5.1A Introduction

Cellulose is the most common organic polymer, representing about 1.5 X 10^{12} tons of the total annual biomass production through photosynthesis especially in the tropics, and is considered to be an almost inexhaustible source of raw material for different products (Kuhad et al. 2011). It is the most abundant and renewable biopolymer on earth and the dominating waste material from agriculture (Jarvis 2003). At the molecular level, cellulose is a linear polymer of glucose composed of an hydro glucose units coupled to each other by β-1-4 glycosidic bonds. The number of glucose units in the cellulose molecules varies and degree of polymerization ranges from 250 to well over 10,000 depending on the source and treatment method (Riswan ali et al. 2013). The nature of cellulotic substrate and its physical state are important factors in its enzymatic hydrolysis. Though lignocellulosic biomass is generally recalcitrant to microbial action, suitable pretreatments resulting in the disruption of lignin structure and increase accessibility of enzymes have been shown to increase the rate of its biodegradation (Naveena et al. 2005). Microbial degradation of lignocellulosic waste and the downstream products resulting from it is accomplished by a concerted action of several enzymes, the most prominent of which are the cellulases, which comprise several different enzyme classifications. Cellulases hydrolyze cellulose (β-1, 4-D-glucan linkages) and produce as primary products glucose, cellobiose and cello-oligosaccharides. There are three major types of cellulase enzymes [Cellobiohydrolase (CBH or 1,4 D-glucan cellbiohydrolase, EC 3.2.1.91), Endo 1,4-glucanase (EG or endo-1,4-β-D-glucan 4- glucanohydrolase, EC 3.2.14) and β-glucosidase (BG-EC 3.2.1.21)] (Table 5.1A).
Table 5.1A Bacterial cellulase enzyme system (Sadhu and Maiti, 2013)

<table>
<thead>
<tr>
<th>Enzyme Description</th>
<th>E. C. Number</th>
<th>Reaction</th>
<th>Other Names</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) Endo -1,4 β-D-glucan gluconohydrolase</td>
<td>E. C. 3. 2. 1, 4</td>
<td>Cut at random at internal amorphous sites of cellulose generating oligosaccharides of various lengths. It acts on Endo-1, 4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.</td>
<td>Endogluconase, Endo-1,4-β-glucanase, Carboxymethyl cellulase, β -1,4-endoglucan hydrolase, Endocellulase</td>
<td>5, 6, 7, 8, 10, 12, 44, 51, 61, 74</td>
</tr>
<tr>
<td>ii) Exoglucanase or 1,4-β-D-glucan cellobiohydrolases (cellobiohydrolases)</td>
<td>E.C.3.2.1.91</td>
<td>Hydrolysis of 1,4-beta-D glucosidic linkages in cellulose and cellotetraose, releasing cellobiose from the non-reducing ends of the chains</td>
<td>Exoglucanase, Exocellobiohydrolase, 1, 4-β-cellobiohydrolase.</td>
<td>5, 6, 7, 9, 10, 48,</td>
</tr>
<tr>
<td>iii) Exoglucanases or 1,4-β-D-oligoglucan Cellobiohydrolases</td>
<td>EC 3.2.1.74</td>
<td>Removal of cellobiose from cellooligosaccharide or from p-nitrophenyl-β-D-cellobioside</td>
<td>Cellodextrinases</td>
<td>-</td>
</tr>
<tr>
<td>iv) β - Glucosidases or β-Dglucoside gluco-hydrolases</td>
<td>E.C.3.2.1.21</td>
<td>Hydrolysis of terminal non-reducing beta-D-glucose residues with release of beta-D-glucose.</td>
<td>Gentobiase, Cellobiase, Amygdalase.</td>
<td>1, 3, 9</td>
</tr>
<tr>
<td>v) Cellobiose: orthophosphate alfa-Dglucosyl transferase</td>
<td>E.C. 2.4.1.49</td>
<td>It catalyzes the reversible phosphorolytic cleavage of cellobiose</td>
<td>Cellobiose Phosphorylase</td>
<td>-</td>
</tr>
</tbody>
</table>
Enzymes within these classifications can be separated into individual components, such as microbial cellulase compositions may consist of one or more CBH components, one or more EG components and possibly β-glucosidases. The complete cellulase system comprising CBH, EG and BG components synergistically act to convert crystalline cellulose to glucose (Sukumaran et al. 2005). The exo-cellbiohydrolases and the endoglucanases act together to hydrolyze cellulose to small cello-oligosaccharides. The oligosaccharides (mainly cellobiose) are subsequently hydrolyzed to glucose by a major β-glucosidase.

Cellulolytic microbes are primarily carbohydrate degraders and are generally nable to use proteins or lipids as energy sources for growth. Cellulolytic microbes notably the bacteria *Cellulomonas* and *Cytophaga* and most fungi can utilize a variety of other carbohydrates in addition to cellulose, while the anaerobic cellulolytic species have a restricted carbohydrate range, limited to cellulose and or its hydrolytic products. The ability to secrete large amounts of extracellular protein is characteristic of certain fungi and such strains are most suited for production of higher levels of extracellular cellulases. One of the most extensively studied fungi is *Trichoderma reesei*, which converts native as well as derived cellulose to glucose. Most commonly studied
cellulolytic organisms include: Fungal species- *Trichoderma, Humicola, Penicillium, Aspergillus*; Bacteria- *Bacilli, Pseudomonads, Cellulomonas*; and Actinomycetes- *Streptomyces, Actinomucor,* and *Streptomyces.* While several fungi can metabolize cellulose as an energy source, only few strains are capable of secreting a complex of cellulase enzymes, which could have practical application in the enzymatic hydrolysis of cellulose. Besides *T. reesei,* other fungi like *Humicola, Penicillium* and *Aspergillus* have the ability to yield high levels of extracellular cellulases. Aerobic bacteria such as *Cellulomonas, Cellovibrio* and *Cytophaga* are capable of cellulose degradation in pure cultures. However, the microbes commercially exploited for cellulase preparations are mostly limited to *T. reesei, H. insolens, A. niger,* *Thermomonospora fusca, Bacillus sp,* and a few other organisms (Sukumaran et al. 2005).

The fungal pathogens play a major role in the development of diseases on many important field and horticulture crops; resulting in severe plant yield losses. Intensified use of fungicides has resulted in accumulation of toxic compounds potentially hazardous to humans and environment and also in the build up of resistance of the pathogens. In order to tackle these national and global problems, effective alternatives to chemical control are being employed (Anand and Raddy 2009). Biological control is a nature friendly approach that uses specific microorganisms, which interfere with plant pathogens and pests to overcome the problems caused by chemical methods of plant protection (Harman 2004). Commercial preparations of plant disease biocontrol agents are based on the practical application of rhizosphere competent species of bacteria or fungi. Although biological control occurs naturally, and is the principal reason diseases are not usually catastrophic, sufficient knowledge in many cases is not available to explain how biological control operates or how abiotic and biotic factors can be manipulated to affect the economic control of a pathogen. Fungi in the genus *Trichoderma* are among the most promising biocontrol agents against plant pathogenic fungi. Specific strains have the ability to control a range of pathogens under a variety of environmental conditions. Moreover, they may be rhizosphere competent which allows
them to colonize and protect plant roots. Among the action mechanisms proposed is mycoparasitism, with concomitant production of enzymes that degrade cell walls. Chitinolytic enzymes, together with β-glucanase or cellulases are the enzymes most frequently considered critical in biocontrol. In spite of enormous scientific research on biological control of plant pathogens with enzymes the most effective way against a wide range of pathogens is yet to be identified (Anand and Reddy 2009).

Various enzyme preparations consisting of different combinations of cellulases, hemicellulases, and pectinases have potential applications in agriculture for enhancing growth of crops and controlling plant diseases (Bhat 2000). Plant or fungal protoplasts produced using microbial hydrolases can be used to produce hybrid strains with desirable properties. Cellulases and related enzymes from certain fungi are capable of degrading the cell wall of plant pathogens in controlling the plant disease. Fungal β-glucanases are capable of controlling diseases by degrading cell walls of plant pathogens. Many cellulolytic fungi including *Trichoderma* sp., *Geocladium* sp., *Chaetomium* sp., and *Penicillium* sp. are known to play a key role in agriculture by facilitating enhanced seed germination, rapid plant growth and flowering, improved root system and increased crop yields (Bailey et al. 1998). Although these fungi have both direct (probably through growth-promoting diffusible factor) and indirect (by controlling the plant disease and pathogens) effects on plants, it is not yet clear how these fungi facilitate the improved plant performance. It has been reported that β-1, 3-glucanase and N-acetylglicosaminidase from *T. harzianum* strain P1 synergistically inhibited the spore germination and germ tube elongation of *B. cinerea*. Moreover, the exoglucanase promoters of *Trichoderma* are used for the expression of the different proteins, enzymes, and antibodies in large amount. The exoglucanase promoters of *Trichoderma* have been used for the expression of chymosin and other proteins: glucoamylase, lignin peroxidase, and laccase.
Cellulases have also been used for the improvement of the soil quality. Traditionally straw incorporation is considered an important strategy to improve soil quality and reduce dependence on mineral fertilizers. Many studies have attempted to hasten straw decomposition via microbial routes. Cellulolytic fungi applications such as Aspergillus, Chaetomium, and Trichoderma and actinomycetes have shown promising results. Fontaine et al. (2004) showed that exogenous cellulase supplementation accelerated decomposition of cellulose in soil. Therefore, using exogenous cellulase may be a potential means to accelerate straw decomposition and increase soil fertility (Kuhad et al. 2011).

Phytophthora root rot (PRR), caused by P. cinnamomi and other soilborne Phytophthora species, and impairs production in a wide variety of food and ornamental crops worldwide. In some cropping systems, organic mulches have proven to be a valuable tool in suppressing this disease. Suppression has been linked to biological activity within the mulch, and is correlated with total and cellulase enzyme activity. The production of cellulases by mulch microflora has been suggested as a primary mechanism of PRR suppression, but it has not been shown that the levels of cellulase produced in mulch are sufficient to inhibit Phytophthora growth or infectivity. Phytophthora root rot (PRR), caused by P. cinnamomi and other soilborne Phytophthora species, and impairs production in a wide variety of food and ornamental crops worldwide. In some cropping systems, organic mulches have proven to be a valuable tool in suppressing this disease. Suppression has been linked to biological activity within the mulch, and is correlated with total and cellulase enzyme activity. The production of cellulases by mulch microflora has been suggested as a primary mechanism of PRR suppression, but it has not been shown that the levels of cellulase produced in mulch are sufficient to inhibit Phytophthora growth or infectivity. To further investigate the role of cellulase enzyme in suppression of P. cinnamomi, a commercial formulation of cellulase was used to generate a standard curve which could be used to correlate cellulase activity levels in field samples with the enzyme unit concentrations commonly used in laboratory studies. Cellulase
produced by *Penicillium funiculorum* has been shown to reduce production of sporangia, zoospores, and chlamydospores by *P. cinnamomi* at concentrations of 10 units/ml or greater in soil extract (Richter 2009).

Cellulases are used in the textile industry, in detergents, pulp and paper industry, improving digestibility of animal feeds, in food industry, and the enzymes account for a significant share of the world enzyme market. The growing concerns about shortage of fossil fuels, the emission of greenhouse gases and air pollution by incomplete combustion of fossil fuel has also resulted in an increased focus on production of bioethanol from lignocellulosics and especially the possibility to use cellulases and hemicellulases to perform enzymatic hydrolysis of the lignocellulosic material. However, in production of bioethanol, the costs of the enzymes to be used for hydrolysis of the raw material need to be reduced and their efficiency increased in order to make the process economically feasible (Pradeep et al. 2012).

Growth of microorganism and its metabolite synthesis mainly depends on the medium nutrients and the growth conditions (Prescott et al. 2002). Optimization of medium components and process conditions plays a vital role for maximizing the microbial metabolites production and minimizing the production cost (Bezbaruah et al. 1994). The operating conditions of fermentation need to be optimized for maximum production in industrial technology. Such optimization by one factor at a time (OFAT) method involves changing one variable while fixing the others at a certain levels (Xu et al. 2003) which is laborious and time consuming, especially for large number of variables. However, statistical method offers several advantages over conventional method being rapid and reliable, short lists significant nutrients, helps understanding the interactions among the nutrients at various concentrations and reduces the total number of experiments tremendously resulting in saving time, glassware, chemicals and manpower. These limitations of a single factor optimization process can be overcome by using statistical
Statistical optimization of lytic enzyme optimization through fractional factorial design with considering all interactions effects of the factors. Initial screening of the ingredients is done to understand the significance of their effect on the product formation and then a few better ingredients are selected for further optimization. In recent years, the application of statistical experimental design was reported by many authors for optimization of various microbial metabolites (Tanyildizi et al. 2005).

Placket burman design (PB) of Response surface methodology (RSM) is successfully employed to predict the optimal level of the variables within the design space of the study (Box and Behnken 1960). RSM is a statistical technique for the modelling and optimization of multiple variables, which determine optimum process conditions by combining experimental designs with interpolation by first or second polynomial equations in a sequential testing procedure. RSM has already been successfully applied for the optimization of enzymatic hydrolysis of other bioprocesses. It is a useful tool which integrates mathematical and statistical approaches to analyze the effects of defined independent variables on the response without the need for prior knowledge of a predetermined relationship between the response function and the variables. RSM is now considered as a standard statistical approach for designing experiments, building models, evaluating the effects of many factors and finding the optimal conditions for desirable responses and reducing the number of required experiments. In biological processes, especially in the production of cellulase RSM has been adopted to optimize the growth of microorganisms and the production of cellulose (Shankar and Isaiarasu 2012).

In the present study, media components such as CMC, sucrose, yeast extract, K$_2$HPO$_4$, MgSO$_4$, peptone, pH, and temperature were evaluated for increasing cellulase production from our isolate B.amyloliquefaciens MBAA3 and statistically optimized using response surface methodology. Media components have been selected as variables after screening ingredients by Plackett-Burman design. The role of each variable, their interactions and statistical analysis for cellulase production were explained by applying the second-order
Statistical optimization of lytic enzyme

polynomial model. The analysis was done using Design-Expert version 8.0. Finally, a defined composition could be constructed which gives the optimum production of cellulose from MBAA3.

5.2A Materials and Methods

5.2.1A Screening of cellulase production

All the isolates (MBAA1, MBAA2, MBAA3 and MBAAT) were screened for cellulase production on minimal medium containing 1% carboxy methyl cellulose as substrate. Zone of cellulase production was observed around the colony by filling the plate with 1% congo red followed by 1% NaCl (Rathnan et al. 2012). The isolate showing the maximum production of was selected for further study.

5.2.1.1A Quantitative production of cellulase by MBAA3

The bacterial strain was grown in the minimal medium containing 1% CMC. 0.5 ml 1% CMC in acetate buffer (50 mM, pH 5.0) was used as a substrate to assay cellulase activity was incubated with 0.5 ml of enzyme supernatant at 30°C for 45 min. The product was measured by the dinitrosalicylic acid (DNS) method (Miller 1959). The amount of reducing sugars released was calculated from standard curves recorded for glucose and enzyme activity was expressed in U/ml.

5.2.2A Statistical optimization of cellulase production

5.2.2.1A Optimization identifying the significant variables using Plackett-Burman design

In the first step of the optimization process Plackett-Burman experimental design was used to identify the significance of the ingredients of the media for the optimum production of cellulase enzyme. (Plackett and Burman 1944). This experimental design is a 2 factorial design, was used to identify the critical parameters required for prominent cellulase production by screening n variables in n+ 1 experiment (Plackett and Burman, 1944). The variables chosen for the present study were CMC, sucrose, yeast extract,
K₂HPO₄, MgSO₄, peptone, pH, and temperature (Table 5.2A). The experimental design for the screening of the variables is presented in Table 5.3A. All the experiments were performed in triplicate and the average of cellulase activity was used as the response (dependant variable) (Singh and Shrivastava 2013). The main effect of each variable was calculated as the difference between the average of measurements made at the high value (+) and at the low value (-) (Table 5.2A). Each variable represented at two levels, high concentration (+) and low concentration (-) and three dummy variables in 12 trials.

The effects of individual parameters on cellulase production were calculated by the following equation:

\[ E(X_i) = \frac{2(M_+ - M_-)}{N} \]  

(1)

Where, \( E \) is the effect of parameter under study; \( M_+ \) and \( M_- \) are responses (Cellulase) of trials at which the parameter was at its higher and lower levels, respectively; \( N \) is the total number of trials.

Experimental error was estimated by calculating the variance among the dummy variables as

\[ V_{eff} = \frac{\Sigma (Ed)^2}{n} \]  

(2)

where, \( V_{eff} \) is the variance of the effect of level; \( Ed \) is the effect of level for the dummy variables; and \( n \) is the number of dummy variables used in the experiment. The standard error (SE, Es) of concentration effect was the square root of variance of an effect, and the significance level (p-value) of each concentration effect was determined using the student’s t-test:

\[ t(X_i) = \frac{E(X_i)}{Es} \]  

(3)

where, \( E(X_i) \) is the effect of variable \( X_i \).
Table 5.2A Medium components and their variables used in Placket-Burman design for cellulase production using MBAA3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Medium component</th>
<th>+Value</th>
<th>-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CMC</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>B</td>
<td>Sucrose</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>C</td>
<td>Yeast extract</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>D</td>
<td>Peptone</td>
<td>0.5</td>
<td>0.05</td>
</tr>
<tr>
<td>E</td>
<td>K$_2$HPO$_4$</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>F</td>
<td>MgSO$_4$</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>G</td>
<td>Temp.</td>
<td>37</td>
<td>28</td>
</tr>
<tr>
<td>H</td>
<td>pH</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>
Table 5.3A Plackett-Burman design generated by fractional rotation of full factorial design where A to H are independent variables and I to K are dummy variables

<table>
<thead>
<tr>
<th>Run</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>U/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>30.54</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>1.53</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>46.65</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4.02</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>37.41</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5.23</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>25.21</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>45.62</td>
</tr>
<tr>
<td>9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>1.23</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>1.65</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>14.78</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.25</td>
</tr>
</tbody>
</table>

Table 5.4A Experimental range and levels of the independent variables of selected components used for response surface central composite design

<table>
<thead>
<tr>
<th>Variable</th>
<th>Components</th>
<th>-α</th>
<th>-1</th>
<th>0</th>
<th>+1</th>
<th>+α</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CMC</td>
<td>0.159104</td>
<td>0.5</td>
<td>1.0</td>
<td>1.5</td>
<td>1.840896</td>
</tr>
<tr>
<td>B</td>
<td>MgSO₄</td>
<td>-0.1034</td>
<td>0.05</td>
<td>0.275</td>
<td>0.5</td>
<td>0.653403</td>
</tr>
<tr>
<td>C</td>
<td>pH</td>
<td>5.977311</td>
<td>7</td>
<td>8.5</td>
<td>10</td>
<td>11.02269</td>
</tr>
</tbody>
</table>
5.2.2.2A Response surface methodology (RSM)

The next step in formulation of the medium was to determine the optimum levels of significant variables for enhancing cellulase production (Prajapati et al. 2013). The levels of the significant parameters and the interaction effects between various medium constituents which influence the cellulase production significantly were analyzed and optimized by response surface central composite design (CCD). RSM is useful for small number of variables (up to 5) but is impractical for large number of variables, due to high number of experimental runs required (Prajapati et al. 2013). The concentrations of the 3 major components CMC, MgSO₄ and pH (identified by Plackett-Burman design) were optimized, keeping other variables constant. According to the design, the total number of treatment combinations is 2k+2k+no, where k is the number of independent variables and no is the number of repetition of experiments at the central point. Each factor in the design was studied at 5 different levels (−α, −1, 0, +1, +α) as shown in Table 5.4A. All variables were set at a central coded value of zero. The minimum and maximum ranges of variables were determined on the basis of our previous experiments. The full experimental plan with respect to their values in actual and coded form is listed in Table 5.6A. Enzymatic activity was measured in triplicate in 20 different experimental runs. The cellulase production was analyzed by using a second order polynomial equation, and the data were fitted into the equation by multiple regression procedure. The model equation for analysis is given as:

\[ Y = \beta_0 + \beta_i X_i + \beta_{ii} X_i^2 + \beta_{ij} X_i X_j \]  

(4)

where, \( \beta_0 \), \( \beta_i \), \( \beta_{ii} \), and \( \beta_{ij} \) represent the constant process effect in total, the linear, quadratic effect of \( X_i \), and the interaction effect between \( X_i \) and \( X_j \), respectively for the production of cellulase. Later, an experiment was run using the optimum values for variables given by response optimization to confirm the predicted value and cellulase production was confirmed.
5.2.2.3A Software and data analysis
The results of the experimental design were analyzed and interpreted using Design-Expert version 8.0 (Stat-Ease Inc., Minneapolis, MN, USA) statistical software.

5.3A Results and Discussion
5.3.1A Screening of cellulase production
Cellulase production of the all isolates was determined by minimal medium containing 1 % CMC as substrate. Zone of cellulase production was around the colony by filling the plate with 1 % congo red followed by 1 % NaCl (Picture 5.1A). Maximum zone of hydrolysis (63 mm) was observed by MBAA3. A significant zone of 36.5 mm was observed in MBAAT. MBAA1 (25 mm) and MBAA2 (33 mm) also hydrolyzed good amount of CMC.

5.3.1.1A Quantitative production cellulase enzyme
Quantitative cellulase production was estimated by 1 % CMC as a substrate in minimal medium. Maximum production of cellulase was observed in isolate MBAA3 (25.64 U/ml) after 72 hrs of incubation where as MBAA2 and MBAA1 showed 4.67 U/ml and 3.89 U/ml of cellulase production respectively. MBAAT also produce good amount of cellulase (7.28 U/ml) production after 5 days of incubation.
Picture 5.1A Zone of cellulase production by selected isolate *Bacillus amyloliquefaciens* MBAA3
5.3.2A Statistical optimization of cellulase production by MBAA3

5.3.2.1A Optimization identifying the significant variables using Plackett-Burman design

The chemical composition of the culture medium and environmental factors influence cell growth and cellulase production (Chawla et al. 2009). A better understanding of the medium components and environmental factors and their optimal control can, therefore, be used to improve the cellulase enzyme production. Plackett-Burman design was used to screen 8 different medium components as carbon and nitrogen sources as 12 run experiment with 2 level of concentration of each variable. The independent variables and their respective high and low concentrations used in the optimization study are represented in Table 5.2A, whereas the Plackett-Burman experimental design for 12 trials with 2 level of concentration of each variable is given in Table 5.3A, which was followed for the optimization of medium components for cellulase production. The variables A-H represented the medium constituents and I-K represented the dummy variables/unassigned variables. The results of Plackett-Burman experiment with respect to cellulase production, the effect, standard error, t (xi), p, and confidence level of each component are represented in Table 5.5A. The components were screened at a confidence level of 95% on the basis of their effects. When components show significance at or above 95% confidence level and its effect is negative, it is considered effective for production but the amount required may be lower than the indicated as low (−1) concentration in Plackett-Burman experiment. If the effect is found positive, a higher concentration than the indicated high value (+) concentration is required. In our experiment CMC, MgSO$_4$ and pH gave confidence level >95% and could be considered significant. Remaining components as Sucrose, yeast extract, peptone, K$_2$HPO$_4$, and temperature showed confidence level <95% and were considered insignificant in the study.
5.3.2.2A Response Surface Methodology

The central composite design was employed to study the interaction among the significant factors and also determine their optimal levels. In the present work, experiments were planned to obtain a quadratic model consisting of 23 trials. The plan includes 20 experiments and 2 levels of concentration for each factor. In order to study the combined effect of these variables, experiments were performed at different combinations. The central composite experimental plan along with the predicted and observed response for each individual experiment is summarized in Table 5.6A. It shows the production of cellulase (U/ml) corresponding to combined effect of all 3 components in the specified ranges. The production of cellulase may be best predicted by the following model:

\[
\text{Cellulase} = (27.32) + (1.28A) + (0.86B) - (2.27C) - (0.85AB) + (0.98AC) - (0.006BC) - (0.41A^2) - (1.94B^2) - (1.33C^2)
\]  

(5)

Where, Y is cellulase production (U/ml); A is CMC concentration (w/w); B is MgSO\textsubscript{4} (g); C is pH concentration.

The statistical significance of the second-order model equation was evaluated by F-test ANOVA which revealed that this regression is statistically highly significant for cellulase production. The model F-value of 3.79 implies that the model is significant. There is only a 2.47% chance that a large ‘model F-value’ could occur due to noise. Values of ‘prob>F’ less than 0.050 indicate that the model terms are significant. In this case A, AC are significant model terms (Table 5.7A). The ‘lack-of-fit F-value’ of 1.08 implies the lack of fit is not significant relative to the pure error. There is 46.72% chance that a large ‘lack-of-fit F-value’ could occur due to noise. Non-significant lack of fit is good for the model to fit. The R\textsuperscript{2} value (multiple correlation coefficient) closer to 1 denotes better correlation between observed and predicted values. The coefficient of variation (CV) indicates the degree of precision with which the experiments are compared. The lower reliability of the experiment is usually indicated by high value of CV. In the present case a low CV (9.51)
denotes that the experiments performed are reliable. Adequate precision measures the signal to noise ratio. A ratio greater than 4 is desirable. In our case, the ratio is of 6.553, which indicates an adequate signal. This model can be used to navigate the design space. The t-value and probability value (p-value) is a tool for evaluating the significance and contribution of each of the parameters to the statistical polynomial model equation. The pattern of interactions between the variables is indicated by these coefficients. The larger the magnitude of t test and the smaller the p-value are an indication of high significance of the corresponding coefficient (Karthikeyan et al 1996). Variables with very low probability levels (close to 0.00) contribute to the model, while others can be neglected and eliminated from the model. The p-value suggested that the coefficient for the linear effect of CMC, MgSO₄ and pH were most significant.

The effect of interaction of variables on enzyme (cellulase production) yield was studied against any 2 independent variables while keeping the other independent variables at their constant level. These response surface plots or contour plots can be used to predict the optimal values for different test variables. Therefore, 3 response surfaces were obtained by considering all the possible combinations. Three-dimensional response plot shown in Figure 5.1A describes behavior of cellulase production, main effect, interaction effect, and squared effect (nonlinear) of MgSO₄ and pH at different concentrations. The shape of the response surface curves showed a moderate interaction between these tested variables. As observed in the contour plot, the middle level of MgSO₄ and middle level of CMC resulted in higher enzyme production. The 3-dimensional curve and contour plot of the calculated response surface from the interaction between CMC and pH while keeping fixed concentration of MgSO₄ are shown in Figure 5.2A. Both components at their lower level did not result in the higher enzyme yield while CMC at the higher level and pH at their lower level showed the maximum enzyme activity. The interaction plot of MgSO₄ and pH is shown in Figure 5.3A, where the shape of the response surface indicates the of MgSO₄ and pH with the fixed coded value of CMC. An increasing MgSO₄ with
simultaneous decrease in pH led to an increase in cellulase. The enzyme yield was found to increase with higher level of MgSO₄ with lower level of pH.

RSM has been employed for the production of lactic acid from wheat bran by using *Lactobacillus amylophilus* GV6 (Naveena et al. 2004, Gawande and Patkar 1999). It has also been applied for the production of various enzymes, such as cyclodextrin glucanotransferase (CGTase) (Mahat et al. 2004), chitinase (Gohel et al. 2005), α-amylase (Rao and Satyanarayana 2003) and pectinase (Nair and Panda 1997), Lipase (Murthy et al. 2000) and vitamin riboflavin (Pujari and Chandra 2000). RSM can also be useful in optimizing the enzyme reaction conditions. A second order polynomial equation was found to be useful for the development of efficient bioprocess for cellulase production. Second-order polynomial models were used to correlate the abovementioned factors to soluble periplasmic IFN-α₂b formation and percentage of soluble IFN-α₂b translocated to the periplasmic space of *E. coli* and the models were found to be significant and subsequently validated (Azaman et al. 2010). Statistical optimization of fermentation conditions were reported to enhanced the cellulase production to 2995.2 ± 200.0 IU/ml, which was 9.91-fold higher than the activity under unoptimized basal medium (302.2 IU/ml) (Nagar et al. 2010). The biosynthesis of cellulases in *Trichoderma reesei* was very high in medium with carboxymethylcellulose as carbon source (Ahamed and Vermette 2008, Domingues et al. 2000). Li et al. (2008) reported maximum cellulase activity (0.26 U/mL) of a *Bacillus* sp. when the culture was grown in LB medium supplemented with 1% CMC29. It has been reported recently that *Bacillus* sp. (DUSELR13) and *Brevibacillus* sp. (DUSELG12) isolated from gold mine produced maximum CMCase activity 0.12U/mL and 0.02U/mL, respectively, under unoptimised conditions (Rastogi et al. 2010). Considering this property the enzyme isolated from this strain (which is alkaline in nature) can be used for various applications. However, very few *Bacillus* spp. are reported to be able to produce CMCase which has activity at highly alkaline pH (Murthy et al. 2000). Cellulase enzyme from *Bacillus*
subtilis (AS3) had an optimum pH of 9.2 so it was called alkaline cellulase (Deka et al. 2011).

El-Sersy et al. (2010) concluded that the optimum formulation of cellulase production medium as follows (g/l): KH₂PO₄, 1.5; MgSO₄, 0.1; MnSO₄, 0.05; NH₄NO₃, 0.5; NaCl, 1.5 with inoculum size of 0.5 ml. The media for cellulase production was optimized by Response surface methodology based on central composite design (CCD) by using four medium constituents, OPEFB fibers (5-15 g/L), yeast extract (3-9 g/L), CaCl₂ (1-5 mM), MgSO₄ (3-7mM) and to optimize the level of two environmental condition agitation speed (200-300 rpm) and temperature (28-32°C) (Nour et al. 2010). A second order model was proposed to assess the effect of these 6 variables. Based on the proposed model, the optimized conditions for the maximum cellulase production was 13.90 g/L OPEFB fiber, 8 g/L yeast extract, 3.5 mM CaCl₂, 7 mM of MgSO₄, 225 rpm of agitation and 29°C temperature. The production of endoglucanase was studied by Aspergillus terreus by applying the Plackett-Burman design for optimization of process parameters and this study agreed with our results in that, both KH₂SO₄ and MgSO₄ positively affected CMCase production. High levels of KH₂SO₄ and low levels of MgSO₄, maximize enzyme production (Youssef and Berekaa 2009).
Table 5.5A Statistical analysis of components for cellulase production by *Bacillus amyloliquefaciens* strain MBAA3

<table>
<thead>
<tr>
<th>Component</th>
<th>Effect</th>
<th>Standard Error</th>
<th>t-value</th>
<th>P Value</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC</td>
<td>17.49</td>
<td>3.963858</td>
<td>4.412368</td>
<td>0.0216</td>
<td>97.83997</td>
</tr>
<tr>
<td>Sucrose</td>
<td>8.99</td>
<td>5.04375</td>
<td>1.782404</td>
<td>0.1727</td>
<td>82.72997</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>2.778333</td>
<td>5.04375</td>
<td>0.550847</td>
<td>0.620106</td>
<td>37.98936</td>
</tr>
<tr>
<td>peptone</td>
<td>2.896667</td>
<td>5.04375</td>
<td>0.574308</td>
<td>0.605995</td>
<td>39.40054</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>0.643333</td>
<td>5.04375</td>
<td>0.127551</td>
<td>0.906574</td>
<td>9.342577</td>
</tr>
<tr>
<td>MgSO$_4$</td>
<td>18.47333</td>
<td>5.04375</td>
<td>3.662619</td>
<td>0.035181</td>
<td>96.48193</td>
</tr>
<tr>
<td>Temperature</td>
<td>9.23</td>
<td>5.04375</td>
<td>1.829988</td>
<td>0.164672</td>
<td>83.53282</td>
</tr>
<tr>
<td>pH</td>
<td>16.33333</td>
<td>5.04375</td>
<td>3.238331</td>
<td>0.047912</td>
<td>95.20884</td>
</tr>
</tbody>
</table>
Table 5.6A Full experimental central composite design with coded and actual level of variables and the response function

<table>
<thead>
<tr>
<th>Run</th>
<th>Actual A:CMC g%</th>
<th>Coded A:CMC g%</th>
<th>Actual B:MgSO₄</th>
<th>Coded B:MgSO₄</th>
<th>pH Actual</th>
<th>Coded pH</th>
<th>Cellulase Activity U/ml Actual</th>
<th>Predicted Cellulase Activity U/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>-1</td>
<td>0.05</td>
<td>-1</td>
<td>7</td>
<td>-1</td>
<td>23.6</td>
<td>23.88629</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>+1</td>
<td>0.05</td>
<td>-1</td>
<td>7</td>
<td>-1</td>
<td>24.55</td>
<td>26.20009</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>-1</td>
<td>0.5</td>
<td>+1</td>
<td>7</td>
<td>-1</td>
<td>25.05</td>
<td>27.348</td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>+1</td>
<td>0.5</td>
<td>+1</td>
<td>7</td>
<td>-1</td>
<td>24.85</td>
<td>26.23679</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>-1</td>
<td>0.05</td>
<td>-1</td>
<td>10</td>
<td>+1</td>
<td>16.85</td>
<td>17.38949</td>
</tr>
<tr>
<td>6</td>
<td>1.5</td>
<td>+1</td>
<td>0.05</td>
<td>-1</td>
<td>10</td>
<td>+1</td>
<td>24</td>
<td>23.62829</td>
</tr>
<tr>
<td>7</td>
<td>0.5</td>
<td>-1</td>
<td>0.5</td>
<td>+1</td>
<td>10</td>
<td>+1</td>
<td>20.55</td>
<td>20.8262</td>
</tr>
<tr>
<td>8</td>
<td>1.5</td>
<td>+1</td>
<td>0.5</td>
<td>+1</td>
<td>10</td>
<td>+1</td>
<td>22</td>
<td>23.63999</td>
</tr>
<tr>
<td>9</td>
<td>0.159104</td>
<td>-α</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>25.1</td>
<td>24.00702</td>
</tr>
<tr>
<td>10</td>
<td>1.840896</td>
<td>+α</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>29.95</td>
<td>28.3188</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>-0.1034</td>
<td>-α</td>
<td>8.5</td>
<td>0</td>
<td>20.7</td>
<td>20.37752</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0.653403</td>
<td>+α</td>
<td>8.5</td>
<td>0</td>
<td>25.7</td>
<td>23.2983</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>5.977311</td>
<td>-α</td>
<td>29.8</td>
<td>27.3863</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>11.02269</td>
<td>+α</td>
<td>20.05</td>
<td>19.73953</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>24.3</td>
<td>27.31957</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>29</td>
<td>27.31957</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>28.7</td>
<td>27.31957</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>24.75</td>
<td>27.31957</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>26.9</td>
<td>27.31957</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>0.275</td>
<td>0</td>
<td>8.5</td>
<td>0</td>
<td>29.8</td>
<td>27.31957</td>
</tr>
</tbody>
</table>
Table 5.7A Analysis of variance of quadratic model for cellulase production using MBAA3

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Df</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>189.9858</td>
<td>9</td>
<td>21.10953213</td>
<td>3.79448</td>
<td>0.0247</td>
</tr>
<tr>
<td>A-CMC</td>
<td>22.4418</td>
<td>1</td>
<td>22.44179898</td>
<td>4.033958</td>
<td>0.0724</td>
</tr>
<tr>
<td>B-MgSO₄</td>
<td>10.29776</td>
<td>1</td>
<td>10.29776171</td>
<td>1.851043</td>
<td>0.2035</td>
</tr>
<tr>
<td>C-pH</td>
<td>70.58331</td>
<td>1</td>
<td>70.58331317</td>
<td>12.68749</td>
<td>0.0052</td>
</tr>
<tr>
<td>AB</td>
<td>5.865313</td>
<td>1</td>
<td>5.8653125</td>
<td>1.054302</td>
<td>0.3287</td>
</tr>
<tr>
<td>AC</td>
<td>7.702812</td>
<td>1</td>
<td>7.7028125</td>
<td>1.384596</td>
<td>0.2666</td>
</tr>
<tr>
<td>BC</td>
<td>0.000312</td>
<td>1</td>
<td>0.0003125</td>
<td>5.62E-05</td>
<td>0.9942</td>
</tr>
<tr>
<td>A^2</td>
<td>2.410015</td>
<td>1</td>
<td>2.410015225</td>
<td>0.433205</td>
<td>0.5253</td>
</tr>
<tr>
<td>B^2</td>
<td>54.12976</td>
<td>1</td>
<td>54.1297625</td>
<td>9.729932</td>
<td>0.0109</td>
</tr>
<tr>
<td>C^2</td>
<td>25.42233</td>
<td>1</td>
<td>25.42232541</td>
<td>4.569713</td>
<td>0.0583</td>
</tr>
<tr>
<td>Residual</td>
<td>55.63221</td>
<td>10</td>
<td>5.56322108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Fit</td>
<td>28.89013</td>
<td>5</td>
<td>5.778025493</td>
<td>1.080324</td>
<td>0.4672</td>
</tr>
<tr>
<td>Pure Error</td>
<td>26.74208</td>
<td>5</td>
<td>5.348416667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>245.618</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.1A Response surface graph showing interaction effects between concentration of CMC and MgSO₄
Figure 5.2A Response surface graph showing interaction effects between concentration of pH and CMC
Figure 5.3A Response surface graph showing interaction effects between concentration of MgSO₄ and pH
5.3.2.3A Validation of the model

Validation was carried out under conditions predicted by the model. The optimal concentrations estimated for each variable were 1.84 g CMC, 0.275 g MgSO₄ and pH 8.5. The predicted cellulase production obtained from the model using the above optimum concentration of medium components was 29.95 U/ml. To validate the prediction of the model, additional experiments in triplicate were performed with the optimized medium. These experiments yielded the maximum cellulase activity of 30.62 U/ml. Agreements between the predicted and experimental results verified the validity of the model and the existence of the optimal points. Alam et al (2008) used a fractional factorial design with six factors to determine the optimal processing conditions for cellulase production by *Trichoderma harzianum* from domestic wastewater sludge; statistical analysis and surface response show the maximum production of filter paper hydrolysing enzymes to be 10,200 U L⁻¹ after three days of fermentation The data presented are the average of three replicates.

5.4A Conclusion

This study conclude the simultaneous screening of a number of nutrients influencing cellulase production. The enzyme yield and the production were found to be significantly influenced by CMC, MgSO₄ and pH whereas sucrose, yeast extract, K₂HPO₄, peptone, and temperature did not show significant effect on production of cellulase. The data obtained after optimization gives 29.95 U/ml enzyme production and earlier it was 25.64 U/ml. The statistical design of experiment offer efficient methodology to identify the significant variables and to optimize the factors with minimum number of experiments for cellulase production by microorganism. These significant factors identified by Plackett-Burman design were considered for the next stage in the medium optimization by using response surface optimization technique and can be further developed using a fermentor. The conversion of cellulosic biomass by microorganisms is a potential sustainable approach to develop novel bioprocesses and products. The biological aspects of processing of cellulosic biomass become the crux of future research involving
cellulases and cellulolytic microorganisms. Application of RSM was successfully used to define the optimal values of alkaline cellulase production by our novel rhizosphere isolate *B. amyloliquefaciens*. 
Chapter 5 (B) Molecular cloning and in silico analysis of cellulase gene from MBAA3

5.1B Introduction

Biological control is a nature friendly approach that uses specific microorganisms, which interfere with plant pathogens and pests to overcome the problems caused by chemical methods of plant protection. Enzymatic degradation of the cell wall of fungal pathogens by biocontrol agents has been reported. Cellulases and related enzymes from certain fungi are capable of degrading the cell wall of plant pathogens in controlling the plant disease (Kuhad et al. 2011). The present study demonstrates the cloning, expression and characterization of cellulase enzyme which was statistically optimized in chapter 5A from MBAA3. Biotechnological conversion of cellulosic biomass is potentially sustainable approach to develop novel bioprocesses and products. Microbial cellulases have become the focal biocatalysts due to their complex nature and wide spread industrial applications (Kuhad et al. 2011). Gene cloning is recently being employed for studying the structure and function of a number of enzymes and proteins and their over expression. This strategy has been found very efficient as compared to other traditional methods. Present study was designed to layout strategy for enhanced production of cellulases. *E. coli* expression system has several advantages: *E. coli* grows faster and in a cheaper and simpler medium. Also, as *E. coli* does not have endoglucanase, all endoglucanase activity is derived from the cloned gene. Many genes encoding cellulosomal components have been cloned, and their products have been characterized. Surprisingly, a 98-kDa protein, the presence of which, in relatively large amounts, in the cellulosome was described has been neither sequenced nor characterized (Kataeva et al. 1999).

Bioinformatics has revolutionized the field of molecular biology. The raw sequence information of proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools. Prediction of protein function is important application of bioinformatics (Prashant et al. 2010). Many researchers reported the
In silico characterization of gene clone

sequence analysis characterization of proteins using biocomputation tools (Sivakumar et al. 2007). In the present bioinformatics analysis for characterization of cellulases from \textit{B. amyloliquefaciens} was carried out. Protein sequences were subjected to ProtParam to analyze various physicochemical properties, secondary structure was predicted by SOPAMA software, the protein 3D model and its characteristics were predicted by (PS2): Protein structure prediction server. These parameters will assist the biochemist and physiologists in extraction, purification, separation and industrial applications of the enzyme. The aim of this study was to clone and express the genes of selected isolate \textit{B. amyloliquefaciens} MBAA3 encoding cellulases into \textit{E. coli} cells.

Isolate MBAA3 was evaluated for antifungal activity against two fungal pathogens that is \textit{Macrophomina phaseolina} and \textit{Sclerotinia sclerotiorum}. All fungal strains tested showed significant reduction in terms of radial diameter after the treatment with MBAA3, in comparison with the controls. Out of these fungal pathogens studied, MBAA3 was found to be efficient producer mycolytic enzyme cellulase. Furthermore the cellulase gene was cloned, characterized and expressed in \textit{E. coli} and then the CMCase activity of recombinant product was determined in statistically optimized medium described in chapter 5A.

5.2B Materials and Methods

5.2.1B Cellulase gene detection

5.2.1.1B Isolation of DNA for cellulase gene amplification

The isolate was grown aerobically overnight at 37°C in LB broth. Bacterial chromosomal DNA was purified using genomic DNA purification kit (Xcelgen).

5.2.1.2B Quantization and quality assessment of DNA

The DNA stock samples was quantified using nanodrop spectrophotometer at 260 and 280 nm using the convention that one absorbance unit at 260 nm wavelength equals 50 μg DNA per ml. Purity of DNA was judged on the basis of optical density ratio at
In silico characterization of gene clone

260:280 nm (Nurachman et al. 2010). The DNA having ratio between 1.8 to 2.0 was considered to be of good purity. Concentration of DNA was estimated using the formula:

\[
\text{Concentration of DNA (mg/ml)} = \text{OD 260} \times 50 \times \text{Dilution factor}
\]

Quality and purity of DNA were checked by agarose gel electrophoresis. Agarose 0.8% (w/v) in 0.5X TAE (pH 8.0) buffer (Sambrook and Russel 2001) was used for submarine gel electrophoresis. Ethidium bromide (1%) was added @ 10µl /100ml. The wells were charged with 5µl of DNA preparations mixed with 1µl gel loading dye. Electrophoresis was carried out at 80V for 30 min at room temperature. DNA was visualized under UV using UV transilluminator. The DNA was used further for amplification of cellulase gene.

5.2.1.3B Polymerase chain reaction

For the confirmation of cellulase gene the gene fragment was amplified by PCR from genomic DNA using gene specific set of primers: celF (5’-ATGAAACGGTCAATCTC-3’) and celR (5’-CTAATTTGGTTCTGTTCCC-3’) (Nurachman et al. 2010). PCR was carried out in a final reaction volume of 25 µl in 200 µl capacity thin wall PCR tube in Eppendorf Thermal Cycler. Reaction mixture for PCR tubes containing the mixture were tapped gently and spin briefly at 10,000 rpm. The PCR tubes with all the components were transferred to thermal cycler. The PCR protocol designed for 30 cycles for the primers used is given in Table 5.1B.

5.2.1.4B Visualization and purification of PCR product

To confirm the targeted PCR amplification, 5 µl of PCR product from each tube was mixed with 1 µl of 6X gel loading dye and electrophoresed on 1.2 % agarose gel containing ethidium bromide (1 per cent solution @10 µl/100 ml) at constant 5V/cm for 30 min in 0.5 X TAE buffer. The amplified product was visualized as a single compact band of expected size under UV light and documented by gel documentation system (Biorad). Length of the product is approx.1500bp Amplified PCR product was purified using Qiagen Mini elute Gel extraction kit according to the manufactures protocol.
<table>
<thead>
<tr>
<th>Steps</th>
<th>Temperature</th>
<th>Time</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>95°C</td>
<td>2 min</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>94°C</td>
<td>30 Sec</td>
<td>30</td>
</tr>
<tr>
<td>Denaturation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>58°C</td>
<td>30 Sec</td>
<td></td>
</tr>
<tr>
<td>Extention</td>
<td>72°C</td>
<td>90 Sec</td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>72°C</td>
<td>10 min</td>
<td>1</td>
</tr>
<tr>
<td>Extention</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2.1.5B Sequencing of purified gene segment

The concentration of the purified DNA was determined and was subjected to automated DNA sequencing on ABI 3730xl Genetic Analyzer (Applied Biosystems, USA). Sequencing was carried out using BigDye® Terminator v3.1 Cycle sequencing kit following manufacturer instructions. Cycle sequencing was performed following the instructions supplied along with BigDye® Terminator v3.1 Cycle Sequencing Kit. The reaction was carried out in a final reaction volume of 20µl using 200µl capacity thin wall PCR tube. The cycling protocol was designed for 25 cycles Denaturation at 96°C for 10 sec, Annealing at 58°C for 5 sec and extension at 60°C for 4 min with the thermal ramp rate of 1°C per second. After cycling, the extension products were purified and mixed well in 10 µl of Hi-Di formamide. The contents were mixed on shaker for 30 minutes at 300xg. Eluted PCR products were placed in a sample plate and covered with the septa. Sample plate was heated at 95°C for 5 min, snap chilled and loaded into autosampler of the instrument. Electrophoresis and data analysis was carried out on the ABI 3730xl Genetic Analyzer using appropriate Module, Basecaller, Dyeset/Primer and Matrix files.

5.2.1.6B Genomic DNA Isolation and nucleotide sequence accession number

DNA from B. amyloliquefaciens MBAA3 was isolated and amplified using CelF. 5’TGAAACGGTCAATCTC-3’ and CelR: 5’-CTAATTTGGTTCTGTTCCC-3’ primers and sequenced. The sequence was deposited in the gene bank under the accession number KF929416. The nucleotide sequence of protein was translated into protein sequence using Translate tool of the ExPASy server. Comparison to protein databases, check for frameshifts and sequencing errors was done by using BLAST tools. Protein family was determined by using SwissProt software.

5.2.2B Cloning of PCR product

The amplified product of cellulase gene was cloned in E.coli. The ligation of purified PCR product in plasmid vector was done with TOPO® TA Cloning® Kit for Sequencing (Cat no. K1214, MBI), which take advantage of the single ‘A’ overhang at each end of
PCR products generated using *Taq* and other non-proofreading DNA polymerase. Ligation was carried out in a final reaction volume of 6 µl using 200 µl capacity thin walls PCR tube (Table 5.2B). The mixture was then and centrifuged for 3-5 s and incubated at room temperature (22°C) for 5 hour. Ligation reaction was checked by PCR using M13 primers PCR was carried out in a final reaction volume of 25 µl in 200 µl capacity thin wall PCR tube. Composition of reaction mixture is given in Table 5.2B. PCR tubes containing the mixture were tapped gently and spun briefly at 10,000 rpm. The PCR tubes with all the components were transferred to thermal cycler. The PCR protocol designed for 30 cycles for the primers used is given in Table 5.4B. (Nurachman et al. 2010). 6 µL of the TOPOR cloning reaction from Perform the TOPO® cloning reaction was mixed with 18 µL of water. 2 µL of the diluted TOPOR Cloning reaction was added into a vial of one shotR electrocompetent *E. coli* and mix gently. The solution was carefully transferred into a 0.1-cm cuvette to avoid formation of bubbles and electroporated and immediately added to 250 µL of room temperature S.O.C. Medium. The solution was transferred into a 15-mL snap-cap tube (e.g. Falcon) and shaken for at least 1 hour at 37°C to allow expression of the antibiotic resistance genes. 10–50 µL from each transformation was spreaded onto a pre-warmed LB-ampicillin X-Gal/IPTG agar plates and incubated overnight at 37°C. 10 colonies were picked up for analysis of positive clones (Nurachman et al. 2010).

### 5.2.2.1B Screening of colonies using X-gal and IPTG (α- Complementation)

Forty µl of 2% X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside) solution and eight µl of 20% IPTG (isopropyl-D-thiogalactopyranoside) solution were pipetted on to a pre-made 90-mm LB agar plates containing ampicillin (50mg/ml). The solutions were spread with the help of a sterile bent glass rod. The agar surface was divided into sectors by drawing lines on the outer glass surface below the media. Each of the individual colonies obtained following the transformation procedure were picked up with sterile microtip and streaked in regular pattern on to sectors made on the X-gal/IPTG plate. The plates were incubated at 37°C for 12-15 hours. Growth on those sectors showing white
colonies was selected for further screening and confirmation of recombinant clones carrying correct insertion was done. Recombinant clones were further confirmed by colony PCR. The white colonies obtained on X-gal/IPTG plates were screened by colony PCR using gene specific and M13 primers. Briefly, a few cells from a white colony were suspended in a 50 µl of autoclaved distilled water in 200 µl thin walled PCR tubes. The tubes were then subjected to 95°C temperature for 15 minutes. After heat shock the tubes were centrifuged at maximum speed for 2 minutes. 3 µl of supernatant was used in a reaction mixture, prepared as per details given in Table 5.5B. PCR tubes containing the mixture were tapped gently, spun briefly and were transferred to thermal cycler. The amplified products were analyzed by agarose gel (1.2%) electrophoresis along with DNA molecular weight marker (Xcelgen). The clones that showed approximately 1400bp amplicon were considered positive clones carrying the inserted fragment generated by M13 primers. To confirm the targeted PCR amplification, 5 µl of PCR product was mixed with 1µl of 6X gel loading dye and electrophoreted along with DNA molecular weight marker (MBI Fermentas) on 1.2 % agarose gel containing ethidium bromide (one per cent solution at the rate of 10 µl/100 ml) at constant 5V/cm for 20 min in 0.5X TBE buffer (Aftab et al. 2012).

5.2.3B Characterization of clone by using bioinformatics tools

The program ProtParam, a component of the ExPASy server, estimates many basic physicochemical properties of a polypeptide on the basis of its sequence and these properties can be deduced from a protein sequence. The ProtParam includes the following computed parameters: Number of amino acids and their composition, Molecular weight (M.Wt), theoretical pI, Total number of negatively and positively charged residues, Atomic composition and their formula, Instability index, aliphatic index and grand average of hydropathicity (GRAVY). Predicting peptide produced by cutting a cellulase enzyme with specific proteases cutting of cellulase enzyme with two proteases such as trypsin and enterokinase was checked by peptide cutter link of the ExPASy server. Prediction of transmembrane domains in proteins was done by SOSUI
server. Motifs in protein were determined by the ScanProsite server PROSITE operates with two types of motifs Patterns – specific, relatively short sequence motifs. Profiles – takes into account the entire protein to predict the motifs. InterProScan server was used to find protein domains. Secondary structure was predicted by using SOPAMA software and 3D structure was predicted by using (PS2) Protein structure prediction server. A Phi/Psi Ramachandran plot was obtained to validate the backbone structure of cellulase. QMEAN Z-score was used to estimate the degree of nativeness of the predicted structure (Seniya et al. 2012).

5.3B Results and Discussion
5.3.1B Cellulase gene detection
5.3.1.1B Isolation and quantification of genomic DNA
The genomic DNA of MBAA3 was isolated by the method given by Yamada et al. (2002) and concentration was 0.498 mg/ml was confirmed by gel electrophoresis shown in Picture 5.1B.

5.3.1.2B PCR amplification of cellulase genes
Amplification of cellulase gene among several sets of primers tried during the study, the best amplification was observed by the primer set CelF This CelF: 5’-ATGAAAAACGGTCAATCTC-3’ and CelR 5’-CTAATTGGTTCTGTTCCCC-3’ set of primer was tried following the sequence of cellulase gene from \textit{B. amyloliquefaciens} PSM 3.1 (Nurachman et al. 2010). Among rest of sets of primers, sufficient amplification of cellulase gene could not be achieved as some set of primers showed non specific amplification of gene while a few others amplified DNA fragments of very minute base pair size (~300). The amplification of 1500 bp gene of cellulase by primer set CelF & CelR is shown in Picture 5.2B.

Cellulase gene detection screening showed that Cel genes was present in the MBAA3 and PCR products were sequenced and analyzed using the National center of
biotechnological information nBlast database (blast.ncbi.nlm.nih.gov). The PCR product showed very high homology to the nearest sequence of the respective gene. Moreover, a protein BLAST search indicated that this gene show similarity to protein sequence of cellulase. The nucleotide sequence of cellulase gene was submitted in the gene bank under the accession number KF929416.

The PCR products obtained using B. amyloliquefaciens PSM 3.1 chromosomal DNA as template were about 1500 bp and other non-specific bands were obtained (Nurachman et al. 2010). Amplification of cellulase gene was reported in which sequence specific primers for cellulase genes were used and significant amplification was achieved from the cDNA and the corresponding bands were found in the gel (Ahmed et al. 2005).

5.3.2B Cloning of cellulase gene

The cloning of purified PCR product in plasmid vector was done with TOPO® TA Cloning® Kit for Sequencing by using TA cloning vector pCR™4-TOPOR. Transformed solution was spread on LB-ampicillin X-Gal/IPTG agar selective plates (Picture 5.3B). Recombinant clones were confirmed by colony PCR. The amplified product was visualized as a single compact band of expected size (approx. 1300bp) under UV light and documented by gel documentation system (Bio-RAD) (Picture 5.4B). The amplified product was further confirmed sequencing the cellulase gene.
In silico characterization of gene clone

Picture 5.1B Genomic DNA of MBAA3

Picture 5.2B Amplification of cellulases gene from MBAA3. Lane I: amplified cellulase gene using CelF & CelR primers Lane II: DNA marker at annealing temperature of 58°C
### Table 5.2B Ligation reaction mixture

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresh PCR product</td>
<td>0.5-4 µl</td>
</tr>
<tr>
<td>Salt Solution</td>
<td>1 µl</td>
</tr>
<tr>
<td>Water</td>
<td>Add to total volume of 5 µl</td>
</tr>
<tr>
<td>TOPO Vector</td>
<td>1 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>6 µl</td>
</tr>
</tbody>
</table>

### Table 5.3B Composition of reaction mixture for PCR using M13 primers

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligated PCR product</td>
<td>3 µl</td>
</tr>
<tr>
<td>PCR Master Mix (2X)</td>
<td>12.5 µl</td>
</tr>
<tr>
<td>M13/pUC sequencing primer, 10 µM</td>
<td>1 µl</td>
</tr>
<tr>
<td>M13/pUC reverse sequencing primer, 10 µM</td>
<td>1 µl</td>
</tr>
<tr>
<td>Water, nuclease-free</td>
<td>7.5 µl</td>
</tr>
<tr>
<td>Total Volume</td>
<td>25 µl</td>
</tr>
</tbody>
</table>
Table 5.4B Steps and conditions of thermal cycling for M13 primers in PCR

<table>
<thead>
<tr>
<th>Steps</th>
<th>Temperature</th>
<th>Time</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Denaturation</td>
<td>95°C</td>
<td>2 min</td>
<td>1</td>
</tr>
<tr>
<td>Final Denaturation</td>
<td>94°C</td>
<td>30 Sec</td>
<td>30</td>
</tr>
<tr>
<td>Annealing</td>
<td>56.5°C</td>
<td>30 Sec</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>1:30 min</td>
<td></td>
</tr>
<tr>
<td>Final Extension</td>
<td>72°C</td>
<td>10 min</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.5B Composition of reaction mixture for colony PCR

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colony Supernatant</td>
<td>3 µl</td>
</tr>
<tr>
<td>PCR Master Mix (2X)</td>
<td>12.5 µl</td>
</tr>
<tr>
<td>M13/pUC sequencing primer, 10 µM</td>
<td>1 µl</td>
</tr>
<tr>
<td>M13/pUC reverse sequencing primer, 10 µM</td>
<td>1 µl</td>
</tr>
<tr>
<td>Water, nuclease-free</td>
<td>7.5 µl</td>
</tr>
<tr>
<td>Total Volume</td>
<td>25 µl</td>
</tr>
</tbody>
</table>
In silico characterization of gene clone

Picture 5.3B  Plates showing  blue and white clones after transformation of ligated products

Picture 5.4B Amplified gene from cloned white colony
Qualitative determination of the cellulase activity of the cloned gene was carried out using CMC agar plate and quantitative determination was carried out by using statistically optimized medium. The optimal concentrations estimated for each variable were 1.84 g CMC, 0.275 g MgSO₄ and pH 8.5. These experiments yielded the maximum cellulase activity of 48.51 U/ml which was higher than cellulase production by wild type MBAA3 (29.95 U/ml) (Figure 5.1B).

Cellulase enzyme was successfully cloned and expressed in a suitable host with better efficiency. The recombinant *E. coli strain* showed higher cellulase activity than the wild type. The amplification of complete DNA sequence of the gene Aucel12A from the genomic DNA of *A. usamii* E001 by using the conventional PCR and pUCm- T vector-mediated PCR and 1,576 bp long clone DNA sequence was expressed in a fully active form in *Pichia pastoris* (Shi et al. 2012). It was reported that the sequence which code for CMCase from the DNA of *Salmonella typhimurium* URI and the protein was overproduced in *E. coli* using T7 expression system, and its activity was confirmed by CMC-SDS-PAGE and revealed that the enzyme belongs to the cellulase family 8 and D (Yoo et al. 2004). celB and celC genes encode 2 cellulases were amplified (Ibrahim et al. 2013), using designed primers and cloned into the pTAC-MAT-2 expression vector and transformed into *E.coli*. The activities of the cloned cellulase genes were determined using agar diffusion assays. The SUMO (Small Ubiquitin-like Modifier)-bsBAFF was efficiently expressed in *E.coli* BL21 (DE3) and confirmed by SDS-PAGE and western blotting analysis. Similarly results were reported the amplification of endoglucanase gene from *B. pumilus* and cloned into *E. coli* DH5a using a pDrive vector (Telang et al. 2014). It was subsequently sub-cloned into *E.coli–Deinococcus* shuttle vector pRAD1 downstream of the *Deinococcus* heatshock promoter, groESL, and the construct was inserted into *D. radiodurans*. Functional endoglucanase enzyme was expressed in both *E.coli* and *D.radiodurans*. 
Figure 5.1B Cellulase activity of wild type and recombinant on statistically optimized media
The results also supported (Piriya et al. 2012), who reported the cloning and transformation of ethanol fermenting genes pyruvate decarboxylase (pdc) and alcohol dehydrogenase II (adh II) from *Zymomonas mobilis*. The recombinant *E. cloacae* strain produced twofold higher percentage of ethanol than the wild type.

**5.3.3B Structure of gene clone by using bioinformatics tools**

Translation of protein sequence was done into all possible 5’- 3’ reading frames (Figure 5.2B). Three forward and three reverse reading frame were observed. First forward reading frame was longest which is having desire length and remaining frames were very short. The sequence was compared to protein databases and there was no frameshift and sequencing error found. The sequence was familiar with Glycoside hydrolase family protein. The family of our cellulase enzyme sequence Glycosyl hydrolase family 5. It was found that the enzyme is not a multi domain but it is a single domain enzyme. The sequence has highest hit and good hit coverage and similarity with bacteria so it confirmed that it is a bacterial enzyme.

The computer analysis of sequence of endocellulase gene from a novel streptomycete and deduced that the Cel1 2A mature enzyme is a protein of 340 amino acids. The protein contained a catalytic domain, a glycine-rich linker region, and a cellulase-binding domain of 221, 12, and 107 amino acids, respectively (Solingen et al. 2001). New cellulase gene, *celK* has an open reading frame (ORF) of 2,685 nucleotides coding for a polypeptide of 895 amino acid residues with a calculated mass of 100,552 Da and was expressed in *E. coli* and purified (Kataeva et al. 1999).
In silico characterization of gene clone

EGVF.DNRVLTMGGLQASPASAAGTTPAASKNGQLSIKGTQLVNRDGKAVQLKGISSHGLQWYGDFVKDSLKLWRLDDWGITVFRAAIYTAYGGYIDNPSVKNKVEAVEAAKELGIYAIIDWHILNPNQNEKEAKEFKEKEMSSLYGNTPNIYIEIAN
EPNGDVNWKRDIKPYAEEVISVIRKNDPDDIIIVRTGTSQDVNDAAAYDQLKDA
NVMYALHFYGATHGLSVPDKANYASKGAPIFVTEWGTSDASNGGGVFLDQSR
EWLYLDCKNIIWVNWNSDKRESSSALKPGASKQRLAAYRSTASGTFVEKLSR
HKINERRS.NV

Figure 5.2B Translated protein sequence from nucleotide sequence

<table>
<thead>
<tr>
<th>Hits by patterns: [1 hit (by 1 pattern) on 1 sequence]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits by PS00659 GLYCOSYL_HYDROL_F5 Glycosyl hydrolases family 5 signature:</td>
</tr>
<tr>
<td>ruler:</td>
</tr>
</tbody>
</table>

EGVF
DNRVLTMGGLQASPASAAGTTPAASKNGQLSIKGTQLVNRDGKAVQLKGISSHGL

(2 aa)

154 - 163: [level tag: (0)] VIYEIANEPN

Figure 5.3B Motifs in cellulase enzyme
### Table 5.6B Amino acid composition (in%) in cellulase protein

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A) 31</td>
<td>9.4%</td>
</tr>
<tr>
<td>Arg (R) 14</td>
<td>4.2%</td>
</tr>
<tr>
<td>Asn (N) 26</td>
<td>7.9%</td>
</tr>
<tr>
<td>Asp (D) 23</td>
<td>6.9%</td>
</tr>
<tr>
<td>Cys (C) 1</td>
<td>0.3%</td>
</tr>
<tr>
<td>Gln (Q) 10</td>
<td>3.0%</td>
</tr>
<tr>
<td>Glu (E) 17</td>
<td>5.1%</td>
</tr>
<tr>
<td>Gly (G) 27</td>
<td>8.2%</td>
</tr>
<tr>
<td>His (H) 5</td>
<td>1.5%</td>
</tr>
<tr>
<td>Ile (I) 22</td>
<td>6.6%</td>
</tr>
<tr>
<td>Leu (L) 21</td>
<td>6.3%</td>
</tr>
<tr>
<td>Lys (K) 26</td>
<td>7.9%</td>
</tr>
<tr>
<td>Met (M) 3</td>
<td>0.9%</td>
</tr>
<tr>
<td>Phe (F) 9</td>
<td>2.7%</td>
</tr>
<tr>
<td>Pro (P) 11</td>
<td>3.3%</td>
</tr>
<tr>
<td>Ser (S) 25</td>
<td>7.6%</td>
</tr>
<tr>
<td>Thr (T) 14</td>
<td>4.2%</td>
</tr>
<tr>
<td>Trp (W) 10</td>
<td>3.0%</td>
</tr>
<tr>
<td>Tyr (Y) 14</td>
<td>4.2%</td>
</tr>
<tr>
<td>Val (V) 22</td>
<td>6.6%</td>
</tr>
<tr>
<td>Pyl (O) 0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Sec (U) 0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(B) 0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(Z) 0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(X) 0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
Amino acid composition of cellulase protein was analyzed which contain 9.4% of alanine was found and 0.3% of cystein was observed (Table 5.6B). The average molecular weight of cellulase calculated is 36732.1 Da. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of protein is zero. At pI proteins are sable and compact. The computed pI value of cellulase 7.21 reveals that the protein is basic in character. The computed isoelectric point (pI) will be useful for developing buffer system for purification by isoelectric focusing method. Extinction coefficient of cellulase at 280 nm is 75860 M⁻¹ cm⁻¹ high extinction coefficient is only due to the high concentration of Tyr and Trp and not due to the Cys because Cys is very low in concentration in cellulase. The computed protein concentration and extinction coefficients help in the quantitative study of protein-protein and protein-ligand interactions in solution. The biocomputed half-life is 1 hours. The average hydrophobicity of protein is -0.517825 indicates protein is soluble in water. Stability index is 33.62 which is <40 indicates that protein is stable. The aliphatic index (AI) which is 79.31 is defined as the relative volume of a protein occupied by aliphatic side chain (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. So the protein is stable at wide range of temperature. Grand average hydropathy (GRAVY) index of cellulase indicates that the protein has better interaction with water (Table 5.7B).

Very high aliphatic index of all cellulases infers that they may be stable for a wide range of temperature. The very low GRAVY index of cellulases ADK13057.1, EGH91691.1 indicates that these cellulases could result in a better interaction with water (Pradeep et al. 2012). The presence of 0.3 % Cys residues in cellulase indicates the absence of disulphide bridges ("SS" bonds 14 tyrosine hydroxylase protein sequences of teleost fishes were retrieved from Uniprot Database were analyzed by In silico tools (khanam et al. 2012). Similar trend was generated for protein feature based identification of cell cycle regulated proteins in Yeast (Lichtenberg et al. 2003).
Table 5.7B Parameters computed using Expasy's ProtParam tool

<table>
<thead>
<tr>
<th>Num of amino acids</th>
<th>M. wt. (R)</th>
<th>pI (+R)</th>
<th>Formula</th>
<th>Total num of atoms</th>
<th>GRA (EC)</th>
<th>SI (SI)</th>
<th>AI (AI)</th>
<th>VY (VY)</th>
<th>Hal life</th>
<th>Hydrophobicity</th>
<th>Atomic composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>331</td>
<td>3673</td>
<td>7.0</td>
<td>(Asp + Glu)</td>
<td>C_{1637}H_{2533}N_{455}O_{501}S_{4}</td>
<td>5130</td>
<td>758</td>
<td>33.79</td>
<td>-1.0</td>
<td>-0.517825</td>
<td>16 25 45 50</td>
<td>C  H  N  O  S</td>
</tr>
<tr>
<td>2.1</td>
<td>21</td>
<td></td>
<td>(Arg + Lys)</td>
<td></td>
<td>60</td>
<td>62</td>
<td>31</td>
<td>0.518</td>
<td>hou</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glu</td>
<td>37</td>
<td>33</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>rs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ly</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.8B Cutting a cellulase enzyme with specific proteases

<table>
<thead>
<tr>
<th>Name of enzyme</th>
<th>No. of cleavages</th>
<th>Positions of cleavage sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>38</td>
<td>7 25 30 37 44 47 52 68 72 75 84 101 103 105 112 136 138 142 169 170 184 185 195 212 234 241 267 276 288 289 301 303 308 318 321 323 327 328</td>
</tr>
</tbody>
</table>

Figure 5.4B  Secondary structure of cellulase enzyme
Cutting of cellulase enzyme with two proteases such as trypsin and enterokinase was checked which is shown in Table 5.8B. Enterokinase cut the cellulase enzyme while 38 number cleavage was observed by trypsin.

In protscale component low score indicates the absence of transmembrane region. Many specific biological functions are associated with relatively short amino acid sequence motifs in proteins. Finding such motifs tells us a lot about the protein. For example, posttranslational modifications are often associated with specific sequence motifs. One can see the motifs present and the location of motif in our cellulase enzyme (Figure 5.3B).

Secondary structure prediction was done by SOPAMA software. The main secondary structure elements of the proteins are: α-helices, β-sheets (or β-strands), random coils predicting the secondary structure elements of the protein is the first step towards understanding its structure and eventually, functions (Figure 5.4B). Secondary structure possess Alpha helix (Hh) : 114 is 34.44%, 310 helix (Gg) : 0 is 0.00%, Pi helix (Ii) : 0 is 0.00%, Beta bridge (Bb) : 0 is 0.00%, Extended strand (Ee) : 59 is 17.82%, Beta turn (Tt) : 0 is 0.00%, Bend region (Ss) : 0 is 0.00%, Random coil (Cc) : 158 is 47.73%, Ambiguous states (?) : 0 is 0.00% and Other states : 0 is 0.00% at Parameters : Window width : 17 Similarity threshold : 8 Number of states : 3. The results of Ramachandran plot suggested that 91.8%, 7.6%, 0.3%, and 0.3% of the residues in model were in the most favored, additional allowed, generously allowed and disallowed regions, respectively. Thus, overall 99.7% of the residues were placed into the favored and allowed regions which indicates high quality model structure in terms of protein folding (Figure 5.5B).

The information about 3D structure of a molecule consists of coordinates of each (or most) atoms which defines a precise position of an atom in an imaginary space. Homology models were predicted using SWISS- MODEL (Figure 5.6B). The model showed least RMSD value compared to other models. QMEAN scoring function is independent of the size of the protein and may therefore be used to assess both monomers and entire oligomeric assemblies. The resulting QMEAN Z-score -1.19 provides an
estimate of the ‘degree of nativeness’ of the structural features observed in a model and indicates that model is of comparable quality to experimental structures. QMEAN scores of the reference set and QMEAN model accuracy estimation parameters (Figure 5.7B a and b).

The secondary structure indicates whether a given amino acid lies in a helix, strand or coil (Choubey et al. 2010). Similarly SOPMA was used for secondary structure prediction of thioredoxin protein in wheat (Prabhavathi et al. 2011). A two-dimensional plot of the values of the backbone torsion angles phi and psi, with allowed regions indicated for conformations where there is no steric interference. Ramachandran plots are used as a diagnosis for accurate structures: when the phi and psi torsion angles of an experimentally determined protein structure are plotted on such a diagram, the observed values should fall predominantly in the allowed regions (Seniya et al. 2012). However the percentage of residues occupying the most favored regions is 84.5 while 13.9% and 1.9% resides in additional allowed and generously allowed regions respectively (Arora et al. 2010). A model structure for the Manduca chitinase was predicted computationally from the amino acid sequence deduced from its cDNA clone by using SWISS-MODEL software and catalytic domain with two highly conserved regions found in all family 18 chitinase (Huang et al. 2000). Three-dimensional (3D) structure of the bsBAFF monomer, analyzed by comparative protein modeling revealed that it was very similar to its counterparts (Yoo et al. 2004). The QMEAN Z-score provides an estimate of the absolute quality of a model by comparing it to same sized reference structures. The QMEAN6 server standing for qualitative model energy analysis along with its clustering method QMEANclust provides a composite scoring function to measure both global and local model accuracy on the basis of single models which can assist in model selection (Seniya et al. 2012).
Figure 5.5B  Ramachandran plot of the cellulase model. The most favored regions are colored black, additional allowed, generously allowed and disallowed regions are indicated as light black and white fields, respectively

Figure 5.6B  Homology Model of MBAA3 cellulase predicted using SWISS MODEL. Helices are presented in green and sheets in blue
Figure 5.7B  (a) QMEAN scores of the reference set. (b) QMEAN model accuracy estimation parameters
In silico approaches assisting protein design and engineering are being developed back to back with experimental techniques. The bioinformatics approaches are most successfully used for engineering protein stability. Computational tools are useful for characterization of the relationships among protein sequence, structure, and function spaces.

5.4B Conclusion
In this study efforts were made to clone and express gene encoding cellulase enzyme in a suitable host with better efficiency and to study about some functional characteristics of mycolytic cellulase enzyme by using computational tools. In silico approaches assisting protein design and engineering are being developed back to back with experimental techniques. The bioinformatics approaches are most successfully used for engineering protein stability. Physico chemical characterization study give better idea about the structural properties of protein such as pl, EC, AI and GRAVY. Secondary structure give the idea about α and β strands of protein. Ramchandran plot revealed the quality of model structure RQMEAN Z-score provides an estimate of the degree of nativeness. Computational tools are useful for characterization of the relationships among protein sequence, structure, and function spaces. The recombinant E. coli strain showed higher cellulase activity (48.51U/ml) in statistically optimizes media than the wild type (29.95 U/ml). This study provides insight about the properties of the enzyme cellulase and thus aid in formulating their use in industries.