# Table of contents

## 1 Introduction
- 1.1 Composition of lignocellulose ................................................. 1
- 1.2 Carbohydrate-Active enZyme (CAZy) database .................................... 3
  - 1.2.1 Glycoside hydrolases .................................................................. 4
  - 1.2.2 Enzymes hydrolyzing lignocellulosic plant material .......................... 5
- 1.3 Applications of lignocellulose hydrolyzing enzymes .............................. 6
- 1.4 Metagenomics and enzymatic discoveries by metagenomic analysis .......... 7
- 1.5 Functional metagenomics ................................................................... 8
- 1.6 Marine environment and Global Ocean Sampling (GOS) project ............... 9
- 1.7 β-glucosidases .............................................................................. 10
- 1.8 Structure and amino acid conservation in GH1 β-glucosidases .................. 12
- 1.9 Structural stability of proteins ......................................................... 14
- 1.10 Aims and objectives of this study .................................................. 16

## 2 Materials and methods ........................................................................ 17
- 2.1 Microbiological methods ................................................................. 17
  - 2.1.1 Cultivation and storage of cultures ............................................... 17
  - 2.1.2 Preparation of chemically competent cells .................................... 17
  - 2.1.3 Heat shock transformation of competent cells ................................. 18
- 2.2 Molecular biology methods ............................................................... 18
  - 2.2.1 Vectors used in the study ............................................................ 18
  - 2.2.2 Codon optimization .................................................................... 20
  - 2.2.3 Isolation of plasmid DNA ........................................................... 20
  - 2.2.4 Site directed mutagenesis ............................................................ 20
  - 2.2.5 Agarose gel electrophoresis ......................................................... 21
  - 2.2.6 Molecular cloning ........................................................................ 21
- 2.3 Biochemical and biophysical methods ................................................ 22
  - 2.3.1 Overexpression and purification of recombinant proteins: .................. 22
  - 2.3.2 Size exclusion chromatography .................................................. 23
2.3.3 Determination of protein concentration ................................................................. 24
2.3.4 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) ........................................ 24
2.3.5 Determination of enzyme activity on solid agar media ........................................... 25
2.3.6 Preparation of cell-free extracts ................................................................................ 25
2.3.7 Analysis of enzyme activity of cell-free extract ......................................................... 26
2.3.8 Determination of optimum pH, temperature and stability ....................................... 26
2.3.9 Substrate specificity of purified proteins .................................................................... 27
2.3.10 Hydrolysis of disaccharide and glucosyl oligosaccharides .................................... 27
2.3.11 Transglycosylation activity ...................................................................................... 28
2.3.12 Kinetic analysis ........................................................................................................ 28
2.3.13 Enzyme inhibition .................................................................................................... 28
2.3.14 Circular Dichroism (CD) spectroscopy ................................................................... 29
2.3.15 Calculation of the melting point ($T_m$) ................................................................. 29
2.4 Bioinformatics methods ............................................................................................... 30
2.4.1 Bioinformatic analysis of protein sequences from GOS ........................................ 30
2.4.2 Analysis of selected metagenomic protein sequences .............................................. 30
2.4.3 Multiple sequence alignment .................................................................................... 31
2.5 Protein crystallization .................................................................................................. 31
2.5.1 Screening of optimal buffer conditions for crystallization ..................................... 32
2.5.2 Protein crystallization with inhibitor ........................................................................ 32
2.5.3 x-ray diffraction and refinement .............................................................................. 33

3 Results .............................................................................................................................. 34
3.1 Carbohydrate-active enzymes from a marine metagenome ....................................... 34
   3.1.1 A bioinformatic analysis of the GOS dataset revealed about 25,000 proteins as putative Carbohydrate-Active enZYmes (CAZymes) ......................................................... 34
   3.1.1.1 The GOS dataset shows wide diversity of glycoside hydrolases ...................... 35
   3.1.1.2 Molecular analysis of selected metagenomic proteins .................................... 37
   3.1.2 Assessment of function of selected metagenomic sequences .............................. 38
   3.1.2.1 Detection of endoglucanase and endo-hemicellulase activity using non-chromogenic substrates ............................................................................................. 39
   3.1.2.2 Detection of endoglucanase and endo-hemicellulase activity using chromogenic substrates ........................................................................................................ 40
   3.1.3 Detection of oligosaccharide degrading enzymes .................................................. 42
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3.1</td>
<td>Utilization of para-nitrophenyl (pNP) glycosides</td>
<td>42</td>
</tr>
<tr>
<td>3.1.3.2</td>
<td>Esculin hydrate as a substrate for β-glucosidase</td>
<td>43</td>
</tr>
<tr>
<td>3.1.4</td>
<td>The selected metagenomic proteins show a broad range of enzyme activities.</td>
<td>44</td>
</tr>
<tr>
<td>3.2</td>
<td>Comprehensive analysis of a naturally occurring mutation of the conserved amino acid in BglM-G1</td>
<td>46</td>
</tr>
<tr>
<td>3.2.1</td>
<td>BglM-G1 showed presence of a naturally occurring mutation of an otherwise invariant amino acid</td>
<td>46</td>
</tr>
<tr>
<td>3.2.2</td>
<td>BglM-G1and H75R are bona fide β-glucosidases</td>
<td>49</td>
</tr>
<tr>
<td>3.2.3</td>
<td>BglM-G1and its H75R were purified to homogeneity</td>
<td>49</td>
</tr>
<tr>
<td>3.2.4</td>
<td>BglM-G1and H75R exhibited similar biochemical characteristics</td>
<td>50</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Structural analysis of BglM-G1and H75R</td>
<td>56</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Loss of the salt bridge and electrostatic interaction network reduced structural stability of the protein</td>
<td>63</td>
</tr>
<tr>
<td>3.2.7</td>
<td>Kinetic analysis demonstrated a higher catalytic efficiency of BglM-G1 when compared to H75R due to a more efficient substrate binding</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>Discussion</td>
<td>68</td>
</tr>
<tr>
<td>4.1</td>
<td>GOS metagenome: a source of carbohydrate-active enzymes (CAZymes)</td>
<td>68</td>
</tr>
<tr>
<td>4.2</td>
<td>The use of multiple substrates is advantageous to detect multiple enzyme activities</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>The evolutionary loss of hydrogen bond interactions in BglM-G1 reduced its structural stability</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>The unresolved electron density near the active site does not affect the electrostatic network</td>
<td>72</td>
</tr>
<tr>
<td>4.5</td>
<td>The substitution of conserved Arg75 influenced turnover of substrates</td>
<td>74</td>
</tr>
<tr>
<td>4.6</td>
<td>The stability/catalytic efficiency trade-off in BglMG-1 is advantageous in cold marine environment</td>
<td>74</td>
</tr>
<tr>
<td>5</td>
<td>Summary</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>Zusammenfassung</td>
<td>79</td>
</tr>
<tr>
<td>7</td>
<td>References</td>
<td>81</td>
</tr>
<tr>
<td>8</td>
<td>Appendix</td>
<td>95</td>
</tr>
</tbody>
</table>