論 文 の 内 容 の 要 旨

水圏生物科学専攻
平成 22 年度博士課程進学
氏　　名　ホネイン　カリム　サード
指導教員名　潮　秀樹

論文題目　　A holistic approach to the endogenous contribution of *Teredo navalis* in the biodegradation of lignocellulose using cDNA cloning and pyrosequencing
（フナクイムシのリグノセルロース分解に関する分子生物学的研究）

Introduction

The current energetic problem, resulting from the dependence of our societies and economies on fossil fuels as well as the increasing demand from developing countries have resulted in an increased interest and focus on renewable energies and a renewed interest in biomass as a potential source of alternative energy source. The ubiquitous availability of lignocellulose in addition to its high content of sugars makes it attractive for investigating new alternative sources of energy. However, an essential obstacle to the optimal use of lignocellulose as a source of sugar is the toughness and resilience of the molecular structure to the chemical and enzymatic treatments currently in use.

The traditional way of studying the degradation of lignocellulosic biomass involved essentially fungi and microorganisms. In fact, it was long believed that wood-feeding animals such as termites were able to feed on lignocellulose solely because of a close symbiotic relation with lignocellulolytic protists and microbia. However, in the last two decades, endogenous cellulases have been identified in different eukaryotic including termites, nematode, and in several species of mollusks. With the discovery of endogenous animal cellulases and other lignocellulose-degrading genes, a new integrated outlook at the lignocellulolytic systems in animal started to appear in the field.

Teredo navalis is a xylotrophic bivalve of the Teredinidae family commonly known as shipworms and of the order Myoida. *T. navalis* is suspected of

feeding exclusively on wood, hence to effectively break lignocellulose down to its building block sugars. Our interest in the great shipworm comes from the reputation of this animal as a great wood borrower and of the scarcity of xylotrophy in the animal world. We believe that a study of an efficient lignocellulolytic system from an animal can clarify many aspects of the whole process and allow us to understand the reaction from a vitally important holistic approach. Chapter 1 described the first full-length cDNA cloning and in silico structural analysis of an endogenous cellulase of the glycoside hydrolase family (GHF) 45 in *T. navalis*. The first Expressed Sequence Tag (EST) study of the lignocellulolytic system of the shipworm using the Roche GS FLX system was carried out in Chapter 2. Finally, the xylotrophic abilities of the shipwrom were investigated through a whole transcriptome analysis using the recent Ion Torrent PGM system in Chapter 3.

The PCR experiments using degenerate primers yielded a cDNA fragment of 488 nucleotides. This initial fragment was amplified from both its 5` and the 3`cDNA ends reaching 740 nucleotides. The BLAST of this sequence showed a high similarity with endoglucanase of the GHF45. This sequence is the first reported endogenous cellulase cloned from *T. navalis*. The resulting nucleotide sequence was translated into a 236 amino acid long sequence including a signal peptide of 20 amino acids and remarkably high number of aspartic acid resides compared to other mollusk GHF45 endoglucanases. The essential role of the catalytic proton donor and the catalytic nucleophile were assigned to two catalytic aspartates separated by a distance of approximately 8.5 Å. *T. navalis* GHF45 alignment with that of molluskan GHF45 clearly shows the conservation of the first Asp at position 27. A model constructed from the deduced amino acid sequence based on the known model of *M. edulis* clearly shows a single characteristic cleft on the surface and Asp27 sits inside the cleft. The GHF45 classification also requires the second catalytic domain to reside too in the surrounding with at a distance of 8.5 Å. *T. navalis* GHF45 amino acid sequence was estimated to contain 20 Asp. The only possible candidate for our model is Asp139 sitting at an estimated distance of 8.86 Å from Asp 27. The model proposed for *T. navalis* GHF45 is therefore proposing a possible answer to the prerequisite of GHF45 classification. Just like *M. edulis*` GHF45, the cleft is clearly visible on the surface of the protein and in both cases the Asp residues of the catalytic domain sit on the cleft and are

separated by a distance of about 8 to 9 Å.

Furthermore, there are molecular evidences demonstrating that these marine lignocellulolytic systems are quite different from that of other systems observed in xylotrophic terrestrial animals. The shipworm symbiotic community is far simpler and is phylogenetically distinct from those found in termites and ruminants. The first deep sequencing of ESTs from *T. navalis* was carried out using a Roche GS FLX 454 next-generation sequencer. We constructed the EST library by sequencing the 3` end reverse transcribed poly(A)+ RNA fragments extracted from the whole body of *T. navalis*. After an RNA adaptor ligation to the 5`-phosphate of the Poly(A)+ RNAs, cDNAs were produced by reverse transcription of the isolated RNA, using the oligo(dT)-adapter primer and used for the 454 pyrosequencing.

A total of 165564 reads with an average length of 348 bp were obtained and 3720 reads were rejected for too short. The redundant reads were assembled into 12879 contigs and the remaining 80, 891 unassembled reads termed singlets, were grouped separately. The average length of contigs is 484 bp with the longest sequence of 1656 bp and the shortest assembled sequence of 41 bp. The complete dataset was queried by using the BLAST2GO software. Each unique sequence was subjected to similarity searches and compared with NCBI database using the BLASTx software. Through the positive hits looking for annotations related to the biodegradation of lignocellulose, several cellulases, hemicellulases and lignin degrading laccases genes were identified, where 11 sequences code cellulase genes were identified. These sequences were endoglucanases of the GHF9 and GHF45, as well as β-glucosidases of the GHF1 and GHF3. Twenty-five sequences with hits related to the degradation of hemicellulose were also identified such as β-endoglucanases, β-endoxylanases, β-galactosidases, β-endomannanase, and β-xylosidase. Hemicellulose degrading genes were distributed amongst GHF2, 3, 10, 16, 17, 26, 30 and 43. Four sequences coding for carbohydrate binding domains (CBD) were also seen. Some of these CBDs belonged to family 2, involved in the degradation of cellulose and that of xylan. Ten sequences with positive hits coding for laccases were identified. Out of all the annotated EST sequences corresponding to GHs, the GHF17 was found to be the dominant family accounting for 26% of all reads suspected of coding for glycosyl hydrolases. The hemicellulose-degrading group has the highest expression with 53.5% of batch, the laccase group comes second with 33% and the cellulases comes third with only

13.25%. It seems that *T. navalis* invests the largest portion of biomass degrading efforts into breaking down hemicellulose and to a lesser extent laccase with cellulose representing the smallest portion.

A phylogenetic analysis of the GHF9 cellulase showed that *T. navalis* lignocellulolytic system shares similarities with different lignocellulose-degrading groups. Five sequences of the GHF9 endoglucanases were distributed amongst bacterial and mollusk cellulases. Four sequences belonged to the same group as the bacterial GHF9 and *teredinobacter turnerae* GHF9, a notorious symbiotic bacterium found in the shipworms. GHF9 animal cellulase genes, especially that of termites, sea squirts and abalones were suggested to have a common ancestor. The presence of cellulases, hemicellulases and ligninases in the EST suggests that the shipworm plays an important role in the degradation of wood, its main source of energy. The input from the shipworm itself is probably not sufficient for an optimal degradation, but rather complements the enzymatic secretion from symbiotic organisms.

Finally, the Ion PGM system which works by measuring the direct release of H+ (protons) from the DNA polymerization reaction was used to sequence a whole transcriptome library from the whole body of *T. navalis* in order to complete the previous EST analysis. A representative whole transcriptome cDNA library was constructed using the Ion Total RNA-Seq Kit V2 (Life Technologies, USA). The sequencing resulted in 5433102 reads with an average of 162 bp. The EST dataset consisting of contigs was assembled with GS De Novo. CLC Genomics Workbench (version 4.6.1; CLC Bio, Denmark) was used for the trimming of the contigs and the subsequent mappings, resulting in 8826 contigs with an average length of 411bp and a maximum of 6946. BLASTx was used to query the dataset for lignocellulolytic related genes. The investigation involved 160 sequences and resulted with two contigs coding for GHF9 cellulase, one coding for carbohydrate binding domain and several hemicellulose degrading genes. The GHF9 cellulases were investigated from a phylogenetic perspective, along with the previously identified partial sequences using a maximum likelihood method. The results confirmed our previous finding in regards to the distribution of GHF9 cellulase genes in *T. navalis*.