Access to Biodiversity and New Genes from Thermophiles by Special Enrichment Methods

PhD Thesis

Cédric F.V. Hobel

Department of Biology

Faculty of Sciences

University of Iceland

Reykjavik, 2004

Academic Dissertation

Submitted to the Department of Biology, Faculty of Sciences, University of Iceland
for the degree of *Doctor of Philosophy* (PhD) in Biology

Presented on Monday, December 20th 2004 at the University of Iceland

The Jury was composed of: Prof. Guðni A. Alfreðsson
Dr. Joël Querellou
Dr. Hörður Filippusson

Supervisors

Dr. Jakob K. Kristjánsson
Dr. Viggó Þór Marteinsson

PhD Advisory Committee

Dr. Jakob K. Kristjánsson
Dr. Viggó Þór Marteinsson
Dr. Guðmundur Óli Hreggvidsson

Doctoral Thesis
December 2004

ISBN:
Printed by Gutenberg, Reykjavik, Iceland, 2004

Faculty of Sciences, Department of Biology
University of Iceland
Askja, Sturlugata 7
101 Reykjavik, Iceland

Prokaria Ltd,
Biodiversity / Microbiology Dept.
Gylfaflöt 5
112, Reykjavik, Iceland

ÁGRIP

Almennt er talið að einungis sé hægt að rækta og einangra minna en 1% af allri náttúrulegri örveruflóru sem er til staðar í náttúrunni. Með þróun á aðferðafræði sameindalíffræðinnar hafa komið fram mikilvægar aðferðir sem byggja á kjarnsýrumögnun (PCR). Þessi tækni hefur hjálpað vísindamönnum til að lýsa örverusamfélögum í náttúrunni betur en áður var unnt án þess að rækta eða einangra örverurnar. Aðferðirnar veita hinsvegar oft takmarkandi upplýsingar vegna aðstæðna við sýnatöku og ýmissa eðlis- og lífefnafræðilegra áhrifaþátta í þeim. Þess vegna er nauðsynlegt auka fjölbreytileika aðferða sem notaðar eru við sýnatökur. Í þessari ritgerð lýsi ég þróun á aðferðum sem stuðla að myndun lífmassa og auknum örveruvexti við leiðandi aðstæður. Með því að beita þessum aðferðum var hægt að fá fram þær örverur sem voru hlutfallslega yfirgnæfandi en einnig að draga þær fram sem voru í minni hluta í náttúrunni.

Þessu doktorsverkefni má skipta í þrjá hluta. Í fyrsta hluta er fjallað um örverufjölbreytileika í tveimur heitum jaðarumhverfum (Extreme environments) sem rannsakaður var með sameindalíffræðilegum aðferðum. Borun niður í heit jarðlög og notkun jarðhita á Íslandi hafa skapað einstakar aðstæður til að nálgast heitt neðanjarðarumhverfi. Erfðagreiningar á fjölbreytileika örveranna í jarðhitaborholunum sýndu tilvist hitakærra örvera sem voru mjög fjarskyldar örverum sem finnast á yfirborðinu. Niðurstöðurnar gáfu einnig vísbendingu um að hitakærar örverur geti ferðast með neðanjarðarrásum milli jarðhitasvæða. Nýlega var neðansjávar háhitahverasvæði uppgötvað við Grímsey sem er á heimskautasvæði mið-Atlandshafshryggjarins. Lífmassa var safnað "*in situ*" úr uppstreymi jarðhitavatnsins sem kom úr neðansjávarhverunum. Rannsókn á lífmassanum leiddi í ljós mikinn örverufjölbreytileika. Þessar niðurstöður eru þær fyrstu sem sýna slíkan fjölbreytileika með sameindalíffræðilegum aðferðum á þessum slóðum.

Í öðrum hluta verkefnisins voru ræktunaraðferðir notaðar til að auðga fyrir örverum með sértæka eiginleika. Hvatinn að þessum hluta verkefnisins er þörf iðnaðarins fyrir ný ensím sem virka best við erfiðar lífeðlis- og efnafræðilegar

aðstæður (háan hita, hátt og lágt sýrustig, háan saltstyrk o.s.frv.). Slík ensím má finna í hitakærum örverum. Auðgað var fyrir örverum sem brjóta niður sterkju í brennisteinsríkum hver með því að nota hveravatnið og lágan styrk af sterkju. Með því að skima ræktirnar með sértækum kjarnsýruþreifurum tókst að fá kjansýruraðir af 18 nýjum α-amylasa genum en aðeins eitt þessara gena kom fram í "*in situ*" lífmassa. Blönduð aðferðafræði var notuð til að skoða fjölbreytileika kítínasagena í örverum sem lifa í strandhverum þar sem sjávarfalla gætir við Reykjanes í Ísafjarðardjúpi. Unnt var að einangra 36 ný kítínasagen úr auðgunum og örveruþekjum. Þessar niðurstöður sýndu fram á að hægt sé að fiska út ný og verðmæt gen úr náttúrunni með markvissum "*in vitro*" og "*in situ*" auðgunum.

Í lokahluta verkefnisins var genamengi salt- og hitakæru bakteríunnar *Rhodothermus marinus* skimað fyrir kítínasageni. Það var síðan einangrað, klónað og tjáð og genaafurðin hreinsuð. Ensímið reyndist vera hitastöðugast allra kítínasa sem hafa fundist í raunbakteríum og geta breytt kítínfjölliðum í tvíliður. Sá eiginleiki kítínasans að vinna best við hátt hitastig og lágt sýrustig er eftirsóttur til vinnslu fáliðaðra kítína á iðnaðarskala.

ABSTRACT

It is generally accepted that less than 1% of the naturally occurring microorganisms can be isolated and grown in pure culture. The development of PCR-based molecular phylogenetic techniques has provided scientists with tools to describe the natural diversity in microbial communities without the need of cultivation and isolation of single organisms. However, because of intrinsic biases of the method, only limited information can be accessed without improvements of various sampling strategies prior to the analysis of the sample. In this thesis, I describe the development and implementation of microbial colonization techniques and enrichment techniques that allow the access to new thermophilic microorganisms in nature.

This project had three focal themes. In the first part, the microbial diversity of two extreme thermophilic environments was described by using a state-of-the-art molecular phylogenetics method. The drilling into hot water reservoirs and harvest of geothermal energy in Iceland has created an unprecedented opportunity to reach subterranean thermophilic environments. Phylogenetic analyzes of the microbial diversity emphasized the presence of distinct populations apparently endemic to the subsurface. The results also showed that noticeable dissemination of thermophilic microbes occurs via subterranean conduits within subsurface volcanic areas. The Grimsey high-temperature hydrothermal vent field (North of Iceland) has been recently discovered on the subpolar Mid-Atlantic Ridge. By deploying *in situ* colonization devices, large amounts of biomass from the subsurface fluid were recovered and examined. This study brought the first results on the microbial diversity from the Mid-Atlantic Ridge submarine vents located near the Arctic Circle.

Secondly, alternative cultivation methods were developed to direct the enrichment of specialized microbes. This project was motivated by the need for new enzymes from thermophiles because they are best active under harsh biochemical conditions (high temperature, high and low pH, high salinity, etc.) and meet thus the technical requirements for industrial processes. Starch-degrading microorganisms were enriched *in vitro* from a sulfide-rich hot spring using low concentration of starch and untreated hot spring water. Molecular screening of the cultures with designed primers yielded 18 novel α-amylase genes whereas only one was obtained from the

original biomass sample. Similarly, a combined approach of *in situ* enrichment and *in situ* colonization was used to examine the chitinase genes diversity in intertidal hot springs on the seashore of North-West Iceland. A total of 36 novel chitinase genes were recovered. Both *in vitro* and *in situ* directed enrichment experiments proved highly efficient for accessing novel and valuable genes from nature.

In the final part, screening of the sequenced genome of the halophilic and thermophilic *Rhodothermus marinus* resulted in the isolation, cloning and expression of a chitinase gene. The purified enzyme was found to be the most thermostable chitinase known from *Bacteria*, being able to degrade chitin homopolymer into dimers. Moreover, the chitinase showed attractive applicability for the industrial production of chitin oligomers at high temperature and acid pH.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the following people who have been helping and supporting me for the completion of this thesis:

- Dr. Jakob K Kristjánsson for offering me the opportunity to do this PhD within Prokaria Ltd, for making possible for me to work in this fascinating world of thermophiles and for guiding me as a very patient leading supervisor.

- My supervisor Dr. Viggó Þór Marteinsson without whom none of this work would have been thought of, conducted, finished, analyzed, discussed, re-discussed, written about, corrected, re-written about, re-corrected and finally published. But this is only the tip of the iceberg…

- Dr. Gudmundur Óli Hreggvidsson for being such an amazing support of my work in practical as well as theoretical terms, and for being such a well of knowledge and inspiration.

- Sólveig Ólafsdóttir, Sigurbjörg Hauksdóttir and Sigrún Sveinsdóttir for being so reliably helpful in the laboratory, anytime.

- Ólafur Héðinn Friðjónsson, Sveinn Ernstsson and Farah Bahrani-Mougeot for everything regarding the secrets of molecular biology and good laboratory practice.

- Hákon Birgisson, Jón Óskar Wheat and Anna Guðný Hermansdóttir for all the really valuable help regarding protein characterization, purification and analysis.

- Hafliði, Solveig P, Arnthor, Sigurlaug, Jón M, Sigga Hj. for their support and help throughout my projects.

- The non-scientific part of Prokaria Ltd, Pétur Stefánsson and Valborg Guðmundsdóttir, for helping me so often resolving all my administrative problems.

- All the people, technicians and trainees at Prokaria and the Líftæknihús for their very good company and who have all one way or an other contributed to my success inside and outside the laboratory.

- I would like to thank Prof. Jean-Marc Jeltsch in Strasbourg for "helping" me choosing Iceland in the first place and for his very valuable time and discussions during my short visits on the Continent.

- My stay in Iceland has been rich in personal experiences. I would like to give a special mention to the scientific shipboard parties from the BIOICE campaign 2000, in particular Dr. Jörundur Svavarsson aboard R.V. Bjarni Sæmundsson, and from the Grimsey & Kolbeinsey cruise 2002 aboard R.V. Poseidon.

- All the friends that I made in Iceland, who chose to stay around or whose path have led somewhere else: Rannveig, Víktoria, Magali, Wilko, Brian, Lina, Óli, Bente, Rahul, Cédric, Christoph, Ute, Pratik, Konstantin, Kerstin, Hulda and Alma, Maggi and many, many, many more.

- All my friends in France, Vincent, Isabelle, Alexandre, Jean-Francois, Sophie, Olivier, and many more from the ESBS, from the Maitrise and school and university.

- My girlfriend Martina for being so supporting everyday, despite not being present in Iceland for the end of this long journey.

- My parents, my sister Réjane, Jérome her husband and their children, my grand-parents, and all my relatives “back there” for their encouragement and unconditional support throughout my studies and stay in Iceland.

- And of course, all of you that I did not specifically list.

This thesis has been financially supported by RANNÍS (Center for Research of Iceland), grants no. 03084003 and 040200004.

TABLE OF CONTENTS

List of Papers

This thesis is based on the following papers, referred to by their Roman numerals throughout the text and included at the end of the thesis.

I. Marteinsson VT, Hauksdottir S, **Hobel CFV**, Kristmannsdottir H, Hreggvidsson GO & Kristjansson JK (2001) Phylogenetic diversity analysis of subterranean hot springs in Iceland. *Appl Environ Microbiol* 67: 4242-4248.

II. **Hobel CFV**, Marteinsson VT, Hreggvidsson GO & Kristjansson JK (2004) Microbial diversity of the Grimsey high-temperature vent field on the subpolar Mid-Atlantic Ridge. *Environ Microbiol*, submitted.

III. **Hobel CFV**, Marteinsson VT, Hauksdottir S, Fridjonsson OH, Skirnisdottir S, Hreggvidsson GO & Kristjansson JK (2004) Use of low nutrient enrichments to access novel amylase genes in silent diversity of thermophiles. *World J Microbiol Biotechnol* 20: 801-809.

IV. **Hobel CFV**, Marteinsson VT, Hreggvidsson GO & Kristjansson JK (2005) Microbial ecology of intertidal hot springs by diversity analysis of 16S rRNA and chitinase genes. *Appl Environ Microbiol*, in press.

V. **Hobel CFV**, Hreggvidsson GO, Marteinsson VT, Bahrani-Mougeot F, Einarsson JM & Kristjansson JK (2004) Cloning, expression, and characterization of a highly thermostable family 18 chitinase from *Rhodothermus marinus*. *Extremophiles*, in press.
DOI: 10.1007/s00792-004-0422-3.

Reprints are published with kind permission of the journals concerned.

1. EXTREMOPHILES

1.1 Definition and terminology

Life and growth of living organisms are governed by numerous physical and chemical factors in their environment, each organism being defined and characterized according to specific parameters which are essential for its development. Man thrives on the surface of the Earth where temperatures are generally temperate, i.e., from 4 to 40°C, pH is around neutrality, or between 5 and 8.5, and where salinity, hydrostatic pressure and ionizing radiations are low. It is thus natural to consider that every ecological system that is not compatible with the growth and survival of humans as being "extreme" [21, 102]. Unlike many organisms that cannot survive outside of "normal" conditions, extremophilic microorganisms thrive optimally when one or several of these parameters are in the extreme range.

The most important factors for the growth of extremophilic microorganisms are high or low temperature, high or low pH, and high salinity. The resulting environments, based on the elevated or low conditions, are then qualified as thermophilic, psychrophilic, alkalophilic, acidophilic and halophilic, respectively. This classification encompasses several natural biotopes in which extreme environmental conditions are more prevalent than usually found in nature. Evidently, considering the high variety of biotopes on Earth, the physiological responses to the environmental extremes can be observed on a gradual scale from tolerance to absolute requirement.

1.2 Habitats of extremophiles

The most important extreme environments found in nature and typical microbial groups or species that have been observed in them are listed in Table 1.1.

Much of Earth's surface experiences low temperatures. The oceans which cover 71% of the surface have an average yearly temperature of 5°C and the ocean depths have a constant temperature between 1 and 4°C throughout the year due to the combined hydrostatic pressure and water density. Polar regions, representing 14% of the surface of the Earth, are permanently frozen or above freezing temperatures only for a short period of time. These environments, which are dominant on Earth, are

favorable to psychrophiles able to grow at any cold temperature at which water is still liquid [102].

Seawater contains on average 3% NaCl (wt/vol) and salt concentrations in water above 3% are rare in nature. The Dead Sea, the Great Salt Lake in the USA are large natural hypersaline bodies on Earth and the Blue Lagoon Spa in Iceland, or solar salterns in Western France are few examples of typical man-made saline or hypersaline ecosystems. Halotolerant and halophilic microbes grow in 1 to 20% NaCl and some can grow in saturated NaCl (>30%) [52]. Furthermore, saline lakes contain high amounts of dissolved organic matter and are regarded as such as extremely productive environments [52].

Table 1.1 Characteristics of extreme environments in which microorganisms can grow (modified from Prescott *et al.* [134]).

Stress	Environmental conditions	Ecosystem	Microorganisms observed	Refs
Temperature	-2.5 - 0°C	Polar regions	*Flavobacterium*	[154]
	0 - 4°C	Deep marine trenches	*Pseudomonas*	[140]
	65 - 95°C	Terrestrial hot springs	*Thermus spp.*	[95]
	>100 - 121°C	Submarine vents	*Thermococcus barophilus*	[108]
Salt	> 6%	Salt brines, salterns	*Halobacterium* sp.	[158]
pH	pH 3 or lower	Sulfide-rich geothermal zones	*Thiobacillus*	[95]
	pH 10 or above	Soda lakes	*Bacillus* sp.	[142]
Pressure [a]	500 - 1034 atm	Deep marine trenches	*Moritella yayanosii*	[124]
Radiation	3 - 5 Mrad	Nuclear power plants	*Deinococcus radiodurans*	[138]

[a] High pressure is in conjunction with temperature. At the bottom of the oceans, the hydrostatic pressure results in a uniform temperature of about 2 to 4°C.

Extreme pH environments restrict the occurrence of microbial life. Acid environments are widely distributed in food and soils all over the world. Large geographic zones rich in sulfur compounds are linked to volcanic activity. Hydrogen sulfide (H_2S) is released in large amounts and oxidized chemically and biologically into sulfuric acid resulting into the decrease of the pH in the soil. On the other hand, there are few examples of highly alkaline biotopes in nature. Soda lakes in West-Africa, Tibet, China or California have pH values reaching 11 to 12 but also contain high concentrations of salts. Unlike natural hypersaline lakes and seas they are depleted in Ca^{2+} and Mg^{2+} ions which disappeared at the early stages of the creation of the lake by precipitation of carbonates [52].

Extreme emissions of radiation are mostly found in man-made nuclear facilities. Very few microbes have been identified originating from the pools in which fissile material is stored. Among these, *Deinococcus radiodurans* can survive as much as 3 to 5 Mrad when a lethal dose for humans is 0.0001 Mrad. Recent research indicates that its ability to resist radioactivity may results from the presence of multiple copies of the chromosome and the ability to repair severely damaged DNA [134].

High-temperature environments are generally found associated with volcanic activity and in man-made industrial complexes. The most important biotopes are terrestrial geothermal fields, with alkaline freshwater hot springs and solfatara, and marine environments with coastal, shallow and deep hydrothermal systems. Hot environments display a complete range of pH, from acid to alkaline, depending on temperature, water availability, and gases and ion concentration [95].

2.1 Definition of thermophily

The classification of living organisms based on their relation to temperature has always been considered as the most basic element of biological systematics [93]. Three major groups were created with respect to optimum growth temperature (T_{opt}), e.g., psychrophiles that have a T_{opt} below 20°C, mesophiles that grow optimally between 20°C and 45 to 55°C and thermophiles (Figure 2.1) [102]. Traditionally, organisms with a maximal growth temperature T_{max} (i.e., above which no growth occurs) higher than 50°C have been described as thermophiles. Brock [21] suggested to set the boundary of thermophily above 60°C based on two arguments. First, temperatures below this boundary are common in nature whereas higher temperatures are mainly associated with geothermal and industrial activities. Second, certain invertebrates can survive exposures to temperatures close to 100°C but cannot grow above 50°C (Table 2.1). The thermophilic world would therefore only be prokaryotic. Interestingly, the thermophilic field covers the upper half of the currently known temperature span of life (-10 to 120°C) [86, 93].

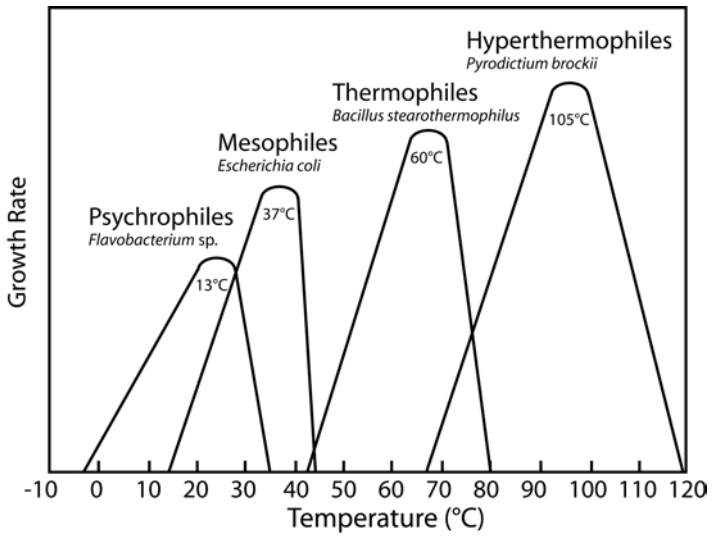

Figure 2.1 Relation of temperature and growth rates for a typical psychrophilic, mesophilic, thermophilic and hyperthermophilic microorganism. The respective optimal growth temperatures T_{opt} are indicated on the graph (modified from Madigan *et al.* [102]).

The thermophiles can be further divided into two groups due to the expansion of the upper limit of life, i.e., the discovery of the archaeon *Pyrolobus fumarii* or strain 121 that can grow optimally at temperatures of 113°C and 121°C, respectively [15, 86]. Currently, it is generally accepted to classify microbes growing optimally above 80 to 85°C as hyperthermophiles [93]. Most hyperthermophiles belong to the archaeal domain. Among *Bacteria* there are only few species that can be called hyperthermophiles, such as *Thermotoga* and *Aquifex* which have a T_{opt} in the range of 90 to 95°C [77]. It is thought that the upper limit of life has not yet been reached and may well be 140 to 150°C, assuming that the chemical reactions such as repair and re-synthesis would become unsustainable above those temperatures [29]. For the sake of simplicity, the terms thermophile and thermophilic will be used here for both physiological groups, unless designated specifically.

Table 2.1: The maximum temperature for growth of main groups of organisms (modified from Hreggvidsson & Kristjansson [72]).

Organisms	T_{max} (°C)	Organisms	T_{max} (°C)
Animals		***Eukaryotic microorganisms***	
Fish	38	Protozoa	56
Insects	45-50	Algae	55-60
Crustacean	48-50	Fungi	60-62
Plants		***Prokaryotes***	
Vascular plant	45	*Cyanobacteria*	70-72
Bryophytes	50	Green bacteria	70-72
		Bacteria	95
		Archaea	121

2.2 Habitats of thermophiles

Natural geothermal areas are widely distributed around the globe but they are primarily associated with tectonically active zones at which the movements of the Earth's crust occur. Due to this localization of geothermal heat sources, hot springs are generally restricted to few concentrated areas. Figure 2.2 describes the major terrestrial and submarine geothermal zones where thermophilic microorganisms have been isolated. The best known terrestrial sites, and most studied biologically, are Iceland, the Naples area in Italy, the Yellowstone National Park in USA, Japan, New Zealand and the Kamchatka Peninsula in Siberia.

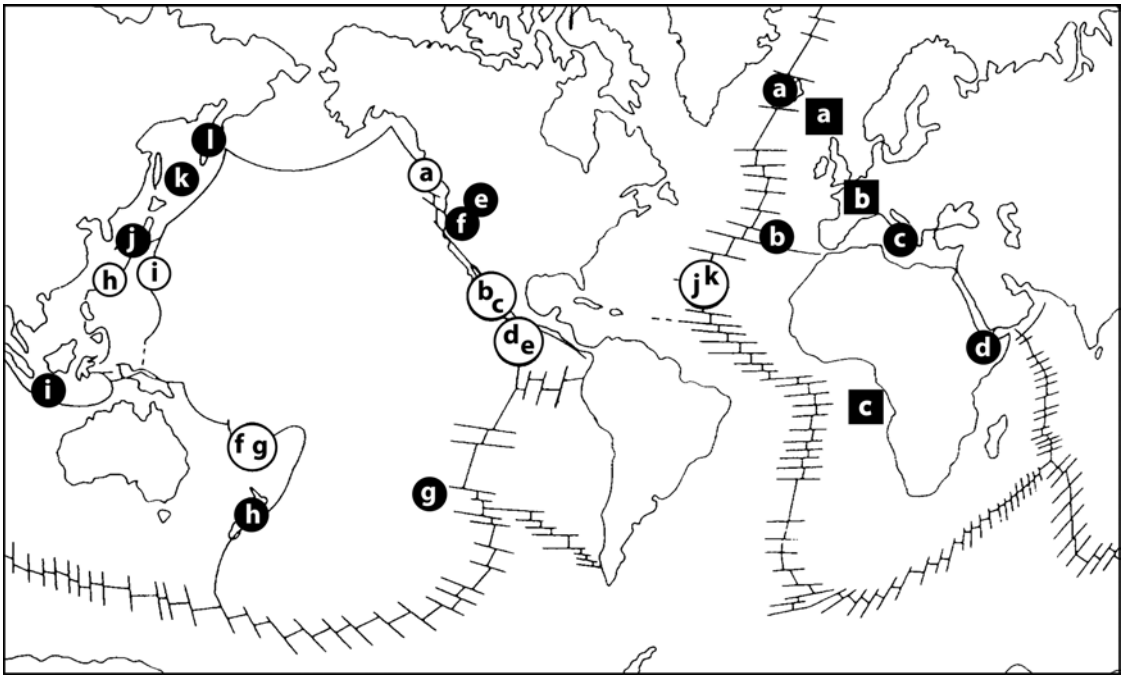

Figure 2.2 Geographical location of major geothermal sites where thermophiles and hyperthermophiles were isolated (modified from Prieur *et al.* [135]).

Legend:

○ Deep-sea hydrothermal vents: a - Juan de Fuca Ridge (45’’57’N to 46’’53’N, 129’’17’W to 130’’01’W), 1544 m to 2370 m deep; b - Guaymas basin (27’’02’N, 111’22’W), 2020 m deep; c - East Pacific Rise (EPR), 21°N (20’’5O’N, 109”06’W), 2600 m deep; d - EPR, 13°N (12’’48’N, 103”56’W), 2630 m deep; e - EPR, 11°N (10’’57’N, 103”46’W), 2505 m deep; f - North Fiji Basin (16’59’S, 173”55’W), 2000 m deep; g - Lau Basin (22’13’S, 176 36’W), 2000 m deep; h - Mid-Okinawa Trough (27”33’N, 126’’58’E), 1395 m deep; i - Ogasawara Trough (28’’N, 142’’E), 1400 m deep; j - Mid-Atlantic Ridge (MAR) (Snake Pit: 23’’22’N, 44’’57’W), 3500 m deep; k - MAR (TAG: 26’’08’N, 44’’49’W), 3630 m deep.

● Coastal marine hydrothermal vents and terrestrial hot springs: a - Iceland; b - Azores; c - Italy; d - Djibouti; e - Yellowstone National Park (USA); f - Mount Lassen National Park (USA); g - Mac Donald Seamount; h - New Zealand; i – Indonesia; j – Japan; k - Kuril Islands; l – Kamchatka.

■ Oil-field reservoir: a - East Shetland Basin (North Sea), 3000 m deep; b - Paris Basin, 2000-2500 m deep; c - West African Coast, 2500 m deep.

2.2.1 Terrestrial hot springs

Terrestrial geothermal areas can be generally divided into two classes according to the nature of the heat source and pH.

High-temperature fields

High temperature vent fields are located within the active volcanic zones and the heat source is a magma chamber at a depth of 2 to 5 km. In these areas, the water temperature reaches 150 to 350°C at the depths of 500 to 3000 m and steam and volcanic gases are emitted at the surface. Mainly, the gases consist of N_2 and CO_2 but H_2S and H_2 can make up to 10% of the total amount of gases produced. Traces of CH_4, NH_3 and CO can also be found. On the surface, H_2S is oxidized chemically and biologically first to sulfur and then to sulfuric acid which acts as the buffering agent in the hot spring environment. As a result, the pH often stabilizes at 2 to 2.5 [21]. Because of the high temperature, little liquid water comes to the surface and the hot springs are usually in the form of fumaroles and steam holes, or grey and brown mud pots resulting from the corrosion of surrounding rocks by the high concentrations of sulfuric acid [93].

Neutral to slightly alkaline sulfide-rich hot springs may also co-exist in high-temperature fields but are more rare. They appear on the periphery of the active zone and are created if water is abundant at low depths, i.e., by melting of snow or rain or with high levels of the groundwater table. Icelandic hot springs have sulfide concentrations as high as 30 mg L^{-1} and under such conditions thick bacterial mats are formed with precipitated sulfur and make spectacular bright yellow or white colors [156].

Low-temperature fields

The low temperature hot spring fields are located outside the active volcanic zones. Extinct or deep lava flows and dead magma chambers serve as heat sources and the water temperature is usually below 150°C at depths of 500 to 3000 m. Groundwater percolating through these zones warms up and returns to the surface enriched with high concentrations of dissolved minerals (silica) and gases (mainly CO_2 and little H_2S). On the surface, CO_2 is blown away and the silica precipitates resulting in an increase in pH, often stabilizing at 9 to 10. The hot springs in the low

temperature field are characterized by a general stability in both temperature, water flow and pH [93].

2.2.2 *Subterranean hot springs*

Deep subsurface environments and hot geothermal water reservoirs

Deep sediments, rocks and minerals offer environments for life that are very different from terrestrial and aquatic habitats. Water is common but generally there is little space for water and life per volume of subsurface, and microbes mainly dwell in the pores of consolidated sediments, in fractures in hard rock and in fluid inclusions. Almost all very deep environments are anoxic. Water usually contains H_2, CH_4 and CO_2 that are believed to promote primarily chemolithoautotrophic life. Gas concentrations are low and microbial growth is active but at low rates only [7].

A wide range of temperatures is measured depending on the geographical location and drilling depth of boreholes, therefore boreholes drilled with the purpose of exploring microbial life are rarely deeper than 1000 m and *in situ* temperatures are lower than 110°C [7].

In volcanic areas, however, heat from the mantle is transferred in the subsurface to aquifers which can serve as large reservoir of geothermal energy. Such energy has been harnessed in Iceland for applications such as heating and electricity production. Steam and water are collected in boreholes 1500 to 2000 m deep and temperatures are ranging between 50 to 130°C. These environments have been minimally investigated [168] yet constitute one of the few opportunities to examine the deep subsurface thermophilic microbial ecosystems (**Paper I**).

Marine and terrestrial oil reservoirs

Oil fields, resulting from the transformation of organic matter into hydrocarbons, are considered as new habitats for thermophiles. Depending on the geographic location, reservoirs are at depths of 1.5 to 4 km below the Earth surface or the sea floor, with pressures between 15 and 40 MPa and temperatures ranging from 60 to 130°C [9]. The reservoirs contain a layer of water known as connate water trapped beneath the gases and the oil, which is the habitat for a large diversity of microorganisms. The connate waters contain varying concentrations of sulfur compounds (S^0, SO_4^{2-}), metals, and n-alkanes. The water is rich in dissolved gases

such as CO_2, CO, CH_4, H_2 and H_2S and is slightly acidic due to the biological oxidation of H_2S into sulfuric acid (pH 3 to 7). Salt content varies from site to site, depending on the nature of the surrounding rock. Concentrations between 0.5% and 3% are common but several oil fields were shown to contain salt concentrations substantially higher than in sea water [10, 26].

2.2.3 *Marine hot springs*

Coastal, intertidal and shallow submarine hot springs

Several geologically and tectonically active sites around the globe offer the possibility to have microbial communities influenced by both terrestrial and marine environmental conditions. The Azores, the Aegean Sea, the bay of Naples, New Zealand and most particularly Iceland display a wide range of terrestrial hot springs as well as hot springs located directly by the sea shore or at a short distance off the coast, at various depths under the sea level.

Both high-temperature and low-temperature characteristics (cf. § 2.2.1) add to the presence of salt and influence the physical, chemical and biological features of the submarine vents. A large majority of submarine hot springs is characterized by high fluid temperatures, a feature typical of solfataric or high temperature fields, and salt concentrations close to the one found in seawater. Seawater percolates through the seafloor and is heated with the rising steam. Also, the geothermal fluid contains mainly CO_2, H_2 and H_2S which serve as energy source for chemosynthetic bacteria but the pH remains relatively stable at the vent outlet due to the high buffer effect of the surrounding seawater [92]. In Iceland, however, intertidal hot springs and hot springs at shallow-depth (0 to 20 m) release mainly neutral to alkaline freshwater originating from the mainland although minor mixing with seawater is sometimes detected. Input of organic matter and surface effect generate substantial influence on the microbial diversity with the appearance of larger amounts of photosynthetic and chemoheterotrophic microorganisms. Alkaline submarine vents and intertidal hot springs are very rare in nature but are easily accessible in Iceland. Giant submarine cones, discovered about 2 km of the coast in the North of Iceland, were shown to discharge large amounts of hot, alkaline freshwater originating from the mainland into the sea and supported purely freshwater microbial habitats [109]. Comparably, alkaline hot springs situated on the seashore in a deep fjord in the North-West region were

shown to be covered periodically at high tides and displayed a large diversity of both marine and terrestrial, psychrophilic, mesophilic and thermophilic species (**Paper IV**). Alkaline submarine hot springs are thus characterized by microbial communities which cope with extreme environmental conditions such as concurrent steep temperature and salinity gradients. Notably, bacterial strains like the highly versatile *Rhodothermus marinus* have been isolated from this site [2] (cf. §2.3.2; **Paper V**) but also from submarine solfataric hot springs located at shallow depths [126].

Deep-sea hydrothermal vents

The ocean floor is continually being renewed at seafloor spreading centers. In these tectonically active areas, seawater infiltrates through cracks in the ocean floor, possibly down to several kilometers. It is then chemically altered by the interaction of heat and the surrounding rock within the crust (Figure 2.3). The low-density fluid thereby modified is forced back to the ocean floor due to convection. The fluid is very hot (250 to 400°C) and remains liquid because of the hydrostatic pressure. It is also acidic, rich in metals and dissolved reduced compounds, such as iron, CH_4 and H_2S [137]. If the hydrothermal fluid mixes with cold seawater just before emission, its temperature is moderate to hot (5 to 100°C). If no mixing occurs, the fluid vents at temperatures of 350°C and above. The contact with cold oxygenated deep-ocean water results in the precipitation of the minerals and this causes mineral structures or chimneys which can reach up to 10 to 15 m in height [136]. The deep sea vents are also known as black smokers, due to the constant discharge of precipitated minerals in seawater that take the aspect of thick, black clouds. Hydrothermal vent fields are generally tens of meters in diameters and may be interconnected via subterranean conduits. Most of the known hydrothermal vent sites in the ocean are located along global mid-ocean ridge systems, back-arc basins and hot spots in the deep-sea [137] (Figure 2.2) and can be found at depths greater than 3500 m and as shallow as 400 m (**Paper II**).

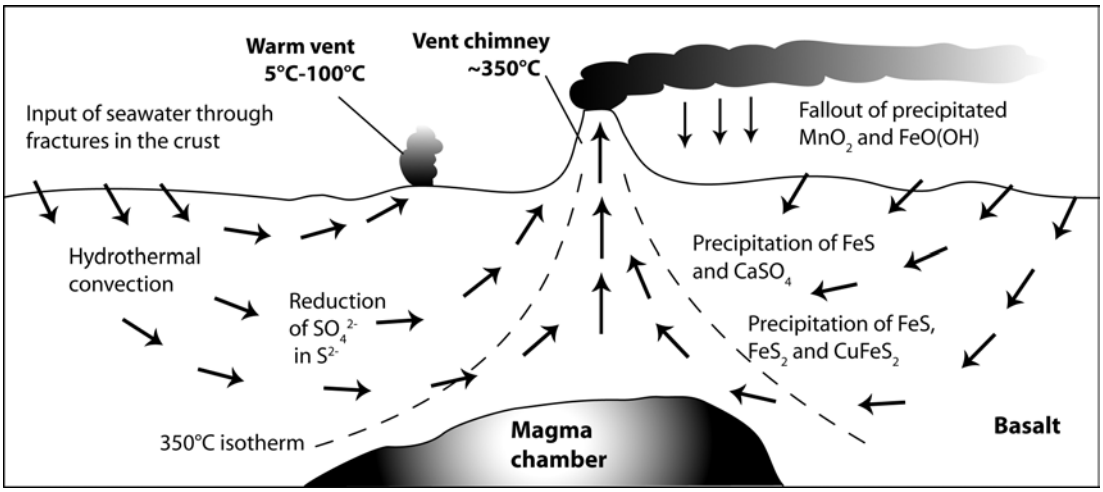

Figure 2.3 Schematic diagram showing the plumbing system, the morphology and the major chemical compounds occurring at warm vents and black smokers. Ocean water is convectively circulated through mid-ocean ridge fractures and emerges as a heated solute-enriched fluid with various temperatures. At warm vents, the hot hydrothermal fluid is cooled by cold (2-3°C) seawater permeating the sediments. In black smokers, hot hydrothermal fluid reaches the seafloor directly and the metal sulfides precipitate in contact of the seawater creating large chimney structures (modified from Madigan *et al.* [102]).

2.2.4 Other geothermal habitats

Constant hot habitats other than geothermal are very few in nature. Solar-heated ponds and biologically-heated composts, hay, litter or manure may cause high temperature but these are very transient ecosystem that will be mostly inhabited by rapidly-growing spore formers. Man-made, constant hot environments have also been created. These include hot water pipelines (cf. §2.2.2), burning coal refuse piles, wastes from treatment plants or industrial processes in the food or chemical industry. The microbial communities in man-made environments are given constant interests and several well-known thermophiles, such as several *Thermus* species [196], as well as unique site-specific microbes like *Thermoplasma acidophilum* [34], have been primarily isolated from those systems [93]. For instance, *Thermus scotoductus*, a pigment-producing rod was isolated first from hot tap water in Iceland [94], and was shown later on to be endemic to a large majority of Icelandic neutral to alkaline hot springs [67, 156].

2.3 Phylogeny, taxonomy, and physiology of thermophiles

Microbes have been traditionally classified according to characters such as morphology and physiology, but it has not been possible to determine the evolutionary relationships between the different microbial groups. The entry of natural phylogeny into microbial systematics based on nucleic acid sequence data enabled the classification of microorganisms on evolutionary terms as well as the clarification of the phylogenetic lineages between them [129]. Consequently, unique phylogenetic groups could be determined at each taxonomic level, i.e., species, genus, family, etc., based on comparative analysis via DNA sequence databases. Moreover, molecular phylogenetics resulted in a major scientific breakthrough in the late 1970's by Carl Woese's discovery of *Archaea* as a third domain of Life in addition to *Eukarya* and *Bacteria* [198, 199] (Figure 2.4). The prokaryotic *Archaea* and *Bacteria* domains display common phenotypic characteristics but have also very specific features which are listed in Table 2.2.

Table 2.2: Comparison of *Bacteria* and *Archaea* (modified from Kristjansson & Stetter [93]).

Character	*Bacteria*	*Archaea*
Cell wall components	Murein	Pseudomurein proteins, polysaccharides
Membrane lipids	Glycerol fatty acid esters	Glycerol isopranyl ethers
Square and flat structures	-	+
Endospores	+	-
tRNA -common arm- contains	Ribothymidine	Pseudo-uridine or 1-methylpseudo-uridine
Methyonyl initiator tRNA formylated	+	-
Introns in genes	-	+
Eukaryotic RNA polymerase	-	+
Special coenzymes	-	+
Max growth temperature	95°C	121°C
Complete photosynthesis	+	-
Methanogenesis	-	+
Calvin cycle used in $C0_2$ fixation	+	-

The phylogeny of extremophiles has been very closely linked to questions on the origin and early evolution of life on Earth. Many thermophiles and hyperthermophiles seem to have very "old" (or deep) lineages in the so-called Tree of Life (Figure 2.4). This is particularly evident among the archaeal thermophiles. It is also very striking that 11 of the 23 major cultivated bacterial phyla contain thermophilic representatives. Four of these, i.e., the *Aquificae*, the *Thermotogae*, the green-non-sulfur bacteria including the *Thermales* and the *Thermodesulfobacter*

group comprise all of the deepest phylogenetic branches in the bacterial tree. This supports the hypothesis that thermophiles and hyperthermophiles represent the most ancient forms of life now present on Earth [102].

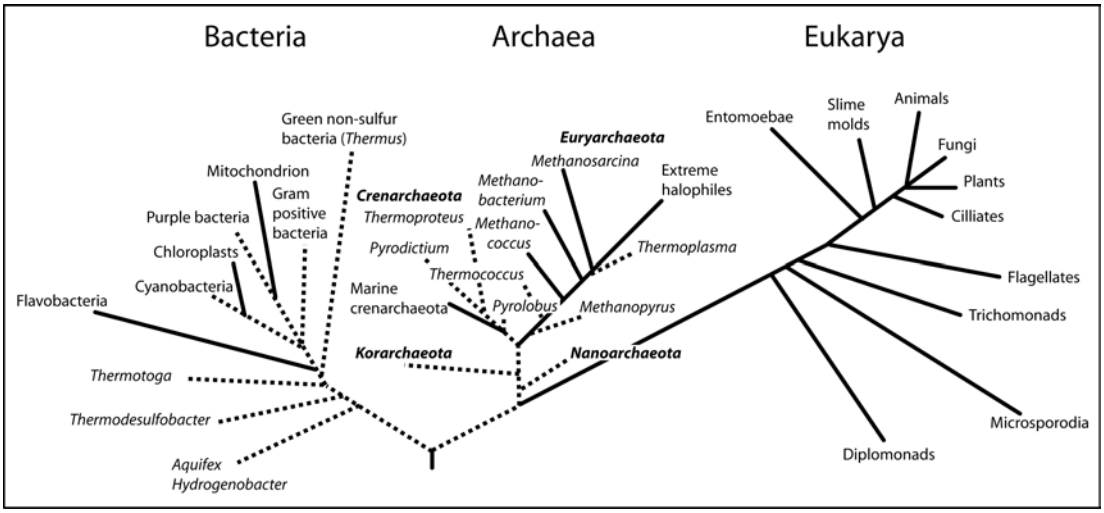

Figure 2.4 Rooted universal phylogenetic tree as determined by comparative analysis of ribosomal genes sequences. The data supports the discrimination of three domains, two of which contain prokaryotic representatives (*Bacteria* and *Archaea*). The root represents the position of a suspected universal ancestor of all cells. In dashed lines are indicated phylogenetic groups which are exclusively thermophilic or contain few thermophilic representatives (modified from Madigan *et al.* [102]).

The current, validly accepted species of bacterial thermophiles that grow at ≥65°C are diverse and distributed across 52 genera while the thermophilic *Archaea* cover 33 genera (Table 2.3, in appendix to this chapter). On a physiological point of view, aerobic and anaerobic *Archaea* distinguish themselves by their optimal pH, the former being very acidophilic. Both aerobic and anaerobic *Bacteria* are neutrophilic or slightly acidophilic (Figure 2.5).

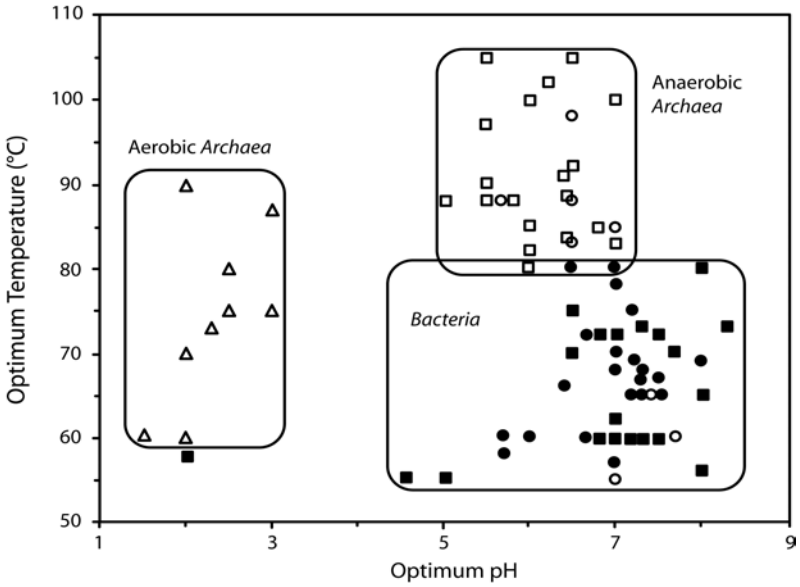

Figure 2.5 The distribution profile of thermophilic prokaryotes according to their respective optimum temperature and pH. *Archaea* are marked with open symbols and *Bacteria* with closed symbols. Legend: ○, methanogens; ρ, aerobic *Archaea*; □, anaerobic *Archaea*; ■, aerobic *Bacteria*; ●, anaerobic *Bacteria* (from Kristjansson & Stetter [93]).

2.3.1 Archaea

As a group the *Archaea* (from the Greek *archaios*, ancient or primitive) are quite diverse, both in morphology and physiology. They can stain either Gram positive or negative and may be spherical, rod-shaped, spiral, lobed, plate-shaped, irregularly shaped, or pleiomorphic. Some are single cells, whereas other form filaments or aggregates. They range in diameter from 0.1 to 15 μm, and some filaments can grow up to 200 μm in length. Multiplication may be by binary fission, budding, or other mechanisms yet unknown [134]. All known *Archaea* belong to four taxonomic phyla, *Crenarchaeota, Euryarchaeota* and *Korarchaeota* and the very recently described *Nanoarchaeota* [68, 73]; three major metabolic and/or physiological groups are found: methanogens, extreme halophiles and extreme thermophiles (Figures 2.4 and 2.5).

The *Euryarchaeota* (Greek *eurus*, wide, and *archaios*) are named as such because they encompass the greatest phenotypic diversity among known culturable species with the halophiles, some thermoacidophiles and some hyperthermophiles, and the methanogens which is the most dominant group in this phylum [44]. The extreme halophiles or halobacteria contain currently 14 known genera. They are

aerobic chemoheterotrophs with a respiratory metabolism and require complex nutrients for growth [134]. No thermophilic halophiles have been isolated so far. The methanogens are strict anaerobes that obtain energy by converting CO_2, H_2 or simple methyl compounds into CH_4. Five orders compose the methanogens, the *Methanobacteriales*, the *Methanopyrales*, the *Methanococcales*, the *Methanomicrobiales* and the *Methanosarcinales*. The three first orders are exclusively hyperthermophilic (*Methanopyrales*) or regroup both mesophilic and thermophilic genera, and but no thermophilic *Methanomicrobiales* or *Methanosarcinales* are known as yet (Table 2.3). All five orders reduce CO_2 as carbon source for the synthesis of CH_4, except for the *Methanosarcinales* which also use methyl group-containing compounds in addition to CO_2 [17]. The thermoacidophiles are represented by the order *Thermoplasmatales*, which contains only two genera, *Thermoplasma* and *Picrophilus* (Table 2.3). *Thermoplasma* grows on refuse piles of coal mines, under extreme acidic and moderately hot conditions (pH 1-2, 55-65°C). Interestingly, at 59°C, *Thermoplasma* takes the form of an irregular filament, whereas at lower temperatures it is spherical [134]. *Picrophilus* is the most extreme archaeon, in terms of affinity to acid conditions. The organism will grow only below pH 3.5, has a pH_{opt} of 0.7 and still grows at pH 0 [17, 152]. The exclusively hyperthermophilic representatives of the *Euryarchaeota* are the *Thermococcae*, divided into three orders, *Archaeoglobales*, *Thermococcales*, and *Methanopyrales* (Table 2.3). The *Thermococcales* are anaerobic fermentatives and have their T_{opt} ranging from 88 to 100°C [134]. They release gaseous H_2 as a by-product of fermentation, which is then detoxified by reducing elementary sulfur to sulfide. The *Archaeoglobales* can extract electrons from a variety of electron donors (e.g. H_2, lactate, glucose) and reduce sulfate, sulfite or thiosulfate to sulfide. The third order, *Methanopyrales* occupies the deepest and most ancient branch of the *Euryarchaeota* and its presence is exclusively restricted to deep hydrothermal ecosystems.

The *Crenarchaeota* (Greek *crene*, spring or fountain, and *archaios*) are thought to resemble the ancestor of the *Archaea*, and almost all well-characterized species are thermophilic or hyperthermophilic. The most extreme example is *Pyrodictium*, which has a minimum growth temperature of 82°C, a T_{opt} of 95°C and a T_{max} above 110°C [163]. There are three major branches among the *Crenarchaeota*. The *Thermoproteales* are Gram-negative anaerobic to facultative, hyperthermophilic rods

which can grow chemolithoautotrophically by reducing sulfur to hydrogen sulfide. The *Sulfolobales* are coccoid-shaped thermoacidophiles with optimum temperature around 70 to 80°C and optimum pH around 2 to 3. The order *Desulfurococcales* contains coccoid- or disc-shaped anaerobic, facultative anaerobic and aerobic hyperthermophiles. Both autotrophic and/or heterotrophic metabolism is found, at neutral pH and temperatures between 85 to 106°C. Nevertheless, mesophilic or psychrophilic species have been identified by molecular analysis of pelagic mud, marine waters and soil, constituting new undefined lineages among the *Crenarchaeota*. This emphasizes that a much wider diversity of physiological and metabolic types occur within this phylum [74, 175, 188]. A recent study has estimated that mesophilic *Crenarchaeota* account for nearly 20% of the total picoplankton in oceanic biomass worldwide [85].

The *Korarchaeota* phylum is thought to be the most ancient one among the *Archaea*. It was discovered originally by the use of molecular techniques in hot springs biomass samples originating from the Yellowstone Park [8]. Yet, *Korarchaeota* have eluded cultivation and only uncultured clones are known from this phylum so far [156] (**Paper II**).

Finally, the recently discovered *Nanoarchaeota* phylum is represented by one species only, the symbiont *Nanoarchaeum equitans*. The cells grow exclusively attached to the surface of the specific hyperthermophilic archaeal host *Ignicoccus* and its distribution is so far unknown. *N. equitans* harbors the smallest archaeal genome, which is only 0.5 megabases in size, but is undetectable by traditional molecular analysis strategies using universal archaeal probes. Its high growth temperature and anaerobic mode of life correlates with probable early environmental conditions which suggest that the *Nanoarchaeota* are possible primitive form of microbial life [68, 73].

2.3.2 Bacteria

Thermophilic bacteria can be found in most bacterial metabolic groups. Yet, it is clear from Table 2.3 that many of the *Bacteria* are not adapted to grow at extremely high temperatures. Most of the bacterial thermophilic representatives have an optimum temperature below 75°C, with *Thermotogae* and *Aquificae* the only one with a T_{opt} above 85°C. Similarly, the pH range for bacterial thermophiles is relatively

narrow between 5 and 9, with few exceptions like *Hydrogenobaculum* spp. or some *Bacillus* species.

The families *Aquificae*, *Thermotogae*, *Thermodesulfobacteria*, *Thermomicrobia*, and the *Thermales* among the *Deinococcus-Thermus* are exclusively thermophilic whereas all other families listed in Table 2.3 contain both mesophilic and thermophilic representatives.

Species belonging to the *Aquificae* are mainly obligately chemolithoautotrophic, thermophilic and aerobic bacteria, using H_2 or reduced sulfur compounds as energy sources. They belong to one of the deepest branching orders of the *Bacteria*, and accordingly are thought to be the most direct bacterial descendant of the ancestor of all cells. The *Thermotogae*, also branching deeply in the Tree of Life, are all anaerobic and fermentative species. They are Gram negative cells with a distinct outer sheath-like envelope of "toga" [17]. The phylum *Thermomicrobia* consists of a single known representative and is distantly related to the *Chloroflexi* but branches deeply in the major reference trees. Cells are obligately aerobic and grow only on complex nutrients. The *Thermodesulfobacteria* also contain only one representative. Cells are rod-shaped, strictly anaerobic, chemoheterotrophic exhibiting a dissimilatory sulfate-reducing metabolism. The order *Thermales* contains the most known thermophilic bacterium, *Thermus aquaticus*, isolated from the Yellowstone Park in the USA [20]. The *Thermales* are predominantly aerobic and heterotrophic although species able to grow mixotrophically have been described [115, 116, 157]. The *Thermus* genus is not restricted to natural hydrothermal areas and can be found in various environments such as hot tap water, thermally polluted streams or compost piles [17]. Several *Thermus* species are also endemic to marine hydrothermal vents and shallow marine hot springs [92, 106, 107] (**Papers II and IV**). Other genera compose the order *Thermales*, such as *Meiothermus*, and the recently appended *Marinithermus*, *Oceanithermus* and *Vulcanithermus* all three isolated from deep marine vents [115, 116, 151].

The thermophilic Gram-positives are scattered among 22 genera, and 20 thereof are exclusively thermophilic (Table 2.3). They are in a large majority anaerobic with a fermentative metabolism and offer an untapped source of carbohydrate-degrading enzymes such as amylases or chitinases for industrial applications (**Papers III and IV**).

The *Cyanobacteria* are oxygenic photosynthetic prokaryotes using H_2O as the electron acceptor in photosynthesis. The *Cyanobacteria* are in a large majority mesophilic with the few exceptions of moderate thermophiles such as *Fischerella* or *Oscillatoria*, and the thermophilic *Synechococcus* which grows optimally above 55°C.

The genera *Rhodothermus* and *Thermonema* are the unique thermophilic representatives among the *Cytophaga* / *Flexibacter* / *Bacteroides* group. *Rhodothermus* is an extremely versatile bacterium and has been used extensively as a source of new enzymes [2, 54, 125, 132, 160, 177] (**Paper V**). It grows optimally above 70°C and requires minimum salt concentrations of 1% for growth [2]. *Thermonema* is a moderate halophile and thermophile with a T_{opt} at 60 to 65°C. The genus has been isolated exclusively in New Zealand, Italy and Iceland. The three isolates however grow at different salt concentrations [109, 174].

In the very large phylum of the purple bacteria or *Proteobacteria*, few thermophilic genera scattered among the α, β, γ, δ and ε-subdivisions co-exist with both psychrophilic and mesophilic representatives. Cells grow either autotrophically or mixotrophically and all excepted the β-proteobacterial genera are growing strictly under anaerobic conditions. The most thermophilic ones are the hydrogen-oxidizing *Hydrogenophilus* and the sulfur-oxidizing *Thermothrix*, as well as the sulfur-reducers *Desulfurella* and *Thermodesulfobacterium*. Representatives of the ε-subdivision are mostly moderate thermophiles with T_{opt} varying around 55°C but growth is sometimes observed above 60°C to 65°C in the laboratory [1, 114, 117].

Table 2.3 List of thermophilic *Archaea* and *Bacteria* with a T_{opt} above 65°C, and their respective physiological characteristics (modified from Kristjansson & Stetter [93], Marteinsson [105], Kristjansson *et al.* [96], Boone *et al.* [17], and Hreggvidsson & Kristjansson [72]). Legend: H, heterotrophic; A, autotrophic; photo; photosynthesis; aer, aerobic; an, anaerobic.

Organisms	T_{max}	Physiology	pH_{opt}	Aerobic / Anaerobic
ARCHAEA				
Crenarchaeota				
Archaeoglobales				
Archaeoglobus	80-82	H/A	6-7	An
Ferroglobus	85	A	7	An
Desulfurococcales				
Aeropyrum	90-95	H	7	An
Desulfurococcus	85-92	H	6	An

Hyperthermus	95-106	H	7	An
Ignicoccus	90	A	5.8-6	An
Pyrodictium	97-105	A	5.5-6	An
Pyrolobus	106	A	5.5	Aer/An
Staphylothermus	92-95	H	6-6.5	An
Stetteria	95	H	6	An
Sulfophobococcus	85	H	7	An
Thermodiscus	90	H	5.5	An
Thermosphaera	85	H	6.5-7.2	An
Euryarchaeota				
Methanobacteriales				
Methanothermus	80-88	A	6.5	An
Methanococcales				
Methanocaldococcus	80-85	A	5.2-7.6	An
Methanothermococcus	60-65	A	5.1-7.5	An
Methanotorris	88	A	5.7	An
Methanopyrales				
Methanopyrus	98	A	6.5	An
Sulfolobales				
Acidianus	70-90	H/A	1.5-2.5	Aer/An
Metallosphaera	75	H/A	1-4.5	Aer
Sulfurococcus	60-75	H/A	2-2.6	Aer
Stygiolobus	80	A	2.5-3	An
Sulfolobus	65-85	H/A	2-4.5	Aer
Sulfurisphaera	84	H	2	Aer/An
Thermococcales				
Thermococcus	75-88	H	6-8	An
Pyrococcus	100	H	7	An
Thermoproteales				
Caldivirga	85	H	3.7-4.2	An
Pyrobaculum	100-102	H/A	6-7	Aer/An
Thermocladium	75	H	4.2	Aer/An
Thermophilum	85-90	H	5	An
Thermoproteus	90	H/A	5-6.5	Aer/An
Thermoplasmatales				
Thermoplasma	60	H	2	Aer/An
Picrophilus	60	H/A	0-3.5	Aer
BACTERIA				
Aquificae				
Aquifex	95	A	6.8	Aer
Hydrogenobacter	77-85	A/H	6-7.5	Aer
Hydrogenobaculum	70	A	3-4	Aer
Hydrogenothermus	80	A	5-7	Aer
Persephonella	75-80	A	6	An/Aer
Thermocrinis	89	A	7-8.5	Aer
Gram positives				
Alicyclobacillus	70	H	2	Aer
Ammonifex	77	A	7.5	An
Bacillus	65-80	H	4.5-9	Aer
Caldicellulosiruptor	80	H	7-7.5	An

Caloramator	68-80	H	7-7.5	An
Caloranaerobacter	65	H	7	An
Carboxydobrachium	80	H	6.8-7.1	An
Carboxydothermus	78	A	6.8-7	An
Clostridium	66-91	H	5-7	An
Desulfotomaculum	65-85	H/A	7-7.4	An
Dictyoglomus	80-86	H	7	An
Geobacillus	68-78	H	6.2-7	Aer
Moorella	65-70	H/A	5.7-6.8	An
Saccharococcus	65	H	7-7.5	Aer
Thermaerobacter	80	H	7-7.5	Aer
Thermanaerobacter	75-85	H/A	5.8-7.5	An
Thermanaerovibrio	65	H/A	7.3	An
Thermoanaerobacterium	74	H/A	6-6.2	An
Thermobrachium	75	H	8.2	An
Thermohalobacter	70	H	7	An
Thermosyntropha	70	H	8.1-8.9	An
Thermoterrabacterium	66-75	H	5.2-7	An
Cyanobacteria				
Synechococcus	73	photo	8	An
Cytophaga* / *Flexibacter* / *Bacteroidetes				
Rhodothermus	72	H	6.5	Aer
Thermonema	65-70	H	6.5-7.5	Aer
Chloroflexi				
Chloroflexus	70	H / photo	8	Aer
Themomicrobia				
Thermomicrobium	80	H	8.3	Aer
Thermooleophilum	70	H	6.8-7.3	Aer
Deinococcus-Thermus				
Marinithermus	67	H	7	Aer
Meiothermus	65-70	H	7-8.5	Aer
Oceanithermus	60	H/A	7.5	Aer/An
Thermus	80-85	H	7.3-7.5	Aer
Vulcanithermus	70	H/A	6.7	Aer/An
Nitrospira				
Thermodesulfovibrio	70	H	6.8-7	An
β-*Proteobacteria*				
Hydrogenophilus	80-86	H/A	6.8-7.5	Aer
Thermothrix	65	A	6.5	Aer
δ-*Proteobacteria*				
Desulfurella	70-77	H/A	6.4-7.2	An
Thermodesulforhabdus	74	H	6.9	An
ε-*Proteobacteria*				
Caminibacter	55-70	H/A	5.5-7	An
Thermodesulfobacteria				
Thermodesulfobacterium	74-85	H/A	6.5-7.5	An
Thermotogae				
Fervidobacterium	80	H	6.5-7.2	An
Marinitoga	65-70	H/A	6-7	An
Thermotoga	72-90	H	6.5-7.4	An

Thermosipho	70-80	H	6.5-7.5	An
Not classified				
Coprothermobacter	70	H	7.5	An
Desulfurobacterium	75	A	6-6.5	An

3. ACCESSING DIVERSITY AND METHODS TO DESCRIBE MICROBIAL ECOSYSTEMS

Life on Earth depends upon microbial processes. Microbes grow across a broad range of temperature, salinity, oxygen levels, and pH. They have an enormous impact and role in our daily lives, including everything from maintaining the biosphere to sustaining our lifestyle. Because of this, diversity analysis is important in order (i) to increase the knowledge of the diversity of genetic resources in microbial communities, (ii) to understand patterns in the relative distribution of microorganisms, (iii) to increase the knowledge of the functional role of the diversity, and therefore (iv) to understand the regulation of biodiversity, as well as assessing to what extent the function and sustainability of an ecosystem depend on maintaining a certain level of genetic diversity [128].

3.1 Cultivation-dependent methods

3.1.1 Traditional taxonomy and phenetic classification

The establishment of techniques to achieve pure cultures by Robert Koch as well as the invention of the double-sided dish by Richard Petri in the late 19th century marked important turning points for the science of microbiology by supplying critically needed tools for development of the fields of bacterial taxonomy, genetics and several related disciplines [102].

Microbial diversity has been traditionally studied by the assessment of the occurrence of particular species, i.e., with a specific phenotype, obtained by enrichments and pure isolates, followed by the comprehensive identification of their phenotypes. A medium and a set of incubation conditions would be used that are selective for specific organisms with a phenotype of interest, and are counter-selective for the undesired organisms. Subsequently, the enrichments would be re-conducted using the previous culture as a new inoculum, preferably on agar plates by series of streaking and for the isolation of single colonies, or in liquid media by dilution to extinction and microscopic observations. Table 3.1 lists some typical cultivation conditions used to obtain specific phenotypes.

Table 3.1 Examples of enrichment culture methods for prokaryotes (from Madigan *et al.* [102]).

Light phototrophic bacteria; C source: CO_2		**Organism enriched**
Aerobic conditions		
N_2 as N source		*Cyanobacteria*
NO_3^- as N source		Thermophilic *Cyanobacteria*
Anaerobic conditions		
H_2, or organic acids; N_2 as sole N source		Non-sulfur *Proteobacteria*
H_2S as e^- donor		*Proteobacteria* and green sulfur bacteria
Chemolithotrophic bacteria; C source: CO_2 (medium lacks organic C)		**Organism enriched**
Aerobic incubation		
e^- donor	*e^- acceptor*	
NH_4^+	O_2	Nitrosifying bacteria
H_2	O_2	Hydrogen-oxidizers
H_2S, S^0, $S_2O_3^{2-}$	O_2	*Thiobacillus* spp.
Anaerobic incubation		
S^0, $S_2O_3^{2-}$	NO_3^-	*Thiobacillus denitrificans*
H_2	NO_3^-	*Paracoccus denitrificans*
Chemoorganotrophic bacteria; C source: organic compounds		**Organism enriched**
Aerobic incubation (respiration)		
e^- donor and N source	*e^- acceptor*	
Lactate + NH_4^+	O_2	*Pseudomonas fluorescens*
Starch + NH_4^+	O_2	*Bacillus* spp.
Cellulose + NH_4^+	O_2	*Cytophaga* spp.
Anaerobic incubation (anaerobic respiration)		
e^- donor	*e^- acceptor*	
Organic acids	KNO_3	Denitrifying *Pseudomonas* spp.
Yeast extracts	KNO_3	Denitrifying *Bacillus* spp.
Acetate, Ethanol	S^0	*Desulfuromonas*
Anaerobic incubation (fermentation)		
e^- donor and N source		
Starch + NH_4^+		*Clostridium* spp.
Glucose or lactose + NH_4^+		*Escherichia*, other fermentative organisms
Glucose + Yeast extract (pH 5)		Lactic acid bacteria

Many different phenotypic and genotypic features have been used to classify bacteria. The term polyphasic taxonomy was suggested in the 1970's aiming at the integration of different kinds of consensual data and information on microorganisms (phenotype, genotype, phylogeny), and eventually at the determination of typical patterns specific to the isolates [187]. Table 3.2 lists the various phenotypic and genotypic information that are generally used as polyphasic taxonomical tools. Previous to the advent of molecular phylogenetics, the taxonomic patterns determined by polyphasic taxonomy in the laboratory served for the classification of microbes, known as the phenetic classification. Yet even if highly informative, the phenetic

classification often resulted into anecdotal sorting of the species and was therefore unreliable [38]. It's only after standardizing the molecular methods as a routine technique that clear phylogenetic and thus taxonomic classes were obtained.

Table 3.2 Overview of the most common phenotypic and genotypic methods used in polyphasic taxonomy (from Vandamme *et al.* [187]). Legend: MLEE, Multilocus Enzyme Electrophoretic Pattern, EP, Electrophoretic Pattern, RFLP, restriction fragment length polymorphism; RAPD, random amplified polymorphic DNA; DGGE, denaturing gradient gel electrophoresis; TGGE, temperature gradient gel electrophoresis.

Methods			
Phenotypic methods		**Phenotypic methods (continued)**	
Expressed features	Color of colonies	Chemotaxon. markers	Cellular fatty acids
	Colony morphology		Cell wall compounds
	Gram staining	Protein analysis	MLEE
	Cell morphology		EP of total cellular proteins
	Cell size	**Genotypic methods**	
	Sporulation	Total DNA	G+C% mol content
	Mobility (flagella)		Total DNA RFLP
	Inclusion bodies		DNA:DNA hybridization
	Temp. range and optimum		Genome sequencing
	pH range and optimum	DNA fragments	RAPD
	Salinity tolerance/needs		DGGE
	Antibiotic sensitivity		TGGE
	Aerobicity		DNA probes
	Single carbon sources		16S rRNA sequencing
	Metabolism of carbon hydrates		23S rRNA sequencing
	Nitrate reductase		ITS sequencing
	Oxidase		
	Catalase		

3.1.2 New rational enrichments strategies

Although enrichments and cultivation techniques are necessary tools to attain phenotypic characteristics, they only enable the access to few species from a microbial ecosystem. This is particularly true in extremely complex communities such as soil or mud which are thought to contain from 1,000 to 10,000 different microbial species [178]. It has been frequently reported that direct microscopic count exceeds viable cell count by several orders of magnitude. Consequently, it is believed that less than 1% of existing microbes are currently culturable, although obvious differences are seen according to the origin of the sample (Table 3.3) [3].

Table 3.3 Culturability determined as a percentage of culturable bacteria in comparison with total cell count (from Amann *et al.* [3]).

Habitat	Culturability (%)
Seawater	0.001-0.1
Freshwater	0.25
Mesotrophic lake	0.1-1
Unpolluted estuarine waters	0.1-3
Activated sludge	1-15
Sediments	0.25
Soil	0.3

This failure to recover larger amount of isolates by cultivation is directly linked to the intrinsic selectivity of any given media and cultivation conditions. Above all is the inability to reproduce or mimic *in situ* growth conditions existing in the habitat from which the samples were obtained. Nevertheless, the suggestion that *in vitro* growth studies of organisms are passé or irrelevant is unjustified. Recently there has been an increasing interest in developing rational, alternative cultivation techniques in order to obtain higher yields of recovery as well as to cultivate more specialized cells [99]. These new cultivation techniques usually apply to two parameters: (i) chemical modifications of the culture medium, and (ii) physical alterations of the growth conditions.

In vitro enrichments usually involve highly artificial growth conditions in terms of nutrients availability (electrons donors and acceptors), concentrations and growth co-factors. A dominant part of the microbial world is thought to grow under chemolithoautotrophic conditions but a majority thereof has not been accessed so far [129]. Enrichment strategies using diversified electron donors and acceptors, in particular metal ions or pollutants have made possible to isolate new microorganisms with very peculiar phenotypes (for review, see Leadbetter [99]). In a similar manner, organotrophic conditions are often in such excess of nutrients that growth may be inhibited due to metabolic stress, as well as promote the selection of few opportunistic cells such as *Thermus* or *Bacillus* species for instance. The use of low to very low nutrient concentrations can be a useful strategy to prevent the growth of those opportunistic heterotrophs and therefore increase the diversity in enrichments (**Papers II, III and IV**). Furthermore it can be a unique and elegant way to cultivate some microorganisms [139]. Recreating *in situ* conditions in the laboratory can also be obtained by the use of additional untypical growth co-factors or signalling molecules

[23], and/or by the use of (un)treated water from the sampling sites (**Paper III**). Evidently, the most suitable experiments should be conducted, if possible, directly in the environment itself by *in situ* enrichments (**Paper II**) or substrate colonization (**Papers II and IV**). Thereby, only one parameter would be modified, e.g., presence of a nutrient-rich matrix or nature of the artificial support for colonization, and all required growth factors and unknown natural *in situ* conditions would still be present [105].

Physical alterations of the enrichments encompass the development and design of atypical growth conditions. Huber and colleagues have succeeded in the cultivation of the hyperthermophilic chemolithotroph *Thermocrinis ruber* by the design of a continuous flow chamber, where the constant flow of oxygenated fluid was apparently necessary to trigger the growth of the bacterium [76]. A new approach involving the encapsulation of single cells in a gel and incubation in a constant flow of nutrient-depleted medium resulted in a high-throughput cultivation of highly diverse microbes [201]. However, low-technology methods may also bring substantial results. A simple way to influence enrichments is by treating the inoculum previously to the inoculation of the media. Simple strategies generally use dilution, filtration, density-gradient centrifugation, cell-sorting using flow cytometry, micromanipulations and optical tweezers, chemical treatment with ethanol or ammonia or even sterilization by autoclaving [80, 102]. As one of such alternative procedure, we have performed cyclic temperature disturbances (from 4 to 80°C) of thermophilic cultures using nutrient rich media but our results showed that the cultures were mainly dominated by *Thermus* species (Hreggvidsson & Hobel, unpublished data). Nevertheless, encouraging results could be obtained by conducting similar experiments with a more varied set of cultivation conditions (hypoxia or anoxia) and inoculum. Various experiments regarding the supply of specific light conditions were tried in our laboratory as well as by other groups. For instance, Glaeser and Overmann [50] have demonstrated that particular phototrophic bacteria from intertidal sandy sediments could be isolated by supplying one specific wavelength of light.

Even though the impediments to cultivate new organisms represent a substantial bottleneck, significant advances have been made in our understanding of microbes as well as ecosystems mechanisms in nature. However, one very important feature still remains to be cleared. The persistence of bacteria in an ecosystem in

response to environmental stress, e.g., changes in nutrient availability, temperature, salinity, etc., is to a great part determined by their ability to endure the stress. Yet, investigations of bacterial survival in natural environments has indicated that some organisms lose culturability on appropriate media under certain conditions but still exhibit signs of metabolic activity and thus viability, a physiologic state known as "viable but non-culturable" (VNBC) (for review see McDougald *et al.* [111]). This state, which appears to be specific to each species, is induced by one or several stress factors such as temperature variations, salt supply or depletion, cycles of freeze/thaw, addition or presence of antibiotics, etc. Various protocols have been developed to resuscitate microbes that have entered this state of non-culturability but similarly they appear to species-specific. Some have suggested the hypothesis that the VNBC might be a programmed physiological response to stress but since the physiological and molecular basis for entry and exit from the non-culturable state are still obscure no genetic determinants that might regulate this response have been discovered. As a result, no universal strategies exist so far to prevent non-culturability [111].

3.2 Molecular methods

Until the mid-1960's, microbiologists were simply distinguishing prokaryotes from eukaryotes. Things improved with the development of robust phenetic tools, and in particular the use of molecular biology which dramatically boosted the field of microbial taxonomy. The situation changed in 1965 when Zuckerhandl and Pauling suggested that molecules could be used as documents for evolutionary history. Evolutionary relationships could now be deduced from sequence differences observed between homologous macromolecules, enabling universal comparison of homologous macromolecular features from virtually all cellular life-forms [38].

Scientists, still today, do not agree about the ideal molecule to use for evaluating the true evolutionary relationships between organisms but they agree that there are some necessary features that an evolutionary chronometer must have. The molecule should be universally distributed across all organisms under study and it must be functionally homologous to all organisms. Furthermore, it is important that lateral gene transfer may not have occurred, as that will represent wrong evolutionary relationships. In addition, the sequences of the molecules must change at a rate that commensurates with the evolutionary distances that are measured [102].

3.2.1 16S ribosomal RNA gene analysis: origins and methodology

The breakthrough formulation was reached by Carl Woese during the 1970's who, by comparison of ribosomal RNA sequences, established a molecular sequence-based phylogenetic tree that would be used to relate all organisms and reconstruct the history of Life [129] (Figure 2.4). Because of the likely antiquity of the protein-synthesizing process, ribosomal RNA turned out to be an excellent evolutionary chronometer. Ribosomal RNA is an ancient molecule, functionally constant, universally distributed and moderately well conserved across broad phylogenetic distances [102]. Moreover, there is no evidence of lateral gene transfer of rRNA genes between different species and therefore rRNA genes can bring true information regarding evolutionary relationships [129].

There are 3 types of RNAs found in the microbial ribosomes, 5S rRNA, 16S rRNA and 23S rRNA [102]. The first attempts to characterize microbes by studying rRNA began by extracting the 5S rRNA molecules directly from the cells. However, the information content in the approximately 120-nucleotide long molecule is relatively small and was abandoned to the benefit of the 1,500 nt long 16S rRNA gene, and to a lesser extent to the 3,000 nt long 23S rRNA. The 16S rRNA molecule has several advantages. Some regions of the gene are universally conserved and suitable for phylogenetic studies of distantly related organisms. Other regions are semi-conserved and are more useful for the analysis of phylogenetic relationship between phyla and families, and variable and hyper-variable regions in the 16S rRNA enable us to discriminate between organisms belonging to the same genus or even between species, although not between strain within the same species [3]. The length of the gene is convenient so PCR and sequencing are easy. Furthermore, the ends of the gene are highly conserved across all bacterial and archaeal domains, therefore almost the entire gene can be amplified by PCR. This last characteristic is the foundation of the cultivation-independent approach [49].

The characterization of an organism in terms of its phylotype requires only a gene sequence and not a functioning cell. By extension, 16S rRNA genes or gene fragments can be selectively amplified by PCR from complex DNA mixtures obtained directly from the environment. The method circumvents then the need to cultivate microorganisms in order to identify them. Figure 3.1 shows the sequential process to characterize an environmental sample by comparative rRNA analysis. A set of

oligonucleotide probes universal for all *Bacteria* or *Archaea* or targeted at specific taxonomic levels (phylum, families, genus) are used to amplify specifically the genes, which are then shotgun-cloned. The resulting libraries should contain only defined molecules that can be rapidly and easily sequenced from known priming sites. The subsequent phylogenetic analyses are based on the comparison of the ribosomal sequences with previously identified ones, available in large databases accessible worldwide [103]. As a final output, a phylogenetic tree is created, gathering all information regarding the relationship between the newly obtained sequences and the reference sequences. Ideally, the clone library covers the entire population of rRNA molecules that were amplified by PCR from the complex biomass DNA and therefore a snapshot of the microbial diversity can be easily obtained. This technology for accessing previously inaccessible microbes can be applied to virtually any environment. The impact on sensitive environment is also minimized since theoretically only very small samples of DNA are sufficient for PCR.

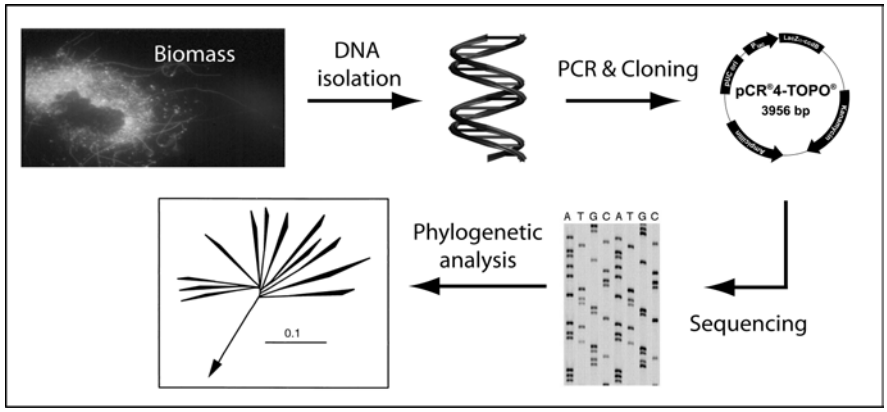

Figure 3.1 Flowchart showing the characterization of the microbial diversity by 16S rRNA analysis

3.2.2 Implementation of the 16S rRNA method and use of other phylogenetic markers

A library of environmental rRNA sequences contains evidently a high number of clones in order to cover the largest part of the rRNA molecule population. Unfortunately, at the time the direct molecular analysis of the microbial diversity was developed, sequencing technology was still time-consuming and only allowed the processing of a limited number of sequences simultaneously. Over the years, various methods have been developed to alleviate the use of large clone libraries and adequately increase the discrimination between closely related clones.

One relatively easy technique is too screen all clones according to their unambiguous restriction pattern on agarose gel electrophoresis [128]. The amplified rDNA restriction analysis (ARDRA) method produces multiple bands from the clones by incubation of the amplified amplicons with one or several DNA restriction enzymes and allows the comparison of the different clones between each other. This method can be quite tedious when considering large number of clones since the restriction patterns are sometimes difficult to read properly on the agarose gels. Nevertheless, the use of ARDRA allows to reduce several-fold the final number of clones that will be sequenced and brings a considerable gain in time. This method however is given less interests due to the latest development in high-throughput clone libraries management systems as well as sequencing techniques which made possible to process more clones in a relatively short period of time.

Terminal restriction fragment length polymorphism (T-RFLP) works in a similar manner as ARDRA, and has been shown to be highly effective at discriminating microbial communities in a range of environments [14]. It involves tagging one end of PCR amplicons through the use of fluorescent molecule attached to a primer. The amplified product is then cut with a restriction enzyme. Terminal restricted fragments (TRF) are separated by electrophoresis and visualized by excitation of the fluorochrome. T-RFLP analysis provides also quantitative data about each TRF detected, including size in base-pairs and intensity of fluorescence (peak heights) [22]. TRF sizes can be compared to a database of theoretical TRFs derived from sequence information. This strategy becomes very valuable when comparing the relative abundance of previously identified species in a given environment, for instance in following environmental disturbances [14].

Independently from cloning, one method allows to examine at a glance the genetic diversity of complex microbial populations. The procedure is based on electrophoresis of PCR-amplified 16S rRNA gene fragments in polyacrylamide gels containing a linearly increasing gradient of denaturant such as urea. In denaturing gradient gel electrophoresis (DGGE), DNA fragments of the same length but with different base-pair sequences have a different melting behavior, and will stop migrating at different positions in the gel [119]. Also, the 16S rRNA gene is coupled to a guanine and cytosine-rich clamp that will play the role of an anchor and therefore stabilize the fragment inside the gel (Figure 3.2). DGGE allows in particular to

compare microbial communities with each other since each band on the gel will have a migration behavior typical for a specific species [83, 119, 118]. More information can furthermore be obtained by hybridization analysis of the DGGE patterns with taxon specific oligonucleotides. Also, bands can be gel-purified and sequenced to finally attribute a phylogenetic assessment of the diversity. This method has many advantages since cloning is not a prerequisite and therefore shows great utility when investigating the dynamics of microbial communities.

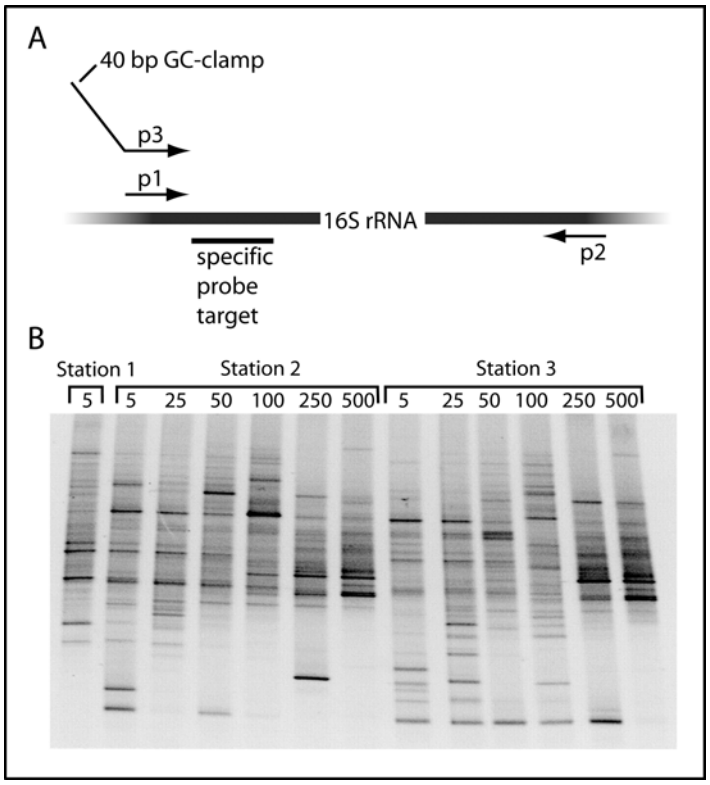

Figure 3.2 (A) Schematic diagram of the methodology of the DGGE strategy. Primers p1 and p2 amplify a fragment of the 16S rRNA gene. Primers p3 and p1 amplify the same region but p3 incorporates a 40-bp long GC clamp at its 5' end. Specific probes can be used later on to identify targeted microbial species after performing a blotting on hybridization membranes (modified from Muyzer *et al.* [118]). **(B)** Example of DGGE fingerprints of microbial population assemblages in the South-West Mediterranean Sea at different times and at different depths (5 to 500 m). Comparison of the band patterns enables to follow the evolution of the species composition of each sample (reproduced from Diez *et al.* [41]).

3.2.3 Pitfalls of the 16S rRNA analysis strategy

Cultivation-independent molecular phylogenetic approaches, like any other methodology, have their own specific pitfalls and biases [38]. This approach is based

on many physical, chemical and biological steps and each of them may be a source of bias that can lead to a distorted view of the "real world" [191]. But only by being aware of these pitfalls can some of them be minimized and other even eliminated.

Sample collection and treatment

The sample collection and the extraction of the DNA thereof are the most crucial steps for the subsequent analysis, yet often ignored as a source of problems and pitfalls. The sample should be representative of the ecosystem under study and sampling may require additional and extensive efforts in difficult and/or extreme habitats such as marine environments for instance. Contamination by the surrounding water, soil, mud, air, or at the laboratory should be prevented in particular by taking special care during transport, storage and processing [3]. The most likely source of contaminants in the laboratory are salts and buffers, lysozyme and proteinase K (used for the chemical lysis of cells) [173]. Cell lysis marks a second critical step in the PCR-based approach. Insufficient or disproportionate disruption of cells in the sample will most likely bias the view of the composition of microbial diversity since DNA or RNA which is not released from the cells will not contribute to the final analysis of diversity. Gram-positives generally need more rigorous disruption than Gram-negative bacteria due to the presence of a thick cell-wall but treatment with lysozyme prevents a possible DNA extraction bias. Extensive physical and chemical treatment may however lead to highly fragmented DNA of Gram-negative bacteria, which can be a source of chimeric artifacts in the PCR reaction [191].

Polymerase Chain Reaction (PCR)

The PCR amplification step is known to introduce several other biases, irrespective of the gene target (i) inhibition of PCR by co-extracted contaminants and production of unreliable PCR products, (ii) differential amplification, and (iii) formation of chimeric artifactual PCR products. Humic substances or polysaccharides co-extracted with nucleic acids strongly inhibit DNA enzymes. The choice of alternative DNA extraction techniques by using polysaccharide-binding molecules such as cetyl-trimethylammonium bromide (CTAB) or polyvinyl-polypyrrolidone (PVPP) may help in the removal of those contaminants [148]. Similarly, additives in the PCR mix like bovine serum albumin (BSA) can also help in preventing the inhibitory effects of the contaminants. Furthermore, the fidelity of the PCR amplification varies depending on the particular DNA polymerase enzyme which is used. Discrepancies

due to misincorporated nucleotides during the amplification process can lead to phylogenetically irrelevant sequence variations. The use of DNA polymerases having a proof-reading capacity (3'-5' exonuclease activity) may prevent misincorporation and it is therefore important to choose carefully the type of enzyme in a PCR-based diversity analysis [191].

Amplified DNA can only reflect the quantitative abundance of species in a sample if the amplification efficiency is the same for all molecules. Yet this is seldom the case since environmental DNA samples constitute evidently a highly heterogeneous mixture with high discrepancies in respective gene copy number and DNA concentrations. Therefore, since all molecules are often not equally accessible to primer hybridization, primer-template hybrids do not form with equal efficiencies and the extension efficiency of the DNA polymerase is not the same for all templates. These problems can also result from the fact that universal primers employed for the amplification of the rRNA genes often contain a degeneracy which may influence the hybridization of the primers to the different templates [191]. The choice of primers, i.e., degenerate- vs. taxon-specific, and number of cycles of replication have a major influence on the final outcome. Suboptimal binding of the primer will result in less efficient amplification of the respective DNA. Moreover, it is thought that the original sample template is mostly amplified during the initial 5-6 cycles of the PCR reaction and that the amplification of the PCR fragments produced at the early stages occurs in the following cycles only. This implies that only sequences with a good primer match and high copy number will be selectively amplified [33]. Additionally, because of the intrinsic biochemical mechanism of the PCR, significant changes in the composition of 16S rRNA libraries can be observed when performing dilutions of the template DNA. Consequently, bacteria that are rare in the ecosystem (e.g., with an abundance below 0.1%) will hardly be detected.

Chimeric PCR products are generated as DNA strands compete with specific primers during the annealing step [191], and as mentioned above, are generally associated with damaged or highly fragmented DNA. This results in the formation of molecules composed of two or more different sequences, with varying sizes. Computer programs such as Check-Chimera from the Ribosomal Database Project [103] or Bellerophon [78] offer practical solutions to recognize chimeric molecules in a pool of 16S rRNA sequences but the clear identification of chimera can sometimes be

difficult to achieve, in particular when obtaining 16S rRNA genes that are very distantly related to any known reference sequence (**Paper V**). Increasing interest is now given to determine the frequency and number of chimeric sequences that have been published as putative novel organisms [81].

Creation of clone libraries of 16S rRNA genes

For almost all analyses of microbial ecosystems amplified 16S rRNA genes have to be separated prior to subsequent sequencing since the PCR generates a heterogeneous mixture of sequences. Cloning in *Escherichia coli* is the most widely used method to separate the PCR products. However the size of the clone library will influence the validity of the results. As we have seen here above, methods such as ARDRA allow discriminating clones from each other and therefore pre-classifying the clones. The advent of high-throughput clone picking and sequencing has alleviated the use of ARDRA but the choice of number of clones to select remains at the foundation of the 16S rRNA method. Considering that the 16S rRNA genes from all organisms are amplified with the same efficiency and assuming that all PCR products are ideally inserted in the cloning vectors, the number of clones sequenced should enable us to cover the largest number of different 16S rRNA sequences, regardless of their abundance or rarity. Extreme ecosystems that are characterized by high dominance of particular organisms require a smaller sampling size to determine the main elements of their community structure than do less extreme environments. However, if the goal is to make an exhaustive mapping of the species composition in these extreme ecosystems, the sampling size needed would be much larger than currently used [156]. Soil or sediments biotopes are known to harbor over 1,000 species [178] and analysis of these environments would require in theory as many clones to detect at least one clone of each species. Several mathematical methods often derived from population genetics studies allow us to evaluate the validity of the coverage of the library [127], such as the calculation of the Coverage or the establishment of rarefaction curves [82, 156] (Figure 3.3).

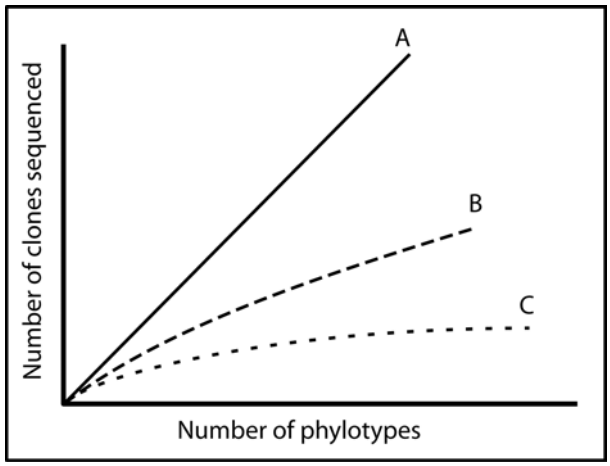

Figure 3.3 Rarefaction curves: case studies of the microbial diversity coverage in 16S rRNA clone libraries from three different types of microbiota. Case A: hot spring sediments; case B: thick filamentous *Chloroflexus* mat and case C: sulfide-rich hot spring. Case A represents the most extreme example where the very high microbial diversity cannot be completely covered unless collecting and sequencing a very high number of clones (up to 10,000 clones). Further cloning is *per se* irrelevant as the extent of the diversity is in theory beyond reach. Case C displays a relatively simple microbial community with high dominance of few species. The slope of the curve has a tangential aspect suggesting that further sequencing will bring little additional information. Case B is in-between. A high diversity has been observed and most of the dominant species may have been identified already. Further sequencing will bring additional information but the slope will reach soon a plateau suggesting that the overall diversity has been covered by the clone library (inspired by results from Skirnisdottir *et al.* [156]).

3.2.4 General considerations, new trends and challenges in molecular diversity analysis

Today's alternative strategies to access microbial diversity

As described above, new cultivation and enrichment strategies have given us an insight into a new portion of the unknown microbial world, but every method has its advantages and disadvantages. For instance, the DGGE method makes possible to follow the evolution of a community at various time points but do not give any information *per se* about the phylogenetic nature of each band observed on the gels. The 16S rRNA gene analysis gives little clues about the spatial and temporal distribution of cells. Vice-versa, the use of fluorescent probes *in situ*, a method known a fluorescent *in situ* hybridization (FISH), reveals the relative abundance of targeted microorganisms but is limited because it does not bring significant information regarding the exact phylogenetic relationships of the identified cells. Moreover, obtaining pure strains by new cultivation methods brings still only a fraction of the

total biodiversity. It is mainly by the adoption of a multidisciplinary approach that globally meaningful information can be obtained. Figure 3.4 shows the various techniques that may be applied to fully investigate microbial diversity in the environment.

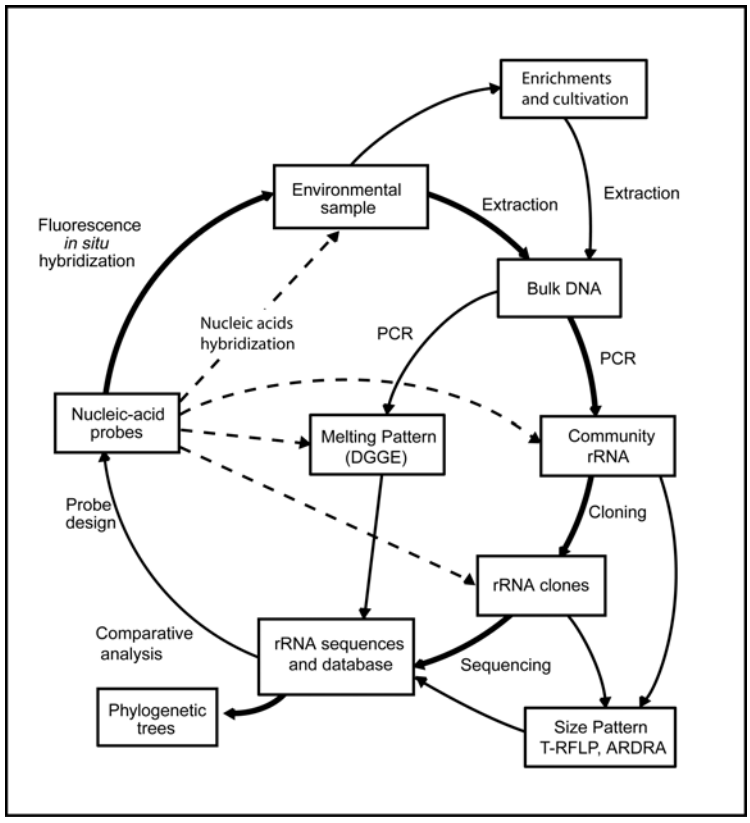

Figure 3.4 Multidisciplinary approach to analyze and characterize the microbial diversity. The bold line shows a typical strategy, from sampling to the construction of phylogenetic dendrograms. The dotted line shows the various steps at which fluorescent probes can be used to identify specific microbial types (modified from Hugenholtz [80]).

Validity of the 16S rRNA strategy

The ultimate goal of the PCR-based analysis of a complex microbiota is the retrieval of sequence information, which allows the determination of the microbial diversity of both cultured and uncultured microorganisms by comparative 16S rRNA sequence analysis. The ease and simplicity of the direct molecular analysis method has resulted in an enormous flow of information from a variety of microbial environments (18,000 entries in GenBank in 1999 [146]). Figure 3.5 shows the evolution of the bacterial Tree of Life since the time of its genesis by Woese and colleagues.

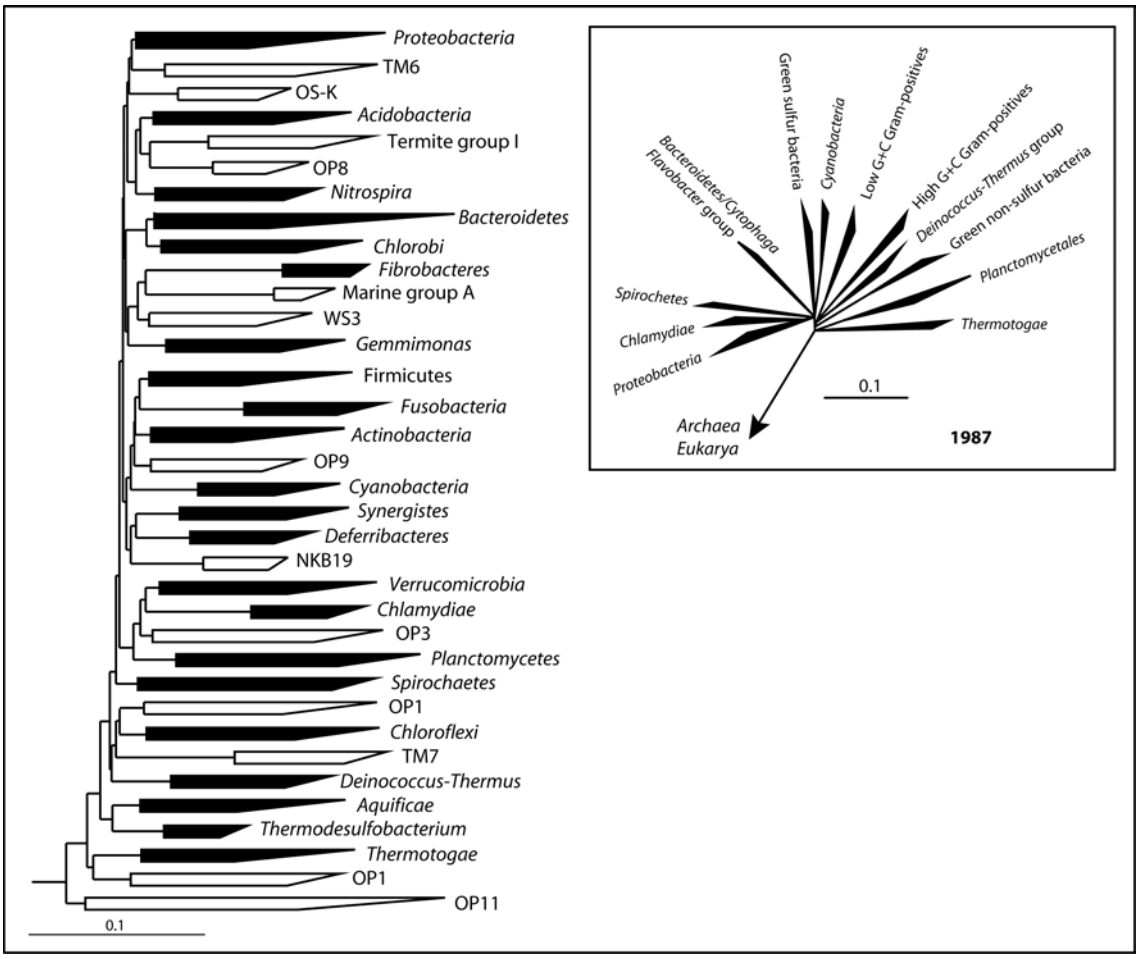

Figure 3.5 Evolutionary distance dendrograms of the bacterial diversity derived from comparative analysis of 16S rRNA gene sequences in 1987 and 1997. Twelve phyla were originally found by analysis of the major known bacterial species. With the rapid evolution of the cultivation-independent 16S rRNA gene analysis, close to 40 bacterial phyla are known today [80]. Phyla with cultivated representatives are in black and phyla known only from environmental sequences are in white; because they are not formally recognized as taxonomic groups, they are usually named after the first clones found from within the group. The scale bar represents 0.1 changes per nucleotide (modified from Hugenholtz *et al.* [79] and Hugenholtz [80]).

This information has importance *per se* because it brings scientific facts about the occurrence of particular types of microbes as well as enables us to discover the largely unknown, "so-far-uncultivated" silent diversity. Nevertheless, nucleic acid databases are flooded with complete or often only partial genes of 16S ribosomal RNAs that either lack proper phylogenetic classification due to their putative novelty, or that are redundant with previously published data. This problem may be inherently a part of the method, since the sequence variation between 16S rRNA is also the basis of its usefulness, i.e., the discrimination of "species" based solely on nucleic acid

similarities between those sequences. Today, microbiologists are using an artificial threshold of DNA similarity to ascribe sequences to a particular phylogenetic position [161] and similarity values of 97% to 98% are most often used to set the limit between different species [109, 156] (**Papers I, II, III and IV**). However, as more sequence information becomes available, we see that the resolving power of 16S rRNA sequences is limited when closely related organisms are inspected. As a result we cannot define bacterial species solely on the basis of sequence similarity of rRNAs or other genes [146]. There are many examples of different species with identical or nearly identical 16S rRNA sequences. A microheterogeneity of the 16S rRNA genes within a single species can also be seen, and in exceptional cases we find single organisms with two or more 16S rRNA genes with relatively high sequence divergence (for review see Rossello-Mora & Amann [146]). Polyphasic taxonomical tools help to assess the exact phylogenetic relationships, in particular DNA:DNA hybridization which appears to be the golden rule to attain a clear distinction between strains [161] but this applies only to pure cultures. The 23S ribosomal RNA genes would support stronger phylogenetic information but their use remains only secondary due to the intrinsic technical bottleneck of sequencing entirely 3,000 nucleotides in one stretch. Sequence polymorphism in the 16S-23S intergenic spacer region (IGS) can also be used to differentiate closely related microbes. This region comprises the signals for transcription initiation and termination. Because different regions of the rRNA repeat units evolve at very different rates, phylogenetic analyses of these rRNA sequences provide an opportunity to investigate organismal evolution [65]. ATPase subunits, elongation factors, RNA polymerases and the σ factor *rpoB* have been assayed as alternatives to ribosomal genes. The valid phylogenetic information deduced from these genes generally confirmed the one obtained by 16S rRNA gene analysis [146]. However, the extent of lateral gene transfer between organisms is unknown and may well affect the information value of such genes. Moreover, reliable detection of lateral gene transfer by comparison of gene trees is complicated by gene duplication and loss, and the various methods for detecting lateral gene transfer are not particularly consistent [80]. In any case, the 16S rRNA strategy remains a very powerful tool to explore the uncultured prokaryotic world and provide preliminary phylogenetic affiliations for unknown and poorly classified microbes.

New trends and challenges in environmental microbiology

The recent advances in the technology of high-throughput sequencing have resulted in an unprecedented innovation in traditional microbiology. Approximately 60 genomes have been completely sequenced and published, and hundreds are in the process of being sequenced. Complete microbial genomes sequencing has made it possible to identify and characterize *in silico*, *in vitro* and *in situ* genes present in a species, including novel metabolic pathways, gene regulatory elements, genes of unknown function and genes for pathogenesis, virulence and drug resistance. This information also provides insights into the evolution of genes and species [179].

All the genomes of a microbiota can be considered to be one large community genome, described as the metagenome [145]. Environmental genetic libraries in vector that carry large segments of DNA (40-150 kb) can now be constructed without the need for cultivation. Relevant clones can be sequenced using shotgun or chromosome-walking and comparatively analyzed [80], or screened for the expression of gene products [144]. On one hand, this direct access to the genomes brings simultaneously information regarding the phylogeny of species but also information regarding their genotype and putatively their phenotype. On the other hand, a more and more defined picture of the whole microbial community rises, including a better understanding of the relationships among the members of the community. Shotgun sequencing allows a relatively unbiased identification of genes and thus a more comprehensive approach to environmental biology than previously possible [188]. Perhaps the most impressive application of this methodology was performed by Venter and colleagues who employed a whole-genome shotgun-sequencing approach to study microbial populations collected from the Sargasso Sea [188]. A total of 1.045 billion base pairs of non-redundant sequences were generated, annotated and analyzed. The data were estimated to be derived from at least 1800 genomic species, including 142 novel bacterial phylotypes [80]. However, if this grandiose project was achieved mostly due to very powerful sequencing capacities those results remain only uncustomary to the technical abilities of common microbiology laboratories. Moreover, this particular strategy is not devoid of major pitfalls and drawbacks [144] which still divide the scientific community [32, 47, 144]. Given the existing technological bottlenecks of sample management relevant to such projects, the choice of the proper microbial community to study is a first priority. Tyson *et al.* have shown that by shotgun sequencing the

metagenome from an acid mine drainage they were able to reconstruct 3 microbial genomes almost entirely. Preliminary study by 16S rRNA gene analysis pointed out that only those 3 species were present in the community and therefore only a comparable effort to those involved for the sequencing of isolated genomes was needed [182].

Assembly of complete genomes from a complex mixture or genomic clones of more than 10 to more than 1,000 different organisms in an environmental sample appears today unrealistic. Even with high-throughput sequencing and gene assembly programs, the interpretation of sequences of cloned community DNA can be complex because communities are comprised of different species that are present in different amounts. Without normalization, most of the sequencing is done on the few most dominant microorganisms, and sequence coverage of any non-dominant genome will be up to several orders of magnitude lower. Moreover, the study of the Sargasso Sea demonstrated that even huge sequencing efforts do not lead to a considerable coverage of the whole genomic diversity in complex populations [87]. Nevertheless, metagenomic analyses provide significant functional information through genomic sequence and expression of traits, although they would need other methods to link specific functions with the groups (species, microcommunities, etc.) responsible for them [179]. Environmental functional gene arrays [203], microsensor measurements, or microradioautography in combination with fluorescent *in situ* hybridization are promising techniques that would enable to shed the light on community functioning and dynamics (for review, see Torsvik & Ovreas [179]).

The greatest challenge in microbiology today is to get access to the tempo and mode of microbial genome and species evolution [39]. The population genetics of bacteria is still in its infancy [144]. The pool of genes shared by prokaryotic strains decreases dramatically when the taxonomic distance increases, even within a single species. For instance, the comparison of genome of three *E. coli* subspecies showed that only about 40% of the total number of genes identified is present in all three. This suggests that the genetic diversity in prokaryotes is mostly due to the different repertoires of genes that each cell carries, regardless of the taxonomic scale. Consequently, the relative influence and importance of vertical inheritance versus lateral gene transfer on microbial genome evolution and speciation remains an open question [39]. Environmental genomics can now provide valuable data on naturally

occurring genomic structure and variation, genetic drift, and lateral gene transfer in natural microbial populations. The natural microbial world provides some ideal case studies for interpreting process and dynamics in microbial genome evolution. Conversely, genomics and metagenomics provide the tools to study the complexity of form, function and process that underpins microbial population structure dynamics and evolution [39]. This however will not occur unless a complete change in structures and fundamentals of bioinformatics occur: management of genomic databases containing complete genomes but also only genome segments; communication and interaction between those databases (for review see Galperin [47]) computational tools to compare and interrelate the datasets and accordingly biological tests for the validity of the theories resulting from the dataset analyses (for an excellent review, see Delong [39]).

4.1 Thermostable enzymes from Extremophiles

Enzymes are versatile tools for sustainable development in a variety of industries because they have important environmental benefits. Enzymes are biodegradable, show improved use of raw materials and decreased amounts of waste products. Extremophilic microorganisms are adapted to live at high or low temperatures, at high or low pH, or other extreme conditions. Metabolic processes and specific biological functions of these microorganisms are, therefore, usually mediated by enzymes that function under these extreme conditions, and these conditions have often to be met in industrial enzymatic process.

Enzymes that are still active at elevated temperatures are beneficial for a large variety of processes because running processes at high temperatures have many advantages. The increase of temperature has a significant influence on the bioavailability and solubility of organic compounds. It is also associated with a decrease in viscosity and increase in diffusion of organic compounds and therefore increased reaction rates. Besides, the risk of contamination by mesophilic microorganisms is reduced. Heat-tolerant enzymes, sometimes called thermozymes, are currently the most investigated of all extreme enzymes (i) because they serve as excellent models for understanding protein stability and (ii) due to their valuable potential in industrial processes [45, 123].

4.1.1 Cellular and molecular basis of protein thermostability

The mechanisms allowing organisms to cope with high temperatures are diverse, including specific adaptations on the levels of (i) the structure and functions of macromolecules and cellular functions, (ii) physiology and metabolism, and (iii) regulation of gene expression and maintenance. Under extreme conditions of temperature, pH and pressure, amino acids can be damaged irreversibly by deamination, β-elimination, hydrolysis, Maillard reaction, oxydations, and disulphide interchange. As a result, at temperatures above the boiling point of water, the half-life of some amino acids is significantly shorter than the generation time of hyperthermophiles. Furthermore, the hydrolysis of peptide bonds sets theoretical

limits to protein stability by directly affecting the integrity of the polypeptide chain at its most basic level. Nevertheless, a large number of enzymes from thermophiles and hyperthermophiles are stable and active at temperatures considerably exceeding the upper growth limit of the producing organism [162].

Considering the cost of protein synthesis and amino acid turnover, keeping proteins in the native state is favorable compared with *de novo* synthesis. The need to stabilize proteins becomes especially important when physiological conditions reach temperatures close to the upper growth limit of an organism or in the temperature range where thermal decomposition of amino acids and protein backbones become significant. Cellular strategies have been developed to enhance the *in vivo* stability of proteins beyond their intrinsic stability by additional factors such as chaperone proteins, protein repair enzymes and high concentrations of small stabilizing solutes (see Sterner & Liebl [162] for an extensive review). Chaperones play essential roles in prokaryotic as well as eukaryotic cells. They are involved in the proper folding of newly synthesized protein and of polypeptides translocated through membranes and can also prevent misfolding and aid in the refolding of denatured proteins. Chaperones prevent the aggregation of proteins that unfold after exposure of the cell and cell compounds to high temperature but additionally help at solubilizing and refolding those proteins that aggregate in the advent of damages by heat exposure [162].

The molecular basis of protein thermostability have been extensively reviewed [41, 45, 162, 190] and only the most common features will be discussed upon here. Much experimental data on thermophilic and hyperthermophilic enzymes have been accumulated in recent years leading to the conclusion that no single intrinsic molecular mechanism can explain the increased thermoresistance of thermophilic enzymes compared to their mesophilic counterparts. With the exception of phylogenetic variations, the major trait differentiating mesophilic, thermophilic and hyperthermophilic enzymes is the temperature range in which they are stable and active. The sequence of homologous enzymes is highly similar between mesophilic and thermophilic types (40 to 85%), their three-dimensional structures are generally superposable and the catalytic mechanisms are also identical [190]. The forces that keep thermophilic and hyperthermophilic enzymes functional are apparently similar to those in mesophilic proteins, and, apart from the help of external cellular compounds, the proteins in thermophilic microorganisms mostly take advantages of variations in

their sequences and secondary structure to deal with the higher temperatures. Analysis and comparison of the amino acid content in partial and complete genomes of mesophilic and thermophilic microorganisms showed that, on average, thermophilic protein contain fewer residues than mesophilic proteins, with a larger proportion of charged residues (Arg, Lys, His, Asp, Glu) and smaller proportion of uncharged polar residues (Ser, Thr, Gln, Asn, Cys). These basic properties bring resistance mechanisms to deleterious changes in protein structure, which would otherwise have a direct negative effect on catalysis due to deformation and/or denaturation. The main characteristics of thermophilic proteins seem to be the following: (i) the hydrophobic core of the protein is strengthened by increasing the number of hydrophobic amino acids with branched side chains; (ii) surface loops are eliminated, reducing the overall conformational entropy caused by the mobility of those loops; (iii) thermophilic enzymes tend to be shorter, thus with a smaller heat exchange capacity; (iv) the structural plasticity is reduced by increasing and optimizing electrostatic interactions by the formation of ion-pair networks at the surface of the proteins [189]. The preferential use of charged residues at the expense of uncharged residues leads to the stabilization of electrostatic interactions in and around the active site, ligand-binging site, or metal-binding site and hence help maintaining the integrity of these sites at elevated temperatures (for review, see Kumar & Nussinov [97]).

4.1.2 Applications of thermophilic and hyperthermophilic enzymes

The synthesis of polymer intermediates, pharmaceuticals, specialty chemicals and agrochemicals is often hampered by expensive processes that suffer from low selectivity and undesired byproducts. Mesophilic enzymes are often not well suited for the harsh reaction conditions required in industrial processes because of the lack of enzyme stability. For this reason, the use of biocatalysts in organic reactions represent only a small fraction of the potential industrial market until now [40]. However, the discovery of thermostable enzymes has resulted in a revolution and re-birth of a large number of industrial processes because of the overall inherent stability of the enzymes, and implying higher applications potentials. The applications of the so-called thermozymes have been extensively reviewed [37, 40, 42, 45, 70, 89, 123, 200]. Table 4.1 summarizes some of the major applications of thermostable enzymes for commercial applications. While the most widely used thermostable enzymes are the amylases in

the starch industry (cf. § 4.3), a number of other applications are in various stages of development: in the food-related industry for the production of amino acids, in the petroleum, chemical, and pulp and paper industries for the elimination of sulfur containing pollutants, in the fine chemical industry by the production of chiral products, or for the modification of DNA (for instance, *Taq* DNA polymerase from *Thermus aquaticus*) and RNA in molecular biology [53].

Table 4.1 Application of enzymes and endogenous compounds from thermophiles and hyperthermophiles (modified from Haki & Rakshit [53]).

Enzyme	Temperature range (°C)	Bioconversions	Applications
α-Amylase (bacterial)	90-100	Starch → dextrose syrup	Starch hydrolysis, brewing, baking, detergents
α-Amylase (fungal)	50-60	Starch → dextrose syrups	Production of maltose
Pullulanase	50-60	Starch → dextrose syrups	Production of glucose syrups
Xylanase	45-105	Craft pulp → xylan+lignin	Pulp and paper industry
Chitinase	65-75	Chitin → chitobiose Chitin → *N*-acetylglucosamine *N*-acetylglucosamine → glucosamine Chitin → chitosan	Food, cosmetics, Pharma-ceuticals, Agrochemicals
Cellulase	45-95	Cellulose → glucose	Cellulose hydrolysis, polymer degradation in detergents
Protease	65-85	Protein → amino acids and peptides	Baking, brewing, detergents, leather industry
Lipase	30-70	Fat removal, hydrolysis, inter-esterification, alcholysis, aminolysis	Dairy, oleo-chemical, detergent, pulp, pharma-ceuticals, cosmetics and leather industry
Alcohol dehydrogenase	80-95	Alcohols → ketones	Fine chemicals
DNA enzymes	90-95	DNA amplification (Polymerase) DNA/RNA ligation (DNA/RNA ligase) DNA restriction (Restr. enzymes) Phosphorylation [a]	Molecular biology, Genetic engineering

[a] Polynucleotide kinase from a thermophilic phage of *Rhodothermus marinus* PRI378 (Blöndal *et al.*, personal communication)

4.1.3 Strategies for enzyme discovery

Enzymes found in nature have been used since ancient times in the production of food products (cheese, wine, sourdough, beer, vinegar, soy sauce, etc.) and in the manufacture of commodities such as leather, indigo and linen. All of these relied on natural enzymes produced by spontaneously growing microorganisms or enzymes present in added natural preparations such as calves' rumen, papaya fruit or pigeon dung [89, 104]. Over the years these cruder enzymes have been replaced, first by partially purified enzymes from the original source and most recently by recombinant enzymes produced in microorganisms.

To date, a large majority of enzymes commercially available are derived from already cultivated microbial strains (fungi, yeasts, *Bacteria*, etc.) that have been obtained by extensive and labor-costing screening procedures. Nevertheless, an increase in availability has been promoted by the development of cultivation and enrichment techniques (cf. § 3) that provided access to a larger panel of new natural microorganisms such as the soil microbes, marine organisms, or extremophiles. In conjunction with the implementation of heterologous expression strategies has large-scale production of enzymes been achieved [104]. Large contigs of the genomes assembled in DNA libraries (genomic libraries or expression libraries) are subsequently screened either by high-throughput sequencing or by functional screening with a designed activity test. Nevertheless, with the advent of direct molecular analysis of DNA without cultivation, the largest untapped part of microbial diversity and therefore unprecedented source of new molecules can be attained by direct PCR amplification. The key to finding new genes in DNA lies in the fact that enzymes that have a designated function contain characteristic amino acid motifs (catalytic residues in the active site, in substrate-binding domains, etc.) that are usually highly conserved among all organisms and that can furthermore be used to classify them and even to recognize them (cf. § 4.2) [58]. The development of amplification strategies by PCR using degenerate oligonucleotide primers, followed by shotgun cloning and sequencing allows a culture-independent direct access to gene diversity in highly diverse DNA mixtures from the environment (**Papers I to IV**). Nevertheless, the genes still need to be expressed to fully evaluate the encoded enzymatic activity and hence the scope of putative applications. The metagenomic approach, discussed earlier in chapter 3, is also gaining importance in enzyme

discovery. The global metabolic *modi operandi* in complex communities may be occurring through sequential or synergetic enzymatic processes between members of the community, and can therefore allow broader applications and more effective pathways than unique mechanisms promoted by single microorganisms alone [145, 153].

Although enzymes have a huge potential in industrial catalytic processes, they do not always meet specified requirements such as substrate specificity, or temperature and pH stability. Recent advances in recombinant DNA technology have fuelled the development of new enzymatic biocatalysts. In particular, directed evolution has emerged as a powerful tool for enzyme engineering [202]. Directed evolution is a fast but expensive way of producing enzyme variants that work better than naturally occurring enzymes under the required conditions. This process mimics natural evolution under *in vitro* conditions in a series of step-wise improvements mainly by error-prone PCR, cassette mutagenesis, or DNA shuffling. This technique is highly effective for the production of thousands if not millions of mutants, yet the choice and ease of screening strategies for those mutants limit considerably the capacity of producing effective enzymes and the development of high-throughput methods is still an active ongoing area of research [16, 25].

4.2 Classification and function of glycoside hydrolases

Glycoside hydrolases (GH) are widespread groups of carbohydrate-active enzymes present in virtually all organisms and are involved in the hydrolysis of glycosidic bonds between carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. Carbohydrates have a remarkable diversity, a consequence of the extensive variety of naturally occurring monosaccharides (glucose, fructose, ribose, xylose, glucosamine, etc.) and of the different glycosidic bonds possible between them (α-(1,6), α-(1,4), β-(1,4), β-(1,6) etc.). They are generally classified into monosaccharides, disaccharides, oligosaccharides and polysaccharides. They play a unique role in diverse processes such as structure, food storage and utilization, cell wall expansion and turnover, viral invasion and highly selected cellular signalling events [61]. They can be of different lengths, of complex composition and can be branched in various ways. Also, this large diversity of carbohydrate structures is accompanied by an equal diversity of the enzymes responsible for their hydrolysis [18, 64, 192].

4.2.1 Classification of catalytic and non-catalytic modules of glycoside hydrolases

Several systems of classification exist for GHs, including those based on substrate or product specificity and stereochemical mechanisms. The simplest classification, expressed in EC numbers, is based on activity, i.e., substrate used and products formed, sometimes the type of linkage forming the glycosidic bonds, or the molecular mechanism of hydrolysis. This system is simple *per se* and is a straightforward mean to describe GHs. However, the intrinsic problem of that classification is that it does not accommodate enzymes which act on several substrates, e.g. highly complex polysaccharides. This classification also fails to reflect the manifest variety of the 3D structural features of these enzymes (Figure 4.1).

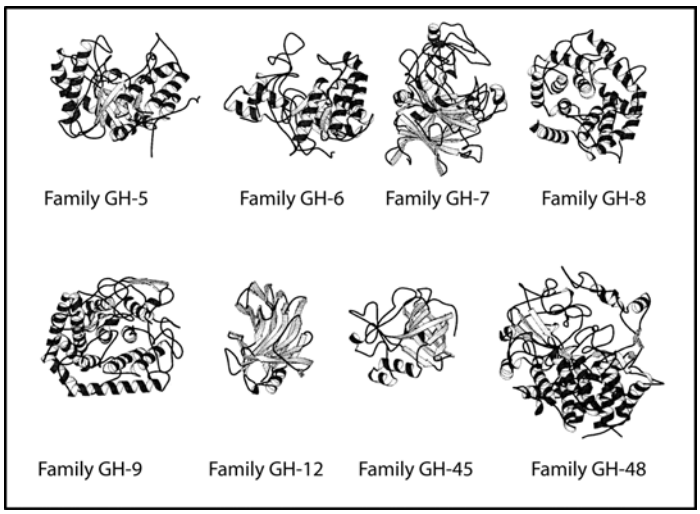

Figure 4.1 Few examples of the 3D structure of β-glucosidases, all described as "cellulases". Such confusion has lead to the re-investigation of the nomenclature used for these enzymes (From Davies & Henrissat [35]).

To cope with this clear limitation, a classification of the catalytic domain of GH into families based on amino acid similarities rather than substrate specificities was introduced a decade ago by Henrissat and is updated regularly (Carbohydrate-Active enZYmes (CAZy), http://afmb.cnrs-mrs.fr/CAZY/index.html) [18, 58]. In the sequence classification system, catalytic modules that display amino acid sequence similarities are grouped into families. The underlying principle of this classification is that because sequence and structure are related, useful structural and mechanistic information can be derived from the amino acid sequence alone. Furthermore, as the primary and secondary structures are conserved, members of a family will adopt

similar 3D structures, and the mechanisms of catalysis will be conserved within each family. By extension, once the stereochemical mechanism is established for one member of a family, it can be safely extended to other members of this family [59, 61, 63]. Over the years, the number of families of GHs has grown constantly and to date (October 2004) there are 97 families listed. These families can be further classified in superfamilies or clans of related families (labeled GH-A to GH-M) in which the catalytic domain fold, the catalytic residues and the catalytic mechanism are conserved [18, 60]. In addition, it is possible to find enzymes of different substrate specificity that display nevertheless similar amino acid sequences. Those families, called polyspecific, represent over one third of all glycoside hydrolase families [63].

Like any other living organisms thermophiles and hyperthermophiles contain GHs. In 1999, the known GHs from thermophiles were scattered among 33 of the 77 known families (Table 4.2) [64].

Table 4.2 Various glycoside hydrolase in thermophiles and hyperthermophiles in 1999. The lack of consensual nomenclature results in a variety of enzyme denominations that may or may not describe the same catalytic activity (modified from Henrissat and Coutinho[64]).

Family	Proteins [a]	Organisms [b]
GH-1	β-glucosidase	*Caldicellulosiruptor saccharolyticus*
		Pyroccus furiosus
		Thermoanaerobacter brockii
		Thermococcus sp. 9ON-7
		Thermotoga maritima
	β-mannosidase	*Pyrococcus furiosus*
	β-glycosidase	*Pyrococcus kodakaraensis*
		Thermus thermophilus
	β-galactosidase	*Pyrococcus woesei*
		Sulfolobus shibatae
GH-2	β-galactosidase	*Thermoanaerobacter ethanolicus*
		Thermotoga neapolitana
GH-3	β-glucosidase	*Thermoanaerobacter brockii*
		Thermotoga neapolitana
	β-xylosidase	*Thermotoga neapolitana*
GH-5	Endo-1,4-glucanase	*Anaerocellum thermophilum*
	Endo-1,4-β-mannosidase	*Caldicellulosiruptor saccharolyticus*
	β-mannanase	*Thermoanaerobacter polysaccharolyticum*
GH-9	1,4-β-glucanase	*Anaerocellum thermophilum*
	Endo-1,4-β-mannosidase	*Caldicellulosiruptor saccharolyticus*
GH-10	Xylanase	*Anaerocellum thermophilum*
		Caldicellulosiruptor saccharolyticus
		Dictyoglomus thermophilus
		Rhodothermus marinus

		Thermoanaerobacter saccharolyticum
		Thermotoga maritima
	Endo-1,4-β-glucanase	*Caldicellulosiruptor saccharolyticus*
GH-11	Xylanase	*Caldicellulosiruptor sp.* Rt69B.1
		Dictyoglomus thermophilus
GH-12	Cellulase	*Rhodothermus marinus*
	Endo-1,4-glucanase	*Pyrococcus furiosus*
		Thermotoga maritima
GH-13	α-amylase	*Dictyoglomus thermophilus*
		Pyrococcus furiosus
		Sulfolobus solfataricus
		Thermoanaerobacter saccharolyticum
		Thermococcus hydrothermalis
		Thermotoga maritima
	Cyclodextrinase	*Thermoanaerobacter saccharolyticum*
	Glycosyltrehalose-producing enzyme	*Sulfolobus solfataricus*
	Isoamylase	*Sulfolobus acidocaldarius*
	Maltogenic amylase	*Thermus thermophilus*
	Pullulanase	*Caldicellulosiruptor saccharolyticus*
		Thermotoga maritima
		Thermus thermophilus
GH-15	Glucoamylase	*Thermoanaerobacter saccharolyticum*
GH-16	β-1,3-glucanase	*Pyrococcus furiosus*
	Laminarinase	*Rhodothermus marinus*
		Thermotoga neapolitana
	Licheninase	*Rhodothermus marinus*
GH-18	Endochitinase - Exochitinase	*Rhodothermus marinus*
		Pyrococcus furiosus
		Thermococcus kodakaraensis
GH-26	β-mannanase	*Caldicellulosiruptor saccharolyticus*
		Dictyoglomus thermophilus
		Rhodothermus marinus
		Thermotoga neapolitana
GH-31	α-glucosidase	*Sulfolobus solfataricus*
GH-36	α-galactosidase	*Thermoanaerobacter ethanolicus*
		Thermotoga maritima
		Thermus brockianus
GH-39	Xylanase	*Caldicellulosiruptor saccharolyticus*
	β-xylosidase	*Thermoanaerobacter saccharolyticum*
GH-42	β-galactosidase	*Thermotoga maritima*
		Thermus sp. A4
GH-43	Xylanase	*Caldicellulosiruptor saccharolyticus*
GH-48	Cellobiohydrolase	*Caldicellulosiruptor saccharolyticus*
GH-57	α-amylase	*Dictyoglomus thermophilus*
		Pyrococcus sp.
	Amylopullulanase	*Thermococcus hydrothermalis*
GH-67	α-glucuronidase	*Thermotoga maritima*
GH-77	Amylomaltase	*Thermus aquaticus*

[a] The protein names used are as listed in the GenBank database.
[b] Some of the enzymes are present in several species scattered within a genus but only one representative for each genus was listed here.

Many GHs contain other modules devoid of catalytic activity in addition to their catalytic module(s). The carbohydrate-binding modules (CBM) are non-catalytic modules that function independently from the catalytic core of the enzymes and have an essential role for the enzyme activity by promoting the adsorption of the enzyme to the substrate. CBMs are about 50 to 150 amino acids in length and display a number of highly conserved residues that are responsible for the attachment of the protein to the polysaccharidic chain. This promotes hence the proper placement of the chain in the vicinity of the catalytic core of the enzyme. Comparably to GHs, CBMs can be classified into families based on amino acid conservation, and 41 distinct families have been identified to date (October 2004). Several other non-catalytic modules are also present and their activity is, with few exceptions, mostly unknown. Substantial data shows however that the deletion of these unknown modules generally results in a global lower catalytic activity. Due to the 3D conformation of the enzymes, the non-catalytic domains are generally set apart from the catalytic core by linker sequences rich in Ser, Thr, Pro and Glu that differ considerably in length, even for GHs belonging to the same family (for review, see Warren [192]). Furthermore, despite the general conservation of the amino acid sequences within the different modules, the overall structural organization of the modules is very variable from one enzyme to the other [63] (Figure 4.2).

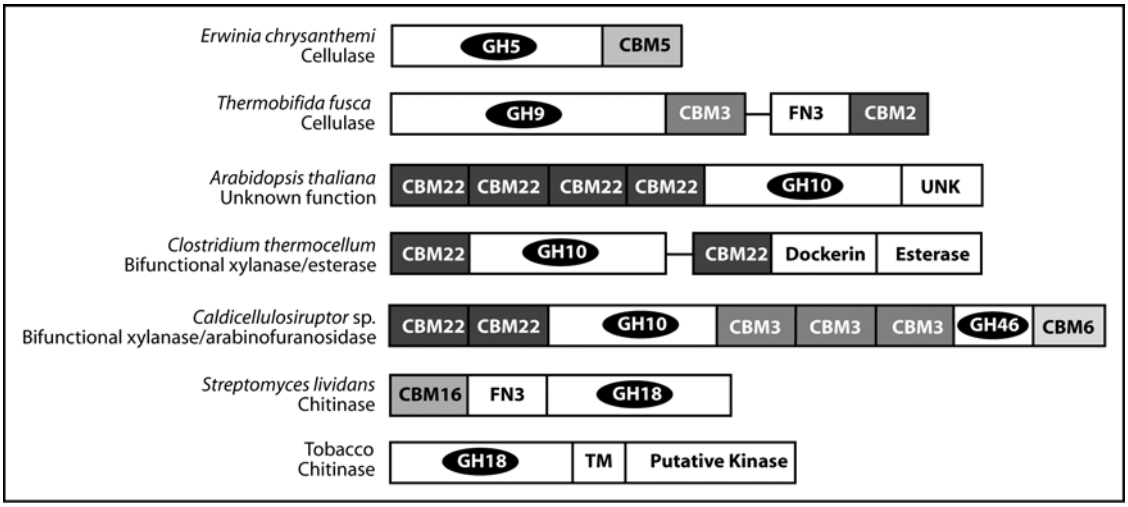

Figure 4.2 Examples of the different modular organization in glycoside hydrolases and related proteins. Legend: GH, glycoside hydrolase; CBM, carbohydrate-binding module; FN3, fibronectin III-like module: UNK, module of unknown function; TM, transmembrane segment and other modules are indicated by their function. When two consecutive modules are separated by a clearly identifiable linker peptide, the peptide is indicated by a horizontal bar between the modules (from Henrissat & Davies [63]).

4.2.2 Catalytic mechanisms of glycoside hydrolases

In all GHs, enzymatic hydrolysis of the glycosidic bond takes place via general acid catalysis that requires generally two invariant carboxylic acids found in the active cleft that act as proton donor or acid, and a nucleophile or base. This hydrolysis occurs via two major mechanisms that result in either inversion (α→β; β→α) or retention (α→α; β→β) of the anomeric configuration of the glycosidic bond [57, 91, 147, 155].

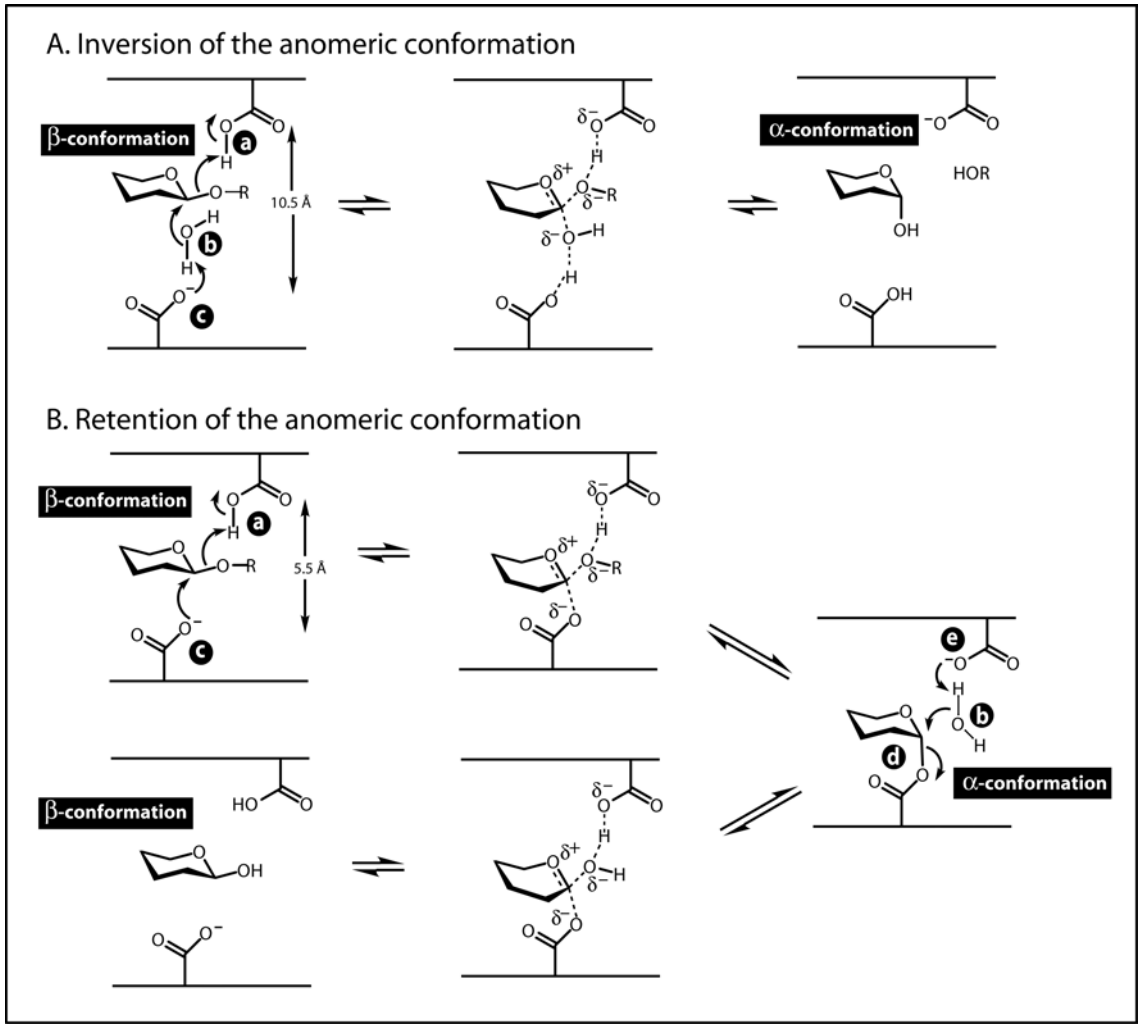

Figure 4.3 General hydrolytic mechanisms for **(A)** an inverting β-glycosidase and **(B)** a retaining β-glycosidase. Legend: a, general acid; b, water-based nucleophilic attack; c, carboxylate base; d, intermediate with the glycosyl residue covalently linked to the enzyme; e, base form of the catalyst acid (modified from Rye & Withers [147]).

Inverting GHs lead to an inversion of the anomeric configuration via a single nucleophilic displacement. In the suggested mechanism of inverting enzymes (Figure 4.3A) the glycosidic oxygen is initially protonated by the general acid catalyst, which

is followed by a nucleophilic attack of a water molecule activated by the carboxylate base. This leads consequently to the inversion of the anomeric conformation. Bond breaking and bond making proceed both in a single concerted step [57, 91, 147, 155]. The retaining reactions proceed via a double-displacement mechanism (Figure 4.3B). The first step involves a similar protonation of the glycosidic oxygen by the general acid as found in the inverting mechanism, creating a covalently-linked glycosyl-enzyme intermediate. In the second step this intermediate is attacked by the water nucleophile assisted by the base form of the acid catalyst. Each displacement step inverts the configuration of the anomeric carbon, resulting in an overall retention of the configuration [57, 91, 147, 155]. In GH family 18 however (chitinases), the reaction intermediate is stabilized intra-molecularly by the *N*-acetyl group in the substrate, and not by the formation of a covalent-glycosyl enzyme intermediate [185] (cf. § 4.4.2). In most inverting enzymes the distance between the two catalytic residues is around 10.5 Å, whereas the distance is about 5.5 Å in retaining GHs [147].

4.3 Starch-degrading enzymes – thermophilic α-amylases from family GH13

Starch from cultivated plants represents a ubiquitous and easily accessible source of energy. Starch-containing crops form thus an important constituent for human diet and a large proportion of the food consumed by the world's population originates from wheat, rice, maize, potatoes, tapioca, potato, cassava, etc. Besides, starch is also harvested and used as such or chemically and/or enzymatically processed into a large variety of products, ranging from the food industry to the detergent production (washing and dish-washing detergents). The necessity to run starch processing at high temperatures has created a need of thermostable biological agents and thermostable amylases have today become irreplaceable in the starch industry [123].

4.3.1 Nature and properties of starch

Starch is synthesized as a storage compound for respiration during the dark periods via photosynthesis in plastids located in the plant leaves. It is also synthesized in amyloplasts found in tubers, seeds and roots as long-term storage compound. In these latter organelles, large amounts of starch accumulate as water-insoluble granules, the shape and diameter of these granules depending on the botanical origin

of the plant. Starch is composed exclusively of α-glucose units that are linked by α-(1,4)- and/or α-(1,6)-glycosidic bonds. The two high molecular weight components of starch are amylose and amylopectin. Amylose consists mainly of linear chains of α-(1,4)-linked glucopyranose residues, approximately 1000 residues long (Figure 4.4A). Pure amylose forms hydrogen bonds between the molecules in solution resulting in rigid gels. Amylopectin is a branched polymer containing, in addition to α-(1,4)-glycosidic linkages, α-(1,6)-linked branches occurring every 17 to 26 glucose units on the linear chain (Figure 4.4B). The branches themselves form an organized 3D structure: A-chains are not substituted at the C_6 positions and inner B-chains are α-(1,6)-branched at one point (B1-chain), or several points (B2, B3, etc.). These branches are not randomly organized but are clustered at 7 nm to 10 nm intervals (Figure 4.4C). In solution amylopectin forms fewer hydrogen bonds than amylose, therefore it remains fluid with a high viscosity and elasticity [11, 110]. Amylopectin doesn't exist *per se* in bacteria but can be found under the equivalent form of glycogen, which is a common substrate for GH13 α-amylases [186]. The amylose/amylopectin ratio in starch varies considerably depending on the origin, plant species, variety within plants, plant organs, age of organ and growth conditions. Consequently, this results in various crystalline organization of starch in the granules and henceforth different sensitivities of starch to enzymatic hydrolysis [110].

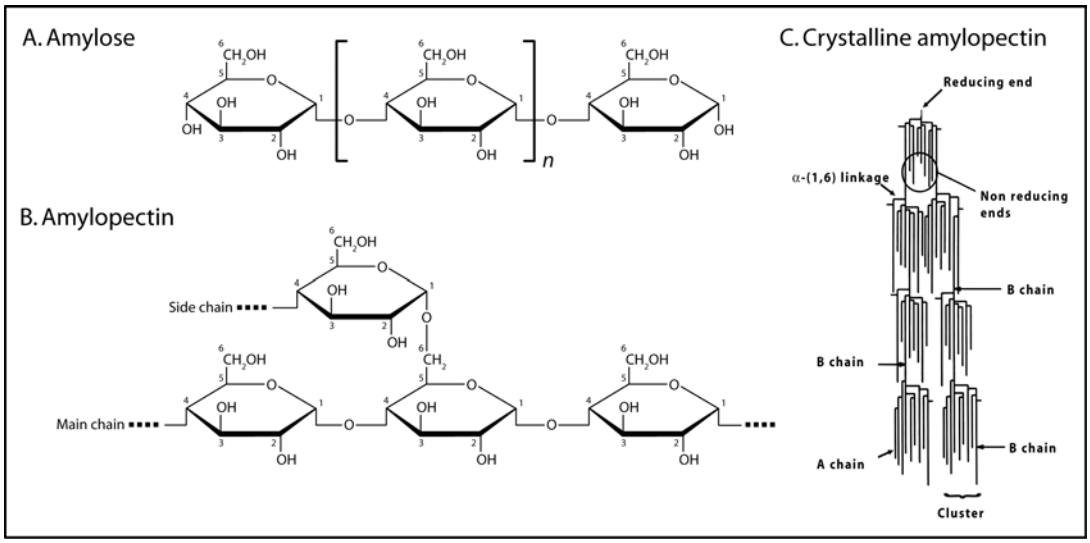

Figure 4.4 Biochemical structures of starch components: (A) amylose and (B) amylopectin. (C) Crystalline structure of amylopectin (from Martin & Smith [110]).

4.3.2 Diversity and action of amylolytic enzymes – the α-amylases from family GH13.

Because of the complex structures of starch, cells require a combination of hydrolytic enzymes for the depolymerization of starch into oligosaccharides and smaller sugars (endoamylases, exoamylases), or to transform starch by transfer of oligoglycosidic branches and residues by the creation of new bonds (debranching enzymes and glycosyl-transferases).

The process of enzymatic starch conversion is displayed on Figure 4.5. Endo-acting enzymes or endoamylases such as α-amylases (EC 3.2.1.1) hydrolyze α-(1,4) linkages in the interior of the starch polymer in a random fashion, which leads to the formation of linear and branched oligosaccharides, or α-limit dextrins [11]. These enzymes are found in a wide variety of microorganisms, belonging to both *Bacteria* and *Archaea* [186]. Exo-acting enzymes include β-amylases, glucoamylases and α-glucosidases. These enzymes cleave exclusively α-(1,4) bonds, such as β-amylases (EC 3.2.1.2), or both α-(1,4) and α-(1,6) bonds such as glucoamylases (EC 3.2.1.3) and α-glucosidases (EC 3.2.1.20). Exoamylases act on the external glucose residues of amylose or amylopectin, starting from the reducing end. They produce well-defined oligosaccharides: glucose by glucoamylases and α-glucosidases, and maltose or β-limit dextrin by β-amylases. β-amylases and glucoamylases promote the inversion of the anomeric conformation from α to β but not α-glucosidases [186]. Debranching enzymes hydrolyze exclusively α-(1,6) glycosidic bonds and are composed of isoamylases (EC 3.2.1.68) and pullulanase type I (EC 3.2.1.41). Isoamylases and pullulanases degrade exclusively amylopectin thus leaving long linear poly-saccharides. Type I and II pullulanases have also the ability to catalyze the hydrolysis of the side-chain in pullulan, a polysaccharide with repeating units of α-(1,6)-branched maltotriose. Additionally, pullulanases type II, also called amylo-pullulanases, are able to hydrolyze both α-(1,6) and α-(1,4) bonds. They debranch pullulan and give maltotriose as final product, but they attack the α-(1,4) bonds in starch, amylose and amylopectin [186]. Amylopullulanases were also shown to have two distinct active sites each promoting the hydrolysis of either α-(1,4) or α-(1,6) linkages [121].

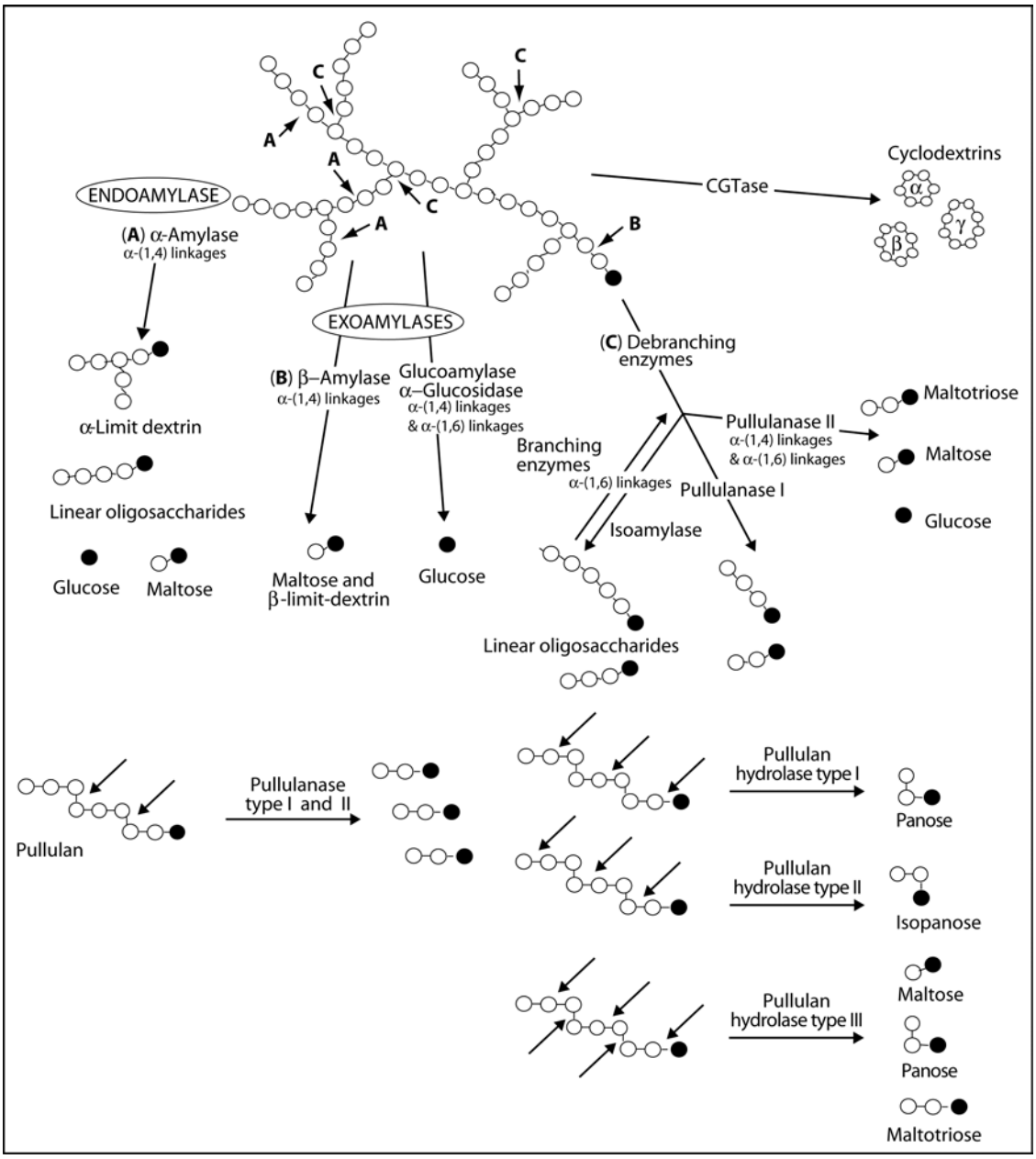

Figure 4.5 Schematic presentation of the action of amylolytic and pullulytic enzymes. Black circles indicate reducing sugars (modified from Bertoldo & Antranikian [12]).

Transferases have little hydrolytic activity. Cyclodextrin glycosyltransferases (CGTases) are unique in the sense that they form circular α-(1,4)-linked oligosaccharides (cyclodextrins) from starch. The major products are α, β, and γ-cyclodextrin with 6, 7 and 8 residues respectively. CGTases have a very low hydrolytic activity and catalyze preferentially the formation of those cyclodextrins. Amylomaltase, or 4-α-glucanotransferases, are very similar to CGTases by promoting the transfer of glycosidic residues through the formation of α-(1,4) bonds, although they yield linear products and not circular products [186]. Finally, neopullulanases

(EC3.2.1.135), maltogenic amylases (3.2.1.133) and cyclodextrinases (EC 3.2.1.-) are enzymes that degrade pullulan, starch and cyclodextrins completely into maltose by catalyzing the hydrolysis of α-(1,4) and α-(1,6) glycosidic bonds. However, whether those enzymes are exo- or endohydrolases is not clear, although they hydrolyze cyclodextrins and may be more inclined towards endohydrolysis. Because these enzymes appear to have very similar amino acid sequences and consequently similar 3D structures as well as common hydrolytic behavior, they can also be cited as cyclomaltodextrinases (EC 3.2.1.54). Interestingly, cyclomaltodextrinases are intracellular enzymes whereas most enzymes involved in starch conversion are extracellular [100].

Table 4.3 Reaction specificities of starch degrading α-amylases from family GH13 (modified from Leemhuis [100] and Henrissat *et al.* [57]).

Enzyme	EC number	H or T [a]	Bonds processed	Preferred Substrate
α-amylase	3.2.1.1	H	α-(1,4)	Starch
Maltotetraose-forming amylase	3.2.1.60	H	α-(1,4)	Starch
Maltopentaose-forming amylase	3.2.1.-	H	α-(1,4)	Starch
Maltohexaose-forming amylase	3.2.1.98	H	α-(1,4)	Starch
α-glucosidase	3.2.1.20	H	α-(1,4)	Terminal non-reducing glucose residues
Cyclodextrin glycosyltransferase	2.4.1.19	T	α-(1,4)	Starch
4-α-glucanotransferase	2.4.1.25	T	α-(1,4)	Starch
Cyclomaltodextrinase	3.2.1.54	H & T	α-(1,4) & α-(1,6)	CDs [b], pullulan, starch
Neopullulanase	3.2.1.135	H & T	α-(1,4) & α-(1,6)	CDs [b], pullulan, starch
Maltogenic amylase	3.2.1.133	H & T	α-(1,4) & α-(1,6)	CDs [b], pullulan, starch
Oligo-1,6-glucosidase	3.2.1.10	H	α-(1,6)	Terminal non-reducing glucose residues
Pullulanase type I	3.2.1.41	H	α-(1,6)	Pullulan, starch
Isoamylase	3.2.1.68	H	α-(1,6)	Starch
Branching enzyme	2.4.1.18	T	α-(1,4) → α-(1,6)	Starch
Glucodextranase	3.2.1.70	H	α-(1,6)	α-(1,6)-glucans
Pullulanase type II [c]	3.2.1.41	H	α-(1,4)	Starch
Pullulanase type II [c]	3.2.1.41	H	α-(1,6)	Pullulan

[a] Main activity, H is hydrolysis and T is transglycosylation
[b] CDs, cyclodextrins
[c] These enzymes have two distinct active sites that are responsible for the two separate activities.

Almost all these enzymes belong to one superfamily, the α-amylase superfamily. This superfamily, also known as the GH-H clan, is composed purely of retaining enzymes belonging to a very large majority to family GH13 (α-amylases, α-glucosidases, pullulanases, cyclodextrin-glucosyl transferases, etc. (Table 4.3)), and

the few representatives from families GH70 and GH77 [101, 192]. The other groups of amylolytic enzymes are the families GH14 of inverting β-amylases, and GH15 of inverting glucoamylases. Retaining α-amylases can also be found in GH family 57 but show distinct amino acid sequences unrelated to those from clan GH-H.

According to Kuriki & Imanaka [98], GH13 α-amylases are defined by (i) the hydrolysis of α-glycosidic bonds, (ii) the hydrolysis or transfer of α-glycosidic bonds by retention of the anomeric conformation, (iii) the conservation of four amino acid sequence motifs containing the catalytic residues and most other residues involved in substrate-binding within the catalytic site, and (iv) the presence of Asp, Glu and Asp as catalytic residues. α-Amylases share a common $(\beta/a)_8$- or TIM-barrel catalytic domain which consists of a closed eight-stranded parallel β-sheet that is surrounded by eight α-helices (Figure 4.6). Usually, the loops that link β-strands to the adjacent helices carry amino acid residues of the active site [100, 101]. The highly conserved residues are located within the motifs of the catalytic domain and additional non-catalytic modules from family GH13 (Figure 4.7). They are used for the classification of the enzymes within the family and can consequently serve for the design of degenerate oligonucleotidic primers for PCR (cf. § 4.1.3; **Paper II**).

Figure 4.6 Three dimensional structure of a representative member of the α-amylase GH13 family, *Bacillus circulans* 251 CGTase. The catalytic $(\beta/\alpha)_8$ barrel domain is shown in black and the other domains in grey (reprinted from Leemhuis [100]).

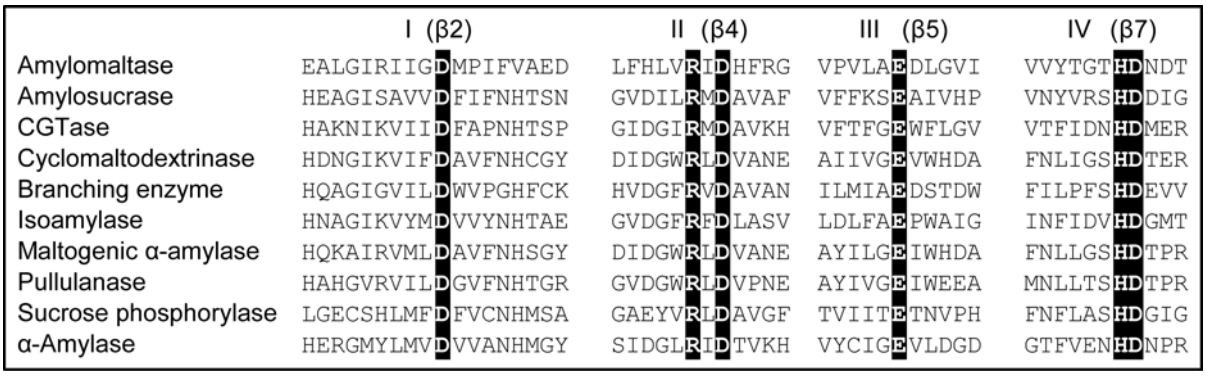

Figure 4.7 The four conserved regions and the corresponding β-sheets found in the amino acid sequence of α-amylase family 13 enzymes (from van der Maarel *et al.* [186]). Highlighted are the conserved catalytic amino acid residues. The following enzymes were used for the alignment: Amylomaltase of *Thermus aquaticus* (GenBank No: BAA33728); Amylosucrase of *Neisseria polysaccharea* (CAA09772); CGTase: cyclodextrin glycosyltransferases of *Bacillus circulans* (CAA55023); Cyclomaltodextrinase of *Clostridium thermohydrosulfuricim* (A42950); Branching enzyme of *Geobacillus stearothermophilus* (AAA22482); Isoamylase of *Pseudomonas amyloderamosa* (S13470); Maltogenic α-amylase of *Bacillus stearothermophilus* (AAC46346); Pullulanase of *Bacillus flavocaldarius* (BAB18516); Sucrose phosphorylase of *Escherichia coli* K12 (BAA14878); α-amylase of *Bacillus licheniformis* (A44326). β2, β4, β5, and β7 indicate the β-sheet in which the region is present.

4.3.3 Thermophilic α-amylases and their applications in the conversion of starch.

The finding of extremely thermostable starch-hydrolyzing enzymes (α-amylases, pullulanases) that are active under similar conditions improved significantly the industrial bioconversion of starch, i.e., liquefaction, saccharification and isomerization (Figure 4.8). Because of the lack of novel thermostable enzymes that are active and stable above 100°C and at acidic pH values, the bioconversion of starch to glucose is performed step-wise at different temperatures and under different pH conditions. This multistep process (step 1: pH 6-6.5, 95 to 105°C; step 2: pH 4.5, 60 to 62°C; step 3: pH 7-8.5, 55 to 60°C) is accompanied by the formation of undesirable high concentrations of salts, which have to be removed by expensive ion exchangers [123]. Ideally, an α-amylase able to work at low pH would reduce substantially the cost of pH adjustments, in particular during the liquefaction. Additionally, they would simplify the process and reduce the formation of high-pH by-products [30]. α-Amylases which do not require added Ca^{2+} and which operate above 100°C at acid pH values are today mostly targeted for improved processing.

The pullulanases, isoamylases, β-amylases and glucoamylases used today in industrial starch processing originate from mesophilic organisms. There is also a need for thermostable enzymatic equivalents because increasing the temperature of the saccharification process would be beneficial. This includes: (i) higher substrate concentration, (ii) decreased viscosity and lower pumping costs, (iii) limited risks of bacterial or fungal contamination, (iv) increased reaction rates and decrease in operation time, (v) lower costs of enzyme purification, and (vi) longer catalyst half-life due to the inherent enzyme thermostability [190].

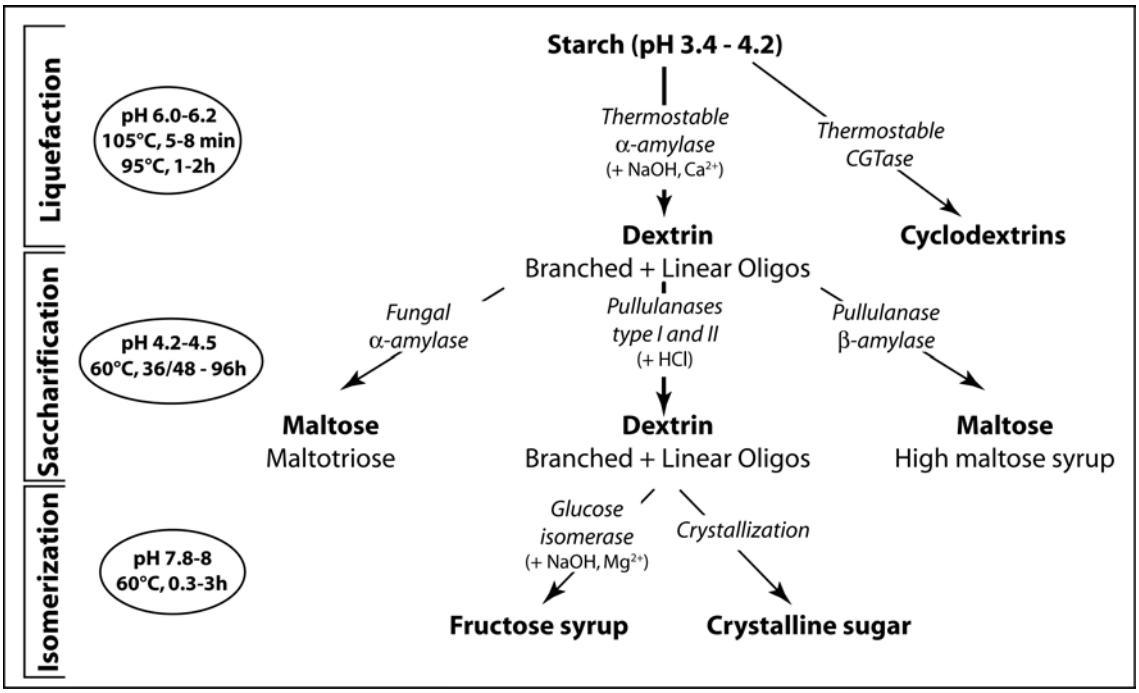

Figure 4.8 Overview of the industrial processing of starch into cyclodextrins, maltodextrins, glucose or fructose and crystalline sugar (from van der Maarel *et al.* [186] and Crabb & Shetty [31]).

Enzymes from thermophiles and hyperthermophiles are regarded as interesting candidates for use in the starch industry and intensive research has been performed aimed at the isolation of thermostable and thermoactive amylases from those microorganisms [123]. As shown on Table 4.4, a large number of enzymes are present in a similarly wide diversity of microbes belonging to both *Bacteria* and *Archaea*. The molecular cloning of the corresponding genes and their expression in heterologous hosts, e.g., *Escherichia coli* or *Bacillus subtillis*, allowed circumvention of the problem of insufficient expression in the original host (*Archaea* mostly) [190].

Highly thermostable α-amylases have been obtained, some even that do not require Ca^{2+} for activity and stability, although none meet the desired feature of

activity at low pH. Substantial data has been gathered on thermophilic and hyperthermophilic exoamylases (β-amylases, glucoamylases, and α-glucosidases) and some of the enzymes show very good potential for the saccharification step.

Table 4.4 Examples of thermophilic and hyperthermophilic enzymes with potential applications in starch processing (data compiled by Vieille *et al.* [190]).

Enzyme	Origin	Properties
α-Amylase	*Desulfurococcus mucosus*	Optimal activity at 100°C, pH 5.5
	Pyrococcus furiosus	Optimal activity at 100°C, pH 5.5–6.0; $t_{1/2}$, 13 h (98°C) [a]
	Pyrococcus woesei	Optimal activity at 100°C, pH 5.5
	Pyrodictium abyssi	Optimal activity at 100°C, pH 5.0
	Staphylothermus marinus	Optimal activity at 100°C, pH 5.0
	Thermococcus profundus	Optimal activity at 80°C, pH 4.0–5.0
	Dictyoglomus thermophilum	Optimal activity at 90°C, pH 5.5
	Thermotoga maritima	Optimal activity at 85–90°C, pH 7.0
Pullulanase	*Bacillus flavocaldarius*	Optimal activity at 75–85°C, pH 6.3; $t_{1/2}$, 10 min (107°C)
type I	*Thermotoga maritima*	Optimal activity at 90°C, pH 6.0
	Thermus caldophilus	Optimal activity at 75°C, pH 5.5; $t_{1/2}$, 72 h (78°C, pH 6.0)
	Fervidobacterium pennavorans	Optimal activity at 80–85°C, pH 6.0; $t_{1/2}$, 2 h (80°C)
Pullulanase	*Desulfurococcus mucosus* ES4	Optimal activity at 100°C, pH 5.0
type II	*Pyrococcus furiosus*	Optimal activity on starch at 120°C, pH 5.5–6.5 $t_{1/2}$, 6.5 h (98°C) and 20 h (98°C +$CaCl_2$)
	Thermococcus celer	Optimal activity at 105°C, pH 6.0 $t_{1/2}$, 44 h (90°C + 5 mM CaCl2)
	Thermococcus litoralis	Optimal activity at 90°C, pH 5.5
	Thermococcus hydrothermalis	Optimal activity at 117°C, pH 5.5 (5 mM $CaCl_2$) $t_{1/2}$, 8 h (95°C) and 1 h (105°C)
	Thermoanaerobacter ethanolicus	Optimal activity at 90°C, pH 5.5; $t_{1/2}$, 40 min (90°C)
Glucoamylase	*Clostridium thermosaccharolyticum*	Optimal activity at 70°C, pH 5.0; 90% active after 6 h (70°C)
	Methanococcus jannaschii	Putative gene identified in the genome sequence
	Thermoanaerobacterium thermosaccharolyticum	Optimal activity at 50–60°C, pH 4.0–5.5 stable for 8 h (65°C)
α-Glucosidase	*Thermoanaerobacter ethanolicus*	Optimal activity at 75°C, pH 5.0–5.5; $t_{1/2}$, 46 h (60°C) and 35 min (75°C)
β-Amylase	*Thermotoga maritima*	Optimal activity at 95°C, pH 4.3–5.5; $t_{1/2}$, ca. 30 min (90°C)
	Thermoanaerobacterium thermosulfurigenes	Optimal activity at 75°C, pH 5.5–6.0; stable at pH 3.5–6.5; 80% active after 1 h (75°C)
CGTase	*Thermococcus* sp.	Optimal activity at 90–100°C, pH 5.0–5.5 (can work at pH 4.5); $t_{1/2}$, 40 min (110°C)

[a] $t_{1/2}$, Half-life of the enzyme is given in hours/minutes at the described temperature.

Thermophilic pullulanases have been isolated from aerobic thermophiles. These pullulanases are optimally active at low pH but their application for the starch processing remains to be tested. Cyclodextrins have numerous applications outside the production of sugars because the internal cavities of CDs are hydrophobic and

therefore can encapsulate hydrophobic molecules. This property makes CDs suitable in the food, cosmetic, and pharmaceutical industries where they are used to capture undesirable tastes and odors, stabilize volatile compounds, increase the water solubility of the hydrophobic substance, and protect a substance against unwanted modifications. The thermophilic CGTase from *Thermococcus* could thus be used to develop a one-step CDs production in which it would replace α-amylase for starch liquefaction [190].

4.4 Chitin-degrading enzymes – chitinases from family GH18

Chitin, a Greek work for envelop, was discovered in 1811 as a substance occurring in mushrooms [176]. It is the second most abundant natural biopolymer on Earth after cellulose, with an estimated production of 10^{10} to 10^{11} tons *per annum* [51]. The study of the biological processes yielding to the enzymatic hydrolysis of chitin by microorganisms is motivated by two things. First, the omnipresence of chitin in the environment makes it important to understand the role and the underlying metabolic processes of chitin turnover in the environment. Second, chitin and its derived compounds have become of great interest not only as an under-utilized bioresource, but also as a new functional material of high potential in various fields. The diverse uses of chitin and chitin-derived products have resulted in a demand for enzymes that can be used to tailor them to specific needs [71]. Additionally, chitinases have been shown to have significant other biotechnological applications such as in the fields of agriculture or food processing [130]. The development of chitinases for industrial uses is still at an early stage but few enzymes have been isolated and were shown to have high potentials as biocontrol agents and pesticides against plant pathogens and fungi, as control agents against disease-transmitting mosquitoes, for the lysis of fungal cells and the production of single-cell proteins or for the production of chitooligo-saccharides from chitin (for review, see Patil *et al.* [130]; **Paper V**).

4.4.1 Nature and properties of chitin and chitosan

Nature and molecular structure of chitin

Chitin is present in the cell wall of most fungi and yeasts, in the cuticle of arachnids and insects, and in the shell of crustaceans, diatoms, protozoans and nematodes [120, 176]. Chitinous structures are mainly of ectodermal origin in

multicellular animals and form the characteristic exoskeleton of most of the invertebrates, in which it constitutes exceptionally more than half of the total organic matter [176]. The occurrence of chitin in various organisms is given in Table 4.5. Due to its ubiquitous presence in the shells of crustaceans and molluscs, the largest amount of free chitin is found in the marine environment where it serves as a nutritional source of carbon and nitrogen compounds for chitinolytic microorganisms [112]. In the oceans, dynamic processes such as senescence and molting of shells and cuticles result in a continuous rain of chitin to the ocean floor known as "marine snow". Despite this continuous deposition of highly insoluble polymer, ocean sediments contain only traces of free chitin. The ubiquity of marine chitinolytic microorganisms in the water column can be directly linked to a significant portion, if not the bulk, of the chitin recycling process [88, 143]. Also, anaerobic degradation and utilization of chitin in the ocean sediments is thought to be coupled to processes such as methanogenesis or sulfate-reduction via interspecies hydrogen transfer [19]. Measurement of the dissolution of crab shells immersed in sea water showed that the half-life of a shell is approximately 2 weeks [133].

Table 4.5 Chitin content of selected crustaceans, insects, molluscan organs and fungi (data collected by Tharanathan & Kittur [176]).

Organism	Chitin Content (%)	Organism	Chitin Content (%)
Crustacean		Insects (continued)	
Cancer (crab)	72.1[c]	*Bombyx* (silk worm)	44.2[c]
Carcinus (crab)	64.2[b]	*Calleria* (wax worm)	33.7[c]
Paralithodes (king crab)	35.0[b]	Molluscan organs	
Callinectes (blue crab)	14.0[a]	Clamshell	6.1
Crangon (shrimp)	69.1[c]	Oyster shell	3.6
Alaskan shrimp	28.0[d]	Squid, skeletal pen	41.0
Nephrobs (lobster)	69.8[c]	Krill, deproteinized shell	40.2
Homarus (lobster)	60-75[c]	Fungi	
Lepas (barnacles)	58.3[c]	*Aspergillus niger*	42.0[e]
Insects		*Penicillium notatum*	18.5[e]
Periplaneta (cockroach)	2.0[d]	*Penicillium chrysogenum*	20.1[e]
Blatella (cockroach)	18.4[c]	*Saccharomyces cerevisiae*	2.9[e]
Coleoptera (beetle)	27-35[c]	*Mucor rouxii*	44.5
Diptera (true fly)	54.8[c]	*Lactarius vellereus* (mushroom)	19.0
Pieris (sulphur butterfly)	64.0[c]		

[a] Wet body weight
[b] Dry body weight
[c] Organic weight of cuticle
[d] Total weight of cuticle
[e] Dry weight of the cell wall

Chitin is a white, hard, inelastic nitrogenous high-molecular weight linear homopolymer of *N*-acetyl-β-D-glucosamine units (GlcNAc) linked by β-(1,4) bonds (Figure 4.9A). It has a compact crystalline structure and is highly insoluble resembling cellulose in its solubility, low chemical activity and natural functions as a structural polysaccharide [130]. The most common allomorph exhibited by chitin in nature is known as α-chitin [24], where the unit cell is orthorhombic and the individual chains are arranged in antiparallel fashion (Figure 4.10A). A less common allomorph, known as the β-conformation [149], corresponds to a monoclinic unit cell with the polymer chains arranged in a parallel fashion (Figure 4.10B). The microcrystalline structure of chitin varies between both α and β conformations: α-chitin is found in the calyces of hydrozoa, mollusks, plankton, and as a component of the cuticles of arthropods; β-chitin, a less stable and more degradable form of chitin, is found in mollusks, squid pen, diatoms, and insect exoskeletons and cocoons and is the major component of fungal cell walls [120]. A third form, γ-chitin, has been reported in the past from the stomach of the squid *Loligo* and is thought to be characterized by a three-chain unit cell in which two right-ward chains are followed by a left-ward chain (↑↑↓) [176]. This form has not been subjected to detailed structural analyses and its existence appears to be controversial today [166]. Incidentally, little information is currently available regarding the distribution of γ-chitin in nature. In most organisms, chitin is modified by forming linkages with other polymers, i.e., glucans or proteins, due to the availability of reactive free amino groups on the C_2. Chitinous structures also have the capacity to engulf minerals, salts, waxes, carotenoids, etc. This gives rise to a large variety of macro-molecules with new structure and physiochemical characteristics. For instance, when chitin is linked to proteins, the chitinous complex exhibits high tensile strength but is essentially pliable and flexible allowing movements and limited expansion, such as in the cuticle of insects [112].

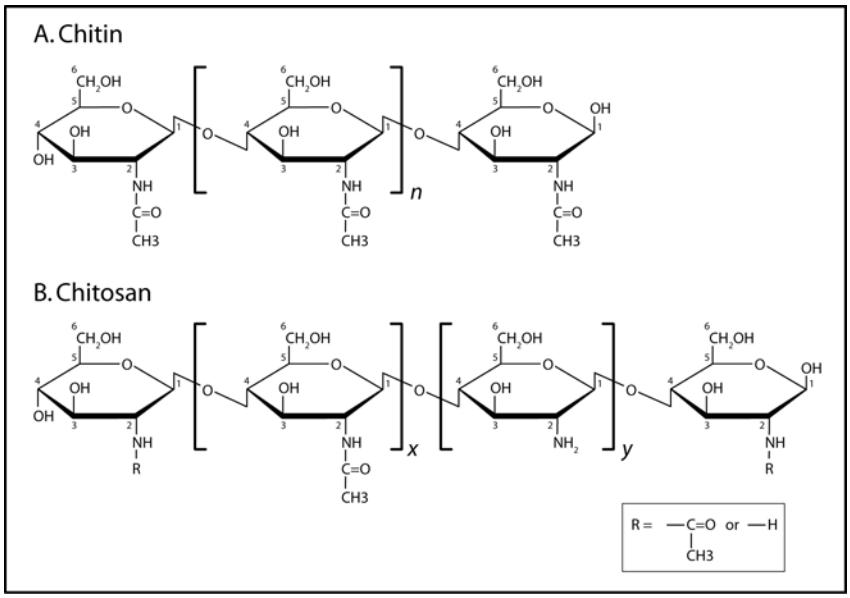

Figure 4.9 Molecular structures of chitin and chitosan. Chitin is a homopolymer of *N*-acetyl-D-glucosamine. Chitosan is a partially deacetylated co-polymer of chitin where x represents the fraction of *N*-acetyl-D-glucosamine residues and y represents the fraction of D-glucosamine residues (x+y=1).

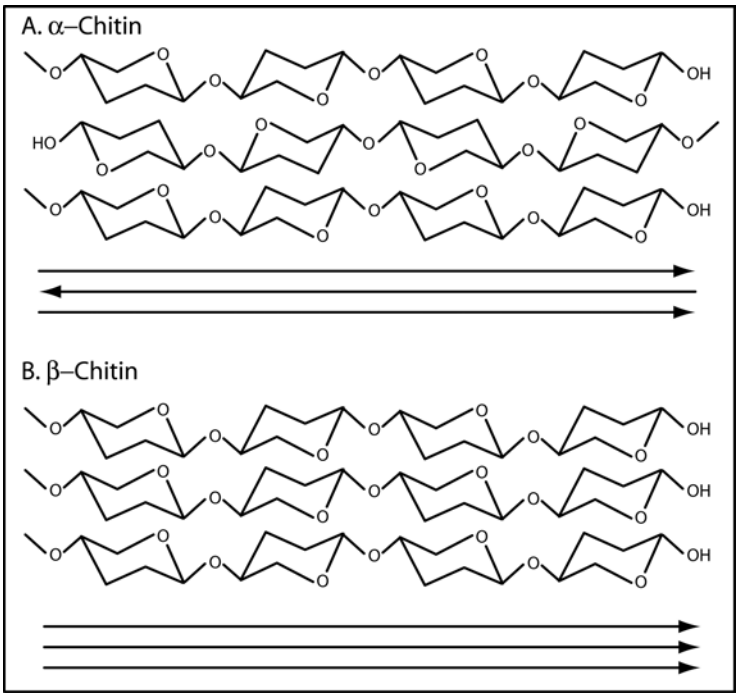

Figure 4.10 Molecular structure of α- and β-chitin (from Tharanathan & Kittur [176]).

Chitin is a major source of surface pollution in coastal areas, in particular in zones of intensive seafood production such as in Asia (Japan, China, Korea, and India), North America (USA, Mexico), Chile or Norway. The crab and shrimp shells can however be used as a bioresource. It is estimated that more than 80,000 tons of chitin is extracted per year from the marine waste [130]. The industrial treatment of

chitin has been ongoing for several decades for various industrial purposes, such as the extraction of carotenoids that cannot be synthesized artificially, or the extraction of chitin itself. Yet, chitin *per se* has relatively few applications and it is as a raw resource material for chitosan that it shows major industrial interests [176].

Chitosan, a partially deacetylated form of chitin

Chitin does not exist in the environment as a 100% acetylated polymer but as a co-polymer of randomly acetylated and deacetylated glucosamine residues with varying degree of *N*-acetylation and polymerization, depending on the biological source of the polysaccharide. When the number of *N*-acetylglucosamine units is below 50%, the co-polymer is named chitosan (Figure 4.9B). Chitosan is in fact a collective name representing a family of deacetylated chitins, deacetylated to different degrees. Chitosan is the major component in the cell wall of fungi. The extraction of chitosan directly from fungi is however dwarfed by several orders of magnitude compared to the production of chitin from seafood waste [176].

Although both chitin and chitosan share a similar molecular structure, chitosan is water-soluble under weakly acidic conditions and forms gels with a large range of viscosity. It is highly versatile because of the powerful chelating capacity of the free amino group. The viscosity is directly linked to the degree of acetylation, where the reduction in number of acetyl group leads to an augmentation in solubility. This major property has shown increasing interests over the last decades and chitosan is now used in countless industrial applications in domains as varied as photography, cosmetics, food and nutrition, agriculture, waste-water treatment, textile finishing, cements, heavy metal chelation, and in medical or veterinary sciences [53, 141]. Because of those highly attractive applications and the availability of cheap chitin sources, the development of production processes of chitosan by a controlled deacetylation of chitin is a very active field of research. Chemical deacetylation of chitin is economically achieved by chemical processes although the degrees of deacetylation and polymerization are difficult to control. Recently, new interest has been shown for the development of enzymatic alternatives, i.e., with the use of fungal or bacterial chitin deacetylases, that could promote a better controlled deacetylation of chitin [84, 180].

4.4.2 Chitin hydrolysis and diversity of chitinolytic microorganisms in the environment

In fungi, protozoa and invertebrates, chitin degrading enzymes are involved in morphogenesis and in plants, chitinases are involved in defense mechanisms. However, most chitinolytic bacteria rely on chitin hydrolysis for nutritional needs as it is a rich source of carbon and nitrogen in the marine environment, or as a preying technique in soil to lyse fungal hyphae and obtain thereby other nutrients from the fungal cytoplasm in addition to chitin itself [36, 130].

Chitin hydrolysis in bacteria results from a succession of enzymatic reactions which promote the depolymerization of the homo- or co-polymer of chitin/chitosan into single units of either *N*-acetyl-glucosamine (GlcNAc) or *N*-glucosamine (GlcN), respectively (Figure 4.11). Chitin hydrolysis is mediated outside the cell environment by a synergetic and/or sequential action of endochitinases (reaction 1; EC 3.2.1.14) and exochitinases or chitobiohydrolases (reaction 2; EC 3.2.1.52), which release GlcNAc multimers ($GlcNAc_n$) and dimers ($GlcNAc_2$), respectively. β-N-Acetylglucosaminidases (reaction 3) further hydrolyze the dimer into monomers of GlcNAc or release GlcNAc units from chitooligosaccharides ($GlcNAc_n$), in a sequential process starting from the reducing end of the polymer. Some organisms degrade $GlcNAc_2$ to GlcNAc and GlcNAc-1P by $GlcNAc_2$-phosphorylases (reaction 4) or convert the dimer to GlcNAc-6P-GlcNAc by a $GlcNAc_2$-phosphotransferase system (reaction 5) followed by the degradation to GlcNAc and GlcNAc-6P mediated by 6-phospho-β-glucosaminidases (reaction 6). Another pathway for chitin degradation is suggested to occur via the deacetylation of chitin into chitosan by chitin deacetylases (reaction 7). Chitosan is then degraded into glucosamine (GlcN) by endo-type chitosanases (reaction 8) in cooperation with exo-β-D-glucosaminidases (GlcNase; reaction 9) [51, 159, 172]. All aminosugar monomers are finally transported inside the cytoplasm where they will be integrated in the general metabolic pathways yielding to both energy production and cell compound synthesis. One alternative pathway has been recently proposed to occur in the chitinolytic archaeon *Thermococcus kodakaraensis*. Tanaka *et al.* [172] have suggested that chitin hydrolysis might be supported by the hydrolysis of the chitin polymer into $GlcNAc_2$, followed by the deacetylation of $GlcNAc_2$ into $GlcN_2$ and final split of $GlcN_2$ into monomers of

GlcN by an exo-β-glucosaminidase, although, to date, no evidence of the presence a chitin deacetylase has been found.

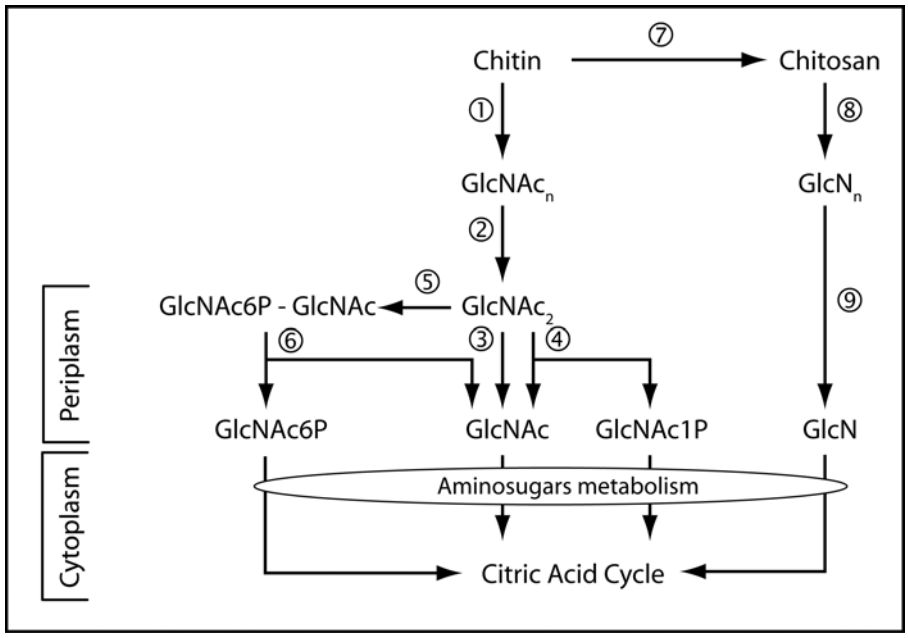

Figure 4.11 Know pathways for the catabolic degradation of chitin and chitosan into monosaccharides for nutritional requirements. Enzymes are displayed as: 1, endochitinase; 2, exochitinase; 3, *N*-acetylglucosaminidase; 4, GlcNAc2 phosphorylase; 5, GlcNAc2 phosphotransferase system; 6, 6-phospho-β-D-glucosaminidase; 7; chitin deacetylase; 8; chitosanase; and 9, glucosaminidase. Legend for the intermediates: GlcNAc$_n$, N-acetylchitooligosaccharide; GlcN$_n$, chitosan oligosaccharides; GlcNAc1P, GlcNAc-1 phosphate; GlcNAc6p-GlcNAc, GlcNAc-6 phosphate- GlcNAc; and GlcNAc6P; GlcNAc-6 phosphate (modified from Tanaka *et al.* [172]).

Endochitinases and exochitinases are generally found in a very large variety of organisms and can be grouped into two families of the glycoside hydrolases. The family of interest, GH18, consists of enzymes found in *Bacteria*, *Archaea*, fungi, protozoa, insects, mammals, and viruses. The second family GH19 is mainly composed of plant chitinases (subgrouped into four classes I to IV) and few enzymes isolated from bacterial species. Since chitin/chitosan is a major component of cell wall in fungi, chitinolytic bacteria have received growing attention due to their ability to lyse the hyphae of fungal crop pathogens and serve hence as biocontrol agents. Similarly, because of the implications on the understanding of global ecological mechanisms of chitin turnover in the oceans, marine bacteria have been investigated for their ability to degrade and use chitin.

Table 4.6 Phylogenetic origin of most of the microbial chitinases from family GH18 among *Bacteria* and *Archaea*, listed as such in GenBank (modified from Henrissat [62]).

Organism	Class or Order	Enzyme
Archaea		
Halobacterium sp. NRC-1	*Halobacteriales*	chitinase
Thermococcus kodakaraensis	*Thermococcales*	Multidomain chitinase
Pyrococcus furiosus	*Thermococcales*	Multidomain chitinase
Thermococcus chitonophagus	*Thermococcales*	Multidomain chitinase
Bacteria		
Aeromonas caviae	*γ-Proteobacteria*	Chitinase A
Aeromonas sp 10S-24	*γ-Proteobacteria*	Chitinase 1
Alteromonas sp.	*γ-Proteobacteria*	Chitinase A
Arthrobacter sp.	*Actinobacteria*	Chitinase A
Bacillus circulans	*Firmicutes*-Gram Positives	Chitinase A1
Bacillus licheniformis	*Firmicutes*-Gram Positives	Chitinase
Bacillus thurigiensis	*Firmicutes*-Gram Positives	Chitinase
Clostridium paraputrificum	*Firmicutes*-Gram Positives	Chitinase B
Clostridium thermocellum	*Firmicutes*-Gram Positives	Chitinase
Enterobacter agglomerans	*γ-Proteobacteria*	Chitinase A
Flavobacterium meningosepticum	*Flavobacteriales*	Endo-*N*-acetylglucosaminidase F1
Flavobacterium sp.	*Flavobacteriales*	Endo-*N*-acetylglucosaminidase F1
Janthinobacterium lividum	*β-Proteobacteria*	Chitinase
Kurthia zopfii	*Firmicutes*-Gram Positives	Chitinase
Microbulbifer degradans	*γ-Proteobacteria*	Chitinase B
Pseudomonas sp. PE2	*γ-Proteobacteria*	Chitinase A
Ralstonia sp. A-471	*β-Proteobacteria*	Chitinase A
Rhodothermus marinus	*Sphingobacteria*	Family 18 chitinase
Serratia liquefaciens	*γ-Proteobacteria*	Chitinase B
Serratia marcescens	*γ-Proteobacteria*	Chitinase B
Stenotrophomonas maltophilia	*γ-Proteobacteria*	Chitinase A
Streptomyces coelicolor	*Actinobacteria*	Chitinase
Streptomyces erythraeus	*Actinobacteria*	Chitinase C
Streptomyces olivaceoviridis	*Actinobacteria*	Exo-chitinase
Streptomyces plicatus	*Actinobacteria*	Chitinase-63
Streptomyces thermoviolaceus	*Actinobacteria*	Chitinase
Vibrio furnissii	*γ-Proteobacteria*	Chitodextrinase
Vibrio harveyi	*γ-Proteobacteria*	Chitinase A

However, general information on the diversity of microbial chitinases from family GH18 has been prominently restricted to few culturable microorganisms belonging to *β*- and *γ-Proteobacteria*, *Actinobacteria*, Gram-positives and *Archaea* (Table 4.6) that were primarily obtained due to their ability to grow on chitinous substrates [27, 28, 62]. Nevertheless, the use of chitinase genes as molecular markers have demonstrated that large pools of uncultured GH18 chitinases-producing micro-

organisms still remain to be discovered in marine and soil environments [27, 28, 113, 197] (**Paper IV**).

4.4.3 Molecular mechanism of chitin hydrolysis and properties of GH18 chitinases

Family GH18 gathers a large number of enzymes catalyzing the hydrolysis of the β-(1,4) bonds between GlcNAc residues in chitin or chitinous substrates, i.e., chitinase, chitodextrinases, exo-chitinases, di-N-acetylchitobiosidases (EC 3.2.1.30), endo-N-acetylglucosaminidases (EC 3.2.1.96). It also contains proteins of unknown functions such as mammalian oviduct glycoproteins, unidentified proteins obtained from the *Caenorhabditis elegans* genome, or plant proteins such as concanavalin B and narbonin which are devoid of catalytic activity (for an extensive review, see Henrissat [62]).

Microbial GH18 chitinases generally consist of multiple functional domain proteins, like a vast majority of glycoside hydrolases. In addition to one or sometimes several catalytic modules, GH18 chitinases contain chitin-binding modules and in some cases fibronectin type III-like modules [62]. The catalytic domains of GH18 chitinases adopt a $(\beta/\alpha)_8$- or TIM-barrel structure, identical to the one found in GH13 α-amylases [131]. Catalytic domains of GH18 chitinase can be further classified into 3 subfamilies, subfamilies A, B and C, based on the presence of conserved amino acid motifs around the catalytic residues [193]. Figure 4.12 shows multiple alignments of the most conserved motifs for each subfamily. While linking phylogenetic information with the occurrence and distribution of the three subfamilies among microbes, Suzuki *et al.* [165] have demonstrated that bacterial chitinases are clearly segregated between the three subfamilies. Subfamily A gathers the largest diversity of phylogenetically-distinct species, whereas subfamilies B and C appear to be more rare and consists only of limited number of species and genera. Interestingly however, some bacterial species produce multiple chitinases that belong to two or more subfamilies of catalytic domains [165]. This may suggest with high probability that those bacteria have inherited the genes via lateral gene transfer.

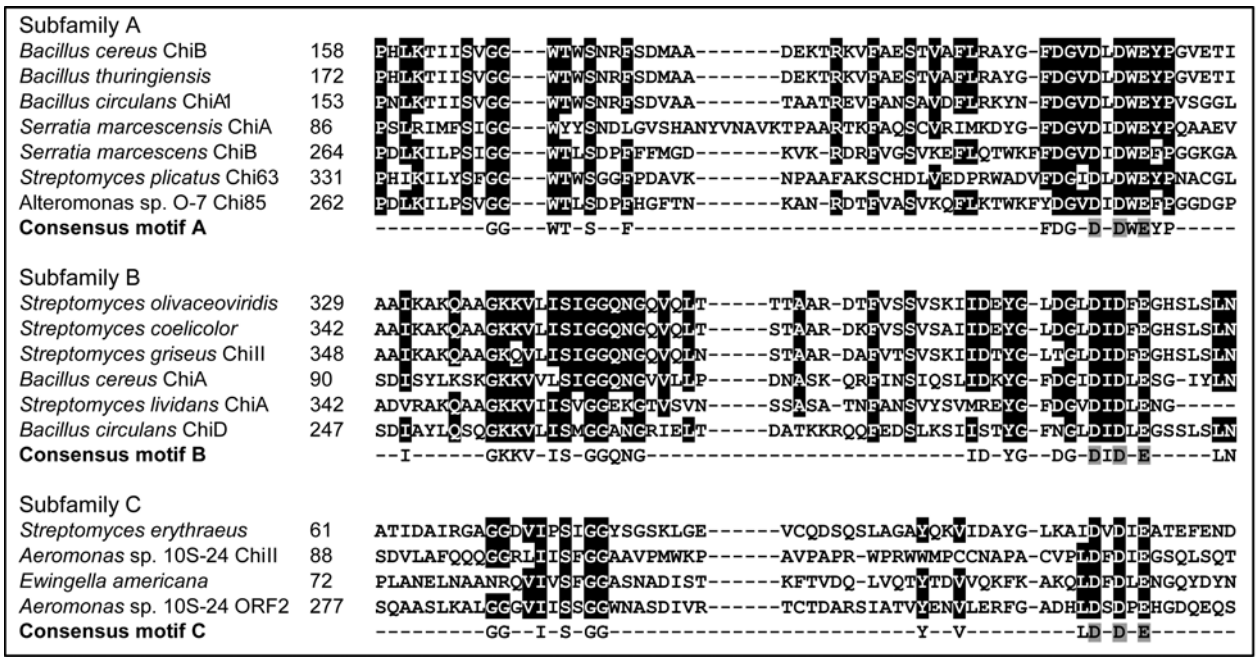

Figure 4.12 Sequence comparison of typical conserved regions of the catalytic domains in subfamilies A, B and C of the catalytic domains in GH18 chitinases. Numbers in each line indicate the position of amino acids for each polypeptide. Highly conserved residues are highlighted in black. The consensus sequence with the respective typical residues for each subfamily is given at the bottom of each multiple alignment and the three Asp, Asp and Glu residues involved in the hydrolysis of the β-(1,4) bond are shaded in grey in the consensus sequence.

GH18 chitinases have been intensively studied to determine the identity and role of the amino acid residues responsible for the hydrolysis of the β-(1,4) glycosidic bond, the residues that promote the binding of the substrate inside and around the catalytic cleft, and consequently the molecular mechanisms occurring during the hydrolysis of the bond [46, 69, 167, 195, 194, 193]. The most prominent motif of GH18 chitinases is DxDxE which lies on strand β4 of the TIM barrel (that will be named here Asp1, Asp3 and Glu5 for the sake of clarity). The Glu5 residue acts as the catalytic acid in a retaining mechanism. However, unlike other retaining GHs, chitinase from family 18 lack a carboxylate that is properly positioned for acting as nucleophile and the hydrolysis is performed via a substrate-assisted mechanism [69]. As shown in Figure 4.13, a putative oxazolinium ion intermediate is stabilized by an anchimeric assistance of the sugar *N*-acetyl group after donation of a proton from the catalytic carboxylate (Glu5) to the leaving group. Simultaneously, the first aspartate Asp1 assists both Asp3 and Glu5 in the hydrolysis by providing a negative charge and hence keeping both Asp3 and Glu5 protonated. The mechanism can thus rationalize

the anomer retaining reaction of the enzymes without a second carboxylate, as usually found in classic retaining enzymes [167].

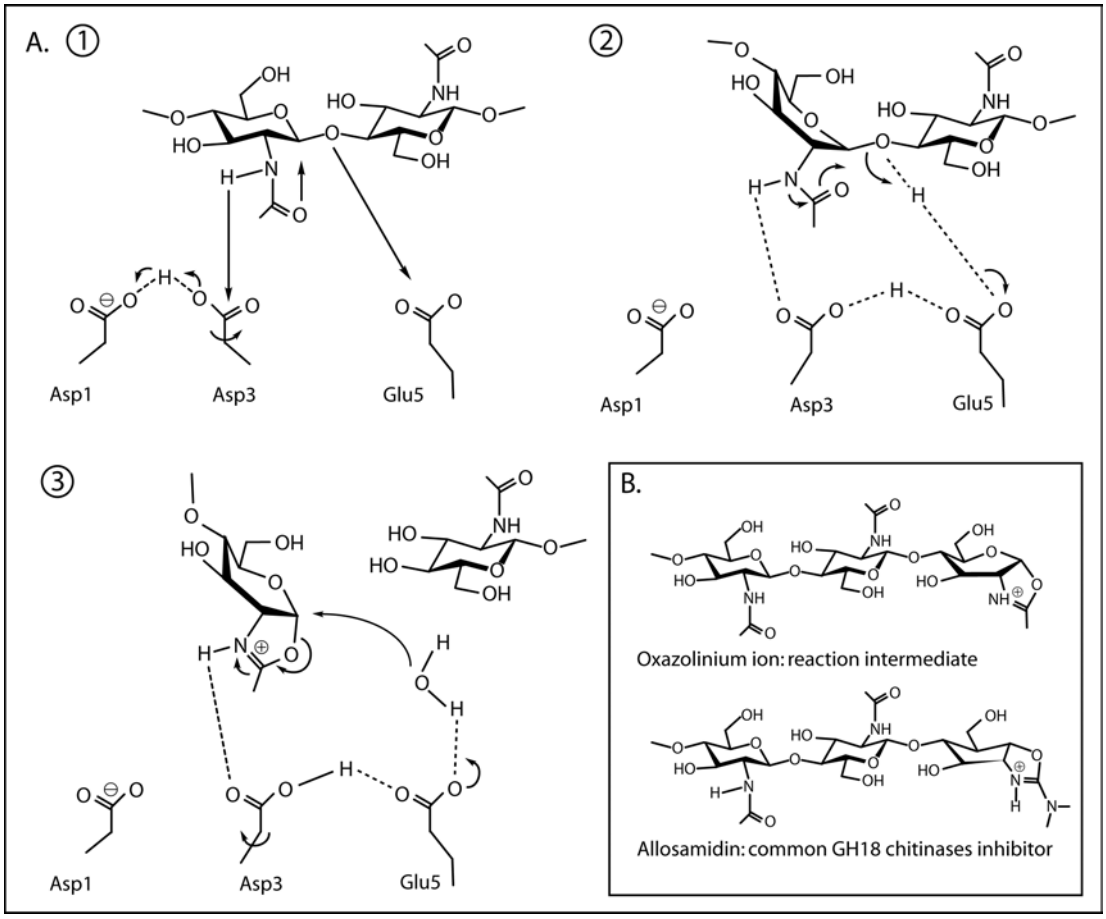

Figure 4.13 (A) Catalytic mechanism of family 18 chitinases. 1, substrate binding; 2, substrate distortion and protonation of the glycosidic bond; 3, formation of the oxazolinium ion intermediate and leaving group departure – a water molecule destined to hydrolyze the oxazolinium ion intermediate is approaching the anomeric carbon. **(B)** Compared molecular structures of the oxazolinium ion reaction intermediate (positively charged) and a general inhibitor of family GH18 chitinases, allosamidin (modified from Vaaje-Kolstad [183]).

The catalytic module contains aromatic residues on its surface (mostly Trp and Tyr) which are not involved in the hydrolysis of chitin but the binding and containment of the substrate around and inside the catalytic cleft [167, 194]. Interestingly, substantial data has proved that crystalline chitin and chitooligosaccharides ($GlcNAc_n$) are not hydrolyzed by GH18 chitinases following a similar mechanistic procedure [195]. The active site in chitinases is composed of a long, tunnel- or cleft-shaped pocket. Aromatic residues strategically placed in the axis of the pocket provide hydrophobic interactions with the pyranose rings of the polymer and thus stabilize the entry of the crystalline chitin microfibrils, reducing-end first, inside the catalytic cleft

[184], Trp and Tyr residues located inside the pocket also having an important role by stabilizing the sugar rings around the catalytic Glu5 [195]. However, according to the model recently proposed by Watanabe and colleagues [195], chitooligosaccharides may enter the catalytic side via the opposite side of the pocket, the hydrolysis of the bond occurring not from the reducing end of the oligomer but from the non-reducing end. Mutagenesis study demonstrated notably that those conserved Trp and Tyr residues inside the catalytic pocket are not involved in the binding of chitooligosaccharides.

The optimum pH of chitinases are pH 4-9 for higher plants and algae, pH 4.8-7.5 for animals and 3.8-8.0 for microorganisms although the values appear to be substrate-dependent. Allosamidin is a general inhibitor of chitinases (Figure 4.13D). It has a similar structure as the oxazolinium ion intermediate of the substrate and is only inhibiting GH18 chitinases because of their highly peculiar hydrolytic mechanisms. GH19 chitinases will not be affected by the addition of this compound. Metal ions sometimes also inhibit or decrease the catalytic activity. Mercury II (Hg^{2+}) and silver II (Ag^{2+}) are commonly inhibiting chitinase activity but copper II (Cu^{2+}), however, can either inhibit or activate the hydrolytic action of GH18 chitinases [90].

The study and classification of the hydrolytic properties of GH18 chitinases is quite difficult given the large variety of activity which they promote. In fact, GH18 chitinases show a very large diversity of substrate specificities, some hydrolyzing virtually any type of chitinous substrates (chitin, chitosan, crystalline chitin, colloidal chitin, chitooligosaccharides, etc.) or even non-chitinous substrates like cellulose, or hydrolyzing only a few or limited number of specified compounds. Also, this large span of substrate specificity is accompanied by an equal range of hydrolytic behavior, i.e., endo-acting enzymes, exo-acting enzymes from either reducing- or non-reducing end, exo-enzyme releasing chitobiose or chitotriose as a final product, etc. Analysis of the amino acid sequence alone gives only little information and it is only by classical biochemical characterization that the properties of these enzymes can be determined. Nevertheless, as mentioned here above, GH18 chitinases are inclined to have a different catalytic behavior according to the nature of the substrate which is used to characterized them (**Paper V**) and the characterization of these properties should be then performed with care throughout the use of a large variety of substrates. Because of this apparent diversity in substrate-related behavior the classification of chitinases as either endochitinase or exochitinase, solely on the basis of catalytic patterns with

few types of substrates is therefore doubtful. The lack of consensus in nomenclature brings additional difficulty and no general naming system has been adopted to date. The universal naming as family 18 chitinase, or GH18 chitinase, as suggested by Henrissat [62], would thus be most appropriate although less informative.

4.4.4 GH18 chitinases from thermophiles and their applications

Due to the ecological role of chitinases and growing interests in biotechnology such as for chitinous waste recycling [43], chitooligosaccharides production [164], or crop protection [130], a large number of chitinases from microorganisms have been isolated, or their genes cloned and expressed. The study of this growing number of enzymes has emphasized a wide diversity of biochemical properties. Some chitinases have been shown to be moderately thermostable, e.g., with optimal temperature at around 50 to 60°C [13, 122, 150, 169, 181]. Until now these relatively thermostable chitinases have been mainly isolated from mesophilic or moderately thermophilic microorganisms such as *Bacillus* spp. [13, 150, 169], *Microbispora* spp. [122], *Streptomyces thermoviolaceus* [181]. Their enzymes do not show resistance to long exposure to heat, resistance to pH variations or resistance to denaturants typically encountered by enzymes from thermophilic microorganisms [40, 190, 200].

To date, the information regarding the distribution of chitinases among thermophiles is limited and only a few true thermophilic chitinases have been isolated or detected in thermophiles or hyperthermophiles. The chitinase ChiA from the moderately halophilic *Rhodothermus marinus* PRI378 [2] is the only representative of highly thermostable GH18 chitinase among *Bacteria* (**Paper V)**. The chitinase was detected by sequencing of the whole genome and subsequently retrieved, cloned and expressed in *E. coli*. Further analysis of the genome of *R. marinus* showed that a β-*N*-acetylglucosaminidase gene was also present on the chromosome (Hobel *et al.*, unpublished data). Both enzymes would thus possibly act in a synergistic manner for the hydrolysis of chitin into GlcNAc of GlcN and provide hence a metabolic pathway for the uptake of carbon and nitrogen from chitinous substrates. However, experiments in the laboratory were unsuccessful to obtain growth under various cultivation conditions using chitinous compounds as the only carbon source. Bacterial chitinase gene expression *in vivo* can be regulated under several nutritional conditions that appear to be species-specific [43]. It is very likely that none of the conditions that

we used to cultivate *R. marinus* did indeed meet the bacterium's requirements for chitinase expression and hence prevented growth.

Chitinases from hyperthermophilic *Archaea* belonging to the *Thermococcales*, i.e., *Thermococcus kodakaraensis*, *Thermococcus chitonophagus*, and *Pyrococcus furiosus* represent so far the most extremely thermostable chitinases of GH18. They have optimal temperatures of about 90°C or more, and half-lives of 30 min up to 1 h at 120°C [4, 6, 5, 48, 170, 171]. *Pyrococcus furiosus* and *Thermococcus kodakaraensis* contain both very similar multidomain chitinases which appear to be unique among the family GH18, both in terms of size and catalytic activity. The *P. furiosus* gene is however interrupted by 37 nucleotides, which splits the encoded protein in two parts. Nevertheless, analysis of the gene and further characterization of the protein showed that both parts together retained a behavior significantly similar as the single one found in *T. kodakaraensis* [48]. Those multidomain chitinases consist of two active sites located at both N-terminal and C-terminal ends of the protein, separated by two or three identical chitin-binding modules, all linked by Pro-, Thr- and Ser-rich linkers. The size of the proteins reaches 1,200 amino acids and the calculated molecular weight is over 140,000 Da. The catalytic and non-catalytic modules of both chitinases show very high sequence similarity between each other (≥70%-80%). Heterologous expression of the enzymes in *E. coli* demonstrated that one of the catalytic domains was acting as an endochitinase, hydrolyzing β-(1,4) bonds inside chitin oligomers, whereas the other module acted as an exochitinase, releasing mostly chitobiose and partially chitotriose from the non-reducing end of the chitooligosaccharides. The authors established that both catalytic domains were working synergistically for the hydrolysis of chitin [48, 171]. Interestingly however, the publications describing the two chitinases reported opposite results for each catalytic module, i.e., in terms of endo- vs. exo-activity, although the high sequence similarity figures would suggest an identical behavior. The presence of these chitinases in the genomes was further confirmed by growth of both strains on chitin as the only nutrient source [48, 171].

In 1995, *Thermococcus chitonophagus* was described as the first archaeon that uses chitin for growth [75]. Since then, at least two chitinases have been isolated or partially sequenced from this organism. Andronopoulou & Vorgias [4] isolated from cell extracts a hyperthermostable chitinase which is also insensitive to allosamidin and denaturants. The chitinase Chi70 displayed a wide substrate specificity, being able to

hydrolyze cellulose and release cellobiose [4]. The same group also reported recently another gene fragment, detected in a genomic library [6]. The gene fragment showed high DNA sequence similarity with the gene encoding for the multidomain chitinases in *T. kodakaraensis* and *P. furiosus*. But since no complete gene had been obtained, there was little data on possible identity and activity of this gene. In my studies however, I identified the same gene in *T. chitonophagus* by PCR using degenerate oligonucleotidic primers, sequenced it but was not able to get expression successfully (Hobel *et al.*, unpublished data). Sequence analysis demonstrated an identical multidomain pattern as observed in *T. kodakaraensis* and *P. furiosus*. The gene encoded a protein of over 1,200 amino acids, with an estimated molecular weight of 143,000 Da. The protein structure followed the chitinase from *T. kodakaraensis* but the linkers between all modules were almost exclusively composed of Pro and Thr residues only. As observed previously, the catalytic and non-catalytic domains of the multidomain chitinase of *T. chitonophagus* show high amino acid sequence similarity. I considered to established a putative hydrolytic behavior *in silico*, based on the data obtained from Tanaka *et al.* [171] and Gao *et al.* [48] for the chitinases of *T. kodakaraensis* and *P. furiosus*, respectively. However, the contradiction in their respective results brings possible doubts and therefore these results cannot be used as a basis for assigning activities to the *T. chitonophagus* enzyme.

5. SUMMARY OF THE PRESENT INVESTIGATION

This thesis describes a multidisciplinary approach for the study and characterization of the microbial genetic diversity in thermophilic microbial ecosystems in Iceland. The ease of access to terrestrial hot springs fields in Iceland in conjunction with the unique wealth in hot springs variety, i.e., with wide ranges of temperatures, pH and chemical compositions, grant significant assets for sampling and studying an extensive diversity of thermophilic microbes.

Sampling substantial amounts of biomass can sometimes be difficult because of environmental conditions that limit the access to the hot springs. For instance, subterranean or submarine milieux are extreme environments and for obvious reasons microbial mat samples are not easily obtained there from. The rich and diversified microbial world that thrives in those environments remains hence largely unknown. Sampling strategies were thus attempted or improved aiming at increasing our capabilities to obtain larger quantities of biomass, and therefore enabling us to evaluate the scope of microbial diversity in these environments (**Papers I and II**).

Paper I describes for the first time the access and analysis of the microbial biosphere in subterranean hot springs in Iceland. Iceland is one of the few countries in the world relying on the use of geothermal energy for heating and electricity production. Also, the boreholes used to pumping hot water from the ground can be considered as windows into the springs that enable an easy access to the thermophilic biomass. Large volumes of water from hot water reservoirs and hot tap water were filtered and the biomass obtained thereby was analyzed by microscopy, enrichment cultures, and direct molecular analysis of 16S rRNA genes. The results demonstrated that the underground thermophilic biosphere was rich, with species belonging to both bacterial and archaeal kingdoms. Although most of the species identified were closely related to surface terrestrial representatives, the microbial diversity in the wells was mainly indigenous. Our findings confirmed furthermore that thermophilic microorganisms may disseminate through the Earth's crust via subsurface water channels, a phenomenon which was previously observed in cold aquifers.

In **Paper II**, the microbial diversity from submarine hot springs located in the North of Iceland was evaluated. The Grimsey hydrothermal vent field has been

recently discovered and was shown to be unique in many aspects, such as high-temperature venting or the presence of large deposits of pelagic and glacial sediments [56]. Recovering microbial samples from hydrothermal vents require the use of submarine vehicles but also the design of sampling strategies that avoid contamination and dilution of the samples by the surrounding seawater. Two woolen nests were used as *in situ* colonization devices providing attachment surfaces for microbes. A solid matrix amended with a 0.25% wt/vol yeast extract and tryptone in one of the nests in order to promote the local growth of heterotrophic microorganisms. The samples were deployed in a zone of 50°C to 80°C percolating water and kept *in situ* for over 50 hours. Following colonization, the biomass was concentrated and the scope of microbial diversity was examined by 16S rRNA analysis. A majority of ε-*Proteobacteria* and *Thermococcales* were recovered from the samples by analysis of the bacterial and archaeal diversity, respectively. The nutrient-containing wool nest supported the growth of 16 out of 17 non-ε-proteobacterial phylotypes. The inferred phylogenetic affiliation of the phylotypes suggested a large dominance of thermotolerant, thermophilic and hyperthermophilic species in both samples. My results were in accordance with previous observations established in several hydrothermal sites. Yet, the low similarity of the detected ε-proteobacterial phylotypes with precedently reported sequences strongly suggested that the microbial flora in the Grimsey vent field was endemic to the site and did not represent the expression of a local selection of globally widespread species. This experiment also demonstrated the efficiency of this sampling methodology, which has been used successfully since then in various terrestrial and submarine hot spring environments (Marteinsson *et al.*, personal communication).

Because a very large majority of bacteria are still "as-yet-uncultivated" due to the lack of information regarding their needs in terms of nutritional requirements and growth conditions, I focused on the development and implementation of alternative enrichment strategies aiming at improving our ability to grow microbes and also increasing the access to specialized microbes. Significant efforts were made to generate a rich diversity of microbes producing hydrolytic enzymes with high potentials for biotechnological applications such as α-amylases (**Paper III**) or chitinases (**Paper IV**).

Paper III describes the development of an *in vitro* directed enrichment method that was used to obtain starch-degrading microorganisms from a sulfide-rich hot spring. The hot spring under study had been previously examined in our laboratory via direct 16S rRNA gene analysis by Skirnisdottir *et al.* [156], who showed that the hot spring microbiota was dominated by few bacterial and archaeal species. By using untreated hot spring water and low concentrations of nutrients, i.e., 0.1% or 0.02% wt/vol starch and 0.005% wt/vol yeast extract, I attempted to prevent common opportunistic strains of *Thermus* and *Bacillus* to overgrow in enrichment cultures (cf. § 3.1.2) and to promote thereby the enrichment of slow-growing, starch-degrading thermophilic microbes. Six cultures were conducted under either aerobic or anaerobic condition and the microbial diversity therein was assessed by 16S rRNA gene analysis. Seventeen bacterial phylotypes were obtained scattering among *Thermus-Deinococcus* group, green non-sulphur bacteria, gram positives, and uncultivated new candidate divisions. No *Archaea* were found. The DNA from 4 cultures was further screened by PCR for amylases genes from family 13 of the glycoside hydrolases. A total of 18 different gene fragments were recovered, including α-amylases, α-glucosidases, 1,4-α-glucan branching enzymes, cyclomaltodextrin hydrolases, maltogenic amylases and neopullulanases, and unspecified family 13 glycoside hydrolases. A control experiment was performed by screening, under identical conditions, the DNA sample primarily used by Skirnisdottir and coworkers to examine the natural diversity in the hot spring [156]. Only one pullulanase was detected in the hot spring biomass demonstrating the efficiency of the directed enrichment method. Moreover, this experiment proved that a combination of enrichments and cultivation-independent studies could give a more comprehensive view of the overall microbial diversity in hot springs.

The Hveravik geothermal field is located within the intertidal zone on the seashore of the Reykjanes peninsula in North-West Iceland. It can be regarded as one of the most extreme environment on Earth. The water originates on the mainland and is essentially deep fresh groundwater. The water is alkaline, contains low-sulfide concentrations and is only slightly mixed with seawater in the hot spring outlets. The temperature in the hot-water outlets ranges from 45°C to 95°C in different hot springs. The tides can be as large as 4 m in this area and therefore the organisms living in those hot springs can be subject to almost 100°C fluctuation twice per day in addition

to dramatic changes in salinity, light and oxygen concentration. This type of habitat is quite rare, although they are found in few other places in Iceland [66]. The global microbial diversity therein had been poorly studied so far. The aim of this experiment was to examine the microbial diversity and to promote high chitinase gene diversity in this extreme environment **(Paper IV**). Little biomass was directly accessible by the hot springs ducts and the constant flushing of the tides was calling for alternative sampling methods to gain sufficient biomass. Colonization experiments were thus conducted by inserting two woolen nests in two different hot springs. Simultaneously, three *in situ* directed enrichments were performed in three hot springs, including one in which a colonizing device was inserted. Raw chitin was used as single carbon source in polyethylene flasks filled with hot fluid. Finally, samples of a green/orange bacterial mat were collected. The six samples had in total 38 bacterial phylotypes as found by 16S rRNA gene analysis, related to both marine and terrestrial, thermophilic, mesophilic and psychrophilic microorganisms and scattering among 11 bacterial division. No *Archaea* were detected. The use of temperature loggers showed that the samples were subject to dramatic temperature shifts that followed the tidal movements. Chitin promoted the growth of substantially different microbial communities in the enrichments. This was confirmed by PCR screening of all DNA samples for the presence of chitinase genes from family GH18. A total of 36 different chitinase gene fragments were detected and comparison between the samples showed that the enrichments contained in average twice as many chitinase genes than the other samples. As emphasized previously, the combined methodologies enabled not only to gather inaccessible microbes but also bring substantial information regarding the diversity and distribution of catabolic genes among thermophiles.

In **Paper V**, I report on the cloning, expression and characterization of a highly thermostable chitinase ChiA from the halophile and thermophile *Rhodothermus marinus* PRI378. *R. marinus* has been previously isolated from the Hveravik intertidal hot springs [2]. *R. marinus* has been shown to be a rich source of various hydrolytic enzymes, including e.g., a xylanase [125], a cellulase [55] or a β-mannanase [132]. The partial *chiA* gene sequence was primarily detected in a Prokaria Ltd. in-house genome database. I proceeded to the recovery of the complete gene by chromosome walking. The complete gene *chiA* and a signal-peptide-lacking mutant *chiAΔsp* were cloned and expressed in *E. coli*. The gene encoded for a protein of 377 amino acids with a calculated molecular weight of 42,341 Da. Sequence analysis

showed that the chitinase was non-modular and belonged to the GH18 family. Expression was poor for the wild-type enzyme but good for the mutant ChiAΔsp. Preliminary biochemical experiments suggested that the deletion of the signal peptide did not affect the activity. The complete characterization of the enzyme was hence performed using only the mutant chitinase. The purified chitinase had an optimum temperature of 70°C and an optimum pH between 4.5 and 5. The enzyme was highly thermostable, maintaining 100% activity after 16 h incubation at 70°C, and had half-lives of 3 h at 90°C and 45 min at 95°C. Several chitinous substrates were used to establish the catalytic pattern of the chitinase. ChiAΔsp was an endochitinase, releasing chitobiose as a major end-product. However, the enzyme acted as exochitinase when using chitin oligomers shorter than 5 GlcNAc units as substrate. Chitosan polymer (73% deacetylated) was hydrolyzed, offering an attractive enzymatic process for the production of valuable chitooligosaccharides at high temperature and low pH [164]. The chitinase from *R. marinus* was found to be the most thermostable family GH18 chitinase known among *Bacteria*.

6 CONCLUDING REMARKS

The main goal of this project was to conduct a multidisciplinary approach to study the undiscovered microbial diversity and hence genetic diversity in thermophilic environments. In short terms, “from the hot springs to the enzymes”.

In view of the results presented here, the general perception that the majority of microbes are just non-culturable should be redefined as “as-yet-uncultured”. It is evident that numerous unknown chemical, physical and biological factors are necessary for microbes in order to live and divide in the environment. The approach described here has demonstrated that conducting enrichments *in situ* or reproducing nature-like growth conditions in the laboratory brings significant advantages to promote microbial growth. This was made possible by applying simple procedures such as the use of a neutral colonization support inserted or maintained inside hot springs, or the use of untreated hot spring geothermal fluid for *in vitro* enrichment cultures. Such strategies enable to obtain larger quantities of biomass which in turn can be better analyzed with all possible microbiology and phylogenetics techniques.

These alternative sampling techniques were designed primarily as a base for the discovery of new genes with high potential for biotechnological applications. They were successfully used for the promotion and selection of starch-degrading as well as chitin-degrading microbes from thermophilic microorganisms. The α-amylases and chitinases discovered hereby were in a large majority new. This underlines therefore the concept that most of the valuable genetic in nature still remains to be discovered. Concurrently, the detection of those genes brings also valuable information as little is known regarding the extent of the distribution and the ecological significance of these metabolic genes among thermophiles in the environment.

Gene discovery and new enzyme development can also be easily and elegantly achieved with cultured strains. The sequencing of genomes allows a significant access to known and unknown genes and the main advantage of this methodology is that the amount and quality of DNA that can be used is not a limiting factor. As an example, I describe in this thesis the cloning and expression of a highly thermostable chitinase from the halophilic and thermophilic bacterium *Rhodothermus marinus* which

genome has been partially sequenced in our laboratory. The ease to recover the complete gene and its expression emphasize the rapid access to new biological material for potential biotechnological applications. Various strategies could also be used to link genotypic and phenotypic information, and establish for instance the role of the cloned chitinase in the global metabolic behavior of *R. marinus*.

The information on microbial diversity in the environment has been largely improved by the development of molecular phylogenetic techniques over the years. However, the results gathered in this thesis strongly support that it is by the combination of several approaches of traditional microbiology with state-of-the-art molecular biology techniques that we can successfully gain access to new microbes and hence novel valuable genes, rather than by using either method alone.

7. REFERENCES

[1] Alain K, Querellou J, Lesongeur F, Pignet P, Crassous P, Raguenes G, Cueff V & Cambon-Bonavita MA (2002) *Caminibacter hydrogeniphilus* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium isolated from an East Pacific Rise hydrothermal vent. *Int J Syst Evol Microbiol* 52: 1317-23.

[2] Alfredsson G, Kristjansson J, Hjörleifsdottir S & Stetter K (1988) *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland. *J Gen Microbiol* 134: 299-306.

[3] Amann RI, Ludwig W & Schleifer K-H (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.

[4] Andronopoulou E & Vorgias CE (2003) Purification and characterization of a new hyperthermostable, allosamidin-insensitive and denaturation-resistant chitinase from the hyperthermophilic archaeon *Thermococcus chitonophagus*. *Extremophiles* 7: 43-53.

[5] Andronopoulou E & Vorgias CE (2004) Multiple components and induction mechanism of the chitinolytic system of the hyperthermophilic archaeon *Thermococcus chitonophagus*. *Appl Microbiol Biotechnol* 65: 694-702.

[6] Andronopoulou E & Vorgias CE (2004) Isolation, cloning, and overexpression of a chitinase gene fragment from the hyperthermophilic archaeon *Thermococcus chitonophagus*: semi-denaturing purification of the recombinant peptide and investigation of its relation with other chitinases. *Protein Expr Purif* 35: 264-71.

[7] Bachofen R, Ferloni P & Flynn I (1998) Review: microorganisms in the subsurface. *Microbiol Res* 153: 1-22.

[8] Barns SM, Fundyga RE, Jeffries MW & Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* 91: 1609-1613.

[9] Beeder J, Nilsen RK, Rosnes JT, Torsvik V & Lien T (1994) *Archaeoglobus fulgidus* isolated from hot North Sea oil field waters. *Appl Environ Microbiol* 60: 1227-1231.

[10] Bernard FP, Connan J & Magot M (1992) Indigenous microorganisms in connate water of many oil fields: a new tool in exploration and production techniques, SPE 24811. In: Annual technical conference and exhibition of the Society of Petroleum Engineers.

[11] Bertoldo C & Antranikian G (2001) Amylolytic enzymes from hyperthermophiles. *Method Enzymol* 330: 269-289.

[12] Bertoldo C & Antranikian G (2002) Starch-hydrolyzing enzymes from thermophilic *Archaea* and *Bacteria*. *Curr Opin Chem Biol* 6: 151-160.

[13] Bhushan B & Hoondal GS (1998) Isolation, purification and properties of a thermostable chitinase from an alkalophilic *Bacillus* sp. BG-11. *Biotechnol Lett* 20: 157-159.

[14] Blackwood CB, Marsh T, Kim S-H & Paul EA (2003) Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Appl Environ Microbiol* 69: 926-932.

[15] Blochl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW & Stetter KO (1997) *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of *Archaea*, extending the upper temperature limit for life to 113 degree. *Extremophiles* 1: 14-21.

[16] Bolon DN, Voigt CA & Mayo SL (2002) *De novo* design of biocatalysts. *Curr Opin Chem Biol* 6: 125-129.

[17] Boone DR, Castenholz RW & Garrity GM (2001) *Bergey's manual of systematic bacteriology - The Archaea and the deeply branching and phototrophic Bacteria*. New-York, USA: Springer-Verlag.

[18] Bourne Y & Henrissat B (2001) Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr Opin Struct Biol* 11: 593-600.

[19] Boyer JN (1986) End products of anaerobic chitin degradation by salt marsh bacteria as substrates for dissimilatory sulfate reduction and methanogenesis. *Appl Environ Microbiol* 52: 1415-1418.

[20] Brock TD & Freeze H (1969) *Thermus aquaticus* gen. nov., and sp. nov., a non-sporulating extreme thermophile. *J Bacteriol* 98: 289-297.

[21] Brock TD (1978) *Thermophilic microorganisms and life at high temperatures*. New-York, USA: Springer-Verlag.

[22] Bruce K (1997) Analysis of *mer* gene subclasses within bacterial communities in soils and sediments resolved by fluorescent-PCR-restriction fragment length polymorphism profiling. *Appl Environ Microbiol* 63: 4914-4919.

[23] Bruns A, Nubel U, Cypionka H & Overmann J (2003) Cyclic AMP and acyl homoserine lactones increase the cultivation efficiency of heterotrophic bacteria from the central Baltic Sea. *Appl Environ Microbiol* 68: 3978-3987.

[24] Carlström D (1957) The crystal structure of α-chitin (poly-*N*-acetyl-D-glucosamine). *J Biophys Biochem Cytol* 3: 669-683.

[25] Cherry JR & Fidantsef AL (2003) Directed evolution of industrial enzymes: an update. *Curr Opin Biotechnol* 14: 438-443.

[26] Corre E (2000) PhD Thesis. *Approche moleculaire de la diversite microbienne de deux environnements extremes: les sources hydrothermales profondes et les reservoirs petroliers*. Roscoff, France: Universite de Bretagne Occidentale.

[27] Cottrell MT, Moore JA & Kirchman DL (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* 65: 2553-2557.

[28] Cottrell MT, Wood DN, Yu LY & Kirchman DL (2000) Selected chitinase genes in cultured and uncultured marine bacteria in the alpha- and gamma-subclasses of the *Proteobacteria*. *Appl Environ Microbiol* 66: 1195-1201.

[29] Cowan DA (2004) The upper temperature for life - where do we draw the line. *Trends Microbiol* 12: 58-60.

[30] Crabb WD & Mitchinson C (1997) Enzymes involved in the processing of starch to sugars. *Trends Biotechnol* 15: 349-352.

[31] Crabb WD & Shetty JK (1999) Commodity scale production of sugars from starches. *Curr Opin Microbiol* 2: 252-256.

[32] Curtis TP & Sloan WT (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr Opin Microbiol* 7: 221-226.

[33] Dahllöf I (2002) Molecular community analysis of microbial diversity. *Curr Opin Biotechnol* 13: 213-217.

[34] Darland G, Brock TD, Samsonoff W & Conti SF (1970) A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. *Science* 170: 1416-1418.

[35] Davies G & Henrissat B (2002) Plant glyco-related genomics. *Biochem Soc Trans* 30: 291-297.

[36] De Boer W, Gunnewiek P, Kowalchuk GA & Van Veen JA (2001) Growth of chitinolytic dune soil beta-subclass *Proteobacteria* in response to invading fungal hyphae. *Appl Environ Microbiol* 67: 3358-3362.

[37] de Vos WM, Kengen SW, Voorhorst WG & van der Oost J (1998) Sugar utilization and its control in hyperthermophiles. *Extremophiles* 2: 201-205.

[38] DeLong EF & Pace NJ (2001) Environmental diversity of *Bacteria* and *Archaea*. *Syst Biol* 50: 470-478.

[39] DeLong EF (2004) Microbial population genomics and ecology: the road ahead. *Environ Microbiol* 6: 875-878.

[40] Demirjian DC, Moris-Varas F & Cassidy CS (2001) Enzymes from extremophiles. *Curr Opin Chem Biol* 5: 144-151.

[41] Diez B, Pedros-Alio C, Marsh TL & Massana R (2001) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine pico-eukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbiol* 67: 2942-2951.

[42] Eichler J (2001) Biotechnological uses of archaeal extremozymes. *Biotechnol Adv* 19: 261-278.

[43] Felse PA & Panda T (1999) Regulation and cloning of microbial chitinase genes. *Appl Microbiol Biotechnol* 51: 141-151.

[44] Forterre P, Brochier C & Philippe H (2002) Evolution of the *Archaea*. *Theor Pop Biol* 61: 409-422.

[45] Fujiwara S (2002) Extremophiles: Developments of their special functions and potential resources. *J Biosci Bioeng* 94: 518-525.

[46] Fukamizo T (2000) Chitinolytic enzymes: catalysis, substrate binding, and their applications. *Curr Prot Pept Sci* 1: 105-124.

[47] Galperin MY (2004) Metagenomics: from acid mine to shining sea. *Environ Microbiol* 6: 543-545.

[48] Gao J, Bauer MW, Shockley KR, Pysz MA & Kelly RM (2003) Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. *Appl Environ Microbiol* 69: 3119-3128.

[49] Giovanonni SJ, Britschgi TB, Moyer CL & Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345: 60-63.

[50] Glaeser J & Overmann J (1999) Selective enrichment and characterization of *Roseospirillum parvum*, gen. nov., sp. nov., a new purple non-sulfur bacterium with unusual light absorption properties. *Arch Microbiol* 171: 405-416.

[51] Gooday GW (1994) Physiology of microbial degradation of chitin and chitosan. In *Biochemistry of microbial degradation*. Ratledge C (ed). Dordrecht, The Netherlands: Kluwer Academic Press pp. 279-312.

[52] Grant WD (1991) General view of halophiles. In *Superbugs - Microorganisms in extreme environments*. Horikoshi K & Grant WD (eds). Tokyo, Japan: Springer-Verlag pp. 15-37.

[53] Haki GD & Rakshit SK (2003) Developments in industrial important thermostable enzymes: a review. *Biores Technol* 89: 17-34.

[54] Halldorsdottir S, Thorolfsdottir ET, Spilliaert R, Johansson M, Thorbjarnardottir SH, Palsdottir A, Hreggvidsson GO, Kristjansson JK, Holst O & Eggertsson G (1998) Cloning, sequencing and over-expression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. *Appl Microbiol Biotechnol* 49: 277-284.

[55] Halldorsdottir S, Thorolfsdottir ET, Spilliaert R, Johansson M, Thorbjarnardottir SH, Palsdottir A, Hreggvidsson GO, Kristjansson JK, Holst O & Eggertsson G (1998) Cloning, sequencing and overexpression of a Rhodothermus marinus gene encoding a thermostable cellulase of glycosyl hydrolase family 12. *Applied Microbiology and Biotechnology* 49: 277-284.

[56] Hannington M, Herzig P, Stoffers P, Scholten J, Botz R, Garbe-Schonberg D, Jonasson IR & Roest W (2001) First observations of high-temperature submarine hydrothermal vents and massive anhydrite deposits off the north coast of Iceland. *Mar Geol* 177: 199-220.

[57] Henrissat B CAZy - Carbohydrate Active enZYmes. URL http://afmb.cnrs-mrs.fr/CAZY/index.html

[58] Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280: 309-316.

[59] Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon J-P & Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA* 92: 7090-7094.

[60] Henrissat B & Romeu A (1995) Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochem J* 311: 350-351.

[61] Henrissat B & Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7: 637-644.

[62] Henrissat B (1999) Classification of chitinase modules. In *Chitin and chitinases*. Muzzarelli RAA (ed). Basel, Switzerland: Birkhäuser-Verlag pp. 137-159.

[63] Henrissat B & Davies G (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol* 124: 1515-1519.

[64] Henrissat B & Coutinho PM (2001) Classification of glycoside hydrolases and glycosyltransferases from hyperthermophiles. *Method Enzymol* 330: 183-201.

[65] Hillis DM & Doxon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66: 411-453.

[66] Hjorleifsdottir S, Kristjansson JK & Alfredsson GA (1989) Thermophilic organisms in submarine freshwater hot springs in Iceland. In *Microbiology of extreme environments and its potential for biotechnology*. da Costa MS, Duarte JC & Williams RAD (eds). London, U.K.: Elsevier Applied Sciences pp. 109-112.

[67] Hjorleifsdottir S, Skirnisdottir S, Hreggvidsson GO, Holst O & Kristjansson JK (2001) Species composition of cultivated and non-cultivated bacteria from short filaments in an Icelandic hot spring at 88 degrees C. *Microb Ecol* 42: 117-125.

[68] Hohn MJ, Hedlund BP & Huber H (2002) Detection of 16S rDNA sequences representing the novel phylum *Nanoarchaeota*: indication for a wide distribution in high temperature biotopes. *Syst Appl Microbiol* 25: 551-554.

[69] Honda Y, Kitoaka M & Hayashi K (2004) Kinetic evidence related to substrate-assisted catalysis of family 18 chitinases. *FEBS Lett* 567: 307-310.

[70] Hough DW & Danson MJ (1999) Extremozymes. *Curr Opin Chem Biol* 3: 39-46.

[71] Howard MB, Ekborg NA, Weiner RM & Hutcheson SW (2003) Detection and characterization of chitinases and other chitin-modifying enzymes. *J Ind Microbiol Biotechnol* 30: 627-35.

[72] Hreggvidsson GO & Kristjansson JK (2004) Thermophily. In *3.9 Extremophiles - Encyclopedia of life support systems*. Gerday C (ed). Oxford, UK: EOLSS Publishers.

[73] Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC & Stetter KO (2002) A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* 417: 63-67.

[74] Huber JA, Butterfield DA & Baross JA (2002) Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Appl Environ Microbiol* 68: 1585-1594.

[75] Huber R, Stohr J, Hohenhaus S, Rachel R, Burggraf S, Jannasch HW & Stetter KO (1995) *Thermococcus chitonophagus* sp. nov., a novel, chitin-degrading, hyperthermophilic archaeum from deep-sea hydrothermal vent environment. *Arch Microbiol* 164: 255-264.

[76] Huber R, Eder W, Heldwein S, Wanner G, Huber H, Rachel R & Stetter KO (1998) *Thermocrinis ruber*, gen. nov., sp. nov., a pink-filament-forming hyperthermophilic bacterium isolated from Yellowstone Park. *Appl Environ Microbiol* 64: 3576-3583.

[77] Huber R, Huber H & Stetter KO (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol Rev* 24: 615-623.

[78] Huber T, Faulkner G & Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*: bth226.

[79] Hugenholtz P, Pitulle C, Hershberger KL & Pace NJ (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* 180: 366-376.

[80] Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* 3: 1-8.

[81] Hugenholtz P & Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int J Syst Evol Microbiol* 53: 289-293.

[82] Hughes JB, Hellmann JJ, Ricketts TH & Bohannan BJM (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 67: 4399-4406.

[83] Jackson CR, Roden EE & Churchill PF (1998) Changes in bacterial species composition in enrichment cultures with various dilutions of inoculum as monitored by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 64: 5046-5048.

[84] Kafetzopoulos D, Martinou A & Bouriotis V (1993) Bioconversion of chitin to chitosan - Purification and characterization of chitin deacetylase from *Mucor rouxii*. *Proc Natl Acad Sci USA* 90: 2564-2568.

[85] Karner MB, DeLong EF & Karl DM (2002) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409: 507-510.

[86] Kashefi K & Lovley DR (2003) Extending the upper temperature limit for life. *Science* 301: 934.

[87] Keller M & Zengler K (2004) Tapping into microbial diversity. *Nat Rev Microbiol* 2: 141-150.

[88] Keyhani NO & Roseman S (1999) Physiological aspects of chitin catabolism in marine bacteria. *Biochim Biophys Acta* 1473: 108-122.

[89] Kirk O, Vedel Borchert R & Crone Fuglsang C (2002) Industrial enzyme applications. *Curr Opin Biotechnol* 13: 345-351.

[90] Koga D, Mitsutomi M, Kono M & Matsumiya M (1999) Biochemistry of chitinases. In *Chitin and chitinases*. Jolles P & Muzzarelli RAA (eds). Basel, Switzerland: Birkhäuser-Verlag pp. 111-123.

[91] Koshland DE (1953) Stereochemistry and mechanisms of enzymatic reactions. *Biol Rev Camb Philos Soc* 28: 416-436.

[92] Kristjansson JK, Hreggvidsson GO & Alfredsson G (1986) Isolation of halotolerant *Thermus* spp. from submarine hot springs in Iceland. *Appl Environ Microbiol* 52: 1313-1316.

[93] Kristjansson JK & Stetter K (1992) Thermophilic bacteria. In *Thermophilic bacteria*. Kristjansson JK (ed). Boca Raton, Fl., USA: CRC Press, Inc. pp. 1-18.

[94] Kristjansson JK, Hjorleifsdottir S, Marteinsson VT & Alfredsson GA (1994) *Thermus scotoductus*, sp. nov., a pigment-producing thermophilic bacterium from hot tap water in Iceland and including *Thermus* sp X-1. *Syst Appl Microbiol* 17: 44-50.

[95] Kristjansson JK & Hreggvidsson GO (1995) Ecology and habitats of extremophiles. *World J Microbiol Biotechnol* 11: 17-25.

[96] Kristjansson JK, Hreggvidsson GO & Grant WD (2000) Taxonomy of extremophiles. In *Applied microbial systematics*. Priest FG & Goodfellow M (eds). Dordrecht, The Netherlands: Kluwer Academic Press pp. 231-291.

[97] Kumar S & Nussinov R (2001) How do thermophilic proteins deal with heat? *Cell Mol Life Sci* 58: 1216-1233.

[98] Kuriki T & Imanaka T (1999) The concept of the α-amylase family: Structural similarity and common catalytic mechanism. *J Biosci Bioeng* 87: 557-565.

[99] Leadbetter JR (2003) Cultivation of recalcitrant microbes: cells are alive, well and revealing their secrets in the 21st century laboratory. *Curr Opin Microbiol* 6: 274-281.

[100] Leemhuis H (2003) PhD Thesis. *What makes cyclodextrin glycosyltransferase a transglycosylase?* Groningen, The Netherlands: University of Groningen.

[101] MacGregor EA, Janecek S & Svenson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim Biophys Acta* 1546: 1-20.

[102] Madigan MT, Martinko JM & Parker J (1997) In *Brock Biology of Microorganisms*. Madigan MT, Martinko JM & Parker J (eds). Upper Saddle River, NJ, USA: Prentice Hall International Editions.

[103] Maidak B, Cole J, Parker Jr. C, Garrity G, Larsen N, Li B, Lilburn T, McCaughey M, Olsen G, Overbeek R, Pramanik S, Schmidt T, Tiedje J & Woese C (1999) A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res* 27: 171-173.

[104] Marrs B, Delagrave S & Murphy D (1999) Novel approaches for discovering industrial enzymes. *Curr Opin Microbiol* 2: 241-245.

[105] Marteinsson VT (1997) PhD Thesis. *Isolement et caracterisation de microorganismes extremophiles originaire de site hydrothermaux oceaniques. Influence du couple temperature-pression sur la physiologie*. Roscoff, France: Universite de Bretagne Occidentale.

[106] Marteinsson VT, Birrien JL & Prieur D (1997) *In situ* enrichment and isolation of thermophilic microorganisms from deep-sea vent environments. *Can J Microbiol* 43: 694-697.

[107] Marteinsson VT, Birrien JL, Raguenes G, da Costa MS & Prieur D (1999) Isolation and characterization of *Thermus thermophilus* Gy1211 from a deep-sea hydrothermal vent. *Extremophiles* 3: 247-51.

[108] Marteinsson VT, Birrien JL, Reysenbach AL, Vernet M, Marie D, Gambacorta A, Messner P, Sleytr UB & Prieur D (1999) *Thermococcus barophilus* sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* 49: 351-359.

[109] Marteinsson VT, Kristjansson JK, Kristmannsdottir H, Dahlkvist M, Saemundsson K, Hannington M, Petursdottir SK, Geptner A & Stoffers P (2001) Discovery and description of giant submarine smectite cones on the seafloor in Eyjafjordur, northern Iceland, and a novel thermal microbial habitat. *Appl Environ Microbiol* 67: 827-833.

[110] Martin C & Smith AM (1995) Starch Biosynthesis. *Plant Cell* 7: 971-985.

[111] McDougald D, Rice SA, Weichart D & Kjelleberg S (1998) Non-culturability: adaptation or debilitation. *FEMS Microbiol Ecol* 25: 1-9.

[112] Merzendorfer H & Zimoch L (2003) Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J Exp Biol* 206: 4393-4412.

[113] Metcalfe AC, Krsek M, Gooday GW, Prosser JI & Wellington EMH (2002) Molecular analysis of a bacterial chitinolytic community in an upland pasture. *Appl Environ Microbiol* 68: 5042-5050.

[114] Miroshnichenko ML, Kostrikina NA, L'Haridon S, Jeanthon C, Hippe H, Stackebrandt E & Bonch-Osmolovskaya EA (2002) *Nautilia lithotrophica* gen. nov., sp. nov., a

thermophilic sulfur-reducing epsilon-*Proteobacterium* isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 52: 1299-1304.

[115] Miroshnichenko ML, L'Haridon S, Jeanthon C, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E & Bonch-Osmolovskaya EA (2003) *Oceanithermus profundus* gen. nov., sp. nov., a thermophilic, microaerophilic, facultatively chemolithoheterotrophic bacterium from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 53: 747-752.

[116] Miroshnichenko ML, L'Haridon S, Nercessian O, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA & Jeanthon C (2003) *Vulcanithermus mediatlanticus* gen. nov., sp. nov., a novel member of the family *Thermaceae* from a deep-sea hot vent. *Int J Syst Evol Microbiol* 53: 1143-1148.

[117] Miroshnichenko ML, L'Haridon S, Schumann P, Spring S, Bonch-Osmolovskaya EA, Jeanthon C & Stackebrandt E (2004) *Caminibacter profundus* sp. nov., a novel thermophile of *Nautiliales* ord. nov. within the class '*Epsilonproteobacteria*', isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 54: 41-45.

[118] Muyzer G, de Waal E & Uitterlinden A (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59: 695-700.

[119] Muyzer G (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Curr Opin Microbiol* 2: 317-322.

[120] Muzzarelli RAA (1999) Native, industrial, and fossil chitins. In *Chitin and chitinases*. Muzzarelli RAA & Jolles P (eds). Basel, Switzerland: Birhäuser-Verlag pp. 1-5.

[121] Nakayama A, Yamamoto K & Tabata S (2001) Identification of the catalytic residues of bifunctional glycogen debranching enzyme. *J Biol Chem* 276: 28824-28828.

[122] Nawani NN, Kapadnis BP, Das AD, Rao AS & Mahajan SK (2002) Purification and characterization of a thermophilic and acidophilic chitinase from *Microbispora* sp. V2. *J Appl Microbiol* 93: 965-975.

[123] Niehaus F, Bertoldo C, Kähler M & Antranikian G (1999) Extremophiles as a source of novel enzymes for industrial applications. *Appl Microbiol Biotechnol* 51: 711-729.

[124] Nogi Y & Kato C (1999) Taxonomic studies of extremely barophilic bacteria isolated from the Mariana Trench and description of *Moritella yayanosii* sp. nov., a new barophilic bacterial isolate. *Extremophiles* 3: 71-77.

[125] Nordberg-Karlsson E, Bartonek-Roxa E & Holst O (1997) Cloning and sequence of a thermostable multidomain xylanase from the bacterium *Rhodothermus marinus*. *BBA - Gene Structure and Expression* 1353: 118-124.

[126] Nunes OC, Donato MM & da Costa MS (1992) Isolation and characterization of *Rhodothermus* strains from S. Miguel, Azores. *Syst Appl Microbiol* 15: 92-97.

[127] Olszewski TD (2004) A unified mathematical framework for the measurement of richness and evenness within and among multiple communities. *Oikos* 104: 377-387.

[128] Ovreas L (2000) Population and community level approaches for analyzing microbial diversity in natural environments. *Ecol Lett* 3: 236-251.

[129] Pace NJ (1997) A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.

[130] Patil RS, Ghormade V & Deshpande MV (2000) Chitinolytic enzymes: an exploration. *Enzyme Microb Technol* 26: 473-483.

[131] Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS & Vorgias CE (1994) Crystal structure of a bacterial chitinase at 2.3Å resolution. *Structure* 2: 1169-1189.

[132] Politz O, Krah M, Thomsen KK & Borriss R (2000) A highly thermostable endo-(1,4)-beta-mannanase from the marine bacterium *Rhodothermus marinus*. *Appl Microbiol Biotechnol* 53: 715-721.

[133] Poulicek M & Jeuniaux C (1988) Chitin biomass in marine sediments. In *Chitin and chitosan*. Skjak-Braek G, Anthonsen T & Sandford P (eds). London, U.K.: Elsevier Applied Sciences pp. 151-155.

[134] Prescott LM, Parley JP & Klein DA (1999) In *Microbiology*. Prescott LM, Parley JP & Klein DA (eds). Boston, Ma, USA: WCB McGraw-Hill.

[135] Prieur D, Erauso G & Jeanthon C (1995) Hyperthermophilic life at deep-sea hydrothermal vents. *Planet Sci* 43: 115-122.

[136] Prieur D (1997) Microbiology of deep-sea hydrothermal vents. *Trends Biotechnol* 15: 242-244.

[137] Prieur D, Voytek M, Jeanthon C & Reysenbach AL (2001) Deep-sea thermophilic prokaryotes. In *Thermophiles - Biodiversity, ecology, and evolution*. Reysenbach AL, Voytek M & Mancinelli R (eds). New-York, USA: Kluwer Academic/Plenum Publishers pp. 11-22.

[138] Rainey FA, Nobre MF, Schumann P, Stackebrandt E & da Costa MS (1997) Phylogenetic diversity of the *Deinococci* as determined by 16S ribosomal DNA sequence comparison. *Int J Syst Bacteriol* 47: 510-514.

[139] Rappé M, Connon SA, Vergin KL & Giovanonni SJ (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418: 806-810.

[140] Ravenschlag K, Sahm K & Amann R (2001) Quantitative molecular analysis of the microbial community in marine Arctic sediments (Svalbard). *Appl Environ Microbiol* 67: 387-395.

[141] Ravi Kumar MNV (2000) A review of chitin and chitosan applications. *Reactive Functional Polymers* 46: 1-27.

[142] Rees HC, Grant WD, Jones BE & Heaphy S (2004) Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles* 8: 63-71.

[143] Reguera G & Leschine SB (2001) Chitin degradation by cellulolytic anaerobes and facultative aerobes from soils and sediments. *FEMS Microbiol Lett* 204: 367-374.

[144] Rodriguez-Valera F (2004) Environmental genomics, the big picture? *FEMS Microbiol Lett* 231: 153-158.

[145] Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J & Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66: 2541-2547.

[146] Rossello-Mora R & Amann R (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* 25: 39-67.

[147] Rye CS & Withers SG (2000) Glycosidase mechanisms. *Curr Opin Chem Biol* 4: 573-580.

[148] Saano A, Tas E, Pippola S, Lindstöm K & Van Elsas JD (1995) Extraction and analysis of microbial DNA from soil - Methods and applications. In *Nucleic acids in the environment*. Trevors JT & Van Elsas JD (eds). Berlin, Germany: Springer-Verlag pp. 49-67.

[149] Saito Y, Okano T, Gaill F, Chanzy H & Putaux J-L (2000) Structural data on the intra-crystalline swelling of β-chitin. *Int J Biol Macromol* 28: 81-88.

[150] Sakai K, Yokota A, Kurokawa H, Wakayama M & Moriguchi M (1998) Purification and characterization of three thermostable endochitinases of a noble *Bacillus* strain, MH-1, isolated from chitin-containing compost. *Appl Environ Microbiol* 64: 3397-3402.

[151] Sako Y, Nakagawa S, Takai K & Horikoshi K (2003) *Marinithermus hydrothermalis* gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* 53: 59-65.

[152] Schleper C, Puhler G, Klenk H-P & Zillig W (1996) *Picrophilus oshimae* and *Picrophilus torridus* fam. nov., gen. nov, sp. nov., two species of hyperacidophilic, thermophilic, heterotrophic, aerobic *Archaea*. *Int J Syst Bacteriol* 46: 814-816.

[153] Schloss PD & Handelsman J (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14: 303-310.

[154] Shi T, Reeves RH, Gilichinsky DA & Friedmann EI (1997) Characterization of viable bacteria from Siberian permafrost by 16S rDNA sequencing. *Microb Ecol* 33: 169-179.

[155] Sinnott ML (1990) Catalytic mechanisms of enzymic glycosyl transfer. *Chem Rev* 90: 1171-1202.

[156] Skirnisdottir S, Hreggvidsson GO, Hjörleifsdottir S, Marteinsson VT, Petursdottir SK, Holst O & Kristjansson JK (2000) Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Appl Environ Microbiol* 66: 2835-2841.

[157] Skirnisdottir S, Hreggvidsson GO, Holst O & Kristjansson JK (2001) Isolation and characterization of a mixotrophic sulfur-oxidizing *Thermus scotoductus*. *Extremophiles* 5: 45-51.

[158] Sleator RD & Hill C (2002) Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* 26: 49-71.

[159] Souza RF, Gomes RC, Coelho RRR, Alviano CS & Soares RMA (2003) Purification and characterization of an endochitinase produced by *Colletotrichum gloeosporioides*. *FEMS Microbiol Lett* 222: 45-50.

[160] Spilliaert R, Hreggvidsson GO, Kristjansson JK, Eggertsson G & Palsdottir A (1994) Cloning and sequencing of a *Rhodothermus marinus* gene, *bglA*, coding for a thermostable beta-glucanase and its expression in *Escherichia coli*. *Eur J Biochem* 224: 923-930.

[161] Stackebrandt E & Goebel BM (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44: 846-849.

[162] Sterner R & Liebl W (2001) Thermophilic adaptation of proteins. *Crit Rev Biochem Mol Biol* 36: 39-106.

[163] Stetter KO, König H & Stackebrandt E (1983) *Pyrodictium* gen. nov., and new genus of submarine disc-shaped sulfur-reducing archaebacteria growing optimally at 105°C. *Syst Appl Microbiol* 3: 535-551.

[164] Sutrisno A, Ueda M, Abe Y, Nakazawa M & Miyatake K (2004) A chitinase with high activity toward partially *N*-acetylated chitosan from a new, moderately thermophilic, chitin-degrading bacterium *Ralstonia* sp. A-471. *Appl Microbiol Biotechnol* 63: 398-406.

[165] Suzuki K, Taiyoji M, Sugawara N, Nikaidou N, Henrissat B & Watanabe T (1999) The third chitinase gene (*chiC*) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases. *Biochem J* 343: 587-596.

[166] Synowiecki J & Al-Khateeb NA (2003) Production, properties, and some new applications of chitin and its derivatives. *Crit Rev Food Sci Nutr* 43: 145-171.

[167] Synstad B, Gaseidnes S, van Aalten DMF, Vriend G, Nielsen JE & Eijsink VGH (2004) Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. *Eur J Biochem* 271: 253-262.

[168] Takai K & Horikoshi K (1999) Molecular phylogenetic analysis of archaeal intron-containing genes coding for rRNA obtained from a deep subsurface geothermal water pool. *Appl Environ Microbiol* 65: 5586-5589.

[169] Takayanagi T, Ajisaka K, Takiguchi Y & Shimahara K (1991) Isolation and characterization of thermostable chitinases from *Bacillus licheniformis* X-7u. *Biochim Biophys Acta* 1078: 404-410.

[170] Tanaka T, Fujiwara S, Nishikori S, Fukui T, Takagi M & Imanaka T (1999) A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *Appl Environ Microbiol* 65: 5338-5344.

[171] Tanaka T, Fukui T & Imanaka T (2001) Different cleavage specificities of the dual catalytic domains in chitinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Biol Chem* 276: 35629-35635.

[172] Tanaka T, Fukui T, Atomi H & Imanaka T (2003) Characterization of an exo-β-D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* 185: 5175-5181.

[173] Tanner MA, Goebel BM, Dojka MA & Pace NJ (1998) Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol* 64: 3110-3113.

[174] Tenreiro S, Nobre MF, Rainey FA, Miguel C & da Costa MS (1997) *Thermonema rossianum* sp. nov., a new thermophilic and slightly halophilic species from saline hot springs in Naples, Italy. *Int J Syst Bacteriol* 47: 122-126.

[175] Teske A, Hinrichs KU, Edgcomb V, de Vera Gomez A, Kysela D, Sylva SP, Sogin ML & Jannasch HW (2002) Microbial diversity of hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl Environ Microbiol* 68: 1994-2007.

[176] Tharanathan RN & Kittur FS (2003) Chitin - the undisputed biomolecule of great potential. *Crit Rev Food Sci Nutr* 43: 61-87.

[177] Thorbjarnardottir SH, Jonsson ZO, Andresson OS, Kristjansson JK, Eggertsson G & Palsdottir A (1995) Cloning and sequence analysis of the DNA ligase-encoding gene of *Rhodothermus marinus*, and overproduction, purification and characterization of 2 thermophilic DNA ligases. *Gene* 161: 1-6.

[178] Torsvik V, Goksoyr J & Daae FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56: 782-787.

[179] Torsvik V & Ovreas L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* 5: 240-245.

[180] Tsigos I, Martinou A, Kafetzopoulos D & Bouriotis V (2000) Chitin deacetylases: new, versatile tools in biotechnology. *Trends Biotechnol* 18: 305-312.

[181] Tsujibo H, Minoura K, Miyamoto K, Endo H, Moriwaki M & Inamori Y (1993) Purification and properties of a thermostable chitinase from *Streptomyces thermoviolaceus* OPC-520. *Appl Environ Microbiol* 59: 620-622.

[182] Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhshar DS & Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37-43.

[183] Vaaje-Kolstad G, Houston DR, Rao FV, Peter MG, Synstad B, van Aalten DMF & Eijsink VGH (2004) Structure of the D142N mutant of the family 18 chitinase ChiB from *Serratia marcescens* and its complex with allosamidin. *BBA - Protein and Proteomics* 1696: 103-111.

[184] van Aalten DMF Chitinases - reaction mechanism and structure-based drug design (ICS2002 7/2002). URL http://davapc1.bioch.dundee.ac.uk/talks/index.html

[185] van Aalten DMF, Komander D, Synstad B, Gaseidnes S, Peter MG & Eijsink VGH (2001) Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc Natl Acad Sci USA* 98: 8979-8984.

[186] van der Maarel MJEC, van der Veen B, Uitdehaag JCM, Leemhuis H & Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the α-amylase family. *J Biotechnol* 94: 137-155.

[187] Vandamme P, Pot B, Gillis M, de Vos P, Kersters K & Swings J (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60: 407-438.

[188] Venter CJ, Remington K, Heidelberg J, Halpern A, Rusch D, Eisen J, Wu D, Paulsen I, Nelson K, Nelson W, Fouts D, Levy S, Knap A, Lomas M, Nealson K, White O, Peterson J, Hoffman J, Parson R, Baden-Tillson H, Pfannkoch C, Rogers Y-H & Smith H (2004) Environmental genome shotgun sequencing of the Sargasso sea. *Science* 304: 66-74.

[189] Vetriani C, Maeder DL, Tolliday N, Yip KS-P, Stillman TJ, Britton KL, Rice DW, Klump HH & Robb FT (1998) Protein thermostability above 100°C: A key role for ionic interactions. *PNAS* 95: 12300-12305.

[190] Vieille C & Zeikus GJ (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65: 1-43.

[191] von Wintzingerode F, Gobel UB & Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* 21: 213-229.

[192] Warren RAJ (1996) Microbial hydrolysis of polysaccharides. *Annu Rev Microbiol* 50: 183-212.

[193] Watanabe T, Kobori K, Miyashita K, Fujii T, Sakai H, Uchida M & Tanaka H (1993) Identification of Glu^{204} and Asp^{200} in chitinase A1 of *Bacillus circulans* WL12 as essential residues for chitinase activity. *J Biol Chem* 268: 18567-18572.

[194] Watanabe T, Ishibashi A, Ariga Y, Hashimoto M, Nikaidou N, Sugiyama J, Matsumoto T & Nonaka T (2001) Trp^{122} and Trp^{134} on the surface of the catalytic domain are essential for crystalline chitin hydrolysis by *Bacillus circulans* chitinase A1. *FEBS Lett* 494: 74-78.

[195] Watanabe T, Ariga Y, Sato U, Toratani T, Hashimoto M, Nikaidou N, Kezuka Y, Nonaka T & Sugiyama J (2003) Aromatic residues within the substrate-binding cleft of *Bacillus circulans* chitinase A1 are essential for hydrolysis of crystalline chitin. *Biochem J* 379: 237-244.

[196] Williams RAD (1992) The genus *Thermus*. In *Thermophilic bacteria*. Kristjansson JK (ed). Boca Raton, Fl, USA: CRC Press, Inc. pp. 51-62.

[197] Williamson N, Brian P & Wellington EMH (2000) Molecular detection of bacterial and streptomycete chitinases in the environment. *Antonie Leeuwenhoek* 78: 315-321.

[198] Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74: 5088-5090.

[199] Woese CR, Kandler O & Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria* and *Eukarya*. *Proc Natl Acad Sci USA* 87: 4576-4579.

[200] Zeikus GJ, Vieille C & Savchenko A (1998) Thermozymes: biotechnology and structure-function relationships. *Extremophiles* 2: 179-183.

[201] Zengler K, Toledo G, Rappé M, Elkins J, Mathur EJ, Short JM & Keller M (2002) Cultivating the uncultured. *Proc Natl Acad Sci USA* 99: 15681-15686.

[202] Zhao H, Chockalingam K & Chen Z (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr Opin Biotechnol* 13: 104-110.

[203] Zhou J (2003) Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* 6: 288-294.