MATÍS OHF

Örorka til etanólframleiðslu

Áfangaskýrsla til Orkusjóðs

Styrktímabil: 2010-2011

Guðmundur Óli Hreggviðsson

3/1/2011

Efnisyfirlit

Inngangur
Verkefnið „Örorka til etanólframleiðslu“ er fólgið í að þróa með erfðatækni hitakæra örveru, Thermoanaerobacterium isalndcium, sem gæti brotið niður sellulósa og framleitt etanól sem orkugjafa. Hitastöðugir sellulasar hafa verið auðkenndir í annarri örveru tegund og er markmiðið að flytja gen þeirra inn í etanólframleiðslustofninn. Matís hefur stundað rannsóknir á hitakærum ensímum og bakteríum, sem framleiða etanól, um árabil. Þessar rannsóknir hafa verið stundaðar af miklum metnaði og er Matís í víðtæku samstarfi um þessar rannsóknir bæði innan- og utanlands.

Staða og samhengi rannsóknar

Eins fram kom í umsókn er verkefnið hluti af stærri verkefnapakka sem öll hafa að markmiði að vinna eldsneyti/ eða önnur efnasambönd úr lífmassa. Var sótt um styrk til bæði erlendra og innlendra rannsóknarsjóða með góðum árangri. Styrkurinn frá Orkusjóði er mikilvægt mótframlag í þessi verkefni en þess er sérstaklega krafist, og er það talið fram sem slíkt í reikningshaldi erlendu verkefnanna.

2011-2014. Árið 2010 fékkst mikilvægur öndvegisstyrkur frá Norræna ráðherraráðinu (TFI, Topforskningsinitivatetet) til þessa verkefnis sem veittur var til fjögurra ára. Þátttakendur eru stofnanir og háskólar frá öllum norðurlöndunum.Verkefnisheiti: *Sustainable Biofuel: Innovations in Bioethanol Production Technologies*

2010-2012. Öndvegisstyrkur frá Tækniþróunarsjóði. Samstarf við HA, Sorpu, Mannvit, Landbúnaðarháskólann á Hvanneyri og Nýsköpunarmiðstöðina. Verkefnisheiti: *Lífeldsneyti.*

2010-2011. Samvinnuverkefni Matís við Háskólann í Lundi á framleiðslu á etanóli með hitakærum bakteríum sem styrkt var af þarlendum sjóði FORMAS (Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning). Verkefnisheiti: *Metabol utveckling av termofiler för effektivisering av såväl etanolproduktion som utnyttjande av biomassa.*

Þess má einnig geta að Matís hefur sótt um Evrópusambandsstyrk á þessu sviði (fyrir tímabilið 2012-2015) og er það til frekari rannsókna á möguleikum hitakærra baktería til eldsneytis- og efnaframleiðslu. Öll þessi verkefni byggja á mjög sérstökum rannsóknarefniviði, hitakærum lífverum sem finna má í jarðhitaumhverfi á Íslandi en óvíða annars staðar, sem og öflugri rannsóknaruppbyggingu og samfelldu starfi á þessu sviði hjá rannsóknarhópi Matís, allt frá árinu 1985 (hófst á Iðntæknistofnun Íslands)´

Starfsmenn í verkefninu:

Ásamt starfsmönnum Matís þá hefur meistaraneminn Emanuel Ron frá Háskólanum í Lundi unnið í verkefninu í fimm mánuði frá september 2010.

Sótt var um styrkt til að fjármagna doktorsnema í þetta verkefni. Það akkur fyrir Matís að þjálfa nýtt fólk og tryggja þar með framhald og samfellu á þessu mikilvæga rannsóknarsviði. Því hefur verið ráðinn að verkefninu ungur nýdoktor Bryndís Björnsdóttir. Hún lýkur doktorsprófi í maí á þessu ári

(2011). Bryndís hefur störf að loknu barnsburðarleyfi þann 1. júní . Hún er mjög hæfur starfsmaður og fengur að fá hana í þetta verkefni .

Þá hefur verið ráðin í verkefnið Beata Wawiernia MSc. sem unnið hefur að sambærilegum rannsóknum hjá Próf. Jóhanni Örlygssyni við Háskóla Akureyrar. Hún hefur lýst yfir áhuga á því að fara doktorsnám og mund það skírast í lok þessa árs. Hún mun einnig hefja störf 1. júní 2012.

Hvorir tveggja þessara starfsmanna. Bryndís og Beata eru mjög öflugir og lítum við til þeirra sem framtíðarstarfsmanna hjá Matís.

Tíma- og verkáætlun:.

Styrkur var veittur í júní og var áætlað í umsókn að verkefnið hæfist 1. Spetember. Verkefnið hófst einum mánuði síðar 1. október 2010 –heldur seinna en ráð var fyrir gert. Það skýrist af þeirri áherslu sem lögð var á að nýta þennan styrk frá Orkusjóði í að veita ungum vísindamönnum tækifæri til að koma að þessum rannsóknum (helst í doktorsnám). Það var á hinn bóginn erfiðleikum bundið að fá nema/ungan vísindamann til verksins með stuttum fyrirvara.

Verkefnið er á áætlun miðað við að upphaf sé 1. október.

1. *Verkþáttur:* **Endurbætur á gagnagrunni yfir erfðamengi *T. islandicus***

   Upphaf: Mánuður 1
   Áætluð lok: Mánuður 2.
   ***Gangur:* Á áætlun (og umfram)**
   Búið er að raðgreina erfðamengi *T. islandicus* að fullu og skilgreina efnaskiptaferla, en auk þess var erfðamengi annarrar hitakærra tegundar *Calormator* spp. raðgreint og skilgreint. Þessum verkþætti er því lokið og umfram það (Fylgiskjal 1). Lokahnykkurinn var uppsetning erfðamengjanna í sérstöku gagnagrunnsforriti. Keypt var sérstakt forrit Geneious™ til þessa sem auðveldar yfirsýn, umsjón og túlkun gagna. Í þessum verkþætti vann Emanuel Ron meistaranemi við Háskólann í Lundi ásamt föstum starfsmönnum Matís.
   Þetta er mjög umfangsmikill og dýr verkþáttur. Aðrir styrkveitendur til þessa verkþáttar: NordicEnergy (2007-2010) FORMAS (2010). og Tækniþróunarsjóður (2006-2010 (tveir styrkir)).

2. *Verkþáttur:* **Frekari breytingar á efnaskiptaferli gerjunar í *T. islandicus* með erfðatækni**

   Upphaf: Mánuður 2
   Áætluð lok: Mánuður 36.
   ***Gangur:* Á áætlun**
   Þessi vinna er samfelld , viðvarandi og sameiginleg öllum verkefnum. Tekist hefur að slá út efnaskiptarásir sem leiða til myndunar tveggja aukaefna annars vegar mjólkursýru og hins vegar ediksýru, en í sitt hvorum stofninum. Næst skref er að koma þessum stökkbreytingum fyrir í einum og sama stofni. Vandamálið við það er að valgen eru af skornum skammti til að koma þessu í kring. Úrval valgena er forsenda áframhaldandi árangursríkrar erfðatæknivinnu á þessum stofnum, þ.e.s. ef gera á fleiri en eina breytingu í sama stofni. Því var farið í ítarlega undirbúningsúttekt á möguleikum í þessum efnum (Fylgiskjal 2).

   Áætlað er að Beate Wawiernia vinni að mestu í þessum verkþætti frá 1. júní. Fram að því verður vinna í þessum verkþætti eingöngu í höndum fastra starfsmanna Matís.

3. *Verkþáttur:* **Greining og val á sellulasagenum *Caldicellulosiruptor kristjanssoni***

   Upphaf: Mánuður 3
   Áætluð lok: Mánuður 6.
   ***Gangur:* Á áætlun (en viðbótarvinnu er þörf)**
   Þessi vinna hófst í verkefni sem styrkt var af NordicEnergy sjóðnum (2007-2010) *„New, innovative pretreatment of Nordic wood for cost-effective fuel-ethanol production“* Fjöldi sellulasa og hemi-sellulasa gena var þar skilgreindur og voru nokkur þeirra klónuð og eiginleikar ensímanna kannaðir (Fylgiskjal 3). Verkin var framhaldið í þessu verkefni „Örorka til etanólframleiðslu“ og var í höndum meistaranemans Emanuels Ron, þar sem við bættist eitt gen til viðbótar. Í því tilviki tókst að klóna ensímið en tjáningu var ábótavant þar eð

virkni fékkst ekki (Þessari vinnu verður lýst í meistararitgerð hans sem er í smíðum). Vinnunni verður haldið áfarm en ljóst er að meira úrval sellulasa gena er þörf. Þessi verkþáttur verðu því efldur að muna og verður vinnan fjármögnuð af öndvegisstyrknum frá Norræna ráðherraráðinu sem fékkst fyrir tímabilið 2011-2014. Verður leitað í öðrum erfðamengjum hitakærra baktería, þar ámeðal í *Caloromator* tegundinni sem var raðgreind í þessu verkefni. Í verkþættinum vann Emanuel Ron meistaranemi við Háskólann í Lundi ásamt föstum starfsmönnum Matís. Áætlað er að Bryndís Björnsdóttir vinni í þessum verkþætti frá 1. júní 2011ásamt föstum starfsmönnum Matís.

4. *Verkþáttur.* **Smíði á innsetningarkassettum sem innihalda Caldicellulosiruptor kristjanssoni sellulasa-gen með seytikerfi og undir stjórn T. islandicus stýriraða**

Upphaf: Mánuður 7
Áætluð lok: Mánuður 13.
Gangur: **Mun seinka um tvo mánuði.**

Þessi verkþáttur er ekki hafinn og ætti samkvæmt verkáætlun að hefjast í 7. mánuði verkefnisins eða 1. apríl. Áætlað er að nýdoktor Bryndís Björnsdóttir í þessum verkþætti, en hún mun hefja störf 1. júní að lokinni doktorsútskrift og að lokun barnsburðarleyfi. Þessi verkþáttur byggir á verkþáttum, 1, 2 og 3. og forsenda árangurs er góður skilningur á efnaskiptabrautum (Verkþáttur 1), góður árangur í þróun valgena (Verkþáttur 2) og úrval sellulasa-gena, (Verkþáttur 3). Unnið verður markvisst í þessum undanfarandi verkþáttum fram að því að verkþáttur 4 getur hafist.

Verkþættir 5-7.
Vinna við þessa verkþætti er á síðara ári styrktímabils (sjá umsókn)

Fylgiskjal 1.

Genome sequencing

The genome sequencing had different purposes:

1. To identify genes encoding for cellulases and hemicellulases that could be used in pre-processing of lignocellulose material (see.
    a. *Caldicellulosiruptor kristjanssonii,* grows well on cellulose as well as hemicelluloses, produces little ethanol but large amounts of acetate
    b. *Thermoanaerobacterium islandicum*., does not grow on cellulose, but is efficient producer of ethanol on glucose and xylose

2. To map metabolic pathways of organisms selected for the metabolic engineering for more efficient production of ethanol
    a. Thermoanaerobacterium islandicum

3. To identify cellulase genes of different acitivites (in *Caldicellulosiruptor kristjanssonii*) that can cloned into ethanol producing thermophile lacking cellulose degrading ability such as *Thermoanaerobacterium islandicum*.
    a. *Caldicellulosiruptor kristjanssonii*
    b. *Caloromator spp*. good producer og ethanol growing on hemicelluloses and degradation products of cellulose

In an initial screening for good lignocellulose degraders, a strain belonging to the species *Caldicellulosiruptor kristjanssonii* was identified (data not shown) as a good degrader of cellulose. The genome was partially sequenced (6 fold coverage) and a number of cellulase and hemicellulase genes were identified (section 3. Sequenced base genome screening). This starin was sequenced very early in the project and thus beyond this project supported by Orkusjóður..

Thermoanaerobacterium islandicum is a good producer of ethanol and was identified in a previous project in cooperation with the University of Akureyri. The species was capable of degrading hemicelluloses, but not cellulose. This species was selected or genetic engineering for more efficient production of ethanol. The genome was sequenced (20 fold coverage) and critical genes in pathways leading to ethanol production and the formation of the side products lactate and acetate were identified. Hemicellulase genes and one putative cellulase gene were also identified (section 3. Sequenced base genome screening ans section 6. A putative endocellulase from the thermophilic anaerobe *Thermoanaerobacterium islandicum*: CeluB).

Caloraomator spp. was identified in a later screening and no close relative has previously been sequenced. It is a good fermenting species and alcohol producer.

Genome sequencing of *Thermoanaoerobacterium islandicum*

Varða 15/12 2010

Genomes of *Thermoanaerobacterium islandicum* had been partially sequenced with 5.5 fold overage by GSM in Germany. In November 2010 it was resequenced with an additional coverage of 26. Total coverage amounts therefore to 31.5. The organism is an efficient producer of ethanol and grows on the pentoses, xylose and arabinose. It could also grow on cellobiose, galactose, and glucose, xylan and pectin. It does not grow on cellulose. Genes have been annotated and key metabolic genes identified. Metabolic pathways will next be deduced and pathways to critical undesirable by-products defined. Data has been put into Geneious softwaref ro handlin gand interpretation. Beleow is andditonal sequencing data:

Coverage = 26fold

Coverage in previous FLX sequencing = 5,5 fold

```
** Consensus results.
*/
consensusResults
        numAlignedReads    = 204396, 97.41%;
        numAlignedBases    = 62569026, 98.17%;
        inferredReadError = 1.16%, 724763;

        numberAssembled = 197227;
        numberPartial   = 7169;
        numberSingleton = 3673;
        numberRepeat    = 156;
        numberOutlier   = 368;
        numberTooShort  = 1229;

    largeContigMetrics
        numberOfContigs   = 115;
        numberOfBases     = 2365808;

        avgContigSize     = 20572;
        N50ContigSize     = 57538;
        largestContigSize = 205019;

        Q40PlusBases      = 2359049, 99.71%;
        Q39MinusBases     = 6759, 0.29%;

    allContigMetrics
        numberOfContigs = 191;
        numberOfBases   = 2383686;
```

Genome sequencing of *Calormator* spp.

Varða 30.01.2011

Strain 2989 was selected for genome sequencing. Genomic DNA was isolated and 16S rRNA analysis was performed. The analysis revealed the strain to be closely related to *Caloramator australicus.* The DNA was sheared by nebulization and fragments of the correct size were isolated from an agarose gel. A genome library was constructed by ligating adaptors to the fragments and single stranded DNA molecules were subsequently isolated. The library was quantified, its quality assessed using the Agilent 2100 Bioanalyzer and it was titrated prior to full scale sequencing. Emulsion PCR and sequencing were performed with the GS FLX Titanium system according to instructions from ROCHE. Results from the sequencing run were analysed by both image processing and signal processing with the GS Run Browser. Subsequently, the sequence reads were assembled into contigs using the GS De Novo Assembler. The sequencing and assembly results are summarized below: Genes were annotated using the software Diya and data set up in the Geneious™ format.

Total number of sequenced bases: 64.180.790

Total number of sequence reads: 195.793

Total number of contigs: 74

Number of assembled bases: 2.597.446

Number of large contigs (larger than 500 bp): 57

Number of bases in large contigs: 2.592.655

Average contig size: 45.485

N50 contig size: 82.666

Largest contig: 199.648

Estimated genome coverage: 24,7

Fylgiskjal 2.

Genetic engineering tools for thermophiles – progress beyond the state of the art

Genetic engineering of thermophiles has been hindered by the paucity of selectable markers because of the thermolability of many antibiotics as well as of the products of the most commonly used resistance genes [22-24]. Thermolability is also dependent on other factors, such as pH and presence/absence of oxygen. For example, several aminoglycosides remain effective following prolonged incubation in anaerobic growth medium at 75°C. At 72°C and pH 7.3, chloramphenicol, streptomycin, penicillin G, tetracycline, and ampicillin lose activity while kanamycin and neomycin remain active, whereas, at 72°C and pH 5 only chloramphenicol and streptomycin remain active [22]. Also, thermolability of the corresponding antibiotic marker proteins has prevented their use for genetic engineering of thermophiles. Still, antibiotic selection has been successfully applied in some moderate thermophiles [25-27] which subsequently enabled the *in vivo* selection for more thermostable forms of resistance determinants [28, 29]. The thermo-adapted genes in turn have been used as selectable markers in more extreme thermophiles (Table 1). However, due this thermolability of the antibiotics and the corresponding marker proteins, selection in thermophiles is often based on complementation of auxotrophy (Table 1).

Metabolic engineering may require several genetic manipulation steps for improving or instituting production of a target metabolite. The success therefore depends on the availability of multiple functional markers or on availability of efficient marker recycling strategies. Marker recycling enables the reuse of a particular marker in subsequent genetic engineering work. The marker gene is then removed following the gene insertion/deletion step, by a deletion in a second recombination step. Consequently, the insertion/deletion event becomes "unmarked" and the marker can be used for another genetic manipulation. The scarcity of markers may limit the scope of what can be achieved in reconstruction of a cellular network for more effective production of a target metabolite. Therefore, one of the main objectives of the THERMOFINERY proposal is to expand and improve the genetic tools available for the model organisms. The interchangeability of markers between the different species will be examined, including for the new thermo-adapted antibiotic markers in Table 1. Also public and proprietary genome databases will be examined for the presence of putative antibiotic resistance genes. However, the main focus will be on adapting or establishing new marker recycling methods for the model bacteria. In this respect, the applicants intend to develop a markerless disruption method similar to the Flp recombinase method [30] which has been established and used in various mesophiles. Applying a corresponding method to a thermophile, a *Thermus* transposase recombination recognition sites will be placed into an insertion cassette flanking a selection marker, yet within the flanking sites used for the insertion by homologous recombination. Following insertion, expression of the transposase will induce deletion of the selection marker through recombination at the transposase recognition sites. This transposase system in *Thermus* is currently under study by Matis. (unpublished results). Other, recycling methods will be based on the *pyr* markers. A toxic uracil analog, 5-fluoroorotic acid (5-FOA) is used to select for deletion of the *pyr* genes (*pyrE* or *pyrF*). A Δ*pyr* strain is a uracil auxotroph that can be restored to a prototroph in a second step involving insertion of the corresponding *pyr* marker. Again insertion into *pyr* (inserted in step 2) gives again *ura/pyr*- selection on 5-FOA in the presence of uracil. This method was applied by the targeted gene knockout in *Clostridium thermocellum* of the *pta* gene that encodes the enzyme phosphotransacetylase [31]. Similar to this, the applicants intend to delete other genes in the *Thermoanaerobacterium islandicum* coding enzymes from particular metabolic pathways of compounds, which can be replaced by halogenic derivatives. The gene will then be used as a selection marker in a following engineering approach. By using the halogenic derivatives in the medium, the presence of active enzyme will generate a toxic compound from that derivative. Thus, a positive selection for a marker loss will be created comparable to *pyr* recycling system.

Table 1. Functionality of genetic markers in thermophiles. Red circles indicate thermo-adapted resistance determinants obtained by applying selection in thermophiles.

	Prototrophic selection					Selection based on resistance							
Genus	His^{+}	Lac^{+}	Leu^{+}	Trp^{+}	Ura^{+}	Adh^{+}	Blm^{R}	Cm^{R}	Ery^{R}	Hyg^{R}	Km^{R}	Mev^{R}	Tc^{R}
Geobacillus								●			●		●
Thermus			●	●	●		●			●	●		
Clostridium					●				●				
Thermoanaerobacter								●			●		
Thermoanaerobacterium		●			●	●			●	●	●		
Pyrococcus					●	●						●	
Thermococcus	●			●	●							●	

His^{+}: Histidine selection in *Thermococcus kodakarensis* based on the native *hisD* gene [32]. Lac^{+}: Lactose selection in *Sulfolobus solfataricus* based on the native *lacS* gene [33]. Leu^{+}: Leucine selection in *Thermus thermophilus* based on the native *leuB* gene [34]. Trp^{+}: Tryptophan selection in *T. thermophilus* and *T. kodakarensis* based on the native *trpB* [35] and *trpE* [32] genes, respectively. Ura^{+}: Uracil selection in *T. thermophilus* based on the native *pyrE* [34], in *S. solfataricus* [36] and *Sulfolobus islandicus* [37] on the *S. solfataricus pyrEF*, in *Pyrococcus abyssi* on *pyrE* from *Sulfolobus acidocaldarius* [38] and in *T. kodakarensis* [32] and *Clostridium thermocellum* [39] on the native *pyrF*. Adh^{+}: Butanol and benzyl alcohol resistance in *S. acidocaldarius* and *Pyrococcus furiosus* conferred by the Adh from *S. solfataricus* [40]. Blm^{R}: Bleomycin resistance in *T. thermophilus* conferred by the thermoadapted Shble from *Streptoalloteichus hindustanus* [23].

Cm^{R}: Chloramphenicol resistance in *Geobacillus stearothermophilus* [41] and *Thermoanaerobacter thermohydrosulfuricus* [27] conferred by the pC194 *cat*. Ery^{R}: Erythromycin resistance in *Thermoanaerobacterium thermosaccharolyticum* [25], *Thermoanaerobacterium saccharolyticum* [42] and *Clostridium thermocellum* [43] conferred by the plasmid pCTC. Hyg^{R}: Hygromycin resistance in *T. thermophilus* [44] and *S. solfataricus* [45] conferred by the thermoadapted Hph from *E. coli*.

Km^{R}: Kanamycin resistance in *G. stearothermophilus* [26] and *T. thermohydrosulfuricus* [27] based on the KNTase from pUB110 and in *G. stearothermophilus* [46] and *T. thermophilus* [47] based on the thermoadapted KNTase [28, 29]. Also in *T. saccharolyticum* based on Aph from *Streptococcus faecalis* [48] and in *T. thermophilus* conferred by the thermoadapted HTK [49]. Mev^{R}: Mevinolin/Simvastatin resistance in *T. kodakarensis* [50] and *P. furiosus* [51] conferred by the species' *hmg*. Tc^{R}:Tetracyclin resistance in *G. stearothermophilus* mediated by plasmid pTB19 [26].

Fylgiskjal 3.

Identification of cellulases and hemicellulases in the sequenced genomes.

Thermoanaerobacterium islandicum, Calcicellulosiruptor kristjanssonii and Rhhodothermus marinus.

Genomes of two lignocellulosic thermophilic anaerobes were partially sequenced gens using GFX pyrosequencer from Roche and their gene content analysed for polysaccharide degrading genes. One was a novel ethanol producing thermophilic anaerobe isolated in a parallel project belonging to the genus *Thermoanaerobacterium.* The coverage was 13 fold. The other one was a cellulosic, hydrogen producing organisms belonging to the species *Caldicellulosiruptor kristjanssonii.* The sequence coverage was 6 fold.

Genes were annotated and cellulolytic enzymes and hemicellulases were noted. Only one putative cellulolytic gene was found in *T. islandicum* but many hemicellulases were detected in the genome including xylanases and mannanasesA number of putative cellulase and heicellulae genes were identified in the *C. kristjanssonii* strain. Table xxx shows the genes identified in these strain, *Rhodthermus marinus* from a previous sequencing project and from biomass,

ENZYME	SIZE	GH-family	Identiy %	Closest relative (species)
beta-D-glucan exohydrolase/beta-glucosidase	600 aa	3+3	62	Solibacter usitatus
beta-glucosidase	782 aa	3+3	53	Sphingopyxis alaskensis
xylosidase (b-glucosidase)	780 aa	3+3	49	Thermoanaaaerobacter brockii
xylosidase (b-glucosidase)	800 aa	3+3	54	Sphingomonas sp. SKA58
xylosidase	800 aa	3+3	53	Pseudoalteromonas atlantica
beta-glucosidase	800 aa	3+3	48	Aeromicrobium erythreum
xylanase	285 aa	10	90	Geobacillus stearothermophilus
xylanase or beta-1,4-cellobiosidase	535 aa	10	40	Solibacter usitatus
xylanase or beta-1,4-cellobiosidase	300 aa	10	38	Pseudomonas sp. PE2
xylanase	380 aa	10	72	Caldocellulosiruptor saccharolyticus
xylanase	688 aa	10	93	Anaerocellum thermophilum from
b-glucosidase	455	1	93	Caldicellulosiruptor saccharolyticus
endoglucanase/ cellulase	337	5	82	Caldicellulosiruptor saccharolyticus
endoglucanase/ cellulase	235	5	62	Verminephrobacter eiseniae
endoglucanase/ cellulase	269	5	49	Prevotella ruminicola
endo-Cellulase		5	55	Cellvibrio mixtus
active on microcrystalline cellulose	1711	9	83	Anaerocellum thermophilum
Active on avicel CMC and xyloglucans	839	74	85	Caldicellulosiruptor saccharolyticus
xylanase	1060 aa	10	55	Thermotoga sp. strain FjSS3-B.1
xylosidase	540 aa	43	63	Sinorhizobium meliloti
endo-b-mannosidase	670	10	54	Bacillus licheniformis
mannosidase	1050 aa	2	49	Roseiflexus sp. RS-1
mannosidase	911	26	90	Caldicellulosiruptor saccharolyticus
mannanase	389	26	45	Flavobacterium johnsoniae

Table 1. Genes identified in in genome and metagenome libararies.

Selected genes for cloning and expression and production

Overview

Six putative cellulolytic genes, **CeluB , CeluC, CeluD, CeluE, CeluF** and **GluA** ,were selected cloning, expression and characterization. Three encode for putative endo-cellulases, one was a potential cellobiohydrolase and two were putative beta-glucosidases. Two, enzymes, **CeluA** and **CeluC**, were part of a production studies in *Streptomyces lividans.* One hemicellulase gene was selected for cloning and expression and characterization, **ManA,** a putative mannosidase.

Cellulolytic enzymes

CeluA: Family **GH12**, an endocellulase from the thermophilic bacterium *Rhodothermus marinus*. This gene was already cloned, expressed and characterized at the onset of this project. . It will be examined further in the project on pre-treated biomass. Status: has been produced and results published.

CeluB: Family **GH5**, a partial gene from a sequenced genome of an anaerobic thermophilic bacterium that appears to be attached to three cellulose binding domains. Homologues in Genbank have no assigned activity and its function is therefore unknown. It is likely that it has cellulolytic activity and novel as such. Status: The gene has been completely retrieved, cloning and expression is ongoing.

CeluC: Family **GH5**, an endocellulase from the aerobic thermophilic bacterium *Rhodthermus marinus*Status: Retrieved, cloned and characterized regarding temperature and pH optima and stability.

CeluD: Family **GH3**, a beta-glucosidase that was detected in the partially sequenced genome of the thermophilic aerobic bacterium *Rhodothermus marinus*

Status: The gene has been completely retrieved, cloned and characterized.

CeluE: Family **GH9**, a putative endocellulase gene from a partially sequenced genome of an anaerobic thermophile *Caldisellulosiruptor kristjanssonii.* Status: The gene has been completely retrieved, cloned and expressed. Activity could not be assigned.

CeluF: Family GH1, a beta-glucosidase that was detected in the partially sequenced genome of the thermophilic aerobic bacterium *Caldisellulosiruptor kristjanssonii*

Status: The gene has been completely retrieved, cloned and characterized.

Hemicellulases

GluA: Family **GH9**, Putative endocellulase gene from a partially sequenced genome of an anaerobic thermophile *Caldisellulosiruptor kristjanssonii* that may be an endocellulase. Only part of the gene

was available, : The complete gene was retrieved, cloned and expressed. Activity was only found on mixed linkage glucan could be found.

An endocellulase from *Rhodothermus marinus*: CeluA

Summary of obtained results:

A) **The enzyme was produced in gram quantities in *E. coli* for further testing in pre-processing**

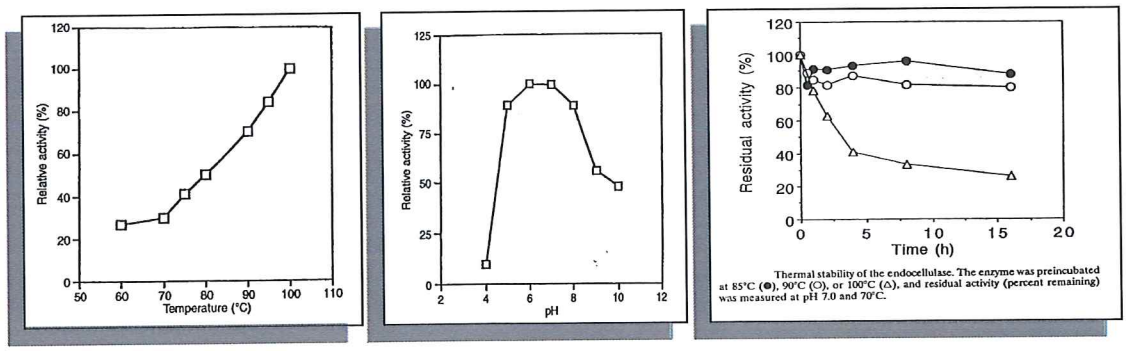
B) **The enzyme was also produced in *Streptomyces lividans*, where it is expressed in greater quantities than in *E. coli* and, furthermore, secreted into the medium.**

Introduction

CeluA from *Rhodothermus marinus* is an endocellulase (EC 3.2.1.4) originally cloned and expressed in *E. coli* from the thermophilic aerobic bacterium *Rhodothermus marinus.* The bacterium belongs to the phylum of *Bacteroidetes* and it is a versatile heterotroph. This is a relatively short protein, 260 amino acids. It has a signal sequence and also a short N-terminal membrane attachment part/hook. Investigations indicate that this N-terminal part is cytotoxic in *E. coli*. Interestingly this enzyme belongs to the glycosyl hydrolase family **GH12** and is a homologue to an endocellulase (**Cel12A)** found in *Thriocderma reesei,* one of the benchmark enzymes for cellulose degradation. A truncated version, missing the N-terminal signal and attachment parts was patented by Prokaria.

Properties of CeluA

The pH optimum for this enzyme is relatively broad, between 6 and 8, with the highest activity at pH 7.0. At pH 4.4 and 8.8, the relative activity is more than 50% of the maximum activity. Using CMC as the substrate, 10-fold linear increase in activity was observed between 60 and 95°C, with the highest initial activity level at 95°C. Thermostability was determined by preincubating the enzyme (4 mg/ml) at 80, 90, and 100°C and measuring residual activity. The half-life of the enzyme was 3.5 h at 100°C, and 80% of the activity remained after 16 h at 908C. No inhibition was observed with 200 mM glucose or cellobiose in CMC plate assays. Specific activity is rather low. (GO.



Figures 1, 2 and 3. Temperature, pH optima and thermostability of CeluA.

A putative endocellulase from the thermophilic anaerobe *Thermoanaerobacterium islandicum*: CeluB

Summary of obtained results

A) Complete gene retrieved.
B) Gene cloned and expressed .
C) Activity on celllose not detected.

Introduction

Genome of the thermophilic ethanol producing anaerobe *T. islandicum* was partially sequenced. Following the sequencing a putative cellulase gene belonging to glycosyl hydrolase family **GH5**, was detected. It was incomplete as the 5' upstream region was missing

Results

The **CeluB** gene was retrieved by PCR gene walking in three runs as illustrated in figure 5, where the thin line indicates the known part of the gene and three arrows indicate the direction of retrieval using two of the mentioned primers. The retrieval procedure was terminated when a putative start codon with an upstream Shine Delgarno sequence was identified. The expressed enzyme was tested on various substrates but t**he activity could not be assigned to it. The deduced amino acid sequence showed highest identity with a putative cellulase enzyme family GH5.** However, it also shows high similarity with an endo-beta-mannanase and a beta-galactosidase. .

The putative cellulolytic gene, **CeluB,** belonged to family **GH5.** No close relative has been characterize yet. Different activities are found in this group, chitosanases (EC 3.2.1.132); β-mannosidases (EC 3.2.1.25); (EC 3.2.1.4);), glucan 1,3-β-glucosidases (EC 3.2.1.58); licheninases (EC 3.2.1.73); glucan endo-1,6-β-glucosidases (EC 3.2.1.75); mannan endo-1,4--mannosidases (EC 3.2.1.78); Endo-1,4-β-xylanases (EC 3.2.1.8); cellulose 1,4-β-cellobiosidases (EC 3.2.1.91); endo-1,6-β-galactanases (EC 3.2.1.-); β-1,3-mannanases (EC 3.2.1.-) and xyloglucan-specific endo-β-1,4-glucanase (EC 3.2.1.151). Most of the enzymes are however cellulase or endoglucanases (EC 3.2.1.4);),

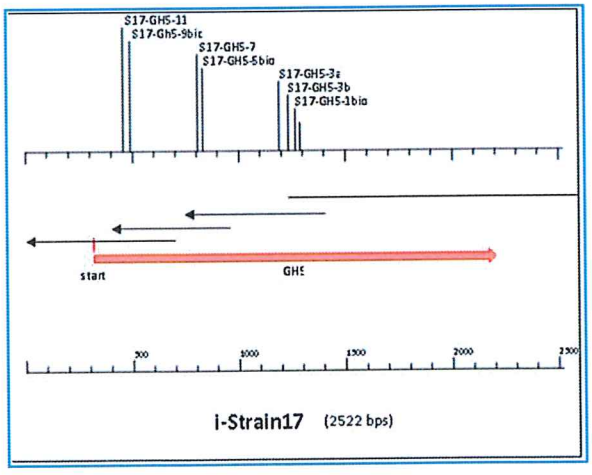


Figure 5. Retrieval of GH5 sequence from a thermophilc anaerobe

An endocellulase from *Rhodothermus marinus*: CeluC

Summary of obtained results

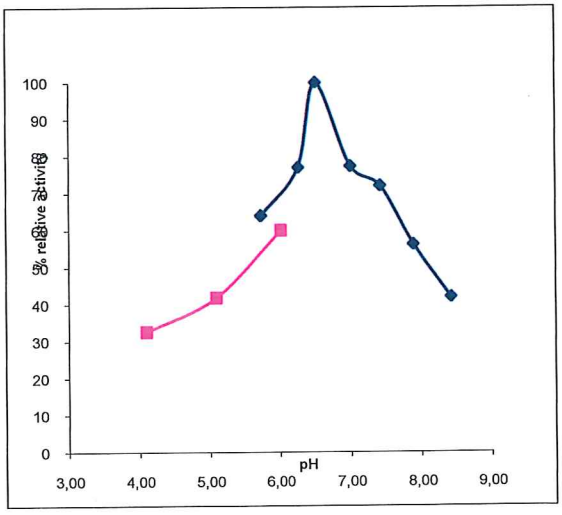
A) Complete gene retrieved.
B) Gene cloned and expressed in *E. coli.*
C) Activity on cellulose detected and different optima determined,
D) Enzyme cloned and expressed in *Streptomyces lividans*

Introduction

CeluC was detected in the partially sequenced genome of the thermophilic aerobic bacterium *Rhodothermus marinus*. It has 357 amino acids and 56% identity with the closest known relative (uncharacterized so far). It had a distinct signal peptide sequence. This enzyme belongs to the glycosyl hydrolase family GH5 as the CeluB enzyme (described avove).

Results

The gene was cloned into both *E. coli* and *S. lividans* hosts. In the latter system it was cloned with optimized codon composition and also the native signal peptide was substituted with one from *S. lividans*. Expression was little in *E. coli* but fairly good in *S. lividans* where it was furthermore secreted into the medium. The expression of CeluC in *Streptomyces* is less than that of CeluA (described above). Figure 9 shows the expression in *Streptomyces lividans*.
The enzyme shows activity on beta 1,4 and mixed 1,3,-1,4 linkage glucans. It was active on beta-glucan, lichenan glucomannan and CMC-cellulose. Specific activity was high compared with CeluA enzyme from the same species. Figures 7 and 8 show pH and temperature optima of the enzyme. The pH optimum for this enzyme is at pH 7 and 50% of the maximum activity is observed between pH 5.5 and pH 8. Using CMC as substrate, temperature optimum was found to be at 80-85°C. 50% of maximum was observed between 60°C and 90°C.



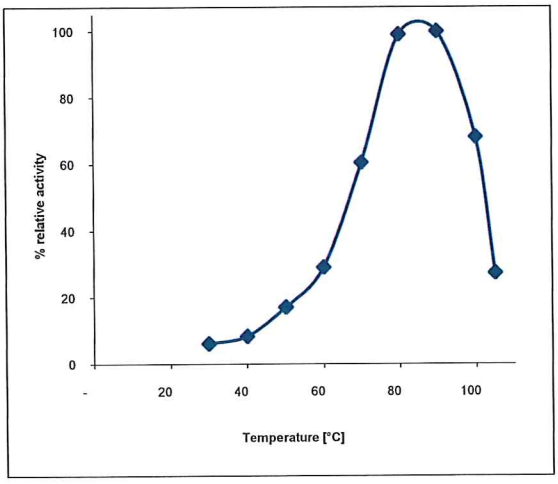


Figure 7 and 8. Temperature and pH optima of CeluC.

A beta-glucosidase from Rhodothermus marinus: CeluD

Summary of obtained results

A) **Complete gene retrieved.**
B) **Gene cloned and expressed in *E. coli.***
C) **Activity on cellobiose detected and different optima determined,**
D) **Enzyme also cloned and expressed in *Streptomyces lividan***

Introduction

CeluD is a beta-glucosidase that was detected in the partially sequenced genome of the thermophilic aerobic bacterium *Rhodothermus marinus*. The enzyme is 764 amino acids long and has 56% identity with the closest known relative (from *Sphingopyxis alaskensis*). It had a distinct signal peptide sequence. It belongs to family GH3 and supposedly degrades cellobiose. The beta-glucosidase is expected to complement in degradation of cellulose the two endocellulases, CeluA and CeluC, from the same species. At least these two different activities are needed for complete degradation of cellulose provided that the chemical pre-processing step has disrupted the cellulose crystalline structure.

Results

The gene was expressed in *E. coli* using rhamnoseThe pH optimum for this enzyme is 6 - with 50% of the maximum activity between 4 and pH 8. Using PNP-D-D-glucopyranoside as substrate, temperature optimum was found to be at 80°C. 50% of maximum was observed between 65°C and 85°C.

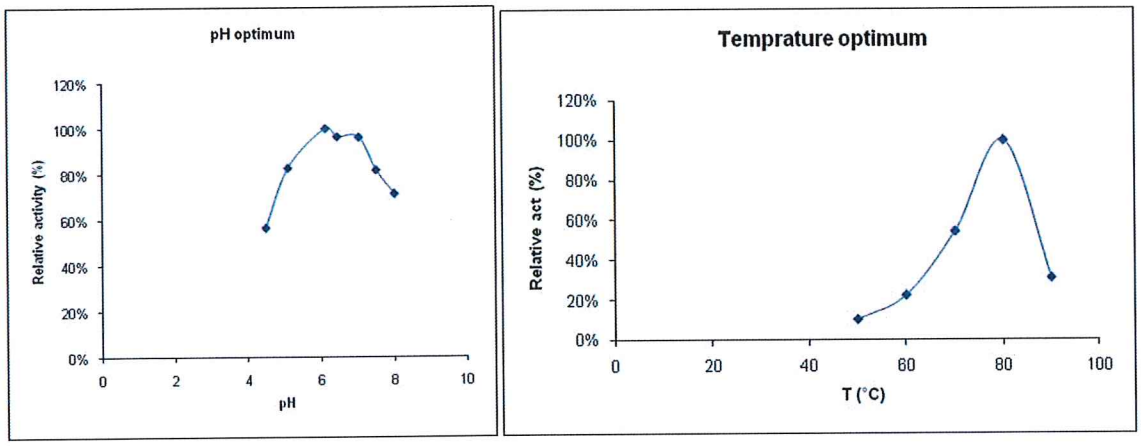


Figure 10 and 11. Temperature and pH optima of CeluD.

A putative endocellulase from *Caldicellulosiruptor kristjanssonii*: CeluE

Summary of obtained results.

A) **Complete gene retrieved.**
B) **Gene cloned and expressed in *E. coli.***
C) **No polysaccharide degrading activity could be detected**

Introduction

Putative endocellulase gene was detectged in a partially sequenced genome of an anaerobic thermophile *Caldisellulosiruptor kristjanssonii.* **CeluE** is a 328 amino acids long putative cellulase that has 82% peptide sequence similarity to its closest relative, a glycoside hydrolase family 5 (**GH5**) protein from *Caldicellulosiruptor saccharolyticus* DSM 8903 (Genbank accession number YP_001178980). According to conserved domains search (NCBI) CeluE is classified as a cellulase and has highest similarity to glycosyl hydrolase **GH5** proteins.

Results

CeluE was cloned into the pJOE3075 with a putative signal peptide on the N-terminal and without. Expression was obtained ednd the enzyme was screened for activity on different substrates, alginic acid, amylose, laminarin, b-glucan, CMC, cellulose, Avicel, crystalline cellulose, lichenan, pustulan, cudlan, or cellooligosaccharides, No activity could be detected.

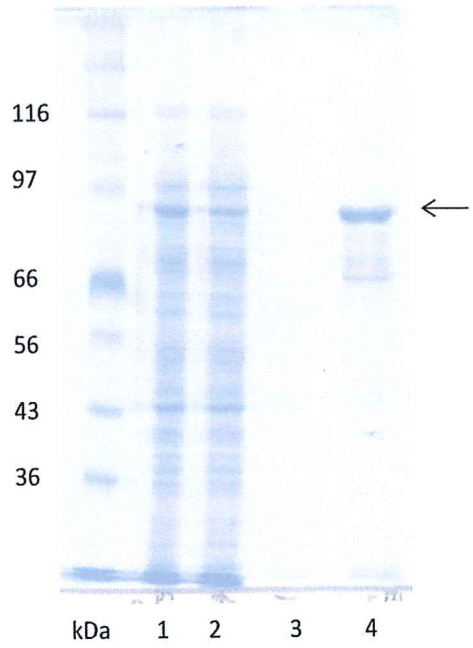


Figure 13. His-trap column purification of CeluE. Lanes: protein marker, (1) crude extract, (2) flowthrough, (3) eluted with buffer containing 75mM imidazole and (4) 200mM imidazole (purified protein (arrows)). (kDa) prestained protein marker

A beta-glucosidase from *Caldicellulosiruptor kristjanssonii*: CeluF

Summary of obtained results.

A) Complete gene retrieved.
B) Gene cloned and expressed in *E. coli.*
C) Activity on cellulooligosaccharides
D) Different optima determined,

Introduction

Putative beta glucosidase gene was detectged in a partially sequenced genome of an anaerobic thermophile *Caldisellulosiruptor kristjanssonii.*. It belongs **GH1** family enzymes according to conserved domains (NCBI). The 452 amino acids long β-glucosidase studied in here has 93% peptide sequence similarity to a β glucosidase (GenBank accession number YP001179893) in *Caldicellulosiruptor saccharolyticus* DSM8903 characterized in Hong et al., 2009. The β-glucosidase in *C. saccharolyticus* was shown to be a broad substrate specificity enzyme with highest activity on cellobiose. It had an optimum temperature of 70˚C and pH of 5.5 on para-nitrophenyl β-D-glucopyranoside and hydrolyzed cello-oligosaccharides in an exo-acting manner.

Results

The whole gene sequence for CeluF was retrieved by using a reverse primer based on the sequence of β glucosidase from homologous gene in *Anaerocellum thermophilum*. The enzyme that was cloned and expressed was functional and therefore it was believed to be similar to the native CeluF enzyme and could be used for further studies.

TLC showed that CeluF had activity on cellobiose, cellotriose, cellotetraose and cellopentaose and that cellobiose and the cello-oligosaccharides in the end were all fully hydrolyzed to glucose (figure 11). The β-glucosidase from *C. Kristjanssonii* had a higher temperature optimum and broader pH range than its homolog in *C. saccharolyticus* on p-NPGp. It showed highest activity towards p-NPGp at 80°C and at a pH of 6.5 (figure 12). The β-glucosidase also showed traces of activity on laminarin, lichenan and (barley) β-glucan, but not on carboxymethyl-cellulose, curdlan or pustulan (data not shown). These results suggests that the enzyme can catalyze the hydrolysis of β 1,3-1,4 and β 1,3-1,6 linkages. The thermal stability at 80˚C was very short but at 60˚C it still had over 60% activity after 48h (figure 13).

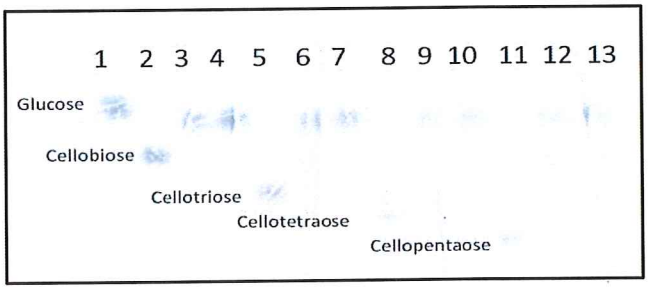


Figure 11. TLC of CeluF activity on cellobiose, cellotriose, cellotetraose and cellopentaose. 1: glucose standard, 2: cellobiose, 3: 10 min, 4:2h, 5: cellotriose, 6: 10 min, 7: 2h, 8: cellotetraose, 9: 10 min, 10: 2h, 11: cellopentaose, 12: 10min, 13: 2h. 10min and 2h represent length of reaction time before the samples were spotted on a TLC plate.

A) B)

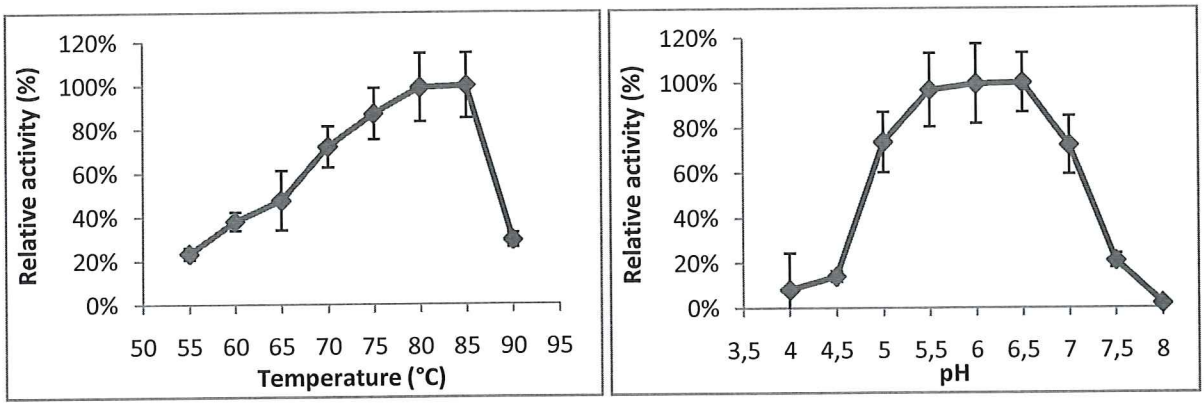


Figure 12. (A) Effect of temperature and (B) of pH on activity for CeluF. Data represents means from three separate experiments.

Specific activity for CeluF on nitrophenyl chemicals showed that CeluF had highest activity towards p-NPGp (Table 2). CeluF had about a 6-fold lower activity on o-nitrophenyl-β-D-galactopyranoside and only some faint activity on p-nitrophenyl-α-D-glucopyranoside compared to p-NPGp.

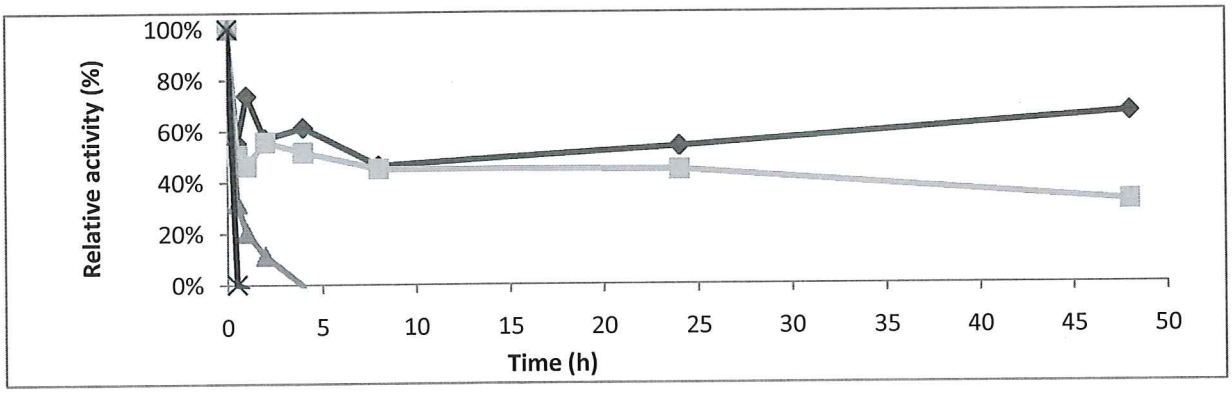


Figure 13. Thermal inactivation for CeluF. The enzyme was incubated from 0-48h and aliquots were taken at various time points and tested for activity on p-NPGp. 60˚C (tilted squares), 70˚C (squares), 80˚C (triangles) and 85˚C (crosses).

Substrate	Specific activity (U/mg)
p-Nitrophenyl-β-D glucopyranoside	340
p-Nitrophenyl-alpha-D glucopyranoside	3.80
p-Nitrophenyl-alpha-D galactopyranoside	NA
o- Nitrophenyl-β-D-galactopyranoside	53.5

Table 2. Specific activity for CeluF on different substrates. NA= no activity detected with this method. Data represents means of three separate experiments.

Conclusion

The substrate specificity of CeluF suggests that it belongs in the group of broad substrate specific β-glucosidases. The fact that it has faint activity on Laminarin (β 1,3-1,6 linkages) and the β 1,3-1,4 linked substrates lichenan and β-glucan (barley) further shows that it has a broad substrate specificity. Activity on more substrates should be screened for to find out just how versatile this enzyme is. CeluF seems to have similar characteristics compared to the β-glucosidase in *C. Saccharolyticus* regarding stability and activity. CeluF had a higher optimum temperature (80°C compared to 70°C) and had highest activity at a pH of 6.5 compared to pH5.5 for the β glucosidase in *C. Saccharolyticus*. However, the thermal stability was about the same in both enzymes suggesting that the 7% difference in peptide sequence does not make a big difference between the enzymes regarding stability.