Molecular Characterization of Zinc- and Iron- Containing Alcohol Dehydrogenases from Anaerobic Hyperthermophiles

by

Liangliang Hao

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2009

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Hyperthermophiles grow optimally at 80 ^{0}C and above, and many of them have the ability to utilize various carbohydrates as carbon source and produce ethanol as an end product. Alcohol dehydrogenase (ADH) is a key enzyme responsible for alcohol production, catalyzing interconversions between alcohols and corresponding ketones or aldehydes. ADHs from hyperthermophiles are of great interests due to their thermostability, high activity and enantioselectivity. The gene encoding ADH from hyperthermophilic archaeon *Thermococcus guaymasensi*s was cloned, sequenced and over-expressed. DNA fragments of the genes encoding the ADHs were amplified directly from the corresponding genomic DNA by combining the use of conventional and inverse PCRs. The entire gene was detected to be 1092 bp and the deduced amino acid sequence had a total of 364 amino acids with a calculated molecular mass of 39463 Dalton. The enzyme belonged to the family of zinc-containing ADHs with catalytic zinc only. It was verified that the enzyme had binding motifs of catalytic zinc only ($GHEX_2GX_5GX_2V$, residues 62-76) and coenzyme NADP ($GXGX_2G$, residues 183-188). The tertiary structural modeling showed two typical domains, one catalytic domain close to amino-terminal (N-terminal) end and one coenzyme-binding domain close to carboxy-terminal (C-terminal) end. Since its codon usage pattern seemed to be different from that of *Escherichia coli*, the enzyme was over-expressed in the *E. coli* codon plus strain using pET-30a vector. The recombinant enzyme was detected to be soluble and active (1073 U/mg), which was virtually the same to the native enzyme (1049 U/mg). The recombinant ADH possessed almost identical properties with the native enzyme. The optimal pHs for ethanol oxidation and acetaldehyde reduction were 10.5 and 7.5 respectively, while the activity for alcohol oxidation was much higher than that of aldehyde reduction. The enzyme activity was inhibited in the presence of 100 μM Zn^{2+} in the assay mixture and it has a half-life of 6 hours after exposure to air.

Thermotoga hypogea is an extremely thermophilic anaerobic bacterium capable of growing at 90 ^{0}C. The gene encoding an alcohol dehydrogenase from *T. hypogea* was cloned, sequenced and over-expressed. The gene sequence (1164 bp) was obtained

successfully by sequencing all the DNA fragments amplified from PCR. The deduced amino acid sequence was found to have high degrees of identity (~72%) to iron-containing ADHs from *Thermotoga* species and harbored typical iron and NADP-binding motifs, $Asp_{195}His_{199}His_{268}His_{282}$ and $Gly_{39}Gly_{40}Gly_{41}Ser_{42}$, respectively. The structural modeling showed that N-terminal domain of *Th*ADH contained α/β-dinucleotide-binding motif and its C-terminal domain was α-helix-rich region including iron-binding motif. The gene encoding *T. hypogea* ADH was functionally expressed in *E. coli* using the vector pET-30a. The recombinant protein was expressed optimally in *E. coli* grown in the presence of 1 mM ferrous and induced by 0.4-0.6 mM IPTG. The recombinant enzyme was found to be soluble, active and thermostable, and had a subunit size of 43 kDa revealed by SDS-PAGE analyses. The native ADH from *T. hypogea* was purified to homogeneity for comparative analysis using a three-step liquid chromatography while the recombinant ADH over-expressed in *E. coli* was isolated by a simpler procedure including one-hour heat treatment. The activity of the purified recombinant enzyme was 69 U/mg and presented almost identical properties with the native enzyme. The optimal pHs for ethanol oxidation and acetaldehyde reduction were 11.0 and 8.0 respectively, while activity for alcohol oxidation were higher than that of aldehyde reduction. The enzyme was oxygen sensitive and it had a half-life ($t_{1/2}$) of 20 minutes after exposed to air. The enzyme remained 50% activity after incubation at 70 ^{0}C for 2 hours. Successful high-level expression of *T. hypogea* ADH in *E. coli* will significantly facilitate further study on the catalytic mechanism of iron-containing ADHs.

In summary, both zinc- and iron-containing ADHs from two hyperthermophiles were successfully cloned, sequenced and overexpressed in mesophilic host *E. coli*, and such a high-level expression of ADH genes provides possibilities for three dimensional structural analysis by X-ray crystallography and enzyme modification by mutagenesis, which will help further explore mechanisms of catalysis and protein thermostability of iron and zinc-containing ADHs and their potential applications in biotechnology.

Acknowledgements

I would like to thank my wonderful supervisor, Dr. Kesen Ma, for his enthusiasm and guidance throughout my degree. I would never forget the efforts you offered that cultivated me from a young girl to a thoughtful graduate student. I would like to thank my committee members, Dr. Trevor Charles and Dr. Owen Ward, for their invaluable discussions regarding my research and all the help for my future career.

I would like to thank all of my former and present labmates and colleagues that I have worked with during the past two years: Dr. Xianqin Yang, Mohammad Eram, Hongbin Zhu, thank you for your friendship and willingness to share your knowledge with me. Particularly, I would like to thank Dr. Xiangxian Ying who did excellent work in alcohol dehydrogenase research for sharing his scientific ideas and helping me through the hard time during my study. I would thoroughly like to thank Xuan Huang and Yili Sun for all the joys and sadness we shared, and the invaluable friendship that is so rare to meet nowadays. Friends in the Department of Biology, Jin Duan, Youai Hao, Zhenyu Cheng, offered me a lot of help and cheerful talks. Thank you for your kindness and care.

I would especially like to thank my fiance, Houqing Yu, for his love, understanding and supporting in academic exploration. Lastly, I would like to thank my dearest dad and mom for their endless love and supporting during my lifetime. I thank them for being so proud of all my achievements.

Dedication

This thesis is dedicated to who guided, supported and helped me during my lifetime.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADH	Alcohol dehydrogenase
***Ao*ADH**	*Alkaliphilus oremlandii* alcohol dehydrogenase
CAPS	3-(Cyclohexylamino)-1-propanesulfonic acid
***Cb*ADH**	*Clostridium beijerinckii* alcohol dehydrogenase
***Cp*ADH**	*Carboxydibrachium pacificum* alcohol dehydrogenase
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EPPS	N-(2-hydroxyethyl)-piperazine-N'-(3-propanedulfonic acid)
Fe	Iron
***Fn*ADH**	*Fervidobacterium nodosum* alcohol dehydrogenase
FPLC	Fast Protein Liquid Chromatography
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
MS	Mass spectrometry
NAD(H)	Nicotinamide adenine dinucleotide (reduced)
NADB_Rossman	Rossmann fold NAD(P)-binding proteins
NADP(H)	Nicotinamide adenine dinucleotide phosphate (reduced)
PCR	Polymerase chain reaction
PIPES	1, 4-Piperazine-bis-(ethanesulfonic acid)
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SDT	Sodium dithionite
***Tb*ADH**	*Thermoanaerobacter brockii* alcohol dehydrogenase
***Te*ADH**	*Thermoanaerobacter ethanolicus* alcohol dehydrogenase
***Tg*ADH**	*Thermococcus guaymasensis* alcohol dehydrogenase
***Th*ADH**	*Thermotoga hypogea* alcohol dehydrogenase
Tris	2-Amino-2-hydroxymethyl-1,3-propanediol
Zn	Zinc

Chapter 1 General Introduction

1.1 HYPERTHERMOPHILES

1.1.1 Thermophiles and hyperthermophiles

Thermophilic and hyperthermophilic organisms are isolated from geothermal and hydrothermal environments, which have the ability to thrive at high temperatures. Thermophiles are defined as organisms with optimal growth temperatures between 45 and 80 ^{0}C, which have been isolated from various environments including geothermal springs, sunlight-heated soils etc. Most thermophiles known are moderate, and show an upper temperature range of growth between 50 and 70 °C, although they are still able to grow slowly at 25-40 °C (Stetter, 1996).

Hyperthermophiles are considered to be organisms with optimal growth temperatures at 80 ^{0}C and above (Stetter, 2006). They have been found in a number of hydrothermal environments including both hydrothermal and geothermal areas, i.e. deep sea and shallow thermal sediments, hydrothermal vent systems, geothermal springs and solfataras. To date, more than 100 species of hyperthermophilic bacteria and archaea are known (Fukami-Kobayashi et al., 2007), which had been isolated from different terrestrial and marine thermal areas in the world. Hyperthermophiles grow fastest from 80 °C to 110 °C (Stetter, 1996 and 1998); however, they are unable to grow below 60 °C. These organisms are able to grow not only at high temperatures, but also extremes of pH, redox potential, pressure and salinity. Hyperthermophiles belong to phylogenetically distant groups and can be divided into two groups: *bacteria* and *archaea* (Stetter, 1996). Hyperthermophilic *bacteria* and *archaea* represent the organisms at the upper temperature border of life (**Fig. 1-1**). Their outstanding heat resistance makes them

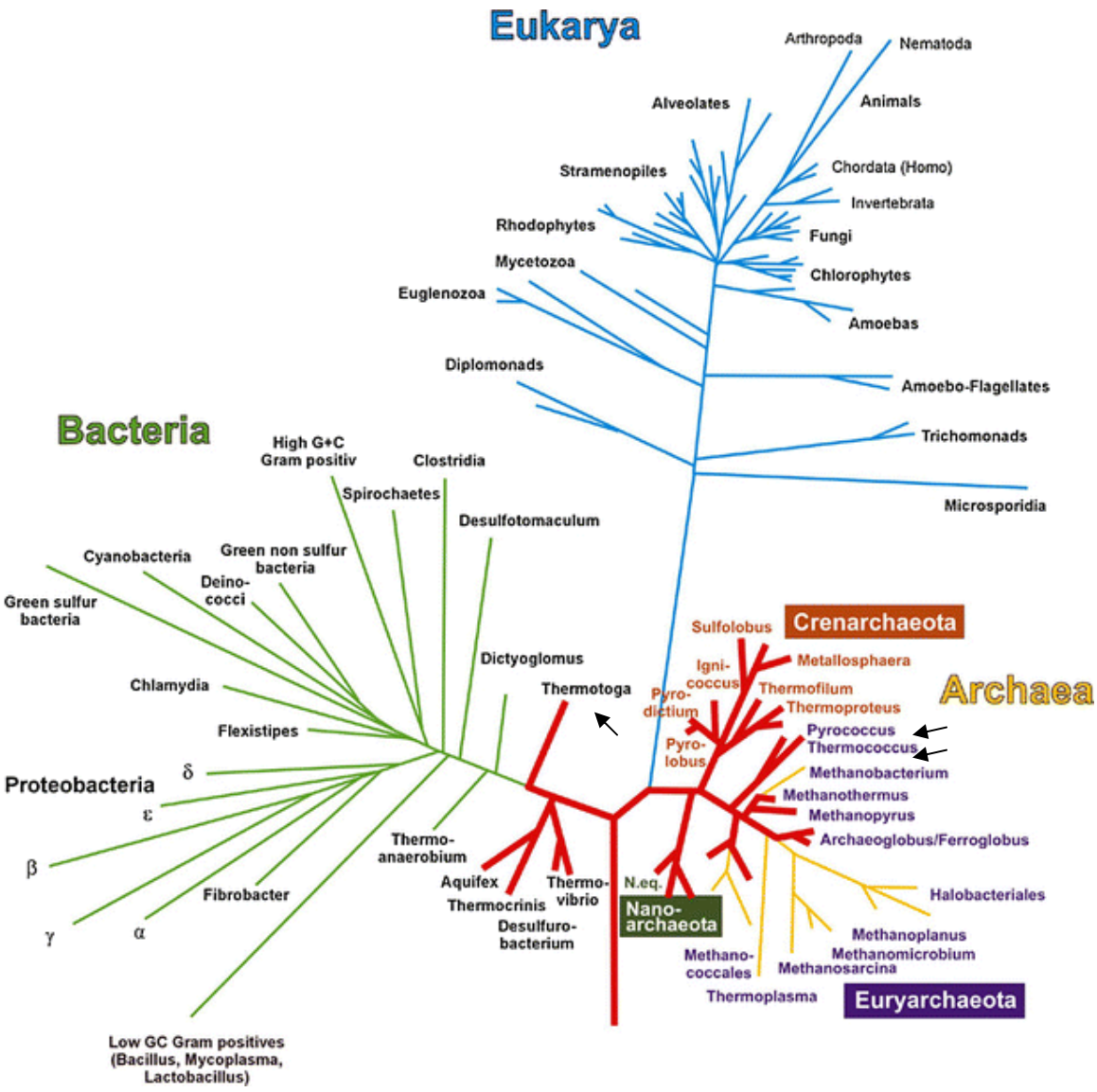


Figure 1-1 Small subunit 16s rRNA based phylogenetic tree

(Adapted and modified from Stetter 2006). The thick lineages represent hyperthermophiles, and the arrows indicted the well-studied species of *Pyrococcus*, *Thermococcous* and *Thermotoga*.

interesting objects for basic research as for biotechnology. So far, in addition to the traditional taxonomic characteristics such as GC-contents of DNA, DNA-DNA homology, morphology, and physiological features, 16S rRNA sequence-based classification of prokaryotes appears to be imperative for the recognition and characterization of novel taxonomic groups (Radianingtyas et al., 2003).

1.1.2 The genus *Thermococcus*

Thermococcus is one of the three genera (*Thermococcus, Pyrococcus, Paleococcus*) from the euryarchaeal order *Thermococcales* (Huber and Stetter 2001; Itoh 2003, Takai et al., 2000). The genera *Pyrococcus* and *Thermococcus* have been well studied, especially for the species *Pyrococcus horikoshii* (Kawarabayasi et al., 1998), *Pyrococcus furiosus* (Robb et al., 2001), *Pyrococcus abyssi* (Cohen et al., 2003) and *Thermococcus kodakaraensis* (Fukui et al., 2005) whose complete genomes have been sequenced and published. The genus *Thermococcus* contains the biggest number of characterized isolates among archaea. Deposited in German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ), this group contains 27 members. Except *Thermococcus sibiricus* (Miroshnichenko et al., 2001), *Thermococcus celer* (Zillig et al., 1983) and *Thermococcus stetteri* (Miroshnichenko et al. 1989), most of *Thermococcus* species are hyperthermophiles and are usually spherical and obligately anaerobic. *Thermococcus guaymasensis* is a member of obligately heterotrophic and strictly anaerobic archaea. It grows on organic substrates, primarily in the presence of elemental sulfur (S°) at temperatures from 56 ^{0}C up to 90 ^{0}C, with no growth occurring at 95 ^{0}C. The optimum temperature for growth is 88 ^{0}C. The

pH range for growth is 5.6-8.1, with the optimum at 7.2 (Canganella et al., 1998). Acetate, CO_2 and H_2S are formed as the main end products of metabolism, and ethanol can also be produced (Ying et al., unpublished results). The G-C content of its genome is found to be 46.0 mol% while no genomic sequence is available presently.

1.1.3 The genus *Thermotoga*

Besides the obligately organotrophic, strictly anaerobic features, members of *Thermotoga* genus have characteristic rodshaped cells surrounded by a sheath known as "*Toga*". The genus *Thermotoga* consists of nine species: *Thermotoga maritima* (Huber et al., 1986), *Thermotoga neapolitana* (Jannasch et al., 1988), *Thermotoga thermarum* (Windberger et al., 1989), *Thermotoga subterranean* (Jeanthon et al., 1995), *Thermotoga elfii* (Ravot et al., 1995), *Thermotoga hypogea* (Fardeau et al., 1997), *Thermotoga petrophila* and *Thermotoga naphthophila* (Takahata et al., 2001) as well as *Thermotoga lettingae* (Balk et al., 2002). Among them, *T. maritima* is the first species whose genome has been sequenced. Up to date, completed genome information of *T. petrophila* and *T. lettingae* is also available in GenBank (http://www.ncbi.nlm.nih.gov/Genbank/).

T. hypogea is an anaerobic, extremely thermophilic bacterium. This strain is rod-shaped and has a characteristic outer sheathlike structure. The optimum temperature for *T. hypogea* growth is 70 ^{0}C at pH 7.0; the growth temperature range is from 56 to 90 ^{0}C. The optimum pH is 7.3 to 7.4 at 70 ^{0}C; growth occurs between pH 6.1 and pH 9.1. It can utilize carbohydrates including xylan as carbon and energy sources and producing acetate, CO_2 and hydrogen as the major end products. Furthermore, ethanol is also produced as an end product of glucose/xylose fermentation (Fardeau et al., 1997).

Therefore, *T. hypogea* may have potential application in biomass conversion to ethanol and hydrogen, which are alternative sources of fuel.

1.2 THERMOSTABLE ENZYMES

1.2.1 Thermostable enzymes

Since hyperthermophiles grow optimally at high temperatures, enzymes from these organisms (or named hyperthermophilic enzymes) have unique structure-function properties with high thermostability and optimal activity at temperatures above 80 ^{0}C, and some of these enzymes are active at temperatures as high as 110 ^{0}C and even above (Vieille et al., 1996). Enzymes from thermophilic organisms (thermophilic enzymes) are usually optimally active between 60 and 80 ^{0}C and show thermostability properties that fall between those of hyperthermophilic and mesophilic enzymes. Active at high temperatures, both thermophilic and hyperthermophilic enzymes typically do not function well below 40 ^{0}C (Vieille et al., 2001).

Since some enzymes remain active for extended periods of times at temperatures even higher than 100 ^{0}C (Kelly et al., 1994), thermostable enzymes have great potentials in applications as biocatalysts, and the research on the thermostable proteins has been a quite dynamic area at present. The discovery of novel enzymes from extremophilic microorganisms greatly promotes industrial application of enzymes that can withstand harsh conditions; in particular, enzymes from hyperthermophilic organisms have the most practical commercial use to date because of their overall inherent stability. Additionally, it was proposed that hyperthermophiles were the first life forms to have arisen on Earth

(Stetter, 1996). Hyperthermophilic enzymes can therefore serve as model systems for understanding enzyme evolution, molecular mechanisms for protein thermostability, and adaptation of enzymes to their unusual environments.

Hyperthermophilic enzymes have unique characteristics such as temperature, chemical, and pH stability; they have been used in several industrial processes, in which they replace mesophilic enzymes or chemicals (Demirjian et al., 2001). These thermostable enzymes are often used when the enzymatic process is compatible with existing (high-temperature) process conditions. The main advantages of performing processes at higher temperatures are reduced risk of microbial contamination, lower viscosity, improved transfer rates, and improved solubility of substrates. However, cofactors, substrates, or products might be unstable or other side reactions may occur. Recent developments show that hyper/thermophiles are good sources of novel catalysts that are of great industrial interest (Gomes et al., 2004). Thermostable polymer-degrading enzymes such as amylases, pullulanases, xylanases, proteases, and cellulases are expected to play an important role in food, chemical, pharmaceutical, paper, pulp, and waste-treatment industries (**Table 1-1**). Lots of molecular adaptations have been identified which involve thermal stabilization of major cell components of thermophiles and hyperthermophiles: nucleic tides, proteins and so on. In particular, considerable research efforts have been made to better understand the stability of hyper/thermophilic enzymes.

Table 1-1 Industrial applications of the enzymes isolated from extremopiles
(Demirjian et al., 2001)

Extremophile	Habitat	Enzymes	Applications
Hyper/thermophile	High temperature	DNA polymerases	Genetic engineering
		Proteases	Baking, brewing, detergents
		Xylanases	Paper bleaching
		Amylases	Glucose, fructose for sweeteners
Psychrophile	Low temperature	Proteases	Cheese maturation, dairy production
Acidophile	Low pH	Sulfur oxidation Chalcopyrite concentrate	Desulfurization of coal, Valuable metals recovery
Alkalophile	High pH	Cellulases	Polymer degradation in detergents
Halophile	High salt concentration	Whole microorganism	Ion exchange resin regenerant disposal, producing poly(γ-glutamic acid) (PGA) and poly(β-hydroxy butyric acid) (PHB)
Piezophile	High pressure	Whole microorganism	Formation of gels and starch granules
Metalophile	High metal concentration	Whole microorganism	Ore-bioleaching, bioremediation, biomineralization
Radiophile	High radiation levels	Whole microorganism	Bioremediation of radionuclide contaminated sites

1.2.2 Factors affecting enzyme thermostability

Attributes governing protein thermostability have not been completely defined yet, while the recent work on thermophilic enzymes helps to understand the types of general trends that factor into stability. Studies of thermostable enzymes suggest that there is no single factor that has been identified to cause stability, and a number of noncovalent features are important (Demirjian et al., 2001).

Extreme environmental conditions require optimized interactions within the protein, at the protein–solvent boundary, or with the influence of extrinsic factors such as metabolites, cofactors, and compatible solutes (Ladenstein et al., 1998). Factors that contribute to the remarkable stability of enzymes could include an increased number of ion pairs, reduction in the size of loops and in the number of cavities, reduced ratio of surface area to volume, changes in specific amino acid residues, increased hydrophobic interaction at subunit interfaces, changes in solvent-exposed surface areas, increase in the extent of secondary structure formation and truncated amino and carboxyl termini (Baneyx, 2004).

From the view of primary structure, protein amino acid composition has long been thought to correlate to its thermostability (Vieille et al., 2001). Researchers have found high numbers of hydrophobic residues in the protein cores, reduced surface-to-volume ratios, decreased glycine contents, high numbers of surface ionic interactions and shortened loose N- and C-terminal regions, all of which contribute to overall thermostability in specific proteins (Adams et al., 1995; Ladenstein et al., 1998). The

overall protein stability could be increased upon internalization of the more labile amino acids in the hydrophobic core. The frequency of occurrence of labile amino acids such as cysteine, asparagine, and aspartic acid is significantly lowered in thermostable proteins with respect to their mesophilic counterparts (Robb et al., 1999). Additionally, hyperthermophilic proteins have the reduced average length compared to mesophilic proteins (Baneyx, 2004), which could be related to the reduction of loose N- and C-terminal regions and the loss of destabilizing loops. It is also known that the general trend of increased surface ionic interactions is accomplished by the replacement of uncharged polar residues (Gln, Asn, Ser, Thr and Cys) with the charged residues (Arg, Lys, His, Asp and Glu) (Chakravarty et al., 2000).

The protein thermostability has been studied on a level of secondary structure as well. It was observed that helices of hyperthermophilic proteins were generally more stable than those of mesophilic proteins and a trend was detected as a decreasing content in β-branched residues (Val, Ile, and Thr) in the helices of thermophilic proteins (β-branched residues are not as well tolerated in helices as linear residues are) (Facchiano et al., 1998). Researchers have compared the temperature dependence of the unfolding kinetics of rubredoxins from the hyperthermophile *P. furiosus* and the mesophile *Clostridium pasteurianum* to probe kinetic stability factors. The results indicated that the more thermostable protein unfolds at a much slower rate, but exhibits increased sensitivity to pH variation (Cavagnero et al., 1998). Addtionally, cystal structure analysis of the proteins shed lights on the factors of thermostability. The crystal structures of the multisubunit glutamate dehydrogenase isolated from two closely related

hyperthermophiles *T. litoralis* (grow optimally 88 °C) and *P. furiosus* (grow optimally 100 °C) were compared. The less stable *T. litoralis* enzyme had a decreased number of ion-pair interactions, modified patterns of hydrogen bonding, substitutions that decrease packing efficiency and substitutions that give rise to subtle shifts in main- and side-chain elements of the structure (Britton et al., 1999).

On the other hand, thermostable enzymes show considerable similarity with their mesophilic homologous enzymes, and share similar catalytic mechanisms. For instance, the recombinant β-glucosidase from the hyperthermophile *P. furiosus* and the mesophilie *Agrobacterium faecalis* were compared. The enzymes were found to exhibit similar broad substrate specificities and nearly identical pH dependencies with several different substrates, as well as similar inhibition constants with various inhibitor types, which indicate the enzymes exhibit similar catalytic mechanisms despite of a large difference in temperature tolerance (Demirjian et al., 2001).

1.3 ALCOHOL DEHYDROGENASES

1.3.1 Alcohol dehydrogenases from hyperthermophilic bacteria and archaea

One of the most interesting enzyme groups from hyperthermophilic microorganisms is alcohol dehydrogenases (ADHs, EC 1.1.1.1). ADHs are ubiquitous in nature--widely distributed in all three domains of life (Bacteria, Archaea, Eucarya) and have been characterized from diverse sources including bacteria, yeasts, plants, and tissues from several mammalian species (Hirakawa et al., 2004). In a number of hyperthermophiles, including species of *Pyrococcus*, *Thermococcus* and *Thermotoga*, ADH has been

considered as a key enzyme responsible for the production of alcohol including ethanol, one of the most desirable renewable energy sources, as end product. In general, as an important family of oxidoreductases, ADHs can catalyze the inter-conversion between alcohols and the corresponding aldehydes or ketones. Morevoer, they display different physical and enzymatic properties, show a wide variety of substrate specificities (Radianingtyas et al., 2003), and react with primary and secondary, linear and branched chain, aliphatic and aromatic alcohols and with their corresponding aldehydes and ketones. ADHs play considerable roles in processing and producing, for example, the generation of potable alcohol (Lamed et al., 1980) and solvents (Reid et al., 1994). ADHs are also involved in the growth of methylotrophs, oxidation of alcohols and catalyse lignin degradation (Reid et al., 1994). Many studies have been undertaken to characterize ADHs from hyperthermophiles to better understand their activities and thermostability as well as industrial application in alcohol and enzyme production (Lamed et al., 1980). There is considerable interest in the use of ADHs as potential biocatalysts in the chemical synthesis industry, especially the chiral chemical production (Simon et al., 1985).

The classification of ADHs is usually based on their substrates or cofactors. The functionally and presumably structurally similar ADHs are classified as primary (P-ADH) or secondary (S-ADH) based on their higher catalytic efficiencies toward primary (1°) or secondary (2°) alcohols (Burdette, 1997). The secondary alcohol dehydrogenases, which show stereoselective reaction mechanisms, are highly desirable for synthetic chemistry (Whitesides et al., 1985). On the other hand, several categories of ADHs can be distinguished based on their cofactor specificity, these being: (i) NAD or NADP, (ii)

the pyrrolo-quinoline quinine, haem or cofactor F_{420}, and (iii) FAD. ADHs can also be classified into three categories by the metal ions contained: short-chain ADHs (lack of metal ions), zinc-containing ADHs and iron-dependent ADHs (Reid and Fewson, 1994). The phylogenetic analysis revealed that ADH subgroups are classified according to their cofactor specificity in the NAD(P)-dependent group, the ADHs are orderly clustered according to enzyme type, i.e. zinc-dependent ADHs, short-chain ADHs, and Fe-containing ADHs, irrespective of archaeal or bacterial origin, location of isolation or growth conditions (**Fig. 1-2**; Jeon et al., 2008).

Up to date, there are quite a few alcohol dehdyrogenases identified from thermophiles or hyperthermophiles, and several of the enzymes have been biochemically characterized or functionally expressed in foreign hosts to further explore the potential application in industry, since some of thermostable dehydrogenases that are useful in stereoselective transformation of ketones to alcohols have aroused large research interests (Demirjian et al., 2001). Multiple NAD(P)-dependent ADHs which contain zinc or iron at the active site have been studies. In hyper/thermophilic archaea, several kinds of ADHs have been discovered, i.e. the ADH from *Sulfolobus solfataricus* and *Aeropyrum pernix* K1 are NAD-dependent zinc-containing ADHs (Giordano et al., 1999; Hirakawa et al., 2004); ADH from *Pyrococcus furiosus* detected to be an unusual iron- and zinc containing enzyme is liable when expose to oxygen (Ma et al., 1999). Putitive NADP-dependent *Thermotoga maritia* ADH (TM0820) was detected to be iron-containing based on the crystal structure. This ADH variation within hyper/thermophiles from different sources

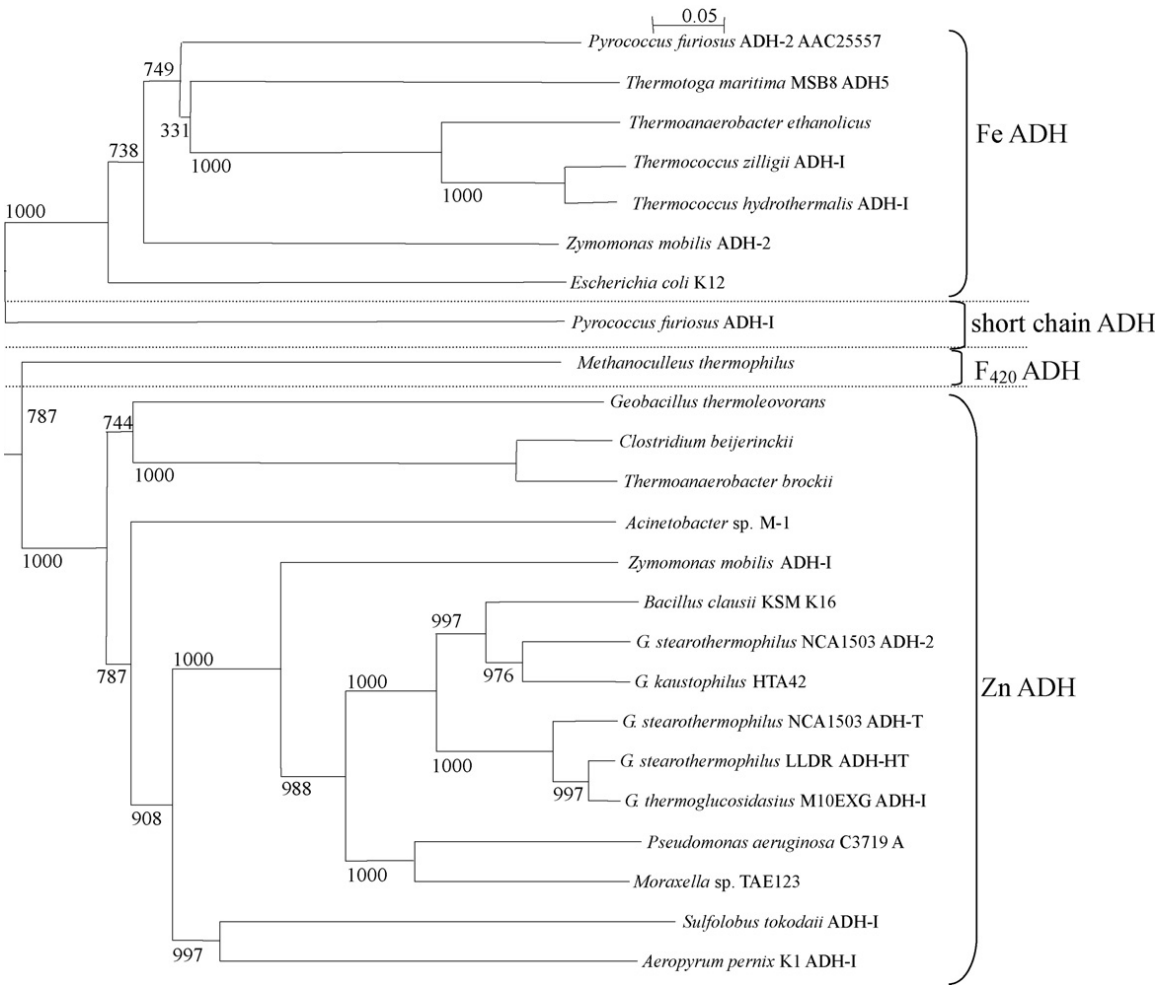


Figure 1-2 Phylogenetic tree derived from NAD(P)-dependent ADHs

Phylogenetic analysis was adapted performed with a selection of 23 sequences covering all representatives of the NAD(P)-dependent class of this family (Fe-dependent, short-chain, and Zn-dependent) from thermophilic, mesophilic, and psychrophilic organisms (Adapted from Jeon et al., 2008).

could provide evidence that the genetic and metabolic diversity present in high-temperature environments reflects the stress components that characterise the environments (Radianingtyas et al., 2003).

1.3.2 Zinc-containing ADHs

Most NAD (P)-dependent ADHs contain zinc at the active site (Radianingtyas et al., 2003). Generally, most zinc-containing ADH enzymes found exist as tetramers, such as those present in yeast and bacteria; however, some of them were detected to be dimmers (Radianingtyas et al., 2003). Zinc-containing ADHs that are dimmers usually found in higher plants and mammals, while some from the microorganisms were also identified, such as the ADH from *Rhodococcus rubber* DSM 44541 (Kosjek et al., 2004), *Thermomicrobium roseum* (Yoon et al., 2002), ADH from *Pyrococcus furiosus* (Ma et al., 1999), and recently even a dodecameric ADH were identified from *Picrophilus torridus* (Hess et al., 2008).

Zinc ions have catalytic or structural functions in several enzymes including hyperthermophilic zinc-containing ADHs. Generally, zinc is essential for the activity; however, several zinc-containing ADHs contain structural zinc. The accumulation of structure data about zinc-containing ADHs leads to further classification based on the role of zinc: ADHs with both catalytic zinc and structural zinc or ADHs with only catalytic zinc.

Thermophilic NAD-dependent homotetrameric alcohol dehydrogenase from the archaeon *Sulfolobus solfataricus* was the firstly well-studied thermophilic ADH containing both catalytic and structural zinc ion (Ammendola et al., 1992). Each monomer contains two Zn ions with catalytic and structural function, respectively. Despite few structural variations, all of the coordinating residues of *S. solfataricus* are located within typical sequence motifs of all zinc-containing ADHs (Vallee et al., 1990). Most Zinc-containing ADHs have three residues, including cystine or histidine, as active-site ligands, while in *S. solfataricus* the catalytic zinc is bound to Cys_{38}, His_{68}, Cys_{154} (numbering *in S. solfataricus*) which are conserved coordinating residues of the catalytic zinc atom, and another protein Zn-ligand, Glu_{69}, is also found to be strongly bound to Zn (Raia et al., 2001). The other example from hyperthermophiles is $_L$-threonine dehydrogenase (TDH) from *P. horikoshii* involving in the oxidation of $_L$-threonine to 2- amino-3-ketobutyrate, which is an NAD-dependent homotetramer containing one catalytic zinc and one structural zinc of each subunit (Ishikawa et al., 2007). The structural zinc ion exhibits coordination with four cysteine ligands (Cys_{97}, Cys_{100}, Cys_{103} and Cys_{111}, numbering *in P. horikoshii*) conserved throughout the structural zinc-containing ADHs and TDHs. The residues Cys_{42}, His_{67} and Glu_{68} coordinated to a functional catalytic zinc ion. However, the catalytic zinc ion has a larger interdomain cleft and is not coordinated at the bottom of the cleft in the crystal of *Pyrococcus horikoshii* TDH, which is a significant difference in the orientation of the catalytic domain (Ishikawa et al., 2007).

Besides the ones with both catalytic and structural zinc ion, some members of the zinc-containing ADH were found to harbor catalytic zinc only. The representatives of this

group are ADHs from thermophiles *Thermoanaerobacter brockii* and *Thermoanaerobacter ethanolicus*. Both of the enzymes are NADP-dependent and contain one zinc ion per subunit (Peretz et al., 1989; Burdette et al., 1997). Cys_{37}, His_{59} and Asp_{150} coordinate the single catalytic zinc. The structures of NADP-dependent ADHs from *T. brockii* has been determined (Bogin et al., 1997; Korkhin et al., 1998), and the monomers are composed of two domains, a coenzyme-binding domain and a catalytic domain, separated by a deep cleft at the bottom of which a single zinc atom is bound in the catalytic site. The *T. ethanolicus adhB* gene was cloned and expressed in *Escherichia coli* (Burdette et al., 1996) and *T. ethanolicus* ADH was characterized to be a medium chain, zinc-containing, tetrameric ADH composed of identical 40 kDa subunits. This NADP-dependent enzyme contains single catalytic zinc coordinated by Cys_{37}, His_{59} and Asp_{150}, conserved in the *T. brockii* ADH (Bogin et al., 1997).

Thermococcus species are the best-studied hyperthermophilic anaerobes and a large number of species have been already described. Many *Thermococcus* ADHs contain iron, such as *Thermococcus strain* ES-1 has the sulfur-regulated, nonhaem iron ADH (Ma et al., 1995), and ADH from *Thermococcus hydrothermalis* is iron-dependent enzyme using NADP as cofactor (Antoine et al., 1999). Interestingly, an ADH purified from *Thermococcus guaymasensis* was determined to be a zinc-containing, NADP-dependent secondary alcohol dehydrogenase (Ying et al., unpublished results). Containing one single zinc atom at the catalytic core, the *T. guaymasensis* ADH has the outstanding thermostability and a higher resistance to oxygen than the iron-containing ADHs from *Thermococcus* species.

1.3.3 Iron-containing ADHs

Although many zinc-containing ADHs have been well studied, only a few iron-containing ADHs are known, few have been described in hyperthermophiles. And also, the importance and catalytic mechanism of iron in such enzymes are not well understood. So far, most of iron-dependent ADHs that still contain iron after purification are found in hyperthermophilic archaea, including members of the genus *Thermococcus*: *T. litoralis* (Ma et al., 1994), *Thermococcus* strain ES1 (Ma et al., 1995), *T. zilligii* (Li et al., 1997; Ronimus et al., 1997), and *T. hydrothermalis* (Antoine et al., 1999). In one of the well-studied hyperthemophilic bacterial genera *Thermotoga*, the gene TM0820 was determined to encode a NADP-dependent iron-containing enzyme (Schwarzenbacher et al., 2004). The bacterial *T. hypogea* ADH is the first purified and characterized iron-containing ADH from the hyperthermophilic bacteria and is one of the most thermostable iron-containing ADHs (Ying et al., 2007).

Several genes encoding metal free and Zn-containing ADHs from hyperthermophiles have been successfully expressed in *E coli* to yield recombinant forms (Cannio et al., 1996; van der Oost et al., 2001; Hirakawa et al., 2004; Kube et al., 2006; Machielsen et al., 2006), while the expression of genes encoding iron-containing ADHs from hyperthermophiles has remained a challenge. Expression of the gene encoding the iron-containing ADH from *T. hydrothermalis* in *E. coli* is the first example of heterologous production of an iron-containing ADH from a hyperthermophilic archaeon, however, the native and recombinant *T. hydrothermalis* ADH showed clear differences in catalytic

activity and thermostability (Antoine et al., 1999). Recently, an iron-containing ADH from *Thermococcus* strain ES1 was successfully expressed in *E. coli*, and the recombinant and native enzyme were shown to have similar catalytic characteristics (Ying et al., 2008).

Only few of the 3-D structure of iron-containing alcohol dehygrogenases are available currently. $_{L}$-1, 2-propanediol dehydrogenase is the first bacterial iron-containing ADH whose 3-D structure was analyzed. $_{L}$ -1, 2-propanediol dehydrogenase expressed in *E. coli* forms a dimer, in which each monomer folds into an α/β dinucleotide-binding N-terminal domain and an all-α-helix C-terminal domain that are separated by a deep cleft. The metal ion is coordinated with an aspartate residue (Asp_{196}) and three histidine residues (His_{200}, His_{263}, and His_{277}, numbering in *E. coli*). The coenzyme NAD binding site is harbored in the α–helix close to the N-terminal site and conserved among dehydrogenases. The enzyme encoded by gene TM0920 from the hyperthermophile bacterium *T. maritima* is a putative NADP-dependent ADH and the crystal structure has been obtained (Schwarzenbacher et al., 2004). Each monomer of the enzyme has two distinct N-terminal and C-terminal domains, separated by a deep cleft. The ferrous ion is deeply located in the catalytic cleft and has a square pyramidal coordination with Asp_{189}, His_{193}, His_{256}, and His_{270} (numbering in *T. maritima*), all of which are situated in C-terminal domain.

An alcohol dehydrogenases from *T. hypogea* represents the first hyperthermophilic bacterial ADH that contains iron with full activity after purification, whose catalytic

properties show similarities to the enzymes in archaea (Ying et al., 2007). The N-terminal sequence of *T. hypogea* ADH has no similarity to archaeal ADHs, which is an indication of the divergence of iron-containing ADHs from hyperthermophilic archaea. However, there is still a need of more sequence information for understanding the evolutionary relationship between the bacterial iron-containing ADH and the archaeal ADHs. Because of the oxygen sensitivity, the iron-containing fully active ADHs including the *T. hypogea* ADH cannot be easily purified from hyperthermophiles, which is probably one of the reasons why there is a lack of understanding of the iron-containing ADH and the catalytic mechanism of iron in such enzymes. To achieve better understanding of these types of ADHs, sufficient amount of pure enzymes that can be obtained by over-expressing hyperthermophilic enzymes in mesophilic host is needed for further study.

1.4 RECOMBIANT HYPER/THERMOPHILIC ENZYMES

ADHs from thermophiles or hyperthermophiles can be used for the study of enzyme evolution, enzyme stability and catalytic mechanisms, protein structure-function relationships, and biocatalysis under extreme condition. However, it is difficult to obtain native enzymes because of the difficulties in cultivating the organisms on the large scale, low basal level expression of the enzymes in native hosts, and complex purification process (Demirjian et al., 2001). In addition, the lack of suitable expression systems in particular extremophiles has prevented the high level production of the target enzymes from the source microorganisms. Recently, based on the discovery of molecular and biochemical studies such as protein purification and characterization, the applications of the hyperthermophilc enzymes are greatly facilitated by the cloning and expressing of

genes from hyperthermophiles in mesophilic hosts, which allows for the tagging or modification of the recombinant proteins for the purpose of simplifying downstream purification (Baneyx, 2004).

Compared to the native enzymes from hyperthermophiles, the recombinant enzymes have the benefits of purity, consistency, and affordability. Although recombinant enzymes and native enzymes are manufactured to meet the same rigorous quality control standards, it is recombinant enzymes that produce a more pure product with less processing time (Baneyx, 2004). In particular, hyperthermophilic enzymes that are intrinsically stable and active at high temperatures can be easily purified by heat treatment. Typically, the yields obtained for recombinant and over-expressed enzymes are significantly larger than those produced by native strains. Moreover, in industry, the introduction of recombinant enzymes has resulted in lower cost that spreads the application of the enzymes (Vieille et al., 2001).

Though initially thought that some of the factors such as difference in codon usage or GC content of the hyperthermophilic genes compared to the host would pose obstacles for successful high-level expression, however, with the continuous improvements of new systems, over-expression of various proteins from hyper/thermophiles have conducted successfully in traditional expression hosts such as *E. coli*, yeast and *Bacillus subtilis*, and in general, most recombinant enzymes are obtained using the mesophilic *E. coli* as the host system. The most commonly used vector of the *E. coli* systems is the T7 promoter based expression vectors (Tabor et al., 1985; Studier et al., 1986), with ~60% of the

recombinant extremophilic proteins beings produced with this system. The second most often used *E. coli* expression system for extremophilic proteins production is *lac*-promotor-based expression vectors. More than 100 genes from hyperthermophiles have been cloned and expressed in mesophiles (Hirakawa et al., 2004). Most of the genes from hyperthermophiles have been directly cloned after PCR amplification or isolated by hybridization (Cannio et al., 1996), only few other genes have been isolated by direct expression and activity screening (i.e., by complementation of growth or activity assay) of a genomic library of the host cell (Vieille et al., 2001).

When the properties of the native and recombinant hyperthermophilic enzymes are compared, the majority of hyperthermophilic enzymes expressed in *E. coli* retain all of the native enzyme's biochemical properties, including proper folding (Grättinger et al., 1998), thermostability, and optimal activity at high temperatures (Arnone et al., 1997). The crystallization studies of crystal structures of both the native and recombinant enzymes indicated recombinant hyperthermophilic proteins are typically similar to that of their mesophilic homologues (Knapp et al., 1996; Yip et al., 1995). Thus, while a few proteins from hyperthermophiles might require extrinsic factors (e.g., salts or polyamines) or posttranslational modifications (e.g., glycosylation) to be fully thermostable, most proteins from hyperthermophiles are intrinsically thermostable, and they can fold properly even at 60 °C, below their physiological conditions (Danson et al., 1996). The fact that most hyperthermophilic enzymes are properly expressed and folded in *E. coli* has greatly facilitated their study, since they can be cultured under aerobic

growth conditions and purified from *E. coli* rather than from an often hard-to-grow hyperthermophilic organism.

On the other hand, difficulties also appear in the over-expression of hyperthermophilic enzymes in mesophiles because of their evolutionary gap. Genes from hyperthermophiles in *E. coli* can be low expression due to a significantly different codon usage in the expressed gene. This difficulty is often alleviated by the expression in *E. coli* of rare tRNA genes together with the target gene (Uriarte et al., 1999; Machielsen et al., 2006). Furthermore, since archaeal transcription systems (including promoter sequences) are more closely related to eucaryal than to bacterial systems, most archaeal genes can be expressed in *E. coli* only when they are cloned under the control of strong promoters, such as Plac, Ptac, or T7 RNA polymerase promoter.

There are quite a few ADHs from hyper/thermophilic sources have been over-expressed in mesophilic host *E. coli* (**Table 1-2**). *T. hydrothermalis* ADH is the first *Thermococcale* ADH to be cloned and over-produced in a mesophilic heterologous expression system (Antoine et al., 1999), the *adhB* gene encoding thermophilic bacterium *T. ethanolicus* 39E secondary-alcohol dehydrogenase (*S*-ADH) has also been cloned, sequenced and expressed in *E. coli* (Burdette et al., 1996). The recombinant *T. ethanolicus* protein, with kinetic properties similar to those of the native enzyme, is a thermophilic and thermostable NADP-dependent enzyme that exhibits significantly greater catalytic efficiency towards secondary than primary alcohols. The molecular cloning, complete nucleotide sequences, and high-level expression have been done in

Table 1-2 Over-expressed ADHs from hyper/thermophiles in mesophilic host *E. coli*

Metal ion	Source Organism	Expression System	Reference
Zinc	*Aeropyrum pernix*	*E. coli*- pTRC99 system (T7 promoter)	Guy et al., 2003
Zinc	*Geobacillus thermoglucosidasius*	*E. coli*-pET system (T7 promoter)	Jeon et al., 2008
Zinc	*Pyrococcus horikoshii*	*E. coli*-pET system (T7 promoter)	Higashi et al.,2005
Zinc	*Sulfolobus solfataricus*	*E. coli*- pTRC99 system (T7 promoter)	Cannio et al., 1996
Zinc	*Thermoanaerobacter brockii*	*E. coli*-pET system (T7 promoter)	Peretz et al., 1997
Zinc	*Thermoanaerobacter ethanolicus*	*E. coli*-pET system (T7 promoter)	Ziegelmann-Fjeld et al., 2007
Zinc	*Thermococcus hydrothermalis*	*E. coli*-pET system (T7 promoter)	Antoine et al., 1999
Iron	*Thermotoga maritima*	*E. coli*	Schwarzenbacher et al., 2004
Iron	*Thermococcus strain ES-1*	*E. coli*-pET system (T7 promoter)	Ying et al., 2008
None	*Pyrococcus furiosus*	*E. coli*-pTZ19R(T7 promoter)	van der Oost et al., 2001

E. coli of the cloned genes encoding ADH in the thermophile *T. brockii* (Peretz et al., 1997). It was found that the level of *T. brockii* ADH expression in *E. coli* was 30-fold higher than in the native bacterium *T. brockii*, while the kinetic properties with C_3 to C_5 substrates and the thermostability of the purified recombinant ADHs were equivalent to those of the enzymes purified from the respective native organisms. And recently, a thermostable NAD-dependent Zinc-containing alcohol dehydrogenase from *Geobacillus thermoglucosidasius* has been functionally expressed in *E. coli* and the catalytic properties were characterized (Jeon et al., 2008).

A few genes from hyperthermophilic archaea have been successfully expressed in yeast systems as an alternative of *E. coli* (Riley, 1993). That is because archaea have some eukaryotic-like characteristics, such as having histone-like proteins and eukaryotic-like transcription and translation initiation (Bell et al., 1998). The majority of the obstacles of yeast expression were reported as protein failed to fold properly into monomeric subunits in the endoplasmic reticulum (ER) and, therefore, was not subsequently trafficked to the yeast secretory pathway. *S. cerevisiae* has been used to express the β-glucosidase from *P. futiosus* (Smith et al., 2002). In this case, the β-glucosidase gene was expressed using the yeast Gal1-10 promotor and 10 mg/L of recombinant protein was obtained. It was determined that increasing the cultivation temperatures from 30 to 37 ^{0}C improved the protein yield, so higher expression temperatures could be necessary for the proper folding of the β-glucosidase subunits. Other yeast expression strains have also been used for the recombinant expression of extremophilic proteins. Xylanase from the hyperthermophilic

stain *Thermotoga* strain FjSS3B.1 was successfully expressed in the yeast strain *Kluyveromuces lactic* (Walsh et al., 1998).

Another alternative expression host *Bacillus subtilis* has recently been used, especially for the production of proteins from alkaliphiles. For example, the recombinant expression of the alkaliohilic pectate lyase in *Bacillus subtilis* resulted in high yield using the *Bacillus* shuttle vector pHY300PLK (Ishiwa et al., 1986). *B. subtilis* has also been used to expression proteins from thermophiles, and become a powerful approach when expression level in *E. coli* is poor. The pullulanase from *Desulfurococcus mucosus* was expressed in very low level in the host *E. coli* because of the different codon biases between thermophiles and *E. coli*. However, the *D. mucosus* gene was expressed successfully when cloned into the *B. subtilis* expression vector pJA803 and under control of the *Bacillus* maltogenic α-amylase (P*amyM*) promoter (Duffner et al., 2000). The pullulanase gene reflected the codon biases present in *B. subtilis* better than that of *E. coli* since it contains argnine, isoleucine and leucine codons that are considered rare in *E. coli* (Duffner et al., 2000). With the increasing of potentially useful alkaliphilic proteins, it is anticipated that the recent development of new *Bacillus* expression systems will continue to occur.

Due to the evolution gap between hyperthermophiles and mesophilc hosts, especially difference between archaeal and bacterial transcription systems, the using of anaerobic thermophilic or hyperthermophilic archaea as cell factories for thermostable protein production has been constructed. However, this system is limited because of the low cell density of these archaea in fermentation processes (Biller et al., 2002). Up to date, few

high-temperature expression systems are available. A shuttle vector that maintained a high copy number in *Pyrococcus abyssi* was constructed (Lucas et al., 2002). Alternatively, *Sulfolobus* species can be used as thermophilic host organism since it grows to higher cell densities (Krahe et al., 1996; Schiraldi et al., 1999). Researchers have developed a generic system for a *Sulfolobus sulfataricus* host and over-expressed ADH from the moderate thermophile *Bacillus stearothermophilus* (Contursi et al., 2003). It is obvious that the thermophile expression systems have their advantages but also require optimization.

In conclusion, higher level production of ADHs using heterologous over-expression is obviously essential for biotechnological application, and can also be extremely useful for further characterization studies, such as determining three-dimensional (3-D) structures and structure/function studies using mutagenesis. However, no general rules can be provided for the production of different proteins, and sometimes a single amino acid change results in dramatic differences. Still, there is a need for the optimization of the heterologous production of thermostable enzymes in general.

1.5 AIM OF THIS STUDY

Hyperthermophiles are located at the bottom of the phologenetic tree in both domains of bacteria and archaea, and represent a broad spectrum of growth physiologies. Hyperthermophilic enzymes may serve as model systems in understanding enzyme evolution, molecular mechanisms for protein thermostability, and adaptation of enzymes to their unusual environments. ADHs are ubiquitous in nature and responsible for diverse metabolic functions. The study of ADHs is important to expand not only the understanding of hyperthmophilic physiology but also their potentials as potent biocatalysts in industry. The aim of this research is to obtain a new version of ADH from both hyperthermophilic bacteria and archaea as model systems for studying thermostability and catalytic mechanisms. This work mainly focused on zinc-containing ADH from *T. guaymasenis* and iron-containing ADH from the bacterium *T. hypogea*. Since metal ions are essential for maintaining the activity of some biocatalysts *in vitro*, and their presence may result in enhancements of selectivity, activity as well as stability. As the *T. guaymasenis* ADH and *T. hypogea* ADH contain zinc and iron respectively, it is interesting to investigate effect of different divalent ions on catalytic characteristics or oxygen sensitivity of the enzymes.

To simplify the cell cultivation and protein purification as well as to increase the protein yield, the entire encoding gene was cloned, sequenced and over-expressed in the mesophilic host *E. coli* following the strategy outlined in **Figure 1-3**.

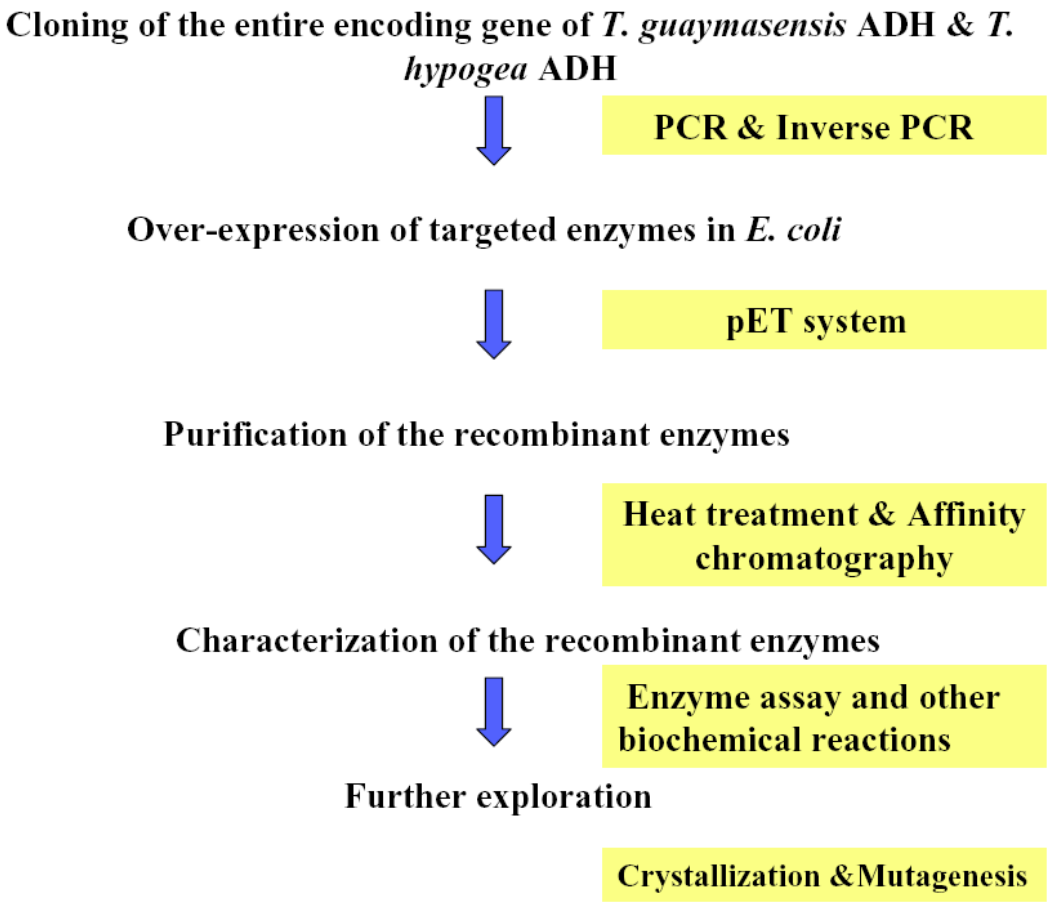


Figure 1-3 Brief strategy of this study

Further biochemical characterization would be effective in determining catalytic properties of recombinant enzymes, including optimal catalytic conditions, thermostabilities, and comparision with the native enzymes from hyperthermophiles. Differences between archaeal and bacterial hyperthermophilic enzymes will be analysed, which will be helpful to understand the evolutionary relationship between the bacterial iron-containing ADH and the archaeal ADHs. Finally, potential approaches in industrial process of thermostable enzymes will be discussed. Further studies on crystal structures of the recombinant enzymes and engineering of the enzymes may provide comprehensive understanding of the thermostable enzymes. Site-directed mutagenesis on the surface areas, cofactor binding motif and ion binding site may modify the catalytic characteristics, substrate selectivity or even create additional nonlocal ion pairs, and it is expected to obtain biocatalysis under extreme conditions with broader substrate specificities and new functions targeted at industrial applications.

Chapter 2: Sequence determination and functional expression of a Zinc-containing alcohol dehydrogenase from *Thermococcus guaymasensis*

2.1. OVERVIEW

Thermococcus guaymasensis is a hyperthermophilic starch-degrading archaeon producing acetate, CO_2, H_2, ethanol and acetoin as end products. An alcohol dehydrogenase (ADH) was characterized previously to be a primary-secondary ADH and exhibited stereoselectivity on interconversion between 2, 3-buanediol and acetoin. The gene encoding the enzyme was cloned using both conventional and inverse PCRs. The deduced amino acid sequence had a total of 364 amino acid residues with a calculated molecular mass of 39463 Da. The enzyme belongs to the family of zinc-containing ADHs with catalytic zinc only. It was verified that the enzyme had binding motifs of catalytic zinc only ($GHEX_2GX_5GX_2V$, residues 62-76) and cofactor NADP ($GXGX_2G$, residues 183-188). The tertiary structural modeling showed two typical domains, one catalytic domain close to N-terminal and one coenzyme-binding domain close to C-terminal end. Since its codon usage pattern of which seemed to be different from that of *E. coli*, the enzyme was over-expressed in the *E. coli* codon plus strain using pET-30a vector. The recombinant enzyme was soluble and active (1073 U/mg), which was virtually the same to the native enzyme (1049 U/mg). The recombinant possessed almost identical properties with the native enzyme. The optimal pH values for ethanol oxidation and acetaldehyde reduction were 10.5 and 7.5 respectively, while activity for alcohol oxidation was much higher than that of aldehyde reduction. The enzyme activity was inhibited in the presence of 100 μM Zn^{2+} in the assay mixture and it has a half-life of 6 hours after exposed to air. The enzyme had outstanding thermostability with 60% activity after incubation at 80 ^{0}C for 40 hours.

2.2 INTRODUCATION

Interconversions of alcohols, aldehydes, and ketones are essential processes in both prokaryotes and eukaryotes. As one of the key enzyme in metabolism pathways, ADHs are involved in a wide range of metabolism, especially in microorganisms (Radianingtyas et al., 2003). Hyperthermophiles are a group of microorganisms growing optimally at temperatures higher than 80 ^{0}C, of which anaerobic heterotrophs have attracted increasingly attention for the use of fermentation at the elevated temperatures (Karakashev et al., 2007). *Thermococcus guaymasensis* is a strictly anaerobic archaeon isolated from marine sediments (Canganella et al., 1998). It is a member of obligately heterotrophic and strictly anaerobic archaea with the optimum temperature for growth at 88 ^{0}C. Acetate, CO_2 and H_2 are formed as the main end products of metabolism, and ethanol can also be produced.

Up to date, alcohol dehydrogenases have been characterized from many sources. Based on the cofactor specificity, they can be divided into 3 types: NAD(P)-dependent, FAD-dependent, and pyrro-quinoline quinine (Duine 1980), haem or cofactor F_{420}. Those NADP-dependent ADHs can be further classified into 3 groups based on the metal content: zinc-containing ADHs, iron-containing or activated ADHs and the enzymes without metal ions. The native ADH purified from *T. guaymasensis* was determined to be a Zinc-containing, NADP-dependent secondary alcohol dehydrogenase (Ying et al., unpublished results).

Zinc-containing ADHs from hyperthermophiles are highly desired as promising catalysts in industry because of the features such as solvent tolerance, high stereoselectivity as well as thermostability. For instance, a zinc-containing ADH from anaerobic archaeon *Pyrococcus furiosus* underwent asymmetric ketone reduction to the corresponding chiral alcohols (Machielsen et al., 2006; Zhu et al., 2006). The *T. guaymasensis* ADH (*Tg*ADH) has the outstanding thermostability and stereoselectivity as well. So there is a need to produce large amount of recombinant ADH for further biophysical and biochemical studies as well as for industrial applications.

Complete genomic sequences of some hyperthermophilic archaea are available in the Genebank (http://www.ncbi.nlm.nih.gov/Genbank/) and several Zinc-containing alcohol dehydrogenases have been successfully expressed in mesophilic heterologous expression systems. Further analysis of 3-D structure also provided valuable information of their catalytic mechanisms. The NADP-dependent ADH from the thermophilic bacterium *Thermoanaerobium brockii* was over-expressed in *Escherichia coli* and crystallized in its holo-enzyme form (Korkhin et al., 1998). *T. hydrothermalis* ADH is the first *Thermococcale* ADH to be cloned and over-produced in mesophilic host *E. coli*. Additionally, the X-ray structure of a tetrameric NAD-dependent ADH from the hyperthermophilic archaeon *Sulfolobus solfataricus* has been solved (Esposito et al., 2002). The gene encoding ADH from *Aeropyrum pernix* was over-expressed in *E. coli* and the 3-D structure of the enzyme was obtained by crystallization with presence of its cofactor NADH (Guy et al., 2003).

In this study, the structure gene encoding *Tg*ADH was cloned, sequenced and over-expressed in mesophilic host *E. coli*, and the resulting enzyme was purified and its catalytic properties were characterized. The over-expression of *Tg*ADH was a successful fundament for further redesign of the enzyme that could better fit the application purpose.

2.3 MATERTIALS AND METHODS

2.3.1 Materials

All the chemicals in this research were purchased from commercially available sources (**Table 2-1**).

Table 2-1 Major chemicals used in this research

Chemicals *	Corporation
Agarose	Fermentas Canada Inc. (ON, Canada)
Acrylamide-Bisacrylamide-Solution	MP Biomedicals (OH, USA)
BIO-RAD Protein Assay	Bio-Rad Laboratories, Inc. (ON, Canada)
5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside	Fermentas Canada Inc. (ON, Canada)
Dithiothreitol	Fisher scientific company (ON, Canada)
Isopropyl β-D-1-thiogalactopyranoside	Fermentas Canada Inc. (ON, Canada)
ß- nicotinamide adenine dinucleotide (NADP)	Sigma-Aldrich Canada Ltd. (ON, Canada)
ß-nicotinamide adenine dinucleotide; di-sodium salt (NADPH)	Sigma-Aldrich Canada Ltd. (ON, Canada)

*, all other chemicals not mentioned here were of high technical grade and obtained from Sigma-Aldrich Canada Ltd. (Oakville, Ontario) or Fisher scientific company (Ottawa, Canada).

Restriction enzymes and DNA ladders for molecular cloning and recombinant plasmid construction enzymes were commercially available (**Table 2-2**), and were used according to the manufacture's instructions. Major instruments used were listed in **Table 2-3**.

Table 2-2 Major chemicals for molecular biology work

Restriction enzymes and reagents *	Cooperation
GeneRuler 100 bp DNA Ladder	Fermentas Canada Inc. (Burlinton, ON, Canada)
KOD Hot Start DNA Polymerase	EMD Chemicals, Inc. (NJ, USA)
PCR Gel Extraction Kit	Qiagen (ON, Canada)
Perfect DNA 1 kb DNA Ladder	Novagen (WI, USA)
Restriction DNA restriction endonuclease	Fermentas Canada Inc. (ON, Canada)
Taq Polymerase	Fermentas Canada Inc. (ON, Canada)
T4 DNA Ligase	Fermentas Canada Inc. (ON, Canada)

*, All other chemicals not mentioned above were obtained from Sigma-Aldrich Canada Ltd. or Fisher scientific company (ON, Canada).

Table 2-3 Major instruments used in this research

Instrument	Corporation
Agarose gel electrophoresis chamber	Bio-Rad Laboratories, Inc. (ON, Canada)
Acrylamide-Bisacrylamide-Solution	MP Biomedicals (OH, USA)
Centrifuge (Allegra 21R Centrifuges)	Beckman Coulter (ON, Canada)
Centrifuge (Sorvale® RC6-Refrigerated Superspeed Centrifuges)	Mandel Scientific Company Inc. (ON, Canada)
FPLC	Amersham Biotech (QC, Canada).
FluorChem 8000 Chemiluminescence and Visible Imaging System	Alpha Innotech Corporation (CA, USA)
Incubation shaker	New Brunswick Science (NJ, USA)
Incubator	Fisher scientific company (ON, Canada)
Microscope	Fermentas Canada Inc. (ON, Canada)
Protein gel chamber	Bio-Rad Laboratories, Inc. (ON, Canada)
Spectrophotometer (GENESYS 10 UV)	VWR Canlab (ON, Canada)
Table centrifuge	Eppendorf (ON, Canada)
Thermal-PCR-cycler TC-312	Techne incorporated (NJ, USA)
Vortex	Fisher scientific company (ON, Canada)
Waterbath	Fisher scientific company (ON, Canada)

2.3.2 Microorganisms

Microorganisms used for this study are listed below:

Thermococcus guaymasensis DSM 11113^{T} was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany.

E.coli DH5α [*(supE44 ΔlacU169 Φ80 lacZΔM15) hsdR17 recA1gyrA96 thi-1 relA1*] (BRL, CA, USA)

E.coli BL21(DE3) [*B F⁻ ompT hsdS$_B$ (r_B^- m_B^-) gal dcm (DE3)*] (Novagen, WI, USA)

E.coli BL 21(DE3)-RIL [*F⁻, ompT, hsdS$_B$ (r_B^- m_B^-) gal dcm lacY1, pRARE (CamR)*] (Stratagene, CA, USA)

2.3.3 Cultivation media and growth conditions

For the growth of *E. coli*, 2YT-medium was used (Sambrook et al., 1989). All media were autoclaved for 30 minutes at 121 ^{0}C.

<u>2YT medium (per liter)</u>

Tryptone 16 g

NaCl 5 g

Yeast extract 10 g

Deionized water was added to 1 L. The solution was autoclaved for 30 minutes at 121 °C.

<u>SOB medium (per liter)</u>

Bacto-peptone 20 g

Bacto-yeast extracts 5 g

NaCl	5.8 g
KCl	0.19 g

Adjust the pH to 7.5 with KOH. After autoclave and cool down, 2.5 ml autoclaved 1 M $MgCl_2$ and 2.5 ml autoclaved 1 M $MgSO_4$ were added.

Medium for *T. guaymasensis* (per liter)

KCl	0.32 g
$MgCl_2.2H_2O$	2.7 g
$MgSO_4.7H_2O$	3.4 g
NH_4Cl	0.25 g
$CaCl_2.2H_2O$	0.14 g
K_2HPO_4	0.14 g
Na_2SeO_3	100 µl (100 mg/ml)
$NiCl_2 \cdot 6H_2O$	100 µl (100 mg/ml)
NaCl	18 g
Bact-yeast extract	5 g
Tryptone-peptone	5 g
Trace mineral	10 ml
Resazurin	2 ml (500 mg/L)
Glucose	5 g

Adjust the pH to 7.0, dispensed 50 ml medium to a 160 ml serum bottle. After autoclave the medium turned to pink. When it cooled down, the medium was degassed and

pressured with nitrogen gas. Then 0.14 ml 15% cystein and 0.17 ml 7% Na_2S were added to each bottle. After incubating the medium at 88 ^{0}C for 5 min, the medium turned to light yellow and it was ready for use.

Stock solutions of antibiotics and reagents

Ampicillin: 1 g /ml in deionized water. Filter sterilized through 0.2 μm filter membranes (Corning NJ, USA). Stored at –20 ^{0}C. Used at 1mg /ml.

Kanamycin: 500 mg/ml in deionized water. Filter sterilized through 0.2 μm filter membranes. Stored at –20 ^{0}C. Used at 0.5 mg/ml.

IPTG (Isopropyl β-D-1-thiogalactopyranoside): 0.1 M in deionized water. Filter sterilized through 0.2 μm filter membranes. Stored at –20 ^{0}C.

X-gal (5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside): Dissolved in dimethylformamide at 20 mg/ml. Stored at –20 ^{0}C in dark.

2.3.4 Cell cultivation

Cell cultivation was carried out either in shake-flasks with baffles or in solidified agar plates with the media composition as described in section 2.3.3.

2.3.4.1 *E. coli* cultivation

E. coli strains were grown in 2YT medium at 37 ^{0}C, 140-200 rpm, in a volume ranging from 500-1000 ml in shake flasks with baffles by inoculating one colony from agar plate or with previous culture in a ratio of 1:100 or 1:50 (1-2% v/v). The ratio between the volume of the media and the volume of the shake flask was 1:5 (e.g. 100 ml media was

used for cultivation in a shake flask having a volume of 500 ml). For isolation of vector material or screening of clones, *E. coli* strains harboring plasmids were grown overnight at 37 ^{0}C in a shaker incubator at 140-160 rpm containing necessary antibiotic in the media composition.

2.3.4.2 *T. guyamsensis* cultivation

T. guyamsensis was cultured at 88 ^{0}C in the medium as described previously (Canganella et al. 1998) with modifications that elemental sulfur and HEPES were omitted. The preparation of trace metal and vitamin solutions was described as previously (Balch et al., 1979). The growth was monitored using cell counting. The culture was taken out from water bath after 18-20 hours incubation, and the culture could be stored at room temperature. For long-term storage, the resulting cell pellet by centrifugation (5,000 × *g*) was frozen in liquid nitrogen immediately and stored at -80 ^{0}C until use.

2.3.5 Conservation and storage of microbiological strains

For long-term storage of *E. coli* cells, cryo-cultures were made with glycerol at –20 ^{0}C or –80 ^{0}C. This method was used for preparation of competent cells as well as preparing stock cultures of *E. coli* cells harboring either pET vectors or pET recombinants. For this, a single colony was picked up from 2YT-agar plate and inoculated into a 5 ml liquid 2YT medium containing appropriate antibiotic if required, incubated at 37 ^{0}C on a shaker with vigorous shaking until the OD_{600} reached 0.6-0.8. Then 0.8 ml of the culture was removed and transferred to a sterilized cryo-vial, and 0.2 ml of 50 % glycerol was added. The culture was mixed well and stored at –20 ^{0}C. Or cells with glycerol (10 %) were frozen in liquid nitrogen quickly and then stored for long-term at –80 ^{0}C.

2.3.6 Preparation for competent cells

2.3.6.1 *E. coli* DH5α high efficiency competent cells

For construction of the cloning vector pGEM-Teasy carrying the TGADH coding gene, the *E. coli* DH5α high efficiency competent cells were prepared following the standard protocol. *E. coli* DH5α cells were amplified into 250 ml SOB medium and grew at 18 ^{0}C (room temperature) and were shaken with 100-110 rpm until a cell density of 0.6 OD_{600nm} was reached.

After harvesting at 4000 rpm for 10 minutes at 4 ^{0}C, the pellet cells were re-suspended in 80 ml pre-cooled transformation buffer (10 mM PIPES, 10 mM $CaCl_2$, 250 mM KCl, 100 mM $MnCl_2$, pH 6.7) slightly, followed by incubation for another 10 minutes on ice. After centrifugation, the pellet was carefully re-suspended in 18.6 ml ice-cold transformation buffer and then 1.4 ml DMSO was slowly added with gentle stirring to obtain a final concentration of DMSO at 7% which is critical for transformation efficiency and long term storage. Cells were incubated for another 10 minutes on ice and dispensed 100 ul each in 1.5 ml centrifuging tube. The tubes were frozen immediately in liquid nitrogen, and stored at -80 ^{0}C (the cells are viable for at least 4 months).

2.3.6.2 *E. coli* BL 21 (DE 3) and *E. coli* BL 21 (DE 3)- RIL competent cells

E. coli BL 21 (DE 3) or BL 21 (DE 3)-RIL single colonies (2-3 mm) from overnight growth on 2YT-agar-plate could be taken as inocula into 5 ml 2YT medium without antibiotics and incubated with shaking in a small scale at 37 ^{0}C for 3-4 hours. Then the bacteria could be amplified into 100 ml 2YT medium. The cells grew at 37 ^{0}C with

shaking at 100-110 rpm until a cell density of OD_{600nm} 0.6 was reached. Then the cells were centrifuged at 4,000×g for 10 minutes at 4 ^{0}C and the supernatant was then discarded. The cells were then re-suspended in 10 ml pre-cooled 0.1 M $CaCl_2$ slightly, followed by incubation for another 10 minutes on ice. After washed using 0.1 M $CaCl_2$ again, cells were incubated on ice for at least 30 minutes. Finally, the pellet was re-suspended carefully using 2 ml ice-cold 0.1 M $CaCl_2$ solution and 200 µl 50% glycerol with gentle stirring. Cells were then dispensed in 1.5 ml centrifuging tubes at 100 µl/tube and were frozen immediately in liquid nitrogen, and finally stored at -80 ^{0}C.

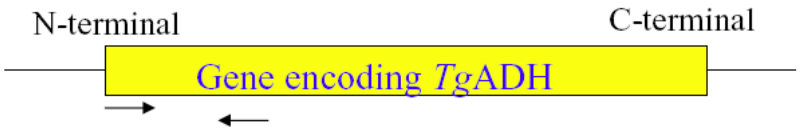
2.3.7 Gene cloning for *Tg*ADH

Since the genome sequence of *T. guaymasensis* is unknown, the *Tg*ADH encoding gene should be isolated and sequenced before over-expression of enzyme in mesophilic host. The PCR amplified fragments were obtained using genomic DNA as template, and the strategy of gene cloning was shown in **Figure 2-1**.

2.3.7.1 Preparation of *T. guaymasensis* genomic DNA

The total genomic DNA from *T. guaymasensis* was isolated by lysis of the cells. The cells harvested and resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), which was followed by the addition 0.5 ml of 10% SDS and 5 mg of protease K and incubation at 60 ^{0}C for 30 minutes. To this, 0.5 ml of 3 M sodium acetate was added and stored on ice for 1 hour. This was centrifuged at 10,000×g for 10 min. The supernatant was transferred to a fresh tube and DNA was isolated from the mixture using the solution

Step 1 PCR for nucleic acids coding the N-terminal of *Tg*ADH



Step 2 Inverse PCR for the entire coding gene of *Tg*ADH

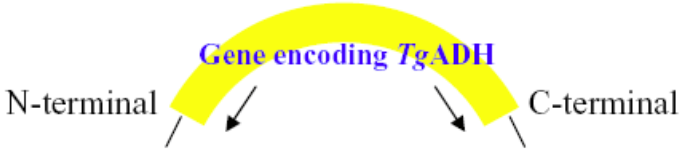


Figure 2-1 Strategy for *T. guaymasensis* ADH coding gene sequencing

Black arrows, location and directions of major oligonucleotide primers for PCR amplification.

of phenol, chloroform and isoamyl alcohol (25: 24: 1). Participated by 100% isopropanol and washed by 75% ethnol, the DNA was finally resuspended in 0.2 ml TE buffer. DNA concentrations of samples were quantified using NanoDrop Spectrophotometer (NanoDrop Technologies, DE, USA), and purified DNA was stored at –20 ^{0}C.

2.3.7.2 Cloning of the entire gene encoding *Tg*ADH

The coding gene of *Tg*ADH was cloned from the *T. guaymasensis* genomic DNA isolated from the *T. guaymasensis* cells directly using Polymerase Chain Reaction (PCR) performed by a thermal cycler termed TC-312 (Techne incorporated, NJ, USA), each reaction had a volume of 25 μl and all the reagents were added following the standard conditions recommended by the suppliers (**Table 2-4**). An error-free amplification was expected to protect occurrence of accidental mutations in the process of PCR, which could finally lead to an inactive recombinant protein. The Taq-polymerase possesses no "proof-reading activity" (Flaman et al., 1994) and therefore the appearance of mutations during the process of amplification cannot be avoided. So the amplification of the PCR-products for cloning in the expression vectors was carried out by the proof-reading polymerase. Unless otherwise stated, PCR was performed using KOD Hot Start DNA Polymerase, and all the reaction parameters were set as conditions suggested by the suppliers (**Figure 2-2**). Two bioinformatic tools were used for primer designing; software DNASTAR was used to predict the potential locations of the primers and GENERUNNER (version 3.01, downloaded from http://www.generunner.net) was used to optimize the parameters of oligonucleotides (Burland 1999). After typing in the nucleic acids sequence in DNASTAR, major properties of the primers were set as the

Table 2-4 PCR mixture

Reagents	Final concentration	Volume (μl)
Polymerase buffer (10×)	1×	2.5
Template	50~100 ng/μl	2
Primer1	1 μM	1
Primer2	1 μM	1
DNTP	0.2 mM	2.5
$MgSO_4$ / $MgCl_2$	1.5 mM	1.5
Polymerase	0.5 unit	0.5
ddH_2O	-	14
Total	-	25

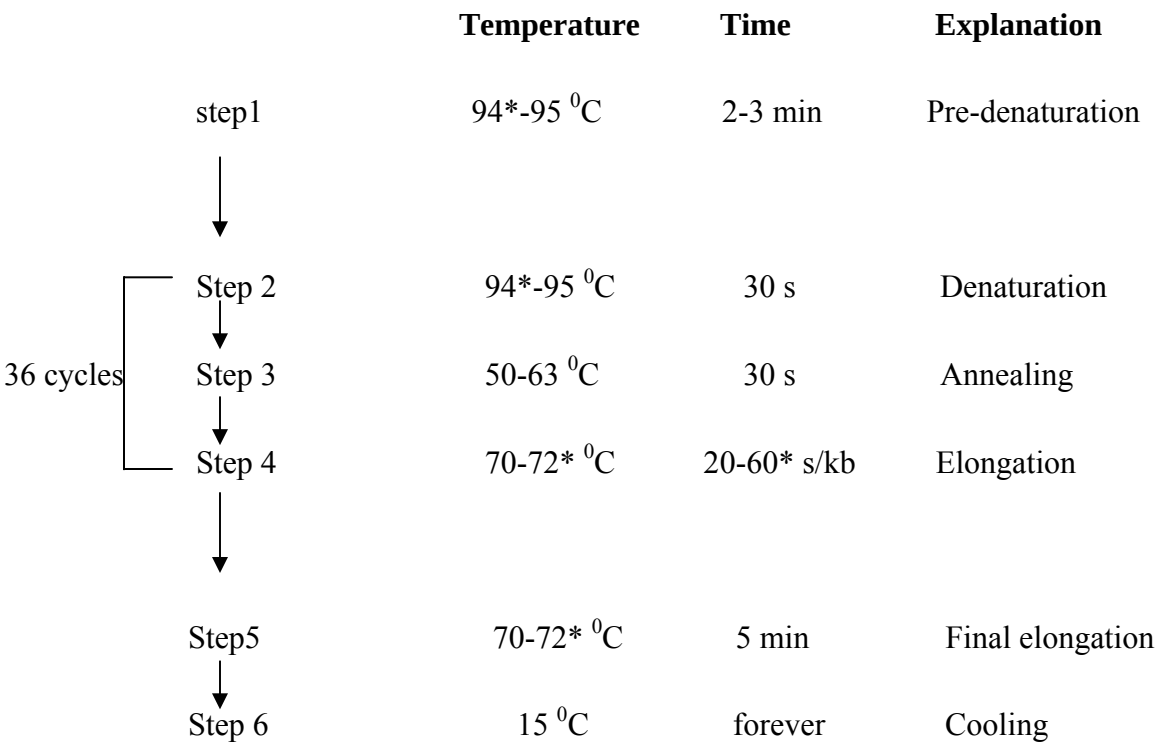


Figure 2-2 PCR parameters setting

*, parameters for Taq polymerase. Parameters without definition were for KOD polymerase.

following: primer length between 18- 30 bp, melting temperatures (Tm) between 45-65 ^{0}C. The software then calculated the locations of proper forward and reverse primers that were consequently optimized by GENERUNNER. The optimal primer pairs had similar melting temperatures (difference ≤ 3 ^{0}C) and harbored no hairpin loops, dimmers, bulge loops or internal loops.

Analytical as well as preparative gel electrophoresis of double-stranded DNA fragments were performed in 0.5-1.5% agarose gels (Sambrook et al., 1989) supplemented with ethidium bromide (final concentration 0.5 μg/ml). The agarose was dissolved in 1×TAE buffer (Diluted from the 50×stock solution: Tris base 1 M, Glacial acetic acid 57.1 ml/L, EDTA 50 mM, pH 8.0). Before loading on the gel, the DNA samples were mixed with 6× DNA loading buffer. For determination of fragment size and concentration estimation, a defined amount of DNA size marker was included. DNA bands were visualized and graphed by FluorChem 8000 Chemiluminescence and Visible Imaging System (Alpha Innotech Corporation, San Leandro, CA, USA). For preparative methods such as cloning of DNA fragments, all the PCR products were purified with the PCR Purification kit (Qiagen, ON, Canada).

The nucleic acids sequence encoding the N-terminal of the enzyme was obtained by PCR using both normal and degenerate primers (**Table 2-5**). The degenerated forward primer TGADHNF was designed based on its N-terminal amino acid sequence considering the codon bias of *Thermococcus kodakaraensis* KOD1 while the reverse non-degenerated

Table 2-5 Primers designed for cloning and sequencing the gene encoding *T. guaymasensis* ADH

Name of primers*	Nucleotide sequence (5'-3')	Restriction enzyme Sites (underlined)
TGADHNF	AARATGMGNGGTTTTGCAATG	
TGADHIR	GGAGTGCTGGTGATATCC	
TGMAYN01	TCTCCTTCTCAATCCACTCG	
TGMAYC02	GCAATAACTCCCGACTGG	
TGMAY28C01	TGCCGAAGTAGTTGATGTTG	
TGMAY28C02	GAGGTCAAGCAGGCGNTC	
TGJL1N1	ATGTCNAAGGATGCGCGGT	
TGJL1N2	ATGAGYAAGGATGCGCGGT	
TGECN	TAGAATTCATGAGCAAGATGCGCGGTTTTC	*Eco*RI
TGXHR	ACCTCGAGTCACTCCTCTATGATGACC	*Xho*I

*, Primer properties such as melting temperature (Tm), GC content (GC%), primer loops and primer dimmers were evaluated by a DNA analysis tool Gene Runner (Hastings Research, Inc., Las Vegas, USA). The table indicates all the key primers used for both fragments cloning and specific amplification. The forward and the reverse primers with the restriction enzyme sites were the specific primer designed based on the confirmed sequence for the amplification of the entire *Tg*ADH encoding gene.

primer TGADHIR was designed based on one of the internal sequences (GYHQHSGGMLAGW) by mass spectrometry and the conserved nucleotide sequence in the gene encoding the ADH from *Thermoanaerobacter brockii*. Both sequence analysis using bioinformatics tool BLAST, as well as the matching of the amino acid sequence finally confirmed the nucleic acids sequence. For the further cloning of the downstream sequence till the coding sequence of C-terminal of *T. guaymasensis* ADH, a process termed Inverse PCR was applied. Firstly, the isolated genomic DNA was digested by the DNA restriction enzyme that was not included in the known DNA sequence, including *Eco*RI, *Hind* III and *Bam* HI, respectively. After incubation at 37 ^{0}C for 1-2 hours, the partially digested samples were incubated at 65 ^{0}C for half an hour to denature the enzymes. Then, the digested product was ligated to circle DNA by using T4 DNA ligase at 16 ^{0}C overnight, which was used as the template in inverse PCR (Triglia, 2000). After amplification, the resulting product of inverse PCRs was sequenced by the dye-termination method using several primers designed on raw sequence information (Molecular Biology Core Facility, University of Waterloo, ON, Canada). The nucleotide sequences were analyzed with the program GENERUNNER and its deduced amino acid sequence was compared to the GenBank Data Base by BLASTP (Altschul et al., 1997). Finally a 1.4 kb fragment amplified from the DNA template digested by *Hind* III was confirmed to be the one carrying the target *Tg*ADH encoding gene.

2.3.8 Data mining

The homologues of the encoding gene sequence of *Tg*ADH and the deduced amino acid sequence were identified using the BLAST program

(http://www.ncbi.nlm.nih.gov/BLAST) with the default parameters (Altschul et al., 1997). Additional sequences were retrieved from the Pfam database (Bateman et al., 2002). Sequence alignments and phylogenetic trees were constructed by the neighbor-joining method of Clustal W with default parameters (Thompson et al., 1994). Theoretical molecular weight was calculated using the ProtParam program at the ExPASy Proteomics Server with standard parameters (Gasteiger et al., 2005). A 3-D structure of *Tg*ADH monomer was modeled using the Swiss Model server (Guex and Peitsch 1997; Peitsch 1995; Schwede et al. 2003), and then the PDB file obtained was used in the PyMOL software to visualize and analyze the 3-D structure.

2.3.9 Construction of the recombinant plasmid

2.3.9.1 Vectors

pGEM-Teasy, 3015 bp, P_{T7}, Amp^R (Promega,WI, USA)

pET-30a, 5360 bp, P_{T7}, Kan^R (Novagen, WI, USA)

2.3.9.2 Plasmid isolation by the alkaline lysis method

Alkaline lysis (Birnboim et al., 1979) was done for plasmid isolation. The method involves 3 steps: washing with RNase solution, lysis of the cells with lysis buffer and precipitation of the plasmid DNA. A single colony was inoculated in 5 ml of 2YT medium with corresponding antibiotics and grew overnight at 37 ^{0}C. The culture was harvested by centrifugation for 10 min at 5,000×g and the pellets were suspended by vortexing in 400 µl of ice-cold Solution A (Glucose 50 mM, EDTA 10 mM, Tris-HCl 25 mM, pH 8.0, RNase A 100 µg/ ml). To the suspension, 800 µl of Solution B (NaOH 0.2

M, 1% SDS) was added and mixed by inverting the tube several times slightly. After incubation on ice for 5 min, 600 µl of ice-cold Solution C (Sodium acetate, 3 M) was added and mixed by inverting the tube. After the centrifugation at 4 ^{0}C, the supernatant containing the plasmid DNA was taken and DNA participated by 100% ethanol and collected by centrifugation. The isolated plasmid DNA was then suspended in 50-100 µl TE buffer.

2.3.9.3 DNA restriction digestion

Digestion of the DNA with restriction endonucleases was performed in the buffer supplied with the restriction enzyme and in accordance with the suppliers' recommendations for temperatures and duration of digestion. Mostly digestion was done for 2-4 hours using 10-20 U of the enzyme for 0.5-1.5 µg DNA. The digestion reaction was incubated at 37 ^{0}C. After completion of the restriction digestion the reaction mixture was analyzed by agarose gel electrophoresis. For preparative restriction digestion e.g., DNA fragments to be inserted into the vector, reaction mixture was purified with the PCR Purification kit (Qiagen, ON, Canada) and quantified by agarose gel electrophoresis.

2.3.9.4 Ligation of DNA fragments

To keep a high efficiency of the ligation, all the restriction endonucleases selected in this research could provide the "sticky end", either 5' or 3' extension. For successive ligation reactions of the inserts to vectors, a 10 µl reaction volume was used with 3:1 molar ratio of insert to vector, 1 U of T4 DNA Ligase and 1 µl 10× Ligation buffer (Fermentas, ON, Canada). The ligation mixture was incubated at 16-20 ^{0}C overnight. After ligation reaction was completed, the mixture was used for transformation.

2.3.9.5 Transformation and selection

As *T. guaymasensis* ADH coding gene preferred a very different codon usage compared to the mesophilic host *E. coli*, higher expressed gene tend to have a greater degree of codon bias. To overcome this big problem, confirmed by sequencing, the isolated recombinant plasmids carrying *Tg*ADH coding gene were transformed into both *E. coli* BL21 (DE3) and the codon-plus *E. coli* BL 21-RIL expression strains. Transformation of the plasmid to *E. coli* host cells followed the standard heat shock method. When *E. coli* competent cells with the constructed plasmids were subjected to 42 ^{0}C heat, a set of heat shock genes would be expressed which aid the bacteria in surviving at such temperatures that was necessary for the uptake of foreign DNA. In heat shock method, 10 µl of the ligation mixture was added to 100 µl competent cells, after incubation on ice for 30 minutes, a heat shock was given at 42 ^{0}C for 90 seconds followed by a second incubation on ice for 5 minutes. To this, 300 µl of blank 2YT medium was added and regeneration was done at 37 ^{0}C for 0.5 to 1 hour. The transformation mixture was plated onto 2YT agar plates containing appropriate antibiotic for selection and incubated at 37 ^{0}C overnight (Hanahan, 1983).

Normally the vector molecule carries a gene whose product confers a selectable or identifiable characteristic on the host cell, or either oppositely, one gene is disrupted when new DNA is inserted into a vector, and the host cell does not display the relevant characteristic. pGEM-Teasy and pET-30a plasmids contained gene giving resistance to ampicillin and kanamycine, respectively, which the intended recipient *E. coli* strain is sensitive to. For the construction of expression vector pET-30a-*Tg*adh, a final

concentration of 50 mg/ml kanamycine was added to the medium as the resistance selection, while for the pGEM-Teasy cloning vector, a blue-white selection together with the ampicillin resistance was used. The blue-white selection is a method of differentiating transformants that carry the vector-insert construct to those that do not carry any insert by using X-Gal (5-bromo-4-chloro-3-indolyl-[beta]-D-galactopyranoside) and IPTG (Isopropyl-[beta]-D-thiogalactopyranoside) as selection markers (Ullmann et al., 1967). After transformation, host cells were plated on 2YT agar containing X-Gal (80 μg/ml 2YT agar), IPTG (final concentration 0.05 mM) along with ampicillin (100 μg/ml 2YT agar). The plates were incubated at room temperature until the transformation mixture had absorbed into the agar. After that, the plates were inverted and incubated at 37 ^{0}C overnight followed by a second incubation at 4 ^{0}C for 5-6 hours. This cold incubation enhances blue color development and thereby facilitates differentiation between blue colonies and white colonies.

2.3.10 Optimization of growth condition for the recombinant *E. coli*

For expression of recombinant ADH gene in *E. coli,* plasmid *Tgadh*-pET-30a was transformed into *E. coli* BL21 (DE3) and transformants were grown in 2YT medium at 37 ^{0}C before induction. The recombinant *Tg*ADH was expressed driven by the T7-lac promoter and the recombinant enzyme was obtained in the periplasmic space when IPTG was added. Both the concentration of the inducer as well as the growth phase of recombinant cells at which it was added affected the final yield of the protein. To optimize the yield, inducer IPTG was added in the exponential phase when OD_{600} of the cell culture reached 0.4-1.0, wheras an ideal yield with high activity presented when the

cell density reached OD_{600} 0.8. The optimum concentration of the inducer was detected by amount of recombinant protein on the SDS-PAGE. The culture was then induced with 0.2 mM IPTG and cultivated at 30 ^{0}C for 12-16 hours prior to harvesting the cells.

2.3.11 Determination of protein concentration

Bradford method was used to measure the protein concentration of solutions using a spectrophotometer or microplate reader (Bradford, 1976). This is based on the coomassie blue dye-binding assay in which a differential color change of the dye occurs in response to various concentrations of protein. The maximum absorbance of the dye shifts from 465 nm to 595 nm when binding of protein occurs and the measurements were at 595 nm. 200 µl of Bio-Rad reagent was mixed with 800 µl of protein solution, and a control was set by mixing 200 µl of Bio-Rad reagent with 800 µl pure water. Since the absorption is proportional to the protein quantity, the concentration of the protein solution can be determined over a linear calibration curve. The calibration curve was obtained with known protein concentrations of standard protein bovine serum albumin (BSA, albumin fraction V) by reading the absorbance of the diluted BSA at 595 nm. The absorbance versus protein concentration curve was linear in the restricted protein concentration range (between 1 mg and 20 mg protein/ml sample solution).

2.3.12 Determination of enzyme activity

The activity of ADH was determined spectrophotometrically by measuring the rate of consumption of the cofactor NADPH. The *in vitro* enzyme assays were anaerobically at 80 ^{0}C by measuring the ethanol-dependent reduction of NADP or the acetaldehyde-dependent oxidation of NADPH at 340 nm. Unless specifically stated, the enzyme assay

was carried out in duplicate using the standard reaction mixture (2 ml) for ethanol oxidation, which contained 100 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer (pH 10.5), 90 mM butanol, and 0.4 mM NADP. For determination of reducing activity of the enzyme, the reaction mixture for acetaldehyde reduction was composed of 100 mM 4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid (HEPES) buffer (pH 7.0), 0.4 mM NADPH and 90 mM acetaldehyde. The reaction was initiated by adding enzymes. One unit (U) is defined as the production of 1 μmol of NADPH per minute.

2.3.13 Protein purification

2.3.13.1 Preparation for the cell-free extract

All procedures for the preparation of cell-free extracts were carried out anaerobically. Frozen *E. coli* cells (5 grams, wet weight) carrying the recombinant expression vector *Tg*ADH-pET-30a were resuspended in 25 ml buffer A [50 mM Tris buffer containing 5% (v/v) glycerol, 2 mM dithiothreitol (DTT), 2 mM sodium dithionite (SDT) and 0.01 mg/ml DNase I, pH 7.8]. The cell suspension was incubated with stirring for 2 hours at 37 ^{0}C. After centrifugation at 10,000 × g for 30 minutes, the supernatant was collected as cell-free extract for further use.

2.3.13.2 Purification of the recombinant *Tg*ADH

Purification of the recombinant enzyme from *E. coli* was carried out anaerobically using the FPLC system. Since the enzyme was thermostable, a step of heat precipitation was applied prior to the column. To optimize the heat treatment time, the cell free extracts were incubated at 60 ^{0}C for 0.5 hour, 1 hour and 2 hours respectively. After incubated at

60 ^{0}C for 0.5 to 1 hour, the solution turned gel-like. The denatured proteins and cell debris in the cell-crude extract were removed by centrifugation at 10,000 × *g* for 30 min at room temperature. The supernatant containing enzyme activity were collected and pooled to a Phenyl-Sepharose column (2.6 × 10 cm) equilibrated with 0.8 M ammonia sulfate in buffer A. A linear gradient (0.82-0 M ammonia sulfate in buffer A) was applied at a flow rate of 2 ml/min and the ADH started to elute at a concentration of 0.4 M ammonia sulfate.

2.3.13.3 Size exclusion chromatography

The recombinant *Tg*ADH was purified after Phenyl-Sepharose column, while a part of sample was loaded onto the Superdex 200 gel filtration column (2.6 × 60 cm; Amersham Biosciences) in order to determine the molecular mass of its native form. Size exclusion chromatography on a Superdex 200 (Amersham Biosciences, USA) equilibrated in 50 mM Tris-HCl (pH 7.8) containing 100 mM KCl. Size of the native form of enzymes was calculated based on the elution volume of standard proteins (Pharmacia, NJ, USA) that contained blue dextran (molecular mass, Da, 2,000,000), thyroglobulin (669,000), ferritin (440,000), catalase (232,000), aldolase (158,000), bovine serum albumin (67,000), ovalbumin (43,000), chymotrysinogen A (25,000) and ribonuclease A (13,700).

2.3.13.4 Protein gel electrophoresis

The fraction containing the dominated activity was loaded to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli (Laemmli, 1970) to examine the purity and analyzed protein composition. Protein samples for SDS-PAGE were prepared by heating for 10 min at 100 ^{0}C in the presence of sample buffer

(0.1 M sodium phosphate buffer, 4% SDS, 10% 2-mercaptoethanol, 20% glycerol, pH 6.8). A low range molecular weight protein marker (Bio-Rad Laboratories Inc., ON, Canada; containing the bands 97 kDa, 66 kDa, 45 kDa, 31kDa, 20 kDa, 14 kDa) was used to estimate the molecular mass of the proteins.

2.3.14 Characterization of catalytic properties

2.3.14.1 Optimum pH

All the catalytic properties of native and recombinant *Tg*ADH were determined using the *in vitro* enzyme assay described above. The optimal pH of ethanol-dependent oxidation of native and recombinant *Tg*ADHs was determined by testing and comparing the enzyme activity at a series of pHs. Standard enzyme assay at 80 ^{0}C were applied using a set of 100 mM buffers: HEPES (pH 6.5, 7.0, 7.5 and 8.0), EPPS (pH 8.0, 8.5, 8.8, 9.0), glycine (pH 9.0, 9.5, 10.0) and CAPS (pH 10.0, 10.5, 11.0). The optimal pH of acetaldehyde-dependent reduction of native and recombinant *Tg*ADH was measured between pH 5.5 and 9.5 using the following buffers (100 mM): citrate (pH 5.5 and 6.0), PIPES (pH 6.0, 6.5 and 7.0), HEPES (pH 7.0, 7.5 and 8.0), EPPS (pH 8.0, 8.5, and 9.0), and glycine (pH 9.0 and 9.5).

2.3.14.2 Temperature dependence and thermostability

The effect of the temperature on the enzyme activity was examined at temperatures from 30 to 95 ^{0}C. The activities were measured using standard assay conditions. Enzyme thermostability was evaluated incubating the enzyme in sealed serum bottle at 80 ^{0}C and 95 ^{0}C respectively, and measuring the residual activities at different time intervals under the standard assay conditions.

2.3.14.3 Oxygen sensitivity

The effect of oxygen on enzyme activity was investigated by exposing the enzyme samples in the air at room temperature and determining the residual activity after oxygen exposure. The exposure was performed in the presence or absence of DDT and SDT. The residual activities of each sample at different time intervals were tested parallelly under the standard assay conditions.

2.3.14.4 Effect of metal ions

Considering the low solubility of cations at alkaline environment, the effect of cations on enzyme activities was carried out only by measuring the reduction of 2-butanone optimal at a moderate pH and alcohol oxidation activity optimal at a pH higher than 10 was not measured. Metal ions were added in the enzyme assay mixture at a final concentration of 100 μM.

2. 4 RESULTS

2.4.1 Cloning of *T. guaymasensis* ADH encoding gene

2.4.1.1 N-terminal sequencing of *Tg*ADH encoding gene

Based on the method as described in section 2.2.3, the genomic DNA (gDNA) from *T. guaymasensis* strain was isolated with a high purity (**Fig. 2-3**). The isolated gDNA had a concentration of 400 ng/μl. The isolated genomic DNA was used as the template for the amplification of ADH encoding gene directly by PCR. In previous research, native *Tg*ADH was purified and the N-terminal sequence of mature enzyme was detected to be SKMRGFAMVDF, which started from serine, indicating the presence of N-terminal methionine excision after translation (**Table 2-6**). The PCR directed by primer pair TGADHNF and TGADHIR produced a single band on 1% agarose gel with the size of approximately 300 bp (**Fig. 2-3**). The nucleotide sequence were confirmed by DNA sequencing and the deduced amino acid sequence of the 321 bp PCR product was applied in the BLAST tool and they aligned the N-terminal sequence of ADH from thermopiles.

2.4.1.2 The entire coding gene of *Tg*ADH obtained by inverse PCR

The SDS-PAGE indicated that native enzyme purified from *T. guaymasensis* has the molecular weight of 40 kDa (Ying et al., unpublished results), so the complete nucleic acid sequence of *Tg*ADH gene should involved in a complete open reading frame with an approximate length of 1.1 kb, encoding a polypeptide of about 360 amino acid residues.

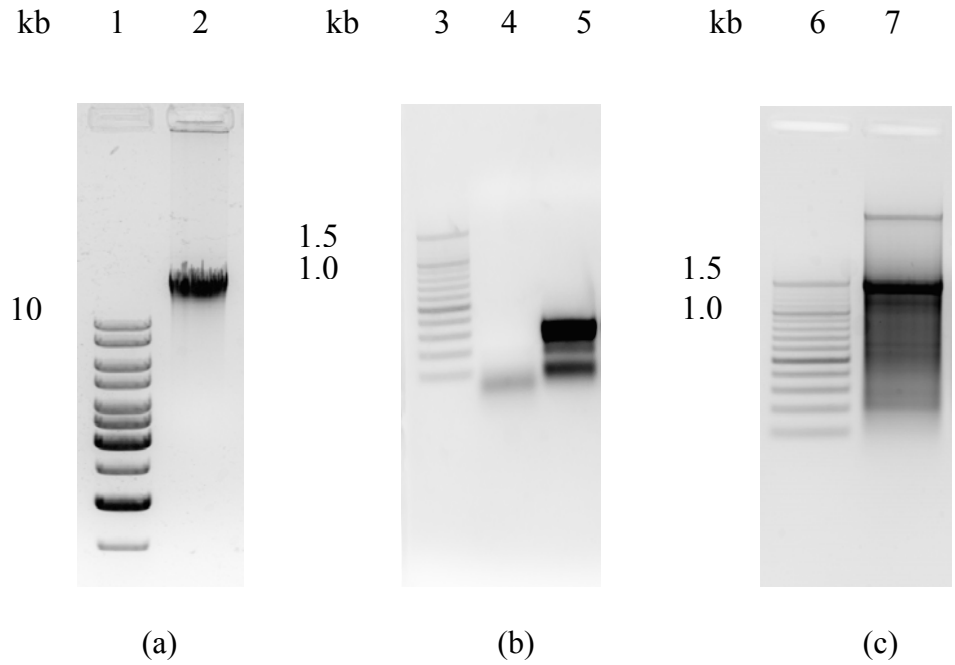


Figure 2-3 PCR amplification of the gene encoding *T. guaymasensis* ADH

(a) 0.8% agarose of isolated *T. guaymasensis* genomic DNA; (b) 1% agarose of the PCR product using primers TGADHNF and TGADHIR; (c) 1% agarose of inverse PCR products using primers TGMAYN01 and TGMAYC02. Lane 1, 1 kb DNA ladders (Novogan, WI, USA); lane 3 and 6; 100 bp DNA ladders (Fermentas Canada Inc., ON, Canada) ; lane 2, *T.guaymasensis* genomic DNA; lane 4, negative control (PCR without template); lane 5, PCR products of 300 bp; lane 7, fragments amplified from inverse PCR containing *Tg*ADH coding gene.

Table 2-6 N-terminal and internal sequences of *Tg*ADH

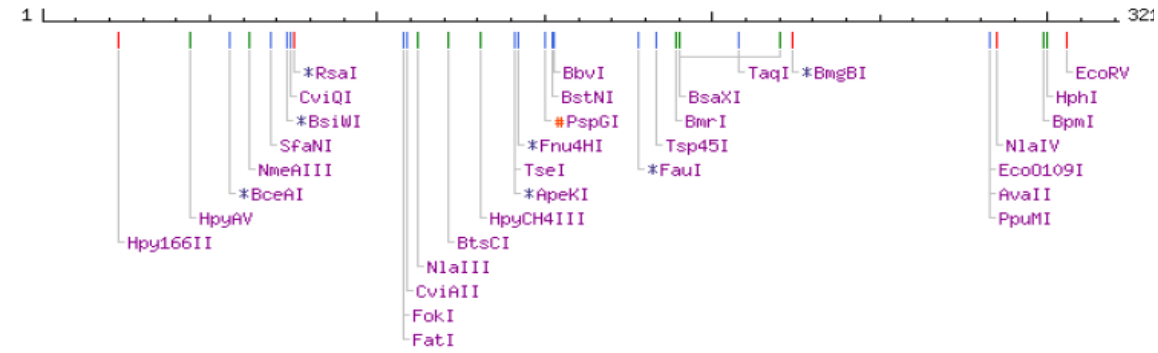
Location	Sequences
N-terminal[a]	SKMRGFAMVDF
Internal 1[b]	DFKPGDR
Internal 2[b]	VVVPAITPDWR
Internal 3[b]	GYHQHSGGMLAGW

[a] amino-terminal sequence was determined by using Edman-degradation
[b] internal sequences were determined by using mass spectrometry

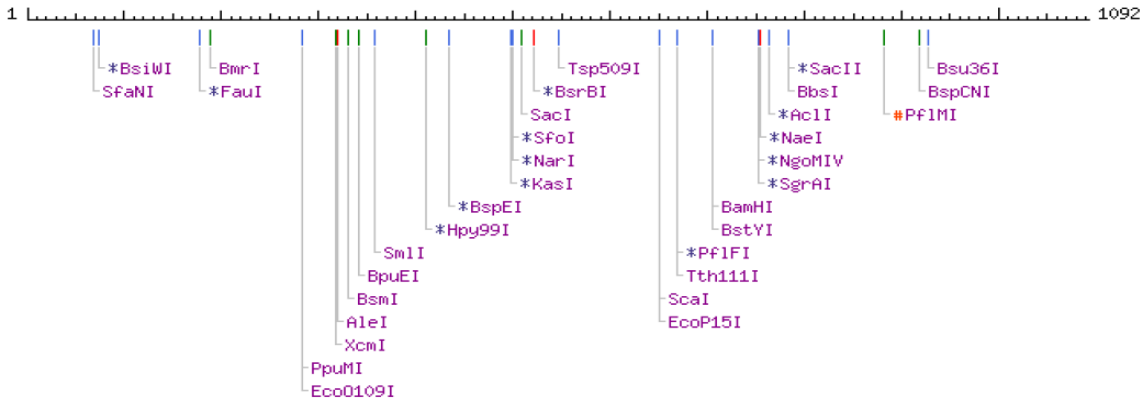
To amplify the upstream and downstream sequences of the known fragment and finally get the sequence of the entire gene, an inverse PCR-based method was used (Benkel et al., 1996; Nie et al., 2007). Inverse PCR uses the polymerase chain reaction, but the template for the reverse primers is a restriction fragment that has been ligated upon itself to form a circle and it has the primers oriented in the reverse direction of the usual orientation. Templates for inverse PCR were obtained from the total genomic DNA of *T. guaymasensis* by partial digestion of the gDNA with *Eco*RI, *Bam*HI and *Hind*III respectively, which were not involved in the known sequence (**Fig. 2-4**). After ligation by T4 DNA ligase, DNA samples were subjected to inverse PCR with the gene-specific primers designed from the known 300 bp sequence encoding N-terminal of *Tg*ADH. For inverse PCR using *Eco*RI or *Bam*HI digested DNA as templates, non-specific bands were found in each lane after agarose gel electrophoresis; while there was one 1.4 kb specific band produced by PCR using the *Hind* III digested genomic DNA as template and driven by the specific primer pair TGMAYN01 and TGMAYC02 (**Fig. 2-3**). The 1.4 kb PCR product was fully sequenced by primer walking, and the alignment revealed that entire coding gene was located in the 1.4 kb fragment.

2.4.2 Sequence analysis

The entire structural gene encoding *T*gADH was detected to be 1092 bp with a deduced 364 amino acids sequence. The molecular weight was calculated to be 39463 Da. Interestingly, the nucleic acid sequence ended at two consecutive stop codons TGA and



(a)



(b)

Figure 2-4 Restriction map of the *T. guaymasensis* ADH coding gene

On-line analysis was conducted using http://tools.neb.com/NEBcutter2/index.php; (a), restriction endonucleases involved in initially sequenced 321 bp fragment; (b), restriction map of the entire *Tg*ADH encoding gene. Red: blunt end cut; blue, 5' extension; green, 3' extension; *, cleavage affected by CpG methylation; #, cleavage affected by other methylation

TAA, and the putative archaeal terminator sequence, TTTTTCT, found 24 bases downstream of the stop codon TGA (Reiter et al., 1988). The downstream sequence of *Tg*ADH encoded a putative gene encoding archaeal hydrogenase (**Fig. 2-5**). Analyzed by the on-line BLAST tool, deduced amino acid sequence of *T. guaymasensis* ADH showed high overall identities to threonine dehydrogenase or zinc-containing ADHs from thermophilic bacteria, e.g., ADHs from *Thermoanaerobacter tengcongensis* MB4 (77% identity, AAM23957), *Thermoanaerobacter brockii* (77% identity, CAA46053), *Thermoanaerobacter pseudethanolicus* ATCC 33223 (77% identity, EAO63648), *Thermoanaerobacter ethanolicus* X514 (76% identity, EAU57308), *Thermosinus carboxydivorans* Nor1 (72% identity, EAX46383), and it also showed high similarity to ADH from mesophile *Clostridium beijerinckii* (67% identity, EAX46383). The deduced *Tg*ADH sequence was classified as zinc-related. The N-terminal region showed homology to ADH_N, the alcohol dehydrogenase GroES-like domain; while the C-terminal region belonged to NADB_Rossmann superfamily indicating the dependence of NAD(P) as coenzyme; the central domain was aligned to Tdh, $_{L}$-threonine dehydrogenase (**Fig. 2-6**). The central region spanning the majority of peptides showed homology to the domain of TDH, so all the three domains were classified as zinc-related. From the conserved motif comparision with its thermophilic and mesophilc counterparts, *Tg*ADH was found to be belonged to the family of zinc-containing ADHs with catalytic zinc only, which was verified by motif searches that the enzyme had binding motifs of catalytic zinc only ($GHEX_2GX_5GX_2V$, residues 62-76) and coenzyme NADP ($GXGX_2G$, residues 183-188)(**Fig. 2-7**).

```
1          ATGAGCAAGATGCGCGGTTTTGCAATGGTGGACTTCGGCAAGGCCGAGTGGATTGAGAAG

1           M  S  K  M  R  G  F  A  M  V  D  F  G  K  A  E  W  I  E  K

61         GAGAGGCCGAAGCCCGGGCCGTACGATGCAATCGTCAAGCCCATTGCAGTCGCCCCATGC
21          E  R  P  K  P  G  P  Y  D  A  I  V  K  P  I  A  V  A  P  C

121        ACCTCGGACATCCACACGGTCTTTGAGGCAGCGTTTCCCAGGGAGATGTGTGAGTTCCCG
41          T  S  D  I  H  T  V  F  E  A  A  F  P  R  E  M  C  E  F  P

181        CGCATACTGGGTCACGAAGCAGTCGGAGAGGTAGTCGAGGTCGGAAGCCACGTCAAGGAC
61          R  I  L  G  H  E  A  V  G  E  V  V  E  V  G  S  H  V  K  D

241        TTCAAGCCCGGGGACAGGGTTGTTGTCCCGGCAATAACTCCCGACTGGAGGACCCTTGAC
81          F  K  P  G  D  R  V  V  V  P  A  I  T  P  D  W  R  T  L  D

301        GTTCAGAGGGGCTACCACCAGCACTCCGGTGGAATGCTCGCCGGATGGAAGTTCAGCAAC
101         V  Q  R  G  Y  H  Q  H  S  G  G  M  L  A  G  W  K  F  S  N

361        CCCCTCAAGGAGGGCGGTAAGGACGGTGTGTTTGCAGAATACTTCCACGTCAACGACGCT
121         P  L  K  E  G  G  K  D  G  V  F  A  E  Y  F  H  V  N  D  A

421        GACATGAACCTGGCACACCTTCCGGACGAAATCAAGCCGGAAGTCGCTGTCATGGCCACC
141         D  M  N  L  A  H  L  P  D  E  I  K  P  E  V  A  V  M  A  T

481        GACATGATGACCACGGGATTCCACGGCGCCGAGCTCGCCGACATTCCGCTCGGAGGAACA
161         D  M  M  T  T  G  F  H  G  A  E  L  A  D  I  P  L  G  G  T

541        GTCGCCGTCATTGGAATTGGACCGGTCGGCCTGATGGCGGTTGCCGGGGCAAGACTGCTC
181         V  A  V  I  G  I  G  P  V  G  L  M  A  V  A  G  A  R  L  L

601        GGTGCCGGAAGGATCATCGCGGTCGGCAGCAGGCCGGTGTGCGTTGAGGCCGCTAAGTAC
201         G  A  G  R  I  I  A  V  G  S  R  P  V  C  V  E  A  A  K  Y

661        TACGGAGCCACCGACATAGTCAACCGCAGGGAGCACCCGGACATCGCCGGAAGGATCCTG
221         Y  G  A  T  D  I  V  N  R  R  E  H  P  D  I  A  G  R  I  L

721        GAGCTGACCGGTGGAGAGGGTGTTGATTCGGTGATAATCGCCGGCGGAAACGTTGACGTA
241         E  L  T  G  G  E  G  V  D  S  V  I  I  A  G  G  N  V  D  V

781        ATGAAGACCGCGGTGAAGATAGTCAAGCCCGGAGGAACGGTGGCCAACATCAACTACTTC
261         M  K  T  A  V  K  I  V  K  P  G  G  T  V  A  N  I  N  Y  F

841        GGCAGCGGTGACTACCTCCCGATCCCGAGGATTGAGTGGGGCCAGGGAATGGCCCACAAG
281         G  S  G  D  Y  L  P  I  P  R  I  E  W  G  Q  G  M  A  H  K

901        ACCATCAAGGGAGGGCTCTGCCCAGGCGGACGCCTGAGGATGGAGCGCCTGCTTGACCTC
301         T  I  K  G  G  L  C  P  G  G  R  L  R  M  E  R  L  L  D  L

961        ATCAAGTACGGCAGGGTTGACCCGTCAAGGCTCATAACCCACAAGTTCAAGGGATTCGAT
321         I  K  Y  G  R  V  D  P  S  R  L  I  T  H  K  F  K  G  F  D

1021       AAGATACCAGAAGCCCTCTACCTGATGAAGGACAAGCCCAAAGACCTGATAAAGCCCGTG
341         K  I  P  E  A  L  Y  L  M  K  D  K  P  K  D  L  I  K  P  V
```

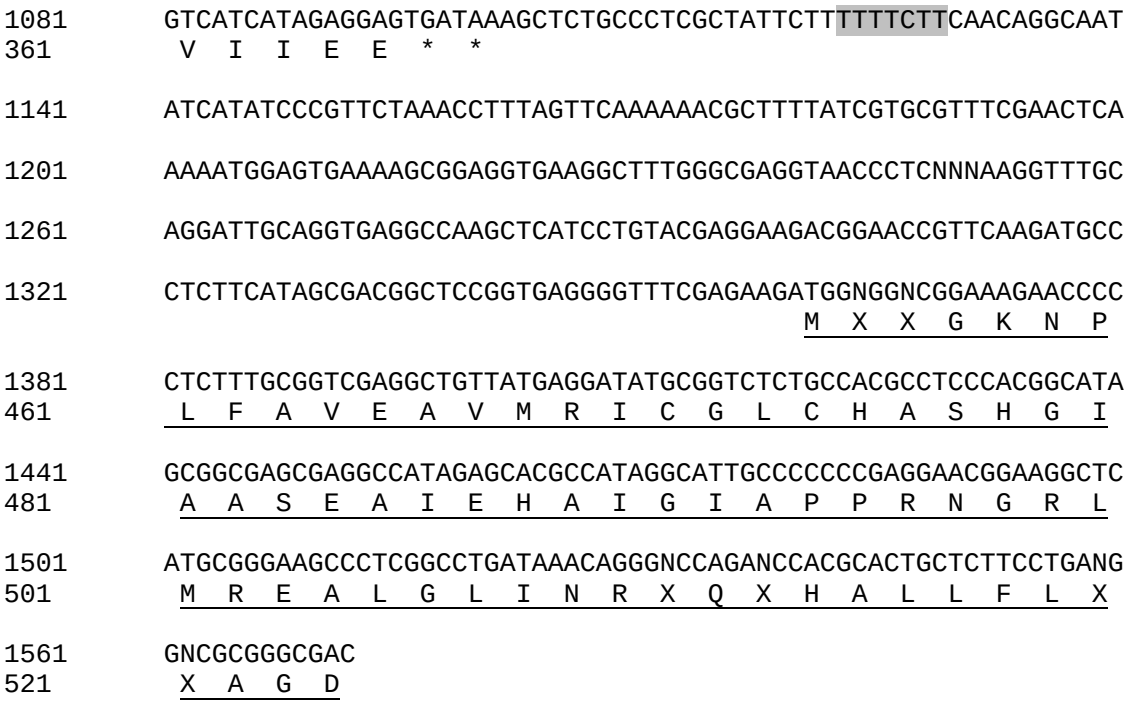


Figure 2-5 Nucleotides and deduced amino acid sequences of *Tg*ADH

The amino acid sequence was deduced using the program DNAMAN (Lynnon Corporation, Vaudreuil-Dorion, Quebec, Canada). The stop codons were marked with asterisk and the possible terminator sequence in the 3'-untranslated region is highlighted. The partial gene encoding a nickel-containing hydrogenase located downstream of the *Tg*ADH was underlined. N in the nucleotide sequence represents either A, T, C or G while X in the deduced amino acid sequence represents unknown amino acid residue.

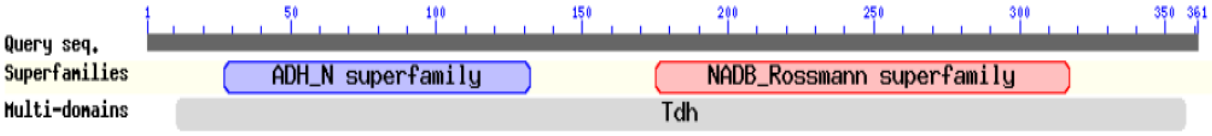


Figure 2-6 Putative conserved domains of *T. guaymasensis* ADH

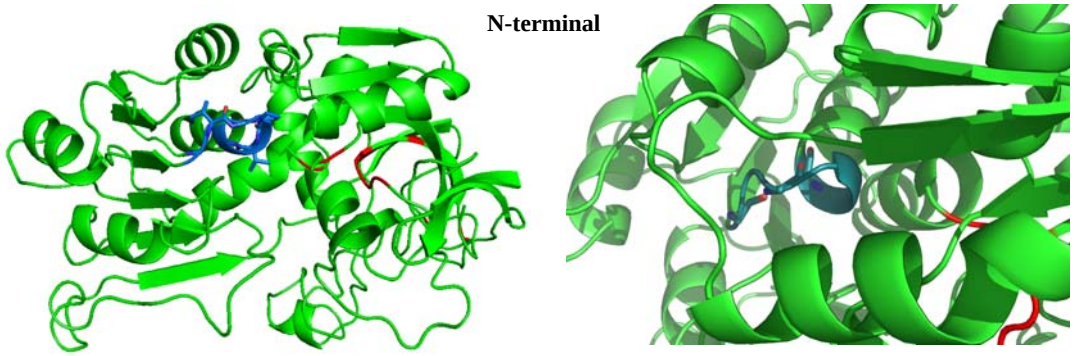
ADH_N, alcohol dehydrogenase GroES-like domain, GroES is an oligomeric molecular chaperone, N-terminal domain of alcohol dehydrogenase-like proteins have a GroES-like fold; NADB_Rossmann superfamily, a large family of proteins that share a Rossmann-fold NAD(P)H/NAD(P) binding domain; Tdh, L-threonine dehydrogenase.

Figure 2-7 Amino acid sequences alignment among *Tg*ADH and its homologues

The sequences were aligned using Clustal W (Thompson et al., 1994). TgADH, *T. guaymasensis* ADH; CpADH, *Carboxydibrachium pacificum* ADH; TbADH, *T. brockii* ADH; CbADH, *C. beijerinckii* ADH. ‘‘*’’, residues or nucleotides that are identical in all sequences in the alignment; ‘‘:’’, conserved substitutions; ‘‘.’’, semi-conserved substitutions; “-”, no corresponding amino acid. Highlighted in yellow, conserved putative binding sites of catalytic zinc; highlighted in black, conserved putative motif of coenzyme binding sites.

To better understand the catalytic mechanism of the enzyme, predicted 3-D structure of *Tg*ADH was shown by PyMOL software (**Fig. 2-8**). The tertiary structural modeling of monomer of *Tg*ADH showed two typical domains. Both domains were separated by the cleft where the active site of the enzyme might be situated, which was responsible for specifically binding a substrate and catalyzing a particular enzymatic reaction. *Tg*ADH and its homologues harbored the highly conserved amino acid residues (**Fig. 2-7**). One putative catalytic domain $G_{62}H_{63}E_{64}X_2G_{67}X_5G_{73}X_2V_{76}$ located close to N-terminal, and one coenzyme NADP-binding domain $G_{183}XG_{185}XXG_{188}$ was close to C-terminal end. In the three dimensional structure, the active site of *Tg*ADH centered around the catalytic zinc ion was close to the pocket containing the cofactor NADP indicating the cofactor binding is essential for the catalysis, as observed for NAD(P)-dependent ADHs (Ishikawa et al., 2007).

The amino acids sequences of *Tg*ADH and its thermophilic and mesophilic homologous were compared. The primary structural analyses revealed that the enzyme and its thermophilic homologue *T. brockii* alcohol dehydrogenase had higher ratio (molar fraction, >0.8% increase or decrease) for Ala, Arg, Glu, Lys and Pro but lower ratio for Asn, Gln, Leu, Ser and Met as compared to the ADH from the mesophile *C. beijerinckii.* (**Table 2-7**). In particular, the amino acid composition of *Tg*ADH had higher ratio for Arg, Pro and Tyr but lower ratio for Ala, Asn and Val than that of the ADH from the *T. brockii*. The 14 most frequently used codons (greater than 2.7%) accounted for 58% of the amino acid residues in *T. guaymasensis*, reflecting its abundant tRNA types (Peretz et al., 1997). Additionally, the codon usage pattern for the *Tg*ADH coding gene was



(a) (b)

Figure 2-8 Predicted tertiary structure of *Tg*ADH monomer (a) and the putative ion/conenzyme-binding site (b)

The structure modeling was run on the Swiss-Model server using *T. brockii* ADH (PDB entry: 1ykf) as template. The structure figure was constructed using the software PyMOL (Delano 2002). (a) Vertical view of predicted tertiary structure of *Tg*ADH monomer; (b) Residues in blue, the putative NADP-binding site; residues in red, the putative zinc-binding site.

Table 2-7 Comparison of the typical amino acids between *Tg*ADH and its thermophilic and mesophilic counterparts

Amino acid	No. of residues in *Tg*ADH	No. of residues in *Tb*ADH	No. of residues in *Cb*ADH	Δ *Tg*ADH & *Cb*ADH	Δ *Tb*ADH & *Cb*ADH
Aromatic					
Phe	13	14	11	+2	+3
Hydrophobic					
Val	35	37	35	0	+2
Leu	23	23	28	-5	-5
Met	14	15	18	-4	-3
Small hydrophobic					
Ala	32	35	27	+5	+8
Pro	25	21	13	+12	+8
Gly	44	43	45	-1	-2
Charged					
Glu	23	21	18	+5	+3
Lys	25	24	21	+4	+3
Hydrophilic					
Ser	9	9	17	-8	-8
Asn	7	10	13	-6	-3
Gln	3	3	6	-3	-3
Total	364	352	351		

*Tb*ADH, ADH from thermophile *Thermoanaerobium brockii*; *Cb*ADH, ADH from mesophile *Clostridium beijerinckii*; amino acids highlighted in grey, amino acids that obviously less in hyper/thermophilic ADH than mesophilic homologous; amino acids highlighted in yellow: amino acids that obviously more in hyper/thermophilic ADH than mesophilic homologous.

analyzed (**Fig. 2-9**). Of the 61 sense codons, thirteen were not used in the *Tgadh* gene, and the comparison between codon usage of *T. guaymasensis* and the mesophilic bacterium *E. coli* indicated an obviously different codon bias between the two species, particularly, AGG (arginine), CUC (leucine), AUA (isoleucine) were rarely used in *E. coli* (**Fig. 2-9**).

2.4.3 Construction of the cloning and expression vectors

The *Tg*ADH gene is difficult to be specifically amplified from the genomic DNA due to the high GC-content, which is a common feature of archaeal genes. Before its cloning to the over-expression vector, the entire encoding gene was inserted to pGEM®-T Easy cloning vector. The PCR amplified *T. guaymasensis* ADH coding gene was inserted to the hanging thymidine (T) with the overhanged restriction sites of *Eco*RI and *Xho*I at the N-terminal and C-terminal respectively (**Fig. 2-10**). The recombinant pGEM vector was selected using the ampicillin resistance and blue-white screening. The *E. coli* DH 5α cells containing transformed recombinant plasmid produced colorless colonies on the agar plate.

After confirmation by sequencing, the insert from T-easy cloning vector was realeased by *Eco*RI digestion. The entire *Tg*ADH encoding gene with overhanged primer including *Eco*RI and *Xho*I restriction sites was then inserted to the *Eco*RI and *Xho*I double digested pET-30a-*Tg*adh expression vector (**Fig. 2-10**). The recombinant plasmid was selected

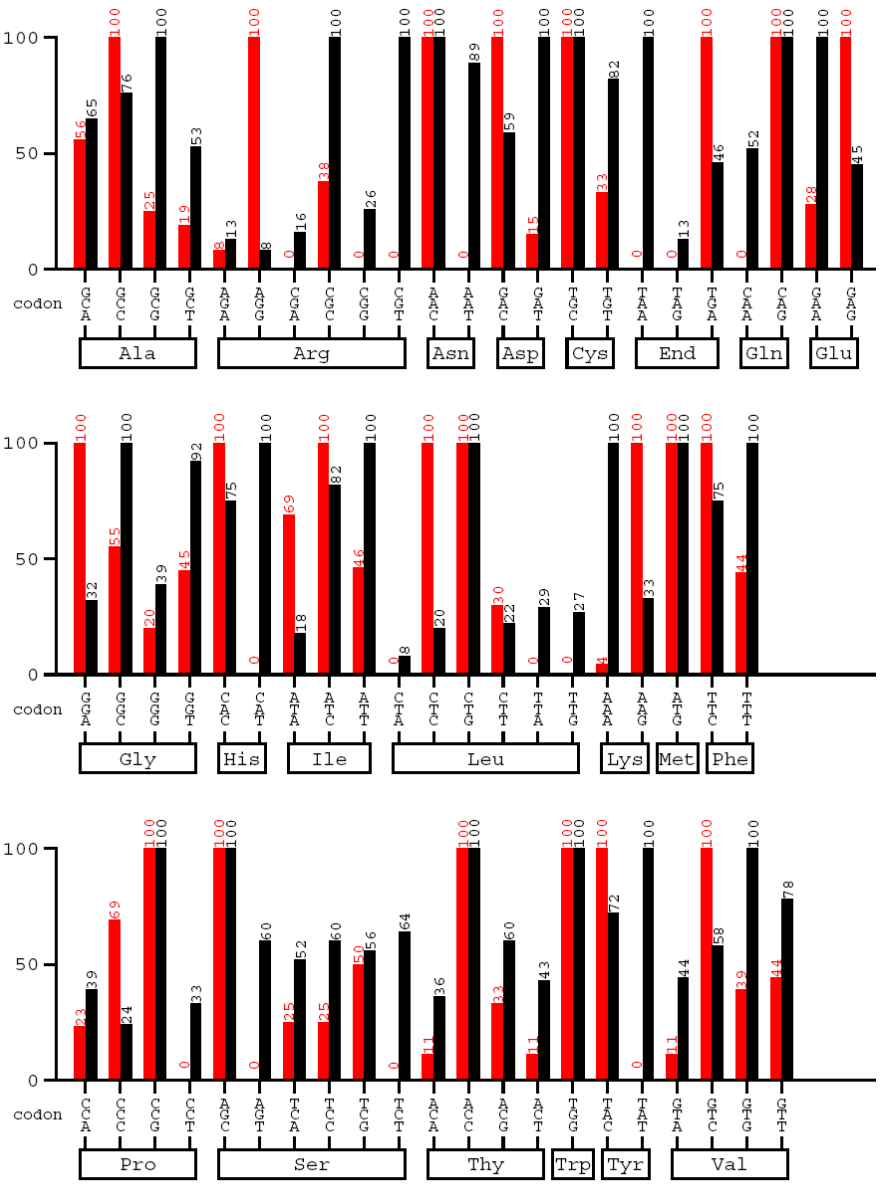
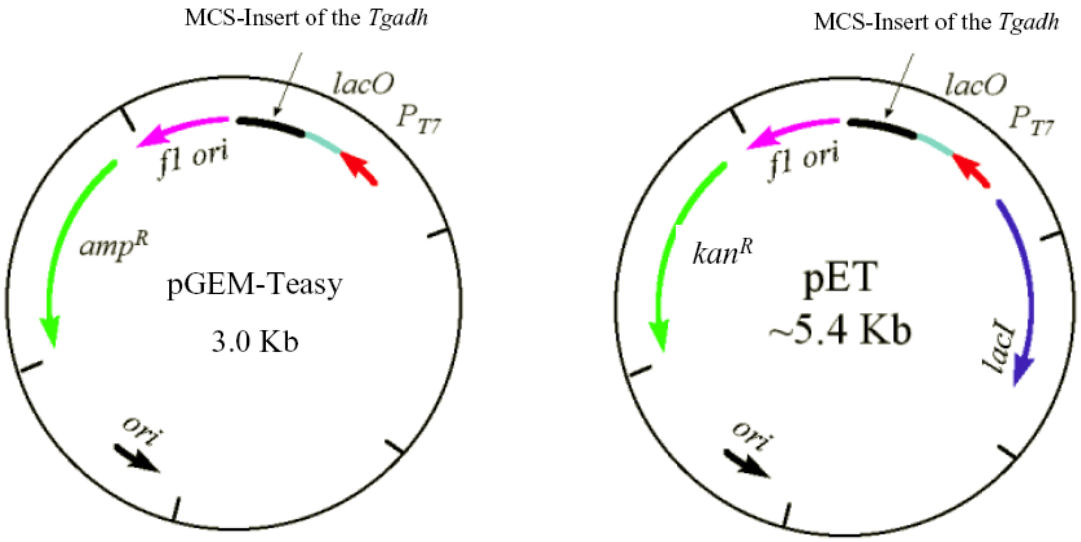


Figure 2-9 Codon usage comparasion between *T. guaymasensis* and *E.coli*

Codon usage analysis was conducted using Graphical codon usage analyzer at www.gcua.de (Fuhrmann et al., 2004). Columns in red, codon usage pattern of *T. guaymasensis*; columns in black, codon usage pattern of *E. coli*. The numbers of most frequenctly used codons for the specific amino acid was defined as 100.



(a) (b)

Figure 2-10 Recombinant plasmid construction and the location of inserted gene

Figures are modified from http://www.bio.davidson.edu. (a) Cloning vector pGEM®-T Easy; (b) Expression vector pET; black arrows, location of the inserted *Tg*ADH encoding gene.

from the colonies grew on 2YT agar with 50 mg/ml kanamycine, and confirmed by both colony PCR and restriction enzymes digestion. Completely digested by *Eco*RI, recombinant plasmids gave a 6.5 kb band on the agarose gel, 1.1 kb larger than 5.4 kb blank pET-30a vector, indicating the insertion of *Tg*ADH encoding gene (**Fig. 2-11**).

2.4.4 Over-expression of the *T. guaymasensis* ADH in *E. coli*

Driven by T7-lac promoter of pET-30a vector, over-expression of *Tg*ADH encoding gene in the mesophilic host *E. coli* was induced by IPTG. From 10% SDS-PAGE, large amount of recombinant enzymes around 40 kDa were produced in the presence of IPTG as inducer (**Fig. 2-12**). The yield of the enzymes was better in the *E. coli* codon plus strain *E. coli* BL 21-RIL expression strains, containing the extra plasmid for rarely used tRNAs including AGG/AGA/AUA to rescue the poor expression by condon bias, however, no expression was observed in blank host strains or recombinant strains without IPTG as inducer.

2.4.5 Optimum cultivation condition

Recombinant *E. coli* cells were incubated in 2YT medium with 50 mg/ml kanamycine to an OD_{600nm} 0.8 before induction, which took 3.5 to 4 hours. From small scale testing, the *E. coli* carrying the recombinant vector *Tgadh*-pET30a provided optimum yield of the recombinant enzyme when the concentration of inducer IPTG reached 0.2 mM (**Fig. 2-12)**. So, the large-scale (1-2 liters) incubation was at 37 ^{0}C to OD_{600nm} 0.8, the final concentration of IPTG for induction was set at 0.2 mM.

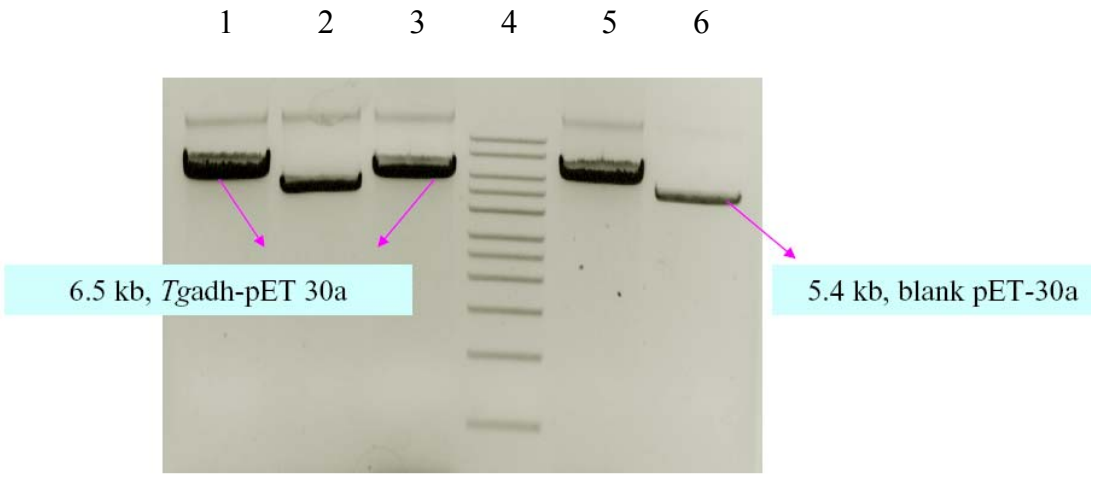


Figure 2-11 Enzyme digestion of the recombinant vector carrying the *Tg*adh by enzyme digestion

The arrows and notes describe behavior of recombinant vector and blank pET-30a vector after enzyme digestion by *Eco* RI. Lane 1, recombinant plasmid *Tgadh*-pET-30a; lane 2, blank pET-30a vector; lane 3, recombinant plasmid *Tgadh*-pET-30a; lane 4, 1 kb DNA ladders (Novogan, WI, USA); lane 5, recombinant plasmid *Tgadh*-pET-30a; lane 6, blank pET-30a vector.

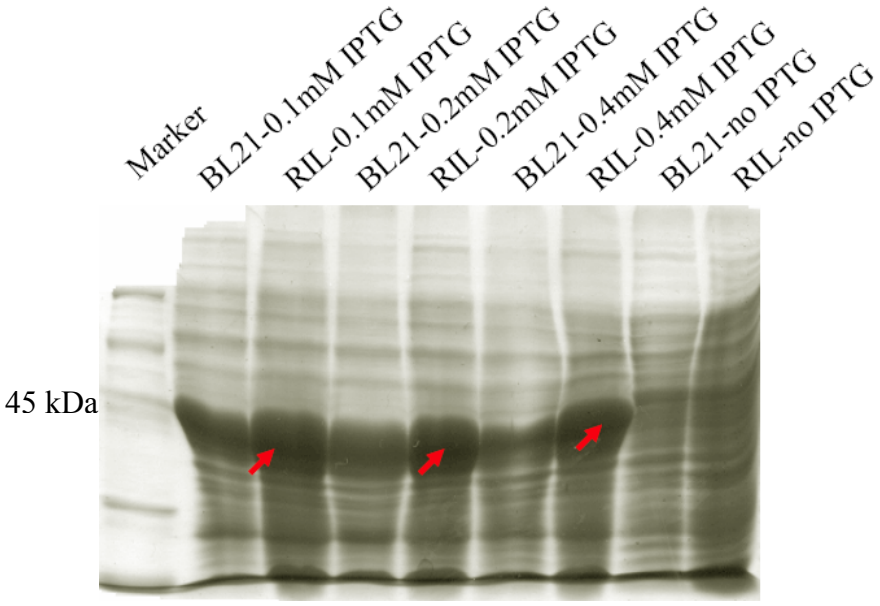


Figure 2-12 Analysis of over-expression level of *Tg*ADH in *E. coli* induced by IPTG using SDS-PAGE (10 %)

Left lane, low molecular weight protein marker (Bio-Rad Laboratories Inc., ON, Canada); other lanes, crude extracts of cells; red arrow, over-expressed *Tg*ADH in the total protein of codon-plus *E. coli* host cells. 30 ul cell crude extracts were loaded per lane.

2.4.6 Purification of the recombinant *T. guaymasensis* ADH from *E. coli*

The recombinant ADH was purified from *E. coli* using a modified procedure. Prior to liquid chromatography, heat treatment was applied to the cell extract. Heating at 60 ^{0}C in half an hour caused no loss of enzyme activity but significantly reduced the protein concentration to 30% (**Fig. 2-13**). The *Tg*ADH activities were dominant in the cell-free extract after heat treatment, subsequently; the recombinant *Tg*ADH was purified to homogeneity after Phenyl-Sepharose column (**Fig. 2-14**). The purified recombinant *Tg*ADH had a specific activity of 1079 U/ mg almost the same as the native protein (Ying et al., unpublished data) but presented a higher yield of 81% (**Table 2-8**). For size exclusion chromatography, ADH activity was assembled in a peak at 170 ml, and molecular mass of the recombinant enzyme was calculated to be 146 ± 6 kDa. The SDS-PAGE analyses showed that both native and recombinant *Tg*ADHs had almost identical subunit size of 40 ± 2 kDa, suggesting that enzyme was homotetramer in the native form (**Fig. 2-14**).

2.4.7 Catalytic properties of the recombinant *Tg*ADH from *E. coli*

The recombiant *Tg*ADH had very similar catalytic properties to the native enzyme purified from *T. guaymansensis* cells, including temperature dependence, optimum pH, thermostability and oxygen sensitivity. Both the native and recombinant *Tg*ADH was thermostable and its activity increased along with the temperature elevated up to 95 ^{0}C (**Fig. 2-15**). The activity values at temperatures higher than 95 ^{0}C were not measured because of the instability of the co-enzyme NADP at those high temperatures. The optimal pHs of the enzyme were tested for the oxidation and formation of 2-butanol using

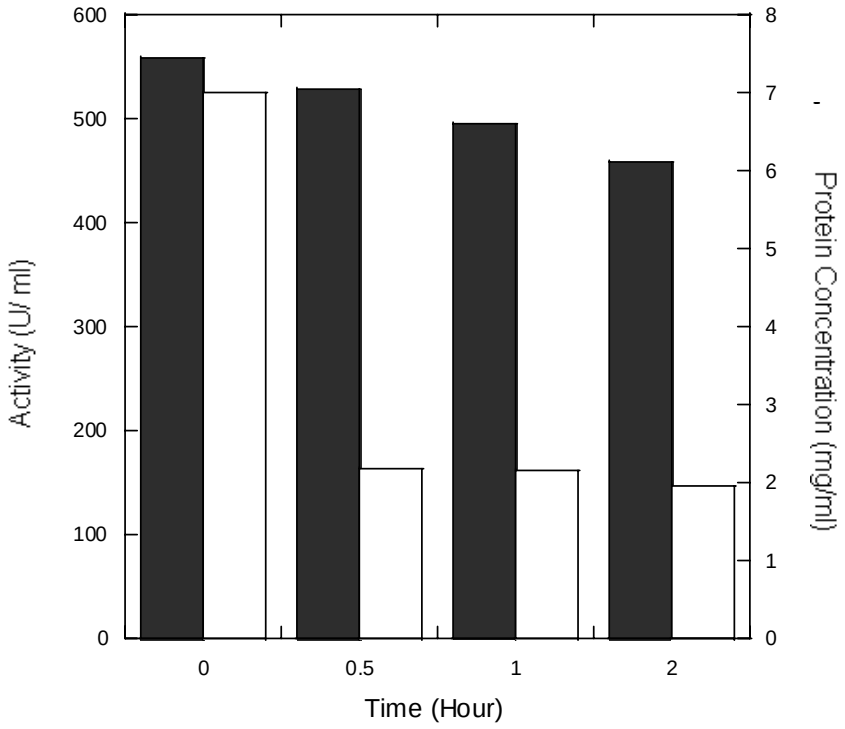


Figure 2-13 Heat treatment of recombinant *E. coli* cell-crude extract

The enzyme activity was assayed in CAPS buffer (100 mM, pH 10.5). Black columns stand for residual activity after heat treatment; white columns stand for residual protein concentration after heat treatment.

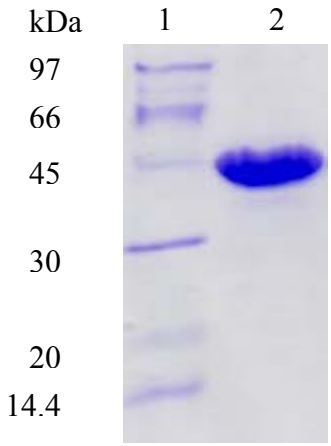


Figure 2-14 10% SDS-PAGE for purified recombinant *Tg*ADH

Lane 1, molecular weight marker; lane 2, 6 μg purified recombinant *Tg*ADH was loaded.

Table 2-8 Purification of recombinant *Tg*ADH

	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Purification fold	Yield (%)
Cell-crude extract	420	28000	67	1	100
Heat-treatment	65	24735	383.4	5.7	88
Phenyl-Sepharose	21	22425	1073.2	16	81

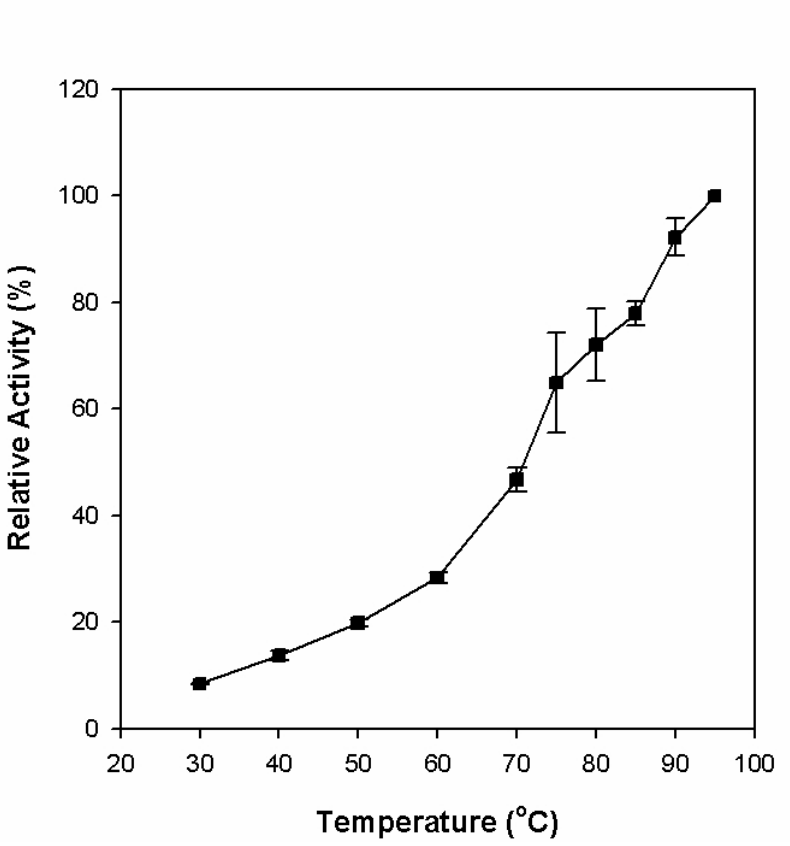


Figure 2-15 Temperature dependence of the purified recombinant *Tg*ADH

The activities were measured in the standard assay conditions except varying assay temperatures from 30 to 95 ^{0}C. The relative activity 100% was defined as the highest activity value achieved in this test (1533 U/mg at 95 ^{0}C). Standard deviations of the measurements are indicated using error bars.

various 100 mM buffers to form a pH gradient from 5.5 to 11.5. The optimal pH value for the 2-butanol oxidation was 10.5 and for the 2-butanone reduction was 7.5 (**Fig. 2-16**). When the buffer pH values were higher than 10.5, the activity of butanol oxidation had a remarkable decrease, similarly, the activity of 2-butanol formation sharply decreased when the buffer pH value higher than 7.5. Recombiant *Tg*ADH presented outstanding stability at high temperatures, which had the same half-life ($t_{1/2}$) of about 26 hours at 95 ^{0}C and the residual activity remained more than 60% of the full activity after 42 hours incubation at 80 ^{0}C (**Fig. 2-17**), revealing the resistance of enzymes to heat. However, both the native and recombinant form of *Tg*ADH presented sensitivity to oxygen. Some of the activity lost after exposure to the air, although they were more resistant to oxidation than that of iron-containing ADHs and the enzyme activity kept consistent in anaerobic conditions. The half-life ($t_{1/2}$) against the oxygen inactivation was about 4 hours, and loss of activity was slightly protected by the presence of 2 mM dithioreitol (**Fig. 2-18**). Metal ions also affected the enzyme activity. The purified enzyme from *T. guaymasensis* was experimentally determined to be zinc-containing; however, the activity of butanol formation was blocked by zinc. When the zinc concentration increased from 20 to 100 μM in the assay mixtures, the corresponding activity obviously decreased (**Fig. 2-19**).

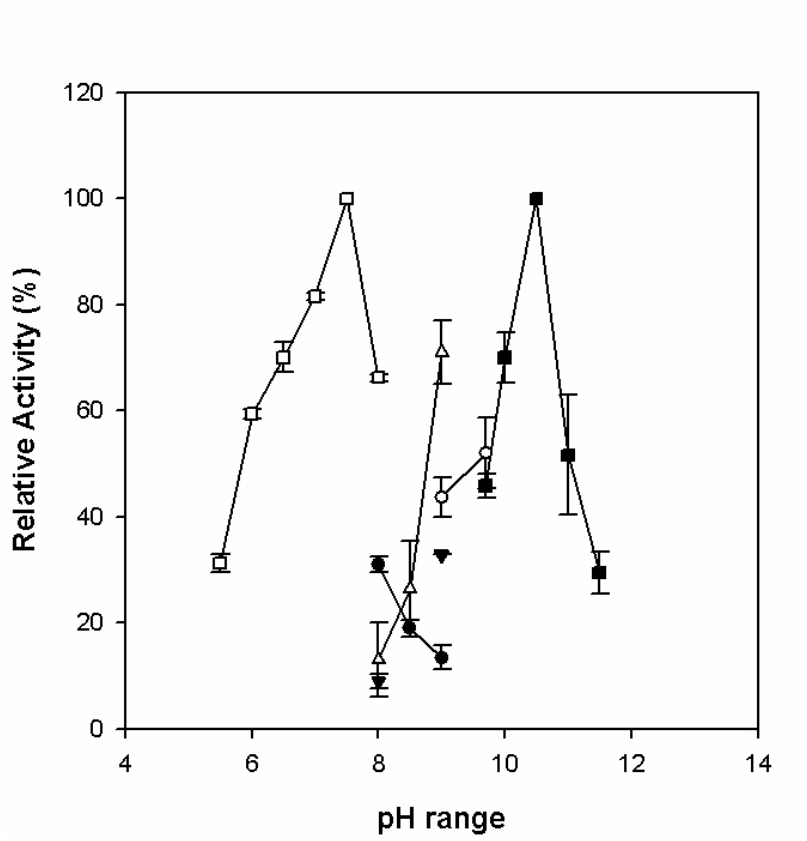


Figure 2-16 Optimal pH of the purified recombinant *Tg*ADH

Optimal pHs for alcohol oxidation and formation were determined by measuring the activities on the oxidation of 2-butanol and the reduction of 2-butanone, respectively. The buffers (100 mM) used were phosphate (unfilled squares), EPPS (filled circles and inverted triangles), Tris (unfilled triangles), glycine (unfilled circles), and CAPS (filled squares). The relative activity of 100% refers to the full activity of the recombinant enzyme that equals to 1073 U/mg of alcohol oxidation activity and 191 U/mg of ketone reduction activity, respectively. Error bars indicate standard deviations of the measurements.

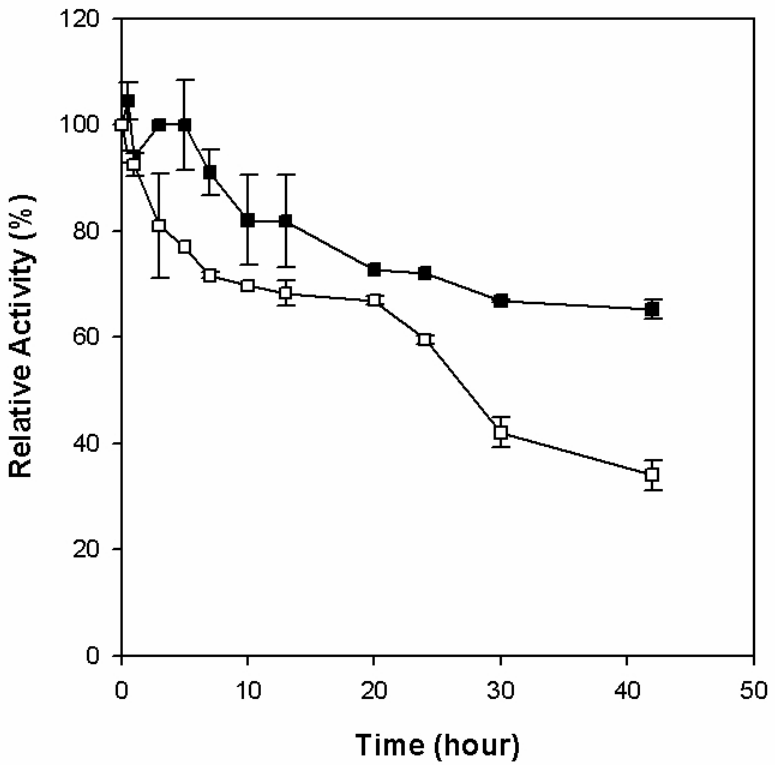


Figure 2-17 Thermostability of the purified recombinant *Tg*ADH

Open squares, incubation at 80 ^{0}C; filled squares, incubation at 95 ^{0}C. The relative activity of 100% equals to the initial ADH activity without heat treatment (1073 U/mg). Standard deviations of the measurements are indicated using error bars.

`

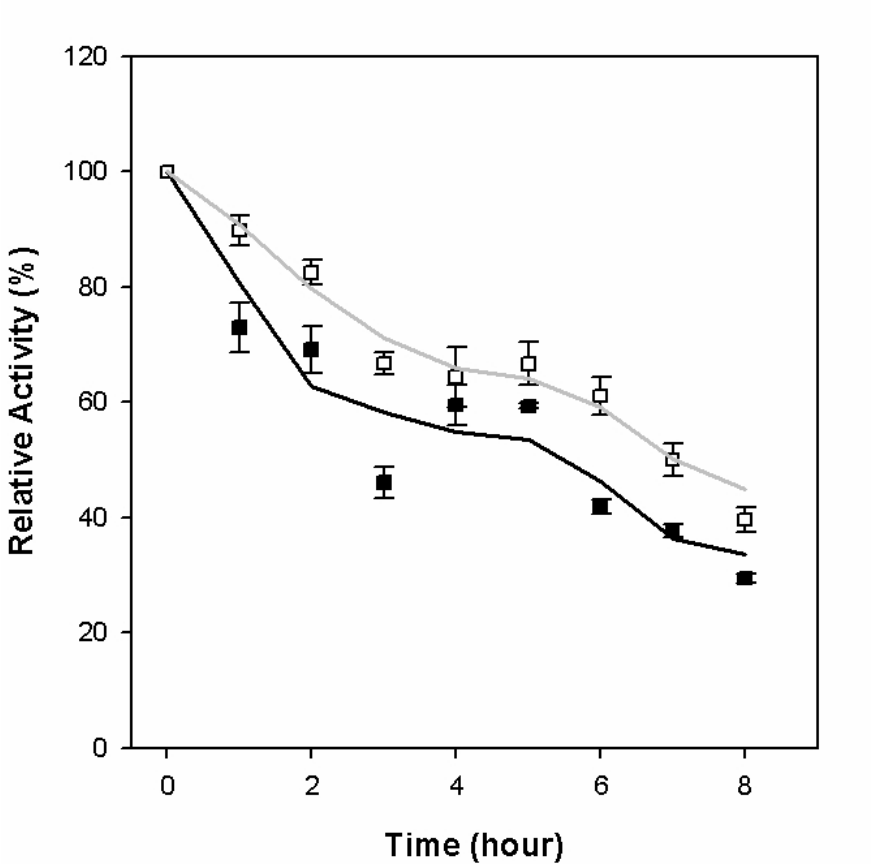


Figure 2-18 Oxygen sensitivity of the purified ADH from *T. guaymasensis*

Open circles, in the presence of 2 mM DTT and 2 mM SDT; filled circles, in the absence of 2 mM DTT and 2 mM SDT. The relative activity of 100% equals to the ADH activity prior to exposure to air (1029 U/mg). Standard deviations of the measurements are indicated using error bars.

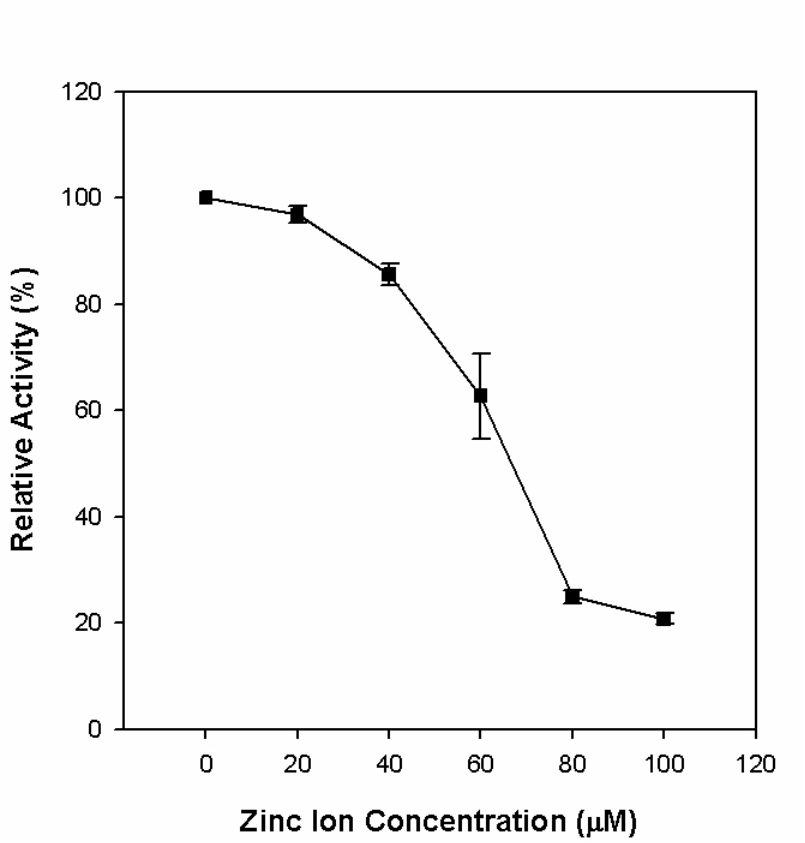


Figure 2-19 Effect of zinc on activity of *Tg*ADH

The activities were measured in the standard assay conditions. The relative activity 100% was defined as the ADH activity prior to zinc treatment (1037 U/mg). Standard deviations of the measurements are indicated using error bars.

2.5 DISCUSSION

ADHs using NAD or NADP as coenzyme are divided into three different groups, the zinc-dependent ADHs, the iron-containing ADHs, as well as the short-chain ADHs that are lack of metal. The native ADH isolated from *T. guaymensensis* represents a NADP-dependent, zinc-containing alcohol dehydrogenase (Ying et al., unpublished).

In hyperthermophilic archaea, several zinc-containing NAD(P)-dependent alcohol dehydrogenases have been discovered, including ADHs from aerobic hyperthermophilic archaea *S. solfataricus* (Ammendola et al., 1992) and *A. pernix* (Guy et al., 2003) or TDHs from anaerobic hyperthermophiles *P. furiosus* (Machielsen et al., 2006) and *P. horikoshii* (Ishikawa et al. 2007). Similar to the above-mentioned enzymes, both the native and recombinant *Tg*ADH were in the ternary structure of homotetramer, which is the usual structural characteristic of the previously characterized zinc-containing ADHs in archaea. However, *Tg*ADH represented different biochemical properties from these hyperthermophilic zinc-containing ADHs or threonine dehydrogenases that preferred to NAD as coenzyme, but native *Tg*ADH was specific for NADP as coenzyme. Interestingly, *Tg*ADH has high overall identities to NADP dependent zinc-containing ADHs from thermophilic bacteria, e.g., ADHs from *T. brockii* and *T. tengcongensis*. The sequence alignment indicated *Tg*ADH shared conserved co-enzyme NADP binding sites and active site predicted harboring catalytic zinc ion, which matched the biochemical characterizations. From the 3-D structure modeling, the monomer of *Tg*ADH folded into two domains, the catalytic domain closing to N-terminal end and one NADP-binding

domain closing to C-terminal end. The phylogenetic analysis between *Tg*ADH and the thermophilic and hyperthermophilic zinc-containing ADH indicated *Tg*ADH to be closer to the ADHs containing catalytic zinc atom only in evolution but further from the ADHs containing both catalytic and structure zinc ions (**Fig. 2-20**).

Cloning and sequencing of the entire encoding gene of *Tg*ADH provided fundamental information for over-expression of the hyperthermophilic enzymes in heterologous hosts. The production of recombinant extremophilic proteins in mesophilic hosts such as *E. coli* is highly desirable due to simpler culture conditions and typically higher yields. Compared to a series of chromatography for the native enzyme purification, only two steps including heat treatment and liquid chromatography were needed for purification. Because of its stability at high temperatures, one heat treatment step could significantly simplify the purification of the recombinant *Tg*ADH from *E. coli*. The over-expression of archaeral genes in bacterium is often challenged by poor yield or loss of activity due to different codon bias. However, the recombinant *Tg*ADH seems soluble, active and thermostable. Although native *Tg*ADH purified form *T. guaymensensis* directly presented a high concentration in the cells (Ying et al., unpublished), *E. coli* provided a much higher yield of recombiant *Tg*ADH at about 4~5 mg per gram cells. The recmombinant *Tg*ADH carried almost same activity and other catalytic properties with the native enzyme purified form *T. guaymensensis* directly. When cloned and expressed in mesophilic hosts, the enzymes usually retain its thermal properties, suggesting that these

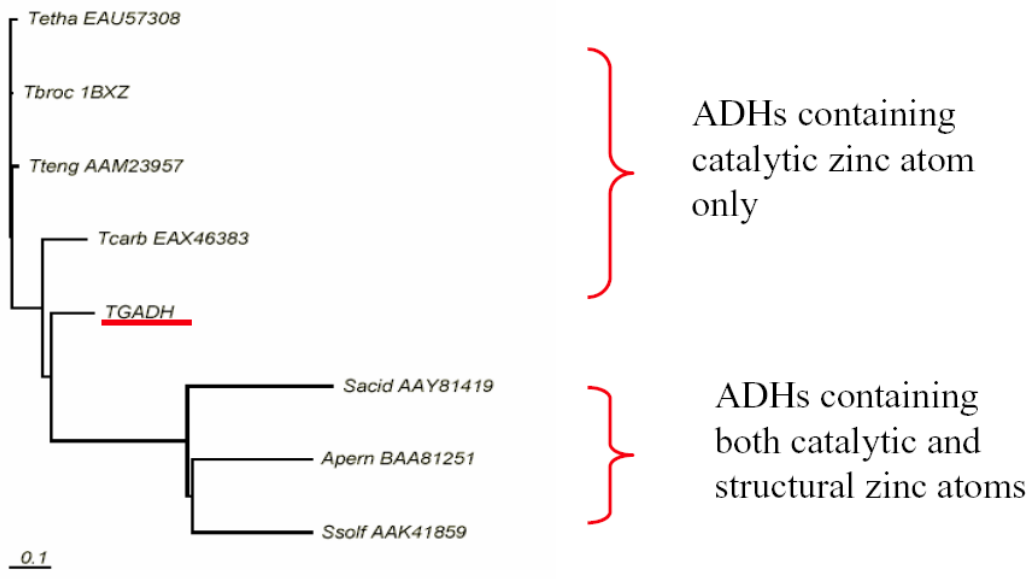


Figure 2-20 Phylogenetic tree of *Tg*ADH and some hyper/thermophilic zinc-containing ADHs

Scale bar indicates 0.1 substitutions per sequence position ClustalW (Thompson et al., 1994). Abbreviation: *Tetha, Thermoanaerobacter ethanolicus; Tbroc, Thermoanaerobacter brockii*; *Tteng, Thermoanaerobacter tengcongensis; Apern, Aeropyrum pernix; Ssolf, Sulfolobus solfataricus; Sacid, Sulfolobus acidocaldarius.*

properties would be genetically encoded.

The enzyme from *T. guaymasensis* possesses several outstanding features to be a competitive biocatalyst. Firstly, the enzyme presents one of the highest specific activities among the zinc-containing ADHs families. The thermoactivity over 1000 U/ mg at 80 ^{0}C was remarkably higher than other zinc-containing ADHs characterized except the TDH from *P. horikoshii. Tg*ADH was active within a broad temperature range from 30 to 95 ^{0}C as tested while the optimal temperature was even over 95 ^{0}C and could maintain a high activity when treated at high temperatures: the half life was around 26 hours when treated at 95 ^{0}C, which is the most thermostable one among the family of zinc-containing ADHs up to date. Thermostability and thermophilicity molecular mechanisms are varied, differing from enzyme to enzyme, which could be a combination of intrinsic stabilizing forces (such as salt bridges, hydrogen bonds, hydrophobic interactions) and extrinsic stabilizing factors. Sequence alignments, amino acid comparisons, and predicted 3-D structure comparisons indicate that *Tg*ADH is, indeed, very similar to mesophilic counterparties. Interestingly, the length of the hyperthermophilic /thermophilic protein was not less than the mesophilic homologous (**Table 2-7**). From the view of primary structure, the ratio of amino acid residues Ala, Arg, Glu, Lys and Pro was increased in *Tg*ADH, whereas that of Ala, Asn Gln, Ser and Val decreased. It was reported that the equal increase of oppositely charged residues especially Arg and Glu, in hyperthermophiles most likely led to the increased amount of ion pairs observed already in their proteins (Cambillau and Claverie 2000). In contrast, the uncharged polar residues Gln, Asn, Ser decreased in *T. guaymasensis* ADH, in which the first two are prone to

deamidation and known to be the most temperature sensitive (Cambillau and Claverie 2000). However, it is likely that other determinants are also critical for thermostability, and detailed structural comparisons between the two types of enzyme are needed.

The optimal pHs of both native and recombinant *Tg*ADH on the oxidation of alcohols are more alkaline than those on the reduction of aldehydes or ketones. Generally, dependence of enzyme activity on pH value is related to protonation at the active site (Koumanov et al., 2003). In a hyperthermophilic L-threonine dehydrogenase from *Pyrococcus horikoshii*, the proton dissociation model with two catalytic forms among three ionizable groups was derived to explain the experimental the examined pH dependence (Higashi et al., 2008). It was also reported in *Drosophila lebanonensis* short-chain alcohol dehydrogenase, the protonation/deprotonation transition was related to the coupled ionization of Tyr_{151} and Lys_{155} in the active site and the pH dependence of the proton abstraction was correlated with a reorganization of the hydrogen bond network in the active site (Koumanov et al., 2003). Likewise, the oxidoreductase activity of *Tg*ADH probably relies on a proton relay mechanism. The conformation of the residues at the catalytic site accomplishes with the deprotonation process, which would be an explanation for dissociation of substracts or cofactor from the enzyme when pH changes.

The sequence analysis indicated *Tg*ADH contained catalytic zinc atom only, which is consistent with the previous determination for the native enzyme (Ying et al., unpublished). Since Zn^{2+} cannot be further oxidized, most zinc-containing ADHs are resistant to oxygen; however, it was unexpected that *Tg*ADH was oxygen sensitive in

both native and recombinant form. The inactivation of the enzyme might be caused by the damage of labile amino acid residues such as cysteine residues. The *Tg*ADH contains four cysteine residues per subunit (Cys_{38}, Cys_{55}, Cys_{212} and Cys_{305}). Cys_{38} was highly conserved in zinc-containing ADHs and a putative active site residue, which has been approved to coordinate the binding of catalytic zinc in *T. brockii* ADH. The residue Cys_{55} was unique in *Tg*ADH and did not exist in the same location of any other zinc containing ADHs sharing high similarities, so site direct mutagenasis at Cys_{55} residue could shed light to the role of Cys_{55} in the oxygen sensitivity.

Produced in relatively high amounts by heterologous expression in *E. coli* and easily purified together with the outstanding stabilities, *Tg*ADH carries an obvious industrial perspective. It is highly *S*-enantioselective (Ying et al., unpublished data) make this enzyme a potential catalyst for industry, especially for the production of chiral compounds. However, the oxygen sensitivity is a disadvantage, and it would be a focus of further study on protein engineering to reconstruct the enzyme and explore the application of *Tg*ADH as one of the biocatalysts for industrial applications.

Chapter 3: Cloning, over-expression and characterization of an iron-containing alcohol dehydrogenase from *Thermotoga hypogea*

3.1 OVERVIEW

Thermotoga hypogea is an extremely thermophilic anaerobic bacterium capable of growing at 90 ^{0}C. The gene encoding *T. hypogea* alcohol dehydrogenase (*Th*ADH) was cloned, sequenced and over-expressed. DNA fragments of the gene encoding *T. hypogea* ADH were amplified directly from the genomic DNA by PCR. The gene sequence (1164 bp) was obtained successfully by sequencing all the DNA fragments. The deduced amino acid sequence was found to have high degrees of identity (~72%) to iron-containing ADHs from *Thermotoga* species and harboured typical iron and NADP-binding motifs, $Asp_{195}His_{199}His_{268}His_{282}$ and $Gly_{39}Gly_{40}Gly_{41}Ser_{42}$ respectively. The structural modeling showed that N-terminal domain of *Th*ADH contained α/β-dinucleotide-binding motif and its C-terminal domain was α-helix-rich region including iron-binding motif. The gene encoding *T. hypogea* ADH was functionally expressed in *Escherichia coli* using the vector pET-30a. The recombinant protein was expressed optimally in *E. coli* grown in the presence of 1 mM ferrous and induced by 0.4-0.6 mM IPTG. The recombinant enzyme was found to be soluble, active and thermostable with a subunit size of 43 kDa revealed by SDS-PAGE analyses. The native ADH from *T. hypogea* was purified to homogeneity for comparative analysis using a three-step liquid chromatography while the recombinant ADH over-expressed in *E. coli* was isolated by a simpler procedure including one-hour heat treatment. The activity of the purified recombinant enzyme was 69 U/mg and presented identical properties with the native enzyme. The optimal pH values for ethanol oxidation and acetaldehyde reduction were 11.0 and 8.0 respectively. The enzyme was oxygen sensitive and it has a half-life ($t_{1/2}$) of 20 minutes upon exposure to air. Compared to the zinc-containing *Tg*ADH, *Th*ADH was less thermostable and retained 50% of the

full activity after incubation at 70 ^{0}C for 2 hours. Successful high-level expression of *T. hypogea* ADH in *E. coli* will significantly facilitate further study on the catalytic mechanisms of iron-containing ADHs.

3.2 INTRODUCTION

Thermotoga hypogea is an anaerobic, extremely thermophilic bacterium. This strain is rod-shaped and has a characteristic outer sheath like structure. The optimal temperature for growth is 70 ^{0}C at pH 7.0 while the temperature range is 56 to 90 ^{0}C. The stain utilizes carbohydrates including xylan as carbon and energy sources and producing acetate, CO_2 and hydrogen as the major end products. Furthermore, ethanol is also produced as an end product of glucose/xylose fermentation (Fardeau et al., 1997).

Alcohol dehydrogenases catalyze the interconversion between alcohols and the corresponding aldehydes. ADHs can be classified into the following categories: zinc-containing ADHs, iron-dependent ADHs and short-chain ADHs that are lack of metal ions (Reid et al., 1994). Zinc ions have catalytic or structural functions in several enzymes including hyperthermophilic zinc-containing ADHs, however, only a few iron-dependent ADHs are known due to its instability (Radianingtyas et al., 2003). And some of them have been over-expressed in the mesophilic heterologous system *E. coli* and detailed 3-D structure was analyzed. For instance, the gene that encodes iron-containing alcohol dehydrogenase II (adhB) from bacterium *Zymomonas mobilis* was expressed well in *E. coli* under control of the lac promoter (approximately 0.25% of the total cell protein) (Conway et al., 1987). Additionally, the *adhE2* gene of *Clostridium acetobutylicum,* a gram-positive spore-forming anaerobic bacterium, coding for an aldehyde/alcohol dehydrogenase, was characterized to be iron-containing ADH and functionally expressed in *E. coli* (Fontaine et al., 2002). Most known hyperthermophilic iron-containing alcohol

dehydrogenases belong to hyperthermophilic archaea, the ADH from *Thermococcus hydrothermalis* was detected to be iron-containg and ADH from the hyperthermophilic archaeon *Pyrococcus furiosus* was purified and biochemical characterized to be an unusual oxygen-sensitive, iron- and zinc-containing ADH. Expression of the gene encoding ADH from *T. hydrothermalis* in *E. coli* is the first example of heterologous production of an iron-containing ADH from a hyperthermophilic archaeon, however, the native and recombinant *T. hydrothermalis* ADH showed clear differences in catalytic activity and thermostability (Antoine et al., 1999). Up to date, only the iron-containing hyperthermophilic alcohol dehydrogenases from *Thermococcus* strain ES-1 (Ying et al., 2008) and NADP-dependent, iron-containing 1,3-propanediol dehydrogenase encoded by TM0920 in *Thermotoga maritima* genome were reported to be successfully expressed in *E. coli* (Schwarzenbacher et al., 2004). And the 3-D structure of the *T. maritima* ADH that has been solved at 1.30 Å resolution is the only available detailed structure of the iron-containing ADH from hyperthermophiles.

T. hypogea ADH (*Th*ADH) is the first iron-containing ADH purified from hyperthermophilic bacteria (Ying et al., 2007). The recombinant *Th*ADH expressed and purified from *E. coli* could replace the enzymes from the native sources, and the higher level of ADH produced in *E. coli* has facilitated preparation of crystals for the determination of the three-dimensional structure of this distinct class of ADH.

3.3 MATERTIALS AND METHODS

3.3.1 Materials and Devices

All chemicals were of high technical grade and obtained from Sigma-Aldrich Canada Ltd (ON, Canada) or Fisher scientific company (ON, Canada). Restriction enzymes, DNA ladders and other reagents for molecular work were purchased from Fermentas Canada Inc. (ON, Canada); all reagents were used according to the manufacture's instructions. Major materials were same as listed in sections 2.2.1.1 and 2.2.1.2). Common used instruments were same as listed in section 2.2.1.

3.3.2 Microorganisms

Thermotoga hypogea (DSM 11164) was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany.

E.coli DH5α [*(supE44 ΔlacU169 Φ80 lacZΔM15) hsdR17 recA1gyrA96 thi-1 relA1*] (BRL, CA, USA)

E.coli BL21 (DE3) [*B F⁻ ompT hsdS$_B$ (r_B^- m_B^-) gal dcm (DE3)*] (Novagen, WI, USA)

E.coli BL 21 (DE3)-Rosetta [F^-, ompT, hsdSB ($r_B^-m_B^-$) gal dcm lacY1, pRARE (CamR)] (Stratagene, CA, USA)

3.3.3 Cultivation media and growth conditions

All media for cell cultivation named 2YT-medium was based on the standard protocols (Sambrook et al., 1989). Solid agar plates were made with 1.5% agar. The preparation of antibiotics and stock solutions were the same as listed in section 2.3.3.

Medium for *T. hypogea* cultivation was described previously (Fardeau et al., 1997; Ying et el., 2007). The medium was autoclaved and degassed before use. Before inoculation, 1 ml 25% $Na_2S_2O_3$, 0.14 ml 15% cysteine and 0.4 ml 3% Na_2S were added to each serum bottle (50 ml medium) and incubation was at 70 ^{0}C.

3.3.4 Cell cultivation

3.3.4.1 *E. coli* cultivation

Cell cultivation was carried out either in shake-flasks with baffles or in solidified agar plates with the medium composition as previously described in section 2.3.3. *E. coli* strains were grown in 2YT medium at 37 ^{0}C in a volume ranging from 500-1000 ml in shake flasks at 140-200 rpm by inoculating one colony from agar plate or with previous culture in a ratio of 1:100 or 1:50 (1-2% v/v). The cells were harvested by centrifugation. Storage of *E. coli* was carried out with a final concentration of 10% glycerol at –20 ^{0}C or –80 ^{0}C.

3.3.4.2 *T. hypogea* cultivation

T. hypogea was cultured in 50 ml serum bottles. Cells were cultured for 16-18 hours anaerobically at 70 ^{0}C and the growth was monitored using cell counting. After

incubation, the cells were harvested by centrifugation at 10,000× *g*. The resulting cell pellet was frozen in liquid nitrogen immediately and stored at -80 ^{0}C.

3.3.5 Cell harvest and storage

The cultivated cells were collected by centrifugation. Glycerol was added to cultures with a final concentration of 10% for storage of *E. coli* cells at –20 ^{0}C or –80 ^{0}C. This method was used for preparation of competent cells as well as preparing stock cultures of *E. coli* cells harboring either pET vectors or pET recombinants. When the *E. coli* grew to OD_{600nm} of 0.6-0.8, 0.8 ml of the culture was removed and transferred to a cryo-vial, and 0.2 ml of 50 % glycerol was added. The culture was mixed well and stored short-term at –20 ^{0}C. Cells with glycerol at a final concentration of 10% were frozen in liquid nitrogen quickly and then stored in –80 ^{0}C for long-term storage.

3.3.6 Preparation for competent cells

Competent cells for transformation were prepared using the same way as described as in section 2.3.6. *E. coli* BL 21 (DE 3) or BL 21 (DE 3)-Rosetta were cultivated in 2YT medium without antibiotics. Then the bacteria were harvested by centrifuging at 4,000×g at 4 ^{0}C and treated by pre-cooled 0.1 M $CaCl_2$ slightly, which stimulated the cells to be more easily altered cell walls that DNA can be passed through easily. Ultimately, the competent cells were frozen immediately in liquid nitrogen, and stored the tubes at -80 ^{0}C.

3.3.7 Gene cloning for *Th*ADH

The genome sequence of *T. hypogea* is unknown yet and the encoding gene of *Th*ADH was cloned and sequenced directly from the genomic DNA. The fragments were obtained by the strategy of primer walking shown in **Fig. 3-1**.

3.3.7.1 Preparation of *T. hypogea* genomic DNA

The genomic DNA from *T. hypogea* as PCR template was isolated by lysis of the cells. The cells were grown in 50 ml degassed liquid broth and harvested by centrifuging. After treated by 25 % SDS, 1mg/ml Proteinase K and 3 M NaAc, DNA was isolated from the mixture using the solution of phenol, chloroform and isoamyl alcohol (25: 24: 1). The genomic DNA was participated by 100% isopropanol and washed by 75% ethanol and concentrations of samples were quantified using NanoDrop Spectrophotometer.

3.3.7.2 Cloning of the entire *Th*ADH encoding gene

DNA fragments carrying target gene were amplified from genomic DNA directly by PCR. The PCR amplification for cloning coding sequence of N-terminal of *Th*ADH was performed by Proof reading KOD Hot Start DNA Polymerase in a volume of 25 μl and all the reagents were added following the standard conditions recommended by the suppliers (**Table 3-1**). Pimer desgining software DNASTAR and GENERUNNER were used to predict the potential locations of the primers and optimize the parameters of oligonucleotides respectively. Principles of primers design were described in sectin 2.3.7. On the basis of N-terminal (MENFVFHNPTKLIFG) and internal sequence (LMLYGGGSI), two oligonucleotides THADHNF and THADHIR (**Table 3-2**) were

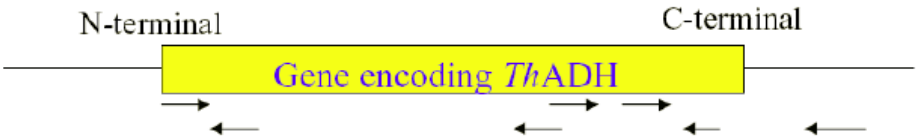


Figure 3-1 Strategy for *Th*ADH coding gene sequencing

Horizontal arrows indicate the direction and location of major oligonucleotide primers for PCR amplification.

Table 3-1 PCR mixture

Reagents	Final concentration	Amount (μl)
Polymerase buffer (10×)	-	2.5
Template	50~100 ng/μl	2
Primer1	1 μM	1
Primer2	1 μM	1
DNTPs	0.2 mM	2.5
$MgSO_4$	1.5 mM	1.5
Polymerase	0.5 unit/μl	0.5
ddH_2O	-	14
Total	-	25

Table 3-2 Primers designed for cloning & sequencing the gene encoding *Th*ADH

Name of primers*	Nucleotide sequence (5'-3')	Restriction enzyme Sites (underlined)
THADHNF	ATGGAGAACTTCGTCTTCCACAATCC	
THADHIR	TATCGATCCACCACCGTATAGCATCAG	
THAUNF	AACTTCGTCTTCCACAATCC	
THAUIR01	TCATCTCCGTTCCTGTCG	
THADHYIR	GTGTGTGCTATTGCGTCG	
THADHYIF1	ACTGAGATGAACGGAAACG	
THADHYIF2	GCGACGATAGCCCTGAAC	
TH11NF	ATGAGYAAGGATGCGCGGT	
TH11CR	GAGCACCATTATTCTTCCC	
TH11CF01	ATACCAGCACAGGACATCG	
TH11CF02	CGATAGCCCTGAACGGTC	
TH11N01	AACTTGGTGGGATTGTGG	
TH11N02	GCACTGGATTGGGTTTGAC	
TH11C01	GAAAGGTGGGGATTGGTC	
TH11C02	CGCAGGTAGGAAAGGTGG	
TH08IF1	GATAATGAAGTCCACAGAAGTTTTG	
TH08IF2	TACGTCTATAAACAAAAACCACAGC	
TH08IR1	CCTTAGCCAGTTCGGACACG	
TH08IR2	TGGATTTGCTGTCTCTTTTCTTSS	
THNCNF	TAGAATTCATGAGCAAGATGCGCGGTTTTC	*Nco*I
THECCR	ACCTCGAGTCACTCCTCTATGATGACC	*Eco*RI

[*], primer properties like melting temperature (Tm), GC content (GC%), primer loops and primer dimmers were evaluated by a DNA analysis tool Gene Runner (Hastings Research, Inc., Las Vegas, USA). The table indicates all the key primers used for both fragments cloning and specific amplification. The forward and the reverse primers with the restriction enzyme sites were the specific primer designed based on the confirmed sequence for the amplification of the entire *Th*ADH encoding gene.

synthesized and used as forward and reverse PCR primers, respectively. Reaction was performed using a thermal cycler TC-312 (Techne incorporated, NJ, USA) with 25 pmol of each primer against 100 ng of genomic DNA isolated from *T. hypogea* cells. After preheating at 95 ^{0}C for 2 min, the cycling program consisted of 36 cycles of denaturation at 95 ^{0}C for 20 seconds, annealing at 55 ^{0}C for 20 seconds, and extension at 70 ^{0}C for 30 seconds. The PCR products were run on a 2% agarose gel and the resulting 150 bp PCR product was sequenced by the dye-termination method at Molecular Biology Core Facility, University of Waterloo. DNA and protein sequencing were analyzed with Gene Runner (Hastings Research Inc., Las Vegas, USA) and compared to the GenBank database by BLAST (Altschul et al., 1997). The BLAST search of the sequenced fragment indicated its highest similarity to the putitive NADH-dependent butanol dehydrogenase from *T. maritima* TM0820, which could be considered as part of the target gene. Considering the high sequence similarity among the *Thermotoga species*, gradual primer walking was selected as the effective strategy. After the fragment was firstly sequenced as if it were a shorter fragment and in order to completely sequence the region of interest, new primers complementary to the final few bases of the known sequence were designed and synthesized to obtain contiguous sequence information. Thus, the new forward primer THAUNF was designed based on the known sequence, while the reverse primer THAUIR01 was designed based on the conserved nucleotide sequence among the *Thermotoga* species whose genome information were already available including *T. maritima*, *T. petrophila* and *Thermotoga* sp. RQ2. A 1.3 kb PCR product covering majority of *Th*ADH encoding gene was obtained after 36 cycles PCR at 95 ^{0}C for 20 seconds, 58 ^{0}C for 20 seconds, 70 ^{0}C for 30 seconds. Then, a series of PCR

were applied under the above condictions and directed by downstream primers including THADHIR, THADHIF1 and THADHIF that, designed based on that preliminary sequencing results of *Th*ADH and conserved sequences analysis (**Table 3-2**). Eventually the nucleotide sequence of 1164 bp encoding 387 amino acid residues was cloned from *T. hypogea* genome.

3.3.8 Data mining

Gene sequence and the deduced amino acid sequence of *Th*ADH were analysed using the BLAST program (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul et al., 1997). Sequence alignments with homologues in iron-containing ADH families and phylogenetic trees were constructed using Clustal W tool with default parameters (Thompson et al., 1994). Theoretical molecular weight was calculated using the ProtParam program at the ExPASy Proteomics Server with standard parameters (Gasteiger et al., 2005). 3-D structure of *Th*ADH monomer was modeled using the Swiss Model server (Guex et al., 1997; Peitsch et al., 1995; Schwede et al., 2003), and visualization and analysis of the 3-D structure were preformed using PyMOL software.

3.3.9 Construction of the recombinant plasmid

3.3.9.1 Vectors and plasmid isolation

Encoding gene of *Th*ADH was inserted into vector pET-30a (5360 bp, P_{T7}, Kan^{R}; Novagen, WI, USA) and over-expressed in *E.coli* host strains. Plasmids were isolated by the alkaline lysis method described in section 2.3.9.2 (Birnboim et al., 1979).

3.3.9.2 DNA restriction digestion

Restriction map of *Th*ADH encoding gene was gained using the on-line tool (http://tools.neb.com/NEBcutter2/index.php). Only the restriction endonucleases that did not cut within the coding gene were selected for recombinant vector construction. Digestion of the DNA was performed in the recommended buffer for 2-4 hours using 10-15 U of the endonucleases (1-1.5 µl; 10 U/µl) for 0.5-1.5 µg DNA. The digestion reaction was incubated at 37 ^{0}C for completion of the restriction digestion and the reaction mixture was analyzed using agarose gel electrophoresis. For DNA fragments used for ligation, reaction mix was purified with the PCR Purification kit (Qiagen, ON, Canada) and quantified by agarose gel electrophoresis.

3.3.9.3 Ligation of DNA fragments

The restriction endonucleases selected in this research, both *Nco*I and *Eco*RI, could provide the sticky end that could keep a high efficiency of the ligation. For the successive ligation reactions of the insert to vectors, a 10 µl reaction volume was used with 3:1 molar ratio of insert to vector together with 1 U of T4 DNA Ligase and 10× Ligation buffer. The ligation mixture was incubated at 16-20 ^{0}C overnight (16-18 hours). After ligation reaction was completed, the mixture was used for transformation.

3.3.9.4 Transformation and selection

In the plasmid transformation, 10 µl of the ligation product was added to 100 µl corresponding competent cells and then heat shock at 42 ^{0}C to the mixture of plasmids and competent cells was applied. After heat shock, intact plasmid DNA molecules

replicated in bacterial host cells. To help the bacterial cells recover from the heat shock, the cells were incubated with non-selective growth media. However, due to the low number of bacterial cells those contain the plasmid and the potential for the plasmid not to propogate itself in all daughter cells, selection for bacterial cells that contain the plasmid were performed with antibiotic selection. *E. coli* BL21 (DE3) strains used in this research sensitive to common antibiotics including kanamycine, while pET-30a has been engineered to harbor the genes for antibiotic resistance. The bacterial transformations were plated onto media containing 50 mg/ml kanamycine and only bacteria possessing the plasmid DNA would have the ability to form colonies.

3.3.9.5 Optimization of induction condition for the recombinant *E. coli*

The chemical induction of the *lac* promoter was accomplished by the addition of IPTG. Under control of the T7-*lac* promoter, the recombinant *Th*ADH was obtained in the periplasmic space when IPTG was added to the 2YT medium with 1 mM ferrous. To optimize the growth condition, both the concentration of the inducer as well as the growth phase of recombinant cells at which it was added were tested. The inducer IPTG was added as a gradient of 0, 0.2 mM, 0.4 mM, 0.8 mM, 1 mM and the optimum concentration of the inducer was detected by amount of recombinant protein on the SDS-PAGE. Then IPTG was added in the exponential phase when OD_{600nm} of the cell culture reached 0.4-1.0, which was recommended by the pET manual. However, induction at OD_{600nm} 0.8 provided an ideal yield with high activity. Therefore, induction in this study was set at OD_{600nm} of 0.8.

3.3.10 Protein assay

The protein concentrations of all samples were determined using the Bradford method and bovine serum albumin (BSA) served as the standard protein (Bradford, 1976). 200 µl of Bio-Rad reagent was mixed with 800 µl of pure water as the blank control, and the assay was performed by mixing 200 µl of Bio-Rad reagent with 800 µl protein solution. The specific absorbance was tested by spectrophotometer (GENESYS 10 Vis, NJ, USA) at 595 nm.

3.3.11 Determination of enzyme activity

The catalytic activity of *T. hypogea* ADH was measured anaerobically at 80°C by specific absorbance change of NADP(H) at 340 nm (ε_{340} = 6.3 $mM^{-1}cm^{-1}$, Ziegenhorn et al., 1976). The catalytic reaction by *Th*ADH involved two directions, ethanol-dependent reduction of NADP or the acetaldehyde-dependent oxidation of NADPH. Unless specified, the enzyme assay was done in duplicate using the assay mixture (2 ml) for the oxidation of alcohol that contained 20 mM 1-butanol and 0.2 mM NADP in 100 mM 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer at pH 11.0. The assay mixture (2 ml) used for the reduction of aldehyde contained 22 mM butyraldehyde and 0.1 mM NADPH in 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 8.0). One unit (U) is defined as the production of 1 µmol of NADPH per minute.

3.3.12 Preparation of cell-free extracts and investigation of oxidoreductase activities

All procedures for the preparation of cell-free extracts were carried out anaerobically. The 5 g frozen cells of *E. coli* carrying *Th*ADH encoding gene were re-suspended in 25 ml of 50 mM Tris–HCl buffer (pH 7.8) containing 2 mM dithiothreitol, 2 mM sodium

dithionite and 5% (v/v) glycerol. Lysozyme and DNase I were added to the cell suspension with the final concentrations of 0.1 mg/ ml and 0.01mg/ml respectively. The mixture was then incubated at 37°C for 2 hours with stirring. After centrifugation at 10,000 × *g* for 30 min at room temperature, the supernatant was collected as cell free extract for further use. ADH activities in the cell free extract were determined spectromatically as described in section 3.3.11.

3.3.13 Purification of the recombinant *Th*ADH

All purification steps were carried out anaerobically using the FPLC system considering the oxygen sensitivity of native *Th*ADH (Ying et al., 2007). Since the enzyme was thermostable, a step of heat precipitation was applied prior to the column chromatography. The cell extract was incubated at 60 ^{0}C for half an hour and the solution turned gel-like. The denatured proteins and cell debris in the cell-crude extract were removed by centrifugation at 10,000 × g for 30 min at room temperature. The supernatant containing enzyme activity were collected and pooled to a DEAE-Sepharose column (2.6 × 10 cm) equilibrated with buffer A. A linear gradient (0- 0.5 M sodium chloride in buffer A) was applied at a flow rate of 2.5 ml per minute and the *Th*ADH was eluted and collected anaerobically.

3.3.14 Size exclusion chromatography

After purification of the recombinant *Th*ADH from *E. coli* DEAE-Sepharose column, size exclusion chromatography was applied in order to determine the molecular mass of its native form. The enzyme sample was loaded onto the gel filtration column Superdex 200 (2.6 × 60 cm) equilibrated in 50 mM Tris-HCl (pH 7.8) containing 100 mM KCl at a

flow rate of 2 ml per minute. The size of the native form of *Th*ADH was calculated based on the elution volume of standard proteins (Pharmacia, NJ, USA) that contained blue dextran (molecular mass, Da, 2,000,000), thyroglobulin (669,000), ferritin (440,000), catalase (232,000), aldolase (158,000), bovine serum albumin (67,000), ovalbumin (43,000), chymotrysinogen A (25,000) and ribonuclease A (13,700).

3.3.15 Protein gel electrophoresis

The fraction containing the dominated activity was loaded to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described previously (Laemmli 1970). Sample preparations were same as described in section 2.3.13.4. The low molecular weight range protein marker (Bio-Rad Laboratories Inc., ON, Canada) was used for indication of the samples on the SDS-PAGE.

3.3.16 Characterization of catalytic properties

3.3.16.1 Optimum pHs

The optimal pHs of ethanol-dependent oxidation of native and recombinant *Th*ADHs were determined by enzyme assay of butanol oxidation or butyraldehyde reduction. Standard enzyme assays at 80 ^{0}C for alcohol oxidation were applied using a set of 100 mM buffers (Ying et al., 2007): Tris/HCl (pH 8.0, 8.5, 9.0) and CAPS (pH9.0, 9.7, 10.0, 10.5, 11.0, 11.5, and 12.0). The optimal pH value of acetaldehyde-dependent reduction of native and recombinant *Th*ADH was measured between pH 6.0 to 9.0 using the following 100 mM buffers PIPES (pH 6.0, 6.5, and 7.0), HEPES (pH 7.0, 7.5, and 8.0), Tris/HCl (pH 8.0, 8.5, 9.0).

3.3.16.2 Temperature dependence

The effect of the temperature on the enzyme activity was examined at temperatures from 30 to 95 ^{0}C by standard essay conditions described in section 3.3.11 except the temperature.

3.3.16.3 Thermostability

Enzyme thermostability was evaluated by incubating the enzyme in sealed serum bottles at 70 ^{0}C and 90 ^{0}C, respectively. The residual activities of each sample at different time intervals were measured parallelly using the standard assay conditions.

3.3.16.4 Oxygen sensitivity

The effect of oxygen on enzyme activity was investigated exposing the enzyme samples in the air at room temperature and determining the residual activity after oxygen exposure. The exposure was performed in the presence and absence of 2 mM DDT and SDT. The residual activities of each sample at different time intervals were measured using the standard assay.

3.4 RESULTS

3.4.1 Cloning of *T. hypogea* ADH

Amino-terminal of the native *Th*ADH sequence was determined by using Edman degradation and five other internal sequences were obtained using mass spectrometry, two of which were aligned and revealed to have high similarity to an iron-containing ADH TM0820 from *T. maritima* (**Table 3-3**). Primers were designed based on the conserved nucleotide sequence among the *Thermotoga* species (**Table 3-2**) and consequently, PCR amplifications resulted in DNA fragments with a length of about 150 bp that was sequenced and showed to have high similarity to the iron-containing ADH TM 0820. Referring to the nucleotide sequence of TM0820, the non-degenerate primers were designed based on conserved fragments and partially sequence fragments of *Th*ADH encoding gene; and PCR amplifications were conducted using the *T. hypogea* genomic DNA as template. PCR using primer THAUNF and THAUIR1 produced a 1.3 kb band on 1% agarose gel and the primary sequencing results indicated 1.1 kb of this fragment had a high similarity to TM0820. PCR product of the entire *Th*ADH encoding gene amplified using specific primer THNCNF and THECR showed a single 1.2 kb band on 1% agarose gel (**Fig. 3-2**). Confirmed by sequencing, the gene consisted of 1164 bp nucleotides, encoding the 387 amino acids peptide chain with a calculated molecular weight of 43374 Da (**Fig. 3-3**).

Table 3-3 N-terminal and internal amino acids fragments of *Th*ADH

Location	Sequences
N-terminal[a]	MENFVFHNPTKLIFG
Internal 1[b]	LPLLLHLE(L)[c]
Internal 2[b]	RAPVSL
Internal 3[b]	LMLYGGGSI
Internal 4[b]	PRSLSLR(A) [c]
Internal 6[b]	LILAS

[a] amino-terminal sequence was determined by using Edman-degradation
[b] internal sequences were determined by using mass spectrometry
[c] amino acid in parentheses indicateed less certainty

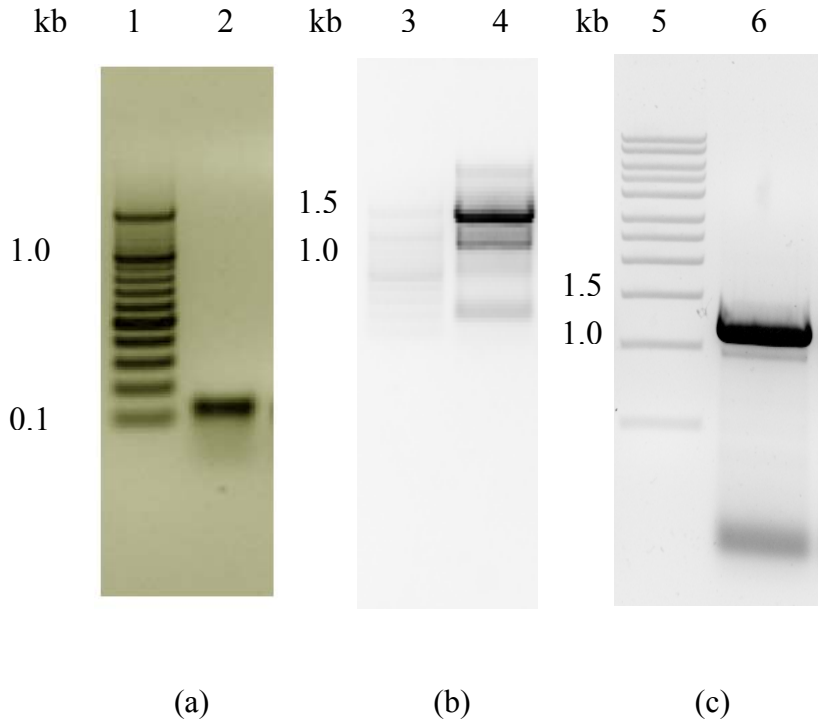


Figure 3-2 Cloning of *Th*ADH encoding gene by PCR

(a), amplification using primers THADHNF and THADHIR; (b), amplification using primers THAUNF and THAUIR1; (c), amplification of entire gene using primers THNCNF and THECR. Lane 1 and 3, 100 bp DNA ladders (Fermentas Canada Inc., ON, Canada); lane 5, 1 kb DNA ladders (Novogan, WI, USA); Lane 2, 4 and 6, PCR products.

```
1          ATGGAGAACTTCGTCTTCCACAATCCCACCAAGTTGATATTCGGTAAAGGAACCATTCCA
1           M  E  N  F  V  F  H  N  P  T  K  L  I  F  G  K  G  T  I  P

61         AAGATAGGCGAGGAGATCAAATCTTTCGGGATCAAAAAGGTTTTGATGCTCTACGGCGGT
21          K  I  G  E  E  I  K  S  F  G  I  K  K  V  L  M  L  Y  G  G

121        GGTTCGATAAAGAAGAACGGTGTCTACGATCAGGTCGTGGAATCTTTGAAAAGAAACGGC
41          G  S  I  K  K  N  G  V  Y  D  Q  V  V  E  S  L  K  R  N  G

181        ATCGAGTGGGTTGAGGTTTCTGGCGTCAAACCCAATCCAGTGCTCTCGAAGGTCCACGAG
61          I  E  W  V  E  V  S  G  V  K  P  N  P  V  L  S  K  V  H  E

241        GCTATAGAAGTTTGCAGAAAAGAGAACGTGGAGGCTGTTCTGGGCGTCGGTGGTGGAAGT
81          A  I  E  V  C  R  K  E  N  V  E  A  V  L  G  V  G  G  G  S

301        GTTATCGATTCAGCTAAAGCGATCGCAGCCGGTGTTCTCTACGAAGGAGACATTTGGGAC
101         V  I  D  S  A  K  A  I  A  A  G  V  L  Y  E  G  D  I  W  D

361        GCATTCGCCGGAAAGCATAAGATCAACAATGCCTTGCCAGTGTTCGCAATTTTGACCATA
121         A  F  A  G  K  H  K  I  N  N  A  L  P  V  F  A  I  L  T  I

421        TCTGCCACTGGAACTGAGATGAACGGAAACGCCGTGGTCACCAACGAAAAGACCCAGGAA
141         S  A  T  G  T  E  M  N  G  N  A  V  V  T  N  E  K  T  Q  E

481        AAGTGGGCGATCAGTGCAAAGTGTCTTTATCCACGAGTTTCTATAATCGATCCCACCGCA
161         K  W  A  I  S  A  K  C  L  Y  P  R  V  S  I  I  D  P  T  A

541        CAGTTTTCTCTACCGAAGGAGCAGACCGTCTATGGTGCGGTCGACGCAATAGCACACACG
181         Q  F  S  L  P  K  E  Q  T  V  Y  G  A  V  D  A  I  A  H  T

601        CTCGAGTACTACTTCGACGGTTCAGACTCGGACATACAGAACCAGATCAGCGAGTCCATT
201         L  E  Y  Y  F  D  G  S  D  S  D  I  Q  N  Q  I  S  E  S  I

661        ATCAGATCGATAATGAAGTCCACAGAAGTTTTGATAGACAATCCACAAGACTACGAGGCG
221         I  R  S  I  M  K  S  T  E  V  L  I  D  N  P  Q  D  Y  E  A

721        AGGGCGAACTTCGCCTGGTGTGCGACGATAGCCCTGAACGGTCTGACCGCCGCAGGTAGG
241         R  A  N  F  A  W  C  A  T  I  A  L  N  G  L  T  A  A  G  R

781        AAAGGTGGGGATTGGTCCTGTCACAAGATAGAGCATTCTCTCAGCGCGCTCTACGACATT
261         K  G  G  D  W  S  C  H  K  I  E  H  S  L  S  A  L  Y  D  I

841        GCTCACGGTGCAGGACTTGCGATCGTTTTCCCCGCGTGGATGAGATACGTCTATAAACAA
281         A  H  G  A  G  L  A  I  V  F  P  A  W  M  R  Y  V  Y  K  Q

901        AAACCACAGCAGTTCGAGAGGTTCGCGAAGCACGTTTTCTCGATCGATGCCGTGGGAGAA
301         K  P  Q  Q  F  E  R  F  A  K  H  V  F  S  I  D  A  V  G  E

961        GAAGCGATCTTGAAAGGTATAGACGCTTTCAAAGCTTGGCTCAGGAAGGTCGGTGCTCCC
321         E  A  I  L  K  G  I  D  A  F  K  A  W  L  R  K  V  G  A  P

1021       GTTTCGTTGAGAGACGTTGGTATACCAGCACAGGACATCGACAGGATCGTCGAGAACGTC
341         V  S  L  R  D  V  G  I  P  A  Q  D  I  D  R  I  V  E  N  V

1081       ATGAAACAGGGTCCATCCTTCGGTGTTCTGAAGAAGCTCGGTAAGGAAGATGTGAAACAG
361         M  K  Q  G  P  S  F  G  V  L  K  K  L  G  K  E  D  V  K  Q
```

```
1141      ATCCTGCTCATAGCTTCCCAATGAGACATTAAAATGATAAAATCAAAGTGTGGAGGCGTG
381        I  L  L  I  A  S  Q  *  D  I  K  M  I  K  S  K  C  G  G  V

1201      T
```

Figure 3-3 Nucleotides and deduced amino acid sequences of *Th*ADH

The amino acid sequence was deduced using the program DNAMAN (Lynnon Corporation, Vaudreuil-Dorion, Quebec, Canada). The stop codons were marked with asterisk. The nucleic acids located downstream of the *Th*ADH were highlighted in grey.

3.4.2 Sequence analysis

The BLAST analysis of the deduced amino acid sequence of *Th*ADH showed the identity as high as 72% to iron-containing alcohol dehydrogenases from hyperthermophilic bacteria in genus of *Thermotoga* such as *T. maritima* (Nelson et al. 1999), *T. petrophila* (Copeland et al., 2007-a), 60% identity to *Fervidobacterium nodosum* Rt17-B1 (Copeland et al., 2007-b) and 58% identity to *Thermosipho melanesiensis* (Copeland et al., 2007-c), moderately high identity (46-58%) to those enzymes from other thermophilic bacteria such as *Symbiobacterium thermophilum* (Ueda et al., 2004) and *Thermoanaerobacter ethanolicus*. The *Th*ADH also showed similarity to ADHs from the mesophiles, e.g. 54% identity to ADH from *Alkaliphilus oremlandii*. *Th*ADH belonged to a group of uncharacterized oxidoreductases of the iron-containing alcohol dehydrogenase family (**Fig. 3-4**). Amino sequences alignment of *Th*ADH with its homologues indicated that *Th*ADH haboured conserved coenzyme NADP binding motif and putative active site for binding iron (**Fig. 3-5**), which matched the results from biochemical characterization of *Th*ADH to be NADP-dependent and iron-containing (Ying et al., 2007).

Though *Th*ADH showed high similarity in amino acids sequence with ADH from thermophile *Fervidobacterium nodosum* and mesophile *Alkaliphilus oremlandii*, the amino acids composition indicated that *Th*ADH and its thermophilic homologues *F. nodosum* alcohol dehydrogenase had higher ratio (molar fraction, >0.8% increase or decrease) for Ala, Arg, Lys, Thr and Val but lower ratio for Asn, Glu, Tyr and Met when

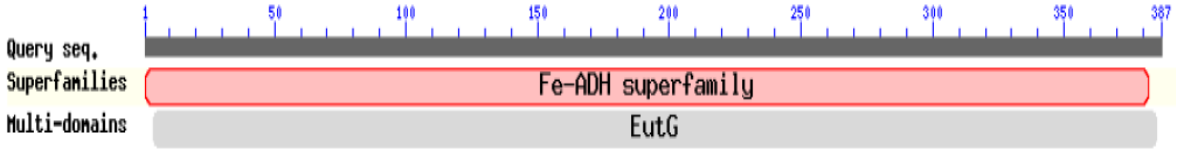


Figure 3-4 Putative conserved domains of *T. hypogea* ADH

Figure was constructed using BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast). Fe-ADH superfamily, a group of uncharacterized iron-containing alcohol dehydrogenase; EutG, alcohol dehydrogenase.

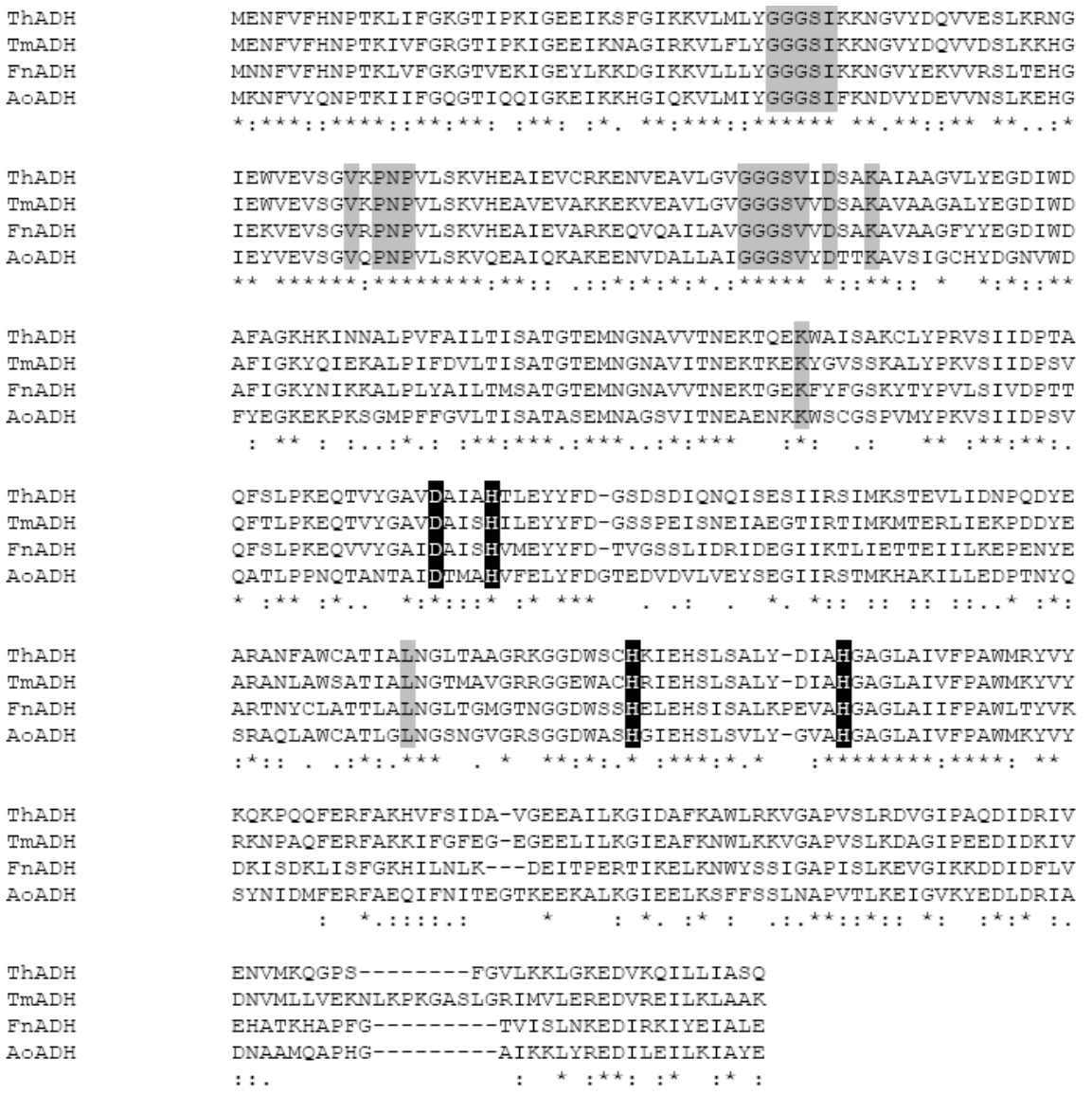


Figure 3-5 Amino acids sequences alignment among *Th*ADH and its homologous enzymes

The sequences were aligned using Clustal W (Thompson et al., 1994). ThADH, *T. hypogea* ADH; TmADH, *T. maritima* ADH; FnADH, *Fervidobacterium nodosum* ADH; AoADH, *Alkaliphilus oremlandii* ADH. ‘‘*’’, residues or nucleotides that are identical in all sequences in the alignment; ‘‘:’’, conserved substitutions; ‘‘.’’, semi-conserved substitutions; “-”, no corresponding amino acid. Highlighted in black, putative catalytic iron binding sites; highlighted in grey, putative motif of coenzyme binding sites.

compared to the ADH from the mesophilic *A. oremlandii* (**Table 3-4**). Particularly, the amino acid composition of *Th*ADH had higher ratio for Ala, Pro, Trp and Val but lower ratio for Glu, Lys and Thr than that of the ADH from thermophile *F. nodosum*. Furthermore, the exchange of hydrophilic and large hydrophobic residues such as Ser, Asn for the small hydrophobic amino acids Ala and Val in *Th*ADH (**Table 3-5**) may also contribute to the higher thermostability of by locking the enzyme in a conformation with a higher density of packing and decreased structural flexibility (Peretz et al., 1997).

The tertiary structural modeling of monomer of *Th*ADH showed two typical domains separating with a deep cleft (**Fig. 3-6**). Similar to the 3-D structure of iron-containing 1, 3-propanediol dehydrogenase (TM0920) from *T. maritima*, the N-terminal domain of *Th*ADH was formed by an α/β region containing the dinucleotide-binding fold, whereas the C-terminal part was an all-helical domain responsible for the iron binding. The *Th*ADH consisted of amino-terminus contained a GGGS motif (residues 39-42) which was well accepted to be involved in the interactions with the pyrophosphate groups of NADP, and the sequence harbored most of amino acid residues responsible for NADP binding that were observed in the crystal structure of TM0920 (Sulzenbacher et al., 2004). The putative active site motif was identified to be $Asp_{195}His_{199}His_{268}His_{282}$. Both the putitive iron binding site and coenzyme binding residues were found to be conserved in *Th*ADH and its thermophilic and mesophilic counterparts (**Fig. 3-5**).

Table 3-4 Amino acids components and abundance of *Th*ADH and its homologous

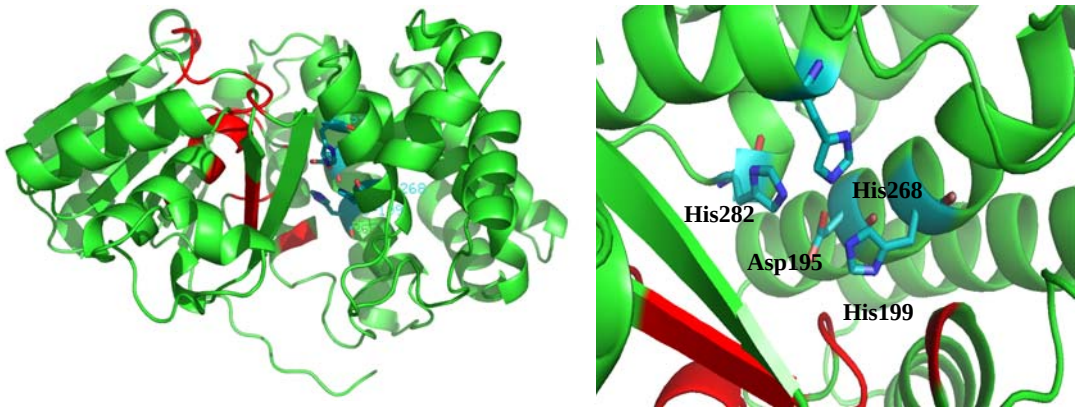
Amino acid	ThADH 387 aa	FnADH 385aa	AoADH 388 aa
Ala (A)	38(9.8)	27 (7.6)	31 (8.0)
Arg (R)	11(2.8)	7 (2.0)	6 (1.5)
Asn (N)	16(4.1)	16 (4.5)	19 (4.9)
Asp (D)	19(4.9)	17 (4.8)	18 (4.6)
Cys (C)	4 (1.0)	1 (0.2)	3 (0.8)
Gln (Q)	14 (3.6)	4 (1.0)	14 (3.6)
Glu (E)	25 (6.4)	32 (8.3)	31(8.0)
Gly (G)	34 (8.9)	34 (8.8)	35 (9.0)
His (H)	8 (2.1)	10 (2.6)	9 (2.3)
Ile (I)	36 (9.3)	36 (9.4)	32 (8.2)
Leu (L)	25 (6.4)	32 (8.3)	25 (6.4)
Lys (K)	33 (8.5)	37 (9.6)	30 (7.7)
Met (M)	6 (1.6)	5 (1.3)	10 (2.6)
Phe (F)	15 (3.8)	13 (3.4)	14 (3.6)
Pro (P)	14 (3.6)	13 (3.4)	14 (3.6)
Ser (S)	24 (6.2)	23 (6.0)	25 (6.4)
Thr (T)	13 (3.4)	25 (6.5)	18 (4.6)
Trp (W)	7(1.8)	4 (1.0)	5 (1.3)
Tyr (Y)	11 (2.8)	18 (4.7)	18 (4.6)
Val (V)	34 (8.9)	31 (8.1)	30 (7.7)

*Fn*ADH, ADH from thermophile *Fervidobacterium nodosum*; *Ao*ADH, ADH from mesophile *Alkaliphilus oremlandii*. Numbers in parenthesis**,** percentage of the amino acid; numbers highlighted in grey, amino acids that are obviously less in thermophilic/hyper thermophilic ADH than mesophilic homologous; numbers highlighted in yellow, amino acids that are obviously more in thermophilic/hyper thermophilic ADH than mesophilic homologous.

Table 3-5 Comparison of the typical amino acids between *Th*ADH and its homologous

Amino acid	No. of residues in *Th*ADH	No. of residues in *Fn*ADH	No. of residues in *Ao*ADH	Δ ThADH & *Ao*ADH	Δ FnADH & *Ao*ADH
Aromatic					
Phe	15	13	14	+1	-1
Hydrophobic					
Val	34	31	30	+4	+1
Leu	25	32	25	0	+7
Met	6	5	10	-4	-5
Small hydrophobic					
Ala	38	27	31	+7	-4
Pro	14	13	14	0	-1
Gly	34	34	35	+1	-1
Charged					
Glu	25	32	31	-6	+1
Lys	33	37	30	+3	+7
Hydrophilic					
Ser	24	23	25	-1	-2
Asn	16	16	19	-3	-3
Gln	14	4	14	0	-10
Total	387	385	388		

*Fn*ADH, ADH from thermophile *Fervidobacterium nodosum*; *Ao*ADH, ADH from mesophile *Alkaliphilus oremlandii*. Highlighted in grey, amino acids that are obviously less in thermophilic and hyperthermophilic ADH than mesophilic homologous, highlighted in yellow, amino acids that are obviously more in thermophilic and hyperthermophilic ADH than mesophilic homologous.



(a) (b)

Figure 3-6 Predicted tertiary structure of *Th*ADH monomer and the putative iron-binding site

The structure modeling was run on the Swiss Model server using an iron-containing ADH from *T. maritima* (TM0820; PDB number: 1vljB) as the template. (a) Residues in red, the putative NADP-binding site; residues in blue, the putative iron-binding site. (b) The vertical view of the putative iron-binding site, $Asp_{195}His_{199}His_{268}His_{282}$.

3.4.3 Construction of the expression vector

Specific primers with the over-hanged restriction sites were designed as THNCNF and THECCR (**Table 3-2**). The restriction enzymes *Nco*I and *Eco*RI were selected since they were not involved in the structure gene (**Fig. 3-7**). The PCR products showed a specific 1.2 kb band on 1% agarose gel, and the DNA was extracted and confirmed by sequencing. The PCR amplified *Th*ADH coding gene was inserted in pET-30a over-expression vector after double enzyme digestion by *Nco*I and *Eco*RI (**Fig. 3-8**). The gene was inserted under the control of T7-*lac* promoter. The recombinant plasmid was selected from the colonies grew on 2YT agar with 50 mg/ml kanamycine by both colony PCR and restriction enzymes digestion (**Fig. 3-9**).

3.4.4 Over-expression of the *Th*ADH in *E. coli*

E. coli was selected as heterlogous expression host in this research. However, *Th*ADH coding gene preferred a different codon usage compared to *E. coli*. Analyzed by Graphical codon usage analyzer (Fuhrmann et al., 2004), the mean difference was 39.98% (**Fig. 3-10**). To overcome the possible poor yield caused by codon bias, confirmed by sequencing, the isolated recombinant plasmids carrying *Th*ADH coding gene were transformed into *E. coli* BL 21-Rossetta expression strains, containing the extra plasmid for rarely used tRNAs codons AGA/AGG/AUA/CUA/GGA/CCC/CGG to rescue the poor expression by codon bias mainly caused by the rare tRNA in *E. coli*: AGG/AGA for arginine, AUA for isoleucine, and CUC for leucine. From 10% SDS-

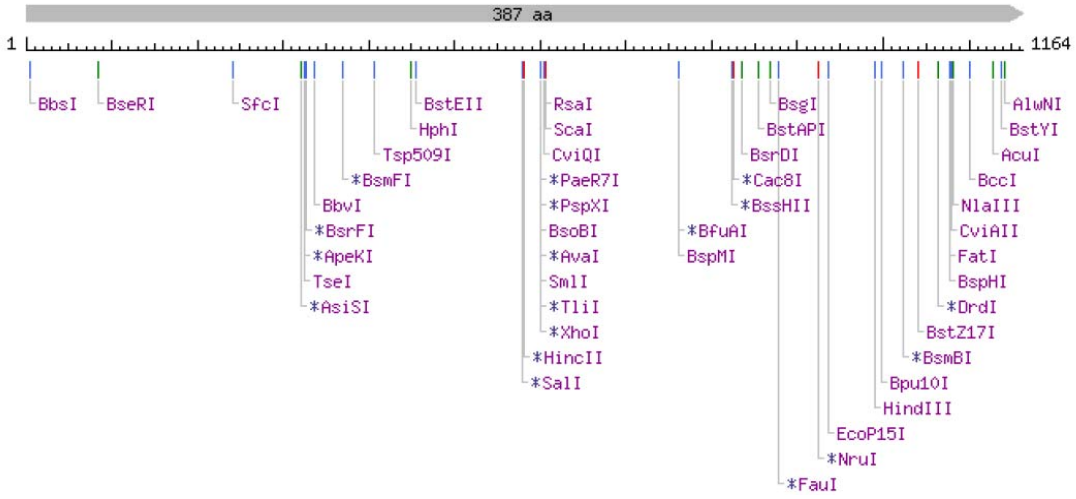


Figure 3-7 Restriction map of the *Th*ADH coding gene

On-line analysis was conducted using http://tools.neb.com/NEBcutter2/index.php; red, blunt end cut; blue, 5' extension; green, 3' extension; *, cleavage affected by CpG methylation; #, cleavage affected by other methylation.

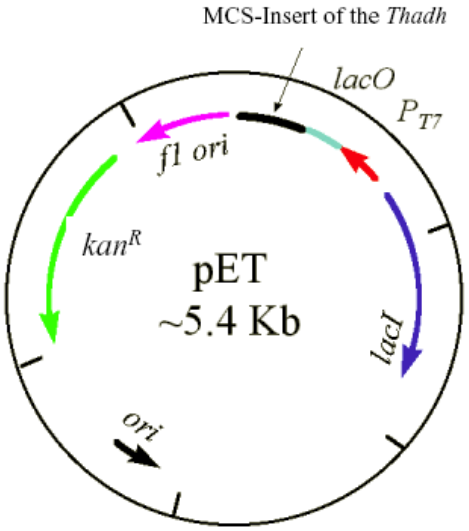


Figure 3-8 Recombinant plasmid construction and the location of inserted gene

Figure was modified from http://www.bio.davidson.edu. Major elements constructed in the pET vector are indicated using color arrows. Black arrow, location of the inserted *Th*ADH encoding gene.

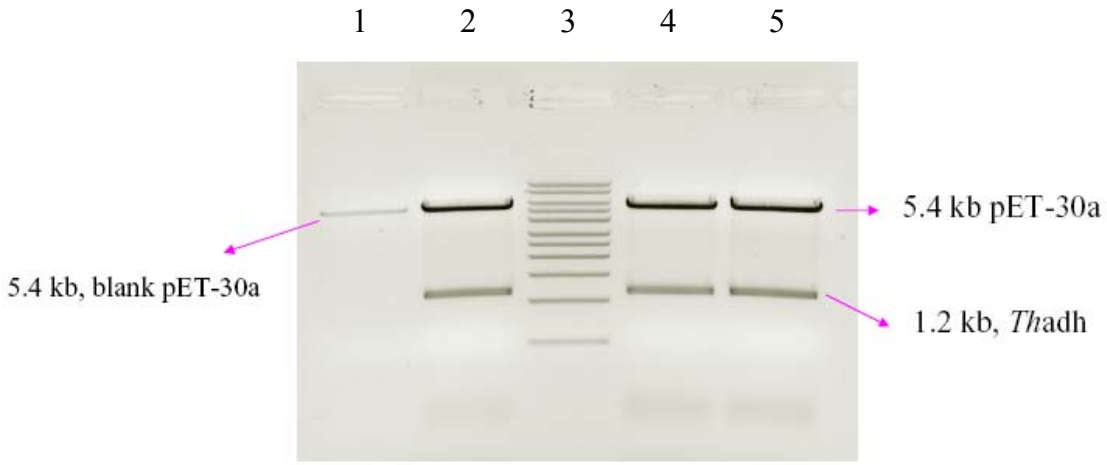


Figure 3-9 Selection of the recombinant vector carrying the gene *Thadh* by enzyme digestion

The arrows and notes describe behavior of recombinant vector and blank pET-30a vector after enzyme digestion by *Nco*I and *Eco*RI. The 1.2 kb bands were released insert gene from recombinant vectors after enzyme digestion. Lane 1, blank pET-30a vector; lane 2, recombinant plasmid *Thadh*-pET-30a; lane 3, 1 kb DNA ladders (Novogan, WI, USA); lane 4, recombinant plasmid *Thadh*-pET-30a; lane 5, recombinant plasmid *Thadh*-pET-30a.

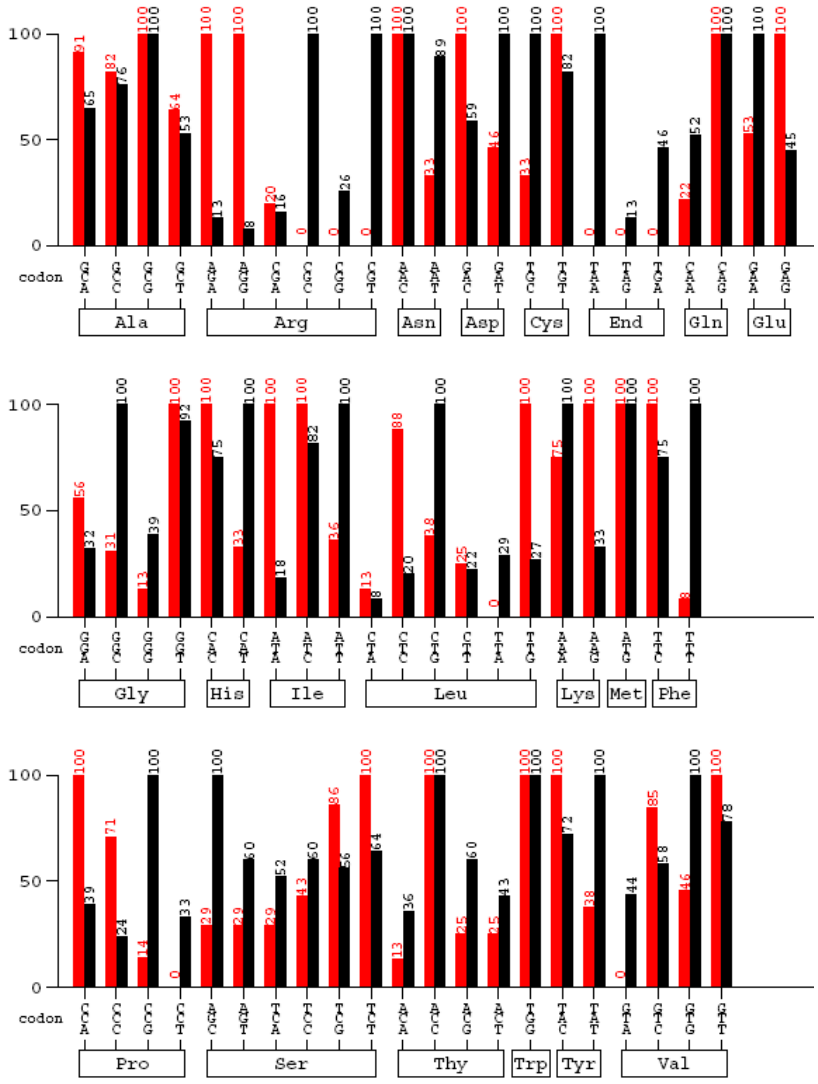


Figure 3-10 Codon usage pattern comparison of *T. hypogea* and *E. coli*

Codon usage analysis was conducted using Graphical codon usage analyzer at www.gcua.de (Fuhrmann et al., 2004). Columns in red, codon bias pattern of sequence derived from *T. hypogea*; Columns in black, codon bias pattern of *E. coli.* The numbers of most frequenctly used codons for the specific amino acid was defined as 100.

PAGE, the recombinant enzyme around 43 kDa was expressed at a high level with the presence of IPTG as inducer (**Fig. 3-11**).

3.4.5 Optimum cultivation condition

E. coli strains carrying the recombinant vector *Tgadh*-pET30a were incubated in 2YT medium with 50 mg/ml kanamycine and 1mM ferrous to an OD_{600} 0.8 before induction, which took 3.5 to 4 hours. The optimal yield of the recombinant enzyme was obtained when the concentration of inducer IPTG was 0.4 to 0.6 mM from SDS-PAGE (**Fig. 3-11)**. Therefore, the concentration of IPTG was chosen to be 0.4 mM.

3.4.6 Purification of the recombinant *T. hypogea* ADH from *E. coli*

Considering the thermostability of *Th*ADH, the recombinant enzyme was purified from *E. coli* using a simplified procedure. Heat treatment was applied to the cell extract prior to liquid chromatography. Heating for 30 minutes at 60 ^{0}C significantly reduced 60% of the total proteins (**Fig. 3-12**). Subsequently, the recombinant ADH was purified to homogeneity after DEAE-Sepharose column. The purified recombinant *Th*ADH had a specific activity of 69 U/mg compared to 57 U/mg of the native enzyme (Ying et al., 2007) and presented a higher purification yield of 49% than 31 % of the native one (**Table 3-6**). Purified recombinant *Th*ADH was eluted from the gel-filltration column Superdex 200, the corresponding ADH activity was collected from a peak at 192 ml, indicating a molecular mass of 84 ± 5 kDa. The SDS-PAGE analyses showed that recombinant *Th*ADH had a single subunit with a molecular weight estimated to be 43 ± 2 kDa, suggesting the recombinant *Th*ADH was a homodimer (**Fig. 3-13**).

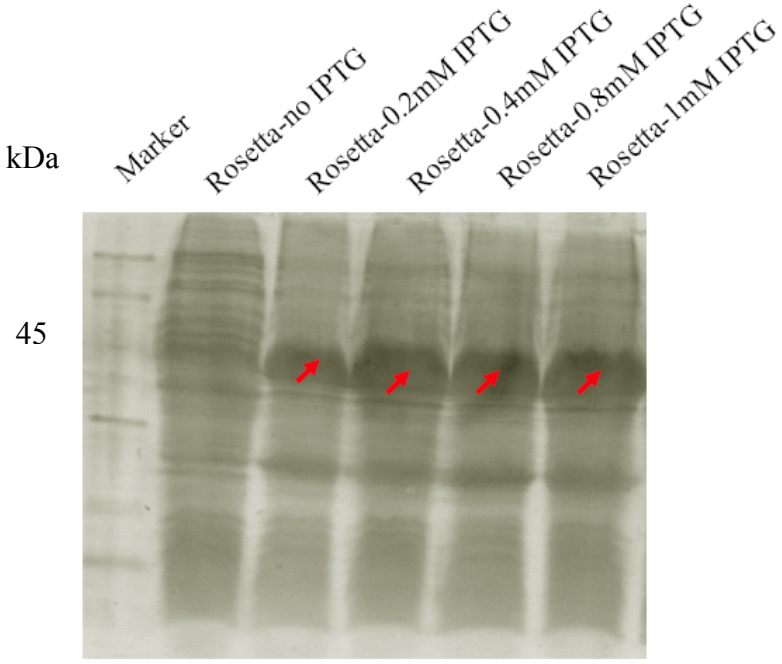


Figure 3-11 10% SDS PAGE for yield of *Th*ADH in *E. coli* induced by IPTG

Left lane, low molecular weight protein marker; other lanes, crude extracts of cells; red arrow, over-expressed *Th*ADH in the total protein of codon-plus *E. coli* host cells. 30 ul cell crude extracts were loaded per lane.

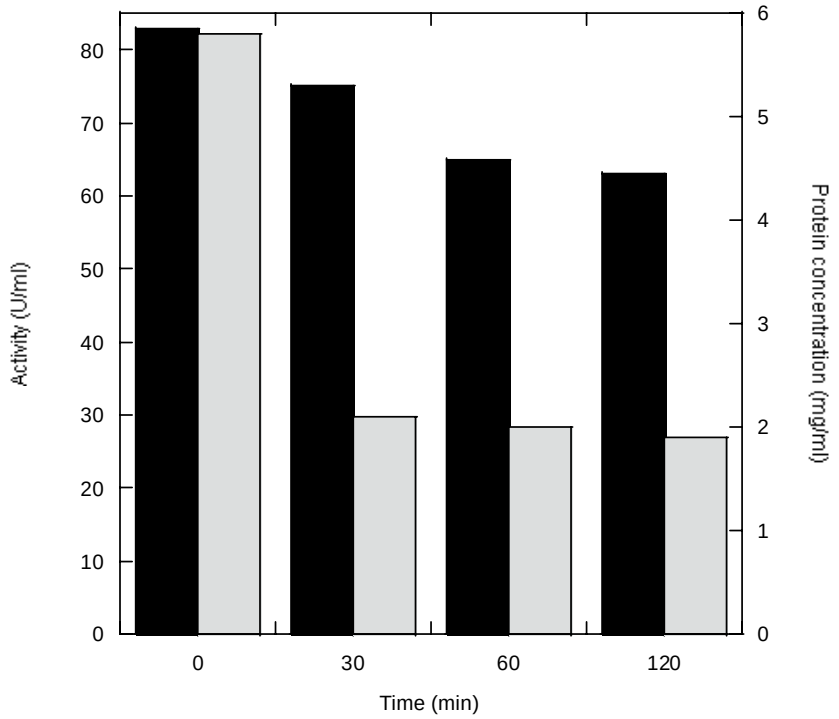


Figure 3-12 Heat treatment of recombinant *E. coli* cell-crude extract

The enzyme activity was measured using standard assay in CAPS buffer (100 mM, pH 11.0). Column in black, residual activity of recombinant *E. coli* cell-crude extract after heat treatment at 60 ^{0}C; column in grey, residual protein concentration recombinant *E. coli* cell-crude extract after heat treatment.

Table 3-6 Purification of the recombinant *Th*ADH from *E. coli*

	Total protein (mg)	Total activity (U)	Specific activity (U/mg)	Purification fold	Yield (%)
Cell-crude extract	240	3552	14.8	1	100
Heat-treatment	62.8	2427.2	38.5	2.6	68
DEAE-Sepharose	25.1	1734.4	69.1	4.7	49

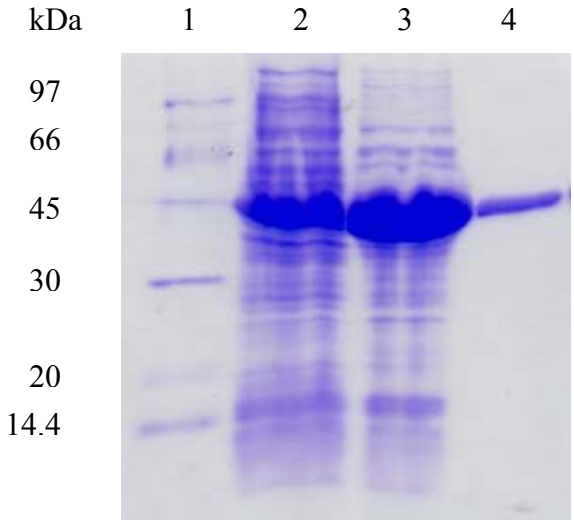


Figure 3-13 10% SDS-PAGE for purified recombinant *Th*ADH

Lane 1, molecular weight marker; Lane 2, cell crude extract; Lane 3, cell crude extract heat treated for 30 min; Lane 4, 2 μg purified recombinant *Th*ADH. In lane 2-3, approximately 10 μg total proteins were loaded per lane.

3.4.7 Catalytic properties and the comparison with native enzyme from *T. hypogea*

Compared to the native ADH purified directly from *T. hypogea*, the recombinant *Th*ADH had very similar catalytic properties to the former. Over a temperature rang from 30 to 95 ^{0}C, the activity of recombinant *Th*ADH increased along with the increase of assay temperature, showing the same trend of the native enzyme (**Fig. 3-14**). The effect of pH on enzyme activities was investigated with a set of 100 mM buffers ranging from pH 6.0 to 12.0. Both the native and recombinant *Th*ADH shared the optimal pH for ethanol oxidation was pH 11.0 while that for butyaldehyde reduction was pH 8.0 (**Fig. 3-15**), additionally, the enzymes worked within quite a small pH range, when the buffer pH value was higher or lower than the optimum value, the activity of ethanol oxidation had a remarkable decreasing. Similar to the native enzyme, when the thermostability was tested at 70 ^{0}C, there was a 50% of residue activity of the recombinant enzyme after 2 hours in the presence of 2 mM DTT, however, the half life ($t_{1/2}$) of the recombinant enzyme was was detected to be less than one hour at 90 ^{0}C in the absence of DTT (**Fig. 3-16**). Both the native and recombinant enzymes were sensitive to oxygen. There was a residual activity of 50% after exposure to the air at room temperature for only 20 minutes (**Fig. 3-17**).

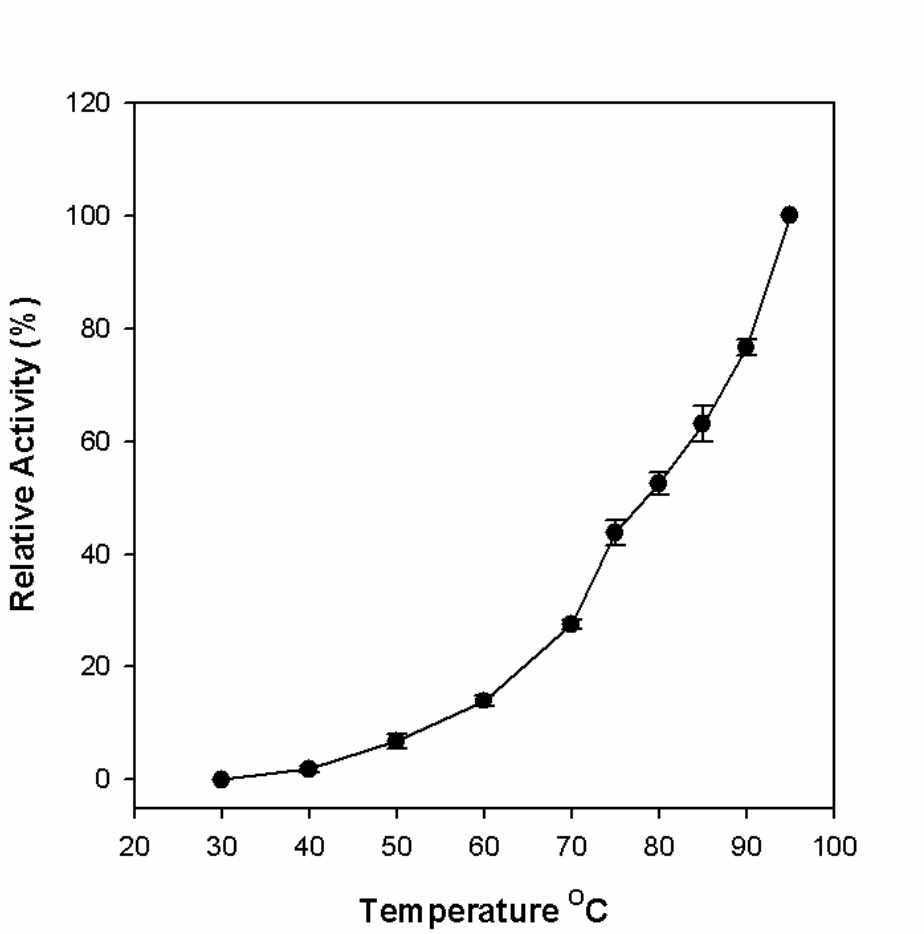


Figure 3-14 Temperature dependence of the purified *Th*ADH

The activities were measured in the standard assay conditions except varying assay temperatures from 30 to 95 ^{0}C. The relative activity 100% was defined as the highest activity value achieved in this test (133 U/mg at 95 ^{0}C). Error bars indicate standard deviations of the measurements.

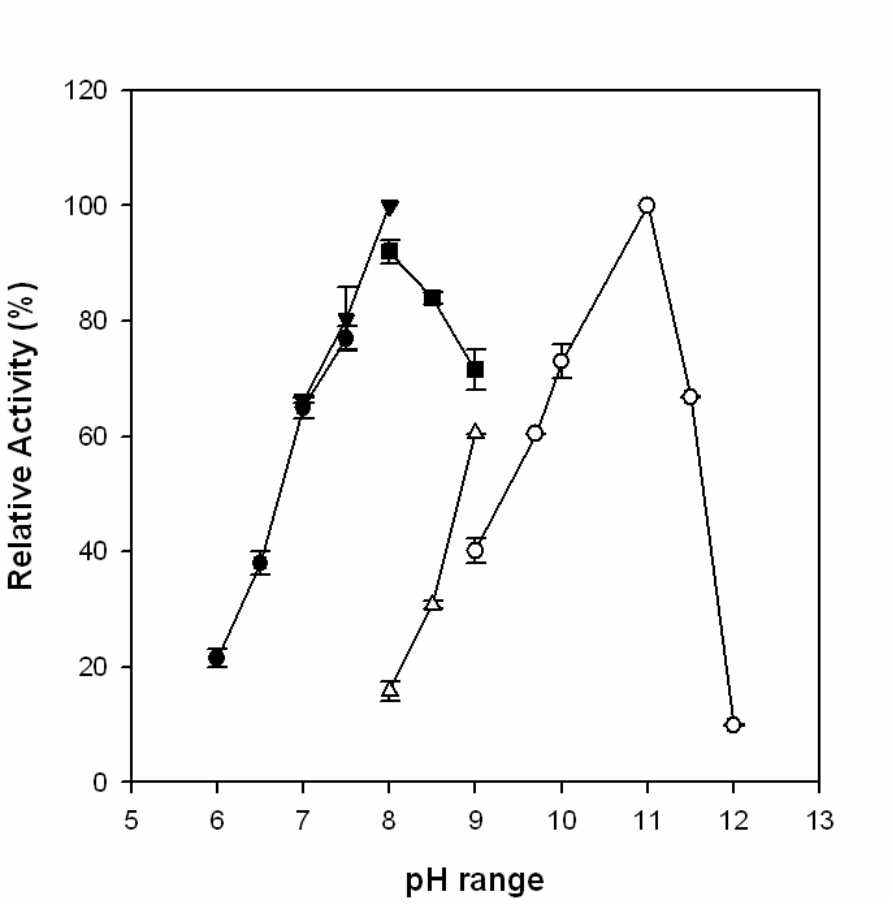


Figure 3-15 Optimal pHs of the purified recombinant *Th*ADH

Optimal pHs for alcohol oxidation and formation were determined by measuring the activities on the oxidation of 1-butanol (unfilled points) and the reduction of butyaldehyde (filled points), respectively. The buffers (100 mM) used were CAPS (unfilled circles), Tirs/HCl (unfilled triangles and unfilled squares), HEPES (filled converted triangles) and PIPES (filled circles). The relative activity of 100% refers to full activity of the recombinant enzyme that equals 69 U/mg of alcohol oxidation activity and means 51 U/mg for of aldehyde reduction, respectively. Error bars indicate standard deviations of the measurements.

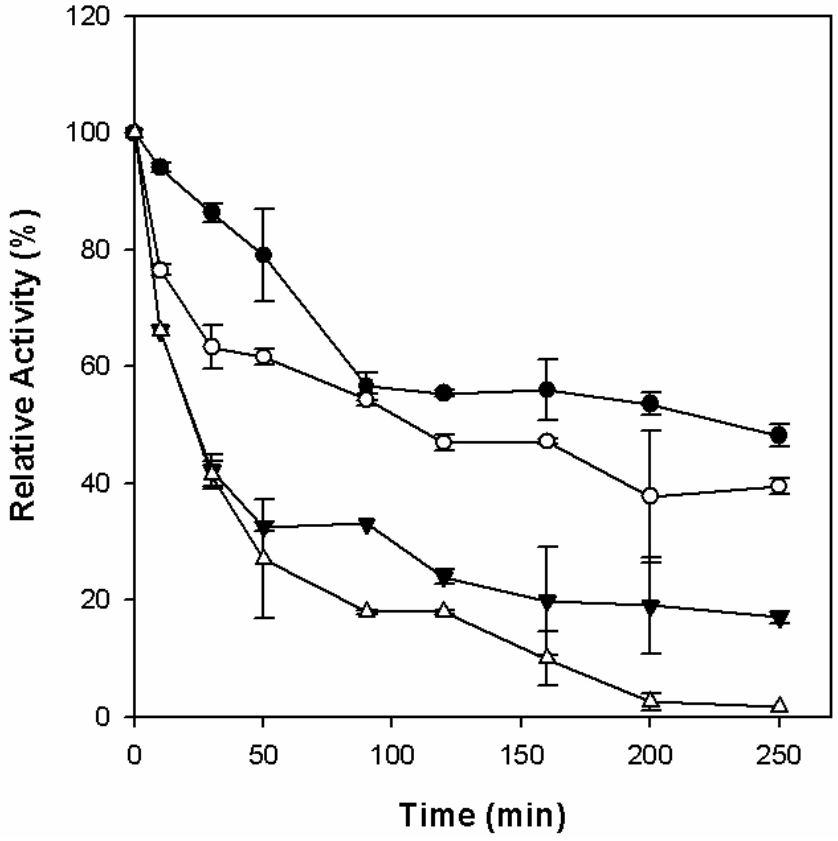


Figure 3-16 Thermostability of the purified recombinant *Th*ADH

The enzyme activity was assayed by 1-butanol oxidation in the CAPS buffer (pH 11.0) at 80 ^{0}C. Filled circles, in the presence of 2 mM DTT at 70°C; open circles, in the absence of DTT at 70 ^{0}C; filled triangle, in the presence of 2 mM DTT at 90 ^{0}C; open triangle, in the absence of DTT at 90 ^{0}C. The relative activity of 100% equals to the initial ADH activity without the heat treatment (62 U/mg). Error bars indicate standard deviations of the measurements.

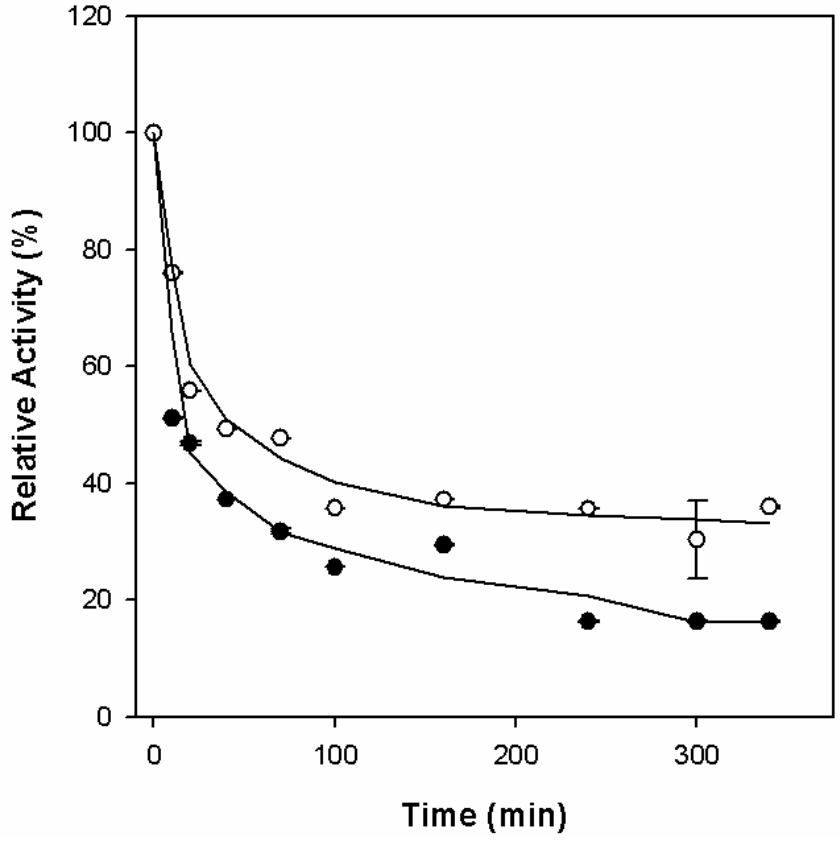


Figure 3-17 Oxygen sensitivity of the purified recombinant *Th*ADH

The enzyme activity was measured by 1-butanol oxidation in the CAPS buffer (pH 11.0) at 80 ^{0}C. Open circles, in the presence of 2 mM DTT and 2 mM SDT; filled circles, in the absence of 2 mM DTT and 2 mM SDT. The relative activity of 100% equals to the ADH activity prior to exposure to air (62 U/mg). Error bars indicate standard deviations of the measurements.

3.5 DISCUSSIONS

Alcohol dehydrogenases are ubiquitous enzymes catalyzing interconversions between alcohols and corresponding ketones or aldehydes. Classified by the metal contained, ADHs are divided into 3 groups. Group I contains long-chain ADHs whose size varies from approximately 350 to 900 amino acid residues and zinc is at their catalytic site and, sometimes, the enzyme also has structural zinc (Littlechild et al., 2004). Group II contains short-chain ADHs with approximately 250 amino acid residues and lacks metals (Reid et al., 1994). Group III only consists of a small number of iron-dependent ADHs. The three groups of ADHs are present in three domains of life: archaea, bacteria and eukarya. There are only a few iron-containing ADHs characterized from hyperthermophiles, though many ADHs from mesophiles, especially those from horse liver, yeast and *E. coli* are well studied. Up to date, only four iron-containing ADHs from hyperthermophiles have been characterized; they are ADHs from hyperthermophilic archaea: *T. hydrothermalis, Thermococcus strain ES-1, Thermococcus litoralis* and *Pyrococcus furiosus*. *T. hypogea* ADH represents the first hyperthermophilic bacterial ADH that contains iron with full activity after purification, whose catalytic properties show similarities to the enzymes in archaea (Ma et al., 1995; Ying et al., 2007).

The sequencing of the entire encoding gene of *Th*ADH provided valuable insights on understanding the evolutionary relationship between this iron-containing ADH and its homologous enzymes (**Fig. 3-18**). The amino acids sequence (387 aa) of *T. hypogea* ADH has similarity of higher than 70% to those within *Thermotoga* species such as *T.*

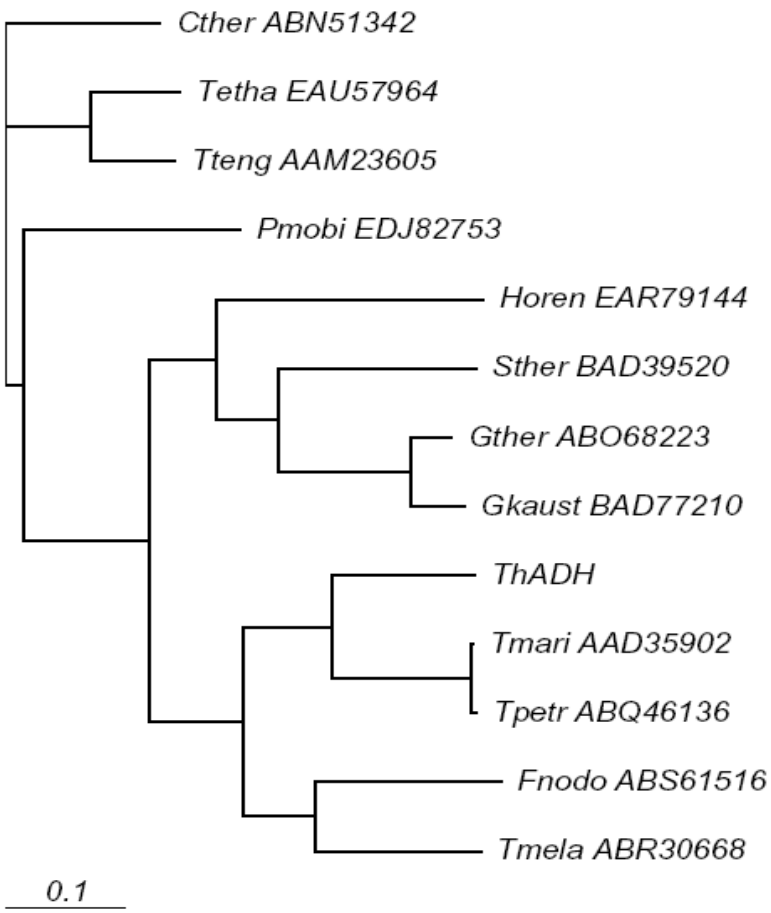


Figure 3-18 Phylogenetic relationships of *Th*ADH and related iron-containing ADHs from bacterial hyper/thermophiles

The sequences were aligned using Clustal W (Thompson et al., 1994). Scale bar indicates 0.1 substitutions per sequence position. The numbers following the enzymes are the accession numbers for the genes encoding ADHs. Tetha, *Thermoanaerobacter ethanolicus* X514; Tteng, *Thermoanaerobacter tengcongensis* MB4; Cther, *Clostridium thermocellum* ATCC 27405; Pmobi, *Petrotoga mobilis* SJ95; Gther, *Geobacillus thermodenitrificans* NG80-2; Gkaust, *Geobacillus kaustophilus* HTA426; Sther, *Symbiobacterium thermophilum* IAM 14863; Horen, *Halothermothrix orenii* H 168; Tmari, *Thermotoga maritima*; Tpetr, *Thermotoga petrophila*; ThADH, ADH from *T. hypogea*; Fnodo, *Fervidobacterium nodosum* Rt17-B1; Tmela, *Thermosipho melanesiensis* BI429.

maritima and *T. petrophila*. And it carries moderately high similarity (50-70%) to those in other bacterial thermophiles or mesophililes, including *Thermoanaerobacter ethanolicus* and *Fervidobacterium nodosum*, and low similarity to ADHs from archaeal hyperthermophiles, The homologues of *T. hypogea* ADH were abundant in bacteria but not archaea, which could be an indication of the divergence of iron-containing ADHs from hyper/thermophiles.

Analysis of both primary and predicted 3-D structure of the enzyme revealed its characteristics of iron-containing ADH. Though sharing a similarity of less than 50% in primary structure, *Th*ADH typically had two domains of three-demonsional structure separated by a cleft where the active site of the enzyme might be situated, similar to that of the *Tg*ADH. The cleft catalytic pockets are commonly found in the metalloenzyme when cofactors are involved the catalytic reactions (Schwarzenbacher et al., 2004). The *Th*ADH consisted of the N-terminal domain formed by an α/β region where contained a $Gly_{39}Gly_{40}Gly_{41}Ser_{42}$ motif, well accepted to be involved in the interactions with the pyrophosphate groups of NADP (Sulzenbacher et al., 2004). On the other hand, the C-terminal part was an all-helical domain responsible for the iron binding. The putative active site motif was identified to be $Asp_{195}His_{199}His_{268}His_{282}$.

The catalytic parameters of the recombinant enzyme were similar to the native form, indicating that this iron-containing ADH was successfully produced in *E. coli* and the heterologous expression in *E. coli* can potentially be used for the large-scale production of this recombinant iron-containing hyperthermophilic ADH. The recombinant enzyme

was periplasmic and soluble, and no inclusion body was formed. The activity of recombinant *Th*ADH from *E. coli* did not require heat activation. However, heat activation was essential to obtain maximally active forms of the iron-containing ADH from *T. hydrothermalis* and 2, 3-butanediol ADH from *P. furiosus* (Antoine et al., 1999; Kube et al., 2006). The activity of the recombinant *T. hydrothermalis* ADH increased 10-25% after 1 minute incubation at 80 ^{0}C while *P. furiosus* ADH was inactive without heat treatment and the highest activity was obtained after 10 minutes incubation at 100 ^{0}C. Recombinant ES1 ADH activity was unaffected by heat treatment of the cell-free extract, similar to what was observed with the recombinant short-chain ADH from *P. furiosus* (van der Oost et al., 2001). Similar to *Tg*ADH, the optimal pHs of both native and recombinant *Th*ADH on the oxidation of alcohols are more alkaline than those on the reduction of aldehydes. One of the reasonable explanations would be the enzyme activity of *Th*ADH relies on the proton relay mechanism, and the pH dependence of the proton abstraction is correlated with a reorganization of specific conformation in the active site (Koumanov et al., 2003).

Both native and recombinant *Th*ADHs were quite oxygen-sensitive, which could be a disadvantage of the enzyme, and construction of a more oxygen tolerant enzyme is the focus of further study. The loss of activity after exposure to the air might be due to the oxidation of ferrous to ferric, and/or loss of ferrous ion, and replacement with other metal ion such as zinc (Ying et al., 2007); however, the structural basis underlying this behavior is not clear yet. Generally, iron-sulfur centers are the oxidation-sensitive sites in several metalloenzymes (Unden et al., 1994), but it is not present in *Th*ADH. Iron substitution by

divalent metal ions that could not be further oxidized (e.g. Zn^{2+}, Co^{2+} etc.) at the catalytic site may provide further mechanistic insights into the nature of oxygen sensitivity. For example, the cobalt-substituted *T. brockii* ADH exhibited an increase in specific activity compared to the native enzyme resulting from higher metal-ligand coordination in the catalytic site (Kleifeld et al., 2004). In addition, irreversible loss of enzyme activity was also likely to be caused by irreversible structure change due to the oxidation of some amino acids such as cysteine residues (Neale et al., 1986). Site-direct mutagenesis worked as a powerful approach that induced changes of the oxygen sensitivity in some iron-containing oxidoreductases. For instance, site-direct mutations of residues near the NAD-binding consensus amino acid sequence (Ile7Leu and Leu8Val) of an oxygen sensitive iron containing propanediol oxidoreductase (FucO protein) in *E. coli* increases resistance to oxidative stress (Lu et al., 1998), however, thermal stability of the mutant enzyme decreased simultaneously. The previous mutagenesis work shed light on the protein engineering of *Th*ADH that could help to clarify the mechanism of oxygen tolerance as well as the relationship between structure and function.

Chapter 4 General Conclusions

Alcohol dehydrogenases are ubiquitous enzymes catalyzing inter-conversions between alcohols and corresponding ketones or aldehydes. ADHs from hyperthermophiles have attracted interests in both fundamental studies and the exploitation of their application potentials because of their outstanding tolerance to high temperatures. This research work aimed at cloning and expression of thermostable alcohol hedydrogenases from anaerobic hyperthermophiles followed by characterization of the recombinant enzyme as well as exploration of their biotechnological potentials.

4.1 Cloning and molecular characterization of ADHs

Both *Tg*ADH and *Th*ADH encoding genes were amplified directly from corresponding genomic DNAs by PCR. Since genome sequences are not available, designing of specific primers for gene amplification remained a challenge. The initial nucleotide sequences of the two genes were amplified using degenerated primers that were designed based on amino acids fragments obtained by Edman degradation and mass spectrometry analysis of the ADHs purified directly from *T. guaymasensis* and *T. hypogea*. When fragments of nucleotide sequences were sequenced, specific primers were designed based on the known sequences, which increased the chance to amplify the target gene. For cloning of the entire *Tg*ADH-encoding gene, inverse PCR was applied because of relatively low identity between *Tg*ADH encoding gene and its homologues. However, detection of DNA fragments carrying the target gene after inverse PCR needed lots of efforts.

Southern hybridization using known nucleic acids as probe would be an effective way to select target fragments.

The 1092 bp *Tg*ADH structural gene encoded a 364 amino acids polypeptide, which was identified to be Zn-binding ADH. *Tg*ADH encoding gene ended with two stop codons TGA and TAA and followed by one of the putative archaeal terminator sequence, which suggested the gene behavior like an independent one but not involved in an operon. Compared to its thermophilic and mesophilic counterparts, hyperthermophilic *Tg*ADH has the amino acids composition of more Ala, Arg, Glu, Pro but less Asn, Ser, Met and Ile, which strengthened ion pairs and might be important determinants of thermostability. The nucleotide sequence analyses of *Tg*ADH showed a different pattern of codon usage from that of *E. coli*.

The *Th*ADH encoding gene was 1164 bp corresponding to 387 amino acids sequence, which contains the conserved domains indicating an uncharacterized iron-containing ADH. In contrast to *Tg*ADH, there is one stop codon TGA. As only a few downstream nucleic acids obtained now and no typical terminal sequence was found, it is not discreet to conclude if *Th*ADH is involved in an operon. While the gene arrangement in the genome of *T. maritima* indicated the encoding gene of TM0920 that have 76% identities with *Th*ADH was not in the operon, so *Th*ADH encoding gene could be independent. Similar to *Tg*ADH, *Th*ADH has higher ratio for Ala, Arg, Lys, Thr but lower ratio for Asn and Met as compared to the thermophilic and mesophilic counterparts. Also, the

nucleotide sequence analyses of *Th*ADH showed the pattern of its codon usage different from that of *E. coli*.

4.2 Heterologous expression of ADHs

*Tg*ADH and *Th*ADH were over-expressed in the mesophilic host *E. coli* using pET-30a vector that contains strong promoter T7 frequently selected for the expression of extremophilic proteins. In both cases, the expression of recombinant enzymes was induced by IPTG. The recombinant enzymes were soluble, and no inclusion body produced. The activity of recombinant ADHs from *E. coli* did not require heat activation, although heat activation was necessary for achiving the maximum activity of some recombinant iron-containg ADHs, including *T. hydrothermalis* ADH and *P. furiosus* 2, 3-butanediol ADH (Antoine et al., 1999; Kube et al., 2006). Purification of the recombinant enzymes was simplified by heat treatment of the cell crude extract and the purified enzymes retained biochemical properties of the native enzymes, such as optimal activity at high temperatures and thermostability, indicating *E. coli* is an effective heterologous expression system for enzymes from anaerobic hyperthermophiles, and under proper cultivation conditions, enzymes from hyperthermophiles could fold correctly in mesophilic host.

4.3 Biochemical and biophysical properties of ADHs

Both the recombinant *Tg*ADH and *Th*ADH have highly similar biochemical properties to characterized native ADHs from *T. guaymasensis* or *T. hypogea*. The recombinant ADH enzymes have full activity of the native forms in alcohol oxidation and aldehydes reduction. Both *Tg*ADH and *Th*ADH show that their optimal temperatures are above 95°C and stay stable after treatment at high temperature (>80 ^{0}C). *Tg*ADH harbors the outstanding thermostability and remain 70% of the original activity when incubated at 80 ^{0}C for 24 hours. Generally, pH dependence of enzymatic activity reflects the ionization of groups involved in the catalysis. The optimal pHs of both *Tg*ADH and *Th*ADH on the oxidation of alcohols are more alkaline than those on the reduction of aldehydes or ketones, which is a common feature among hyperthermophilic ADHs. Additionally, both *Th*ADH and *Tg*ADH are sensitive to oxygen. The oxygen inactivation of the iron-containing *Th*ADH seems to be due to the oxidation of iron atom, while the oxygen-inactivation mechanism of *Tg*ADH has not been well understood yet, which may be associated with oxygen damage on amino acid residues.

4.4 Structural properties of ADHs

Though sharing a similarity of less than 50% in primary structure, both *Tg*ADH and *Th*ADH typically had two domains of three-demonsional structure separated by a cleft where the active site of the enzyme might be situated. The active sites centered on the catalytic ions were dimensionally close to the pocket containing the cofactor NADP indicating the cofactor binding is essential for the catalysis. The *Tg*ADH amino acid sequence shows high similarity to zinc-containing ADHs from hyperthermophiles,

especially those from *Thermoanaerobacter* species. The conserved domain analyses revealed that it belongs to zinc-containing ADHs and the enzyme harboured only catalytic zinc and coenzyme NADP binding sites. Two typical domains located in 3-D structure of *Tg*ADH, the catalytic zinc-binding site ($GHEX_2GX_5GX_2V$, residues 62-76) close to N-terminal and one cofactor NADP binding site ($GXGX_2G$, residues 183-188) close to C-terminal end.

On the other hand, the amino acid sequence of *Th*ADH has high similarity with the iron-containing ADH from *Thermotoga* species. The *Th*ADH consists of two cleft typical domains as well: the N-terminal domain formed by the α/β region which contains $Gly_{39}Gly_{40}Gly_{41}Ser_{42}$ motif, predicted to be conserved NADP binding sites; and the C-terminal part is an all-helical domain responsible for the iron binding, where the putative active site motif $Asp_{195}His_{199}His_{268}His_{282}$ is located.

4.5 Relationship between Zinc-and iron-containing ADHs

The zinc-containing ADH from *T. guaymesensis* and the iron-containing ADH from *T. hypogea* belong to NADP-dependent ADH and they shared some properties: divalent metal ion is necessary for the activity, 3-D structure consists of two cleft typical domains locating coenzyme binding site and catalytic core respectively. However, the two enzymes represent very different biochemical properties due to different divalent metal in the catalytic core. Containing stable zinc ion, the *Tg*ADH was present more stable either at high temperatures or exposure to oxygen than *Th*ADH containing ferrous that is

sensitive to oxidation. From evolutionary point of view, Fe^{2+}-containing ADHs have been proposed to evolve earlier than Zn^{2+}-containing ADHs. It is well accepted that the original global environment was anaerobic and highly reducing and it was possible the supply of ferrous iron was abundant. Fe^{2+}-containing ADHs commonly function in anaerobiosis and such enzymes are proposed to evolve earlier than zinc-containing ADHs (Lu et al., 1998). Later, iron became mostly sequestered in the Fe^{3+}-containing compounds when oxygen accumulated in the earth atmosphere. It is thus logic to speculate that more oxygen-resistible Zn^{2+}-containing ADHs gradually supplanted Fe^{2+}-containing ADHs. Fe^{2+}-containing ADHs persist either because there is no selective pressure such as environments lacking of oxygen or because they play a role in the shift from anaerobic to aerobic metabolism (Lu et al., 1998).

4.6 Future outlooks

ADHs are important for fundamentally scientific studies as well as exploration of their potential in biotechnological applications. *Tg*ADH has the extremely high specific activity (1079 U/mg on the oxidation of 2-butanol), thermostability, stereoselectivity, broad substrate specificity to meet the requirements of good biocatalysts in industry. However, the coenzyme of *Tg*ADH--NADP is commercially expensive, which limits its industrial application to some extent, therefore it is necessary to alter properties of the enzyme, enabling it to transform new substrates or catalyze existing substrates more efficiently associated with a relevantly cheap coenzyme. On the other hand, oxygen sensitivity limited application of *Th*ADH, and construction of a more oxygen tolerant enzyme is the focus of further study. Since protein lability is caused by oxidation of Fe^{2+},

iron substitution by divalent metal ions that could not be further oxidized (e.g. Zn^{2+}, Co^{2+} etc.) at the catalytic site may provide further mechanistic insights into the nature of oxygen sensitivity. Moreover, protein engineering is a powerful approach to alter catalytic properties of enzymes. Site-directed mutagenesis and direct evolution are most commonly used approaches for protein engineering. In addition, site-directed mutagenesis is also widely used for identifying the role of specific amino acid in catalysis. Generally, the 3-D structural information of enzymes is critical to understand the structure-function relationships for enzymes. The successful production of recombinant enzymes provides sufficient amount of active enzymes for crystallography study that will shed light on the design of site-directed mutagenesis.

References

Adams MWW, Perler FB, Kelly RM. (1995). Extremozymes: expanding the limits of biocatalysis. Nature.Biotech. 13: 662-668

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25: 3389-3402

Ammendola S, Raia CA, Caruso C, Camardella L, D'Auria S, De Rosa M, Rossi M. (1992). Thermostable NAD^+-dependent alcohol dehydrogenase from *Sulfolobus solfataricus*: gene and protein sequence determination and relationship to other alcohol dehydrogenases. Biochemistry 31: 12514-12523

Arnone MI. (1997). Stability of aspartate aminotransferase from *Sulfolobus solfataricus*. Protein Engin. 10: 237-248

Antoine E, Rolland JL, Raffin JP, Dietrich J. (1999). Cloning and over-expression in *Escherichia coli* of the gene encoding NADPH group III alcohol dehydrogenase from *Thermococcus hydrothermalis*. Eur J Biochem. 264: 880-889

Balk M, Weijma J, Stams AJM. (2002). *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. Int J Syst Evol Microbiol. 52: 1361-1368

Baneyx F. (2004). Protein expression technologies: current status and future trends. Horizon Bioscience. Chapter 1: 3-47

Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS. (1979). Methanogens: reevaluation of a unique biological group. Microbiol Rev. 43: 260-296

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. (2002). The Pfam protein families database. Nuc Acids Res 30: 276-280

Bell SD and Jackson SP. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol. 6: 222-228

Benkel BF and Fong Y. (1996). Long range-inverse PCR (LR-IPCR): extending the useful range of inverse PCR. Genet Anal Biomol Engin. 13: 123-127

Biller KF, Kato I, Märkl H. (2002). Effect of glucose, maltose, soluble starch, and CO_2 on the growth of the hyperthermophilic archaeon *Pyrococcus furiosus*. Extremophiles 6: 161-166

Birnboim HC and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7: 1513-1523

Bogin O, Peretz M, Burstein Y. (1997). *Thermoanaerobacter brockii* alcohol dehydrogenase: characterization of the active site metal and its ligand amino acids. Protein Sci. 6: 450-458

Bradford MM. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254

Britton KL, Yip KSP, Sedelnikova SE, Stillman TJ, Adams MWW, Ma K, Maeder DL, Robb FT, Tolliday N, Vetriani C, Rice DW, and Baker PJ. (1999). Structure determination of the glutamate dehydrogenase from the hyperthermophile *Thermococcus litoralis* and its comparison with that from *Pyrococcus furiosus*. J Mol Biol. 293: 1121-1132

Burdette DS, Vieille C, and Zeikus JG. (1996). Cloning and expression of the gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. Biochem J. 316: 115-122

Burdetter DS, Secundo F, Phillips RS. (1997). Biophysical and mutagenic analysis of *Thermoanaerobacter ethanolicus* secondary-alcohol dehydrogenase activity and specificity. Biochem J. 326: 717-724

Burland TG. (1999). DNASTAR's Lasergene Sequence Analysis Software. Methods Mol Biol. 132: 71-91

Cambillau C and Claverie JM. (2000). Structural and genomic correlates of hyperthermostability. J Biol Chem. 275: 32383-32386

Canganella F, Jones WJ, Gambacorta A, Antranikian G. (1998). *Thermococcus guaymasensis* sp. nov. and *Thermococcus aggregans* sp. nov., two novel thermophilic archaea isolated from the Guaymas Basin hydrothermal vent site. Int J Syst Bacteriol. 48: 1181-1185

Cannio R, Fiorentino G, Carpinelli P, Rossi M, Bartolucci S. (1996). Cloning and overexpression in *Escherichia coli* of the genes encoding NAD-dependent alcohol dehydrogenase from two *Sulfolobus* species. J Bacteriol. 178: 301-305

Cavagnero S, Debe DA, Zhou ZH, Adams MWW, Chan SI. (1998). Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry. 37: 3369-3376

Chakravarty S and Varadarajan R. (2000). Elucidation of determinants of protein stability through genome sequence analysis. FEBS Lett. 470: 65-69

Cohen GN, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Querellou J, Ripp R, Thierry JC, van der Oost J, Weissenbach J, Zivanovic Y, Forterre P.

(2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol Microbiol. 47: 1495-1512

Contursi P, Cannio R, Prato S, Fiorentino G, Rossi M, Bartolucci S. (2003). Development of a genetic system for hyperthermophilic archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS Microbiol Lett. 218: 115-120

Conway T, Sewell GW, Osman YA, Ingram LO. (1987). Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. J Bacteriol. 169: 2591-2597

Copeland A, Lucas S, Lapidus A, Barry K, Glavina del Rio T, Dalin E, Tice H, Bruce D, Pitluck S, Richardson P. (2007-a). Complete sequence of *Thermotoga petrophila* RKU-1. US DOE Joint Genome Institute.

Copeland A, Lucas S, Lapidus A, Barry K, Glavina del Rio T, Dalin E, Tice H, Bruce D, Pitluck S, Richardson P. (2007-b). Complete sequence of *Fervidobacterium nodosum* Rt17-B1. US DOE Joint Genome Institute.

Copeland A, Lucas S, Lapidus A, Barry K, Glavina del Rio T, Dalin E, Tice H, Pitluck S, Chertkov O, Brettin T, Bruce D, Detter JC, Han C, Schmutz J, Larimer F, Land M, Hauser L, Kyrpides N, Mikhailova N, Nelson K, Gogarten JP, Noll K, Richardson P. (2007-c). Complete sequence of *Thermosipho melanesiensis* BI429. US DOE Joint Genome Institute.

Danson MJ, Hough DW, Russell RJ, Taylor GL, Pearl L. (1996). Enzyme thermostability and thermoactivity. Protein Eng. 9: 629-630

Delano WL. (2002). The PyMOL molecular graphics system. Delano Scientific, Palo Alto, CA

Demirjian DC, Morís-Varasa F, Cassidyb CS. (2001). Enzymes from extremophiles. Curr Opin Chem Biol. 5: 144-151

Duffner F, Bertoldo C, Andersen JT, Wagner K and Antranikian G. (2000). A New Thermoactive Pullulanase from *Desulfurococcus mucosus*: Cloning, Sequencing, Purification, and Characterization of the Recombinant Enzyme after Expression in *Bacillus subtilis*. J Bacteriol. 182: 6331-6338

Duine JA, Frank J, Verwiel PE. (1980). Structure and activity of the prosthetic group of methanol dehydrogenase. Eur J Biochem. 108: 187-192

Esposito L, Sica F, Raia CA, Giordano A, Rossi M, Mazzarella L, Zagari A. (2002). Crystal structure of the alcohol dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus* at 1.85 Å resolution. J Mol Biol. 318: 463-477

Facchiano AM, Colonna G, Ragone R. (1998). Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. Protein Eng. 11: 753-760

Fardeau ML, Ollivier B, Patel BKC, Magot M, Thomas P, Rimbault A, Rocchiccioli F, Garcia JL. (1997). *Thermotoga hypogea* sp. nov., a xylanolytic, thermophilic bacterium from an oil-producing well. Int J Syst Bacteriol. 47: 1013-1019

Flaman JM, Frebourg T, Moreau V, Charbonnier F, Martin C, Ishioka C, Friend SH, Iggo R. (1994). A rapid PCR fidelity assay. Nucleic Acids Res. 22: 3259-3260

Fontaine L, Meynial-Salles I, Girbal L, Yang XH, Croux C and Soucaille P. (2002). Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. J Bacteriol. 184: 821-830

Fuhrmann M, Hausherr A, Ferbitz L, Schödl T, Heitzer M, Hegemann P. (2004). Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. Plant Mol Biol. 55: 869-881

Fukami-Kobayashi K, Minezaki Y, Tateno Y and Nishikawa K. (2007). A tree of life based on protein domain organizations. Mol Biol Evol. 24: 1181-1189

Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. 15: 352-363

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. (2005). Protein identification and analysis tools on the ExPASy Server. In: Walker JM (eds). The proteomics protocols handbook. Humana Press, Totowa, NJ, pp: 571-607

Giordano A, Cannio R, La Cara F, Bartolucci S, Rossi M, Raia CA. (1999). Asn249Tyr substitution at the coenzyme binding domain activates *Sulfolobus solfataricus* alcohol dehydrogenase and increases its thermal stability. Biochem. 38: 3043-3054

Grättinger M, Dankesreiter A, Schurig H, Jaenicke R. (1998). Recombinant phosphoglycerate kinase from the hyperthermophilic bacterium Thermotoga maritima: catalytic, spectral and thermodynamic properties. J. Mol Biol. 280: 525-533

Gomes J and Steiner W. (2004). The biocatalytic potential of extremophiles and extremozymes. Biotechnol. 42: 223-235

Guex N, Peitsch MC. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis. 18: 2714-2723

Guy JE, Isupov MN, Littlechild JA. (2003). The structure of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix*. J Mol Biol. 331: 1041-1051

Hanahan D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol. 166: 557-580

Hess M and Antranikian G. (2008). Archaeal alcohol dehydrogenase active at increased temperatures and in the presence of organic solvents. Appl Microbiol Biotechnol. 77: 1003-1013

Higashi N, Fukada H, Ishikawa K. (2005). Kinetic study of thermostable L-threonine dehydrogenase from an archaeon Pyrococcus horikoshii. J Biosci Bioeng. 99: 175-180

Higashi N, Tanimoto K, Nishioka M, Ishikawa K, Taya M. (2008). Investigating a catalytic mechanism of hyperthermophilic L-threonine dehydrogenase from *Pyrococcus horikoshii*. J. Biochem. 144: 77-85

Hirakawa H, Kamiya N, Kawarabayashi Y, Nagamune T. (2004). Properties of an alcohol dehydrogenase from the hyperthermophilic archaeon *Aeropyrum pernix* K1. J Biosci Bioeng. 97: 202-206

Höllrig V, Hollmann F, Kleeb AC, Buehler K, Schmid A. (2008). TADH, the thermostable alcohol dehydrogenase from *Thermus sp*. ATN1: a versatile new biocatalyst for organic synthesis. Appl Microbiol Biotechnol. 81: 263-273

Huber R, Langworthy TA, König H, Thomm M, Woese M, Sleytr UB, Stetter KO. (1986). *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. Arch Microbiol. 144: 324-333

Huber R and Stetter KO. (2001). Discovery of hyperthermophilic microorganisms. Methods Enzymol. 330: 11-24

Ishiwa H, Shibahara-Sone H. (1986). New shuttle vectors for *Escherichia coli* and *Bacillus subtilis* IV. The nucleotide sequence of pHY300PLK and some properties in relation to transformation. Jpn. J. Genet. 61: 515-528

Ishikawa K, Higashi N, Nakamura T, Matsuura T, Nakagawa A. (2007). The first crystal structure of L-threonine dehydrogenase. J Mol Biol. 366: 857-867

Itoh T. (2003). Taxonomy of nonmethanogenic hyperthermophilic and related thermophilic archaea. J Biosci Bioeng. 96: 203-212

Jannasch HW, Huber R, Belkin S, Stetter KO. (1988). *Thermotoga neapolitana* sp. nov. of the extremely thermophilic eubacterial genus *Thermotoga*. Arch Microbiol. 150: 103-104

Jeanthon C, Reysenbach AL, L'Haridon S, Gambacorta A, Pace NR, Glenat P, Prieur D. (1995). *Thermotoga subterranean* sp. nov., a new thermophilic bacterium isolated from a continental oil reservoir. Arch Microbiol. 164: 91-97

Jeon YJ, Fong JCN, Riyanti EI, Neilan BA, Rogers PL, Svenson CJ. (2008). Heterologous expression of the alcohol dehydrogenase (adhI) gene from *Geobacillus thermoglucosidasius* strain M10EXG. J Biotech. 135: 127-133

Karakashev D, Thomsen AB, Angelidaki I. (2007). Anaerobic biotechnological approaches for production of liquid energy carriers from biomass. Biotech Lett. 29: 1005-1012

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Nakamura Y, Robb TF, Horikoshi K, Masuchi Y, Shizuya H, Kikuchi H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. 5: 55-76

Kelly RM and Adams MWW. (1994). Metabolism in hyperthermophilic microorganisms. Antonie Van Leeuwenhoek. 66: 247-270

Kleifeld O, Rulišek L, Bogin O, Frenkel A, Havlas Z, Burstein Y, Sagi I. (2004). Higher metal-ligand coordination in the catalytic site of cobalt-substituted *Thermoanaerobacter brockii* alcohol dehydrogenase lowers the barrier for enzyme catalysis. Biochem. 43: 7151-7161

Knapp S, Karshikoff A, Berndt KD, Christova P, Atanasov B and Ladenstein R. (1996). Thermal unfolding of the DNA-binding protein Sso7d from the hyperthermophile *Sulfolobus solfataricus*. J. Mol. Biol. 264: 1132-1144

Konings FA, Zhong P, Agwara M, Agyingi L, Zekeng L, Achkar JM, Ewane L, Saa, Ze EA, Kinge T, Nyambi PN. (2004). Protease Mutations in HIV-1 Non-B Strains Infecting Drug-Naive Villagers of Cameroon. AIDS Res Human Retrov. 20: 105-109

Kopp J, Schwede T. (2004). The SWISS-MODEL repository of annotated three-dimensional protein structure homology models. Nucl Acids Res. 32: D230-D234

Korkhin Y, Kalb AJ, Peretz M, Bogin O, Burstein Y, Frolow F. (1998). NADP-dependent bacterial alcohol dehydrogenases: crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii*. J Mol Biol. 278: 967-981

Kosjek B, Stampfer W, Pogorevc M, Goessler W, Faber K, Kroutil W. (2004). Purification and characterization of a chemotolerant alcohol dehydrogenase applicable to coupled redox reactions. Biotechnol Bioeng. 86: 55-62

Koumanov A, Benach J, Atrian S, Gonza` lez-Duarte R, Karshikoff A, Ladenstein R. (2003). The catalytic mechanism of drosophila alcohol dehydrogenase: evidence for a proton relay modulated by the coupled ionization of the active site Lysine/Tyrosine pair and a NAD_ribose OH switch. Protein Struc Fun Genet. 51: 289-298

Krahe M, Antranikian G, Märkl H. (1996). Fermentation of extremophilic microorganisms. FEMS Microbial Rev. 18: 271-285

Kube J, Brokamp C, Machielsen R, van der Oost J, Märkl H. (2006). Influence of temperature on the production of an archaeal thermoactive alcohol dehydrogenase from *Pyrococcus furiosus* with recombinant *Escherichia coli*. Extremophiles. 10: 221-227

Ladenstein R, Antranikian G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. Adv Biochem Eng Biotechnol. 61: 37-85

Laemmli UK. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature. 227: 680-685

Lamed R, Zeikus JG. (1980). Ethanol production by thermophilic bacteria: relationship between fermentation product yields of the catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J Bacteriol. 144: 569-578

Lamed RJ and Zeikus JG. (1981). Novel NADP-linked alcohol--aldehyde/ketone oxidoreductase in thermophilic ethanologenic bacteria. Biochem J. 195: 183-190

Li D, Stevenson KJ. (1997). Purification and sequence analysis of a novel NADP(H)-dependent type III alcohol dehydrogenase from *Thermococcus* strain AN1. J Bacteriol. 179: 4433-4437

Littlechild JA, Guy JE, Isupov MN. (2004). Hyperthermophilic dehydrogenase enzymes. Biochem Soc Trans. 32: 255-258

Lu Z, Cabiscol E, Obradors N, Tamarit J, Ros J, Aguilar J, Lin ECC. (1998). Evolution of an Escherichia coli protein with increased resistance to oxidative stress. J Biol Chem. 273: 8308-8316

Lucas S, Toffin L, Zivanovic Y, Charlier D, Moussard H, Forterre P, Prieur D, Erauso D. (2002). Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. Appl Environ Microbiol. 68: 5528-5536

Ma K, Robb FT, Adams MWW. (1994). Purification and characterization of NADP-specific alcohol dehydrogenase and glutamate dehydrogenase from the hyperthermophilic archaeon *Thermococcus litoralis*. Appl Environ Microbiol. 60: 562-568

Ma K, Loessner H, Heider J, Johnson MK, Adams MWW. (1995). Effects of elemental sulfur on the metabolism of the deep-sea hyperthermophilic archaeon *Thermococcus* strain ES-1: characterization of a sulfur-regulated, non-heme iron alcohol dehydrogenase. J Bacteriol. 177: 4748-4756

Ma K, Adams MW. (1999). An unusual oxygen-sensitive, iron- and zinc-containing alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Bacteriol. 181: 1163-1170

Machielsen R, Uria AR, Kengen SWM, van der Oost J. (2006). Production and characterization of a thermostable alcohol dehydrogenase that belongs to the aldo-keto reductase superfamily. Appl Environ Microbiol. 72: 233-238

Miroshnichenko ML, Bonch-Osmolovskaya EA, Neuner A, Kostrikina NA, Chernych NA, Alekseev VA. (1989). *Thermococcus stetteri* sp. nov., a new extremely thermophilic marine sulfur-metabolizing archaebacterium. Syst Appl Microbiol. 12: 257-262

Miroshnichenko ML, Hippe H, Stackebrandt E, Kostrikina NA, Chernyh NA, Jeanthon C, Nazina TN, Belyaev SS, Bonch-Osmolovskaya EA. (2001). Isolation and characterization of *Thermococcus sibiricus* sp. nov. from a Western Siberia high-temperature oil reservoir. Extremophiles. 5: 85-91

Montella C, Bellsolell L, Pérez-Luque R, Badía J, Baldoma L, Coll M, Aguilar J. (2005). Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L -1,2-propanediol in *Escherichia coli*. J Bacteriol. 187: 4957-4966

Neale AD, Scopes RK, Kelly JM, Wettenhall RE. (1986). The two alcohol dehydrogenases of Zymomonas mobilis. Purification by differential dye ligand chromatography, molecular characterisation and physiological roles. Eur J Biochem. 154: 119-124

Nelson KE, Clayton R, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson J, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature. 399: 323-329

Nie Y, Xu Y, Mu XQ, Wang HY, Yang M, Xiao R. (2007). Purification, characterization, gene cloning, and expression of a novel alcohol dehydrogenase with anti-prelog stereospecificity from *Candida parapsilosis*. Appl Environ Microbiol. 73: 3759-3764

Reid MF and Fewson CA. (1994). Molecular characterization of microbial alcohol dehydrogenases. Crit Rev Microbiol. 20: 13-56

Peitsch MC. (1995). Protein modeling by E-mail. Biotech. 13: 658-660

Peretz M and Burstein Y. (1989). Amino acid sequence of alcohol dehydrogenase from the thermophilic bacterium *Thermoanaerobium brockii*. Biochem. 28: 6549-6555

Peretz M, Weiner LM, Burstein Y. (1997). Cysteine reactivity in *Thermoanaerobacter brockii* alcohol dehydrogenase. Protein Sci. 6: 1074-1083

Raia CA, Giordano A, Rossi M. (2001). Alcohol dehydrogenase from *Sulfolobus solfataricus*. Methods Enzymol. 331: 176-195

Radianingtyas H, Wright PC. (2003). Alcohol dehydrogenases from thermophilic and hyperthermophilic archaea and bacteria. FEMS Microbiol Rev. 794: 1-24

Rao S, Rossmann M. (1973). Comparison of super-secondary structures in proteins. J Mol Biol. 76: 241-56

Ravot G, Ollivier B, Magot M, Patel BKC, Crolet JL, Fardeau ML, Garcia JL. (1995). Thiosulfate reduction, an important physiological feature shared by members of the order *Thermotogales*. Appl Environ Microbiol. 61: 2053-2055

Reid MF, Fewson CA. (1994). Molecular characterization of microbial alcohol dehydrogenases. Crit Rev Microbiol. 20: 13-56

Reiter WD, Palm P, Zillig W. (1988). Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucl Acids Res. 16: 2445-2459

Riley M. Microbiol Mol Biol Rev. (1993). Functions of the gene products of *Escherichia coli*. 57: 862-952

Robb FT and Maeder DL. (1998). Novel evolutionary histories and adaptive features of proteins from hyperthermophiles. Curr Opin Biotech. 9: 288-291

Robb FT and Clark DS. (1999). Adaptation of proteins from hyperthermophiles to high pressure and high temperature. J Mol Microbiol Biotechnol. 1: 101-105

Robb FT, Maeder DL, Brown JR, DiRuggiero J, Stump MD, Yeh RK, Weiss RB, Dunn DM. (2001). Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology. Methods Enzymol. 330: 134-157

Ronimus RS, Reysenbach AL, Musgrave DR, Morgan HW. (1997). The phylogenetic position of the *Thermococcus* isolate AN1 based on 16S rRNA gene sequence analysis: a proposal that AN1 represents a new species, *Thermococcus zilligii* sp. nov. Arch Microbiol. 168: 245-248

Sambrook J, Fritsch EF, Maniatis T. (1989). Molecular Cloning: A Laboratory Manual (second edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York,

25-28

Schiraldi C, Marulli F, Di Lernia I, Martino A, De Rosa M. (1999). A microfiltration reactor to achieve high cell density in *Sulfolobus solfataricus* fermentation. Extremophiles. 3: 199-204

Schwarzenbacher R, von Delft F, Canaves JM, Brinen LS, Dai X and Deacon AM. (2004). Crystal structure of an iron-containing 1, 3-propanediol dehydrogenase (TM0920) from *Thermotoga maritima* at 1.3 Å resolution. Proteins: Struct. Funct. Genet. 54: 174-177

Schwede T, Kopp J, Guex N, Peitsch MC. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucl Acids Res. 31: 3381-3385

Simon H, Bader J, Gunther H, Neumann S, Thanos J. (1985). Chiral compounds synthesized by biocatalytic reductions. Angew Chem-Int Edit Engl. 24: 539-553

Smith JD, Robinson AS. (2002). Overexpression of an archaeal protein in yeast: secretion bottleneck at the ER. Biotech Bioeng. 79: 713-723

Stetter KO. (1989). Extremely thermophilic chemolithoautotrophic archaebacteria. In: Schlegel HG, Bowen Beds Autotrophic bacteria. Science Tech Publishers/Springer, Madison/Berlin, 167-176

Stetter KO. (1996). Hyperthermophilic prokaryotes. FEMS Microbiol Rev. 18: 149-158

Stetter KO. (1998). Hyperthermophiles: Isolation, classification, and properties. In: Extremophiles Microbial Life in Extreme Environments. John Wiley and Sons, New York, 1-24

Stetter KO. (2006). History of discovery of the first hyperthermophiles. Extremophiles. 10: 357-362

Studier FW and Moffatt B. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 189: 113-130

Sulzenbacher G, Alvarez K, van den Heuvel RH, Versluis C, Spinelli S, Campanacci1 V, Valencia C, Cambillau C, Eklund H and Tegoni M. (2004). Crystal Structure of *E. coli* Alcohol dehydrogenase YqhD: evidence of a covalently modified NADP coenzyme. J Mol Biol. 342: 489-502

Tabor S and Richardson C. (1985). A bacteriophage T7 RNA polymerase/ promoter system for controlled exclusive expression of specific genes. Proc Natl Acad Sci. USA. 82: 1014-1078

Takahata Y, Nishijima M, Hoaki T, Maruyama T. (2001). *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. Int J Syst Evol Microbiol. 51: 1901-1909

Takai K, Sugai A, Itoh T, Horikoshi K. (2000). *Palaeococcus ferrophilus* gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. Int J Syst Evol Microbiol. 50: 489-500

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res. 22: 4673-4680

Triglia T. (2000). Inverse PCR (IPCR) for obtaining promoter sequence. Methods Mol Biol. 130: 79-84

Ueda K, Yamashita A, Ishikawa J, Shimada M, Watsuji TO, Morimura K, Ikeda H, Hattori M, Beppu T. (2004). Genome sequence of *Symbiobacterium thermophilum*, an uncultivable bacterium that depends on microbial commensalisms. Nucl Acids Res. 32: 4937-4944

Ullmann A, Jacob F, Monod J. (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. J. Mol. Biol. 24: 339-343

Uriarte M, Marina A, Ramón-Maiques S, Fita I, Rubi V. (1999). The carbamoyl-phosphate synthetase of *Pyrococcus furiosus* is enzymologically and structurally a carbamate kinase. J Biol Chem. 274: 16295-16303

Unden G, Becker S, Bongaerts J, Schirawski J, Six S. (1994). Oxygen regulated gene expression in facultatively anaerobic bacteria. Antonie Leeuwenhoek 66: 3-23

Vallee H and Auld DS. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry 29: 5647-5659

van der Oost J, Voorhorst WGB, Kengen SWM, Geerling ACM, Wittenhorst V, Gueguen Y, de Vos WM. (2001). Genetic and biochemical characterization of a short-chain alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur J Biochem. 268: 3062-3068

Vieille C, Burdette DS, Zeikus JG. (1996). Thermozymes. Biotechnol Annu Rev. 2: 1-83

Vieille C, Zeikus GJ. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev. 65: 1-43

Walsh DJ, Gibbs MD, Bergquist PL. (1998). Expression and secretion of a xylanase from the extreme thermophile, *Thermotoga* strain FjSS3B.1, in *Kluyveromyces lactis*. Extremophiles. 2: 9-14

Whitesides GM and Wong CH. (1985). Enzymes as catalysts in synthetic organic chemistry. Angew Chem-Int Edit Engl. 24: 617-638

Windberger E, Huber R, Trincone A, Fricke H, Stetter KO. (1989). *Thermotoga thermarum* sp. nov. and *Thermotoga neapolitana* occurring in African continental solfataric springs. Arch Microbiol. 151: 506-512

Ying X, Wang Y, Badiei HR, Karanassios V, Ma K. (2007). Purification and characterization of an iron-containing alcohol dehydrogenase in extremely thermophilic bacterium Thermotoga hypogea. Arch Microbiol. 187: 499-510

Ying X, Grunden AM, NieL, Adams MWW, Ma K. (2008). Molecular characterization of the recombinant iron-containing alcohol dehydrogenase from the hyperthermophilic archaeon, *Thermococcus* strain ES1. The 108th annual conference of the American Society of Microbiologists. Boston, MA, U.S.A.

Yip KSP, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R, Consalvi V, Scandurra R, Rice DW. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure. 3: 1147-1158

Yoon SY, Noh HS, Kim EH, Kong KH. (2002). The highly stable alcohol dehydrogenase of Thermomicrobium roseum: purification and molecular characterization. Comparative Biochem Physiology Part B: Biochemistry and Molecular Biology. 132: 415-422

Zhu D, Malik HT, Hua L. (2006). Asymmetric ketone reduction by a hyperthermophilic alcohol dehydrogenase: the substrate specificity, enantioselectivity and tolerance of organic solvents. Tetrahedron Asymmetry. 17: 3010-3014

Ziegenhorn J, Senn M, Bücher T. (1976). Molar absorptivities of β-NADH and β-NADPH. Clin Chem. 22: 151-160

Ziegelmann-Fjeld KI, Musa MM, Phillips RS, Zeikus JG and Vieille CA. (2007). *Thermoanaerobacter ethanolicus* secondary alcohol dehydrogenase mutant derivative highly active and stereoselective on phenylacetone and benzylacetone. Prot Engin Design Select. 20: 47-55

Zillig W, Janekovic D, Schafer W, Reiter WD. (1983). The archaebacterium Thermococcus celer represents, a novel genus within the thermophilic branch of the archaebacteria. Syst Appl Microbiol. 4: 88-94