Data in Brief 35 (2021) 106727

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

# Molecular characterization, phylogenetic and *in silico* sequence analysis data of trehalose biosynthesis genes; *otsA* and *otsB* from the deep sea halophilic actinobacteria, *Streptomyces qinglanensis* NIOT-DSA03

Balakrishnan Meena[a,*], Lawrance Anburajan[a], Nambali Valsalan Vinithkumar[a], Ramalingam Kirubagaran[b], Gopal Dharani[b]

[a] *Atal Centre for Ocean Science and Technology for Islands, National Institute of Ocean Technology, Ministry of Earth Sciences, Government of India, Port Blair-744103, Andaman and Nicobar Islands, India*
[b] *Marine Biotechnology Division, Ocean Science and Technology for Islands Group, National Institute of Ocean Technology, Ministry of Earth Sciences, Government of India, Chennai 600100, Tamil Nadu, India*

ARTICLE INFO

*Article history:*
Received 2 August 2020
Revised 6 January 2021
Accepted 6 January 2021
Available online 21 January 2021

*Keywords:*
Deep sea
Halophilic actinobacteria
Trehalose genes
Compatible solutes
Osmolytes

ABSTRACT

Trehalose, a non-reducing disaccharide ($\alpha$-D-glucopyranosyl-(1→1)-$\alpha$-D-glucopyranoside) is a natural compound, which serves as a protective substance in halophilic bacterial cells. Trehalose biosynthesis genes (*otsA* and *otsB*) were PCR amplified from the genomic DNA of deep sea actinobacteria, *Streptomyces qinglanensis* NIOT-DSA03. The amplified genes were cloned and nucleotide sequences were determined. *In silico* sequence and phylogenetic analysis of nucleotides and amino acids of *otsA* and *otsB* sequences of *S. qinglanensis* were also determined. The experimental data described in this study will be helpful to develop a recombinant expression system to produce trehalose for biotechnological applications.



* Corresponding authors.
*E-mail addresses:* meenasteffy@gmail.com, anbuniot1979@gmail.com (B. Meena).

https://doi.org/10.1016/j.dib.2021.106727

## Specifications Table

| | |
|---|---|
| Subject | Applied Microbiology and Biotechnology |
| Specific subject area | Biotechnology and Bioinformatics |
| Type of data | Images, Figures |
| How data were acquired | Molecular cloning, BLAST program of NCBI (http://www.ncbi.nlm.nih.gov), CLUSTALW program, GeneDoc program (https://genedoc.software.informer.com/2.7/), BioEdit 7.05 program (www.mbio.ncsu.edu/BioEdit/), PROTEAN program, MODELLER program |
| Data format | Raw and Analysed data |
| Parameters for data collection | Data were collected using polymerase chain reaction studies, nucleic acid electrophoresis through agarose gels, molecular cloning, DNA sequencing and *in silico* analysis. |
| Description of data collection | Sequencing results revealed that *otsA* and *otsB* genes contains 1278 bp and 879 bp long ORF encoding 425 and 292 amino acids, respectively. *In silico* sequence and phylogenetic analysis of nucleotides and amino acids revealed that the *otsA* and *otsB* sequences of *Streptomyces qinglanensis* NIOT-DSA03 were conserved in many eubacteria. |
| Data source location | National Institute of Ocean Technology<br>Port Blair<br>India<br>(12 °12.90'N, 093 °48.92'E) |
| Data accessibility | Data is available with this publication |

## Value of the Data

- The dataset describes the importance of major osmolyte, trehalose in protecting the proteins and cellular membranes in prokaryotes from inactivation or denaturation by the environmental stress.
- The dataset provides information about the osmolyte in *Streptomyces qinglanensis* NIOT-DSA03 and its application in derma-pharmacy industries.
- In this data we have provided the detailed information regarding the gene sequence and its protein structure. This data may be used for further heterologous gene expression studies.

## 1. Data Description

The *otsA* and *otsB* genes encode the, $\alpha$-trehalose-phosphate synthase and trehalose-6-phosphate phosphohydrolase respectively. Together these proteins constitute the trehalose biosynthetic pathway. The trehalose biosynthesis genes *otsA* and *otsB* were PCR amplified and are encoded by polynucleotides of 1278 bp and 879 bp (Fig. 1). The *otsA* and *otsB* genes encodes proteins of 425 and 292 amino acids with calculated molecular masses of 46625, 30228 Daltons (Fig. 2a & b). After PCR amplification, the products were purified from the agarose gel and cloned into pDrive cloning vector. The white colonies were selected and screened for the presence of insert by PCR amplification using specific primers, which gave specific product. The recombinant transformants of *otsA* and *otsB* genes were also confirmed by double digestion with *Sac* I and *Bam* HI restriction enzymes, which released full gene along with flanking region of the vector. The *otsA* and *otsB* sequences generated in this study have been deposited in the GenBank database with the accession numbers MN017301 and MN023141.

The *otsA* and *otsB* sequences from *S. qinglanensis* NIOT-DSA03 were analyzed with reported amino acid sequences of other actinobacteria viz. *S.venezuelae,* (GenBank accession no. LN881739.1), *S. lIncolnensis,*(CP016438), *S. clavuligerus,* (CP016559), *S. fradiae,* (CP032266), *Streptomyces* sp., (CP029188), *S. alfalfa,* (CP015588), *S. venezuelae,* (LN881739), *S. clavuligerus*, (CP016559), using Clustal W program [1].

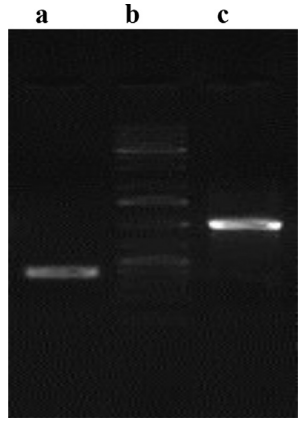


**Fig. 1.** Agarose gel electrophoresis of *otsA* and *otsB* gene amplicons. Lane a: *otsB* amplicon 879 bp, Lane c: *otsA* amplicon 1,278 bp, Lane b: 1 kb DNA ladder.

The phylogenetic tree at nucleotide and amino acid level of *otsA* revealed the phylogenetic similarity of *otsA* gene from *S. qinglanensis* NIOT-DSA03 with other organisms. The bacterial species switched to different clusters for *otsA* gene at nucleotide and amino acid level indicating divergence among the organisms and the degree of divergence in the sequences. *S. qinglanensis, S. venezuelae* and *S. lIncolnensis* were grouped in the same cluster in both the phylogenetic trees (Fig. 3a). The phylogenetic tree of nucleotide and amino acid sequences of *otsB* gene also revealed the grouping of *S. qinglanensis, S. alfalfa* and *S. lIncolnensis* in a single cluster as that of *otsB*. In phylogenetic tree analysis, a diverged mode of clustering was observed (Fig. 3b).

On phylogenetic analysis, the *otsA* and *otsB* genes of *S. qinglanensis* was found to be highly conserved among the bacterial species. The *otsB* gene was found to have highest similarity between bacterial species compared to the *otsA* gene. Based on phylogenetic analysis, *S. qinglanensis* and *S. lincolnensis* were found to be clustered together for *otsA* and *otsB* genes. The genes involved in the biosynthesis of trehalose in *S. qinglanensis* and *S. venezuelae* are comparatively well conserved compared to other bacteria both at nucleotide and amino acid level [2].

Prediction of secondary structure was performed with the PROTEAN program (Discovery studio 3.5). The secondary structure of *otsA* and *otsB* proteins were predicted to have the alpha-helical structure with maximum hydrophilic molecules. The prediction analysis also revealed the presence of many acidic amino acids; regions with high antigenicity and very high backbone chain flexibility. Upon analysis of *otsA* protein, the predicted charge at pH 7.0 was "+9.54" with the isoelectric point of 5.82. Common amino acids include 14% glutamic acid, 10% leucine, 11% each of alanine, threonine, valine and 17% each of phenylalanine, glycine, proline and glutamine (Fig. 4a). In *otsB* protein, the predicted charge at pH 7.0 was "+20.14" with the isoelectric point of 5.64. The amino acid composition includes 52% glycine, 41% glutamic acid, 37% leucine, 39% threonine and 29% lysine (Fig 4b). This prediction result also showed considerable similarity with the reported trehalose biosynthesis enzymes from actinobacteria.

Three dimensional structure prediction of the trehalose synthase enzyme suggests that the tertiary structure was highly compatible with the secondary structure prediction analysis. The structure was validated using Ramachandran plot and the plot suggested that none of the residues were present in the disallowed region. This deduces that the modelled structure shares high level of similarity with the structures that have been already reported. Homology analysis of trehalose synthase enzyme with Protein Data Base (PDB) revealed the maximum of 100% and minimum of 21% identity with the PDB templates (Fig. 5a & b).

```
atg ctg gac atc ccc tcc ggc acc ttc gag gcg gcc tac cac ggc atc gcc aac tcc gta
 M   L   D   I   P   S   G   T   F   E   A   A   Y   H   G   I   A   N   S   V
ctg tgg ttc acc cac cac atg ctg tac cac acg ccg ctg gag ccc gtc ttc gac gag gac
 L   W   F   T   H   H   M   L   Y   H   T   P   L   E   P   V   F   D   E   D
ttc agc ggt cag tgg gcc ggt tac gag gcg tac aac gcc gct ttc gcg gac gcg ctg gcg
 F   S   G   Q   W   A   G   Y   E   A   Y   N   A   A   F   A   D   A   L   A
cag gag gcc gcc gac ggt gcc gcc gtc ctg gtg cag gac tac cat ctc gcg ctg gtc ccc
 Q   E   A   A   D   G   A   A   V   L   V   Q   D   Y   H   L   A   L   V   P
ggc atg ctc cgc gca cgc cgc ccc gac ctg cgg atc ggc cat ttc tcg cac aca ccg tgg
 G   M   L   R   A   R   R   P   D   L   R   I   G   H   F   S   H   T   P   W
gcg ccc gcc gac tac ttc cgg ctg ctc ccg gac gac gtc gcc gcg cag gtg ctg gcc ggc
 A   P   A   D   Y   F   R   L   L   P   D   D   V   A   A   Q   V   L   A   G
atg ctg ggc gcc gac cgc acc gcg ttc ctg acc gcg cgc tgg gcc cgg ctg ttc gcc gag
 M   L   G   A   D   R   T   A   F   L   T   A   R   W   A   R   L   F   A   E
tgc tgc gcg cgg gtg ctg ggc gcg acg gtc gag ggc ggc ggc ccg ccg cgc gag gac gcc
 C   C   A   R   V   L   G   A   T   V   E   G   G   G   P   P   R   E   D   A
gcg ccg gag ggc ctc tcc tcg gac gac gcc gca ccg gag gac ttc tcg ctg acg gtg acc
 A   P   E   G   L   S   S   D   D   A   A   P   E   D   F   S   L   T   V   T
tac gag ggc cgt acg aca cat gtc ggg gtg cat ccg ctg ggt gcg gac ggc gac ttc ctg
 Y   E   G   R   T   T   H   V   G   V   H   P   L   G   A   D   G   D   F   L
cgc gaa cgg gcg cac cgg agc gac gtc gcc gac cgg ctg gcg cag ttg cgc gac cag gtg
 R   E   R   A   H   R   S   D   V   A   D   R   L   A   Q   L   R   D   Q   V
ggc acg ggc ccg gac ggg gcc ccg cgc cgg gtc ctg gtc cgg gtg gac cgc acc gag ctg
 G   T   G   P   D   G   A   P   R   R   V   L   V   R   V   D   R   T   E   L
tcc aag aac atc gtc cgg ggg ctg tac gcc tac cgg cgg ctg ctg gcc gac cgc ccc gag
 S   K   N   I   V   R   G   L   Y   A   Y   R   R   L   L   A   D   R   P   E
tgg cgg gag cgc gtc gtc cac ctc gcg ttc gcc tac ccc tcg cgg cag gac ctc gcc gtc
 W   R   E   R   V   V   H   L   A   F   A   Y   P   S   R   Q   D   L   A   V
tac cgg gac tac acg gcc gag gtg agc cgg gtc gcc gag gag atc aac cgg gag ttc ggg
 Y   R   D   Y   T   A   E   V   S   R   V   A   E   E   I   N   R   E   F   G
```

(a)

```
acc ggc ggc tgg acg ccg gtc gtg ctg cat gtg aag gac gac ttc gcg agg tcg ctg gcc
 T   G   G   W   T   P   V   V   L   H   V   K   D   D   F   A   R   S   L   A
gcc tac cgg atg gcg gac gtc gcg ctg gtg aac ccc gta cgg gac ggg atg aac ctg gtg
 A   Y   R   M   A   D   V   A   L   V   N   P   V   R   D   G   M   N   L   V
gcc aag gag gtc ccc gtg gtc tcg gac gcg ggc tgc gtg ctg gtg ctc tcc cgg gag gcg
 A   K   E   V   P   V   V   S   D   A   G   C   V   L   V   L   S   R   E   A
ggc gcc cat gag gaa ctc gcg ccc gac gtg ctg tcc gtg aac ccg ttc gac gtg cgg gag
 G   A   H   E   E   L   A   P   D   V   L   S   V   N   P   F   D   V   R   E
acg gcg gcc gcc ctg cac tcc gcg ctg gag gcg gac ccg gcc gac cgc gcg gag cgg acc
 T   A   A   A   L   H   S   A   L   E   A   D   P   A   D   R   A   E   R   T
cgc cga ctg gcc gcg acg gcc acc tcc cgg ccg ccc gcc cgc tgg ttc ctc gac cag ttg
 R   R   L   A   A   T   A   T   S   R   P   P   A   R   W   F   L   D   Q   L
gag gcg ctg ccc ggc tga
 E   A   L   P   G   -
```

(b)

**Fig. 2.** (a) Nucleotide sequences of *otsA* gene in *Streptomyces qinglanensis* NIOT-DSB03. (b) Nucleotide sequences of *otsB* gene in *Streptomyces qinglanensis* NIOT-DSB03.

## 2. Experimental Design, Materials and Methods

### *2.1. Bacterial strain, growth conditions, DNA isolation and plasmids*

*S. qinglanensis* NIOT-DSA03 was isolated from the deep sea sediment sample obtained during the cruise of the Barren Island, Andaman and Nicobar (A & N) Islands in the ocean research vessel Sagar Manjusha. Using box cores at a depth of 1,840 m (12° 12.90 'N, 093° 48.92′ E), sediment samples were collected from the seafloor. In the ISP 1 medium, the isolate was grown aerobically and the genomic DNA was isolated following the modified Kutchma et al., procedure. [3]. Using the universal Eubacterial primers, 16S F (5′-ACTCAAGGAATTGACGG-3′) and 16S R (5′-TACGGCTACCTGTTACGACTT-3′), the 16S rDNA was amplified by polymerase chain reaction. According to the instructions of the manufacturer in the InsTAclone PCR Cloning Kit, the 16S rDNA amplicon was cloned into a T/A cloning vector (MBI Fermentas, USA). Using the dye termination process, DNA sequencing was carried out on an ABI PRISM 377 genetic analyzer (Applied

```
atg tcg ctt ccg ccg cct tcc gcg cat ccg acg ctg ccc gaa cct gag acc gag gcc ggg
 M   S   L   P   P   P   S   A   H   P   T   L   P   E   P   E   T   E   A   G
cgc gcc ggg ctc gcc gcc gtc cgc gcc gac ccc gcg cgc acc gtg ctc gca ctc gac ttc
 R   A   G   L   A   A   V   R   A   D   P   A   R   T   V   L   A   L   D   F
gac ggc acc ctc gcg ccc atc gtc ggc gat ccg cgg gac gcc cgg gcg cac ccc gag gcg
 D   G   T   L   A   P   I   V   G   D   P   R   D   A   R   A   H   P   E   A
gtt ccc gtg ctg gcg cgg ctc gcc ccg cgg ctg gcc ggc gtc gcc gtg atc acc ggc cgc
 V   P   V   L   A   R   L   A   P   R   L   A   G   V   A   V   I   T   G   R
ccg gcg gcg gaa gcc gtc cgg tac ggc ggc ctc gaa ggc gcc gcc gga ctg gag gga ctc
 P   A   A   E   A   V   R   Y   G   G   L   E   G   A   A   G   L   E   G   L
acc gtc ctg ggc gcg tac ggg gcc gaa cgg tgg gac gcg gcg gac ccg gtc gtc cgc gcc
 T   V   L   G   A   Y   G   A   E   R   W   D   A   A   D   P   V   V   R   A
ccc gaa ccc ccg gcc ggg gtg gcc gcg gta cgg gcc gaa ctc ccc ggg ctg ctc cgg gag
 P   E   P   P   A   G   V   A   A   V   R   A   E   L   P   G   L   L   R   E
tcc gaa gcg ccg gaa ggc acc tgg acg gag gac aag ggc cgg gcg ctc gct gtg cac acc
 S   E   A   P   E   G   T   W   T   E   D   K   G   R   A   L   A   V   H   T
cgg cgt gcg gcc tcg ccc gac gag gcc ctc gac cgg ctg cgg gag ccg ctg tac acg ctg
 R   R   A   A   S   P   D   E   A   L   D   R   L   R   E   P   L   Y   T   L
gcc gag cgg cac gga ctc gtg gtg gaa ccg ggg cgc atg gtg ctc gaa ctc cgg ccg ccg
 A   E   R   H   G   L   V   V   E   P   G   R   M   V   L   E   L   R   P   P
ggc gcc gac aag ggc gcc gcg ctc acc ggc ttc gtc cgc gag cgg gcc gcc acc gcc gtc
 G   A   D   K   G   A   A   L   T   G   F   V   R   E   R   A   A   T   A   V
gtc tac gcc ggt gac gac cgc ggc gat ctg ccc gcc tac gcc gct gtc acg gcc ctg cgc
 V   Y   A   G   D   D   R   G   D   L   P   A   Y   A   A   V   T   A   L   R
gcc gag ggc gta ccg gga ctg ctg ctc tac agc gcc ccc gag gcg gag gcc gag gcg gtt
 A   E   G   V   P   G   L   L   L   Y   S   A   P   E   A   E   A   E   A   V
ccc gag ctc cgc gac gga gcc gac ctc cgg gtc ccg ggc ccg gcc gga ctg gtg gcc tgg
 P   E   L   R   D   G   A   D   L   R   V   P   G   P   A   G   L   V   A   W
ctg cgg gcg ttg gcg gcc gag atc ccg ccg gtg cgc tga
 L   R   A   L   A   A   E   I   P   P   V   R   -
```

**Fig. 2.** Continued

Biosystems, USA). In a homology search with the available sequences in GenBank using BLAST provided for pair identities by NCBI, the acquired 16S rDNA sequences were used. As the host strain for transformation in cloning, the *Escherichia coli* JM109 strain was used and the plasmid pDrive was used as the cloning vector.

### *2.2. PCR amplification of trehalose biosynthesis genes*

Trehalose biosynthesis genes, *otsA* and *otsB* were individually amplified by PCR using gene-specific primers designed by a program available at http://frodo.wi.mit.edu/primer3 otsA F (5′-ATGCTGGACATCCCCTCC-3′) and otsA R (5′-TCAGCCGGCAGCCCTC-3′) and otsB F (5′-ATGTCGCTCCGCCCGCCCC-3′) and otsB R (5′-CAGCGCCGGGATC-3′). The final volume of PCR was 50 $\mu$l, each containing 0.5 $\mu$M of forward and reverse primers; 1.0 $\mu$l of crude genomic DNA; 200 $\mu$M of dNTP; 1 $\times$ *Pfu* buffer; 2.5 mM MgSO4; 1.0 U of *Pfu* DNA polymerase (MBI Fermentas, USA) and remaining autoclaved Milli Q water. In the Master Cycler (Eppendorf, Germany), amplification was carried out under the following conditions; initial denaturation at 94 °C for 3 min, followed by 30 repeated cycles at 94 °C for 30 sec, 50 °C for 1 min, 72 °C for 2 min and final extension at 72 °C for 5 min. The PCR amplicons were analyzed on 1.5 percent agarose gel along with 100 bp DNA ladder (MBI Fermentas) and recorded in gel documentation system (UVP BioSpectrum Imaging system, USA).

### *2.3. Molecular cloning of trehalose genes*

The *otsA* and *otsB* gene amplicons were purified by the MinElute Gel purification kit (Qiagen, Germany) and as directed by the manufacturer, cloned into pDrive (Qiagen). The gene constructs

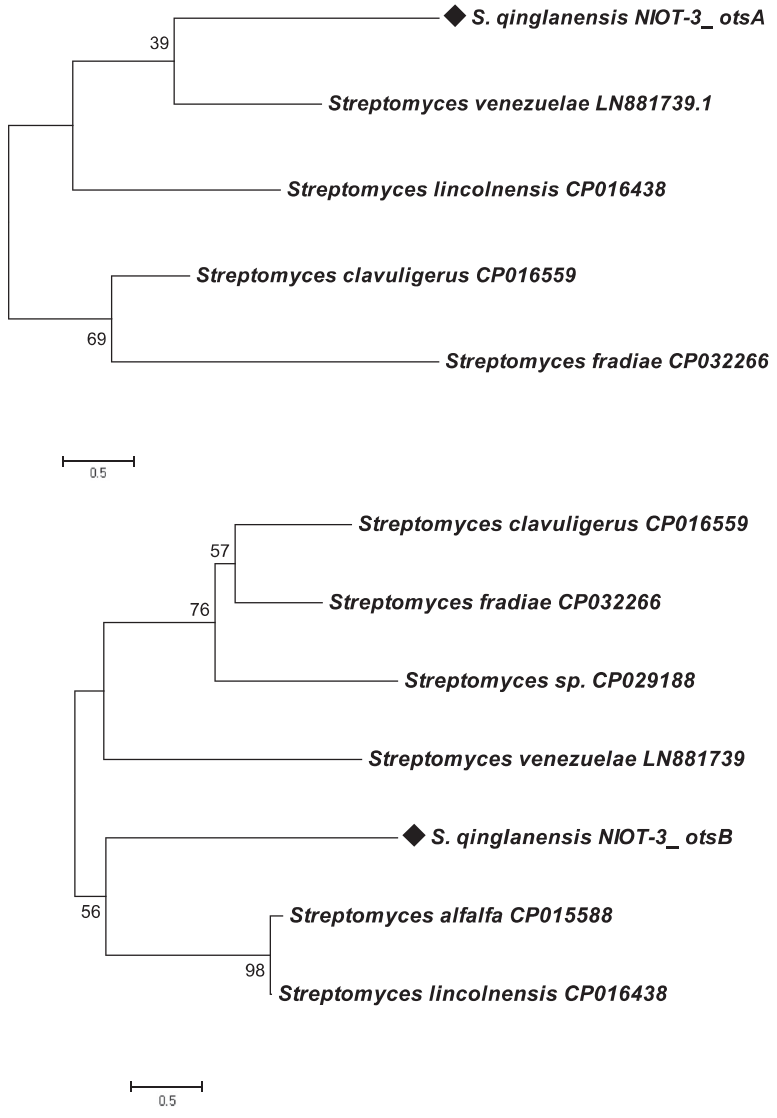


**Fig. 3.** (a) Phylogenetic tree analysis of *otsA* amino acid using MEGA program. (b) Phylogenetic tree analysis of *otsB* amino acid using MEGA program.

of pDrive-*otsA* and *otsB* have been transformed into *E. coli* JM109 (recA1, endA1, gyrA96, thi-1, hsdR17 (rk-mk+), e14-(mcrA-), supE44, relA1, Δ (lac-proAB)/F '[traD36, proAB+, lacIq, lacZ Δ M15]). White colonies were selected for PCR amplification with M13f-M13r (MBI Fermentas) vector primers, and clones with the correct insert were selected for alkali lysis process plasmid isolation as measured by size and correct orientation [4].

### *2.4. Characterization of recombinant plasmids*

The recombinant plasmids with pDrive-*otsA* and *otsB* gene constructs were double digested with *Sac*I and *Bam* HI enzymes. The restriction digestion was carried out with a final volume of

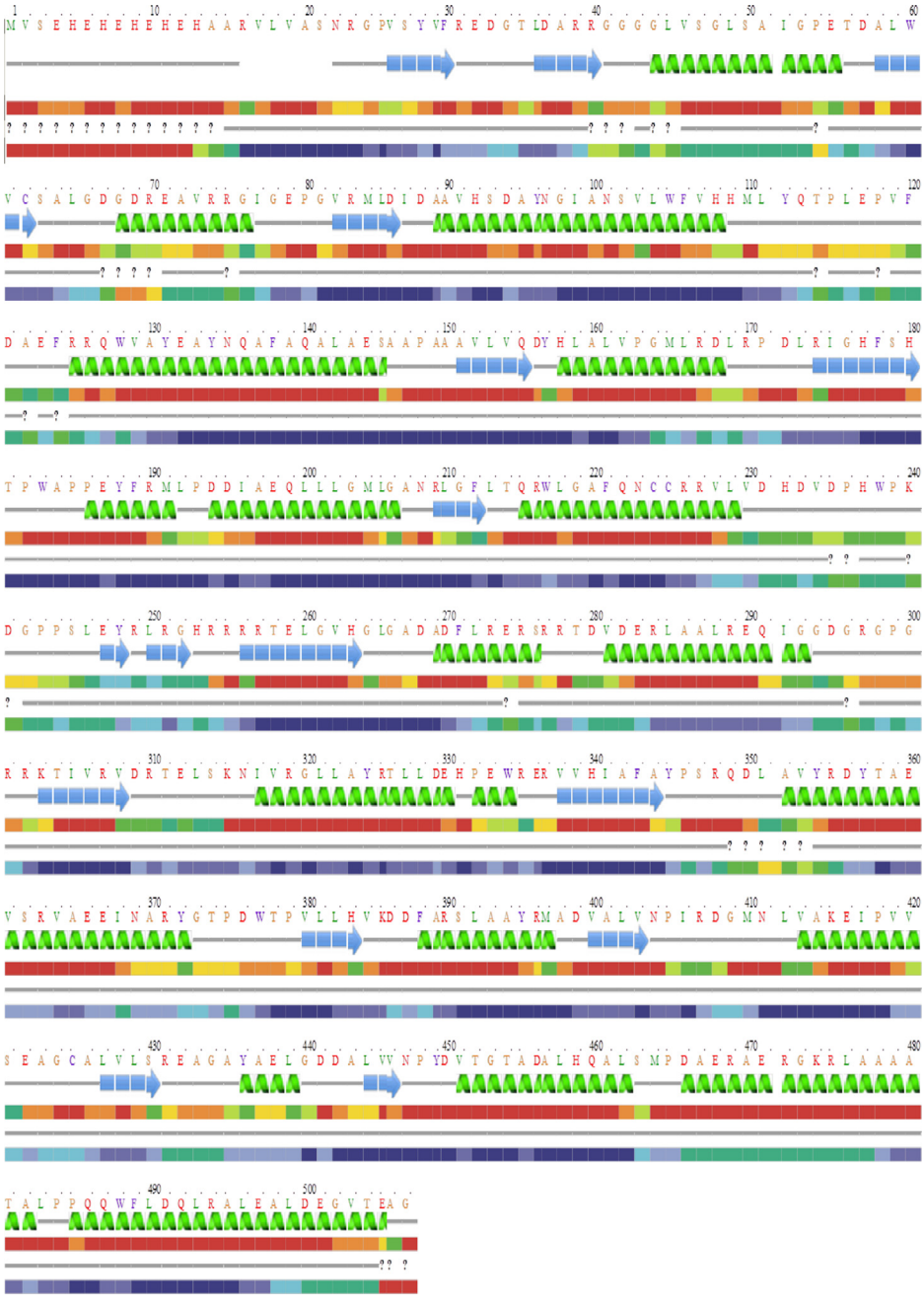

**Fig. 4.** (a) Secondary structure prediction of *otsA* gene (Chou-Fasman and Garnier-Robson algorithms for predicting alpha, beta and turn regions) using PROTEAN program. (b) Secondary structure prediction of *otsB* gene (Chou-Fasman and Garnier-Robson algorithms for predicting alpha, beta and turn regions) using PROTEAN program.

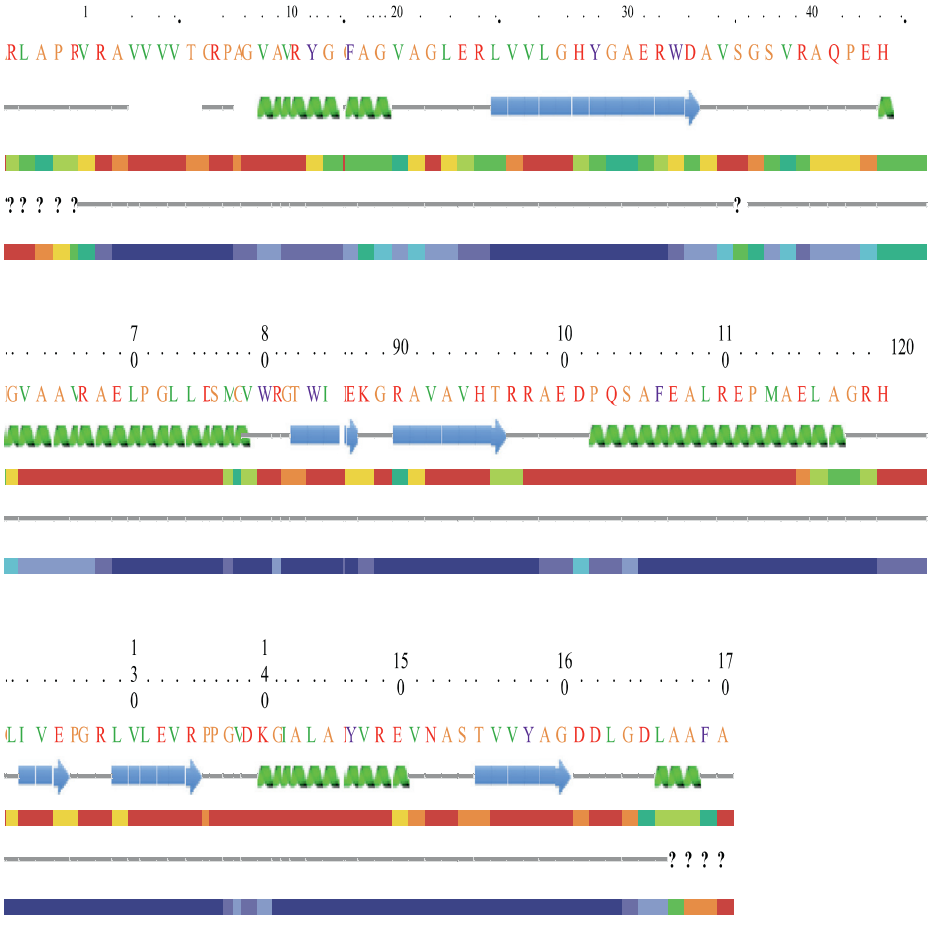

**Fig. 4.** Continued

20 $\mu$l comprising 2 $\mu$l of each recombinant plasmid; 1 $\times$ enzyme buffer; 5 U/$\mu$l of each restriction enzyme and the remaining Milli Q autoclaved water. In a water bath, reaction mixtures were incubated overnight at 37 °C and the digested products were analyzed on 1.5 percent agarose gel along with 100 bp DNA ladder and documented in the gel documentation system (UVP BioSpectrum Imaging system, USA). The restriction digested trehalose biosynthesis genes were gel eluted, purified and sequenced on an ABI PRISM 377 genetic analyzer (Applied Biosystems Inc., USA).

### *2.5. In silico sequence analysis of trehalose biosynthesis genes*

The nucleotide sequences acquired were compared with database sequences using NCBI's BLAST (http://www.ncbi.nlm.nih.gov) program and were aligned and clustered using CLUSTALW [5]. In order to measure the percentage identities between nucleotide and amino acid sequences, alignments were imported into the GeneDoc program (https://genedoc.software.informer.com/2.7/) and the BioEdit 7.05 program (www.mbio.ncsu.edu/BioEdit/). Using the ProtParam tool (http://www.expasy.org/tools/protparam.html), the molecular masses and theoretical pI values of the polypeptides were predicted. Prediction of secondary structure was performed with the

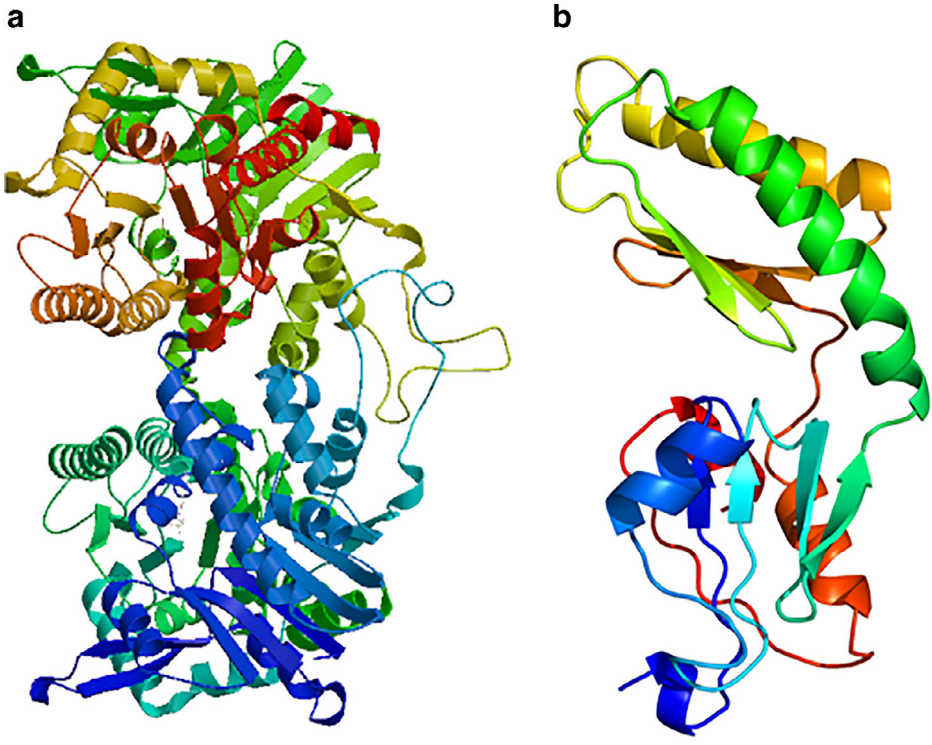


**Fig. 5.** (a) Three dimensional structure prediction of *otsA* gene using MODELLER program. (b) Three dimensional structure prediction of *otsB* gene using MODELLER program.

PROTEAN program (DNASTAR, Inc., Madison). The three dimensional structure was predicted through homology modelling approach using MODELLER program (Discovery Studio Modeling Environment 4.0, San Diego: Accelrys Software Inc., 2013).

## Declaration of Competing Interest

The authors declare no competing interests.

## Acknowledgements

The authors gratefully acknowledge the financial support given by the Ministry of Earth Sciences, Government of India, New Delhi to conduct the research. The authors are thankful to Dr. M. A. Atmanand, Director, Chennai for his constant support and encouragement to perform this research.

## References

[1] S. Kumar, K. Tamura, M. Nei, MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment, Brief. Bioinform. 5 (2004) 150–163.
[2] C. Schiraldi, I. Di Lernia, M. De Rosa, Trehalose production: exploiting novel approaches, Trends Biotechnol. 20 (2002) 420–425.

[3] A.J. Kutchma, M.A. Roberts, D.L.Crawford D.B.Knaebel, Small-scale isolation of genomic DNA from Streptomyces mycelia or spores, Biotechniques 24 (1998) 452–456.
[4] J Sambrook, D.W. Russell, A Molecular Cloning: A Laboratory Manual, 3rd ed., Cold spring harbour laboratory press, New York, 2001.
[5] J.D Thompson, T.J Gibson, F Plewniak, F Jeanmougin, D.G. Higgins, The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, NucleicAcids Res. 25 (1997) 4876–4882.