Chapter 5: Discussion

5.1 Extraction of mgDNA from goat rumen digesta

The goat rumen digesta mgDNA was extracted using the protocol P5. The protocol involves cell lysis to disrupt the bacterial cell wall followed by incubation at 65°C for 2 h. Cell lysis is a critical step in genomic DNA extraction to break open the cell wall and membrane of microbes to release the DNA. Although the efficiency of cell lysis has improved with the newly designed protocols, complete cell lysis during genomic DNA extraction still remains a bottleneck and the extraction bias still exists (Frostegard *et al.*[137]). The extraction protocol P5 is based on chemical and enzymatic lysis that is relatively gentle, exhibits minimal DNA shearing and often discriminates against particular cell types[138]. The polysaccharides and tannins contained within the goat rumen can inhibit PCR amplification and lead to erroneous interpretations[139, 140]. In the present study, the cell lysis buffer contains cetyltrimethylammonium bromide (CTAB) to remove polysaccharides, polyphenols, and other secondary metabolites. Besides this, the long-tail surfactant, CTAB introduces a conformational change from random coil to compact globule in the DNA, making the precipitation of DNA more effective. The removal of tannins was achieved by using polyethylene glycol (PEG). PEG forms an aggregate with tannins and other phenolics, which bind to the protein and cell debris upon lysis, forming a complex. This lysate, when centrifuged in the presence of phenol, results in the accumulation of protein-tannin complex at the interface between the organic and aqueous phases. The supernatant, thus obtained, is largely free from tannins[126].

An important criterion in the environmental microbiology study is the high yield of quality mgDNA. The sheared DNA is not suitable for cloning and can cause PCR amplification artifacts[141]. The P5 protocol was consistently able to recover pure (Table 1) and high-molecular-weight mgDNA (>23kb) having high integrity as represented by non-sheared DNA on the agarose gel electrophoresis (Fig. 4.1).

This is probably the first report of getting contaminant-free and clonable mgDNA from goat rumen digesta with no further need for the purification. The available protocols so far may yield high-molecular-weight mgDNA but

subsequent purification results in loss of DNA thereby negating the prospect of cloning and sequencing.

5.2 Comparison of the extraction methods

Out of five mgDNA extraction protocols (P1-P5), the protocol P5 yielded the highest amount of mgDNA i.e., 240.6 ± 8.1 μg/g goat rumen digesta (Fig. 4.2 and 4.3). The protocols P1, P2 and P3 failed to produce quality mgDNA that could neither be digested with the restriction enzymes nor could be subjected to PCR (Table 4.2, Fig. 4.4). The protocol P4 developed by Popova *et al*.[127] could yield quality DNA suitable for restriction digestion and PCR but its processing time was too long (>8h). The protocol P5 took only 3.5 h and resulted in quality DNA that was suitable for the restriction digestion, PCR amplification as well as cloning (Table 4.2, Fig. 4.4 and 4.5). Thus, the protocol P5 was adjudged to be the best for the extraction of mgDNA from the goat rumen digesta with the least protein and tannin contamination and higher yield of quality DNA. Henderson *et al*.[142] reported that the choice of DNA extraction method affects the rumen microbial community structure in ways that would prove to be statistically and biologically significant, even though the mgDNA was isolated from sub-samples of the same homogenised rumen sample. Thus, the choice of mgDNA extraction method is vital in assessing the true microbial diversity of any environmental sample.

5.3 Genomic DNA isolation from culturable bacteria

The genomic DNA from the culturable bacteria was isolated using the P5 extraction protocol (Fig. 4.2). It could give high yield of DNA from *E. coli* (MTCC 40), thus expanding the scope of usage of this protocol. Sagar *et al*.[130] studied the suitability of the mgDNA extraction protocol for the extraction of genomic DNA from the culturable bacteria. The mgDNA extraction protocol could give high yield of genomic DNA in the case of gram positive and gram negative bacteria.

5.4 Construction of mgDNA library

The mgDNA library was constructed so as to screen it for the cellulase encoding gene(s). For the purpose, the mgDNA was partially digested with *Bam*HI restriction enzyme, electrophoresed and DNA fragments ranging from 0.5-2.0 kb were fractionated and used for mgDNA library construction (Fig. 4.6). The pUC19

plasmid was used to clone the mgDNA fragments (Fig. 4.7 and 4.8) and subsequently transferred to the competent cells of *E. coli* DH5α with the transformation efficiency of ~1 x 10^6 cells/µg DNA. The library consisted of ~24,000 clones. Similar mgDNA libraries from goat rumen digesta were constructed earlier by Lim *et al*.[89], Pushpam *et al*.[90], Wang *et al*.[92], Cheng *et al*.[93] and Denman *et al*.[143].

A total of 11 clones, out of 24,000 clones, were randomly selected to determine the quality of the mgDNA library (Fig. 4.9). The quality of the library could be considered as good if all clones possess inserts, that too of the different sizes. The restriction digestion of the recombinant pUC19 plasmid from all the selected recombinant colonies with *Kpn*I showed shifting of the bands in the gel, indicating the presence of different lengths of DNA inserts (Fig. 4.10). The present study corroborates with the study of Duan *et al*.[119] who analysed the quality and insert sizes of the mgDNA library from the buffalo rumen by restriction digestion of the recombinant cosmids from 14 randomly selected bacterial colonies.

The goat rumen digesta mgDNA library was maintained in *E. coli* DH5α competent cells.

5.5 Screening of mgDNA library for cellulolytic clone(s)

Lee *et al*.[144] reported that the ruminal microbiota is a rich source of cellulolytic enzymes, which is important not only for the feed but also for the bioenergy industry. The goat rumen microbiota is well adapted to convert plant biomass into nutrients. This makes the goat rumen an ideal place to find microorganisms and enzymes specialized in the plant cell wall degradation.

In the present investigation, the mgDNA library was screened for cellulase producing clones using carboxymethyl cellulose (CMC) as the substrate. On screening 2,000 *E. coli* DH5α recombinant colonies, only one was found to exhibit a distinct yellow halo around the colony after 24-36 h of incubation at 37°C, accordingly it was considered to harbour the cellulolytic recombinant plasmid and designated as 'T3' clone (Fig. 4.11). Duan *et al*.[119] screened 15,000 clones of a buffalo rumen mgDNA library and 11 independent clones were found to be cellulolytic. Cheng *et al*.[93] reported six endoglucanase clones after screening 12,672 clones. The reason for the low number of cellulolytic clones in the present

study might be due to the failure of the mgDNA to express heterologously or the expression being too low to be detected[119].

5.6 Cloning, expression and purification of metagenomic cellulase

The recombinant bacterial clone T3 possessing the cellulolytic activity was used for the isolation of the recombinant pUC19 plasmid DNA. The recombinant plasmid was isolated (Fig. 4.12) and sequenced using M13 Forward (-20) and M13 Reverse (-24) primers. The sequencing data revealed the size of the DNA insert to be 1,501 bp. The same was confirmed by digesting the recombinant pUC19 plasmid with *Bam*HI which showed a DNA insert of ~1.5 kb in size (Fig 4.13). The open reading frame (ORF) was identified via ORF Finder of NCBI (https://www.ncbi.nlm.nih.gov/orffinder/). The ORF consisted of 1,122 bp (Fig. 4.14). BLASTn analysis showed the ORF to be a cellulase encoding gene. This cellulase encoding gene was designated as 'CelT3'.

The cellulase encoding gene was amplified from the recombinant clone T3, harbouring the recombinant plasmid, using a set of primers having *Nde*I and *Hin*dIII restriction sites in the forward and reverse primers, respectively (Fig. 4.15). The CelT3 gene digested with *Nde*I and *Hin*dIII restriction enzymes was subsequently cloned into the *Nde*I and *Hin*dIII digested-pET28a(+) expression vector (Fig. 4.16 and 4.17). According to Mierendorf *et al.*[145] pET vectors are the most powerful vectors used for the expression of recombinant proteins in *E. coli*. The desired protein encoding genes are cloned into pET plasmids under the control of strong bacteriophage T7 promoters. Genetically engineered *E. coli* BL21(DE3) host cells provide T7 RNA polymerase. On optimal induction, this enzyme is so active that the desired protein can comprise more than 50% of the total cell protein within few hours after induction.

In the present investigation, the CelT3 gene was cloned into pET28a(+) and then transferred to *E. coli* DH5α cells (Fig. 4.18). The recombinant plasmid was digested with *Nde*I and *Hin*dIII restriction enzymes to confirm the presence of CelT3 gene insert (Fig. 4.19). The recombinant plasmid [pET28a(+)containing CelT3 gene] was isolated from the *E. coli* DH5α transformants and designated as 'pET28CT3' (Fig. 4.20). A genomic library maintained in the expression host *E. coli* BL21(DE3) shows instability owing to the presence of functional endonuclease and recombinase pathways. As a result, plasmid DNA stored in *E.*

coli BL21(DE3) cells gets damaged or rearranged after some time. *E. coli* DH5α cells do not contain the T7 RNA polymerase encoding gene, thus eliminating the chances of plasmid instability due to the production of proteins toxic to the host cells (Mierendorf *et al.*[145]). Therefore, pET28CT3 recombinant plasmid was kept in *E. coli* DH5α cells and preserved at -80°C.

On being established in *E. coli* DH5α cloning host, the CelT3 protein expression was initiated by transferring the pET28CT3 recombinant plasmid into *E. coli* BL21(DE3) expression host containing a chromosomal copy of the T7 RNA polymerase gene under the lacUV5 promoter. The recombinant clones were assayed for cellulolytic activity (Fig. 4.21).

The *E. coli* BL21(DE3) culture harbouring the recombinant pET28CT3 was grown at 37°C in orbital shaker till optical density (OD) reached the log phase i.e. $\mathbf{A_{600}}$**~0.5**. The culture was then induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 5h at 37°C. IPTG is added to the culture medium when there is sufficient population density of *E. coli* BL21(DE3) cells. The early induction with IPTG might be toxic to the cells if the population density is too low. Furthermore, it might result in the absorption and subsequent accumulation of IPTG inside the bacterial cells that might hinder the cell division.

The size of the recombinant CelT3 protein, fused with N- terminal His-tag, was calculated to be ~44.909 kDa including the insert sequence encoding 373 amino acids (42.727 kDa) and pET28a(+) vector backbone sequence contributing 20 amino acids (2181.34 Da). On running the CelT3 protein on SDS-PAGE, a band of ~45 kDa was observed which corresponds to His-tagged CelT3 with vector backbone. Thus the molecular weight of the CelT3 protein observed on the SDS-PAGE corroborates the calculated molecular weight of CelT3 (Fig. 4.22).

In order to determine the total protein content and enzymatic activity of the crude extract, cell lysis was carried out. The total protein concentration and specific activity of cellulase enzyme in the cell lysate was calculated to be 746.32 $\mu g\ ml^{-1}$ and 4.01 $U\ mg^{-1}$, respectively (Table 4.6).

The soluble protein was purified on Ni-NTA agarose column based on the phenomenon of affinity chromatography. The recombinant protein contained a polyhistidine tag to allow highly homogeneous purification by a Ni-NTA column. Poly-histidine tag (His-tag) is the most commonly used affinity tag for collecting

large amounts of highly purified recombinant proteins. The phenomenon is based on the affinity of Nickle (Ni^{2+}) ions for the Histidine residues. Ni^{2+} shows the highest affinity and selectivity for His-tag and is therefore preferred. The hydrophilic and flexible nature of His-tag often increases the solubility of the target protein and rarely interferes with its biological activity. Polyhistidine-tag consists of at least six histidine (His) amino acid residues, either at the N- or C-terminus of the protein. It is known by the trademarked name His-tag. pET28a(+) plasmid DNA contains a His-tag at its N-terminus.

Optimization of protein purification is based on a number of factors, including the amount of 6x His-tagged protein required and the expression level[146]. The imidazole rings in the histidine residues of the 6x-His tag bind to the Ni^{2+} ions immobilized by the NTA groups on the matrix. Imidazole itself can also bind to the nickel ions and disrupt the binding of dispersed histidine residues in non-tagged background proteins. At low imidazole concentration, the non-specific, low affinity binding of the background proteins is prevented, while His-tagged proteins still bind strongly to the Ni-NTA matrix. Therefore, adding imidazole to the lysis buffer leads to greater purity in fewer steps. Since the interaction between Ni-NTA and His-tag of the recombinant protein does not depend on the tertiary structure, proteins can be purified either under the native or denaturing conditions.

In order to assess the purification parameters viz. specific activity, yield (%) and purification fold of the CelT3 recombinant protein, it was purified followed by increase of concentration with the use of Centricon. Enzymatic activity of the purified recombinant protein using Ni-NTA column was calculated to be 2.38 U ml^{-1} with a specific activity of 17.74 U mg^{-1} (Table 4.6). These findings are in agreement with the study of Cheng *et al*.[93] who characterized a non-specific endoglucanase from a metagenomic library of the goat rumen digesta with the highest specific activity of 20.56 U mg^{-1} towards CMC as the substrate.

The nucleotide sequence of the CelT3 cellulases was translated into amino acid sequence using translate tool of ExPASy online server (http://expasy.org/translate) and consisted of 373 amino acid residues (Fig. 4.23). The deduced amino acid sequence of CelT3 was used to perform a BLASTp search and revealed 97% amino acid sequence homology with the cellulase enzyme of the uncultured bacterium (accession no. AFJ05146). The phylogenetic tree of the

CelT3 cellulase showed the closest relationship to a metagenome-derived cellulase 'Cel5a' (Fig. 4.24).

The multiple sequence alignment of the CelT3 cellulase showed that it contained a glycosyl hydrolase family 5 (GH5) catalytic domain (Fig. 4.25). The activities of GH5 have been identified as the most abundant cellulases from both the cultured as well as the uncultured ruminal bacterial species[119]. The GH5 cellulases were also isolated from mgDNA libraries from other environments supporting the view that the GH5 family cellulases tend to predominate in the nature[83, 147].

The homology modelling was performed using sequence alignment of amino acids of CelT3 against the protein data bank. Based on the alignment, a model of the CelT3 cellulase with 21 coils, 11 β-strands and 10 α-helices (Fig. 4.26) was built using uncultured bacterium cellulase (PDB ID: 4HTY) as the template.

The Ramachandran plot for CelT3 predicted the protein structure in the region of possible angle formations by ϕ (phi) and ψ (psi). As per the three regions of the Ramachandran plot, 96.6% amino acid residues were in the favoured, 2.8% in the allowed and 0.6% in the outlier regions (Fig. 4.27). The plot was significant as >96% percentage of residues remained in the favoured region postulating that the protein is stable.

5.7 Biochemical characterization of purified CelT3 cellulase

5.7.1 Effect of substrate, pH and temperature on CelT3 activity

The CelT3 enzyme showed the maximum activity in the presence of carboxymethyl cellulose (Table 4.7) suggesting its endoglucanase nature. The hydrolytic activity towards Avicel, filter paper and laminarin was found to be less as compared to CMC suggesting that the enzyme was more specific to amorphous cellulose than crystalline cellulose. The CelT3 cellulase displayed maximum enzyme activity at pH 6.0 and a relatively high activity at pH 5.0-7.0 varying from 71% to 88%, respectively. The relative enzyme activity was retained significantly. A sharp decline in enzyme activity 2% was observed at pH 4.0. The relative enzyme activity at pH 8.0 was 70% and the same declined to 25% at pH 10.0 (Fig. 4.28). The stability of CelT3 in the alkaline pH demonstrated suitability for its

incorporation in the commercial laundry detergent formulations which requires the enzyme to remain active at alkaline pH.

The CelT3 enzyme exhibited the maximum residual activity at 45°C temperature and 32% relative activity even at a temperature as low as 15°C. Similarly at 65°C, the recombinant cellulase displayed 27% relative enzyme activity and the same declined to 2.5% at 75°C (Fig. 4.29).

Interestingly, the average pH and temperature of the goat rumen is around 6.8 and 39°C, respectively, which means that the CelT3 cellulase enzyme has a higher temperature optimum and lower pH optimum than the rumen environment. However, similar cases were also reported earlier; a lipase from a cow rumen mgDNA library was found to have an optimum pH of 5.5 and temperature of 50°C. Liu *et al.*[148] reported an endoglucanase showing temperature and pH optima of 55°C and 5.5, respectively.

5.7.2 Effect of metal ions

Most of the enzymes tend to incorporate the divalent cations and the transition metal ions within their structure to stabilize the folded conformation. The commercial laundry detergent contains metal chelators as important ingredients to remove metal ions and decrease water hardness[149, 150].

The divalent metal ions Ca^{2+} and Fe^{2+} enhanced the activity of recombinant cellulase. Ni^{2+} mildly inhibited the activity. Most of the cellulases reported in the literature are Ca^{2+} and Mg^{2+} dependent. In the present investigation, the CelT3 enzyme was found to be Ca^{2+} dependent but independent of Mg^{2+} (Fig. 4.30). Since water hardness depends on the presence of Ca^{2+} and Mg^{2+}, the CelT3 enzyme could be applicable in the laundry detergent formulations.

5.7.3 Determination of K_m and V_{max} for the enzyme catalysed reactions

The rate of the chemical reaction is affected by the concentration of the enzyme as well as the substrate. The rate of enzyme reaction increases with the increase of the substrate concentration. At high or low enzyme concentrations, the available enzyme active sites could be occupied by the substrates. Therefore, with further increase in the substrate concentration, there would not be further change in the rate of reaction. In other words, if substrate concentration is constant against the increase of the enzyme there would be increase in the enzymatic activity

linearly upto a particular concentration, beyond which it would remain constant because all the active sites of the enzymes would be occupied by the substrate.

In the present investigation, the recombinant protein CelT3 showed the maximum residual activity in the presence of 1% CMC (Fig. 4.31). The activity increased linearly upto 1% of CMC, beyond which there was no further increase in the rate of the reaction.

The kinetic parameters for the CelT3 enzyme were characterized in the presence of the optimum substrate i.e., CMC using Michaelis–Menten kinetics model and Lineweaver-Burk plot. The apparent V_{max} and K_m values of the enzyme catalysed reaction were determined to be 222.22 U mg^{-1} and 0.35 mg ml^{-1}, respectively (Fig. 4.32). The study demonstrated a lower K_m value and a higher V_{max} value as compared to the earlier reported cellulases through metagenomic approach[119]. The lower K_m values indicated the high binding affinity of the CelT3 enzyme towards the optimum substrate CMC.

The CelT3 cellulase with large scale production may be used for degrading the lignocellulosic biomass, in the production of energy-rich and nutritionally improved animal feed, and as a co-additive in pulp bleaching in pulp and paper industries.