water was equilibrated for 100 ps and the simulation was performed in the local cluster for 50 ns. The trajectory movie was visualized frame-by-frame with PyMOL (DeLano, 2002). The Root Mean Square Fluctuation (RMSF) values were displayed using the b-factor putty representation, showing the backbone as a tube with a diameter and coloring correlating to the RMSF.

3. Results and discussion

3.1. Slr0006-like proteins are found together with both TsaC- and YciO-like proteins

BLAST searches with *Synechocystis* Sll1866, Sll0216 and Slr0006 as queries against cyanobacterial genomes resulted in 136 sequences with significant Evalue (<0.001) and 22–71% sequence identity. These were aligned together with the *E. coli* TsaC and YciO sequences and subsequently used for phylogenetic analysis. Both the ML and the NJ trees present the same topology grouped in three major clades: (1) the TsaC-like proteins, (2) the YciO-like proteins and (3) the Slr0006-like proteins (Fig. 1). The results indicate that an ancestral protein gave origin to the Slr0006-like proteins and to a TsaC – YciO-like precursor through a gene duplication event. The TsaC- and YciO-like proteins have subsequently diverged from this last common ancestor. Subclades containing only cyanobacterial proteins are formed within each major clade and are supported by high bootstrap values: 69% for YciO-like, 99% for TsaC-like and 80% for Slr0006-like proteins. As expected, the bacterial *E. coli* TsaC and YciO form their own branches within the major clades. Interestingly though, *Moorean producens* 3L TsaC-like protein groups together with *E. coli* TsaC, indicating that it is a middle form between the cyanobacterial and bacterial TsaC-like proteins. In the Slr0006-like clade, the Slr0006-like protein from *Pleurocapsa* sp. PCC 7319 branches alone due to the many differences of otherwise completely conserved residues within the clade, but the branching was not elucidated with high confidence. Moreover, the low sequence identity characteristic for the TsaC/TsaC2/YciO family overall (Carmel et al., 2013; Fu et al., 2010; Jia et al., 2002) is also found in the YciO- and Slr0006-like subfamilies, which have an average sequence identity of ~52.2% and ~30.1%, respectively, compared to ~66.1% sequence identity among TsaC-like proteins. Interestingly, TsaC- and YciO-like proteins are found together in 27 cyanobacterial species, while Slr0006-like proteins are found exclusively in 14 of these 27 species. Hence, the function of Slr0006-like proteins is likely needed in addition to the tasks performed by the TsaC- and YciO-like proteins, indicating a distinct function for the Slr0006-like proteins.