Chapter 5

Cloning of amyloid bioflocculant gene of *B. cereus* CR4

Chapter 5. Cloning of amyloid bioflocculant gene of *B. cereus* CR4

5.1 CDAG System

The amyloid fibrils present on the surface of bacteria has major role in attachment of cells on solid surfaces, cell aggregation and formation of biofilms. *E. coli* is much studied bacterium that produces a cell surface amyloid termed as curli. The curli amyloid has two major proteins, CsgA and CsgB (Gebbink *et al.,* 2005). The CsgA component is a major subunit protein that aggregates along the axis of beta sheets and forms long amyloid fibrils that are rigid and stable in harsh physical and chemical conditions. On the other hand, the subunit CsgB remains on the cell surface and anchors the fibrils of CsgA on the cell surface. Without the presence of CsgB the CsgA fibers would be released freely in the outside environment (Smith *et al.,* 2017). Hence CsgA subunit is often termed as major subunit of curli while the CsgB is termed as the minor subunit. Unlike mammalian amyloids that are formed because of mutations and protein misfolding the bacterial amyloids are encoded by the specific genes present in the genome. The genome of *E. coli* has curli operon that regulates the expression of genes required for biogenesis of curli amyloid (DeBenedictis *et al*., 2017). Besides controlled expression the transport of amyloid subunits on the cell surface is also regulated. The subunit CsgA has 22 amino acid signal sequence that ensures that the monomers are exported outside the cell. The signal sequence is identified by a transmembrane protein termed as CsgG which is encoded by the curli operon (Dueholm *et al*., 2012). In the absence of CsgG the CsgA monomers aggregates in the cytoplasm to form amyloid fibrils which inturn becomes lethal for the cells. Hence inspite of the fact that bacterial amyloids are functional in nature and helps the cells to accomplish particular function, they can be lethal if polymerized inside the cell. Thus, while cloning amyloid gene in bacteria, it becomes mandatory to ensure that the cloned amyloid gene (after expression) is secreted outside the cell. To ensure this phenomenon, Sivanathan and Hochschild (2013) have developed an *E. coli*-based system that ensures that the cloned gene product is secreted outside the cell. To ensure the phenotypic detection of clones for the presence of amyloid gene, an *E. coli* curli mutant strain is used. This strain has been named VS16 and it has deleted

curli operon. This deletion ensures that the recipient strain (VS16) is devoid of amyloid and produces white-coloured colonies when grown on Congo red agar. To ensure that the cloned gene product is transported outside the cell the VS16 strain has an over expression plasmid pVS76 that governs the expression of CsgG under IPTG induction. The presence of overexpressed CsgG in the cell membrane ensures that the cloned gene product is excreted outsite the cell (Sivanathan and Hochschild, 2013). The cloning vector pVS72 has the signal sequence of CsgA gene. This singal sequence is translated along with the cloned gene, into a fusion protein. The presence of CsgA signal sequnece with the protein ensures that the protein of interest is transported outside the cell by CsgG protein. The transported protein if amyloid, results in the development of red-coloured colonies on Congo red agar. The cloning system elaborated above is termed as curli dependent amyloid generator or CDAG system.

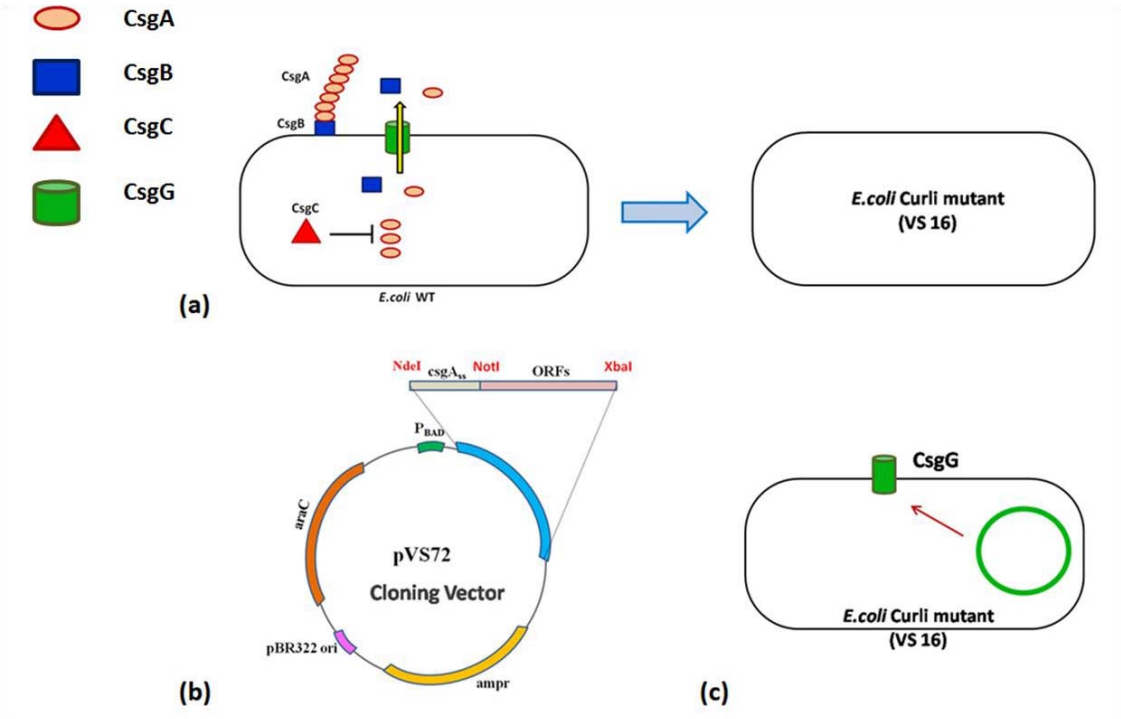

Figure 5. 1 Cloning strategy of amyloid genes in *E.coli* a) The use of *E.coli* curli mutant in cloning b) Map of vector pVS72 used for cloning c) The *E.coli* curli mutant pretransformed with overexpression plasmid that encodes CsgG

5.1.1 Applications of CDAG system

The curli dependent amyloid generator system have been widely employed for the detection of amyloid forming nucleotide sequences in several organisms. As the cloned nucleotide sequence is translated along with the CsgA signal sequence the translated product is exported outside the cell. The exported protein, if amyloid will adsorb Congo red and give red-coloured colonies which can be easily detected. On contrary if the cloned nucleotide sequence translates non-amyloid protein, then this results in white colony phenotype. Hence the system facilitates the identification of amyloid vs non amyloid sequences. Besides detection of amyloidogenic nucleotide sequences the CDAG system also allows the identification of different variants of single protein that shows tendency to form amyloids. The mutations in gene leading to formation of amyloids can be identified. In a similar way the library of chemical agents that can hinder or accelerate amyloidogenicity can be identified. The other widespread application includes the screen for open reading frames of genomic DNA libraries or cDNA libraries for the screening of unknown amyloid proteins. In one of such screening experiments it was demonstrated that the system can effectively identify amyloid forming gene sequences of *E. coli*. The curli dependent amyloid generator provides quick, efficient and unbiased method for detection of amyloid proteins. The screening time for red and white colonies after transformation of strain VS16 is less than 3 days. Besides screening based on colony morphology, the positive screens can be directly observed under transmission electron microscope for amyloid detection.

Having studied the characteristics of the purified amyloid bioflocculant CR4 and its applications, next it was proposed to clone the gene for the amyloid bioflocculant CR4. The CDAG system afforded a direct approach for cloning the amyloid gene which was attempted for *B. cereus* CR4 and the results of the same are discussed in this chapter

5.2 Materials and Methods

5.2.1 Genomic DNA isolation, amplification of ORF and cloning for identification of the amyloid bioflocculant gene

The amyloid bioflocculant gene was identified using CDAG system (curli-dependent amyloid generator) as described by (Sivanathan and Hochschild, 2013). The curli-dependent amyloid generator uses *E. coli* strain VS16 which has deleted curli operon and harbors a plasmid that over expresses csgG under IPTG induction. CsgG is a trans membrane protein that identifies signal sequence of curli amyloid and exports it outside the cell. The gene of interest (in this case the amyloid bioflocculant CR4 gene) is cloned in plasmid pVS72 that translates a fusion product of the gene of interest and the curli signal sequence which ensures its export outside the cell. The gene of interest, if amyloid produces red colored colonies of *E. coli* V16 on Congo red agar plates. *B.cereus* CR4, culture was grown overnight in 5ml Luria broth, pH 7 at 37ºC for 24 h. The culture was collected in 1.5 ml centrifuge tube and centrifuged at 8000g for 15 min to obtain biomass. The biomass was treated with 0.5 ml DNA extraction buffer (100 mM Tris Cl pH 8, 50 mM EDTA pH 8 and 1% SDS) followed by addition of 5µl lysozyme (10 mg/ml). The mixture was kept at 37ºC for1 hr, followed by treatment with equal volume of equilibrated phenol (pH 8.0) followed by gentle mixing. Centrifugation was carried out at 8000g for 15 min, and the aqueous phase was collected and treated with equal volume of chloroform isoamyl alcohol (24:1) and centrifuged at 8000g for15 min. The aqueous phase was collected and treated with 50µl 3M sodium acetate and 300µl chilled isopropyl alcohol. The resulting precipitates of gDNA were collected by centrifugation at 8000g for 15 min, followed by washing with 70% ethanol. The gDNA precipitate was air dried and dissolved in 10mM TE buffer (Sambrook and Russel., 2001).

The plasmid pVS72 was isolated from *E. coli* DH5α by alkaline lysis method (Sambrook and Russell, 2001). In order to amplify the ORFs from the genomic DNA, following primers as suggested by Chen *et al*. (1994) were designed ShnF:AAGCGGGCCGNAGGAGGAAATG.

RanR: GGGGGTCTAGATNGTAGNGCACNG. The forward primer was made specific for the Shine-Dalgarno sequence, to amplify the ORFs so that inframe sequences are amplified. The reverse primer was a random primer. The PCR amplification of the ORFs from the gDNA obtained as described earlier was carried out with following conditions. The PCR system was-Initial denaturation at 94ºC for 5 min, Denaturation at 94ºC for 30 sec, Annealing

at 65ºC for 30 sec and Amplification at 72ºC for 1 min. The amplicons obtained by PCR and the plasmid mixture was digested with restriction enzymes NotI and XbaI respectively for 4 to 5 h at37ºC to obtain the construct pVR72ORF. The restriction enzymes were denatured by heating at 68ºC and the mixture was subjected to ligation at 20ºC, with T4 DNA ligase.

5.2.2 Cloning of bioflocculant CR4 gene

Transformation of pVR72ORF was done by electroporation method. *E. coli* VS16 strain was grown in Luria broth till its OD600nm reached 0.8. The biomass was harvested by centrifugation at 8000g for 15 min at 4ºC and was washed with 10% glycerol, three times at 4ºC, and re-suspended in 1 ml 10% glycerol too obtain electrocompetent cells. For transformation about 50µl electrocompetent cells were taken in prechilled cuvette along with 10µl of ligation mixture for transformation. Electroporation was performed at1.2 kV for 5ms, following which 1 ml Luria broth was immediately added and the mixture was incubated at 37ºC for 1 h. The screening of the transformants was done on selective media of Luria agar supplemented with 100µg/ml Ampicillin and 50µg/ml Chloramphenicol. For screening for amyloid positive clones the colonies obtained after electroporation were screened on induction medium containing Luria agar with Congo red (0.005%), supplemented with Ampicillin (100 µg/ml), Chloramphenicol (50µg/ml), 1mM IPTG and 0.2% (wt/vol) L-arabinose. Red colored colonies on this medium indicate positive clone containing amyloid producing gene in the plasmid. The amyloid positive transformant was grown in Luria broth and plasmid isolation was performed by alkaline lysis method (Sambrook and Russell 2001). The plasmid was used for PCR amplification of the gene of interest using the flanking primers specific for plasmid. The forward primer was CCTGACGCTTTTTATCG and the reverse primer was AGGCTGAAAATCTTCTCTC. The amplicon so obtained was outsourced for sequencing for identification of the gene (SciGenome Labs, Kochi India).

5.2.3 Cloning of tasA gene from *B. cereus* CR4

For the cloning of tasA gene, the genomic DNA from *B. cereus* CR4 was isolated as described in section 5.2.1 followed by PCR amplification of tasA gene using tasA specific Forward

AAAAAAGCGGCCGCGNAGGAANCNATG and Reverse AAAAAATCTAGATNGTAAANCNC primers. PCR amplification and plasmid isolation was carried out as described in section 5.2.1. The PCR amplicon and the plasmid obtained was digested with NotI and XbaI, followed by ligation and transformation (section 5.2.2).

5.3 Results and Discussion

5.3.1 Identification of amyloid bioflocculant CR4 gene

For the identification of amyloid bioflocculant gene present in *B. cereus* CR4 as mentioned above C-DAG system was used. This system comprises a cloning plasmid pVS72 and *E. coli* curli mutant VS16 which gives white-coloured colonies on Congo red agar plates. The VS16 strain harbours a plasmid that over expresses CsgG protein under IPTG induction. This protein is a transmembrane protein that assures the export of cloned protein outside the cell. The screening of clones was accomplished on the Congo red agar plates containing IPTG and arabinose as inducer. The IPTG and arabinose induces csgG and the cloned insert respectively. If the inserted fragment is an amyloid the colonies would appear red in colour (Sivanathan and Hochschild, 2013). Hence the C-DAG system is an effective system for testing particular target proteins for their amyloid-forming tendency, to identify mutations that can modulate target protein amyloidogenicity and to screen open reading frame (ORF), genomic or cDNA libraries, with amyloidogenic proteins. With this purpose the bacterial genomic DNA was isolated and used for PCR amplification of ORFs in the genome. As one of the primers used for the amplification of ORFs was a random primer, the resultant amplicons obtained were of different size. As demonstrated in Figure 5.2b the multiple bands observed after agarose gel electrophoresis showd amplification of ORFs of different sizes giving multiple bands (Figure. 5.2a, b). The cloning vector was digested with NotI and XbaI (Figure. 5.2c, d) respectively and ligated. Similarly, the ORFs were digested with NotI and XbaI and ligated in the cloning vector.

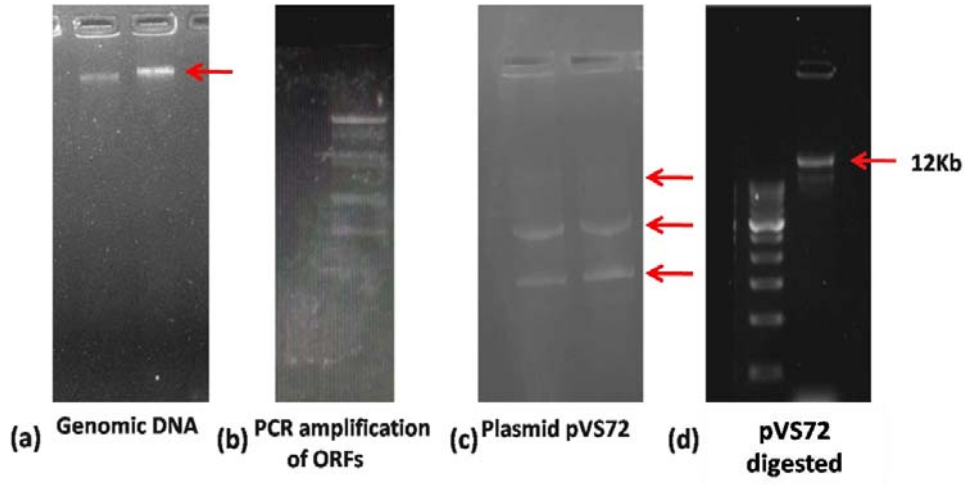

Figure 5. 2 Cloning strategy for the identification of amyloid gene in *B. cereus* CR4 (a) Genomic DNA of *B.cereus* CR4. b) PCR amplification of ORFs from the genomic DNA. c) Isolation of plasmid pVS72. d) Restriction digestion of plasmid pVS72 with NotI and XbaI.

After transformation the colonies obtained were transferred on Congo red induction media containing IPTG and arabinose. Presence of IPTG induces expression of CsgG transmembrane protein while presence of arabinose induces expression of gene of interest ligated in the plasmid. Most of the colonies obtained on congored selection plates were white, indicating absence of amyloid gene. One of the colonies showed red colored and hence was considered as positive clone or transformant (Figure. 5.3, 5.4). This was further subcultured on Congo red induction media along with empty vector as a control (Figure. 5.5a and 5.5b). The red colored colony was further subcultured and tested for flocculation activity. The flocculation activity associated with the positive transformant is depicted in the figure 5.6. The positive clone showed clear flocculation activity after 10 min as depicted in figure 5.6b. Using flanking primers specific for the plasmid the insert was amplified and sent for sequencing. The sequencing results showed the presence of tasA amyloid in the insert (Table. 5.1).

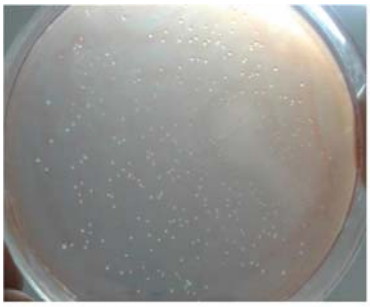

Figure 5. 3 Screeing of transformants on Congo red agar.

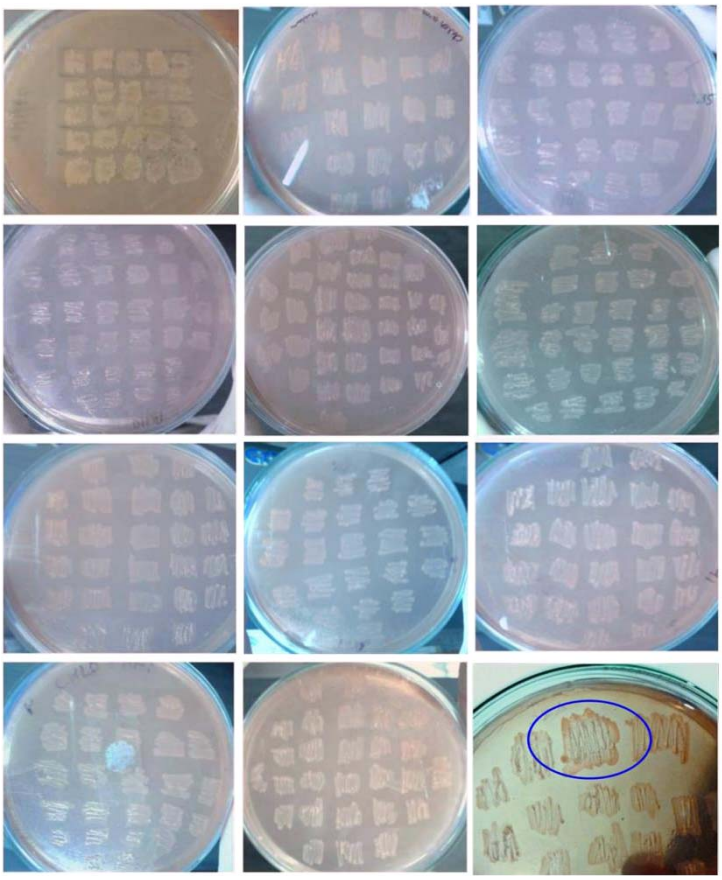

Figure 5. 4 Screening for amyloid positive clone on Congo red agar.

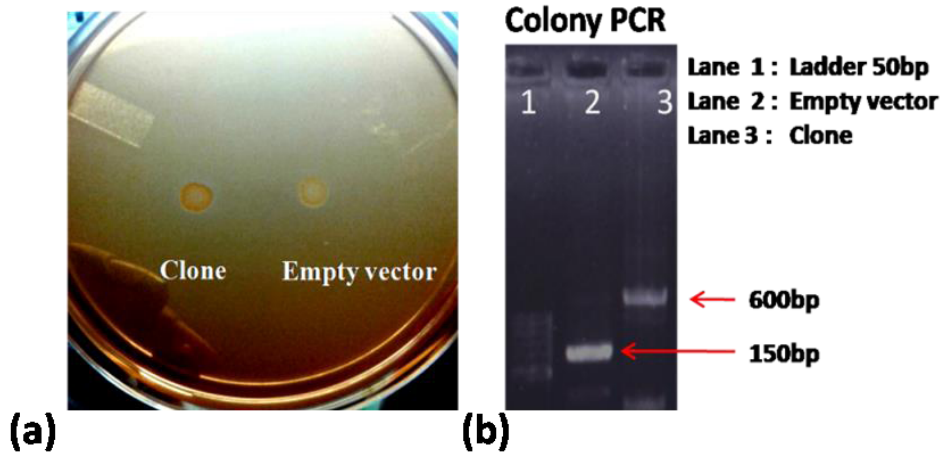

Figure 5. 5 Confirmation of amyloid positive clone on Congo red agar and colony PCR for identification of the insert. a) Positive clone and control on Congo red agar. b) Colony PCR of insert using flanking primers specific for cloning site (between NotI and XbaI) of plasmid pVS72 giving 600bp amplicon of gene of interest.

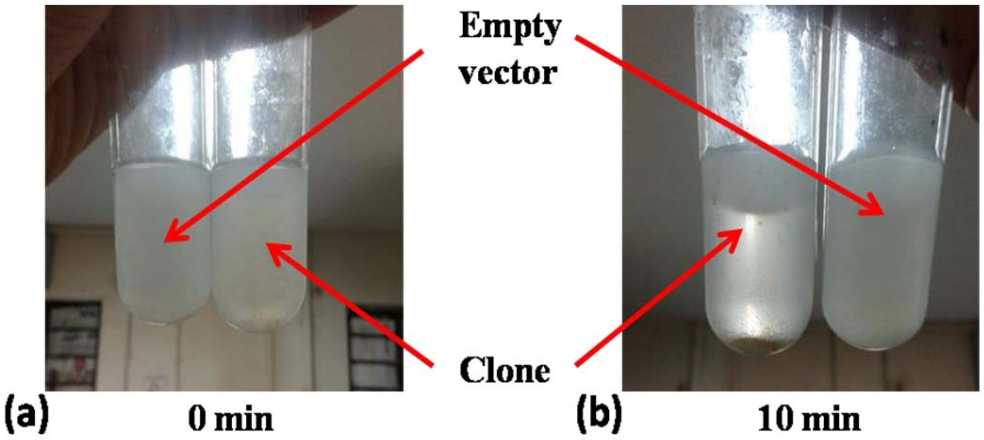

Figure 5. 6 Study of flocculation activity of the positive clone. a) Flocculation of kaolin by *E.coli* carrying empty vector pVS72 (0 min). b) Flocculation of kaolin by positive clone (10 min) showing flocculation activity.

Table 5. 1 BLAST analysis of the amyloid gene from *B. cereus* CR4 cloned in plasmid pVS72 showing flocculation activity.

Description	Max score	Total score	Query cover	E value	Ident	Accession
Bacillus subtilis tasa gene	531	531	99%	8e-147	89%	AJ871386.1
Bacillus subtilis strain ge28, complete genome	529	529	99%	3e-146	89%	CP021903.1
Bacillus subtilis strain TLO3, complete genome	529	529	99%	3e-146	89%	CP021169.1
Bacillus subtilis subsp. subtilis strain D12-5, complete genome	529	529	99%	3e-146	89%	CP014858.1
Bacillus sp. YP1, complete genome	529	529	99%	3e-146	89%	CP010014.1
Bacillus subtilis subsp. subtilis str. BSP1, complete genome	529	529	99%	3e-146	89%	CP003695.1
Bacillus subtilis strain DKU_NT_03, complete genome	523	523	99%	1e-144	89%	CP022891.1
Bacillus subtilis strain DKU_NT_02, complete genome	523	523	99%	1e-144	89%	CP022890.1
Bacillus subtilis strain GS 188 genome	523	523	99%	1e-144	89%	CP022391.1
Bacillus subtilis strain SR1 genome	523	523	99%	1e-144	89%	CP021985.1
Bacillus subtilis subsp. subtilis strain SRCM101392, complete genome	523	523	99%	1e-144	89%	CP021921.1
Bacillus subtilis subsp. subtilis strain SRCM100333, complete genome	523	523	99%	1e-144	89%	CP021892.1
Bacillus subtilis subsp. subtilis strain SRCM100761, complete genome	523	523	99%	1e-144	89%	CP021889.1
Bacillus licheniformis strain SRCM101441, complete genome	523	523	99%	1e-144	89%	CP021507.1

5.4 tasA gene cloning of *B. cereus* CR4

B. cereus CR4 genomic DNA was isolated and used for PCR amplification of tasA gene, using tasA specific primers (Figure. 5.7 a, b). The tasA amplicon and the cloning vector pVS72 were digested with NotI and XbaI respectively (Figure. 5.7c) and ligated. After transformation of ligated plasmid, the colonies obtained were transferred on Congored induction media containing IPTG and arabinose where the positive colony showed red color (Figure. 5.8a). The presence of tasA in vector was confirmed by PCR amplification of tasA from the cloned plasmid (Figure. 5.8b). This confirmed the presence of tasA gene in *B. cereus* CR4.

The gene for bioflocculant CR4 was cloned successfully. It was confirmed that the gene for amyloid bioflocculant CR4 produced by *B. cereus* CR4 was tasA. The tasA gene was further cloned and amplified from *B. cereus* CR4 by tasA genes pecific primers for confirmation of presence of tasA gene. In the earlier chapter the bioflocculant protein was identified as TasA by LC-MS/MS.

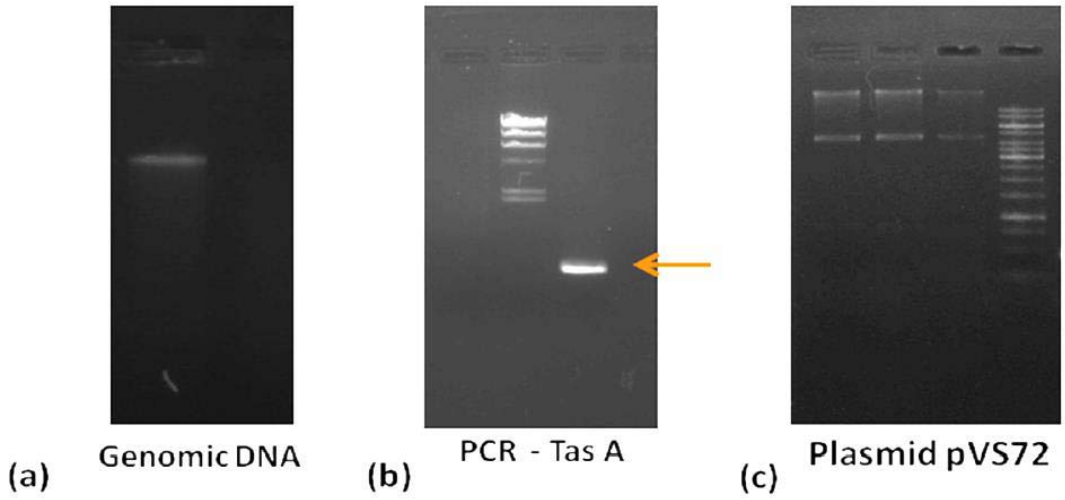

Figure 5. 7 Steps in cloning of tasA gene from *B. cereus* CR4. a) Genomic DNA. b) amplification of tasA from genomic DNA. c) plasmid pVs72 used for cloning.

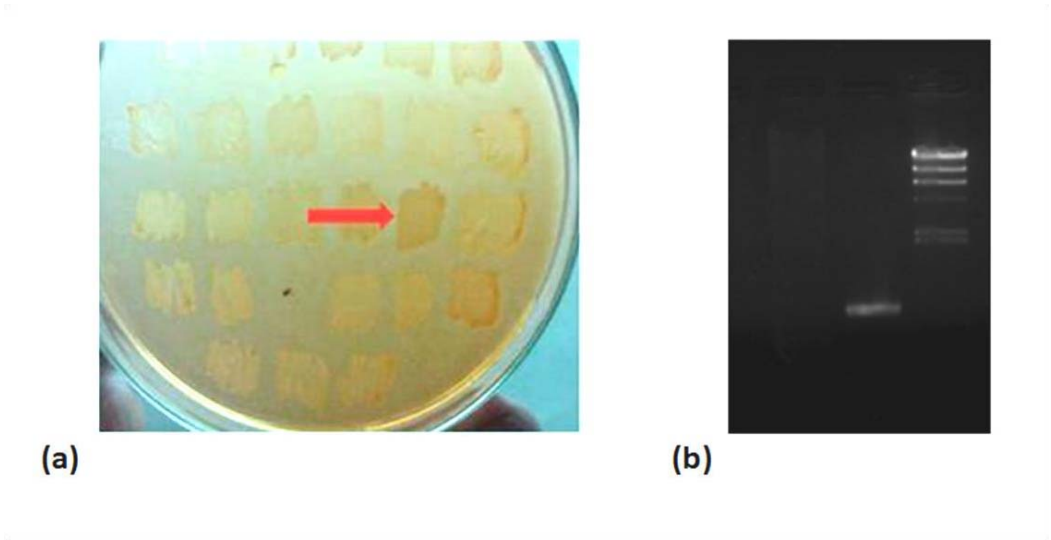

Figure 5. 8 Congo red plate assay for the identification of clone with an insert of amyloid gene. b) Amplification of tasA gene in clone using tasA specific primers. Lane1 – empty vector, Lane 2 – Cloned vector, Lane 3 – Lambda HindIII ladder

References

Sambrook, J. and Russell, D. W. (2001). Molecular cloning: a laboratory manual, Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press.

Sivanathan V. and Hochschild, A. (2013). A bacterial export system for generating extracellular amyloid aggregates. Nature protocols, 8(7), 1381-1390.

Smith, D. R., Price, J. E., Burby, P. E., Blanco, L. P., Chamberlain, J. and Chapman, M. R. (2017). The production of curli amyloid fibers is deeply integrated into the biology of *Escherichia coli.* Biomolecules, 7(4), 75.

DeBenedictis, E. P., Ma, D., and Keten, S. (2017). Structural predictions for curli amyloid fibril subunits CsgA and CsgB. Rsc Advances, 7(76), 48102-48112.

Dueholm, M. S., Albertsen, M., Otzen, D. and Nielsen, P. H. (2012). Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. PloS one, 7(12), e51274.