Enzymatic Production of Cellulosic Hydrogen by Cell-free Synthetic Pathway Biotransformation (SyPaB)

Xinhao Ye

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Biological Systems Engineering

Y.-H. Percival Zhang, Chair

Chenming (Mike) Zhang, Co-Chair

Justin Barone

Timothy J. Larson

July 6, 2011

Blacksburg, Virginia

Keywords: biofuel, directed evolution, hydrogen, phosphorylase, protein engineering, rational design, synthetic pathway biotranformation (SyPaB).

Enzymatic Production of Cellulosic Hydrogen by Cell-free Synthetic Pathway Biotransformation (SyPaB)

Xinhao Ye

ABSTRACT

The goals of this research were 1) to produce hydrogen in high yields from cellulosic materials and water by synthetic pathway biotranformation (SyPaB), and 2) to increase the hydrogen production rate to a level comparable to microbe-based methods (~ 5 mmol H_2/L/h).

Cell-free SyPaB is a new biocatalysis technology that integrates a number of enzymatic reactions from four different metabolic pathways, e.g. glucan phosphorylation, pentose phosphate pathway, gluconeogenesis, and hydrogenase-catalyzed hydrogen production, so as to release 12 mol hydrogen per mol glucose equivalent. To ensure the artificial enzymatic pathway would work for hydrogen production, thermodynamic analysis was firstly conducted, suggesting that the artificial enzymatic pathway would spontaneously release hydrogen from cellulosic materials. A kinetic model was constructed to assess the rate-limited step(s) through metabolic control analysis. Three phosphorylases, i.e. α-glucan phosphorylase, cellobiose phosphorylase, and cellodextrin phosphorylase, were cloned from a thermophile *Clostridium thermocellum*, and heterologously expressed in *Escherichia coli*, purified and characterized in detail. Finally, up to 93% of hydrogen was produced from cellulosic materials (11.2 mol H_2/mol glucose equivalent). A nearly 20-fold enhancement in hydrogen production rates has been achieved by increasing the rate-limiting hydrogenase concentration, increasing the substrate loading, and elevating the reaction temperature slightly from 30 to 32 ^{o}C. The hydrogen production rates were higher than those of photobiological systems and comparable to the rates reported in dark fermentations.

Now the hydrogen production is limited by the low stabilities and low activities of various phosphorylases. Therefore, non-biologically based methods have been applied to prolong the stability of α-glucan phosphorylases. The catalytic potential of cellodextrin phosphorylase has been improved to degrade insoluble cellulose by fusion of a carbohydrate-binding module (CBM) family 9 from *Thermotoga maritima* Xyn10A. The inactivation halftime of *C. thermocellum* cellobiose phosphorylase has been enhanced by three-fold at 70 ^{o}C via a combination of rational design and directed evolution. The phosphorylases with improved properties would work as building blocks for SyPaB and enabled large-scale enzymatic production of cellulosic hydrogen.

What I cannot create, I do not understand.
-- Richard Feynman

Acknowledgement

My time in Virginia Tech has been a gift. I have had the opportunity to work with many great people who have given me much in terms of both personal and professional development.

I would like to thank my advisor, Dr. Percival Zhang. Dr. P. Zhang has shaped my life in ways great (enlightening me with his vision) and small (continuously reminding me '*detail does matter*'), some direct (with sharp advices) and some by example (taking fictional characters to be my role models). I've tried to rise to the standard he sets, and imagine that I will keep trying for the rest of my scientific lifetime.

I am indebted to Dr. Mike Zhang on innumerable levels. He is a generous and calm mentor, a joy to work for, and a friend I can seek suggestions at any time. Without him, I would have given up my Ph.D. study more than a year ago.

I am grateful to my committee members, Dr. Justin Barone and Dr. Tim Larson. They always hold the door open for me. Their priceless advice and constructive comments have guided the direction of this research.

The research would not be possible without support of the Department of Biological Systems Engineering, Virginia Tech. I really want to express my gratitude to Dr. Saied Mostaghimi and Dr. Mary Leigh Wolfe for their encouragement and support. I also appreciate all the people in BSE for their kind help.

I owe a lot to my lab mates. Dr. Jiong Hong and Dr. Wenjin Liu shared many tips and tricks with me and helped solve my problems in molecular biology. I got hand-on experience in enzymology from Dr. Yiran Wang. Despite tense argument, Dr. Xiaozhou Zhang selfishlessly provided his input and suggestions for improving my projects. I'd like to thank Zhiguang Zhu, Noppadon (Tik) Sathitsuksanoh, Joe Rollin, Suwan Myung, Hehuan Liao, and Fangfang Sun for always being there with a warm heart and

sympathetic ears. I am also thankful to Dr. Zuoming Zhang who offered a warm place where I dare to get drunk.

My special thanks go to Dr. Jonathan Mielenz, Dao Zhou, Dr. Jean Peccoud, Chuan Han, and Dr. Pengtao Yue. Every phosphorylase studied here was amplified from the genomic DNA of *Clostridium thermocellum* that is a gift from Dr. Jonathan Mielenz. Dao Zhou designed the printed circuit boards (PCB), through which the hydrogen sensors started to function. The carbohydrate-binding modules (CBM) family 4, family 6, and family 9 in Chapter 4 were synthesized in Dr. Jean Peccoud's group and given freely. Chuan Han and Dr. Pentao Yue helped me fix the stochastic model in Chapter 5, by which the optimal mutagenesis frequency was identified.

I benefit a lot from my former advisors, Prof. Siliang Zhang and Dr. Kean Wang. I'm grateful to their blind investment in me. They opened up my mind and trained me at multiple scales. Much appreciation is given to Gramma (Qilin) Li. I got to know her when I was a sophomore. Since then, she has been the one whom I could contact whenever I had troubles. Another special 'thanks' goes to Simon Kwong. He is a mentor for life, a friend in need, and a chef who finds 'the cure' to my homesickness. I'd also like to extend my thanks to the crew in Virginia Tech and beyond: Dr. Zheying Guo, Dr. Ruijia Wang, Dr. Zhenyu Zhang, Sarah Jennings and her friends, Dr. Pu Wang, Guanying Wang, Dr. Yihong Yang, Dr. Qinqin Chen, Dr. Souwen Lai, Canming Jiang, Jianzhong Hu, Lifeng Li, Dr. Ying Jin, Dr. Hao Yu, Chuan Han, Dr. Kun Zhou, Cunhao Gao, Senlin Peng, Dr. Zengping Xing, Min Li, Dr. Bin Xu, Dr. Bin Xue, Dr. Shuangchuan Yan, Shuchi Wu, Dr. Liang Shan, Luman Chen, Qin Li, Ian Doran, Yunde Zhang, Dr. Kaigui Bian, Xiaomo Liu, Liguang Xie, Ji Wang, Jian Zhang, Bo Wen, Dr. Yong Wang, and Prof. Damai Zhou, etc. In addition to the science, they have made my life in Blacksburg colorful.

Finally, I would like to express my deepest gratitude to my parents for supporting and believing in me as I chase my dreams. I appreciate my sister who takes care of the parents and thus allows me to continue my education abroad.

Table of Contents

List of Figures

List of Tables

Chapter 1

Hydrogen production from carbohydrates: A review

1.1 Introduction

Fossil fuels – oil, nature gas, and coal – are providing more than 85% of the world's total energy (Burgdorf et al. 2002). In 2007, the world consumed approximately 3.11×10^{10} barrels of oil, 2.92×10^{12} m^3 of natural gas, and 3.18×10^{9} tons of coal. As compared to their known reserves (oil, 1.24×10^{12} barrels; natural gas, 1.77×10^{14} m^3, coal, 8.47×10^{11} tons), oil, natural gas, and coal would go to depletion in year 2047, 2068, and 2273, respectively, if the world annual energy consumption rates kept unchanged since 2007. In addition, combustion of fossil fuels also raises concerns such as global climate change, air pollution, and health problems (Bockris 2002; Levin et al. 2004; Pacala and Socolow 2004). Now to identify and build up a sustainable energy system for remedying the depletion of fossil fuels and their environmental misdeeds have become two of the most critical issues of our time (Das and Veziroglu 2001; Turner 2004).

Hydrogen is the most common element in the universe, and has been considered as a promising energy carrier to replace fossil fuel-based liquid fuels, offering a unique capability to address issues of sustainability, environmental emissions, and energy security (Kotay and Das 2008; Penner 2006; Turner 2004). Hydrogen is present in water (most), biomass, and hydrocarbons, but no hydrogen is found in pure form on our planet. Figure 1.1 presents a number of hydrogen generation techniques. They can be classified based on energy sources. From solar energy, hydrogen can be produced by photobiological (e.g., algae) and photochemical (e.g., TiO_2-based catalysis) methods. But advances must be made before these technologies become economically competitive (Milne 2002). Hydrogen can be produced through water electrolysis with consumption of electricity. But water splitting by electricity is too costly even up to date (Barbir 2005). Large-scale hydrogen has mainly been produced from fossil fuels through gasification and water shifting (Das and Veziroglu 2001; Navarro et al. 2007). Among them, steam reforming of natural gas is the most common way to produce hydrogen (Wagner et al. 1998). Today nearly 95 % of all hydrogen in U.S. and 48 % globally is produced from steam reforming of natural gas (Dunn 2002). However, hydrogen generation from depleting fossil fuels is not a sustainable process. It also releases greenhouse gases (Turner 2004).

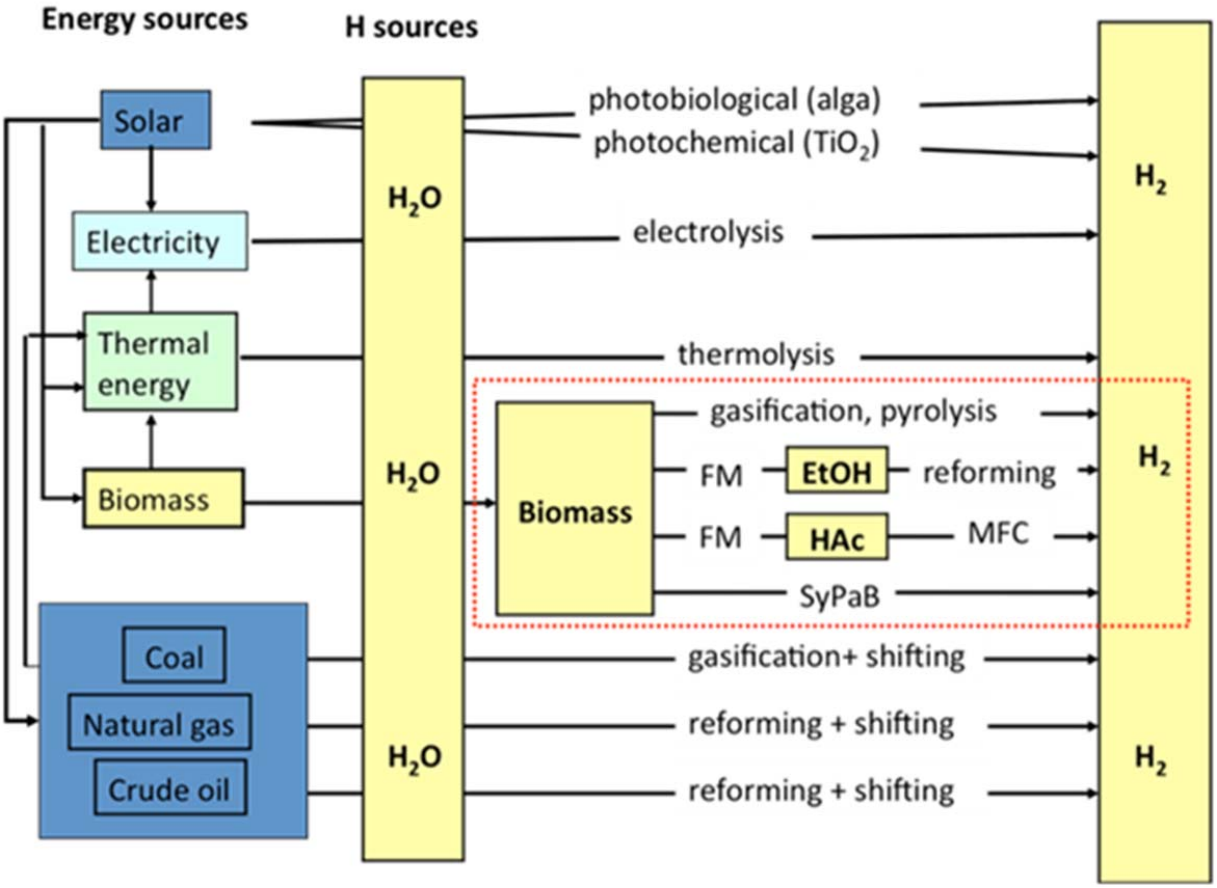


Figure 1.1 Hydrogen production methods. The methods are classified with respect to their energy sources The frame with red dot-line presents the conversion of biomass to hydrogen. EtOH, ethanol; FM, fermentation; HAc, acetic acid; MFC, microbe fuel cell; SyPaB, synthetic pathway biotransformation.

Hydrogen production from low-cost renewable carbohydrates (biomass or its derivatives) is a promising sustainable solution to the hydrogen economy (Huber et al. 2003; Milne 2002). Renewable carbohydrate is produced by photosynthesis of plants, which are natural solar cells that enrich low-energy-density solar energy to chemical energy carriers – carbohydrates. A number of methods have been developed to generate hydrogen from carbohydrates, as shown in Fig. 1.2.

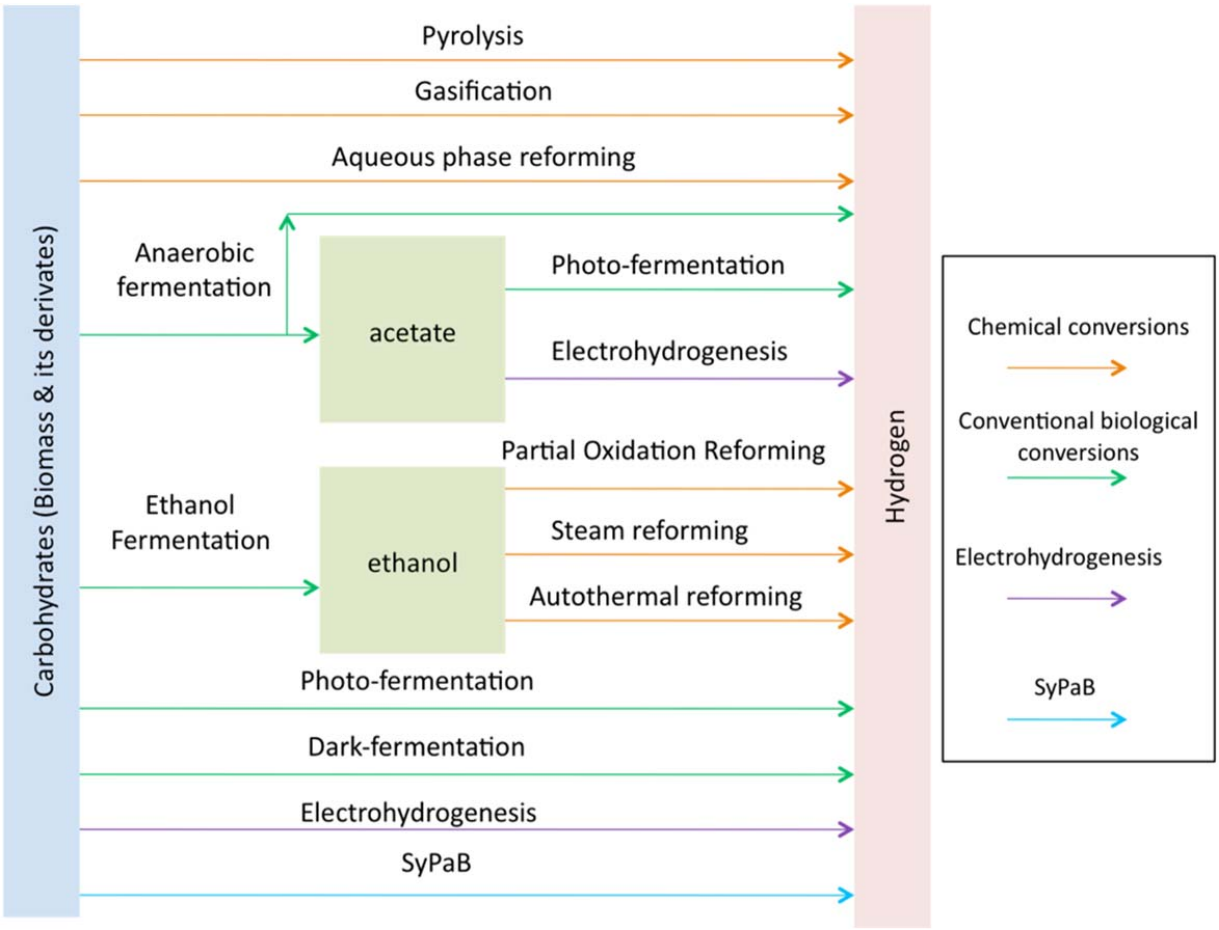


Figure 1.2 The conversion of carbohydrates to hydrogen.

In this article, we reviewed the carbohydrate-based hydrogen technologies, addressing the key barriers to assembly of H_2 infrastructures and the pros and cons associated with these approaches. It's worth noting that any economically viable process must be closely linked to the qualities of locally available biomass feedstock, including cost, distribution, mass, and physical and chemical characteristics (Milne 2002). The conversion yield from carbohydrates to hydrogen thus becomes critically important, and any conversion in low yield strictly means economic infeasible. The US DOE has set a conversion goal at 50% efficiency for hydrogen production from biomass, by which hydrogen costs could be reduced from $6 to $1.50/kilogram (DOE 2003; Logan 2004). Hence, we highlighted the technological possibilities to improve the yield as well as the reaction rate, and tried to provide a relatively clear and systematical picture on which technology (technologies) we could rely on in the future.

1.2 Carbohydrates-to-hydrogen conversion

1.2.1 Chemical catalysis

1.2.1.1 Pyrolysis

Pyrolysis is a chemical decomposition of condensed substance by heat under anoxic (oxygen-free) conditions (Rezaiyan and Cheremisinoff 2005). Pyrolysis of coal and biomass (primarily wood) was popular for producing fuel-related gas and smokeless solid fuel, for example charcoal, from 1700s to early 1900s (Klass 1998). Now it is heavily applied to convert biomass into syngas, to produce coke from coal, and to treat hazardous wastes.

The products of biomass pyrolysis include biochar, bio-oil and gases (e.g. hydrogen, carbon monoxide, and carbon dioxide). The composition is influenced by temperature, heating rate, and particle size of biomass (Probstein and Hicks 2006; Zanzi et al. 2002). High temperatures promote gas production, while lower temperatures favor the formation of char, tar, and heavy oil (Klass and Emert 1981). In the rapid heating experiments, the rapid removal of volatiles from the reactor prevents the formation of secondary char and attributes to a higher volatile yield and thus a lower char yield. Zanzi *et al.* reported the particle sizes of feedstock affected heating rate and altered the product composition of pyrolysis (2002). Since both heat flux and heating rate are lower in the large particles than in the smaller ones, small particles can be heated up more efficiently so that the average particle temperature is higher. It results to a higher devolatilization rate and higher yields. Recently considerable attention is attracted to pyrolysis of carbohydrates by thermal plasma (Gomez et al. 2009; Huang and Tang 2007; Tang and Huang 2005). In favor of the high heating rate and the high energy density of plasma, plasma pyrolysis is able to gasify and dissociate organic matter in a few milliseconds. There is no oxygen available and no burning in any form, ensuring biomass pyrolysis with high yields of gas (~70 wt %), high-quality of char, and little tar formation (Gomez et al. 2009; Rezaiyan and Cheremisinoff 2005; Tang and Huang 2005).

1.2.1.2 Gasification

Conventional biomass gasification

Biomass can also be converted into hydrogen and other useful products by steam/oxygen gasification (Navarro et al. 2007). Unlike pyrolysis, biomass gasification is usually operated at the temperature above 1000 K and in the presence of oxygen and/or water. Pyrolysis of carbohydrates principally generates tar and char, while gasification produces mainly gaseous products (Rezaiyan and Cheremisinoff 2005). Coupled by water-gas shift reaction, gasification is regarded to be more favorable for hydrogen production than pyrolysis (Ni et al. 2006).

Nevertheless, to eliminate tar formation still remains one of the major issues over biomass gasification. Many efforts have been made to meet the demand (Narvaez et al. 1997; Navarro et al. 2007). Operation parameters are key factors in formation and decomposition of tar. In the similar way pyrolysis applied, low tar formation could be achieved by high temperature, rapid heating rate, and small feedstock particle size (Chen et al. 2003; Milne 2002). Additives and catalysts, such as dolomite, olivine, and even char, are helpful for tar reduction (Corella et al. 1999; Corella et al. 2004). Proper selection of gasifier and good process design also remain critical to reduce tar formation. In the downdraft gasifier, the product gas contains the lowest concentration of tar and particles (1 g/m^3) as most of the tars have been combusted (Navarro et al. 2007). However, the disadvantages are also obvious: 1) the overall thermal efficiency is low, and 2) it is difficult in handling ash content and tar. Alternatively, fluidized-bed gasifier is widely used for large-scale biomass gasification. With suitable catalysts, about 60 vol % of hydrogen yield and a medium tar yield of around 10 g/Nm^3 can be achieved by fluidized-bed gasifiers (Navarro et al. 2007; Ni et al. 2006; Weerachanchai et al. 2009).

In spite of the tar formation, the biogas generated from biomass gasification generally contain contaminants, such as particulates, ashes, alkali compounds, nitrogen-containing components, sulfur, and low-molecular-weight hydrocarbons (e.g. methane and ethane) (Rezaiyan and Cheremisinoff 2005). The need to remove the contaminants from biogas steam depends on the end use of the gas. Several reports and reviews are recommended on this topic (Cummer and Brown 2002; Hallgren 1995; Rezaiyan and Cheremisinoff 2005; Stevens 2001).

On the other hand, gasification coupled with water-gas shift is the most widely practiced process route for biomass to hydrogen. Two integrated processes, one consisting of an air-blown bubbling fluidized bed gasifier, a steam reformer, and a water-gas-shift membrane reactor and the other integrating water carbon reaction, water-gas shift reaction and CO_2 adsorbing reaction in a single reactor (HyPr-RING), were recently reported to completely decompose tar and produce ultrapure hydrogen from biomass gasification (Ji et al. 2009; Lin et al. 2005).

Gasification in supercritical water (SCW)

Generally, the gasification process is applicable to biomass having a moisture content of less than 35% (Demirbas 2002). Pre-drying may be required, and moisture content of 10% to 20% is usually preferred (Rezaiyan and Cheremisinoff 2005). Yet it is possible to gasify biomass in supercritical water (SCW) conditions if the moisture content exceeds 35%. The process is operated at pressures and temperature above critical point (221 Bar and 647 K) of water to the supercritical conditions.

SCW gasification is a promising alternative to gasify biomass with high moisture contents because of the high gasification (100% conversion) and hydrogen ratios (50 vol %) (Huber et al. 2006; Matsumura et al. 2005; Xu et al. 1996). Under the low-temperature SCW gasification of biomass (within the range of 623-873 K), lignin and hemicelluloses is firstly decomposed, while cellulose, as the most stable component of biomass, is hydrolyzed to sugar and then forms gases, oil, char, and other water-soluble compounds (Sasaki et al. 2000). In the absence of catalyst, around 60% of carbon is converted to char, 15% of carbon remains in water and 10% goes to gas (mainly as CO_2 and small amounts of CO) (Minowa and Fang 1998). With nickel catalyst, water-soluble compounds can be catalyzed into CO_2, H_2 and CH_4. The yield of char and oil goes down, and CO_2 and H_2 become the primary gas product (Minowa et al. 1998). In most cases, metal catalysts are employed to accelerate the reaction and improve the biomass-to-hydrogen conversion. However, the catalysts are unstable and easily oxidized in the hot water environment (Elliott et al. 1993). In this respect, new catalyst formulations have been developed, such as combinations of stable metals (e.g. ruthenium or nickel bimetallics) and stable supports (e.g. certain titania, zirconia, or carbon) (Matsumura et al. 2005). Among them, ruthenium on rutile titania extrudate is particularly effective in this process and remains active for long periods of

operation (Elliott et al. 2001). Alkali catalysts have also been applied for this purpose. Moreover, sodium carbonate has been found to be helpful to degrade cellulose at substantially lower temperature (453K) and be capable to increase the biomass gasification rates (Minowa et al. 1998).

Nevertheless, complete gasification of feedstock is difficult under the low-temperature conditions. By contrast, high-temperature SCW gasification (in the range of 773-1073 K) is exploited because of the high reactivity of biomass and the potential to achieve high gasification efficiency (Matsumura et al. 2005; Navarro et al. 2007). Note that the efficiency of high-temperature SCW gasification is governed by the characteristics of feedstock, operation parameters, reactor, catalyst, etc. The researcher of University of Twente investigated gasification of glycerol and glucose in detail (Matsumura et al. 2005). They found the concentration of feedstock had a major influence on the gas yield. 100 % gasification efficiency was achieved only at feedstock concentrations below 3%. Bigger water gas shift activities were observed at temperature above 873 K. Once the reaction temperature declined to be lower than 873 K, the yield of CO increased while the yield of H_2 and CO_2 sharply decreased (Matsumura et al. 2005). Yu *et al.* from University of Hawaii conducted gasification of biomass without catalyst by a metallic tubular flow reactor (1993). Their results indicated the efficiency of glucose gasification was affected by the reactor wall material. For example, they pretreated the reactor with sodium chloride to obtain a catalytic effect from the wall. Additionally, the addition of alkali compounds (like KOH, $KHCO_3$, and Na_2CO_3), which accelerated the water-gas shift reaction, in SCW gasification was helpful to increase H_2 yield, to decrease CO yield, and to reduce char/coke formation (Elliott and Sealock 1983).

1.2.1.3 Aqueous-phase reforming

Aqueous-phase reforming (APR) is a unique process where reforming is done in the liquid phase. It's usually carried out at low temperature (400K-550K) and medium pressure (50-70 bar) (Chheda et al. 2007). In 2002, Dumesic *et al.* first demonstrated hydrogen can be produced from biomass-derived hydrocarbons at temperatures near 500 K in a single-reactor aqueous-phase reforming process using a platinum-based catalyst (Cortright et al. 2002). The advantages of APR against vapor-phase reforming of oxygenated hydrocarbons are 1) APR occurs in the

presence of water at a temperature near 500 K and a total pressure near 30 bar, by which water-gas shift (WGS) reaction is favorable, making it possible to generate hydrogen with low level of CO (around 100 ppm); 2) hydrogen gas is generated without vaporizing water which represents major energy savings; 3) by taking place at low temperatures and low vapor pressures, the process minimizes undesirable decomposition reactions typically encountered when carbohydrates are heated up to elevated temperatures; and 4) nonvolatile modifiers are required to control the performance of the catalytic process, for example addition of basic or acidic components to adjust pH of the aqueous carbohydrate feed stream (http://www.wisbiorefine.org/proc/apr.pdf). A disadvantage of the APR process is that leaching of catalyst components into the aqueous phase damages the stability of catalysts under the reaction conditions. Like the treatment of catalysts in SCW process, long-term hydrothermal stable materials, such as carbon, titania, and zirconia, are chosen to support and stabilize the catalysts (Chheda et al. 2007).

Practically, the selectivity towards H_2 by APR is controlled by catalytically active sites, catalyst support, solution pH, feed concentration, process conditions, and reactor design. Pt, Ni, and Ru exhibit relatively high activities for the reforming reaction, but only Pt and Pd show relatively high selectivity for the production of H_2 (Davda et al. 2005). It infers that active catalysts for APR reactions should possess high catalytic activity for C-C bond cleavage and promote water-gas shift to covert CO into CO_2. Furthermore, the catalysts should exhibit low activity for C-O scission and not facilitate further hydrogenation reactions of CO and CO_2 to form alkanes. On this basis, Pt-based catalysts are promising for hydrogen production from biomass by APR. The effect of catalyst supports was investigated on the activity and selectivity for H_2 production by APR of ethylene glycol (Shabaker et al. 2003). It was demonstrated that Pt-black and Pt supported on TiO_2, carbon, and Al_2O_3 exhibited the highest turnover frequencies, while Pt supported on CeO_2, ZnO, and SiO_2 had the lowest activity for production of hydrogen. C/O stoichiometry of the feedstock is also of importance for the high selectivity of H_2 production. The C/O stoichiometry of 1:1 is recommended to obtain high selectivity towards H_2. The organic compounds that do not have a 1 to 1 C/O stoichiometry would lead to the formation of additional amounts of alkanes, because only oxygen-bound carbon atoms would take the function of the formation of H_2 and CO_2 (Chheda et al. 2007; Huber et al. 2005). More advances on hydrogen

production by aqueous-phase reforming can be referred to Davda's (2005) and Jürgen's review (2006).

Although APR process is attractive to produce H_2 from actual biomass, hydrogen has been formed in very low yields (1.05-1.41 mmol per gram of carbohydrates) owing to the formation of cokes and byproducts (Valenzuela et al. 2006). Nevertheless, aqueous-phase processing may play a critical role to catalytically transform carbohydrates in the liquid phase to value-added chemicals and fuels in the future.

1.2.2 Biological catalysis

Biological catalysis, different from chemical and electrochemical processes, is catalyzed by microorganisms or enzymes in an aqueous environment at ambient temperature and pressure. Hence, biological hydrogen production is less energy intensive than the chemical methods (Manish and Banerjee 2008; Nath and Das 2004). Another advantage of biological hydrogen production is that microorganisms can recover and concentrate the energy from high water content organic resources, such as wastewater, sewage-sludge, etc., which, in contrast, is difficult to achieve by chemical and mechanical means (Kotay and Das 2008).

The processes of biological hydrogen production can be broadly classified into two distinct groups. One is light dependent, including direct or indirect biophotolysis and photo-fermentation; the other is a light-independent process normally called dark-fermentation. Water and light are the natural sources for biophotolysis. Despite the excitement over the developments of biophotolysis in recent decades, many obstacles to putting these technologies towards real-world applications remain unsolved (Akkerman et al. 2002). For instance, the oxygen produced by photosynthesis (PS II) inhibits hydrogenase and then disables the hydrogen production (Benemann 1996). Since our goal is to evaluate the carbohydrates-to-hydrogen conversion, biophotolysis will not be addressed in this review. Those who are interested in this area can refer to relevant reports and reviews (Benemann 1998; Hallenbeck and Benemann 2002; Miura 1995).

1.2.2.1 Photo fermentation

Photo-fermentation is a process in which photosynthetic microorganisms, like purple non-sulfur bacteria, directly convert solar energy to hydrogen from organic/inorganic substrates and water (Levin et al. 2004). In these organisms, the photosystem is not powerful enough to split water. Under anaerobic circumstances, however, the bacteria are able to use simple organic acids, like acetic acid or hydrogen disulfide as electron donor and to accumulate extra ATP energy (Akkerman et al. 2002). Under nitrogen-limited circumstances, the extra energy is used to reduce molecular nitrogen into ammonium by nitrogenase, whereas when molecular nitrogen is not present, the nitrogenase will use the energy to reduce protons to generate hydrogen. Since the light energy is not used for water oxidation, the efficiency of light energy for the production of hydrogen via photosynthetic bacteria is, theoretically, much higher than biophotolysis catalyzed by cyanobacteria (Asada and Miyake 1999).

Nitrogenase is the key enzyme that catalyzes hydrogen production by photosynthetic bacteria, whereas hydrogenases may be active for hydrogen uptake and therefore is antagonistic to nitrogenase activity (Koku et al. 2002). Inactivation of hydrogenase will lead to a mutant that is not able to recycle the hydrogen evolved by nitrogenase under the nitrogen-fixing conditions so that the hydrogen productivity can be increased (Tamagnini et al. 2002). For example, a mutant strain of *R. sphaeroides* KD131 lacking uptake hydrogenase (Hup^-) and PHB (polyhydroxybutyrate) synthase (Phb^-) produced 2-3 times more hydrogen than the wild type (Kim et al. 2006). Similar results were also observed in the Hup^- mutant of *A. variabilis* ATCC 29413 (Happe et al. 2000) and *Nostoc sp.* PCC7120 (Happe et al. 2000). On the other hand, the balance between photosystem and hydrogen producing enzymes is required for efficient hydrogen production (Asada and Miyake 1999). The light harvesting systems were rearranged by Miyake et al. using a 'promoter (puf) competition method' to enhance the bacterial light-dependent hydrogen production (Miyake et al. 1999; Nagamine et al. 1996). The study of the physiology of genetically modified photosynthetic microorganisms has shown that electron flux could be redirected to the bidirectional hydrogenase in an *ndhB* (NADPH-dehydrogenase) mutant of *Synechocystis* (Cournac et al. 2004; Vignais et al. 2006). The mutant without NADPH-dehydrogenase (NDH-1) produced only low amounts of O_2 in the light, had a poor capability to

fix CO_2 and to uptake H_2, and thus accumulated NADP, by which electron flux could be redirected to the bidirectional NAD(P)-dependent [NiFe] hydrogenase for hydrogen production.

The activity of nitrogenase is inhibited in the presence of oxygen, ammonia or at high N/C ratios (Kapdan and Kargi 2006; Koku et al. 2003). Most nitrogenases are irreversibly inhibited by oxygen, which degradatively oxidizes the Fe-S cofactors. Therefore, it requires photofermentation as well as photosynthetic bacteria themselves to separate nitrogen fixation and oxygen generation either spatially or temporally (Holladay et al. 2009). However, one known exception is the nitrogenase of *Streptomyces thermoautotrophicus* that is completely insensitive to oxygen and dissimilar from the conventional nitrogenases (Ribbe et al. 1997). In the presence of high nitrogen concentrations, the metabolism shifts to utilization of organic substance for cell synthesis rather than hydrogen production, resulting in excess biomass growth and reduction in light diffusion. Consequently, the process requires ammonium limited and oxygen free conditions (Takabatake et al. 2004; Zhu et al. 2001), and a proper ratio of carbon to nitrogen nutrients must be maintained (Holladay et al. 2009).

In addition, photofermentation is also affected by light intensity, carbon source and the type of microbial culture. Increasing light intensity stimulates the hydrogen yield and production rate, but decreases the light conversion efficiency (Shi and Yu 2005). In favor of the potential that light is able to penetrate to the deep region of the reactor, a pigment-reduced mutant of *R. sphaeroides* was constructed that can improve the hydrogen yield and light conversion efficiency simultaneously (Kondo et al. 2002; Miyake et al. 1999). The depletion of the color intensity of feedstock by clay treatment was found to have the similar effects (Eroglu et al. 2006). Moreover, several types of photo-bioreactors, such as tubular (multi-tubular), flat panel, bubble column reactors, have been developed in order to increase the light intensity and photochemical efficiency of hydrogen production (Akkerman et al. 2002). A 9.23% maximum light conversion efficiency was achieved under light intensity of 300 W/m^2 by using a novel light-induced and diffused photo-bioreactor (IDPBR) in RITE Japan (El-Shishtawy et al. 1997).

The photosynthetic bacteria prefer organic acids as carbon source, such as acetic, butyric, propoinic, lactic and malic acid (Eroglu et al. 1999; Fang et al. 2005; Shi and Yu 2004). Based on available literature, the highest conversion efficiency was obtained with lactic acid (Asada

and Miyake 1999; Kapdan and Kargi 2006). To date, the efficiency has reached as high as 85% (He et al. 2005). Meanwhile, efforts have been made to explore the potential of producing hydrogen from waste organic compounds by photo-fermentation (Bolton 1996; Fedorov et al. 1998; Tsygankov et al. 1998) either in batch processes (Zurrer and Bachofen 1979) or in continuous cultures (Zürrer and Bachofen 1982). For photo-fermentation, the maximum hydrogen production rate of 0.16 mmol H_2/L/h was reported by culturing *Rb. spheroids* immobilized on porous glass (Levin et al. 2004; Polle et al. 2002).

1.2.2.2 Dark fermentation

Hydrogen can be produced by anaerobic bacteria growing on carbohydrates without light. Unlike biophotolysis or photo-fermentation that produces pure H_2, the anaerobic bacteria in dark-fermentation usually produce a mixed biogas containing primarily H_2, CO_2, and in some way small amounts of methane, CO, and/or H_2S (Levin et al. 2004). *Enterobacter*, *Bacillus*, and *Clostridium* are well-known species to produce hydrogen. However, *Clostridium pasteaurianum*, *C. butyricum*, and *C. beijerinkii* are strong H_2 producers, while *C. propionicum* and other non-sporeformers are poor H_2 producers (Hawkes et al. 2002), because Clostridia such as *C. pasteaurianum* and *C. butyricum* produce predominantly butyrate, whereas *C. propionicum* mainly produce propionate. In practice, high H_2 yields are usually associated with butyrate and acetate, and low yields are with the production of propionate or other reduced end products, like alcohols and lactic acid (Vavilin et al. 1995). Another interesting fact of dark fermentation is that certain phototrophic bacteria of the Rhodospirillaceae family can grow in the dark using CO as the sole carbon source to generate H_2, CO_2 and ATP (Champine and Uffen 1987; Kerby et al. 1995). Such a water-gas shift reaction is thermodynamically favorable (-20 kJ/mol) (Levin et al. 2004). Stoichiometric amounts of CO_2 and H_2 were produced by *R. gelatinosa* during CO-oxidation (Uffen 1983).

Dark-fermentation has been regarded as a nearer-term approach to practical biological hydrogen production (Benemann 1996). In literature, hydrogen synthesis rate by dark fermentation was reported as high as 121 mmol H_2/l/h (Levin et al. 2004). However, it's still far from the practical application because anaerobic bacteria can only produce relatively small amounts of hydrogen in the dark, typically only 10-20 % stoichiometrically (Benemann 1996; Kotay and Das 2008;

Solomon et al. 1995). For a thermodynamic perspective, the most favorable products of breaking down 1 mol glucose are 2 mol acetates and 4 mol H_2. The near-stoichiometric yield (4 mol H_2 /mol glucose) is only achievable under near-equilibrium conditions, which implies very slow rates and/or at very low partial pressures of H_2 (Hallenbeck and Benemann 2002; Nath and Das 2004).

In order to approach the theoretical yield, big efforts have been put in process optimization, reactor design, and metabolic engineering recently (Hansel and Lindblad 1998; Hawkes et al. 2002; Nath and Das 2004). Process conditions including nutrient sources, pH, reducing equivalents generated, H_2 partial pressure, hydraulic retention time (HRT), as well as reactor design, have significant effects on H_2 yield, since they influence on the fermentation end products (Hawkes et al. 2002; Levin et al. 2004). The relationship between hydrogen production by *Oscillatoria sp.* Miami BG7 and nitrogen nutrient was studied by Kumazawa and Mitsui (1981). When a combined nitrogen sufficient culture was inoculated into a combined nitrogen-limited medium, the capability of photoproduction of oxygen decreased while the capability of hydrogen production dramatically increased (260 μmol/mg chlorophyll I/h). Lay systemically investigated the effect of feedstock (microcrystalline cellulose) concentrations on biohydrogen generation by mesophilic anaerobic batch fermentation, and found an upper limit of feedstock concentration of 25 g [cellulose] /L , above which specific hydrogen production rate fell (2001). *C. pasteurianum* is a classic volatile fatty acid (VFA) and H_2 producer, but metabolic flux can be shifted to produce solvent rather than H_2 by high glucose concentrations (12.5% w/v) but limited CO and iron supply (Dabrock et al. 1992). pH affects the hydrogen production as well. The reported pH range for the maximum hydrogen yield and specific hydrogen production rate is between pH 5.0 and 6.0 (Kapdan and Kargi 2006). Experiments demonstrated *C. acetobutyricum* favored hydrogen production at pH above 5, phosphate and iron above limiting levels and glucose concentration below 12.5% (Hawkes et al. 2002). Furthermore, hydrogen producing pathways are sensitive to H_2 and subject to end-product inhibition. Hence, hydrogen partial pressure in the liquid phase is one of the key factors affecting H_2 production. Increased agitator speed and sparging with nitrogen or argon are efficient to lower dissolved H_2 concentration (Mizuno et al. 2000; Tanisho et al. 1998). Other potentially useful methods for H_2 removal rely on gas-stripping and membrane-adsorption technologies (Nath and Das 2004). Recently a high-

rate trickling filter was designed for dark fermentation of thermophiles, in which the hydrogen gas can be removed by steam stripping (van Groenestijn et al. 2002). The energy required to run such a process was more than 4 times lower than the combustion value of the hydrogen gas produced. For the immobilized cell culture, gas holdup is a major problem in tubular bioreactors. Tapered and rhomboid bioreactors are used to improve the performance in terms of both the rate of hydrogen production and the gas holdup (Kumar and Das 2001). Kumar & Das founded the gas holdup was reduced by 67% using the rhomboid bioreactor for the continuous hydrogen production by immobilized *E. cloacae* IIT-BT 08, as compared to the tubular one. The maximum hydrogen production rate achieved was 75.6 mmol/l/h at a dilution rate of 0.93 h^{-1} and recirculation ratio (recirculation volume to feed volume ratio) of 6.4 (Kumar and Das 2001).

Beyond that, hydrogenase is unstable, sensitive to oxygen, and inhibited by hydrogen itself. Directed evolution of an oxygen-tolerant hydrogenase has been proposed years ago (Boyer et al. 2005). Unfortunately, little progress is reported mainly due to the lack of convenient high-throughput screening methods. A successful example was to replace *Azotobacter vinelandii* hydrogenase small-subunit cysteines with serines that created insensitivity to oxygen inhibition and preferentially restricted hydrogen oxidation over hydrogen evolution (McTavish et al. 1995). Burgdorf *et al.* created 27 mutants of *Ralstonia eutropha* [NiFe] hydrogenase by site-directed mutagenesis, in which the amino acids replacement occurred in five of six conserved motifs surrounding the active site. Several mutants showed O_2-senstive growth on hydrogen but no O_2-tolerant mutants were found (Burgdorf et al. 2002). Nonetheless, the hydrogen-producing activity of *E.coli* hydrogenase 3 has been enhanced by 17 folds via error-prone PCR (Maeda et al. 2008a). Series of hyperthermophilic hydrogenases have been characterized, heterologously expressed, and purified (Jenney and Adams 2008). Structural and functional analysis of those enzymes will be extremely important not only to dissect the molecular mechanism of hydrogen activation but to improve the enzymatic properties for biohydrogen production as well (Ye et al. 2005).

Another challenge of producing cost-effective hydrogen by dark fermentation is that the high energy requirement limits the efficiency of the hydrogen production, resulting in a low hydrogen yield and a low volumetric hydrogen production rate. Several fermentation and genetic strategies can be employed to maximize the production yield of the selected organism (Claassen et al.

1999; Nath and Das 2004). Overexpression of cellulases, hemicellulases, and lignases will maximize substrate (glucose) availability, while coupling the genes with a strong promoter is useful to over-expression of hydrogenases and thus to increased level of hydrogen produced by the organism (Tamagnini et al. 2002). Low volumetric hydrogen productivity is mainly attributed to a low hydrogen production rate per cell and a low cell density, which resulted from a low growth rate under anaerobic conditions (Yoshida et al. 2005). To develop genetically engineered bacteria (either *E. coli* or yeast) with hydrogen producing pathways, including hydrogenase maturation pathways, is a well-accepted solution for this challenge. Efforts have been made to enhance hydrogen production from formic acid by modifying formate hydrogen lyase (FHL) complex of *E. coli*. The researchers at Texas A&M University altered regulation of FHL complex and then increased the hydrogen production by a strain of *E. coli* 140 times greater than that by the wild type. In addition, the hydrogen yield from glucose was increased by 50%, and a three-fold higher hydrogen production rate was also achieved by this strain (Maeda et al. 2008b).

In spite of the traditional pure culture, dark fermentation by mixed microflora is also attractive for the production of biohydrogen. The specific advantages of such a mixed culture biotechnology include: no sterilization requirements, adaptive capacity owing to microbial diversity, broad substrate spectrum, and the possibility of a continuous process (Kleerebezem and van Loosdrecht 2007). With respect to these properties, mixed culture biotechnology has also been widely applied for waste utilization. However, the H_2 production by mixed microflora is very sensitive to process conditions, such as feedstock, inoculation start-up, pH, temperature, and so on (Hawkes et al. 2002). It is necessary to avoid the presence of methanogens and other organisms that utilize H_2 for maximizing the hydrogen productivity. Low pH and/or short retention times can be implemented to fulfill this task, since methanogens grow more slowly than other fermentative organisms (Das and Veziroglu 2001).

1.2.2.3 Two-stage process (integration of dark- and photo- fermentation)

Owing to biochemical and thermodynamic limitations, the maximum of 4 mol hydrogen can be expected from ideal acetate fermentation. Most often, hydrogen yields [mol H_2 per mol substrate] are much lower than the theoretical value. Since photosynthetic bacteria can utilize organic acids, byproduct produced by fermentative bacteria, a combination of photosynthetic and

anaerobic bacteria in a two-stage hybrid system may improve the overall yield of hydrogen (Fascetti and Todini 1995; Ueno et al. 2007).

This two-step process has become a subject of considerable interest (Nath and Das 2004). In such a system, dark fermentation was firstly employed to generate H_2 and other intermediates such as low-molecular weight organic acids, which were then converted to hydrogen by photosynthetic bacteria in the second step using a photobioreactor. A yield of 6.6 mol H_2/ mol glucose was achieved by the mixed culture of *C. butyricum* and *Rhodobacter sp.* M-19 (Yokoi et al. 1998). Recently, Yokoi et al. have further improved the hydrogen productivity to 7.2 mol H_2/ mol glucose from starch-manufacturing wastes by repeated batch culture with *C. butyricum*, *E. aerogenes* HO-39, and *Rhodobacter sp.* M-19 (Yokoi et al. 2002; Yokoi et al. 2001).

Nevertheless, integration of multiple processes produces significant challenges for the reactor engineering, system design, process control, and operation and maintenance (Holladay et al. 2009). For example, the synergy of the process lies in the maximum utilization of the substrate. The system should be well-controlled to provide optimum media composition and environmental conditions for the two microbial components of the process (Fascetti et al. 1998; Kapdan and Kargi 2006; Yokoi et al. 2001). The ammonia concentration and C/N ratio in the effluent from the first stage should not inhibit the hydrogen production in the second stage (Fascetti et al. 1998; Lee et al. 2002). Therefore, dilution and neutralization are generally required before photo fermentation in order to adjust the organic acid concentration and the pH level (Kapdan and Kargi 2006). Other challenges of integrating these two stages include 1) adjusting photosynthetic and respiration capacity ratio, 2) co-culture balance, and 3) pretreatment of cell biomass from photo fermentation for the dark fermentation (Holladay et al. 2009).

1.2.3 Newly developed methods

1.2.3.1 Electrohydrogenesis

Electrohydrogenesis is a process in which protons and electrons released by exoelectrogenic bacteria in specially designed reactors (based on modifying microbial fuel cells) are catalyzed to form hydrogen gas through the addition of a small voltage to the circuit (Cheng and Logan 2007b). To substantially increase hydrogen production rate and efficiencies, Cheng *et al.* firstly

developed a compact reactor system using chemically modified three-dimensional graphite granule anode and an anion exchange membrane (2007a). Experiments then demonstrated hydrogen production was dependent on the voltage supplied. The addition of at least 0.11 V was required for hydrogen generation at the cathode according to thermodynamic analysis. A yield of 2.01-3.95 mol H_2 /mol acetic acid was achieved by applying voltages of 0.2 to 0.8 V. At an applied voltage of 0.6 V, the overall energy efficiency, for glucose and cellulose respectively, were 266 % and 268 % based solely on electricity applied, and 64% and 63% if combustion heat was included. The final yields were 8.55 mol [H_2] /mol [glucose] for glucose and 8.20 mol [H_2] /mol [hexose equivalent] for cellulose. The corresponding hydrogen production rates were 2.29 mmol/L/h (= 1.23 $m^3/d/m^3$) and 0.20 mmol/L/h (= 0.11 $m^3/d/m^3$) for glucose and cellulose, respectively (Cheng and Logan 2007b).

By electrohydrogensis, hydrogen can be produced from various types of biodegradable organic matter. This technology is the first that makes possible to produce large yield of hydrogen from domestic waste water (Liu et al. 2005). Sustainable hydrogen production from electrohydrogensis also represents a potential solution to meet the energy requirement of transportation systems. However, the practical application remains challenging. For example, the available anode surface, where bacteria attached, dictates the upper limit of power generation for microbial fuel cells, as well as the upper limit of hydrogen production rate for electrohydrogenesis. More concerted efforts are required to develop a less expensive method not only to provide high surface area for bacterial attachment, but also to support efficient electron transfer to the anode (Logan 2004).

1.2.3.2 Enzymatic hydrogen production by cell-free synthetic enzymatic biotransformation

The production of nearly 12 moles of hydrogen generated from glucose unit was a long-term goal. The goal was not achieved until a new biocatalysis pathway biotransfomation (SyPaB) was developed. SyPaB implements complicated biochemical networks by assembling a number of purified enzymes and coenzymes. It is an open system that can be controlled and regulated easily based upon the inputs, such as temperature, substrate, pressure, and so forth. On the contrary, microorganisms have complicated biochemical network but only a small fraction of enzymes (e.g. <1%) are responsible for bioconversion of a substrate to desired product while most of them

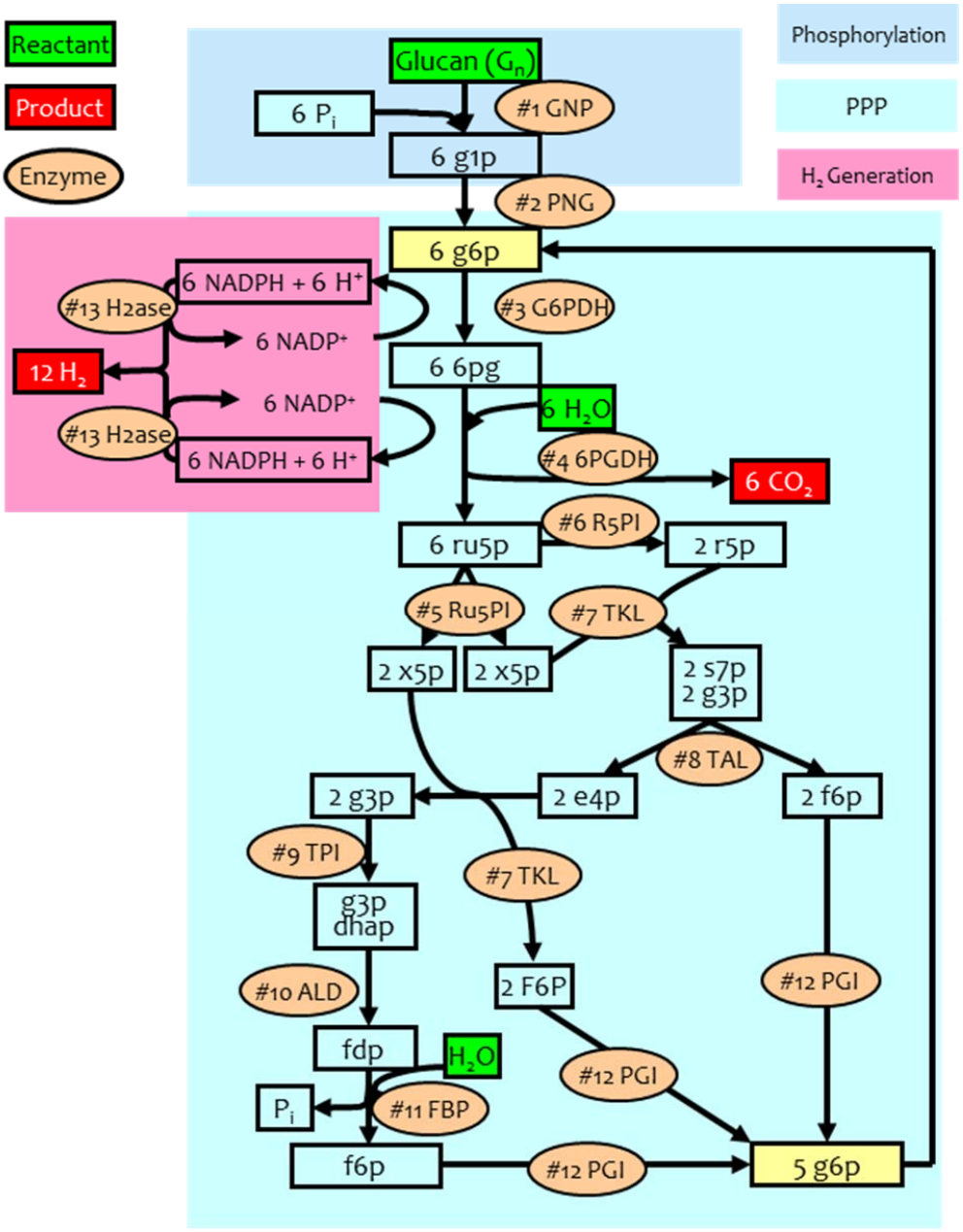


Figure 1.3 The synthetic pathway for conversion of starch and water to hydrogen and carbon dioxide. The abbreviations are: PPP, pentose phosphate pathway; g1p, glucose 1-phosphate; g6p, glucose 6-phosphate; 6pg, 6-phosphogluconate; ru5p, ribulose 5-phosphate; r5p, ribose 5-phosphate; x5p, xyulose 5-phosphate; s7p, sedoheptulose 7-phosphate; e4p, erythrose 4-phosphate; g3p, glyceraldehyde 3-phosphate; dhap, dihydroxyacetonephosphate; f6p, fructose 6-phosphate; fdp, fructose 1,6-bisphosphate; and Pi, inorganic phosphate. The enzymes are: #1 glucan phosphorylase, #2 phosphoglucomutase; #3 g6p dehydrogenase, #4 6-phosphogluconate dehydrogenase, #5 ribulose 5-phosphate epimerase, #6 phosphoribose isomerase, #7 transketolase, #8 transaldolase, #9 Triose phosphate isomerase, #10 aldolase, #11 Fructose 1,6-bisphosphatase, #12 phosphoglucose isomerase, and #13 hydrogenase. Figure is adopted from (Zhang et al. 2007).

are responsible for self-duplication, self-repairing, and non-product-related metabolisms, which inevitably decreases the product yields (Zhang and Mielenz 2011). SyPaB also differs from other *in vitro* synthetic biology systems that synthesize biopolymers from small building blocks (Forster and Church 2007). SyPaB is catabolic wherein biopolymers are completely degraded and the chemical energy in the large molecules is released by a step-wise cascade of enzymatic reactions (Zhang et al. 2007).

Figure 1.3 shows the synthetic enzymatic pathway for conversion of starch and water to hydrogen and carbon dioxide. It contains totally five sub-modules: (i) cellobiose conversion to glucose-1-phosphate (g1p) catalyzed by cellobiose phosphorylase, (ii) glucose-6-phosphate (g6p) generation from G1P catalyzed by phosphoglucomutase, (iii) NADPH production catalyzed by two dehydrogenases of the oxidative phase of the pentose phosphate pathway (PPP), (iv) g6p regeneration from ribulose-5-phosphate catalyzed by the eight enzymes of the non-oxidative phase of PPP, and (v) hydrogen generation from NADPH catalyzed by hydrogenase. The overall cellobiose-to-hydrogen reaction can be summarized as

$$C_{12}H_{22}O_{11}\ (aq) + 7\ H_2O\ (l) \rightarrow C_6H_{12}O_6\ (aq) + 12\ H_2\ (g) + 6\ CO_2 \quad (1)$$

The reaction is a spontaneous and endothermic process (ΔG^o = -48.8 kJ/mol, and ΔH^o = 596 kJ/mol) (Zhang et al. 2007). The gaseous reaction products (H_2 and CO_2) are removed from the aqueous reaction solution, promoting a complete conversion.

At present, the highest carbohydrates-to-hydrogen conversion efficiency was achieved by the synthetic enzymatic pathways. High-yield hydrogen gas production was demonstrated by using glucose 6-phoshphate, starch, cellulosic materials (e.g. cellobiose and cellodextrin) at maximum stoichiometric yields of 50% to 97 % (Woodward et al. 2000; Ye et al. 2009; Zhang et al. 2007) . Within the past two years, overall ~20-fold rate enhancement (from 0.21 mmol/l/h to 3.92 mmol/l/h, Figure 4) has been implemented by increasing the rate-limiting hydrogenase concentration, increasing the substrate concentration, and by elevating the reaction temperature slightly from 30 to 32^{o}C (Ye et al. 2009). So far the production rate of H_2 is higher than that for photobiological systems and comparable to those reported for dark fermentations and electrohydrogensis (Cheng and Logan 2007b; Das and Veziroglu 2001).

To put this new innovative technology into practice, however, several obstacles must be overcome, such as still slow reaction rates, low-cost cellulose utilization, and the high costs of enzymes due to their deactivation, production, and purification. It is anticipated that the synthetic enzymatic hydrogen production rate can be further enhanced by several orders of magnitude with the appropriate optimization. A similar trend has led to a 1,000,000-fold enhancement in the power densities of microbial fuels cells during the past decade (Logan and Regan 2006).

1.3 Concluding remarks

Concerns about energy crisis and climate change have increased interest in hydrogen as fuel. Existing technologies have potential for practical application but they are still in their early stages (Levin et al. 2004). Persistent efforts should be made to increase the hydrogen yield and hydrogen production rates so that biohdyrogen systems could become commercially competitive. Chemical and biological catalysis have been integrated over the last decades. An impressive example is ethanol fermentation followed by ethanol partial oxidation reforming, which resulted in an overall yield of *ca.* 9 moles of hydrogen per mole of glucose (Deluga et al. 2004). Several new technologies are developed, by which carbohydrates-to-hydrogen conversion has been achieved in high yields (> 8 mol H_2/mol hexose equivalent). As compared to other sugar-to-hydrogen approaches, the special features of SyPaB, e.g. the highest hydrogen yield, a mere byproduct (CO_2), and the modest reaction conditions, make it compatible to transportation applications. For practical application, however, several obstacles must be overcome. It is expected that thousands-fold increase in the rate of H_2 synthesis could be achieved by intensive R&D over the next ten years, which would meet the energy requirement for transportation as well as stationary power (Zhang 2010b). With such improvements, this carbohydrate-to-hydrogen technology will address the challenges associated with hydrogen production, storage, safety, distribution, and infrastructure in the hydrogen economy (Zhang 2009).

Chapter 2

Spontaneous high-yield hydrogen production from cellulosic materials and water catalyzed by enzyme cocktails

Material from this chapter appears in Xinhao Ye, Yiran Wang, Robert C. Hopkins, Michael W. W. Adams, Barbara R. Evans, Jonathan R. Mielenz, Y.-H. Percival Zhang: Spontaneous high-yield hydrogen production from cellulosic materials and water catalyzed by enzyme cocktails. *ChemSusChem* 2009, **2**, 149-152, and is reprinted by permission of Wiley-VCH.

2.1 Abstract

Cocktail reception: Biohydrogen is produced in high yield from cellulosic materials and water in a one-pot process catalyzed by up to 14 enzymes and one coenzyme. This assembly of enzymes results in non-natural catabolic pathways. These spontaneous reactions are conducted under modest reaction conditions (32 °C and atmospheric pressure).

2.2 Introduction

Carbon-neutral hydrogen gas is a future energy carrier, especially for the transportation sector (Cho 2004; Jacobson et al. 2005; Schlapbach and Zuttel 2001). Although hydrogen gas can be produced from a number of hydrogen-containing compounds, such as natural gas and water, low-cost hydrogen production from renewable energy sources are in high demand. Production of biohydrogen from less costly and abundant lignocellulosic biomass under modest reaction conditions may be a cost-effective shortcut because it solves a global-scale solar energy collection and storage and has nearly zero net carbon emissions (Zhang 2008b; Zhang 2009).

Numerous carbohydrate-to-hydrogen conversion approaches have been investigated, involving chemical catalysis, biocatalysis, and combinations of both. Chemical catalysis includes gasification (Matsumura et al. 2005), pyrolysis (Navarro et al. 2007), ultra-fast volatilization (Salge et al. 2006), and aqueous phase reforming (Cortright et al. 2002). Most biocatalyzed hydrogen is produced through anaerobic dark fermentation with a theoretical maximum yield of four H_2 per glucose, along with two acetate (Adams and Stiefel 1998). Recently, a bioelectro-chemically-assisted microbial electrolysis cell has been used to convert acetate to hydrogen with help of low-voltage electricity for high-yield hydrogen production (Cheng and Logan 2007b). A theoretical 10 H_2 per glucose could be achieved by a combination of biocatalysis and catalysis involves cellulose hydrolysis and glucose-ethanol fermentation (Zhang et al. 2006b; Zhang and Lynd 2004b), followed by ethanol reforming or partial oxidation reforming (Deluga et al. 2004; Haryanto et al. 2005). However, all of these methods still suffer from much lower hydrogen yields than the theoretical maximum yield of 12 H_2 per glucose.

The feasibility of producing high-yield hydrogen from starch in one reactor has been demonstrated by using an enzyme cocktail containing 13 enzymes (Zhang et al. 2007). But only an approximately half of glucose equivalent of soluble starch, a branched polysaccharide, can be converted to glucose-1-phosphate mediated by starch phosphorylase for hydrogen production. The previously-reported hydrogen yield was only 5.19 moles of hydrogen per glucose equivalent of starch consumed, largely due to incomplete reaction (Zhang et al. 2007). Non-efficient conversion of starch and its limited supplies economically prohibit large-scale hydrogen production through this new approach.

Cellulose, a linear polymer of anhydroglucose, is the most abundant renewable polysaccharides (Hermann 2006; Zhang 2008a; Zhang et al. 2006b). Cellobiose is a dominant product of primary enzymatic cellulose hydrolysis (Zhang and Lynd 2004b), and cellodextrins are prepared in high yields from cellulose or biomass by using mixed acid hydrolysis (Zhang and Lynd 2003).

2.3 Results and Discussion

Our goal was to produce hydrogen in high yield from cellulosic materials and water. To fulfill this task, a new synthetic enzymatic pathway was designed (Fig. 2.1).

The pathway contained five sub-modules: (i) cellobiose conversion to glucose-1-phosphate (g1p) catalyzed by cellobiose phosphorylase, (ii) glucose-6-phosphate (g6p) generation from G1P catalyzed by phosphoglucomutase, (iii) NADPH production catalyzed by two dehydrogenases of the oxidative phase of the pentose phosphate pathway (PPP), (iv) g6p regeneration from ribulose-5-phosphate catalyzed by the eight enzymes of the non-oxidative phase of PPP, and (v) hydrogen generation from NADPH catalyzed by hydrogenase. The overall cellobiose-to-hydrogen reaction can be summarized as

$$C_{12}H_{22}O_{11}\,(aq) + 7\,H_2O\,(l) \rightarrow C_6H_{12}O_6\,(aq) + 12\,H_2\,(g) + 6\,CO_2\,(g) \qquad (1)$$

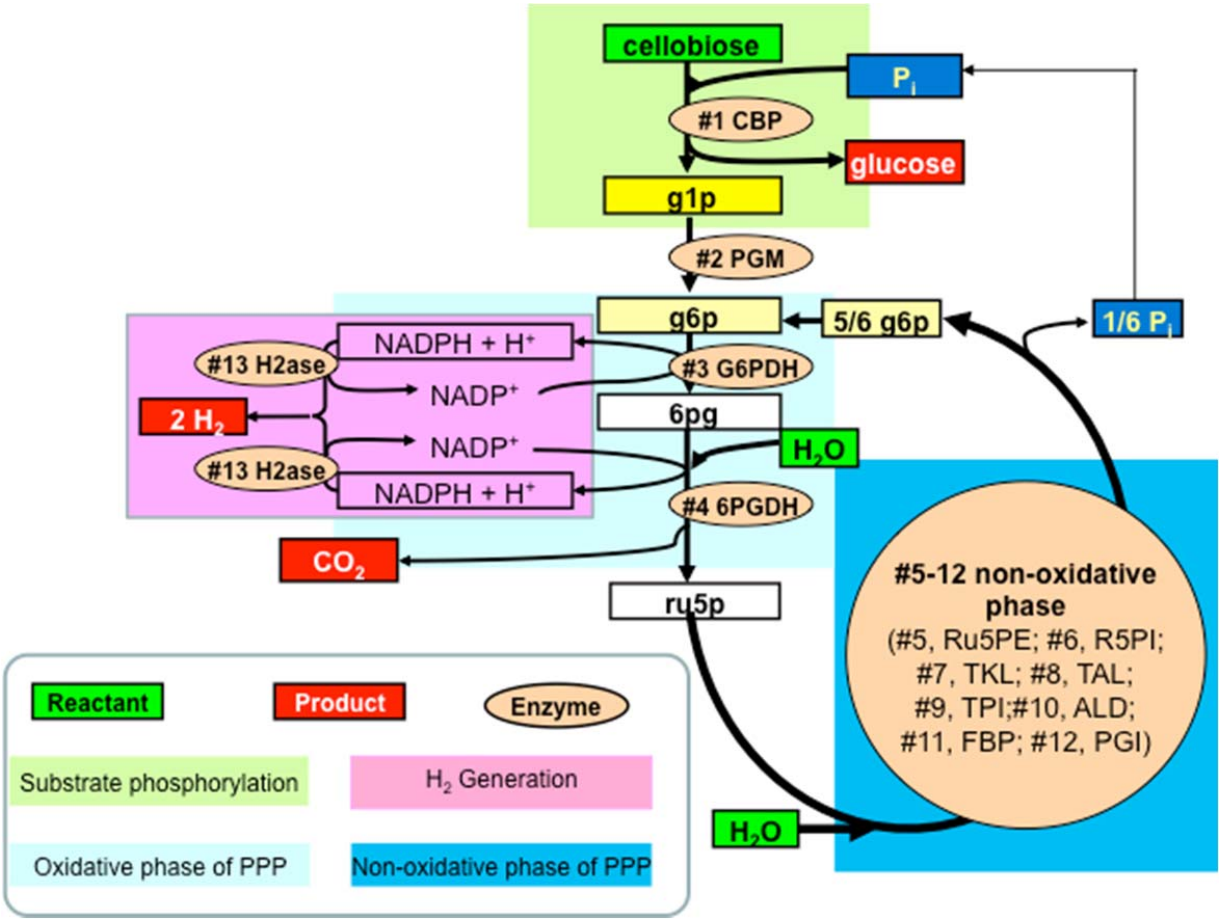


Figure 2.1 Hydrogen production from cellodextrin and water by a synthetic enzymatic pathway. The enzymes are: #1 CBP, cellobiose phosphorylase; #1.1 CDP, cellodextrin phosphorylase #2 PGM, phosphoglucomutase; #3 G6PDH, G-6-P dehydrogenase; #4 6PGDH, 6-phosphogluconate dehydrogenase; #5 R5PI, phosphoribose isomerase; #6 Ru5PE, ribulose 5-phosphate epimerase; #7 TKL, transketolase; #8 TAL, transaldolase; #9 TPI, triose phosphate isomerase; #10 ALD, aldolase; #11 FBP, fructose-1, 6-bisphosphatase; #12 PGI, phosphoglucose isomerase; and #13 H2ase, hydrogenase. The metabolites and chemicals are: g1p, glucose-1-phosphate; g6p, glucose-6-phosphate; 6pg, 6-phosphogluconate; ru5p, ribulose-5-phosphate; x5p, xylulose-5-phosphate; r5p, ribose-5-phosphate; s7p, sedoheptulose-7-phosphate; g3p, glyceraldehyde-3-phosphate; e4p, erythrose-4-phosphate; dhap, dihydroxacetone phosphate; fdp, fructose-1,6-diphosphate; f6p, fructose-6-phosphate; and Pi, inorganic phosphate.

Figure 2.2 shows the profile of H_2 and CO_2 production from a 2 mM cellobiose aqueous solution supplemented with $NADP^+$ and phosphate and catalyzed by 13 enzymes. CO_2 was produced before hydrogen (in the inset figure), in good agreement with the reaction mechanism that the #4

enzyme produces CO_2 and NADPH but the #13 hydrogenase produced hydrogen only when the NADPH concentration was sufficiently high (Ma et al. 1994). The maximal hydrogen production rate was 0.48 mmole H_2/h/L at hour 1.64. After 150 hours, H_2 and CO_2 were not detectable in the constantly-flushing carrier gas N_2. The overall H_2 and CO_2 yields were 11.2 moles of H_2 and 5.64 moles of CO_2 per mole of anhydroglucose unit of cellobiose, which were 93.1% and 94% of the theoretical yields, respectively. The hydrogen yield was comparable to that reported previously (11.6 moles of H_2 per mole glucose-6-phosphate) (Woodward et al. 2000). The slightly less than theoretical value was readily explained by accumulated equilibrium intermediates (e.g., g1p, NADPH) in a batch reaction.

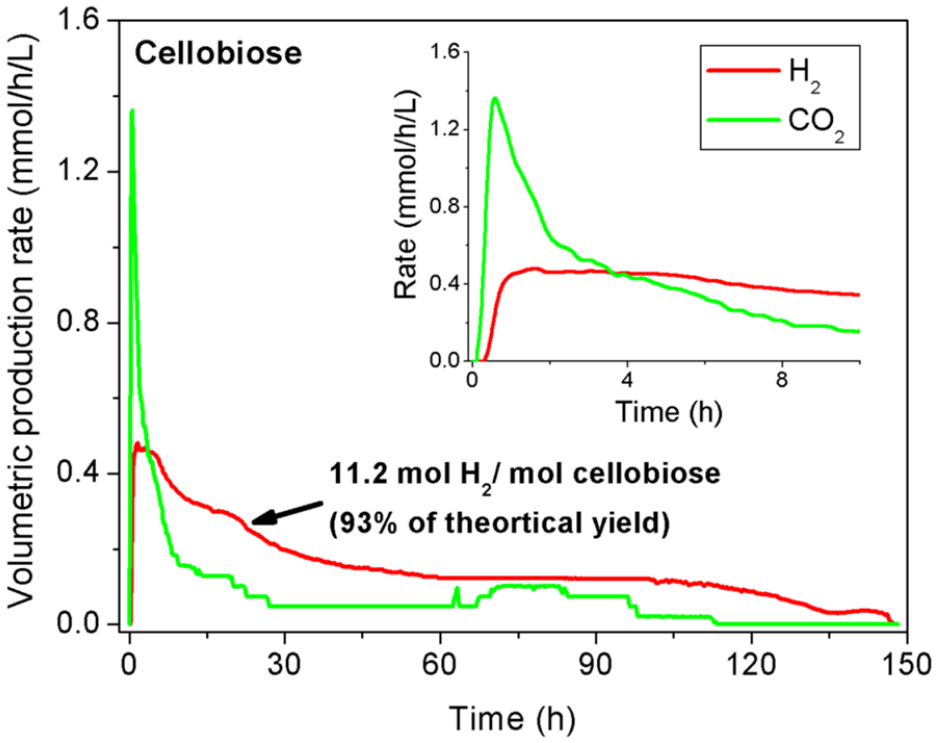


Figure 2.2. Profile of hydrogen and carbon dioxide production from cellobiose and water.

Thermodynamic analysis clearly suggested that this cellobiose-to-hydrogen reaction was spontaneous but endothermic (SOM-2 Fig. I). To our limited knowledge, this reaction was the first chemical reaction that can absorb ambient-temperature heat and convert it to chemical energy that we can utilize, i.e., output/input (chemical energy) > 1. This reaction was spontaneous ($\Delta G < 0$) when the reaction temperatures was higher than 0^oC (SOM-2 Fig. I),

because both gaseous products were released from the aqueous solution under the modest conditions of < 100°C and ~1 atm, accompanied with an entropy gain ($\Delta S >> 0$). Although spontaneous endothermic (entropy-driven) chemical reactions are rare, several examples are reported, such as N_2O_5 (s) → 2 NO_2 (g) + 1/2 O_2 (g).

Removal of the gaseous products from the aqueous phase at low temperature and pressure decreases product inhibition, simplifies product separation, and promotes the nearly complete overall reaction. But only one anhydroglucose (C-6 unit) of cellobiose (C-12) can be used to produce hydrogen so far. The remaining glucose (C-6) may be phosphorylized to glucose-6-phosphate by hexokinase at the cost of one ATP per glucose (Woodward et al. 2000). Unfortunately, utilization of hexokinase and ATP could be economically prohibitive because of (1) costly ATP regeneration system; (2) accumulation of phosphate, an inhibitor of several enzymes (e.g., FBP) (Donahue et al. 2000); (3) Mg^{2+} precipitation due to high phosphate levels (Donahue et al. 2000), since Mg^{2+} is a key co-factor of several enzymes; and (4) a pH shift.

To efficiently utilize glucose equivalent in cellulosic materials, we investigated production of more g1p from longer chain cellulosic fragments catalyzed by cellodextrin phosphorylase (Eq. 2).

$$G_n\ (aq) + (n\text{-}1)\ P_i \rightarrow C_6H_{12}O_6\ (aq) + (n\text{-}1)\ g1p \qquad (2)$$

where n is the degree of polymerization of water-soluble cellodextrins ranging from 3-6 (Zhang and Lynd 2004a; Zhang and Lynd 2005a). With supplement of cellodextrin phosphorylase, the new pathway has a potential to produce more hydrogen (e.g., 12*(n-1)/n) from per glucose equivalent of longer celldextrins (Eq. 3).

$$G_n\ (aq) + 7\ (n\text{-}1)\ H_2O\ (l) \rightarrow C_6H_{12}O_6\ (aq) + 12\ (n\text{-}1)\ H_2\ (g) + 6\ (n\text{-}1)\ CO_2\ (g) \qquad (3)$$

To accelerate hydrogen production rates, we increased hydrogenase loading and substrate concentration because the metabolic mass flux model suggested that the previous hydrogenase loading limited the overall reaction rates, and that the higher substrate concentrate would lead to the higher reaction rates (SOM-2 Fig. II). Figure 2.3 presents the profiles of hydrogen and carbon dioxide production from 2 mM cellopentaose and water. The maximal hydrogen production rate was 3.92 mmole/h/L, 8.2-fold as high as that of cellobiose (Fig. 2.2). The overall H_2 and CO_2

yields were 67.7% and 70.0% of their theoretical yields, respectively, due to the incomplete reaction at hour 150 (Fig. 2.3).

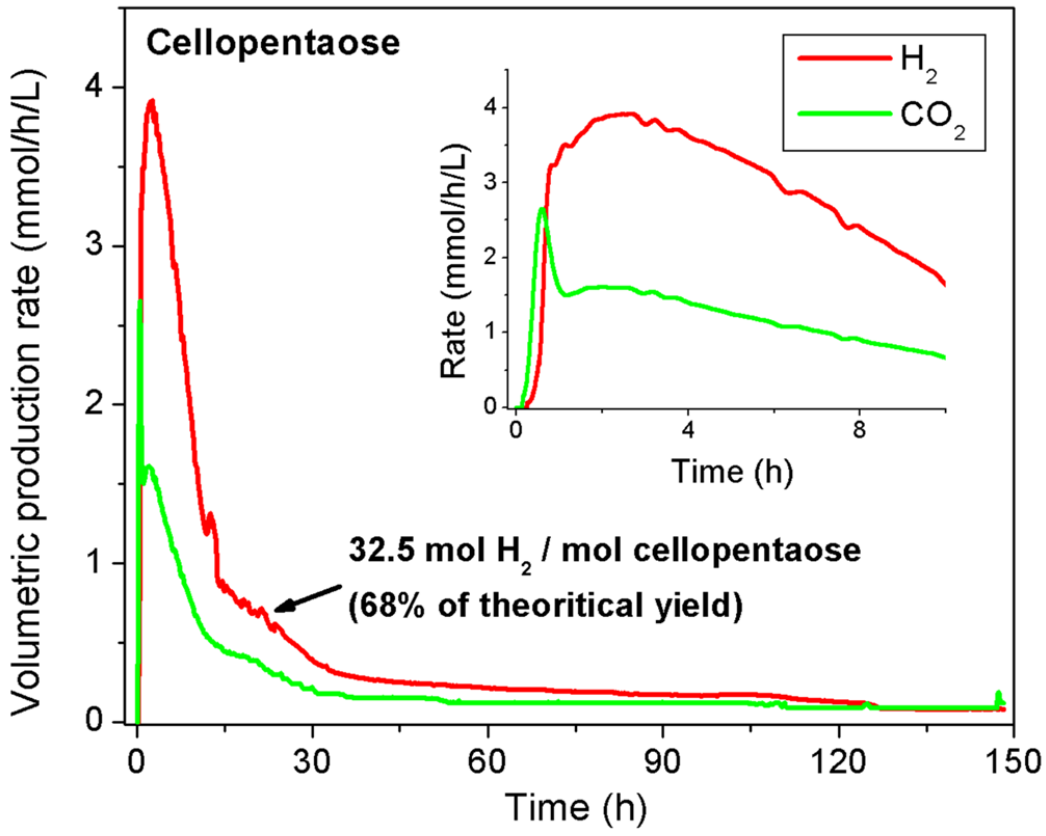


Figure 2.3 Profile of hydrogen and carbon dioxide production from cellopentaose and water.

Two major obstacles must be overcome for future applications, such as high costs of enzymes and slow reaction rates. Enzyme costs can be decreased by use of (hyper)-thermostable enzymes (Woodward et al. 2000; Zhang et al. 2007), enzyme immobilization (Cooney et al. 2008; Moehlenbrock and Minteer 2008), simple enzyme purification (Hong et al. 2008a; Hong et al. 2008b), large-scale production of recombinant protein, and so on. Long-term stability of immobilized thermostable enzymes has been demonstrated industrially. For example, one kg of immobilized glucose isomerase can convert at least 1,500,000 kg of glucose to fructose for several months at ~60^{o}C before its deactivation and replacement (Bhosale et al. 1996; Zhang 2009). It is anticipated that biohydrogen production rate can be accelerated by several orders of magnitude by using a combination of technologies, e.g., (hyper-)thermophilic enzyme

replacement, elevated reaction temperature, optimization of key enzyme ratios, higher substrate concentration, higher enzyme loading, and even metabolite channeling (Conrado et al. 2008; Srivastava and Bernhard 1986; Zhang 2009). For instance, approximately 1,000,000-fold enhancement in the power densities of microbial fuels cells has occurred during the past decade (Logan and Regan 2006). In this study, we have increased the hydrogen production rate by 8.2 fold as compared to our previous results on starch (Zhang et al. 2007) by increasing the rate-limiting hydrogenase concentration, increasing the substrate concentration, and by elevating the reaction temperature slightly from 30 to 32^{o}C. The overall ~20-fold rate enhancement has been implemented in the past two years (Table 2.1).

Table 2.1 Summary of enzymatic hydrogen production rates and yield

Substrate	**Conc. [a] (mM)**	**Temp (^{o}C)**	**$V_{max,H2}$ (mmol/h/L)**	**Y_{H2} (%)**	**References**
G-6-P	2	30	0.21	96.7	(Woodward et al. 2000)
G-6-P	2	30	0.73 [b]	70 [d]	(Zhang et al. 2007)
Starch	1	30	0.48	43 [d]	(Zhang et al. 2007)
Cellobiose	2	32	0.48	93.3	This study
Cellopentaose	8	32	3.92 [c]	67.7 [d]	This study

[a], potential of g6p.

[b], changes compared to the Woodward et al. (2000) experiment are a reduction in buffer salt concentration and replacement of FBP with a recombinant form.

[c], changes compared to the previous results (Zhang et al. 2007) are different combinations of substrate and enzyme, substrate concentration, and concentrations of key enzymes.

[d], incomplete batch reaction.

In conclusion, the entropy-driven reactions mediated by enzyme cocktails have several unique features: (1) a low-temperature endothermic reaction that adsorb low-temperature thermal energy for producing high-quality chemical energy (hydrogen), (2) a very high demonstrated hydrogen yield (~11.2 H_2/anhydroglucose unit of cellulosic materials) in a batch reaction, and (3) a ~10-fold enhancement in hydrogen production rate than before (Zhang et al. 2007).

2.4 Experimental Section

Enzymes and their preparation. All the enzymes are listed in SOM-2 Table 2. Except for the enzymes purchased from Sigma, the recombinant enzymes were prepared as described below. The *cbp*, *cdp*, and *pgm* genes of the thermophilic bacterium *C. thermocellum* ATCC 27405 were amplified by PCR and inserted into the T7-protein expression plasmid pCIG (Hong et al. 2008a). The recombinant proteins were expressed in *E. coli* BL21 rosetta. The recombinant protein was purified by affinity adsorption on RAC followed by intein self-cleavage (Hong et al. 2008a). The recombinant *E. coli* FBP was expressed and purified as described previously (Donahue et al. 2000). The *P. furious* hydrogenase I was prepared as described elsewhere and stored -80 ^{o}C (Ma et al. 1994). The reaction buffer contained the enzymes (SOM-2 , Table 3), 0.5 mM thiamine pyrophosphate, 2 mM $NADP^{+}$, 10 mM MgCl2, and 0.5 mM MnCl2 in a 0.1 M HEPES buffer (pH 7.5). Five mM of 1,4-dithiotreitol (DTT) was added for the cellopentaose experiment.

Experimental conditions and data analysis. The experiments were carried out in a continuous flow system as described previously (Zhang et al. 2007), with the modification that (1) ultra-pure nitrogen (Air Liquide America Corp., Houston, TX) was used as a carrier gas and (2) reaction temperature was increased to 32^{o}C. Hydrogen was measured with a Figaro TGS 822 tin oxide sensor (Figaro Engineering Inc., Osaka, Japan) and carbon dioxide was measured with a LI-COR CO2 Analyzer Model LI-6252 (LI-COR Biosciences Lincoln, NE). Data collection was conducted with the LabView program (National Instruments Corp., Austin, TX) and data analysis was carried out by the MatLab program (Natick, MA).

2.5 Supplementary Materials

I. Calculation of thermodynamics properties

The overall reaction from cellobiose to hydrogen, carbon dioxide, and glucose is

$C_{12}H_{22}O_{11}$ (aq) + 7 H_2O (l) → $C_6H_{12}O_6$ (aq) + 6 CO_2 (g) + 12 H_2 (g) (R1)

This reaction can be written as the consequence of two below reactions

$C_{12}H_{22}O_{11}$ (aq) + H_2O (l) → 2 $C_6H_{12}O_6$ (aq) (R2)

$C_6H_{12}O_6$ (aq) + 6 H_2O (l) → 2 CO_2 (g) + 12 H_2 (g) (R3)

Thermodynamic parameters of Reaction 2 -- hydrolysis reaction of cellobiose at 298.15 K are $\Delta H_2^o = -2.43 \text{ kJ/mol}$, $\Delta S_2^o = 33.8 \text{ J/mol}$, and $\Delta G_2^o = -12.5 \text{ kJ/mol}$ (Tewari and Goldberg 1989). Thermodynamic parameters of Reaction 3 were calculated based on the thermo-chemical properties of reactants as well as products. At the standard states, all the data were listed in SOM Table I-1 from the NIST website (Chase 1998; NIST) listed in.

SOM-2 Table I. Basic properties of hydrogen, carbon dioxide, water and glucose.

Compound	$\Delta_f H^o$ (kJ/mol)	S^o (J/mol)
$C_6H_{12}O_6$ (l)	-1271.1	209.19
H_2O (l)	-285.83	69.95
H_2 (g)	0	130.68
CO_2 (g)	-393.52	213.79

Thus, the enthalpy and entropy of Reaction 3 are

$$\Delta H_3^o = \sum_{\text{product}} \Delta_f H^o - \sum_{\text{reac tan t}} \Delta_f H^o = 626.48 \text{ kJ/mol}$$

$$\Delta S_3^o = \sum_{\text{product}} S^o - \sum_{\text{reac tan t}} S^o = 2221.97 \text{ J/mol}$$

Since $\Delta G = \Delta H - T \cdot \Delta S$, the standard Gibbs free energy (ΔG^o_3) of Reaction 3 was

$$\Delta G_3^o = \Delta H_3^o - T \cdot \Delta S_3^o = -36.38 \text{ kJ/mol}$$

Therefore, thermodynamic properties of Reaction 1 are:

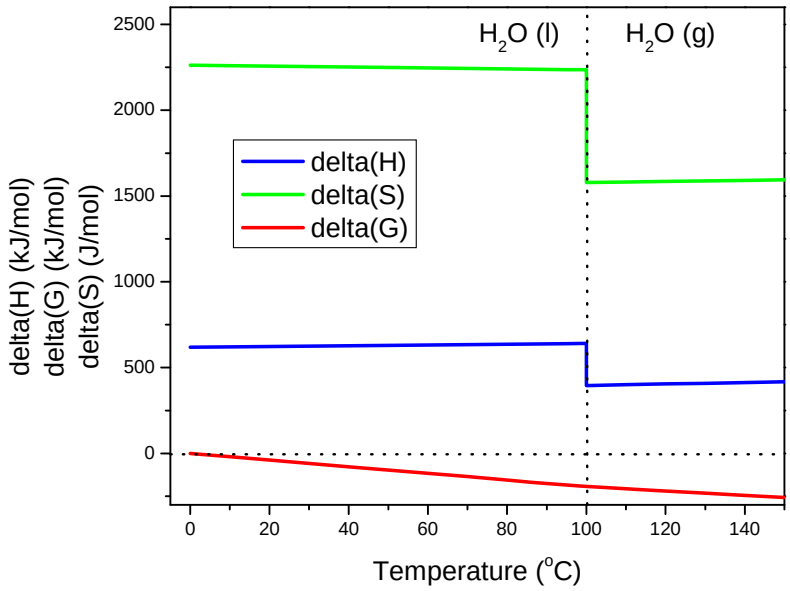
$$\Delta H_1^0 = \Delta H_2^0 + \Delta H_3^0 = 624.05 \text{ kJ/mol}$$

$$\Delta S_1^0 = \Delta S_2^0 + \Delta S_3^0 = 2255.77 \text{ J/mol}$$

$$\Delta G_1^0 = \Delta G_2^0 + \Delta G_3^0 = -48.88 \text{ J/mol}$$

Clearly, Reaction 1 is an endothermic ($\Delta H > 0$) spontaneous ($\Delta G < 0$) entropy-driven ($\Delta S >> 0$) reaction.

We further calculated the properties of Reaction 1 in terms of reaction temperature at 1 atm. Enthalpy change (ΔH) and Gibbs free energy change (ΔG) of Reaction #2 and Reaction #3 can be calculated according to the information (Tewari and Goldberg 1989) and (Chase 1998), respectively. The heat capacity of glucose was calculated as described elsewhere (Douglas et al. 1951; Ginnings et al. 1950); the temperature-dependent enthalpy of glucose can be calculated as before (Douglas et al. 1951); the entropy of glucose can be calculated (Boerio-Goates 1991).



SOM-2 Figure 1. Profiles of enthalpy, entropy, and Gibbs free energy of Reaction 1 in terms of temperature at 1 atm.

SOM-2 Figure 1 shows the property profile of Reaction 1 in terms of temperature at a fixed pressure (1 atm). Clearly, with an increase in temperature, the Gibbs free energy decreased gradually, suggesting that this spontaneous reaction had more potential for completion. When the reaction temperature is higher than 100^{o}C, water vapor will be mixed with the products, resulting in difficulty in product/reactant separation.

II. Model Simulation of in vitro Synthetic Enzymatic Pathway for Novel Hydrogen Production

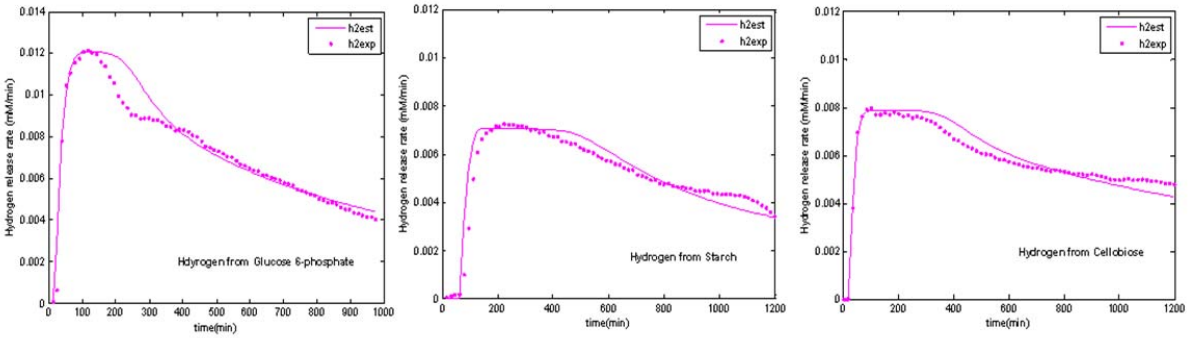
A mathematical model was developed to simulate hydrogen production from glucose 6-phosate (Zhang et al. 2007), starch (Zhang et al. 2007), and cellobiose. The model consisted of the kinetic equations of the individual reactions and a set of mass balance equations for the different concentrations of each metabolite. The kinetics was based on a modified two-substrate Michaelis-Menten equation except three reactions catalyzed by TKL, TAL, and TPI (SOM-2 Table 2). Near-equilibrium conditions were assumed for these reactions owing to the respective relatively high Michaelis constant K_m (Vaseghi et al. 1999). The mass balances take the following form

$$\frac{dC_i}{dt} = \sum_i v_{ij} r_j$$

where C_i denotes the concentration of metabolite i, and v_{ij} is the stoichiometric coefficient for the metabolite i in reaction j, the rate of which is $r_{j.}$ Furthermore sensitive analysis is performed based on flux control coefficient (or response coefficient) which was defined as the percentage change in the flux caused by a 1% modulation of the enzyme activity. All in all, the model containing ODEs was solved by Runge-Kutta method and parameters were estimated using derivative-free method, both of which were performed in Matlab 7.0.

A comparison of hydrogen profiles between simulation and experimental data was shown in SOM-2 Figure 2. The parameters used for computer simulations were listed in SOM-2 table 2, which were estimated based on the curve-fitting of the CO_2 release rates. The model fit the observed trends reasonably well. However, the largest deviations between measured and predicted value were found after the peak appeared around 200 min, especially for the reactions

driven from glucose 6-phosphate. Neither changes in the enzyme loading nor variation in parameter settings leaded to the improvement of the fitting. They might be impaired by the model assumptions that the kinetic parameters kept unchanged if the same enzyme was employed in these three reactions, and that the difference of hydrogen production rate was mainly contributed by the variation of enzyme loading, enzyme activity, and substrates. On the other hand, the remaining deviations between model and measured data also indicated the present limitations in comprehensive modeling bases on mechanistic rate equations (Chassagnole et al. 2002).



SOM-2 Figure 2. Comparison between experimental observation and model predictions for hydrogen evolution rate (mM/min). Left-hand figure illustrates the time curve for G6P reaction, the middle figure is for starch reaction, and the right-hand figure presents the hydrogen evolution from cellobiose. ■: experimental observation; lines: model predictions.

.

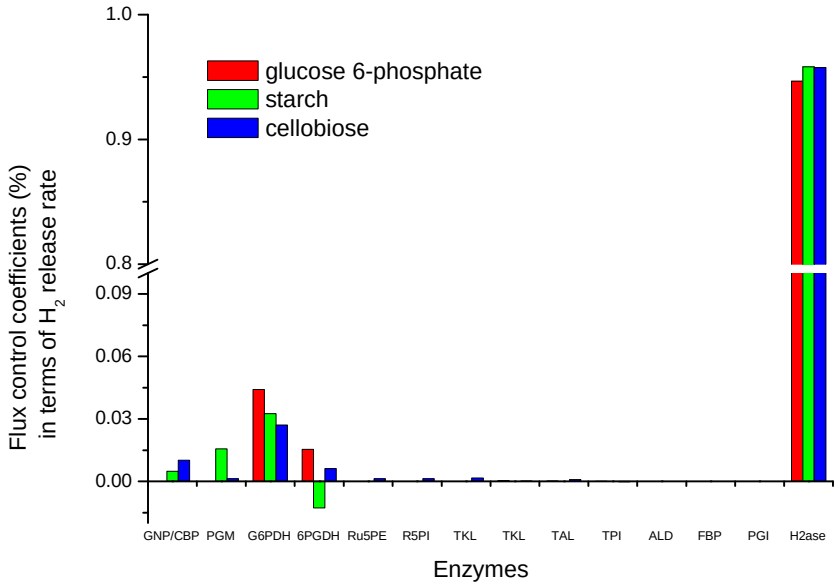
SOM-2 Table 2 Kinetic expression and parameters

No	Equation	Parameters	Liter. Data	Estimation		
				G6P	STARCH	CELLOBIOSE
1	$r_1 = r^{GNP} = \frac{k^{GNP}[GNP][G_N][P_i]}{\left[K^{GNP}_{m,Gn}(1+\frac{[P_i]}{K^{GNP}_{I,Pi}})+[G_N]\right]\times\left[K^{GNP}_{m,Gn}+[P_i]\right]}$	$V^{GNP}_{\max}$	-	-	0.10	-
		K^{GNP}_{m,G_n}	0.15 (Mosi and Withers 1999)	-	0.25	-
		K^{GNP}_{m,P_i}	2.4 (Mosi and Withers 1999)	-	2.4	-
		K^{GNP}_{i,P_i}	0.14 (Mosi and Withers 1999)	-	0.12	-
	$r_1 = r^{CBP} = \frac{k^{CBP}[CBP][G_2][P_i]}{\left[K^{CBP}_{m,G_2}(1+\frac{[glc]}{K^{CBP}_{I,glc}})+[G_2]\right]\times\left[K^{CBP}_{m,G_2}+[P_i]\right]}$	$V^{CBP}_{\max}$	-	-	-	0.43
		K^{CBP}_{m,G_2}	7.3(Alexander 1986)	-	-	7.3
		K^{CBP}_{m,P_i}	2.9(Alexander 1986)	-	-	2.9
		$K^{CBP}_{i,glc}$	1.2(Alexander 1986)	-	-	1.2
2	$r_2 = r^{PGM} = \frac{k^{PGM}[PGM]\left([g1p]-[g6p]/K^{PGM}_{eq}\right)}{K^{PGM}_{m,g1p}\left(1+\frac{[g6p]}{K^{PGM}_{I,g6p}}\right)+[g1p]}$	$V^{PGM}_{\max}$	-	-	0.13	0.81
		$K^{PGM}_{m,g1p}$	0.008(Lowry and Passonneau 1969)	-	8.5	0.41
		$K^{PGM}_{m,g6p}$	0.047(Lowry and Passonneau 1969)	-	47	0.50
		K^{PGM}_{eq}	7(Lowry and Passonneau 1969)	-	4	7
3	$r_3 = r^{G6PDH} = \frac{k^{G6PDH}[G6PDH][g6p][NADP]}{\left(K^{G6PDH}_{m,g6p}+[g6p]\right)*\left[K^{G6PDH}_{m,NADP}\left(1+\frac{[NADPH]}{K^{G6PDH}_{I,NADPH}}\right)+[NADP]\right]}$	$V^{G6PDH}_{\max}$	-	0.0287	0.03	0.4
		$K^{G6PDH}_{m,g6p}$	0.034(Horne et al. 1970)	0.0374	0.0374	0.0374
		$K^{G6PDH}_{m,NADP}$	0.019(Zhang and Lynd 2004b)	0.0223	0.0223	0.0223
		$K^{G6PDH}_{I,NADPH}$	1.7(Vaseghi et al. 1999)	0.101	0.101	0.101

4	$r_4 = r^{6PGDH} = \frac{k^{6PGDH}[6PGDH][6pg][NADP]}{\left(K_{m,6pg}^{6PGDH} + [6pg]\right) * \left[K_{m,NADP}^{6PGDH}\left(1 + \frac{[NADPH]}{K_{I,NADPH}^{6PGDH}}\right) + [NADP] \right]}$	$V_{\max}^{6PGDH}$	-	0.02	0.028	0.13
		$K_{m,6pg}^{6PGDH}$	0.153 (Rendina et al. 1984)	0.274	0.2749	0.2749
		$K_{m,NADP}^{6PGDH}$	0.025 (Rendina et al. 1984)	0.032	0.032	0.032
		$K_{I,NADPH}^{6PGDH}$	0.055 (Vaseghi et al. 1999)	3.37	3.37	0.082
5	$r_5 = r^{Ru5PE} = \frac{k^{Ru5PE}[Ru5pE] * \left([ru5p] - [x5p] / K_{eq}^{Ru5PE}\right)}{K_{m,ru5p}^{Ru5PE} + [ru5p]}$	$V_{\max}^{Ru5PE}$	-	0.0163	0.03	0.20
		$K_{m,ru5p}^{Ru5PE}$	-	0.828	0.828	0.828
		K_{eq}^{Ru5PE}	4(Vaseghi et al. 1999)	3.46	3.46	3.46
6	$r_6 = r^{R5PI} = \frac{k^{R5PI}[R5PI] * \left([ru5p] - [r5p] / K_{eq}^{R5PI}\right)}{K_{m,ru5p}^{R5PI} + [R5PI]}$	$V_{\max}^{R5PI}$	-	0.0308	0.05	0.10
		$K_{m,ru5p}^{R5PI}$	3.3(Woodruff and Wolfenden 1979)	3.075	3.075	3.075
		K_{eq}^{R5PI}	1.4(Vaseghi et al. 1999)	0.25	0.25	0.25
7	$r_7 = r^{TKLa} = k^{TKLa}[TKL] * \left([x5p][r5p] - [s7p][g3p] / K_{eq}^{TKLa}\right)$	$V_{\max}^{TKLa}$	-	0.0392	0.05	0.09
		K_{eq}^{TKLa}	1.2 (Chassagnole et al. 2002)	0.994	0.994	0.994
8	$r_8 = r^{TKLb} = k^{TKLb}[TKL] * \left([x5p][e4p] - [f6p][g3p] / K_{eq}^{TKLb}\right)$	$V_{\max}^{TKLb}$	-	0.0193	0.0193	0.1
		K_{eq}^{TKLb}	10(Chassagnole et al. 2002)	10.3	10.4	10.3
9	$r_9 = r^{TAL} = k^{TAL}[TAL] * \left([xs7p][g3p] - [f6p][e4p] / K_{eq}^{TAL}\right)$	$V_{\max}^{TAL}$	-	0.0729	0.035	0.08
		K_{eq}^{TAL}	1.05(Vaseghi et al. 1999)	1.2744	1.27	1.27
10	$r_{10} = r^{TPI} = -\frac{k^{TPI}[TPI] * \left([g3p] - [dhap] * K_{eq}^{TPI}\right)}{K_{m,g3p}^{TPI}\left(1 + \frac{[dhap]}{K_{m,dhap}^{TPI}}\right) + [g3p]}$	$V_{\max}^{TPI}$	-	0.0304	0.0304	0.1
		$K_{m,dhap}^{TPI}$	-	1.51	1.51	1.51
		$K_{m,g3p}^{TPI}$	3.28(Greenberg et al. 1972)	3.53	3.53	3.53
		K_{eq}^{TPI}	0.045(Noltmann 1972)	0.0543	0.0543	0.0543

11	$r_{11} = r^{ALD} = \frac{k^{ALD}[ALD] * \left([g3p][dhap] - [fdp]/K_{eq}^{ALD}\right)}{\left(K_{m,g3p}^{ALD}(1 + \frac{[P_i]}{K_{I,Pi}^{ALD}}) + [g3p]\right)\left(K_{m,dhap}^{ALD}(1 + \frac{[P_i]}{K_{I,Pi}^{ALD}}) + [dhap]\right)}$	$V_{\max}^{ALD}$	-	0.047	0.047	0.16
		$K_{m,g3p}^{ALD}$	1(Morse and Horecker 1968)	0.790	0.7899	0.790
		$K_{m,dhap}^{ALD}$	2(Morse and Horecker 1968)	1.48	1.48	1.48
		K_{I,P_i}^{ALD}	-	6.70	6.70	6.70
		K_{eq}^{TPI}	0.03(Lehrer and Barker 1970)	0.0241	0.0241	0.0241
12	$r_{12} = r^{FBP} = \frac{k^{FBP}[FBP] * (fdp)}{K_{m,}^{FBP}(1 + \frac{[P_i]}{K_{I,Pi}^{FBP}}) + [fdp]}$	$V_{\max}^{FBP}$	-	0.032	0.05	0.16
		$K_{m,fdp}^{FBP}$	0.035(Donahue et al. 2000)	0.0434	0.0434	0.0434
		K_{I,P_i}^{FBP}	0.35(Donahue et al. 2000)	0.480	0.480	0.480
13	$r_{13} = r^{PGI} = \frac{k^{PGI}[PGI] * \left([f6p] - \frac{[g6p]}{K_{eq}^{PGI}}\right)}{K_{m,f6p}^{PGI}\left(1 + \frac{[g6p]}{K_{m,g6p}^{PGI}}\right) + [f6p]}$	$V_{\max}^{PGI}$	-	0.0329	0.0329	0.16
		$K_{m,f6p}^{PGI}$	0.167(Marchand et al. 1989)	0.159	0.159	0.159
		$K_{m,g6p}^{PGI}$	0.3(Manfred Rizzi 1997)	0.198	0.198	0.198
		K_{eq}^{PGI}	5(Manfred Rizzi 1997)	6.27	6.271	6.271
14	$r_{14} = r^{H2ase} = \frac{k^{H2ase}[H2ase] * \left([NADPH] - [NADP]/K_{eq}^{H2ase}\right)}{K_{m,NADPH}^{H2ase}\left(1 + \frac{[NADP]}{K_{I,NADP}^{H2ase}}\right) + [NADPH]}$	K^{H2ase}	-	0.0128	0.0075	0.0084
		$K_{m,NADPH}^{H2ase}$	-	1.74	1.1	1.1
		$K_{I,NADPH}^{H2ase}$	0.063(Ma et al. 2000)	0.0589	0.0589	0.589
		K_{eq}^{H2ase}	-	15.1	15.1	15.0

One of the goals of our dynamic model was to identify the rate-limiting steps in the synthetic pathway as well as to provide new insights in further enhancement in enzymatic hydrogen production rate and key enzyme properties. Flux control coefficients (FCC) in terms of hydrogen release rate were determined followed by the process simulation. The results were shown in SOM Figure II-2. The highest control was exerted by hydrogenase (#14) as indicated by FCC = 0.95, 0.96, and 0.96 for glucose 6-phosphate reaction, starch reaction, and cellobiose reaction, respectively. Although enzyme loading of hydrogenase was much higher than others (approximately 70 fold), its exhibiting activity was expected to be very low because residual activity of hyper-thermophilic hydrogenase was less than 1 U at 30^{o}C. The low hydrogenase activity not only limited the hydrogen production but also resulted in the accumulation of cofactor NADPH which was a strong inhibitor to other enzymes, such as G6PDH, 6PGDH and hydrogenase itself.



SOM-2 Figure 3. Flux control coefficients of the enzyme activities on maximum hydrogen production rate.

As a result, the hydrogen production rate can be increased simply by increasing the hydrogenase activity. By *in silico* simulation, 2 fold and 5 fold increases in hydrogenase activity could enhance hydrogen production rates by more than 80% and 150%, respectively.

III. Enzyme information.

SOM-2 Table 3. The designated number, enzyme catalogue number, full name and abbreviated name of the enzymes as well as their source, preparation, and units used in the experiments.

		Enzyme				Cellobiose (IU/reactor)		Cellopentaose (IU/reactor)	
NO	**EC**	Full name	Abb.	**Source**	**Vendor**	Added*	T-Adj.**	Added*	T-Adj.**
1	2.4.1.20	cellobiose phosphorylase	CBP	*C. thermocellum*	lab	8 [a]	8	4 [a]	4
1.1	2.4.1.49	cellodextrin phosphorylase	CDP	*C. thermocellum*	lab			0.5 [b]	0.5
2	5.4.2.2	phosphoglucomutase	PGM	*C. thermocellum*	lab	10 [c]	10	4 [c]	4
3	1.1.1.49	glucose 6 phosphate dehydrogenase	G6PDH	*S. cerevisiae*	Sigma	5 [d]	8	5 [d]	8
4	1.1.1.44	6-phosphogluconic dehydrogenase	6PGDH	*S. cerevisiae*	Sigma	1 [e]	0.38	1 [e]	0.38
5	5.1.3.1	ribulose 5-phosphate 3-epimerase	Ru5PE	*S. cerevisiae*	Sigma	1 [f]	1.6	1 [f]	1.6
6	5.3.1.6	ribose 5-phosphate isomerase	Ru5PI	*spinach*	Sigma	1 [g]	1	1 [g]	1
7	2.2.1.1	Transketolase	TKL	*E. coli*	Sigma	1 [h]	1.6	1 [h]	1.6
8	2.2.1.2	Transaldolase	TAL	*S. cerevisiae*	Sigma	1 [i]	1.6	1 [i]	1.6
9	5.3.1.1	triose-phosphate isomerase	TPI	*rabbit muscle*	Sigma	4.8 [j]	7.68	4.8 [j]	7.68
10	4.1.2.13	aldolase	ADL	*rabbit muscle*	Sigma	1 [k]	1.6	1 [k]	1.6
11	3.1.3.11	fructose-bisphosphatase	FBPase	*E. coli*	lab	1 [l]	1.6	1 [l]	1.6
12	5.3.1.9	phosphoglucose isomerase	PGI	*S. cerevisiae*	Sigma	1 [m]	1.6	1 [m]	1.6
13	1.12.1.3	*P. furiosus* hydrogenase I	H2ase	*P. furisous*	lab	~100 [n]	3.65	~200 [n]	7.2

* the enzyme activities were determined based on international unit (IU) definition under their optimal or experimental conditions.

** the T-adjusted enzyme activities were estimated based on the rule of thumb for the relationship between the enzyme activity and reaction temperature (doubled enzyme activity per ten degree increase) (Berg et al. 2007; Zhang and Lynd 2004b). The real enzyme activities cannot be measured exactly due to the presence of other factors (enzymes, cofactors, and so on) (Teusink et al. 2000).

[a], the CBP activity was measured by a discontinuous assay coupled with glucose 6-phosphate dehydrogenase from glucose HK kit (Sigma, MO, US). Formation of glucose 6-phosphate from glucose 1-phosphate was measured by monitoring NADPH formation. The initial reaction was conducted at 32^{o}C with 2 mM cellobiose, 4 mM phosphate, 10 mM Mg^{2+}, 0.5 mM Mn^{2+}, 20 μg/ml phosphoglucomutase (~10 U/ml), and 1 mg/ml BSA in 50 mM HEPES buffer, pH7.5, and stopped by heating the samples in boiling water bath for 5 min. Coupled with glucose 6-phosphate dehydrogenase, the formation of NADPH was then determined at the 2nd reaction by measuring the change in absorbance at 340 nm for 15 min. One unit of CBP is defined as the amount of enzyme that generated 1 μmol of NADPH per min.

[b]. the CDP activity was determined similarly with CBP assay, except using 2 mM cellopentaose rather than cellobiose, and pulsing 5 mM DTT. One unit of CDP is defined as the amount of enzyme that generated 1 μmol of NADPH per min under the conditions described.

[c]. the PGM activity was assayed by the discontinuous assay coupled with glucose 6-phosphate dehydrogenase from glucose HK kit (Sigma, MO, US). The initial reaction was performed at 32 ^{o}C with 50 mM HEPES buffer (pH7.5) with 5 mM glucose 1-phosphate, 10 mM Mg^{2+}, 0.5 mM Mn^{2+} and 1 mg/ml BSA. The reaction was stopped by heating the samples in boiling water bath for 5 min. Then the product glucose 6-phosphate was detected by glucose HK kit. One unit of PGM is defined as the amount of enzyme that generated 1 μmol of NADPH per min under the above conditions.

[d]. one unit of G6PDH can oxidize 1.0 μmole of glucose 6-phosphate to 6-phospho-D-gluconate per min in the presence of NADP at pH 7.4 at 25°C.

[e]. one unit of 6PGDH can oxidize 1.0 μmole of 6-phospho-D-gluconate to D-ribulose 5-phosphate and CO_2 per min at pH 7.4 at 37°C in the presence of $NADP^{+}$.

[f]. one unit of Ru5PE can convert 1 μmole of D-ribulose 5-phosphate to xylulose 5-phosphate per min at pH 7.7 at 25°C when coupled with transketolase, α-glycerophosphate dehydrogenase, and triosephosphate isomerase

[g]. one unit of Ru5PI can convert 1.0 μmole of D-ribose 5-phosphate to D-ribulose 5-phosphate per min at pH 7.7 at 30 °C.

[h]. one unit of TKL can produce 1 μmol of glyceraldehyde-3-phosphate from xylulose-5-phosphate per minute at pH 7.7 and 25°C, in the presence of ribose-5-phosphate, thiamine pyrophosphate and Mg^{2+}.

[i]. One unit of TAL can produce 1.0 μmole of D-glyceraldehyde 3-phosphate from D-fructose 6-phosphate per min in the presence of D-erythrose 4-phosphate, at pH 7.7 at 25°C in a coupled system with GDH/TPI and β-NADH.

[j]. One unit of TPI can convert 1.0 μmole D-glyceraldehyde 3-phosphate to dihydroxyacetone phosphate per min at pH 7.6 at 25 °C.

[k]. One unit of ADL can convert 1.0 μmole of fructose 1,6-diphosphate to dihydroxyacetone phosphate and glyceraldehyde 3-phosphate per min at pH 7.4 at 25 °C.

[l]. One unit of FBP can produce 1 μmol of phosphate in 1 min from fructose-1,6-biphosphate as described before (Donahue et al. 2000).

[m]. One unit of PGI can convert 1.0 μmole of D-fructose 6-phosphate to D-glucose 6-phosphate per min at pH 7.4 at 25 °C.

[n]. One unit of hydrogenase I can 1μmole of H_2 per min at 80 °C, as described elsewhere (Bryant and Adams 1989).

Chapter 3

Thermophilic *α*-glucan phosphorylase from *Clostridium thermocellum*: Cloning, characterization and enhanced thermostability

Material from this chapter appears in Xinhao Ye, Joe Rollin, Y.-H. Percival Zhang: Thermophilic *α*-glucan phosphorylase from *Clostridium thermocellum*: Cloning, characterization and enhanced thermostability. *J Mol Catal B -Enzym* 2010, **65**, 110-116, and is reprinted by permission of Elservier.

3.1 Abstract

ORF Cthe0357 from the thermophilic bacterium *Clostridium thermocellum* ATCC 27405 that encodes a putative *α*-glucan phosphorylase (*α*GP) was cloned and expressed in *E. coli*. The protein with a C-terminal His-tag was purified by Ni^{2+} affinity chromatography; the tag-free protein obtained from a cellulose-binding module-intein-*α*GP fusion protein was purified through affinity adsorption on amorphous cellulose followed by intein self-cleavage. Both purified enzymes had molecular weights of ca. 81,000 and similar specific activities. The optimal conditions were pH 6.0-6.5 and 60 ^{o}C for the synthesis direction and pH 7.0-7.5 and 80 ^{o}C for the degradation direction. This enzyme had broad substrate specificities for different chain length dextrins and soluble starch. The thermal inactivation of this enzyme strongly depended on temperature, protein concentration, and certain additives that were shown previously to benefit the protein thermostability. The half lifetime of 0.05 mg αGP/mL at 50 ^{o}C was extended by 45 fold to 90 hours through a combined addition of 0.1 mM Mg^{2+}, 5 mM DTT, 1% NaCl, 0.1% Triton X-100, and 1 mg/mL BSA. The enzyme with prolonged stability would work as a building block for cell-free synthetic enzymatic pathway biotransformations, which can implement complicated biocatalysis through assembly of a number of enzymes and coenzymes.

3.2 Introduction

α-Glucan phosphorylases (αGP, EC 2.4.1.1) are responsible for catalyzing the reversible phosphorolysis of α-1,4 glucan as shown below

$$\alpha\text{-glucan (DP = n)} + \text{phosphate} \leftrightharpoons \alpha\text{-glucan (DP = n -1)} + \alpha\text{-D-glucose 1-phosphate}$$

where DP denotes the degree of polymerization. These enzymes are widely distributed in microorganisms, plants, and animals. They play important roles in carbohydrate metabolism (Fletterick and Madsen 1980; Newgard et al. 1989). All α-glucan phosphorylases belong to the family 35 glycoyl transferases (www.cazy.org) (Cantarel et al. 2009). Although enzymes from various organisms differ dramatically in their substrate specificity and their modes of regulation, they have similar subunit molecular masses (≈ 90 kDa), have highly conserved sequences in

substrate binding sites and active sites, and require pyridoxal phosphate (PLP) as a cofactor (Schinzel and Nidetzky 1999; Weinhausel et al. 1997). α-Glucan phosphorylases can be used for generation of high-yield hydrogen (Ye et al. 2009; Zhang 2010a; Zhang et al. 2007), synthesis of amylose (Ohdan et al. 2007), production of complicated carbohydrates (Hong et al. 2008c; Yanase et al. 2006), and generation of glucose 1-phosphate (G-1-P) for cell-free protein synthesis (Wang and Zhang 2009a).

Thermophilic enzymes are of importance for industrial applications and molecular biology R&D. In addition to a large number of mesophilic α-glucan phosphorylases that have been characterized from bacteria (Chen and Segel 1968), yeast (Rath et al. 1992), fungi , plants, and animal tissues (Fukui et al. 1982; Rogers et al. 1992), several thermostable α-glucan phosphorylases have been purified and characterized from *Aquifex aeolicus* (Bhuiyan et al. 2003), *Geobacillus stearothermophilus* (Takata et al. 1998), *Sulfolobus tokodaii* (Hong et al. 2008c), *Thermus aquaticus* (Takaha et al. 2001), *Thermus thermophilus* (Boeck and Schinzel 1996), and *Thermotoga maritima* (Bibel et al. 1998). However, most of them are hyperthermophilic enzymes, which are rarely active at relatively low temperatures (Suzuki et al. 2001; Wang and Zhang 2010; Ye et al. 2009). To our knowledge, only one thermophilic α-glucan phosphorylase gene from *G. stearothermophilus* has been cloned, expressed in *E. coli*, and the protein has been characterized (Takata et al. 1998). For reactions in the temperature range of 40-60^{o}C, it is relatively important to discover more thermophilic α-glucan phosphorylases that can be active and stable as building blocks for cell-free enzymatic biotransformation for a long time (e.g., a week or longer).

Clostridium thermocellum is a thermophilic, anaerobic, Gram-positive bacterium with an optimal growth temperature of *ca.* 60^{o}C. It is a cellulose specialist, utilizing beta-1,4-glucosidic bond linked cellodextrins and cellulosic substrates only but neither glucose, maltodextrin nor starch (Lynd et al. 2002). The *C. thermocellum* cellobiose and cellodextrin phosphorylases (CBP and CDP) play important roles in intracellular cleavage of cellodextrins by generating glucose-1-phosphate without ATP expenditure (Zhang and Lynd 2004a; Zhang and Lynd 2005a). According to the *C. thermocellum* genome sequence (http://genome.ornl.gov/microbial/cthe/), ORF Cthe0357 (*αgp*) has been annotated to encode a putative α-glucan phosphorylase (αGP). The biochemical and physiological function of the putative αGP in *C. thermocellum* is not clear.

In this work, we cloned the putative *C. thermocellum agp* gene and expressed it in *E. coli,* and characterized the properties of this recombinant enzyme. We also significantly enhanced the half lifetime of this enzyme by formulating an enzyme buffer, which would be compatible with potential applications.

3.3 Materials and Methods

Chemicals and strains

All chemicals were reagent-grade, purchased from Sigma (St. Louis, MO), unless otherwise noted. *Clostridium thermocellum* ATCC 27405 genomic DNA was a gift from Dr. Jonathan Mielenz at the Oak Ridge National Laboratory (Oak Ridge, TN). *Escherichia coli* DH5α was used for plasmid manipulation; *E. coli* Rosetta BL21 (DE3) containing the gene expression plasmid was employed for producing the recombinant protein. Luria-Bertani (LB) medium was used with 100 μg/mL ampicillin (sodium salt). Microcrystalline cellulose – Avicel PH105 – was purchased from FMC (Philadelphia, PA). Regenerated amorphous cellulose (RAC) with a high external binding capacity was prepared from Avicel after water slurrying, cellulose dissolution in H_3PO_4, and regeneration in water (Zhang et al. 2006a). RAC with a number-average DP of 33 was prepared in concentrated phosphoric acid with partial hydrolysis at 50 ^oC for 10 h. Dextrin DP 4, dextrin DP 14, and dextrin DP 19 denote maltodextrin with average DPs of 4, 14, and 19, respectively.

Plasmid construction

ORP Cthe0357 was inserted into two gene expression plasmids pET21c and pCIG, yielding pET21c-*αgp* and pCI-*αgp*, respectively. For the pET21c-*αgp* plasmid, a pair of primers (p1 -- 5' GGAGG GGAGC TCTGT ATCTT TTTGG AAAAA TTAC 3' and p2 -- 5' GAGCG AGCTC AACTG TACAA TCCAT CTGAT AAGTC 3', SacI site underlined) was employed for PCR amplification based on the genomic DNA. After SacI digestion, the resulting fragment was ligated to the SacI-digested pET21c. After transformation, a clone containing pET21c-*agp* with *agp* in the proper direction was screened by colony PCR and then was validated by DNA

sequencing. Plasmid pCI-*αgp* was constructed based on plasmid pCIG (Hong et al. 2008a) by replacing the green fluorescent protein gene (*gfp*) with *αgp* to produce a fusion protein containing a family 3 cellulose binding module (CBM3), intein, and αGP. The *αgp* gene was amplified with a pair of primers (p3 -- 5' TGGTG GCTCG AGATG TATCT TTTTG GAAAA ATTAC 3', XhoI site underlined and p4 -- 5' AAGAA GGGAT CCTTA CTGTA CAATC CATCT GATAA GTCC 3', BamHI site underlined) based on the genomic DNA of *C. thermocellum*. The PCR product was digested by XhoI and BamHI, and then ligated with the XhoI/BamHI-digested pCIG vector for plasmid pCI-*αgp*, and the plasmid sequence was validated by DNA sequencing.

Protein production and purification

E. coli cultures were grown in the LB medium at 37 ^{o}C with a rotary shaking rate of 250 rpm. Once the absorbance of the culture reached ~0.8, IPTG was added to a final concentration of 0.25 mM. After 4 hours of cultivation at 37 ^{o}C, the *E. coli* cells were harvested by centrifugation and re-suspended in a 50 mM HEPES buffer (pH 7.2). The cells were then lysed by sonication. After centrifugation, the His-tagged αGP in the supernatant was adsorbed to the Bio-Rad Profinity IMAC Ni-resin (Hercules, CA) and was eluted by a 250 mM imidazole buffer. The fusion protein CBM-intein-αGP was specifically bound on RAC. After pH adjustment for intein self-cleavage, the cleaved αGP protein was obtained in the supernatant (Hong et al. 2008a). Both purified proteins were finally dialyzed against a 50 mM HEPES buffer (pH 7.2) and concentrated by using Centriprep centrifugal filter tubes (Millipore, MA) with a molecular weight cut-off of 50,000.

Activity assays

The enzyme activities were assayed at 50^{o}C in 50 mM HEPES buffer (pH 7.20) containing 30 mM Dextrin DP19 and 10 mM P_i or G-1-P unless otherwise noted. All enzymatic reactions were conducted in 5-mL glass tubes (12 × 75 mm, Fisher Scientific). The reactions were stopped by placing the tubes in a boiling water bath for 5 min. For the synthesis direction, the product - inorganic phosphate released from G-1-P was measured by the mild pH phosphate assay as described elsewhere (Saheki et al. 1985). For the degradation direction, the product G-1-P was

measured by using a glucose hexokinase/glucose-6-phosphate dehydrogenase assay kit (Zhang and Lynd 2004a) supplemented with a recombinant phosphoglucomutase (Wang and Zhang 2010).

Optimization of αGP reaction conditions

The effects of metal ions and reagents on αGP activity were examined, and the concentration of potential activators were optimized for both synthesis and degradation directions. The pH effects on the enzyme activities in both directions were determined in 50 mM acetate buffers (pH 3.0-6.0), HEPES buffers (pH 6.5-8.0), and Na_2CO_3-$NaHCO_3$ buffers (pH 9.0-10.0). The temperature ranges of the enzymes were tested from 10 to 90 ^oC.

Enzyme kinetics

For the synthesis direction, kinetic parameters were determined based on the initial rates by measuring the release of inorganic phosphate. The reactions were conducted at 50 ^oC in a 50 mM HEPES buffer (pH 6.8) containing 1 mM Mg^{2+}, 5 mM DTT, and various substrate concentrations between 0.2 and 5 times of their respective K_m values. Typical enzyme concentrations were 9 and 90 μg/mL on starch or dextrin and on RAC, respectively. One unit of phosphorylase in the synthesis direction was defined as the amount of enzyme that generated one μmole of phosphate per min. For the degradation direction, the reactions were conducted at 50^oC in a 50 mM HEPES buffer (pH 7.2) containing 3 mM Mg^{2+}, 1 mM DTT, and G-1-P at different concentrations. The enzyme concentration was 17.8 μg/mL. One unit of phosphorylase in the degradation direction is defined as the amount of enzyme generating one μmole of G-1-P per min.

Thermostability

The thermostability of αGP was studied at 40, 50, 60, 70, and 80 ^oC. The purified enzyme (0.01, 0.05, 0.25 mg/mL) was incubated in a 50 mM HEPES buffer (pH 7.2) without and with 0.1 mM Mg^{2+}, 5 mM DTT, 1% NaCl, 0.1% Triton X-100, and 1 mg/mL BSA at different temperatures for different time periods. The residual enzyme activities were assayed in the direction of dextrin

degradation at 50 ^{o}C in a 50 mM HEPES buffer (pH 7.2) containing 5 μg/mL αGP, 30 mM dextrin DP19, 10 mM P_i, 3 mM Mg^{2+}, and 1 mM DTT.

Other assays

Protein mass concentrations were determined by the Bradford methods with bovine serum albumin (BSA) as the standard. The number average degree of polymerization of RAC was determined as described elsewhere (Zhang and Lynd 2005b). The purity of the enzymes was checked by SDS PAGE.

3.4 Results and Discussion

Sequence analysis

A 2569-bp *C. thermocellum* ORF Cthe0357 was annotated to encode a putative *α*-glucan phosphorylase, according to BLASTP analysis (Altschul et al. 1997). The deduced amino acid sequence has 43% identity with the hyperthermophilic *Thermotoga maritima* αGP, 23% identity with the thermophilic *Geobacillus stearothermophilus* αGP, and 22% identity with the mesophilic *Escherichia coli* maltodextrin phosphorylase (MalP). The domain from residues 114 to 705 has 95% identity to the conserved domains (CDS) of family 35 glycoyl transferases -- α-glucan phosphorylases. Compared to the β-glucan phosphorylases from family 94 glycoside hydrolases, such as *C. thermocellum* cellobiose phosphorylase and cellodextrin phosphorylase, the identities were only 6.0% and 6.4%, respectively. This analysis suggested that Cthe 0357 encoded an α-glucan phosphorylase.

Protein production & purification

Two expression plasmids, pET21c-αgp and pCI-αgp, were constructed for the production of His-tagged and tag-free α-glucan phosphorylase, respectively. The purified His-tagged αGP appeared to be homogeneous with a molecular mass of *ca.* 81 kDa (Fig 3.1), which was close to the estimated value of 97,712 Da from the deduced amino acid sequence. The purified enzyme had a specific activity of 2.22 ± 0.17 U/mg and the purification yield was 43.8% (Table 3.1). Approximately 10.8 mg of the purified enzyme was obtained per liter of the culture.

Table 3.1 Purification of the αGP from *C. thermocellum.*

Fraction	Volume (mL)	Total protein (mg)	Total activity* (U)	Sp. Act. (U/mg)	Yield (%)
His_6-tagged enzyme					
Crude extract	23	40.1	27.2	0.678	100
Fraction eluted from Ni^{2+} resin	15	6.18	13.5	2.18	49.6
Dialysis & concentration	6.1	5.36	11.9	2.22	43.8
Tag-free enzyme					
Crude extract	26	44.1	22.7	0.515	100
Fraction truncated by Intein self-cleavage	6.5	3.74	7.47	2.00	32.9
Dialysis & concentration	3.2	3.16	6.57	2.08	28.9

* The activity was assayed in the synthesis direction at 50 ^{o}C in a 50 mM HEPES buffer (pH 6.8) containing 30 mM Dextrin DP 19 and 10 mM G-1-P.

The tag-free phosphorylase was purified by affinity adsorption of the fusion protein containing a CBM3 tag on an ultra-high capacity adsorbent RAC, followed by intein self-cleavage. The specific activity of the tag-free phosphorylase was 2.08 ± 0.23 U/mg with a yield of 28.9% (Table 3.1). Approximately 6.4 mg of the purified phosphorylase was obtained per L of the culture. The relatively low purification yield of the CBM3-intein-based protein purification was mainly attributed to self-cleavage of the intein *in vivo*. Higher yield could be achieved through short-time protein production at low temperature (Hong et al. 2008a). SDS-PAGE analysis indicated that the purified tag-free phosphorylase was homogenous and had a molecular mass of ~ 81 kDa similar to the His-tagged protein (Fig. 3.1, lanes 2 & 5). Since the two forms of phosphorylase had similar specific activities, the His-tagged phosphorylase was used for further characterization.

Biochemical properties

The effects of metal ions, including Mn^{2+}, Mg^{2+}, Ca^{2+}, Cu^{2+}, Zn^{2+}, and Ni^{2+}, and reagents, including EDTA, AMP, and DTT, on αGP activity were examined. The highest activity was obtained in the presence of 1 mM Mg^{2+} and 5 mM DTT for the synthesis direction, resulting in 1.4-fold increase in αGP activity as compared to those without addition of metal ions or DTT.

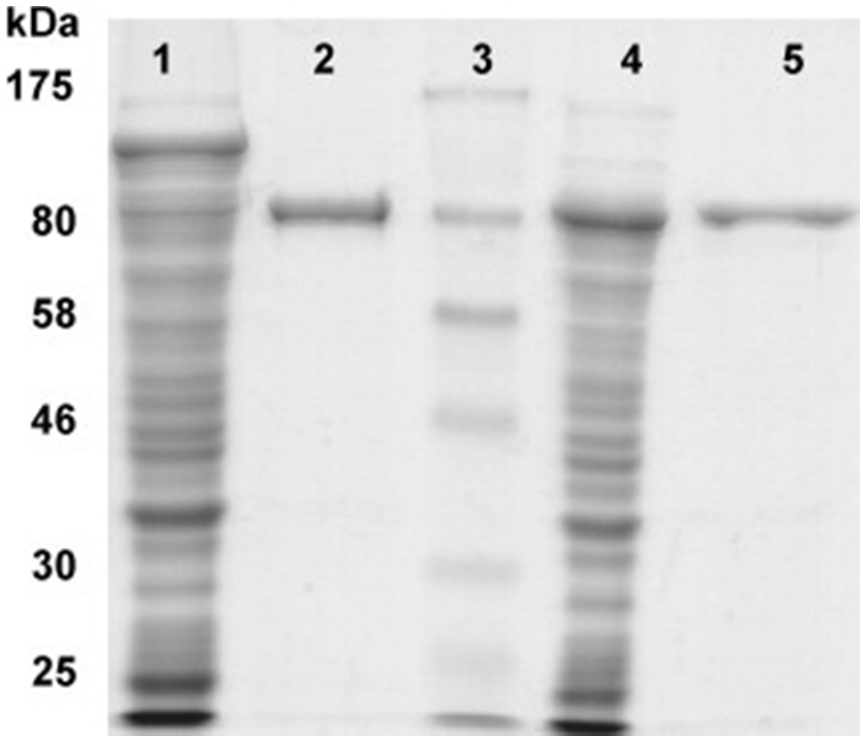


Figure 3.1 Purification of His-tagged and non-tagged *C. thermocellum* αGP. Lane 1, crude cell extracts from *E. coli* containing the CBM3-Intein- αGP fusion protein; Lane 2, final purified tag-free protein; Lane 3, molecular weight marker; Lane 4, crude cell extracts from *E. coli* containing His-tagged αGP; Lane 5. purified His-tagged protein.

But the highest activity in the degradation direction was obtained in the presence of 3 mM Mg^{2+} plus 1 mM DTT, 1.7-fold higher than the blank control. One millimolar Mn^{2+} strongly inhibited αGP activity, decreasing the activity more than 90%. The other metals (1 mM) or AMP (1 mM) or EDTA (1 mM) had no significant influences on its activities. However, high level Ca^{2+} and Ni^{2+} caused obvious salt precipitates and 10 mM EDTA decreased its activities by *ca.* 30%.

Figure 3.2A shows the effects of pH from 3.0 to 10.0. The maximal activity was found at pH 6.0-6.5 and 7.0-7.5 in the synthesis and degradation directions, respectively (Fig. 3.2A). Different pH preferences for different reaction directions were also reported for the αGP from *Corynebacterium callunae*, which had pH optima at 5.7-6.1 and 7.0-7.2 in the synthesis and degradation directions, respectively (Weinhausel et al. 1997).

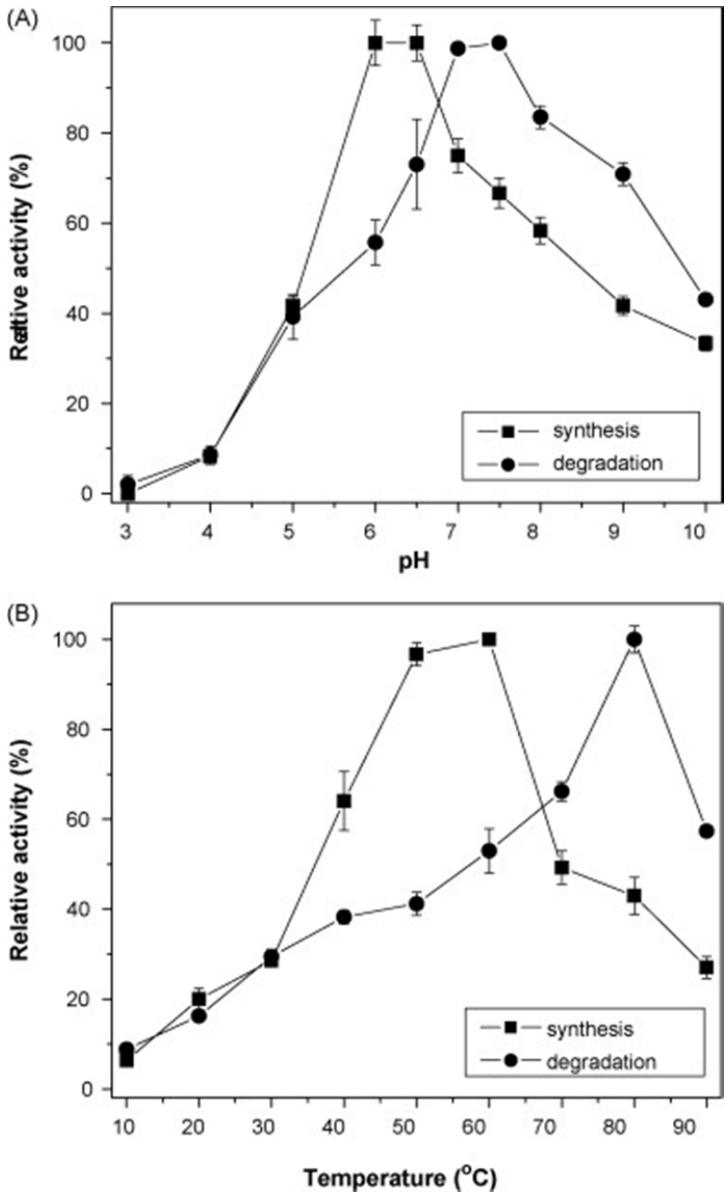


Figure 3.2 The profiles of pH and temperature of the *C. thermocellum* αGP. A: profiles of pH; B: profiles of temperature.

The effect of temperature was investigated from 10 to 90 ^{o}C at pH 7.2. For the synthesis direction, the optimal temperature was 60 ^{o}C; this enzyme retained nearly 97% of its activity at 50 ^{o}C (Fig. 3.2B). For the degradation direction, the optimal temperature was 80 ^{o}C. The enzyme activity sharply declined when the temperatures were higher or lower than 80 ^{o}C (Fig. 3.2B). The activation energies were 25.3 and 42.9 kJ/mol for the degradation and synthesis directions, respectively, according to the Arrhenius plots.

Different optimal pH and temperatures suggested partially different binding and reaction modes in synthesis and degradation (Drueckes et al. 1996), but different pH and temperature optima have often been ignored in studies of αGP (Bhuiyan et al. 2003; Hong et al. 2008c; Takata et al. 1998).

Substrate specificity and kinetics

The substrate specificity and kinetics of αGP were determined in the synthesis and degradation directions on different substrates (dextrin DP19, dextrin DP14, dextrin DP4, soluble starch, and RAC DP33) (Table 3.2). In the synthetic direction, there were no significant differences between the kinetic parameters on dextrin DP14 and dextrin DP19. When the substrate chain length decreased to 4, the K_m value increased 14 fold, to 17.1 mM. Consequently, enzyme efficiency (k_{cat}/K_m= 0.00064 $M^{-1}s^{-1}$) on dextrin DP 4 was approximately 5 fold lower than that on dextrin DP 14 (0.0030 $M^{-1}s^{-1}$) although k_{cat} on dextrin DP 4 (11.0 s^{-1}) was higher than on dextrin DP 14 (3.6 s^{-1}). The similar trend for K_m was found in the degradation direction. αGP exhibited similar kinetic properties for dextrin DP 14 and DP 19, and had a lower catalytic efficiency on the shorter substrate dextrin DP 4. These results indicated that the *C. thermocellum* α-glucan phosphorylase preferred substrates with DP > 4 for both synthetic and phosphorolytic reactions.

The *C. thermocellum* α-glucan phosphorylase showed broad substrate specificities from soluble dextrin to starch. The catalytic efficiencies on dextrins (with DP > 4) and starch were comparable. αGP did not work on amorphous cellulose (RAC DP 33, cellulosic materials linked by beta-1,4-glucosidic bonds) in the degradation direction but it had a weak activity in the synthesis direction (k_{cat}/K_m = 0.00017 $M^{-1}s^{-1}$). Such special properties implied that the enzyme might be responsible for the formation of glycocalyx in *C. thermocellum* (Lynd et al. 2002).

Thermostability

Figure 3.3A shows that αGP deactivation was strongly associated with its concentration. After being incubated at 50 ^{o}C for 2 hours, more than 60% and 85 % of αGP activities remained for enzyme concentrations of 0.05 and 0.25 mg/mL, respectively, but αGP at a low concentration

Table 3.2 Apparent kinetics constants of *C. thermocellum* αGP in the directions of glucan synthesis and degradation.

Variable substrate	Fixed substrate	k_{cat} (s^{-1})	K_m (mM)[a]	k_{cat}/K_m ($M^{-1}s^{-1}$)
a) Synthesis[c]				
Dextrin DP 19	10 mM g1p	6.6 ± 0.3	1.9 ± 0.2	0.0035
Dextrin DP 14	10 mM g1p	3.6 ± 0.1	1.2 ± 0.2	0.0030
Dextrin DP 4	10 mM g1p	11.0 ± 0.4	17.1 ± 0.2	0.00064
Starch [b]	10 mM g1p	2.1 ± 0.2	1.5 ± 0.2	0.0015
Glucose 1-phosphate	40 mM dextrin 19	4.2 ± 0.1	0.054 ± 0.008	0.0077
RAC DP 33	10 mM g1p	0.12 ± 0.05	0.72 ± 0.18	0.00017
b) Degradation[d]				
Dextrin DP 19	30 mM P_i	8.1 ± 0.2	0.39 ± 0.01	0.021
Dextrin DP 14	30 mM P_i	9.3 ± 0.3	0.73 ± 0.03	0.013
Dextrin DP 4	30 mM P_i	8.4 ± 0.8	1.8 ± 0.2	0.0047
Starch [b]	30 mM P_i	6.3 ± 0.6	0.30 ± 0.07	0.021
P_i	40 mM dextrin 19	8.2 ± 0.5	7.7 ± 0.6	0.0011
RAC DP 33	30 mM P_i	ND[e]	ND[e]	ND[e]

[a] The concentration of polysaccharides were given as the molar concentration of the non-reducing ends.

[b] The DP of soluble starch was estimated to be 24.

[c] The activities were assayed at 50 ^oC in a 50 mM HEPES buffer (pH 6.8) containing 1 mM Mg^{2+}, 5 mM DTT and various substrate concentrations between 0.2 and 5 times of their respective *Km* values.

[d] The activities were conducted at 50^oC in a 50 mM HEPES buffer (pH 7.2) containing 3 mM Mg^{2+} and 1 mM DTT at various substrate concentrations.

[e] ND: not detected.

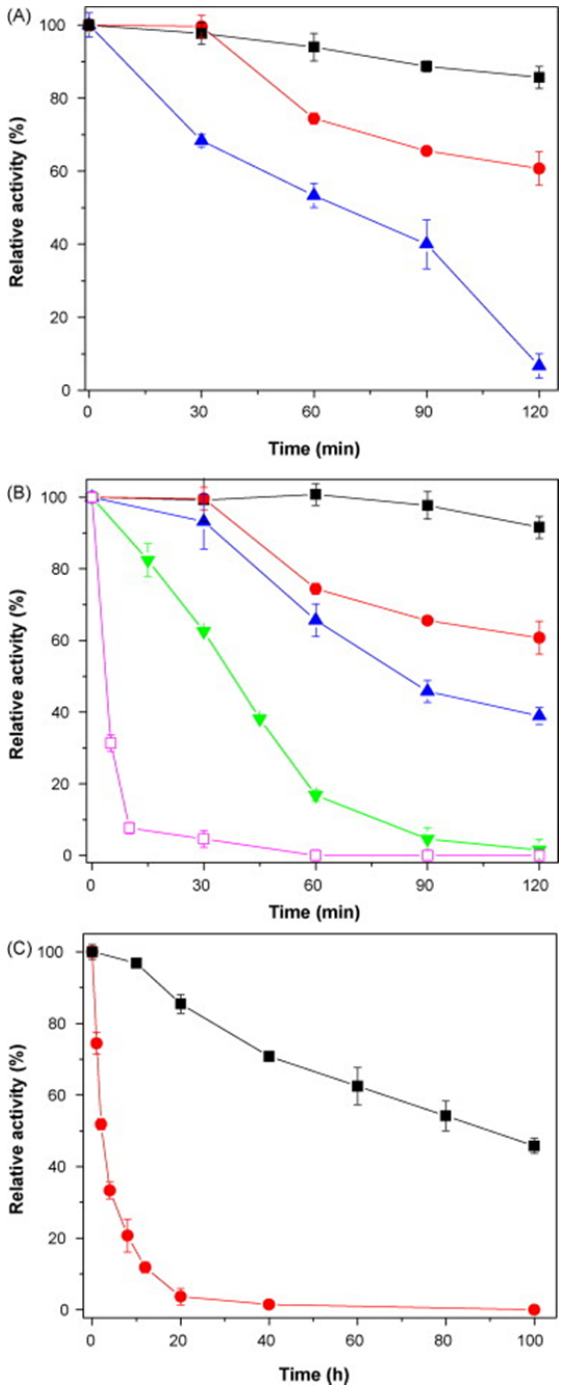


Figure 3.3. Thermostability of the *C. thermocellum* αGP. (A) Thermostability of αGP at different protein concentrations of 0.01, 0.05, 0.25 mg/mL at 50 °C. 0.01 mg/mL (▲), 0.05 mg/mL (●), and 0.25 mg/mL (■); (B) thermostability of 0.05 mg/mL αGP at 40 °C (■), 50 °C (●), 60 °C (▲) , 70 °C (▼), and 80 °C (□); (C) thermostability of 0.05 mg/mL αGP without additives (●) , and with addition of 0.1 mM Mg^{2+}, 5 mM DTT, 1% NaCl, 0.1% Triton X-100, and 1 mg/mL BSA (■) at 50 °C. The residual enzyme activities were assayed in the degradation direction as described in Materials and Methods.

(0.01 mg/mL) had less than 5 % of its initial activity. The half lifetimes of 0.05 mg/mL αGP were around 125, 84, 37, and 4 min at 50, 60, 70, and 80 ^{o}C, respectively (Fig. 3.3B). At 40 ^{o}C, the enzyme was relatively stable with more than 90% activity remaining after a 2 hour heat treatment.

Stabilization of αGP at low concentrations would be important for its potential applications. A myriad of efforts have been made to improve αGP thermostability through immobilization (Friedrich et al. 1977), protein engineering *via* directed evolution (Yanase et al. 2005) or rational design (Griessler et al. 2000), protein formulation (Janecek 1993), and with the help of folding chaperones (Eronina et al. 2009). Compared to these methods, the enhancement of protein thermal inactivation through the addition of protective reagents is easier, but this strategy is often ignored (Wang and Zhang 2010). Therefore, a number of combinations of various concentrations of Mg^{2+}, DTT, NaCl, Triton X-100, and BSA were tested. The presence of metal ions and the reducing agent DTT was reported to reduce the disulfide bridges and suppress thiol/disulfide (SH/SS) interchange -- a major cause of enzyme deactivation (Meng et al. 1993). Magnesium ions may protect the enzyme from thermal denaturation (Liu et al. 2007). High concentration salts are capable of preventing protein aggregation as well as thermally-induced unfolding (Cannon and MacDonald 2009). Triton X-100 has been widely used to stabilize DNA polymerases for PCR, nucleic acid hybridization, and enzyme storage, although its real function remains unclear (Wang and Zhang 2010). After testing, a combination of 0.1 mM Mg^{2+}, 5 mM DTT, 1% NaCl, 0.1% Triton X-100, and 1 mg/mL BSA was found to keep the activity of αGP almost unchanged but to extend the half time of inactivation by 45 fold from ~ 2 h to *ca.* 90 h at 50 ^{o}C (Fig. 3.3C).

Homology modeling and structural analysis

A homology model of the *C. thermocellum* α-glucan phosphorylase was built by using 3D-JIGSAW based on the human liver glycogen phosphorylase (HLGP, PDB: 1l7x), as shown in Figure 3.4. The crystal structure of HLGP was used as a template because it has the highest identity (23%) with the *C. thermocellum* GP based on their catalytic domains. The *C. thermocellum* αGP shares some similar structural properties with HLGP around the catalytic sites. For example, the catalytic domain of HLGP is formed by the 250's loop (residues 250-

260), the GATE (residues 280-289), and the 380's loop (residues 376-386) (Fig. 4A) (O'Reilly et al. 1997; Rath et al. 2000b). Similarly, the catalytic domain of the *C. thermocellum* GP is composed of the 215's loop (residues 215-225), the GATE (residues 239-248), and the 315's loop (residues 315-325) (Fig. 4B). The key residues, involving in intermolecular contacts for enzyme-substrate complex, appear identical in HLGP and αGP, such as His377 (HLGP) to His315(αGP) for sugar subsite -1, Asn339 (HLGP) to Asp276 (αGP) for sugar subsite +1, Tyr280 and His341 (HLGP) to Tyr241 and His279 (αGP) for sugar subsite +2, Arg 292 (HLGP) to Arg 248 (αGP) for sugar subsite +3, and Pro381 (HLGP) to Pro319 (αGP) for sugar sub-site +4 (O'Reilly et al. 1997; Watson et al. 1999). The homology model clearly suggest that the *C. thermocellum* αGP has a five-glucose unit binding site (sub-sites -1 to +4), similar to HLGP. It is also supported by our experimental data that Michaelis-Menten constant (K_m) significantly decreased as substrate chain lengths were longer than four (Table 3.2).

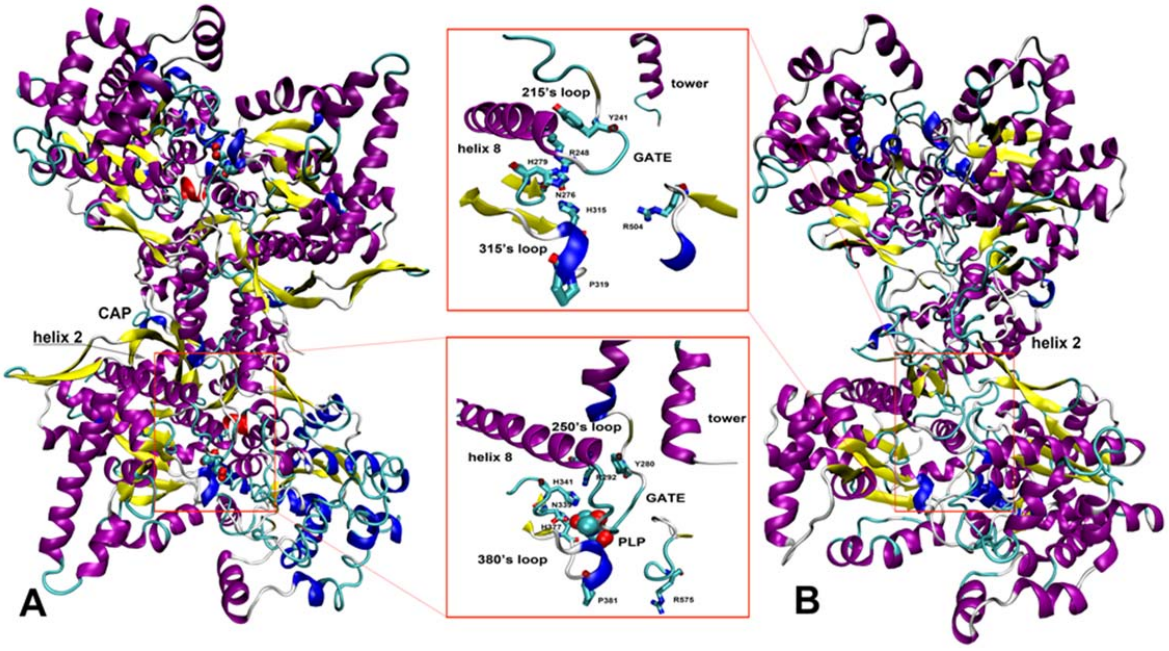


Figure 3.4. Comparison of the crystal structure of human liver glycogen phosphorylase (HLPG) and a homology model of *C. thermocellum* αGP. A: HLPG; B: *C. thermocellum* αGP. The protein structures were illustrated by VMD (UIUC, Illinois, US), where α-helices, 3_{10}-helices, extended β-stands, bridge β-stands, turns, and coils were colored by purple, blue, yellow, tan, cyan and white, respectively. CAP: a subunit (residues 42-45) of HLGP which caps the AMP binding site; GATE: a subunit (residues 280-289 in HLGP, and residues 239-248 in αGP) that takes a critical role in substrate binding; PLP, pyridoxal phosphate.

Some differences between the HLGP structure and the *C. thermocellum* αGP model were observed, e.g. in the region of helix 2 (residues 48-78 of HLGP) and 'CAP' (residues 42-45 of HLGP, so called because it 'caps' the AMP binding site) (Rath et al. 2000b). Helix 2 and CAP structures are believed to form the binding site of AMP which is an efficient activator for HLGP (Maddaiah and Madsen 1966). Hence, the big structural differences in the helix 2 and CAP between HLGP and αGP may explain why AMP, a regulator for most eukaryotic α-glucan phosphorylases, has no influence on the activity of the *C. thermocellum* αGP. Such changes also result in dramatic differences in the catalytic machinery between them. HLGP has two conformational states: a less active T state and a more active R state. In the T state, 280's loop and GATE block access to the catalytic site and create an unfavorable electrostatic environment at the catalytic site for phosphate ions (O'Reilly et al. 1997). Once it binds with AMP or is phosphorylated at Ser14, HLGP is switched to the active R state (Rath et al. 2000a; Rath et al. 2000b). By contrast, the 215's loop and GATE structures of *C. thermocellum* αGP seem to be an open conformation always (Fig. 3.4). It allows the substrate to access the catalytic site so that the *C. thermocellum* αGP is constantly active.

3.5 Conclusion

ORF Cthe 0357 from *C. thermocellum*, which encodes an α-glucan phosphorylase, was cloned and expressed in *E. coli*. The recombinant enzyme showed broad substrate specificities on dextrin and starch. Although its physiological role is still unclear, it may be responsible for the formulation of glycocalyces. Through buffer formulation, the half lifetime of this enzyme was extended by 45 fold to *ca.* 90 h at 50^{o}C. Therefore, the enzyme has the potential for broad applications in cell-free protein synthesis (Wang and Zhang 2009b) and biofuels production *via* cell-free synthetic enzymatic pathway biotransformation (Zhang 2009; Zhang et al. 2008).

Chapter 4

Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose

Material from this chapter appears in Xinhao Ye, Zhiguang Zhu, Chenming Zhang, Y-H. Percival Zhang: Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose. *Appl Microbiol Biotechnol* 2011, **X**, XXXX-XXXX, and is reprinted by permission of Springer-Verlag.

4.1 Abstract

Clostridium thermocellum cellodextrin phosphorylase (*Ct*CDP), a single-module protein without an apparent carbohydrate-binding module, has reported activities on soluble cellodextrin with a degree of polymerization (DP) from 2-5. In this study, *Ct*CDP was first discovered to have weak activities on weakly water-soluble celloheptaose and insoluble regenerated amorphous cellulose (RAC). To enhance its activity on solid cellulosic materials, four cellulose binding modules, e.g. CBM3 (type A) from *C. thermocellum* CbhA, CBM4-2 (type B) from *Rhodothermus marinus* Xyn10A, CBM6 (type B) from *Cellvibrio mixtus* Cel5B, and CBM9-2 (type C) from *Thermotoga maritima* Xyn10A, were fused to the C terminus of *Ct*CDP. Fusion of any selected CBM with *Ct*CDP did not influence its kinetic parameters on cellobiose but affected the binding and catalytic properties on celloheptaose and RAC differently. Among them, addition of CBM9 to *Ct*CDP resulted in a 2.7-fold increase of catalytic efficiency for degrading celloheptaose. *Ct*CDP-CBM9 exhibited enhanced specific activities over 20% on the short-chain RAC (DP = 14) and more than 50% on the long-chain RAC (DP = 164). The chimeric protein *Ct*CDP-CBM9 would be the first step to construct a cellulose phosphorylase for in vitro hydrogen production from cellulose by synthetic pathway biotransformation (SyPaB).

4.2 Introduction

Cellodextrin phosphorylase (CDP, EC 2.4.1.49) is responsible for catalyzing phosphorolysis of soluble cellodextrins with degree of polymerization (DP) of 2-5 (Schomburg et al. 2009; Sheth and Alexander 1967). It belongs to glycoside hydrolase (GH) family 94 (Hidaka et al. 2004) and plays an important role in the energy-efficient metabolism of long-chain cellodextrins (Lou et al. 1996; Zhang and Lynd 2004a). CDP activity has been reported in *Fibrobacter succinogenes*, *Vibrio splendidus*, *Spirochaeta thermophile*, and several species of *Cellulomonas* and *Clostridia* (Reichenbecher et al. 1997; Sheth and Alexander 1967; Wells et al. 1995), but the CDP enzymes have only been isolated from *Clostridium thermocellum* (Arai et al. 1994) and *Clostridium sterocorarium* (Reichenbecher et al. 1997). Recently, a recombinant cellodextrin phosphorylase from *C. thermocellum* (*Ct*CDP) was expressed in *Escherichia coli*, purified, and characterized in detail (Krishnareddy et al. 2002). The recombinant *Ct*CDP has been used for synthesis of

cellulase inhibitors (Kawaguchi et al. 1998), in vitro enzymatic hydrogen production (Ye et al. 2009), and the production of glucose 1-phosphate, cellodextrin, and highly-ordered cellulose (Hiraishi et al. 2009; Samain et al. 1995).

Glycoside hydrolases usually consist of a catalytic module and one or more non-catalytic carbohydrate-binding modules (CBM) (Boraston et al. 2004; Bourne and Henrissat 2001). CBM can enrich enzymes on the surface of solid substrates through affinity adsorption, so as to enhance their activity (Carrard et al. 2000; Shoseyov et al. 2006). Currently, CBMs have been classified into 61 sequence-based families (www.cazy.org) and further grouped into three classes based on their structure, function and ligand specificities (Boraston et al. 2004). Families 1, 2a, 3, 5, and 10 fit in type A CBMs that have a planar hydrophobic ligand binding surface and bind to crystalline cellulose. Type B CBMs, comprising families 4, 6, 17, 28, etc., contain clefts that accommodate single polysaccharide chains, while the ligand binding sites in Type C CBMs, comprising families 9, 13, 14, etc., interact with mono- or disaccharides (Boraston et al. 2004; van Bueren et al. 2005).

Unlike other glycoside hydrolases, the *C. thermocellum* cellodextrin phosphorylase is a single modular protein and does not have the apparent CBM. Recently, several studies have shown that addition of a CBM to single domain enzymes increased their activity toward insoluble substrates (Carrard et al. 2000; Kittur et al. 2003). Ravalason et al. (2009) showed that addition of a family 1 CBM (type A) of *Aspergillus niger* to *Pycnoporus cinnabarinus* laccase significantly improved the delignification capabilities of laccase to softwood kraft pulp. *Ruminococcus albus* endoglucanase fused with a family 6 CBM (type B) of *C. stercorarium* xylanase A exhibited higher molar activity toward insoluble acid-swollen cellulose and ball-milled cellulose (Karita et al. 1996). Adding the CBM6 to *Bacillus halodurans* xylanase also improved the enzyme's activity toward insoluble xylan (Mangala et al. 2003). Moreover, Maglione et al. (1992) reported a 10-fold increase in the specific activity of *Prevotella ruminicola* endoglucanase on insoluble cellulose after being fused with a family 2 CBM (type A) from *Thermomonospora fusca* endoglucanase. It thus would be of great interest to investigate if fusion of a CBM to *Ct*CDP could potentiate its catalytic activity on insoluble cellulose.

Table 4.1 The origins and their functions of the CBMs used in this study

CBM	Type	Origin	Organism	GenBank ID	Binding site	Reference
CBM3	A	CbhA	*C. thermocellum*	ABN51651.1	Crystalline and amorphous cellulose	Carrard *et al.* (2000)
CBM4-2	B	Xyn10A	*R. marinus*	CAA72323.2	Amorphous cellulose and cellodextrin	Hachem *et al.* (2000)
CBM6	B	Cel5B	*C. mixtus*	AAB61462.2	Cello-oligosaccharides	Heshaw *et al.* (2004)
CBM9-2	C	Xyn10A	*T. maritima*	AAD35155.1	Amorphous and crystalline cellulose, Soluble polysaccharides, Mono- and disaccharides	Boraston *et al.* (2001)

Creation of a non-natural cellulose phosphosphorylase would be extremely important for high-yield hydrogen production because it would be possible to produce low-cost hydrogen from less costly pretreated solid cellulosic materials rather than soluble cellodextrins without costly ATP consumption via synthetic pathway biotransformation (SyPaB) (Ye et al. 2009; Zhang 2010b). In this study, we found that *Ct*CDP had weak activities on weakly water-soluble celloheptaose and insoluble regenerated amorphous cellulose (RAC). Then *Ct*CDP was linked with four representatives of the three types of CBM (Table 4.1) to test whether fusion of a CBM could improve the catalytic potential of *Ct*CDP on cellulose. Among the CBMs selected, CBM3, derived from *C. thermocellum* CbhA, belongs to type A CBM and binds to the surface of insoluble cellulose (Carrard et al. 2000; Hong et al. 2007); CBM4-2 from *Rhodothermus marinus* Xyn10A and CBM6 from *Cellvibrio mixtus* Cel5B are type B CBMs, which interact with amorphous cellulose and/or cellodextrin (Henshaw et al. 2004; Simpson et al. 2002); CBM9-2 of Xyn10A from *Thermotoga maritima* pertains to type C CBM, which has affinity with a broad range of cellulose, e.g. amorphous and crystalline cellulose, cello-/xylo- oligomers, and soluble mono- and disaccharides (Boraston et al. 2001; Notenboom et al. 2001). Here we described the production, isolation, and characterization of the chimeric proteins *Ct*CDP-CBM3, *Ct*CDP-CBM4, *Ct*CDP-CBM6, and *Ct*CDP-CBM9. The results demonstrated that *Ct*CDP-CBM9 promoted the catalytic activity toward short-chain and/or long-chain cellulose. The active *Ct*CDP-CBM9 would be the first step to develop high-activity cellulose phosphorylase in the future.

4.3 Materials and Methods

Materials

All chemicals were reagent-grade, purchased from Sigma (St. Louis, MO), unless otherwise noted. *C. thermocellum* ATCC 27405 genomic DNA was a gift from Dr. Jonathan Mielenz at the Oak Ridge National Laboratory (Oak Ridge, TN). *E. coli* JM109 was used for plasmid manipulation; *E. coli* Rosetta BL21 (DE3) containing the gene expression plasmid was employed for producing the recombinant protein. The Luria-Bertani (LB) medium was used with 100 μg/mL ampicillin (sodium salt). Microcrystalline cellulose – Avicel PH105 – was purchased from FMC (Philadelphia, PA). Cellodextrins with DP from 2-7 was prepared by mixed acid

hydrolysis of Avicel and separated by large size chromatographic column, as described before (Zhang and Lynd 2003). Regenerated amorphous cellulose (RAC) was prepared from Avicel after water slurring, cellulose dissolution in H_3PO_4, and regeneration in water (Zhang et al. 2006a). Two types of RAC were generated in terms of different number-average degrees of polymerization (DP). Insoluble RAC DP 164 was prepared in concentrated phosphoric acid with partial hydrolysis in an ice bath for 1 h, while slightly soluble RAC DP 14 was prepared by partial hydrolysis in concentrated phosphoric acid at 50 ^{o}C for 20 h (Zhang and Lynd 2004b).

Plasmid construction

A pair of primers (p1 – 5' TTAAGA CATATG ATTACT AAAGTA ACAGCG AG 3' and p2 – 5' AATTTT CTCGAG GAGCTC GGATCC TTTAAA CTTAAG AGTCAC TATATG TTC 3', restriction enzyme cutting sites underlined) was employed to amplify full-length *cdp* from *C. thermocellum* genomic DNA. The PCR product was digested by *Nde*I and *Xho*I and ligated with the digested plasmid pET21c to give plasmid pET-*cdp*. The gene fragment of *cbm3* was amplified from the gDNA of *C. thermocellum* by primers p3 (5' TTAAGA GAGCTC GTACAG TATTTG TGCGAA AATACG 3', *Sac*I site underlined) and p4 (5' ATTATC CTCGAG TTCCAG CTGCAG ATAATG CTC 3', *Xho*I site underlined). The amino acid sequences of CBM4, CBM6, and CBM9 were taken from GenBank (their accessible numbers were listed in table 1), and back-translated into *E. coli* optimized nucleotide sequences by OPTIMIZER (http://genomes.urv.es/OPTIMIZER/) (Puigbò et al. 2007). The genes were then synthesized in Virginia Bioinformatics Institute (Blacksburg, VA). In particular, the synthesized fragment of *cbm6* was joined with a natural N-terminal linker of family 6 CBM from *C. thermocellum* (GenBank ID: AAB61462.2, residue 310-331) via a *Sac*I site (GAGCTC). The fragment was digested by *Bam*HI and *Xho*I and then inserted into the digested pET21-*cdp* plasmid, yielding the plasmid pET-*cdp-cbm6*. Later, *cbm6* was replaced by *cbm3*, *cbm4* and *cbm9* through *Sac*I and *Xho*I sites, resulting to pET-*cdp-cbm3*, pET-*cdp-cbm4*, and pET-*cdp-cbm9*, respectively. Finally, the nucleotide sequences of the chimeric genes were validated by DNA sequencing in MCLAB (San Francisco, CA).

Protein production and purification

The *E. coli* BL21 strain harboring the expression plasmid was grown in the LB medium at 37 ^{o}C with a rotary shaking rate of 250 rpm. Once the absorbance at 600 nm (OD_{600}) reached ~0.8, IPTG was added to a final concentration of 0.25 mM. After 20 hours of cultivation at 16 ^{o}C, the *E. coli* cells were harvested by centrifugation and re-suspended in a 50 mM HEPES buffer (pH 7.2). The cells were then lysed by sonication. After centrifugation, the soluble His-tagged *Ct*CDP or a chimeric protein was adsorbed to the Bio-Rad Profinity IMAC Ni-resin (Hercules, CA) and was eluted by a HEPES buffer (50 mM, pH 7.2) with 250 mM imidazole and 0.3 M NaCl. The eluate containing the purified proteins was then exchanged into 50 mM HEPES buffer (pH 7.2) through GE Healthcare PD-10 desalting columns (Piscataway, NJ) and concentrated by Centriprep centrifugal filter tubes with a 50,000 molecular weight cut-off membrane (Millipore, MA).

Activity assays

All enzymatic reactions were conducted in 5-mL glass tubes (12 × 75 mm, Fisher Scientific). The enzyme activities were assayed at 30 ^{o}C in 50 mM HEPES buffer (pH 7.2) containing 30 mM cellobiose, 10 mM glucose 1-phosphate, 5 mM DTT, and 1 mM Mg^{2+} unless otherwise noted. Enzyme concentrations were set at 2.5 mg L^{-1} for most assays. To measure the specific activities on RAC DP 14 and RAC DP 164, sugar concentration was increased to 6.3 g [glucose equivalent] L^{-1} and 7.5 g [glucose equivalent] L^{-1}, respectively. Enzyme concentrations were 0.01 g L^{-1} on RAC DP 14 and 0.05 g L^{-1} on RAC DP 164. The reactions were stopped by placing the tubes in a boiling water bath for 5 min. The temperature and pH optimum was determined as described elsewhere (Ye et al. 2010).

Enzyme kinetics

For the synthesis direction, kinetic parameters were determined based on the initial rates by measuring the release of inorganic phosphate (P_i). The reactions were conducted at 30 ^{o}C in a 50 mM HEPES buffer (pH 7.2) containing 1 mM Mg^{2+}, 5 mM DTT, and various substrate concentrations between 0.2 and 5 times of their respective K_m values. The product P_i released from glucose 1-phosphate was measured by the mild pH phosphate assay as described elsewhere (Ye et al. 2010). One unit of phosphorylase in the synthesis direction was defined as the amount

of enzyme that generates one μmole of phosphate per min. For the degradation direction, the reactions were conducted at 30 ^{o}C in a 50 mM HEPES buffer (pH 7.2) containing 1 mM Mg^{2+}, 5 mM DTT, and P_i at different concentrations. The product was measured by using a glucose hexokinase/glucose 6-phosphate dehydrogenase assay kit (Zhang and Lynd 2004a) supplemented with a recombinant *C. thermocellum* phosphoglucomutase (Wang and Zhang 2010). One unit of phosphorylase in the degradation direction is defined as the amount of enzyme generating one μmole of glucose 1-phosphate per min.

Binding assays

Binding assays were carried out in 2 ml glass screw-cap vials. Each vial contained 0.2 g RAC/L, 0.05 - 0.6 g L^{-1} protein, and 50 mM HEPES buffer (pH 7.2). The vials were incubated with slow rotation for 1 h at 30 ^{o}C. The binding substrate was pelleted by centrifugation (5 min, 16,000 g) and the supernatants were tested for protein content. The maximum adsorption capacity A_{max} was determined by regressing binding isotherm data to a modified Langmuir-type binding model as described previously (Hong et al. 2007).

Other assays

Protein concentrations were determined by the Bradford methods with bovine serum albumin (BSA) as the standard. The average degree of polymerization of RAC was determined as described elsewhere (Zhang and Lynd 2005b). The purity of the enzymes was checked by SDS PAGE.

4.4 Results

Construction, production, and purification of the fusion proteins

Expression plasmid pET-*cdp* was first constructed to produce wild-type cellodextrin phosphorylase. Since c*bm3* from the *C. thermocellum Cbh*A had been successfully expressed in *E. coli*, it was directly amplified from the genomic DNA of *C. thermocellum* without further optimization. After codon optimization, however, *cbm*4-2 from *R. marinus Xyn*10A, *cbm*6 from *C. mixtus Cel*5B, and *cbm*9-2 from *T. maritima Xyn*10A were synthesized to overcome the potential codon-bias problem (Kurland 1991). The DNA fragment of each CBM was inserted

into pET-*cdp* to generate the chimeric gene, and a natural linker from *C. thermocellum* xylosidase (Cthe 2196) was placed in the junction between *cdp* and *cbm* (Fig.4.1).

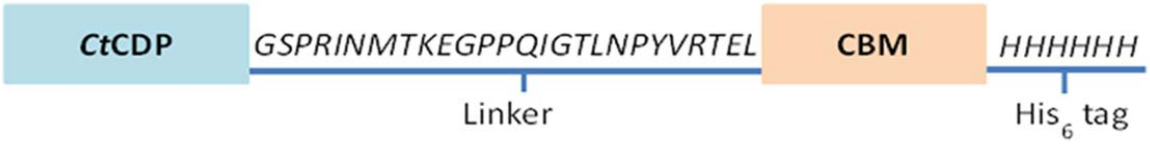


Figure. 4.1 A generalized structure of CDP-CBM fusion proteins. *Ct*CDP, *Clostridium thermocellum* cellodextrin phosphorylase; CBM, carbohydrate-binding module.

Around 40 mg of *Ct*CDP and four different fusion proteins were produced per liter LB medium. They were purified by immobilized metal affinity chromatography (IMAC) with yields of ca. 20%. SDS-PAGE analysis indicated that the purified proteins were homogeneous, and that the observed molecular weights were consistent with the estimated values, i.e. 115 kDa for *Ct*CDP, 126 kDa for *Ct*CDP-CBM3, 134 kDa for *Ct*CDP-CBM4, 130 kDa for *Ct*CDP-CBM6, and 138 kDa of *Ct*CDP-CBM9 (Fig. 4.2).

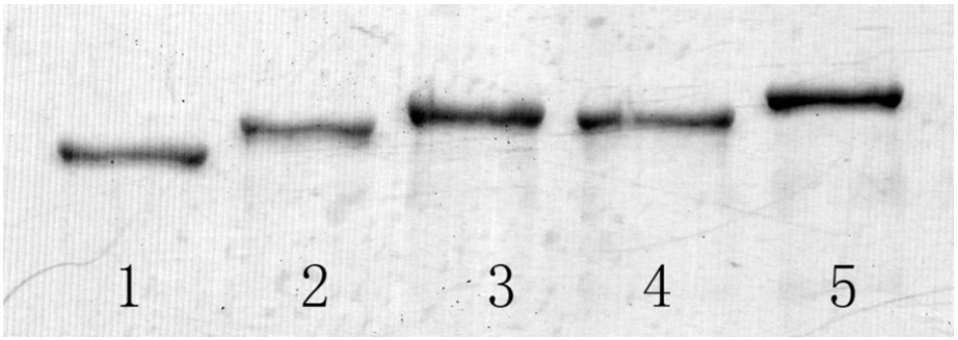


Figure. 4.2 Purity of the fusion proteins. Lane 1, CDP (MW 115 kDa); Lane 2, CDP-CBM3 (MW 126 kDa); Lane 3, CDP-CBM4 (MW 134 kDa); Lane 4, CDP-CBM6 (MW 130 kDa); Lane 5, CDP-CBM9 (MW 138 kDa). MW: molecular weight.

Basic characteristics of the fusion proteins

Since fusion of CBM may alter the activity and thermal stability of the catalytic domains (Karita et al. 1996; Kittur et al. 2003), the influences of pH and temperature on the activities of fusion

proteins were examined. Every chimeric protein displayed the same pH and temperature optimum as the wild-type *Ct*CDP (data not shown). The maximum activities were detected at pH 7.0 and 60 ^{o}C, in agreement with previous report (Krishnareddy et al. 2002).

Kinetics of *Ct*CDP and fusion proteins were determined in both synthesis and degradation directions on cellobiose and celloheptaose. For cellobiose, similar values of K_m and k_{cat} were observed in the synthesis direction among the wild-type *Ct*CDP and the fusion proteins, whereas no activities were detected in the degradation direction (Table 4.2a), consistent with the previous results (Krishnareddy et al. 2002; Sheth and Alexander 1967). The comparable kinetic parameters on cellobiose and optimal pH and temperature of *Ct*CDP and the fusion proteins suggested that fusion of CBM had little influence on the structure and function of the catalytic domain of *Ct*CDP on soluble substrate.

Table 4.2a. Kinetics of CDP-CBM fusion proteins on cellobiose

	Synthesis direction			Degradation direction		
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m
*Ct*CDP	0.78 ± 0.02	2.67 ± 0.04	3.42	ND	ND	-
*Ct*CDP-CBM3	0.59 ± 0.03	1.83 ± 0.08	3.10	ND	ND	-
*Ct*CDP-CBM4	0.53 ± 0.01	2.07 ± 0.08	3.90	ND	ND	-
*Ct*CDP-CBM6	0.76 ± 0.02	2.08 ± 0.05	2.73	ND	ND	-
*Ct*CDP-CBM9	1.08 ± 0.01	2.55 ± 0.02	2.36	ND	ND	-

Table 4.2b. Kinetics of CDP-CBM fusion proteins on celloheptaose

	Synthesis direction			Degradation direction		
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m
*Ct*CDP	1.29 ± 0.02	3.11 ± 0.05	2.41	1.82 ± 0.03	4.14 ± 0.10	2.27
*Ct*CDP-CBM3	1.93 ± 0.04	1.08 ± 0.02	0.56	2.71 ± 0.02	2.70 ± 0.01	1.00
*Ct*CDP-CBM4	0.29 ± 0.05	2.25 ± 0.05	7.76	1.56 ± 0.01	4.53 ± 0.03	2.90
*Ct*CDP-CBM6	5.24 ± 0.05	2.16 ± 0.02	0.41	1.53 ± 0.01	1.37 ± 0.02	0.90
*Ct*CDP-CBM9	1.46 ± 0.02	2.98 ± 0.04	2.04	0.62 ± 0.02	3.77 ± 0.06	6.08

CBMs had various impacts on *Ct*CDP activities on celloheptaose (Table 4.2b). With comparison to the wild-type *Ct*CDP, fusion of CBM3 decreased the turnover rates and the affinity between the enzyme and the substrate, resulting in reduced k_{cat} and increased K_m in both directions. In addition, the catalytic efficiencies (k_{cat}/K_m) were reduced from 2.41 to 0.56 $mM^{-1}s^{-1}$ in the

synthesis direction and from 2.27 to 1.00 $mM^{-1}s^{-1}$ in the degradation direction. Similar effects were also observed in case of CBM6, even though family 6 CBM has different binding properties from CBM3. In contrast with *Ct*CDP, the catalytic efficiencies of *Ct*CDP-CBM6 were decreased to 0.41 and 0.90 $mM^{-1}s^{-1}$ in the directions of glucan synthesis and degradation, respectively. Furthermore, addition of CBM4 and CBM9 attributed little to the turnover rate of *Ct*CDP, but, promoted the Michaelis constant in different directions. Fusion of CBM4 caused a four-fold decrease of K_m in the synthesis direction, while addition of CBM9 lowered K_m by nearly three folds in the degradation direction. Consequently, the catalytic efficiencies were increased to 7.76 $mM^{-1}s^{-1}$ for synthesis and to 6.08 $mM^{-1}s^{-1}$ for degradation by the addition of family 4 and family 9 CBM, respectively.

Binding properties on insoluble cellulose

The binding properties of *Ct*CDP and the chimeric proteins on insoluble cellulose were studied by incubating the proteins with RAC. Fig. 3 showed the binding isotherms of *Ct*CDP and the fusion proteins on RAC DP 14 and RAC DP 164. All of the fusion proteins had smaller dissociation constants (K_p) than *Ct*CDP. It suggested that CBMs enhanced the binding between enzyme and RAC. At 30 ^oC, the free energy of adsorption (ΔG) (Boraston et al. 2001) between *Ct*CDP and RAC DP 14 was -38.2 kJ mol^{-1}, very close to the energy change as *Ct*CDP bound to RAC DP 164 (ΔG = -38.1 kJ mol^{-1}) (Table 4.3). However, the interaction between the fusion proteins with short-chain RAC had smaller energy change (ΔG) than that on long-chain RAC, except in the case of *Ct*CDP-CBM4 that preferentially bound to RAC DP 164.

Table 4.3 Modified Langmuir isotherm parameters for the adsorption of *Ct*CDP and its fusion proteins on regenerated amorphous cellulose (RAC).

	RAC DP 14			RAC DP 164		
	A_{max} [a] (μmol g^{-1})	K_p [b] (L μmol^{-1})	ΔG [c] (kJ mol^{-1})	A_{max} [a] (μmol g^{-1})	K_p [b] (L μmol^{-1})	ΔG [c] (kJ mol^{-1})
*Ct*CDP	0.44 ± 0.04	3.83 ± 0.35	-38.2	0.52 ± 0.07	3.63 ± 0.49	-38.1
*Ct*CDP-CBM3	0.56 ± 0.03	1.51 ± 0.08	-35.9	0.88 ± 0.05	1.58 ± 0.09	-36.0
*Ct*CDP-CBM4	0.32 ± 0.02	2.90 ± 0.18	-37.5	2.92 ± 0.16	0.95 ± 0.05	-34.7
*Ct*CDP-CBM6	0.82 ± 0.06	1.03 ± 0.07	-34.9	1.27 ± 0.09	1.45 ± 0.10	-35.8
*Ct*CDP-CBM9	2.55 ± 0.10	0.71 ± 0.03	-34.0	4.93 ± 0.14	1.28 ± 0.04	-35.4

[a] A_{max}, maximum adsorption capacity (μmol [enzyme] g-1 [RAC]);
[b] K_p, dissociation constant (L μmol-1 [enzyme]);
[c] ΔG, free energy of adsorption (kJ mol-1). ΔG=-RT·ln(Kp) (Boraston et al. 2001), where R is a gas constant, and T represents temperature (K, T=303.15 K).

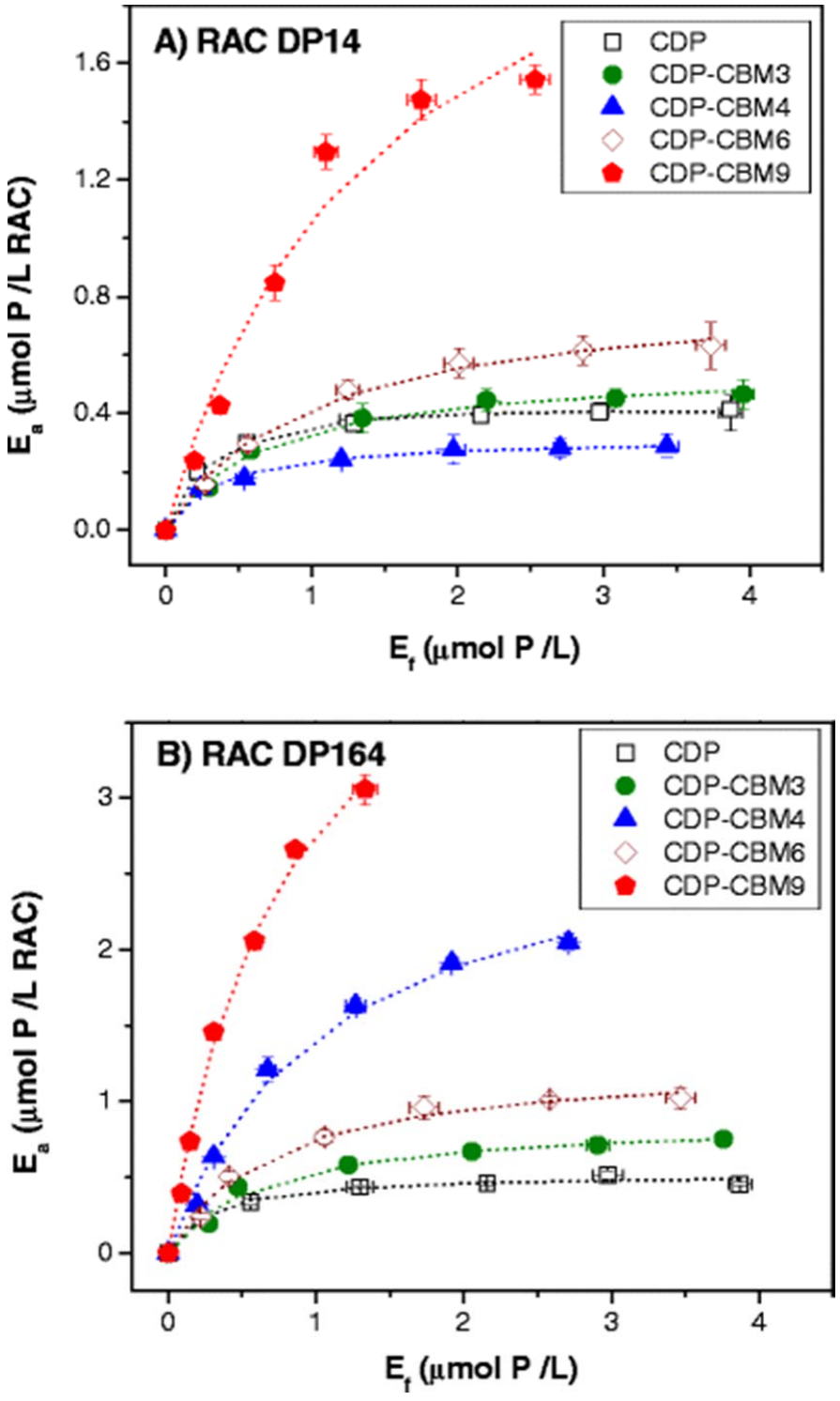


Figure 4.3 Binding isotherms of CDP-CBM fusion proteins to RAC. A, adsorption on RAC DP 14; B, adsorption on RAC DP164.

Addition of CBM3, CBM6, or CBM9 increased the maximum adsorption capacities (A_{max}) of *Ct*CDP to both RACs. In particular, *Ct*CDP-CBM9 had A_{max} approximately 6-fold and more than 9-fold higher than that of *Ct*CDP on RAC DP 14 and RAC DP 164, respectively (Table 4.3). By contrast, fusion of CBM4 lessened the binding capacities on the short-chain RAC (Fig. 4.3a), but promoted binding capacities to the long-chain RAC (Fig. 3b). The corresponding A_{max} were 0.32 μmol [*Ct*CDP-CBM4] g^{-1} [RAC DP14] and 2.92 μmol [*Ct*CDP-CBM4] g^{-1} [RAC DP164] (Table 4.3).

Activities on amorphous cellulose

Preliminary experiments had shown that wild-type *Ct*CDP was able to weakly phosphorolyse insoluble amorphous cellulose (Fig. 4.4). In order to obtain meaningful results for activity assay, the reactions on RAC DP 14 and RAC DP 164 were performed under different substrate and enzyme concentration. It was quite clear that *Ct*CDP, along with the chimeric proteins, had higher activities on RAC DP 14 than on RAC DP 164, especially in the degradation direction (Fig. 4.4). The difference may stem from the fact that reaction on RAC DP 14 started with a higher concentration of reducing end ([RE] = 2.8 mM), where phosphorolysis happens (Samain et al. 1995), than that on RAC DP 164 in which [RE] was 0.28 mM.

On RAC DP 164, *Ct*CDP and its fusion proteins, like other phosphorylases (e.g. maltodextrin phosphorylase), had higher activities of extending the glucan chain than of degrading it (Fig. 4.4b), because the conversion of phosphate into glucose 1-phosphate is thermodynamically unfavorable (Nidetzky et al. 1996). Nevertheless, a reverse trend was found on RAC DP 14. Both *Ct*CDP and CBM-tagged *Ct*CDPs degraded the short-chain RAC more efficiently than extending it (Fig. 4.4a). As the slightly soluble RAC DP 14 was degraded, the cascade reaction generated shorter substrates which were more soluble and more accessible to CDP. Thus the degradation reaction may be accelerated.

In addition, fusion of CBM3 and CBM6 reduced the activities of *Ct*CDP on amorphous cellulose. Fused CBM4 had little impact on *Ct*CDP, although *Ct*CDP-CBM4 displayed nearly 10% increases of catalytic efficiencies on RAC DP 14. Comparatively, CBM9 improved the specific activities of *Ct*CDP in both synthesis and degradation directions over 20% on the short-chain

RAC and more than 50% on the long-chain RAC. The specific activities of *Ct*CDP-CBM9 reached 18.9 and 5.9 U μmol^{-1} toward synthesis and degradation of RAC DP 14, respectively. One μmol *Ct*CDP-CBM9 also possessed 4.1 U for extension and 1.2 U for degradation of RAC DP 164 (Fig. 4.4).

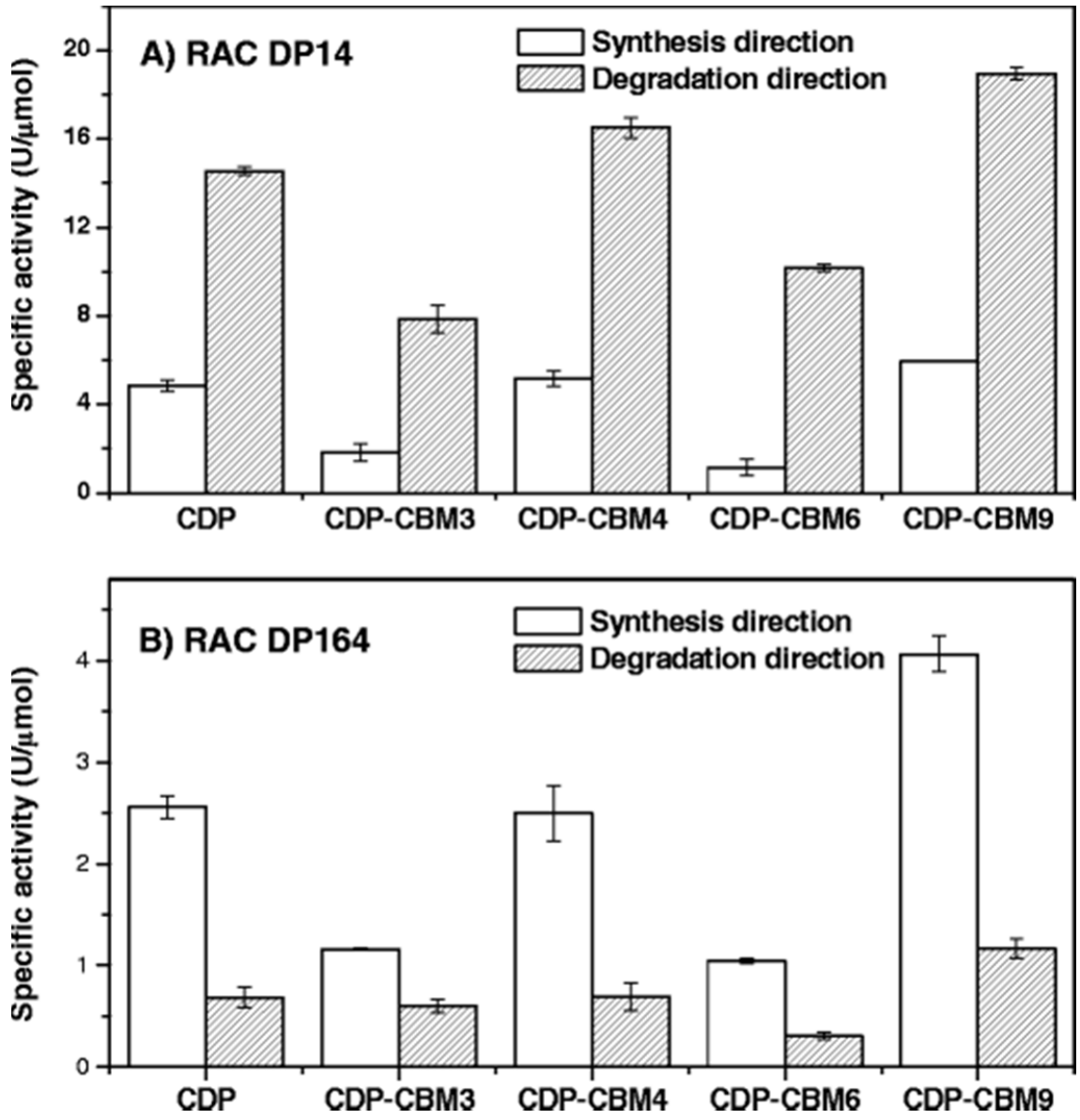


Figure 4.4 Specific activities of CDP-CBM fusion proteins to regenerated amorphous cellulose. A, specific activities on RAC DP14; B, specific activities on RAC DP164.

4.5 Discussion

Surface accessibility and end accessibility are recognized as two important factors in the enzymatic hydrolysis of cellulose (Rollin et al. 2010; Zhang and Lynd 2006). In nature, cellulose hydrolases developed a complex molecular architecture in which catalytic modules are appended to non-catalytic carbohydrate-binding modules that enrich enzyme onto its substrate, thereby increasing possibility of enzyme-substrate complex formation (McCartney et al. 2006). The *C. thermocellum* cellodextrin phosphorylase, a single modular protein, does not have an apparent CBM. We discovered that it has a very weak activity toward amorphous cellulose. To mimic cellulase performance, we constructed four *Ct*CDP-CBM fusion proteins and characterized them. Among them, fusion of a family 9-2 CBM from *T. maritima* Xyn10A expands the catalytic potential of *Ct*CDP towards insoluble cellulose.

Being a type C CBM, the family 9 CBM from *T. maritima* has a broad binding specificity (Table 4.1). Addition of CBM9 to *Ct*CDP did not influence its kinetics on soluble cellobiose, but increased the binding capacities on insoluble cellulose. It likewise improved the catalytic activities, to different extents, on celloheptaose and RAC. The carbohydrate-binding site of CBM9 has been described as a 'blind canyon', a groove on protein surface that is blocked at one end. Such a conformation accommodates only two sugar rings and makes CBM9 specifically bind to the reducing ends of sugars (Notenboom et al. 2001). Although both CBM9 and *Ct*CDP bind to the reducing end of cellulose, it was doubtful the sugar residues could be transferred from the binding site of CBM9 to the catalytic domain of *Ct*CDP. However, fusion of CBM9 may direct the enzyme toward certain reducing end enriched area so that the fusion protein *Ct*CDP-CBM9 could show higher activities on insoluble cellulose than *Ct*CDP. Moreover, fusion of CBM9 considerably increased the specific activities on RAC DP 164 while only marginally improved the activities on RAC DP 14, suggesting that the enriching benefit of CBM9 fusion strategy was limited on RAC DP 14 because of much higher reducing end availability.

Our results also suggested that different CBMs, with different binding specificities, affected *Ct*CDP to different extents. For instance, fusion of family 3 CBM from *C. thermocellum Cbh*A improved enzyme binding to insoluble cellulose but significantly decreased the activities on both celloheptaose and RAC. CBM3, a typical type A CBM, targeted *Ct*CDP-CBM3 to the crystalline

surface of cellulose where the trapped *Ct*CDP lost its freedom to capture the reducing ends of glucan chains, resulting in decreased activities. On the other hand, CBM3 cannot hold cellobiose. Therefore, *Ct*CDP-CBM3 and *Ct*CDP exhibited the similar kinetics toward cellobiose (Table 4.2a).

Type B CBM6 from *C. mixtus* Cel5B has two potential ligand binding sites (Henshaw et al. 2004). One (cleft A) is located within the loops connecting two β-sheets and recognizes two sugar residues from the non-reducing end, whereas the concave surface cleft B accommodates three to four saccharide units in both orientations. When *Ct*CDP-CBM6 was mixed with cellobiose, cleft A and *Ct*CDP may each adopt a cellobiose unit. Given that *Ct*CDP-CBM6 retained the activity of *Ct*CDP on cellobiose, it could be concluded that two binding processes happened in non-interacting sites and were independent of each other. On the contrary, fusion of CBM6 lowered the activities of *Ct*CDP on insoluble celluloses, including celloheptaose, RAC DP 14, and RAC DP 164. Since binding of CBM6 to insoluble cellulose involves synergistic interactions between cleft A and cleft B, cleft A may draw *Ct*CDP-CBM6 to the non-reducing termini once the fusion protein interacts with the cellulose. Similar to the case of *Ct*CDP-CBM3, it lessened the possibility of *Ct*CDP to reach the reducing end of cellulose, resulting in declined activities.

Through the comparison of the unequal impacts of fused CBMs on *Ct*CDP, it was concluded that fusion of CBM that anchors the enzyme at some specific positions may not promote the catalytic potential on the insoluble substrate. A desired CBM should direct the enzyme to the sites where the enzyme would form active enzyme-substrate complex. The success of chimeric *Ct*CDP-CBM9 over the wild-type *Ct*CDP provided a good example here. Nevertheless, it was worth noting that substrate properties also exerted different effects on the fusion protein. For example, CBM 4-2 from *R. marinus Xyn*10A is a type B CBM interacting with a range of single-stranded polysaccharides, including amorphous (but not crystalline) cellulose, β-1,3-glucan, and xylan, with approximately equal affinity in both orientations (Hachem et al. 2000). Different from other fusion proteins, *Ct*CDP-CBM4 exhibited distinct binding and catalytic properties towards RAC DP 14 and RAC DP 164. On RAC DP 14, *Ct*CDP-CBM4 displayed a smaller A_{max} but higher activities than *Ct*CDP, whereas *Ct*CDP-CBM4 had a bigger A_{max} on RAC DP 164, but similar activities as *Ct*CDP in both synthesis and degradation directions. For the latter observation, an

increase of binding capacities may not increase enzyme concentration around its target substrate sites (e.g. reducing ends of β-glucan chain), similar to the above discussion on *Ct*CDP-CBM3 and *Ct*CDP-CBM6. However, *Ct*CDP-CBM4 with low binding capacities but high catalytic activities on RAC DP 14 may be explained in respect of substrate properties, including solubility, chain length distribution, and accessibility of the reducing ends. Substrate solubility may play a critical role. Since RAC DP 14 is partially soluble (Zhang and Lynd 2004b) and the binding cleft of CBM4 prefers soluble substrate (Johnson et al. 1996; Simpson et al. 2002), the soluble fraction of RAC DP 14 attracted a large amount of *Ct*CDP-CBM4 in supernatant, leading to the decrease of A_{max} and the increase of K_p and absolute ΔG. Meanwhile, the interaction between soluble RAC with *Ct*CDP-CBM4 helped the enzyme access to the reducing ends and thus increased its activities.

Moreover, kinetic analysis demonstrated fusion of CBM4 and CBM9 promoted the enzyme affinities on celloheptaose in different directions. Note that CBM4 and CBM9 bind to different sites of cellulose (Blake et al. 2006; McLean et al. 2002). Above result implies CBM4 might direct the enzyme to the sites of celloheptaose where glucan synthesis was preferred, while CBM9 concentrates the fusion protein to the sites that favors glucan degradation. It was difficult to evaluate the validity of this hypothesis. However, it clearly suggested complex interactions between binding modules and catalytic modules in enzyme, between binding sites and catalytic sites in substrate, and between enzymes and substrates (Liu et al. 2010).

The linker joining *Ct*CDP with CBM is another factor that could potentially affect both binding and catalytic properties of the fusion protein (Liu et al. 2010). In this work, a natural linker was cloned from *C. thermocellum* Cthe 2196. It is a short and flexible peptide consisting of repeated units of proline, threonine, and glycine (Fig. 4.1). The sequence shows no putative consensus motif in the GenBank database. For construction of CBM-fused protein, the synthetic CBM domain routinely includes a fragment of its natural linker, if not all, to alleviate the possible steric hindrance of CBM on the activity of chimeric enzymes (Ahn et al. 2004; Kittur et al. 2003; Ravalason et al. 2009). However, it was still unclear the real effect of linker on the catalytic potential of chimeric enzymes. Concerted efforts would be focused on optimizing the linker length and composition for maximizing the catalytic potential of *Ct*CDP-CBM9.

Increasing enzyme activities on insoluble substrates is challenging (Wen et al. 2009; Zhang et al. 2006b). It is difficult to develop an active enzyme against insoluble substrates by directed evolution owing to the lack of reliable and efficient screening/selection methods (Liu et al. 2009). Since the mechanism how enzymes work with insoluble substrates is still unknown, engineering an enzyme toward insoluble substrates by rational design has been scarcely reported so far (Li et al. 2010). In this study, we discovered that *Ct*CDP had weak activities on cellohepatose and regenerated amorphous cellulose (RAC). To test whether addition of a CBM could improve the catalytic potential of *Ct*CDP on insoluble cellulose, *Ct*CDP was linked with four representatives of the three types of CBM (Table 4.1). The results demonstrated that addition of a CBM 9-2 from *T. maritima* Xyn10A significantly improved the catalytic activity of *Ct*CDP on both short-chain and long-chain RAC. The chimeric protein will be the first step to construct a highly active cellulose phosphorylase for the in vitro hydrogen production by SyPaB. The enhanced activity on RAC also makes it possible to develop an efficient screening/selection method for directed evolution of *Ct*CDP-CBM9 (Liu et al. 2010; Zhang and Zhang 2011). The activity of *Ct*CDP-CBM9 on solid cellulose might be further enhanced by rational design of linker length and composition in the future.

Chapter 5

Engineering a large protein by combined rational and random approaches: Stabilizing the *Clostridium thermocellum* cellobiose phosphorylase

5.1 Abstract

The *Clostridium thermocellum* cellobiose phosphorylase (*Ct*CBP) is a large protein consisting of 812 amino acids and has great potentials in the production of sugar phosphates, novel glycosides, and biofuels. It is relatively stable at 50 ^{o}C, but is rapidly inactivated at 70 ^{o}C. To stabilize *Ct*CBP at elevated temperatures, two protein engineering approaches were developed, i.e. site-directed mutagenesis based on structure-guided homology analysis and random mutagenesis depended upon various mutation rates. The former chose substitutions by comparison of the protein sequences of CBP homologs, utilized structural information to identify key amino acid residues responsible for the stability, and then created a few variants accurately. The latter constructed large libraries of random mutants using different mutagenesis frequencies. A novel combinational selection/screening strategy was employed to quickly isolate thermostability-enhanced and active variants. Several stability-enhanced mutants were obtained by both methods. Manually combining the stabilizing mutations identified from both rational and random approaches led to the best mutant (CM3) with the halftime of inactivation at 70 ^{o}C extended from 8.3 to 24.6 min. The temperature optimum of CM3 was increased from 60 to 80 ^{o}C. These results demonstrated that a combination of rational design and random mutagenesis formed a solid basis for engineering large proteins.

5.2 Introduction

Development of thermostable and hypersthermostable enzymes is a very active research area because it not only broadens the industrial applicability of enzymes, but also furthers our understanding of protein structure/function relationships (Kumar et al. 2000; Lehmann and Wyss 2001). Numberous methods have been employed to improve pronounced thermostabilization, including both non-biologically and biologically based methods (Polizzi et al. 2006). Non-biological methods use additives (e.g., surfactant, salt, reducing agents, etc.) or/and immobilize enzymes on a solid support to guard against denaturation (Jiang et al. 2001b; Ye et al. 2010). Biological methods aim to isolate enzyme variants from thermophiles or to enhance enzyme stability by protein engineering (Haki and Rakshit 2003; Hough and Danson 1999).

Protein engineering offers numerous promising advantages over the other methods (Böttcher and Bornscheuer 2010; Wen et al. 2009), often by two distinctive approaches: rational design and

directed evolution (Bornscheuer and Pohl 2001; Chen 2001). Directed evolution is an effective way to evolve biocatalysts with desired performance (Cherry and Fidantsef 2003; Schmidt-Dannert and Arnold 1999). It is independent of the knowledge of three-dimensional structures but does require effective screening or selection systems. The success of directed evolution is restricted by sequences explored, biases in mutagenesis methods, as well as the degeneracy of genetic codes (Lutz 2010). For example, beneficial mutations can be accumulated and be selected or screened from iterative rounds of directed evolution (Giver et al. 1998; Romero and Arnold 2009). On the other hand, thermostable enzyme variants can also be rationally designed by introducing hydrogen bonds, strengthening salt or disulphide bridges, improving core packing, optimizing surface charges, increasing rigidity with preferred residue substitutions, or stabilizing α-helix, β-turns or flexible termini or loops (Jaenicke et al. 1996; Lehmann et al. 2000). Rational protein design is less labor intensive than directed evolution, but requires extensive knowledge of protein structure and function (Polizzi et al. 2006; Zhang et al. 2006b). Among numerous protein design approaches, a semi-rational approach is to design mutants by comparison of the amino acid sequences of homologous enzymes (Perl et al. 2000; Van den Burg et al. 1998). Statistical analysis extracts recurring amino acid replacement trends, from which the respective consensus amino acids are presumed to contribute more than average to protein stability than the nonconsensus residues (Lehmann and Wyss 2001). This 'consensus concept' has been used to produce many new thermostable proteins (Steipe et al. 1994; Vázquez-Figueroa et al. 2007). However, the approach may be hampered by noises accompanied with the high levels of sequence divergence that may not only account for thermal adaptation (signal) but also a background of random genetic drift (noise) (Haney et al. 1999; Lutz 2010). In view of the pros and cons of every protein engineering method, a combination of directed evolution and rational design represents a more effective route to improving the properties and function of enzymes (Bornscheuer and Pohl 2001). An impressive example of successful combination of directed evolution and rational design is the improvement of stability of *Coprinus cinereus* heme peroxidase (Cherry et al. 1999). Other examples include increasing the thermal stability of *Bacillus subtilis* 3-isopropylmalate dehydrogenase by a combination of *in vivo* mutagenesis, *in vitro* evolution, and rational design (Akanuma et al. 1998; Akanuma et al. 1999), and creating a thermostable penicillin G aclylase and a thermostable glucose dehydrogenase by a structure-guided consensus approach (Polizzi et al. 2006; Vázquez-Figueroa et al. 2007).

Cellobiose phosphorylase (CBP, E.C. 2.4.1.20) catalyses the phosphorolysis of cellobiose into glucose 1-phosphate and glucose. It belongs to glycoside hydrolase (GH) family 94 and plays an important role in the energy-efficient metabolism of carbohydrates (Lou et al. 1996; Zhang and Lynd 2004a). CBP has been applied for hetero-oligosaccharide synthesis (Percy et al. 1998), sugar 1-phosphate production (De Groeve et al. 2009), and enzymatic hydrogen production (Ye et al. 2009). The *C. thermocellum* CBP (*Ct*CBP) consists of 812 amino acid residues. It is relatively stable at 50 ^{o}C, but the inactivation halftime is less than 10 min at 70 ^{o}C (Tanaka et al. 1995).

The objective of our work is to create a *Ct*CBP variant with improved thermal stability and comparable activity to the wild-type enzyme. We created the *Ct*CBP mutants by two parallel approaches: site-directed mutagenesis based on a structure-guided homology approach and random mutagenesis of the whole *ctcbp* gene *via* error-prone PCR. The best mutant based on a combination of both approaches significantly extended the inactivation halftime of *Ct*CBP and had a 3-fold increase of the specific activity at 80 ^{o}C. Such a thermostability-enhanced active variant would be of essence to reduce the costs of *in vitro* hydrogen production by Synthetic Pathway Biotransformation (SyPaB) (Zhang 2010a).

5.3 Materials and Methods

Materials

All chemicals were reagent-grade, purchased from Sigma (St. Louis, MO), unless otherwise noted. *Clostridium thermocellum* ATCC 27405 genomic DNA was a gift from Dr. Jonathan Mielenz at the Oak Ridge National Laboratory (Oak Ridge, TN). The Luria-Bertani (LB) medium and M9/cellobiose medium were prepared as described elsewhere (Liu et al. 2009). Regenerated amorphous cellulose (RAC) was prepared from Avicel after water slurrying, cellulose dissolution in H_3PO_4, and regeneration in water (Zhang et al. 2006a).

Bacterial strains and plasmids

Escherichia coli JM109 was used for cloning and mutant library construction; *E. coli* Rosetta BL21 (DE3) was employed for CBP production. The *cbp* gene was amplified from genomic

DNA of *C. thermocellum* using primer P1 and P2 (SOM-5 Table 1). The DNA fragments were digested by *Pst*I and *Bam*HI, and then ligated into the digested plasmid pUC19 (New England Labs, Ipswich, MA) to give plasmid pUCB. For protein characterization, the wild-type and mutant *cbp* genes were amplified from plasmid pUCB with the primers P2 and P3 (SOM-5 Table 1). The PCR product was digested by *Bam*HI and *Xho*I, and ligated with *Bam*HI/*Xho*I digested pCIG vector (Hong et al. 2008a) for the plasmid pCIB. All constructs described herein were verified by DNA sequencing (MCLab, San Francis, CA).

Structure-guided homology analysis

Calculation of the CBP consensus sequence was performed as follows: 55 unique CBP sequences, identified by PSI-BLAST of the amino acid sequence of *Ct*CBP (http://blast.ncbi.nlm.nih.gov/Blast.cgi), were aligned together by the program ClustalW from BioEdit (Carlsbad, CA). Nine CBP sequences, which exhibited the highest identities with the *Ct*CBP (73.9% - 56.4%), were chosen for homology analysis. Among them, *Ia*CBP is from *Ignisphaera aggregans* (growth temperature: 95 ^{o}C), *Tn*CBP is from *Thermotoga neapolitana* (growth temperature: 90 ^{o}C), *Tm*CBP is from *Thermotoga maritime* (growth temperature: 80 ^{o}C), *Dt*CBP is from *Dictyoglomus thermophilum* (growth temperature: 78 ^{o}C), *Csa*CBP is from *Caldicellulosiruptor saccharolyticus* (growth temperature: 70 ^{o}C), *Cst*CBP is from *Clostridium sterocorarium* (growth temperature: 65 ^{o}C), *Cu*CBP is from *Cellulomonas uda* (growth temperature: 50 ^{o}C), *Bf*CBP is from *Butyrivibrio fibrisolvens* (growth temperature: 37 ^{o}C), and *Cg*CBP is from *Cellvibrio gilvus* (growth temperature: 30 ^{o}C) (Table 5.1). Following sequence alignment, the candidates for substitution were selected by applying the following criteria: 1) at the certain residue the substitutions are obviously coupled with the change of growth temperatures; 2) the substitutions prefer the consensus amino acid from thermophiles and hyperthermophiles (Lehmann and Wyss 2001); 3) the substitutions are not located in stabilization center or appear in pairs (SOM-5 Table 2) (Dosztányi et al. 1997; Dosztányi et al. 2003); and 4) the substitutions don't contradict any structural knowledge with respect to stability enhancement. For example, if the mutation is found in a helix, a helix-stabilizer is not changed to a helix-desabilizer (Vázquez-Figueroa et al. 2007).

Site-directed mutagenesis

Site-directed mutagenesis was performed with the Phusion site-directed mutagenesis kit (New England Labs, Ipswich, MA) according to the manufacturer's instructions. Each reaction (50 μL) contained 100 pg of template plasmid pCIB and 25 pmol of each mutagenic primer. The primers were commercially phosphorylated at the 5' end and listed in SOM-Table 1. Thermal cycling was performed in Eppendorf Mastercycler pro Thermal cyclers (Hauppauge, NY) with the following cycling conditions: 98 ^{o}C for 30 s, 25 cycles of 98 ^{o}C for 10 s, 54-61 ^{o}C for 20 s, and 72 ^{o}C for 5 min, and finally 72 ^{o}C for 10 min. The PCR products were gel-purified by Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA), self-ligated by quick ligase kit (New England Labs, Ipswich, MA), and then transformed into *E. coli* Rosetta BL21 (DE3).

Construction of random mutant libraries

Random mutagenesis of the *Ctcbp* gene was performed in a reaction mix containing 0.2 ng/uL plasmid pUCB as the template, 10 mM Tris-HCl (pH 8.3), 0.2 mM dATP, 0.2 mM dGTP, 1 mM dCTP, 1 mM dTTP, 0.4 μM each of the primers P4 & P5 (SOM-5 Table 1), 30 mM KCl, $MgCl_2$, $MnCl_2$, and 2.5 U Taq polymerase (New England Labs, Ipswich, MA). Three mutagenic libraries with different mutation frequencies were made by varying concentration of Mg^{2+} and Mn^{2+}, *i.e.* 5 mM $MgCl_2$ for the library (Library L) with a low mutation frequency, 5 mM $MgCl_2$ plus 3 mM $MnCl_2$ for the one (Library O) with an estimated optimal mutation frequency (SOM-Fig. 4), and 7 mM $MgCl_2$ with 5 mM $MnCl_2$ for the library (Library H) with a high mutation frequency (Arnold and Georgiou 2003). The cycling scheme was 94 ^{o}C for 2 min, 30 cycles of 94 ^{o}C for 1 min, 54 ^{o}C for 1 min, and 72 ^{o}C for 3 min, followed by 72 ^{o}C for 10 min. The 2852 bp PCR products, including the entire *cbp* gene, 244 bp of upstream sequence, and 175 bp of downstream sequence, were purified using Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA). The resulting DNA fragment was then digested by *Pst*I and *Bam*HI, ligated with plasmid pUC19, and transformed into *E. coli* JM109.

Selection and screening

Putative *Ct*CBP mutants were isolated by a combinatorial selection/screening approach, similar to our previous work (Liu et al. 2009). The whole process consisted of two steps that involve the selection for mutants with adequate CBP activity and subsequent screening for improved

thermostability. Pooled transformants were first spread on solid agar plates (selection plates) with 1.5% agar, M9 minimal medium with 0.4% cellobiose, 75μg/mL ampicillin, and 0.1 mM IPTG. Meanwhile, a small fraction of the transformed cell from each library was grown on a LB plate containing 10μg/mL ampicillin, which would be used as the reference plate. All the plates were incubated at 30 ^{o}C until the colonies could be visualized easily. Next, the colonies were transferred to nylon membranes, subjected to heat treatment at 80 ^{o}C for 10 min, and eventually lysed in situ on the nylon membranes, by which intracellular CBP was released. The membranes with released CBP were overlaid on the screening plates that contained M9 minimal medium with 1% cellobiose, 100μg/mL ampicillin, 0.5% agar, and an indicator strain *E. coli* JM109/pUC19. Since the indicator strain is able to utilize glucose and glucose 1-phosphate but not cellobiose, after heat treatment the survived CBP variants could digest cellobiose and then support the growth of indicator strain. Therefore, the colonies (white spots) in the screening plates would help identify the clones expressing thermostable CBP mutants.

Colonies on the reference plates were counted to estimate the library sizes, and the mutation rate of each library was determined by sequencing, in both directions, the *Ctcbp* gene from 6-10 randomly picked clones. The selection power is calculated from ratio of the number of colonies on the M9-cellobiose plates, where only the transformants expressing active CBP can grow, to that on the LB reference plates, where all transformants can grow.

Protein production and purification

Wild-type CBP and its thermostable mutants were produced from the expression vector pCIB in *E. coli*, firstly as fusion protein CBM-intein-CBP. The *E. coli* strain was grown in LB medium at 37 ^{o}C with a rotary shaking rate of 250 rpm. When the absorbance of the culture (A_{600}) reached 0.8, IPTG was added to a final concentration of 0.25 mM and the growth temperature was decreased to 16 ^{o}C. After overnight culture, the *E. coli* cells were harvested by centrifugation and resuspended in a 50 mM HEPES buffer (pH 7.2). The cell suspension was sonicated, and the cell debris was removed by centrifugation (10,000g for 20 min at 4 ^{o}C). Then the fusion protein in the supernatant was bound to regenerated amorphous cellulose. After pH adjustment for intein self-cleavage, the cleaved CBP protein was obtained in the supernatant (Hong et al. 2008a). Finally, the protein solution was dialyzed with a 50 mM HEPES buffer (pH 7.2) and

concentrated with a protein concentration kit from Millipore (Billerica, MA) with a molecular weight cut-off of 50,000. The concentration of purified protein was determined by the Bradford method with bovine serum albumin (BSA) as the standard. Enzyme purity was checked by SDS-PAGE.

Enzyme essay

All enzymatic reactions were conducted in 5-ml glass culture tubes (12 × 75 mm, Fisher Scientific). The enzyme activities were assayed at 40 ^{o}C in 50 mM HEPES buffer (pH 7.2) containing 30 mM cellobiose, 10 mM glucose 1-phosphate, 1 mM DTT, and 1 mM Mg^{2+} unless otherwise noted. Kinetic parameters (K_m and k_{cat}) were determined by non-linear regression of Michaelis-Menten data via CurveExpert v1.4 (Hixson, TN). The kinetics were examined at 40 ^{o}C in a 50 mM HEPES buffer (pH 7.2) containing 1 mM Mg^{2+}, 1 mM DTT, and various substrate concentrations between 0.2 and 5 times of their respective K_m values. Enzyme concentrations were set at 5.0 mg/L for most assays. The reactions were stopped by placing the tubes in a boiling water bath for 10 min. The product Pi was measured by the mild pH phosphate assay as described elsewhere (Saheki et al. 1985). One unit of cellobiose phosphorylase was defined as the amount of enzyme that generates one μmole of phosphate per min.

Temperature effects and thermostability

The effects of temperature were examined by measuring the enzyme activities at various temperatures under standard assay conditions. Halftime of thermal inactivation was determined by incubating the enzymes (0.05 mg/mL) in 50 mM HEPES buffer (pH 7.2) at 70 ^{o}C for different time intervals, as described before (Liu et al. 2009). After the incubation, samples were chilled on ice immediately. Then the residual activity was assayed at 40 ^{o}C as described above. The activation energy (E_a) was determined from the slope of the Arrhenius plot.

Homology modeling

The structural models of *Ct*CBP and its variants were created online by ESyPred3D (Lambert et al. 2002). The crystal structure of *Cellvibrio gilvus* cellobiose phosphorylase (*Cg*CBP, PDB: 2CQT) was chosen as the template because *Cg*CBP is the only CBP with a solved 3D structure,

and shares high catalytic and sequence similarity (63%) with *Ct*CBP (Hidaka et al. 2006; Liu et al. 1998). Before structural analysis, every model was verified for consistency with known protein folds and allowed ϕ and ψ angles by Insight II (Accelrys Inc., San Diego, CA), both with default settings. For each selected protein, the structure was visualized by VMD (UIUC, Urbana-Champaign, IL), and the protein contact map was built by the Contact Map plugin of VMD 1.8.6 (Humphrey et al. 1996).

5.4 Results

Site-directed mutagenesis based on structure-guided homology analysis

The amino acid sequence of *C. thermocellum* cellobiose phosphorylase (*Ct*CBP) was compared with nine homologous proteins for identifying the candidates for site-directed substitutions. The ten CBPs having more than 50% sequence identities were chose from ten species with different growth temperatures. For example, *Ia*CBP, *Tn*CBP, along with *Tm*CBP are from hyperthermophiles (growth optimum > 80 ^{o}C); *Dt*CBP, *Csa*CBP, *Cst*CBP, *Cu*CBP, as well as *Ct*CBP are from thermophiles (growth optimum > 45 ^{o}C); *Bf*CBP and *Cg*CBP are from mesophiles (growth optimum > 25 ^{o}C).

Seven mutations were identified by the structure-guided homology analysis (Table 5.1), and individually introduced to the wild-type *Ctcbp* gene, except Q130H and K131Y. Q130 and K131 interact with each other, and both function as the stabilization center (SOM-5 Table 2). They are surrounded by aromatic amino acids, *e.g.* Y71, F72, Y103, F119, Y258, and F260, as illustrated in contact map (SOM-5 Fig. 1) and in Fig 1A & B. The mutations, Q130H together with K131Y, create new π-π interactions, resulting in an extension of the inactivation halftime at 70 ^{o}C from 8.34 to 14.1 min. K201P, N292K, and S411G, all found in the surface loops. K201P was chosen because the pyrolidine ring of Pro restricts the number of conformations and eliminates entropy gain in the denatured state (Frare et al. 2005). Replacement of Asn 292 by Lys might increase the number of ion pairs (Wong et al. 2003), while substitution of Ser 411 by Gly was because (hyper)thermophilic proteins statistically contain more Gly and less Ser (Kumar and Nussinov 2001). As a consequence, the mutation S411G improved the protein stability (Fig. 5.1 C & D, and Table 5.2), whereas K201P and N292K reduced the thermostability of *Ct*CBP (data not shown). The fifth site-directed mutant A423S was designed to stabilize the α-helix by capping

the interactions among W419, S423, and Y427 (Fig. 5.1 E & F). It practically increased the inactivation halftime to 11.3 min (Table 5.2). The most striking effect on protein stability occurred upon the mutation of A781. The A781K mutant offered a better packed C-terminal domain (*e.g.* smaller and less cavities as shown in Fig 5.1 G & H) and enhanced the inactivation halftime to 15.3 min (Table 5.2).

Table 5.1 Putative stabilizing mutations identified by structure-guided homology analysis

CBP ID	Origin	Growth Temp.	Amino Acids at the Potential Residues[a]					
			130-131	201	292	411	423	781
*Ia*CBP	*I. aggregans*	95 °C	WW	**P**[b]	L	**G**[b]	**S**[b]	**K**[b]
*Tn*CBP	*T. neapolitana*	90 °C	HH	S	**K**[b]	G	S	K
*Tm*CBP	*T. maritima*	80 °C	**HY**[b]	P	K	G	S	K
*Dt*CBP	*D. thermophilum*	78 °C	HY	P	K	S	S	K
*Csa*CBP	*C. saccharolyticus*	70 °C	HY	P	K	S	S	K
*Cst*CBP	*C. stercorarium*	65 °C	NQ	P	I	S	A	N
*Ct*CBP	*C. thermocellum*	60 °C	**QK**[c]	**K**[c]	**N**[c]	**S**[c]	**A**[c]	**A**[c]
*Cu*CBP	*C. uda*	50 °C	QK	Q	A	S	G	G
*Bf*CBP	*B. fibrisolvens*	37 °C	TK	K	Q	S	G	A
*Cg*CBP	*C. gilvus*	30 °C	QK	R	A	S	G	A

[a] The residues were numbered based on the protein sequence of *Ct*CBP;
[b] The chosen amino acids (also marked in red) were introduced to *Ct*CBP to replace their counterparts ;
[c] The residues of *Ct*CBP (also marked in bold) were chosen to mutate.

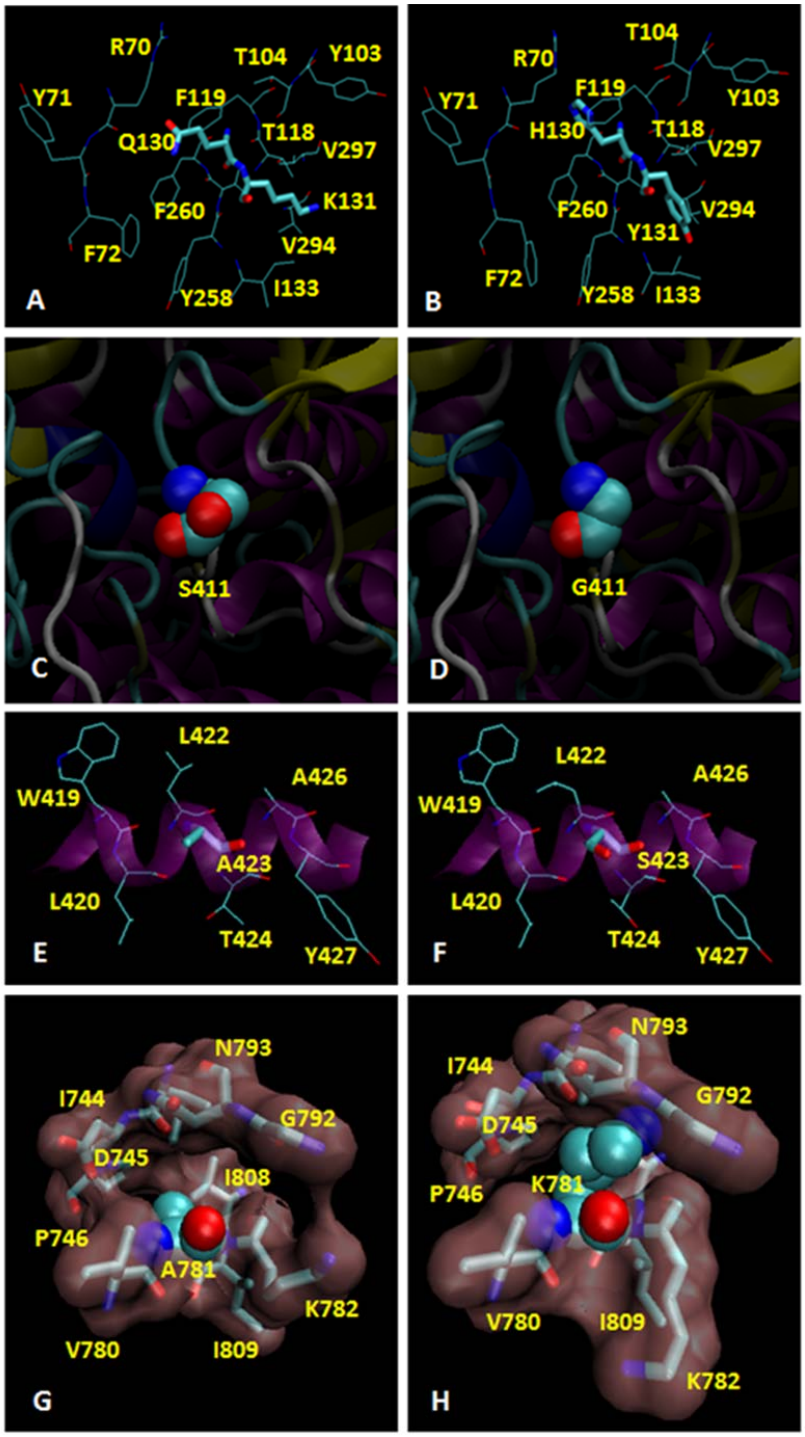


Figure 5.1 Structural interpretation of the residues that contribute the differences in thermostability between *Ct*CBP and its mutants. **A**, the residues surround Q130 and K131 in *Ct*CBP; **B**, the residues surround H130 and Y131 in the mutant Q130H/K131Y; **C**, surface portion around S411 in *Ct*CBP; **D**, surface portion around G411 in the mutant S411G; **E**, an α-helix containing A423 in *Ct*CBP; **F**, an α-helix containing G423 in the mutant A423G; **G**, a structural cavity that surrounds A781 in *Ct*CBP; **H**, a structural cavity that surrounds K781 in the mutant A781K.

As the individual mutations were combined in successive rounds of site-directed mutagenesis they were found to be approximately multiplicative in their effects on both improved thermostability and reduced activity. The thermal stability effects were cumulative. The final combined mutant I (CM1) from the modified structure-guided homology approach (Q130H, K131Y, S411G, A423S, and A781K) lengthened the inactivation halftime at 70 ^{o}C by more than 2 folds to 17.7 min (Table 5.2).

Table 5. 2 Characterization of the thermostability-enhanced mutants.

Enzyme	$T_{1/2}$ (min)[a]	Kinetics		
		k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m
Wild-type	8.3 ± 0.2	3.40 ± 0.11	16.5 ± 0.58	0.206
Q130H & K131Y	14.1 ± 0.3	4.50 ± 0.14	18.8 ± 0.24	0.239
S411G	13.6 ± 0.3	1.32 ± 0.18	6.94 ± 0.20	0.189
A423S	11.3 ± 0.4	2.96 ± 0.38	10.3 ± 0.45	0.287
A781K	15.3 ± 0.2	2.90 ± 0.25	11.0 ± 0.19	0.263
CM1[b]	17.7 ± 0.4	2.13 ± 0.16	15.9 ± 0.33	0.134
M52[c]	16.6 ± 0.8	0.85 ± 0.21	15.3 ± 0.42	0.056
M52m[d]	15.3 ± 0.3	3.12 ± 0.15	19.0 ± 0.61	0.164
CM2[e]	17.4 ± 0.4	1.44 ± 0.21	7.92 ± 0.41	0.182
CM3[f]	24.6 ± 0.3	1.52 ± 0.23	14.7 ± 0.34	0.103

[a] $T_{1/2}$ denotes the halftime of thermal inactivation at 70 ^{o}C ;

[b] CM1 represents the mutant with combined mutations I (Q130H, K131Y, S411G, A423S, A781K);

[c] M52 is a positive mutant obtained from moderate-frequency random mutagenesis, which contained five mutations, i.e. R48K, K142R, R189L, A423S, and V526A;

[d] M52m is an alternative mutant to M52 with a reverse mutation L189R. So it held four mutations, such as R48K, K142R, A423S, and V526A;

[e] CM2 represents the mutant with combined mutations II (R48K, K142R, A423S, V526A, and A781K);

[f] CM3 represents the mutant with combined mutation III (R48R, Q130H, K131Y, K142R, S411G, A423S, V526A, and A781K).

Random mutagenesis with different mutation frequencies

Three *Ctcdp* mutant libraries were generated with different mutation rates ranging from 0.14 ± 0.02 % in Library L, to 0.28 ± 0.07 % in Library O, and to 1.02 ± 0.44 % in Library H. Three libraries had sizes of *ca.* 1.0×10^4. The mutation generated included all possible transitions and transversions with errors biased toward AT to GC changes.

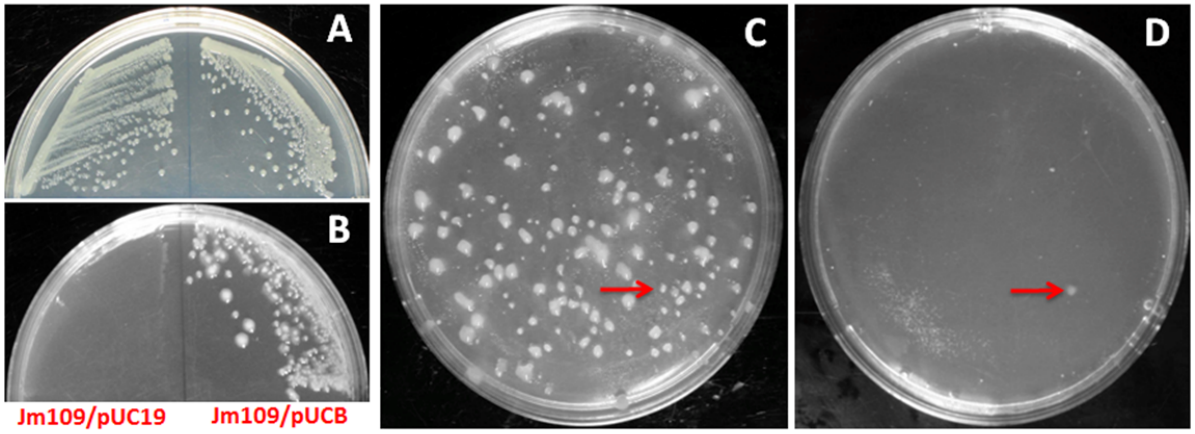


Figure 5.2 The combinatorial selection/screening strategy. *E. coli* strain JM109/pUCB and the control JM109/pUC19 were streaked on the plates with LB medium (**A**) or M9 minimal medium plus 0.4% cellobiose (**B**). **C**, a selection plate with the colonies expressing active cellobiose phosphorylase; **D**, the screening plate corresponding to (**C**). The red arrows, pointing out the growth of indicator strain (JM109/pUC19) in the selection plate, highlighted the putative mutants in the selection plate.

Selection was designed based on the hypothesis that introduction of a heterologous cellobiose phosphorylase enables non-cellobiose-utilizing *E. coli* to grow in the M9 synthetic medium containing cellobiose as the sole carbon source (M9-CB plate). On LB plates both *E. coli* JM109/pUC19 (a negative control) and *E. coli* JM109/pUCB that produces CBP grew well (Fig. 5.2A), while on the M9-CB plates only *E. coli* JM109/pUCB grew (Fig. 5.2B). These results validated the selection hypothesis. Since the selection approach ensured only the tranformants with active CBP expressed could survive, the active CBP mutants could be easily isolated from the mutant libraries (Fig. 5.2C). The difference of colony numbers in selection plates (M9-CB plates) against reference plates (LB plates) inferred the selection powers were 63.7% ± 2.1%, 45.2% ± 3.4%, and 12.1% ± 1.9% for Library L, Library O, and Library H, respectively.

The following screening method was designed to further identify the thermostability-enhanced CBP mutants. The colonies on the selection plates were imprinted on nylon membranes, and lysed in situ on membranes by heat. The heat treatment also deactivated the negative mutants and the wild type. Hence only the desired mutants remained active to degrade cellobiose and support the growth of the indicator strain *E. coli* JM109/pUC19 in the screening M9 plates containing cellobiose as a sole carbon source (Fig. 5.2D).

Four putative CBP mutants were found from the Library L, one was identified from the Library O, but none was isolated from the Library H. All of the five putative mutants were purified and characterized. As a result, only the one (namely M52) from the Library O exhibited a considerably enhanced thermostability with the inactivation halftime at 70 ^{o}C extended to 16.6 min (Table 5.2). However, the turnover number (k_{cat}) and the catalytic efficiency (k_{cat}/K_m) of M52 decreased to 0.85 s^{-1} and 0.056 mM^{-1} s^{-1} from 3.40 s^{-1} and 0.206 mM^{-1} s^{-1} of the wide type, respectively (Table 5.2).

Identification of the beneficial mutations

The mutant M52 contains 7 mutations in DNA sequence, five of which resulted in amino acid substitutions (Table 5.3). Consensus analysis indicated that R189 was strictly conserved among the 10 CBP homologies so that the mutation R189L may be the main reason for decrease in activity. Replacement of leucine back to the conserved arginine did restore the activity of M52, but decreased its thermal stability slightly, as shown as M52m in the Table 5.2.

It is interesting to note that the stabilizing mutation A423S identified by rational design was also found in M52. Reverse mutation R142K of the M52m and single mutation K142R of the wild type slightly reduced the halftime of inactivation at 70 ^{o}C to 12.4 min and 11.3 min, respectively (Fig. 5.3). K142R was confirmed to be a stabilizing mutation. Either reverse mutation K48R or A526V reduced the inactivation halftime of M52m. However, single mutation R48K and V526A cast deleterious effects on the thermostability of *Ct*CBP, shortening the inactivation halftime to 8.0 min and 7.9 min, respectively (Fig. 5.3). The results suggested that the combination of R48K and V526A improved the thermostability of M52m, even though introduction of R48K or V526A, individually, had negative impacts on the stability of *Ct*CBP.

Table 5.3 Consensus analysis of the mutations found in M52

DNA mutation	Amino acid mutation	Consensus analysis [a]
C9T	-	-
G143A	R48K	50% R, 50% K
A425G	K142R	60% K, 30% R, 10% T
G566T	R189L	100% R, 0% L
C681T	-	-
G1267T	A423S	50% S, 30% G, 20% A
T1577C	V526A	60% V, 40% A

[a] Consensus analysis was conducted based on the homology of the ten CBPs listed in Table 5.1.

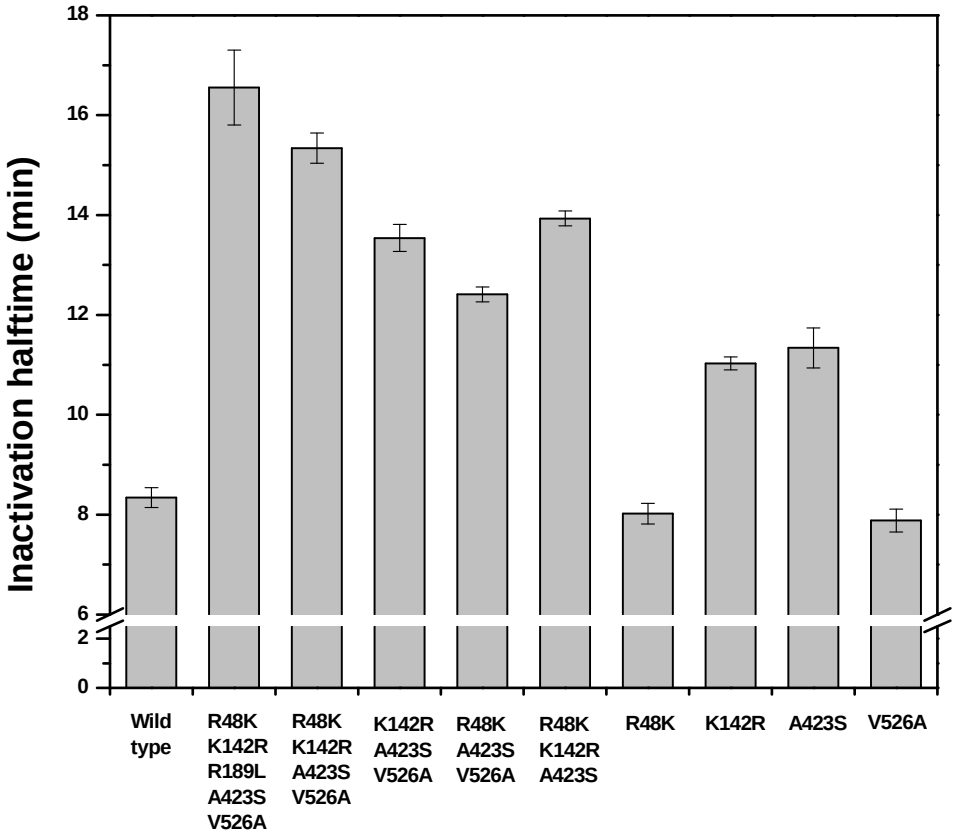


Figure 5.3 Inactivation halftime (at 70 ^{o}C) for the mutants in relation to M52.

Combining the stabilizing mutations from rational design and directed evolution

To further improve the thermal stability, the substitution A781K, with nearly 2 –fold increase of inactivation halftime, was firstly introduced into M52m, generating a new mutant II (CM2, as shown in Fig. 5.4). The inactivation halftime of CM2 was thus extended to 17.4 min (Table 5.2). The incremental effects were more significant as we incorporated all the stabilizing mutations identified by rational design to the M52m. The final mutant with combined mutations III (CM3, with R48K, Q130H, K131Y, Y142R, S411G, A423S, V526A, and A781K) had inactivation halftime at 70 ^{o}C of 24.6 min, nearly 3-fold longer than the wild type (Table 5.2). In contrast to the wild type, the temperature optimum of CM3 had been raised from 60 to 80 ^{o}C (Fig. 5.5), with the activation energy (E_a) decreased from 59.2 to 31.7 kJ mol^{-1}. The catalytic efficiency was 0.103 mM^{-1} s^{-1}, nearly half of the wild type at 40 ^{o}C (Table 5.2). However, the specific activity of CM3 was 3.43 U mg^{-1}, 3-fold higher than the wild type (1.03 U mg^{-1}) at 80 ^{o}C (Fig. 5.5).

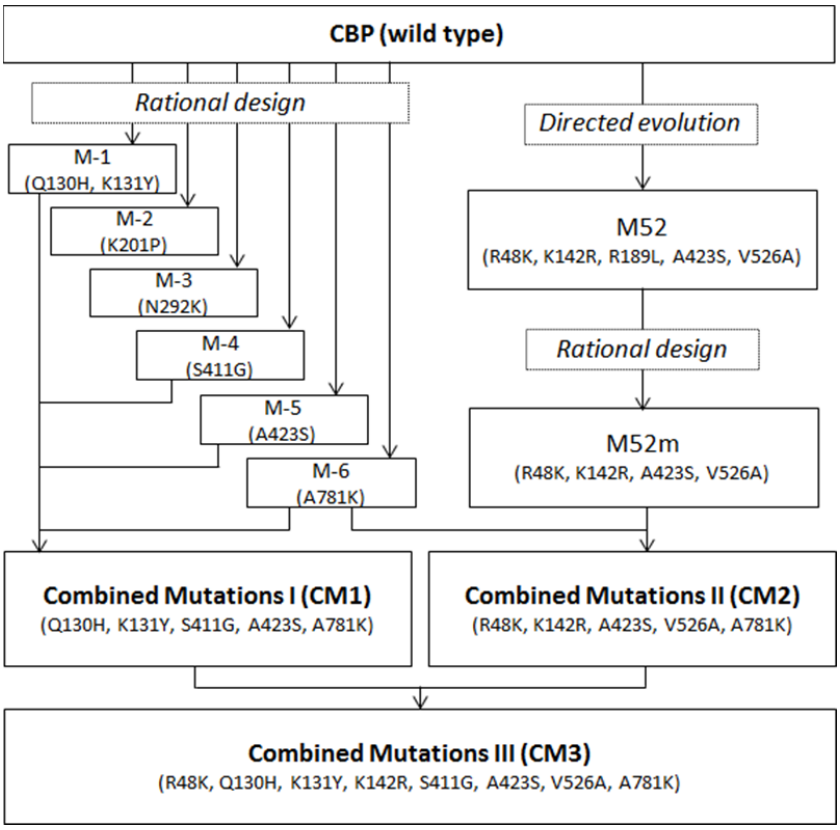


Figure 5.4 Flowchart illustrating the development of thermostability-enhanced active *Ct*CBP mutants by a combination of rational design and directed evolution.

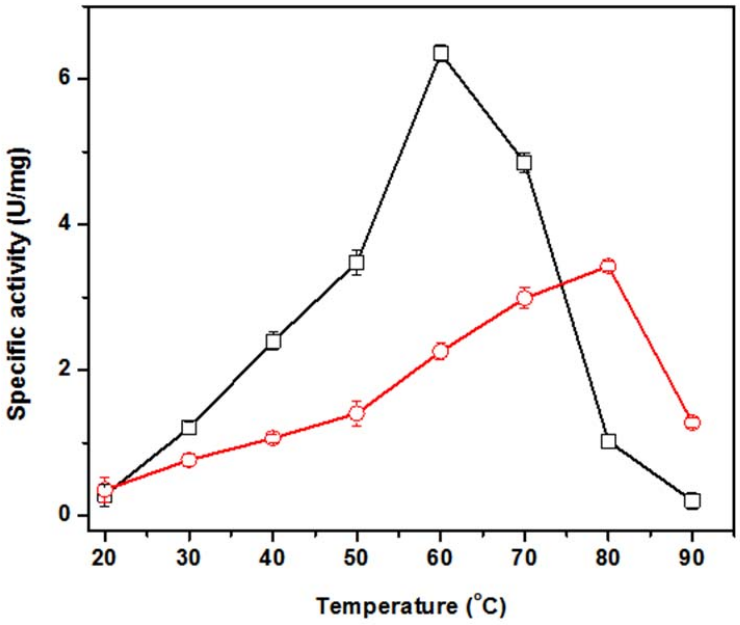


Figure 5.5 Temperature optimum of *Ct*CBP and the mutant CM3. *Ct*CBP: black rectangles, CM3: red circles.

5.5 Discussion

Engineering *Ct*CBP is challenging because it is a large dimeric enzyme. Since the effort to construct combinatorial libraries increases substantially for large proteins and a large number of variants would tend to be inactive, directed evolution of *Ct*CBP has a low probability of success for whole sequence mutagenesis, especially in a traditional manner (Bloom and Arnold 2009; Zeldovich et al. 2007). Therefore, De Groeve et al. converted *Cellulomonas uda* cellobiose phosphorylase to lactose phosphorylase by running a random mutagenesis restricted to the residues between T216 and V757 (542 out of the total 822 residues) (2009). Additionally, *Ct*CBP itself is a thermostable protein (Tanaka et al. 1995). The rational attempts to develop thermostability-enhanced variants is also risky because even the mechanisms that stabilize *Ct*CBP under modest conditions (<50 ^{o}C) remain unclear.

In this study, we fulfilled the task by two complementary approaches: structure-guided homology analysis and random mutagenesis with different mutation rates (Fig. 5.4). For the

structure-guided homology analysis, the homologous set included ten diverse CBPs (to reduce biases), had high sequence identity (to minimize random drift), and adapted to very disparate temperatures (to maximize signal) (Haney et al. 1999). In comparison with previous homology-based protein design that set the vote weight of each species identical and focused on replacing poorly conserved residues at a given position with the most representative (or the consensus) type of residue (Jiang et al. 2001; Lehmann et al. 2000), our method favored the residues where substitutions are tightly coupled with growth temperature and preferred the consensus residue among the thermophilic and hyperthermophilic counterparts. Then structural information was utilized to reduce the number of residues to be mutated for stabilization. Consequently, of the six mutants, four (~67%) rendered more stable effects, and the effects were additive. The best mutant A781K involves three base-pair changes in DNA (from GCT to AAA or AAG), which is often inaccessible by regular error-prone PCR. Two unsuccessful mutations, K201P and N292K, both were located in the protein surface. Despite extensive investigation of protein stability enhanced by surface mutations (Goldstein 2007; Perl et al. 2000; Wong et al. 2003), no general rule was available to predict the stability of proteins upon surface mutations. The incomplete understanding of underlying mechanisms accounts for the failed rational attempts, and further points out a major drawback that restricts the application of rational design.

Random mutagenesis was conducted over the entire *Ctcbp* gene (2433 bp plus 419 bp pre- and pro sequences) to maximize the evolutionary search. A novel combinatorial selection/screening approach was established to meet the challenges of substantially expanded fitness landscapes (Romero and Arnold 2009). The selection was designed for the mutants with adequate CBP activity, while the subsequent screening was developed for the mutants with improved thermostability. The whole process allowed fast and easy identification of the thermostability-enhanced active mutants. Meanwhile, unlike the traditional directed evolution strategy '*climbing fitness peaks one amino acid at a time*' (Tracewell and Arnold 2009), the mutant libraries were constructed at various mutagenesis frequencies here. Low mutagenesis frequency offered Library L a high probability of functional sequences and a low probability of beneficial mutations, whereas high mutagenesis frequency brings out a high probability of lethal mutations with a high probability of unique sequences to the Library H (Drummond et al. 2005a). Although the unique sequences of these two libraries were hard to assess, such a theory resulted in the lowest

selection power (63.7%) to Library L and the highest selection power (12.1%) to Library H. It is also believed an optimal mutation rate practically exist that balanced diversity and retention of the function (Daugherty et al. 2000; Miura and Sonigo 2001; Zaccolo and Gherardi 1999). Following Sun's and Drummond's model (Drummond et al. 2005a; Sun 1995), we predicted the optimal mutation rate of this work was *ca.* 0.25% (SOM-5 Fig. 4), by which the Library O was developed. Four putative mutants from Library L and one from Library O were identified via selection and screening, whereas none was found from Library H possibly owing to the small number of surviving mutants (~1200) screened. Characterization of the purified enzymes demonstrated the mutants from Library L did not have the properties of interest improved, while the one from Library O (M52) gained an enhanced thermostability but significantly lost the catalytic activity. The false positive mutants screened by the *in vivo* approach did not lead to desired mutants *in vitro*, suggesting big differences between *in vivo* and *in vitro* (Bulter et al. 2003). In particular, the evolved improvement *in vivo* may stem from the increase in expression level, polymerase folding or stability that are specific to the context of the cytoplasm, rather than in the *in vitro* features (Esvelt et al. 2011).

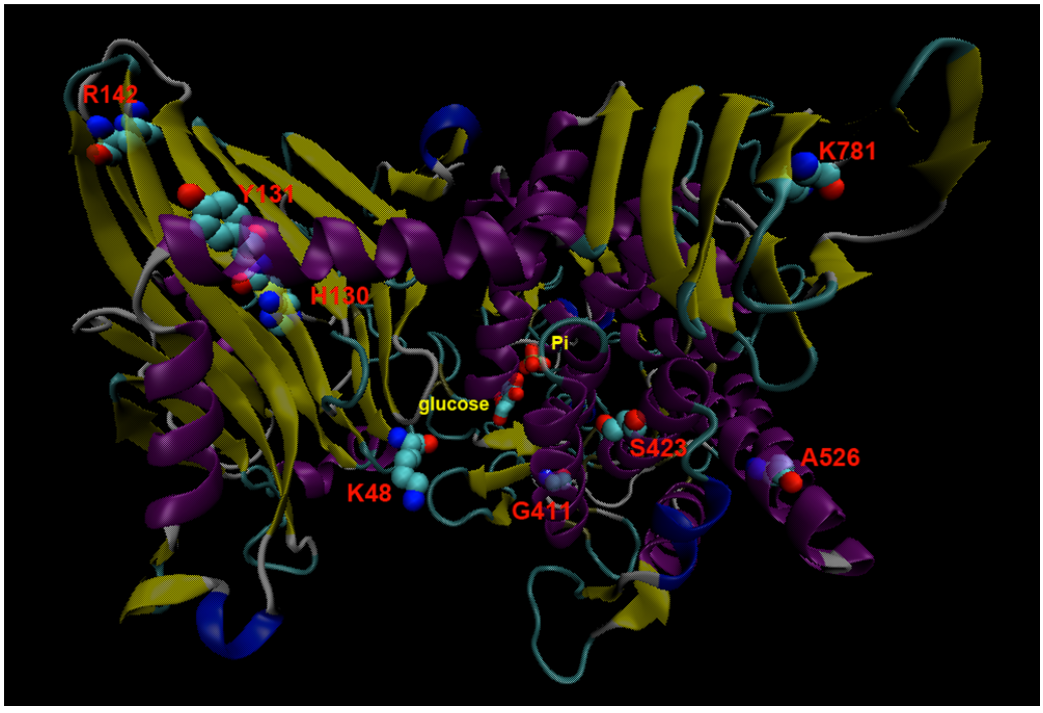


Figure 5.6 Overview of the beneficial mutations identified from rational and random attempts. The ligands glucose and Pi was located according to the crystal structure of *Cg*CBP (PDB: 3AFJ).

The mutant M52 contained five amino acid substitutions, *i.e.* R48K, K142R, R189L, A423S, and V526A. Consensus analysis suggested it was mainly deactivated by R189L (Table 5.3). Reverse mutation L189R then significantly restored the enzyme activity. A423S was a stabilizing mutation, which had been proved by ration design. Replacing Lys 142 by Arg could increase the salt bridges in protein surface and thus strengthened the stability. R48, located in the CBM10 domain and near the entrance of active sites (Fig. 5.6), may affect the formation of enzyme-substrate complex. V526 is in an α-helix, sterically close to S423 (Fig. 5.6). The substitution R48K and V526A individually destabilized *Ct*CBP, whereas the synergistic mutations led to a modest increase in the thermostability. The above results were difficult to explain on structural grounds, but clearly suggested the combination of R48K, V526A, maybe together with A423S had a cooperative effect in stabilizing the *Ct*CBP. Since such kind of mutational epistasis was not implemented by the traditional directed evolution strategies or by current rational attempts (Drummond et al. 2005a; Romero and Arnold 2009), it sets a good example demonstrating the unique advantages of optimal random mutagenesis.

After all, several thermostability-enhanced *Ct*CBP mutants were generated by rational design and directed evolution. The success of each method depended upon the level of understanding of the protein structure and function, or the effectiveness of the selection/screening scheme over large mutant libraries with the proper mutation frequencies. Although either rational design or directed evolution can be effective, a combination of both strategies represents the most successful route to engineer a thermostability-enhanced active *Ct*CBP. The final combination of stabilizing mutations (CM3) identified from rational design and random mutagenesis extended the inactivation halftime at 70 ^{o}C to 24.6 min, three-fold higher than the wild type and more than 2-fold higher than any mutant obtained from the above methods alone. In general, the mutants with enhanced stability displayed smaller k_{cat} and lower enzyme efficiency (k_{cat}/K_m) (Table 5.2), because high activity requires a high flexibility of the protein to undergo conformational changes that cast a concomitant negative impact on thermostability (Bloom and Arnold 2009). Of interest, the mutant CB3 retained the specific activity three-fold higher than the wild type against 80 ^{o}C. The thermostability-enhanced active enzyme would work as a building block for synthetic enzymatic pathway biotransformation (SyPaB) and had the potential for large-scale enzymatic

hydrogen production from cellulosic materials under elevated temperatures (Huang and Zhang 2011; Ye et al. 2009).

5.6 Conclusion

Recent advances in protein engineering expedite the development of robust enzymes tailored for industrial applications (Chen 2001; Schmidt-Dannert and Arnold 1999). However, the knowledge of engineering a large protein (>500 amino acid residues) is still very limited since the space of all possibilities is too large (and expensive) to exhaustively investigate (Zhao et al. 2002). Even though the extremozymes have attracted much attention nowadays (Hough and Danson 1999), it is rarely reported to enhance the stability of thermostable enzymes because of the high risk of failure and the small space for improvement. In this work, we developed two methods to fulfill these challenges. The rational method depended on homology analysis of ten CBP homologs with high sequence identities but from diverse species with distinct living temperatures, by which the signal (stabilizing mutations) was maximized and the noise (ineffective mutations) was reduced. Meanwhile, structural information was facilitated to select the stabilizing mutations as well as to ensure the success rate. As a result, four of six (~67%) were positive mutants. Directed evolution of *Ct*CBP was performed at different mutation rates. In contrast to traditional directed evolution strategies, high frequency mutagenesis, in an adaptive walk (3-25 mutations per *Ctcbp* gene), searched a larger sequence space that was then screened by a high-throughput selection/screening method. One thermostability-enhanced mutant was identified with an apparent manifestation of mutational epistasis. Combining the best mutations from the rational and random attempts generated a *Ct*CBP mutant (CM3) with a three-fold increase in inactivation halftime at 70 ^{o}C. The temperature optimum was raised from 60 to 80 ^{o}C. It was three times more active than the wild type at 80 ^{o}C. These results clearly demonstrated a hybrid approach of rational design and directed evolution enabled great abilities to engineer large proteins and held great potential for creating extremozymes.

5.7 Supplementary Materials

I Primers

SOM-5 Table 1 Primers used in this study

Name	Sequence
For plasmid construction	
P1	5' GAAAGGCTGCAGCATGAAGTTCGGTTTTTTTG 3'
P2	5' CCTTTTGGATCCTTATCCCATAATTACTTCAAC 3'
P3	5' AAACTGCTCGAGGGCTCTTCCATGAAGTTCGGTTTTTTTG 3'
For rational design	
P1_130/131	5' GTGAAGTCCATTATCTTATATTGAAGAATGAAG 3'
P2_130/131	5' CATTGTAGTTTAACGGAACGAAGAAAG 3'
P1_201	5' AAATGCACCGATCAGCGGATTTG 3'
P2_201	5' ACAGAATAGAATGCGTAATGGTTTCTGC 3'
P1_292	5' GATAGAGCAGTTCAAGACTGTTG 3'
P2_292	5' ATTTCATAAGCTTTTTTCTTGTTGATG 3'
P1_411	5' AATGAAATCGGAGGCAACTTCAAC 3'
P2_411	5' GTTACCTTTTTTGGTAAGAGGCTGATAC 3'
P1_423	5' GCTGATTCTTAGCACTGCTGCATATATTAAG 3'
P2_423	5' CACAACGGGTCATCGTTGAAGTTG 3'
P1_781	5' GTATCAAAAGGTGTGAAAAAAATTACTGTTGAC 3'
P2_781	5' ATGGTTCGGATTCTTCACAGTGATTTC 3'
For directed evolution	
P4	5' CGCCCAATACGCAAACC 3'
P5	5' CGCCATTCGCCATTCAGG 3'
P1_48	5' GGATATTGCTTTTACAAGGATGCAAGG 3'
P1_rs_48*	5' GGATATTGCTTTTACAGGGATGCAAGG 3'
P2_48	5' GCCTGCGGTATTCGAAATGAGTGAG 3'
P1_142	5' GGACAGGACAAAAGGAAAATAACTC 3'
P1_rs_142*	5' GGACAGGACAAAAAGAAAATAACTC 3'
P2_142	5' TTCATTCTTCAATATAAGCTTTTGGACTTC 3'
P1_189	5' ACAGAGTACAGAGAGCTCAGAAACC 3'
P1_rs_189*	5' ACAGAGTACAGAGAGCGCAGAAACC 3'
P2_189	5' CTTGTGATAGATAACCGAGCCTTCAATC 3'
P1_526	5' CGGAAAAGACTATGCGAAGCTTTGC 3'
P1_rs_526*	5' CGGAAAAGACTATGTGAAGCTTTGC 3'
P2_526	5' ATGAACACAAACATTCCGGCAATCATAAC 3'

*primers used for reverse mutation

II Stabilization centers

Stabilization center refers the residues that involved in cooperative long-range contacts, and are expected to stabilize proteins structures by preventing their decay with their cooperative long range interactions (Dosztányi et al. 1997). The stabilization centers of *Ct*CBP were determined by SCider (http://www.enzim.hu/scide), in response to its homology modeling (Dosztányi et al. 2003).

SOM-5 Table 2 Stabilization centers of *Ct*CBP

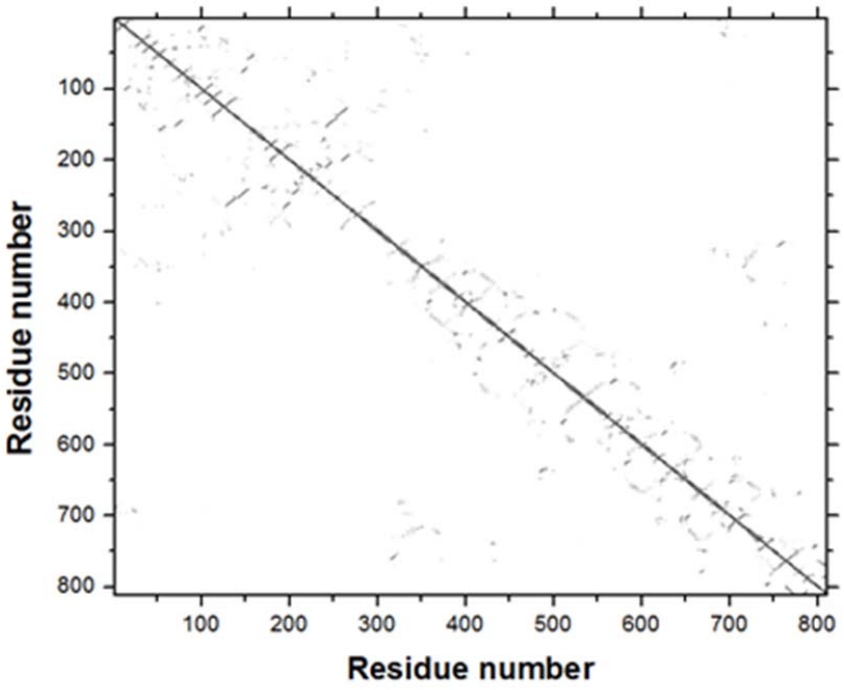
Resid. No.	Resid. Name	List of stabilization Centers	Resid. No.	Resid. Name	List of stabilization Centers	Resid. No.	Resid. Name	List of stabilization Centers
12	GLU	99	202	ILE	245	511	SER	473, 474, 476
13	TYR	98	203	SER	245	514	ILE	474, 475, 476
14	VAL	97, 98	205	PHE	231, 232	517	MET	418
15	ILE	96	206	ASP	231	518	PHE	548
22	TYR	691	207	SER	233	519	VAL	548, 594
26	ASN	36	208	ASP	235	548	MET	518, 519
36	LEU	26	231	ASN	205	551	ALA	472
42	GLY	68, 69, 70	232	ASN	205	561	TRP	575
43	GLY	68, 69	233	SER	206	563	LEU	574
45	CYS	55	234	VAL	207	568	ASP	508
49	ASP	400	235	ALA	207	574	GLY	563
52	LEU	156	239	ALA	207	575	SER	561
54	ARG	155	241	ILE	150	581	GLY	625
55	ILE	45, 453	242	ALA	150	582	LYS	624
56	THR	152, 153	244	HIS	147	583	ILE	624
57	ARG	152	245	SER	146	592	VAL	661
68	GLY	42, 43	253	GLY	135	594	ALA	519
69	GLY	42, 43	254	GLU	135	622	LEU	648
70	ARG	42	256	LYS	133	623	GLN	647
72	PHE	147, 148	257	GLU	132	624	ASN	582, 583
73	TYR	146, 147	258	TYR	131	625	PRO	581
74	ILE	145, 146	259	VAL	130	635	GLY	486
93	GLU	109	261	ILE	196	643	GLY	705
96	GLU	15	264	TYR	193	647	ASN	623
97	CYS	14	265	VAL	193	648	ALA	622
98	ARG	13, 14	315	TYR	758	649	GLY	702
99	HIS	12	336	CYS	371	650	ILE	700, 701, 702
105	LYS	117, 118	337	MET	372	651	PHE	699, 700
109	LYS	93	356	GLY	398	661	ALA	592
112	GLY	136	357	MET	398	668	GLY	748
113	ILE	135, 136, 137	367	LEU	723	691	LYS	22
114	LYS	134, 135	371	HIS	336	697	TYR	715, 716, 718
115	ALA	133	372	GLN	337	698	ALA	714, 715, 716, 720
116	GLU	132	388	GLN	443	699	GLN	651, 713, 714, 715
117	VAL	105	398	TYR	356	700	MET	650, 651, 712
118	THR	105	400	PRO	49	701	VAL	650, 711
119	PHE	129	413	PHE	478	702	ALA	649, 650
129	VAL	119	418	LEU	517	705	ASP	643
130[a]	**GLN**	**259**	420	LEU	460	711	GLU	701
131[a]	**LYS**	**258, 259**	421	ILE	457	712	ALA	700
132	LEU	116, 257	439	GLU	452	713	LYS	699
133	ILE	115, 256	440	GLN	451	714	ASN	698, 699
134	LEU	114	441	VAL	453	715	SER	697, 698, 699
135	LYS	113, 114, 253, 254	443	PHE	388	716	TRP	697, 698
136	ASN	112, 113	444	ASN	388	718	THR	697, 698
137	GLU	113	451	ASP	440	720	THR	698
145	THR	74	452	THR	439	723	TRP	367
146	LEU	73, 74, 245	453	MET	441	740	ASP	796
147	PHE	72, 73, 244	457	LEU	421	741	GLY	794, 795, 796
148	SER	72	460	SER	420	742	LEU	794

150	ILE	241, 242	472	GLY	551	748	ILE	668
151	GLU	241	473	LEU	511	754	GLY	771
152	PHE	56, 57	474	PRO	510	757	VAL	767
153	CYS	55, 56	475	LEU	510	758	THR	315
155	TRP	54	476	ILE	510, 511, 514	767	GLU	757
156	ASN	52	478	ARG	413	771	LYS	754
193	TYR	264, 265	486	ASN	635	794	ILE	741
194	ALA	264	508	VAL	568	795	LEU	741, 742
196	TYR	261	510	GLU	474, 475, 476	796	PRO	740, 741

[a] The resides Q130 and K131 were located in the stabilization centers in pair, both interacting with V259 and/or Y258. Mutation Q130H and K131Y generated strong π-π interactions and expanded the stabilization centers of both resid 130 and 131 to F260. Therefore, they were chosen for site-directed mutagenesis, resulted in a 1.7-fold increase of the inactivation halftime at 70 °C.

III Protein contact map

A protein contact map represents the distance between all possible residue pairs of a 3-D protein structure. The contact map of *Ct*CBP was established by Contact Map plugin of VMD 1.8.6 (Humphrey et al. 1996).



SOM-5 Figure 1 Contact map of *Ct*CBP. A graph square is colored black at 0.0 Angstrom distance, to a linear gray scale between 0.0 and 10.0 Angstroms, and white when equal to or greater than 10.0 Angstroms.

IV Optimal mutation rate for the random mutagenesis of CtCBP

Given a choice of protein scaffold, libraries of fixed sizes, and no reliable knowledge for rational engineering, a simple assessment of library optimality is the number of valuable clones they contain (Daugherty et al. 2000). Therefore, it is hard to tell if we have reasonable mutation rates before qualitative and quantitative evaluation of both low- and high- frequency mutagenesis libraries.

Although the optimal conditions for making mutant libraries are unclear, it is well believed an optimal mutation rate exists that balances diversity and retention of the function (Miura and Sonigo 2001; Zaccolo and Gherardi 1999). In general, low-error-rate mutagenesis results in a high probability of functional sequences with a low probability of beneficial mutations, while high-error-rate mutagenesis brings out a high probability of lethal mutations with a high probability of unique sequences. In the past decades, there have been many attempts to determine the optimum mutation rates for directed evolution theoretically and experimentally (Clune et al. 2008; Kimura 1960; Pritchard et al. 2005; Shafikhani et al. 1997; Zaccolo et al. 1996). Daugherty *et al.* quantitatively analyzed the effect of mutation frequency on the affinity maturation of single chain Fv antibodies (2000). At the low to moderate mutation frequencies with an average mutation rate of $m \leq 8$, the functional fraction of clones decreased exponentially, but the most highly mutated library ($m = 22.5$) had significantly more active clones than expected relative to this trend. The results indicated a preferred mutation frequency for functional improvement of scFv may persist between $m = 4$ and $m = 2$, under which the library includes a larger fraction of active mutants and is more likely to yield improved mutants. Miura and Sonigo proposed a simple model to determine the optimal mutation rate for random mutagenesis (2001). They linked the optimal rate with the number of simultaneous mutations required for possible beneficial and lethal changes. As a result, the model predicted the optimum is a mutation rate that induces at least 63% ($1 - e^{-n}$, *n* represents the required number of positive mutations and $n \geq 1$) of the cloned genes in the library to be non-functional.

Recently, a more inclusive model was presented by Drummond *et al.* (2005a). In their work, experimental data firstly proved that the mutations do not follow the Poisson distribution under the high-error-rate random mutagenesis, but rather a previously proposed distribution derived

from a PCR model (Sun 1995). The PCR-distributed mutations were then modeled to investigate the effect of mutation rates on functionally improved fractions in the random mutagenesis libraries. It was found that the optimal mutation rates depend on the number of tranformants sampled, the PCR protocol used, the wild-type protein being mutated, and other parameters. Considering it's the most realistic model of error-prone PCR by far, we followed Sun's and Drummond's model to estimate the possible optimum in our work as described below.

Assume the *cbp* gene would be amplified by error-prone PCR with a constant efficiency λ. Here λ is set as 0.6 (Drummond et al. 2005a). The thermal cycles are n, resulting in $d = n\lambda$ DNA doublings. Based on Sun's model (Sun 1995), the average of nucleotide mutations per sequence, say $E(m_{nt})$, is defined as

$$E(m_{nt}) = \frac{n\lambda\mu G}{1+\lambda} \tag{1}$$

where μ and G represent mutation rate (mutations per base per PCR cycle) and gene length, respectively. Set

$$x = \mu G \tag{2}$$

which actually presents another type of mutation rate (mutations per PCR cycle).

Then the mutational distribution can be computed as follows,

$$\begin{aligned} P(M = m_{nt}) &= \sum_{k=0}^{n} P(M = m | K = k) P(K = k) \\ &= \sum_{k=0}^{n} \frac{(kx)^{m_{nt}} e^{-kx}}{m_{nt}!} \frac{\binom{n}{k}\lambda^k}{(1+\lambda)^n} = \frac{x^{m_{nt}}}{(1+\lambda)^n m_{nt}!} \sum_{k=n}^{n} \binom{n}{k} \lambda^k k^{m_{nt}} e^{-kx} \end{aligned} \tag{3}$$

The probability a nucleotide mutation produces a non-synonymous change is assumed to be binomial, with parameter p_{ns}. Generally, p_{ns} is around 0.8 (Daugherty et al. 2000). Note that non-synonymous variations include the changes of amino acid encoded as well as the others that truncated or inactivated the protein (e.g. insertions, deletions, and mutations to stop codons). The

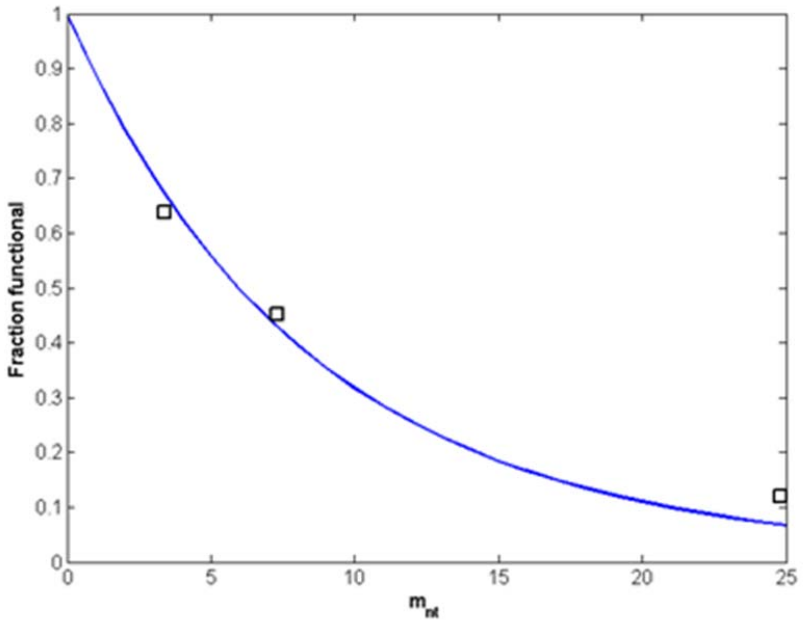
latter constitutes a fraction of mutation with the probability p_{tr}. The p_{tr} is around 0.05-0.07 (Drummond et al. 2005b) and assumed to be 0.06 here. Then the probability a sequence with m_{nt} nucleotide mutations retains function can be defined as

$$P(f \mid m_{nt}) = \sum_{m_{ns}=0}^{m_{nt}} \binom{m_{nt}}{m_{ns}} P(non\ trunc. \mid m_{ns}) \times P(f \mid m_{ns}\ a\,\mathrm{min}o\ acid\ changes)$$
$$= \left(1-\left(1-\nu\left(1-p_{tr}/p_{ns}\right)\right)p_{ns}\right)^{m_{nt}} \tag{4}$$

where m_{ns} denotes non-synonymous mutation, and ν represents the average fraction of functional one-mutant neighors on the protein-sequence-space network. By combining eq.(3) and (4), and assuming gene length $L \to \infty$ since $\langle m_n \rangle << L$, we find the probability a sequence from the library will retain function is:

$$P(f) = \sum_{m_{nt}=0}^{\infty} P(f \mid m_{nt}) P(m_{nt}) = \left(\frac{1+\lambda\left(-\frac{\langle m_{nt} \rangle (1+\lambda)}{n\lambda} \left(1-\nu\left(1-p_{tr}/p_{ns}\right)\right)p_{ns} \right)}{1+\lambda} \right)^{n} \tag{5}$$

Now ν is required to predict the functional fractions. In this work, since the surviving colonies in selection plates represent the functional mutants, the probability *P(f)* would approximate the selection power of each library. Using the $\langle m_{nt} \rangle$ vs. selection power data, we then estimate the random mutagenesis of *Ctcbp* with ν ≈ 0.91 (see SOM-5 Fig. 2). Intriguingly, the ν value is much bigger than previous reports (ν ≈ 0.2 for the antibody binding task, and ν ≈ 0.65 for the subtilisin data), thanks to the larger sequence of *Ctcbp*. It suggests a low probability to inactivate *Ct*CBP by one-mutant-per-step spacewalk, which is helpful to produce functional mutants, but limits the unique fractions.



SOM-5 Figure. 2 Regression analysis of the experimental data (m_{nt} vs. selection power) to estimate ν. The blue line fits the data ('□') in Matlab 2009a (MathWorks, Natick, MA), based on the equation (5).

Next, the probability that a non-truncated sequences with m_{nt} substitutions can be calculated by

$$P(non-truncated|\, M = m_{nt}) = (1 - p_{tr} / p_{ns})^{m_{nt}} \qquad (6)$$

Suppose there are N transformants in the epPCR library. On average,

$$N_m = N \cdot P(M = m_{nt}) P(non-truncated|\, M = m_{nt})$$

are non-truncated proteins with *m* nucleotide mutations.

If with one nucleotide mutation per codon, an average of 5.7 amino acid substitutions (out of a maximum of 19) is accessible due to the conservatism of the genetic code, the *m* nucleotide mutations will conduce to

$$M_m = \binom{L/3}{m} 5.7^m \qquad (7)$$

the average number of unique proteins, where L is the length of the gene in nucleotides.

Give N_m samples, the expected number of unique sequences produced by equal-probably sampling M_m sequences is

$$U_m = M_m - M_m(1-1/M_m)^{N_m} \approx M_m\left(1-e^{-N_m/M_m}\right) \quad (8)$$

Consequently, the total number of unique sequences in a library is the sum over all unique sequences with a specific number of substitutions:

$$U = \sum_{m=0}^{L/3} U_m \quad (9)$$

and the number of unique sequences that retain at least wide-type function is

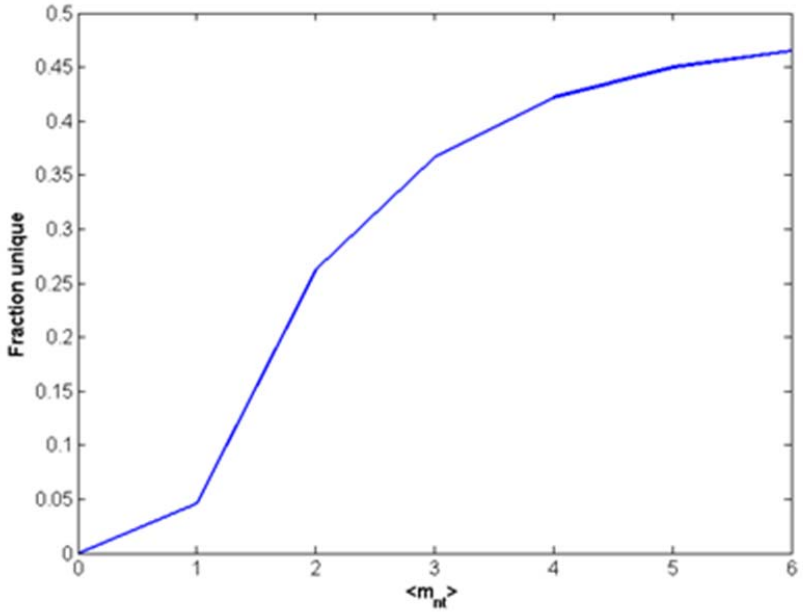
$$U_f = \sum_{m=0}^{L/3} U_m v^m \quad (10)$$

Finally, the fraction of unique sequences (*P(U)*) and unique functional sequences(*P(U$_f$)*) are

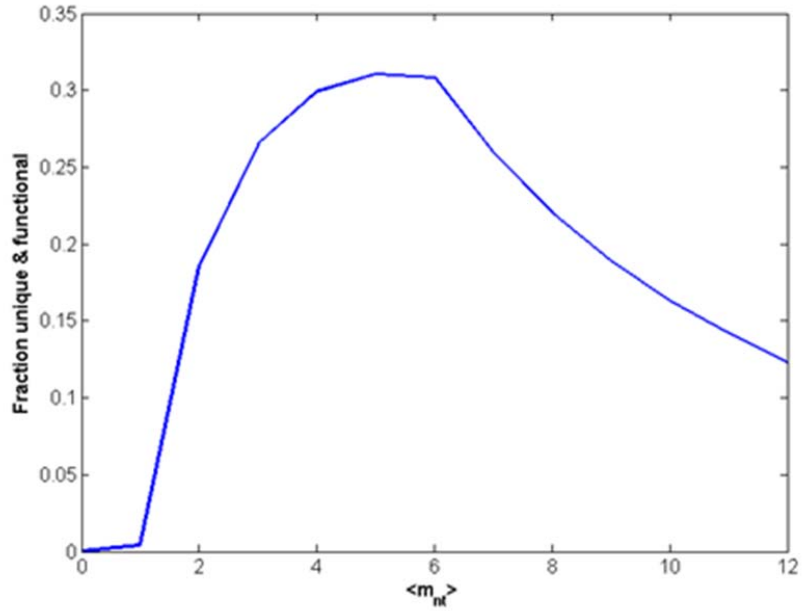
$$P(U) = U/N \quad (11)$$

$$P(U_f) = U_f/N \quad (12)$$

respectively. The relationship between mutation rates and fraction unique sequences is shown in SOM-5 Fig. 3. It is clear the unique sequences are enriched in the high-frequency mutagenesis. However, the maximum fraction of unique sequences is relatively small (*P(U) < 0.5*) even as $\langle m_{nt} \rangle = 6$, because sequences truncated by fameshifts and stop codons accumulate at increasing levels as the mutation rate is increased (Drummond et al. 2005a; Miura and Sonigo 2001).



SOM-5 Fig. 3 The relationship between error rates and fraction of unique sequences. The line is simulated by eq. (11) with n = 25 thermal cycles, efficiency λ = 0.6, and library size N = 10,000 clones.



SOM-5 Figure. 4 The effect of mutagenesis frequencies on fraction of unique and functional sequences. The blue line is estimated by eq. (12) with n = 25 thermal cycles, PCR efficiency λ = 0.6, p_{ns} = 0.8, p_{tr} = 0.07, and library size N = 10,000 clones.

SOM-5 Fig. 4 illustrates the effects of mutation rates on the fraction of unique functional sequences $P(U_f)$. The results suggest optimal mutation rates are both protocol and protein-dependent. Under current circumstances with n = 25 thermal cycle, Taq DNA polymerase (λ = 0.6) and the structural plasticity (v = 0.91, p_{ns} = 0.8, p_{tr} = 0.07), we predict the optimal mutation rate for random mutagenesis of *Ctcbp* is around five to six mutations per gene (*ca.* 2.5 mutations per 1,000 bp, SOM-5 Fig.4), which theoretically yield 4500+ unique mutants and more than 3,000 unique & functional CBP mutants over the 10, 000 clones.

Based on previous work in our lab (Liu et al. 2009), we can achieve such a mutation rate (~ 0.25%) under the condition with regular PCR buffer with 5 mM $MgCl_2$ plus 3 mM $MnCl_2$, 0.2 mM dATP, 0.2 mM dGTP, 1 mM dCTP, and 1 mM dTTP. Practically, it resulted to a mutant library (Library O) with an average mutation rate 0.28%.

Chapter 6

Conclusion and perspectives

6.1 Conclusion

There are numerous advantages to producing cellulosic hydrogen by cell-free synthetic pathway biotransformation (SyPaB), such as high product yield (12 H_2 per glucose unit), nearly 100% selectivity, high-energy conversion efficiency (122%, based on combustion energy), high-purity hydrogen generated, mild reaction conditions, low-cost of reactor, few safety concerns, and nearly no toxicity hazards (Zhang 2010b). The new sugar-to-hydrogen technology has been successfully applied to liberate nearly 100% stoichiometric yield of hydrogen from glucose 6-phosphate and starch previously (Woodward et al. 2000; Zhang et al. 2007). However, glucose 6-phosphate is an unstable and unaffordable substrate for hydrogen production. The limited supply of starch due to competition with food consumption and its increasing prices render starch to hydrogen conversion both expensive and impractical. Alternatively, lignocellulose is the most abundant renewable resource. Cellulose, a polymer of glucose, is the dominant component of lignocellulose. Cellobiose, a disaccharide linked via a $\beta(1\rightarrow4)$ bond, is a dominant product of cellulose hydrolysis. Hydrogen production from these cellulosic materials provides an economic, sustainable, and environmental friendly method to meet increased fuel demand.

The goal of this research is to produce hydrogen in high yields from cellulosic materials and water. It is also intended to increase enzymatic hydrogen production rate to a level comparable to microbe-based methods (~5 mmol H2/L/h) (Das and Veziroglu 2001). The hypothesis is that 1) hydrogen can be produced from cellulosic materials by cell-free SyPaB with the help of β-glucan phosphorylase(s), and that 2) hydrogen production rate can be improved by accelerating the key reactions identified by kinetic analysis, by increasing substrate loading, and by elevating reaction temperature. To test the above hypothesis, four main projects have been completed:

1) **Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails**

Biohydrogen was produced in high yields from cellulosic materials and water in a one-pot process catalyzed by 14 enzymes and one coenzyme. Different from the previous systems (Zhang et al. 2007), the new pathway replaced starch phosphorylase with *Clostridium thermocellum* cellobiose phosphorylase and included a new enzyme *Clostridium thermocellum*

cellodextrin phophorylase. The assembly of these new enzymes catalyzed an overall sugar-to-hydrogen reaction as:

$$G_n\,(aq) + 7\,(n\text{-}1)\,H_2O\,(l) \rightarrow C_6H_{12}O_6\,(aq) + 12\,(n\text{-}1)\,H_2\,(g) + 6\,(n\text{-}1)\,CO_2\,(g)$$

Thermodynamic analysis indicated it is a spontaneous ($\Delta G < 0$), endothermic ($\Delta H >> 0$), and entropy-gained ($\Delta S >> 0$) process under the modest conditions of < 100 ^{o}C and ~1 atm.

A proof-of-principle experiment was first conducted to validate if hydrogen could be produced from cellobiose by SyPaB. As results, the final H_2 and CO_2 yields reached 11.2 moles of H_2 and 5.64 moles of CO_2 per mole of anhydroglucose unit of cellobiose, which were 93.1% and 94% of the theoretical yields, respectively. It was comparable to the yields reported previously (11.6 moles of H_2 per mole glucose-6-phosphate) (Woodward et al. 2000). Next, a mathematical model was established to simulate *in vitro* hydrogen production from glucose 6-phosphate, starch, and cellobiose, as well as to identify the rate-limited step(s). Model results inferred the possibilities to accelerate hydrogen production rates by increasing rate-limiting hydrogenase concentration, increasing the substrate concentration, and elevating the reaction temperature. In such wise, the maximal hydrogen production rate from cellopentaose was enhanced to 3.92 mmole/L/h, nearly 20–fold higher than before (Woodward et al. 2000; Zhang et al. 2007). So far the production rate of hydrogen is higher than those of photobiological systems and comparable to the rates reported in dark fermentations (Das and Veziroglu 2001).

2) Thermophilic *α*-glucan phosphorylase from *Clostridium thermocellum*: Cloning, characterization and enhanced thermostability

A putative *α*-glucan phosphorylase (αGP) was cloned from the thermophilic bacterium *Clostridium thermocellum* ATCC 27405, heterologously expressed in *Escherichia coli*, purified with and without a His tag, and biochemically characterized in detail. The enzyme had broad substrate specificities for different chain length dextrins and soluble starch. The optimal conditions were pH 6.0–6.5 and 60 °C for the synthesis direction and pH 7.0–7.5 and 80 °C for the degradation direction. The thermal inactivation of this enzyme strongly depended on temperature, protein concentration, and certain additives that were shown previously to benefit the protein thermostability. The half lifetime of 0.05 mg *α*GP/mL at 50 °C was extended by 45-

fold to 90 h through a combined addition of 0.1 mM Mg^{2+}, 5 mM DTT, 1% NaCl, 0.1% Triton X-100, and 1 mg/mL BSA. The discovered conditions are suitable to prolong the stability of enzymes for both current and future applications of SyPaB.

3) Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose

This project aimed to construct a non-natural cellulose phosphorylase in order to produce low-cost hydrogen from abundant but insoluble cellulose. *Clostridium thermocellum* cellodextrin phosphorylase (*Ct*CDP), a single-module protein without an apparent carbohydrate-binding module, has reported activities on soluble cellodextrin with a degree of polymerization (DP) from two to five (Alexander 1961; Krishnareddy et al. 2002). In this study, *Ct*CDP was first discovered to have weak activities on weakly water-soluble celloheptaose and insoluble regenerated amorphous cellulose (RAC). To enhance its activity on solid cellulosic materials, four cellulose binding modules, e.g. CBM3 (type A) from *C. thermocellum* CbhA, CBM4-2 (type B) from *Rhodothermus marinus* Xyn10A, CBM6 (type B) from *Cellvibrio mixtus* Cel5B, and CBM9-2 (type C) from *Thermotoga maritima* Xyn10A, were fused to the C terminus of *Ct*CDP. The results demonstrated that addition of the CBM 9-2 significantly improved the catalytic activity of *Ct*CDP on both short-chain and long-chain RAC. The chimeric protein will be the first step to construct a highly active cellulose phosphorylase for the *in vitro* hydrogen production by SyPaB.

4) Engineering a thermostability-enhanced active *Clostridium thermocellum* cellobiose phosphorylase by a combination of rational design and directed evolution

The *Clostridium thermocellum* cellobiose phosphorylase (*Ct*CBP) is relatively stable at 50 ^{o}C, but rapidly inactivated at 70 ^{o}C. In an effort to produce a *Ct*CBP mutant with improved thermostability suitable for the large-scale *in vitro* hydrogen production, two approaches were developed, i.e. site-directed mutagenesis based on structure-guided homology analysis and random mutagenesis with various mutation rates. The former was a data-driven protein design method that chose substitutions by comparison of the protein sequences of CBP homologs that have different stabilities, and utilized structural information to limit the number of variants created. The latter is a modified directed evolution method that created mutant libraries with different mutagenesis frequencies. A novel combinational selection/screening strategy was also

developed and employed to rapidly isolate the thermostability-enhanced active protein variants. Several stabilizing mutations were identified by both methods. The final combination of stabilizing mutations (CM3) identified from rational design and random mutagenesis lengthened the inactivation halftime at 70 ^{o}C to 24.6 min, 3-fold higher than the wild type and more than 2-fold higher than any mutant obtained from the above methods alone. The results demonstrated that the coupling of these rational and random approaches formed a solid basis for engineering large proteins, like *Ct*CBP. The thermostability-enhanced active *Ct*CBP mutant would work as a building block for cell-free SyPaB and had the potential for large-scale *in vitro* cellulosic hydrogen production.

6.2 Perspectives

The results of the research have informed the development of a novel high-efficiency and high-yield enzymatic process for the conversion of cellulosic materials to hydrogen. Although not an explicit aim in the thesis, the logic long-term goal of this work is to improve cellulose-to-hydrogen efficiency by SyPaB so as to make it more appealing for transportation application (Zhang 2009). However, there are many challenges ahead in order to fulfill this goal.

Practical applications of cellulose-to-hydrogen systems are tightly correlated with hydrogen production rates. The current hydrogen production rate of SyPaB is 3.92 mmol per liter of reactor per hour, where the substrate concentration is 8 mM (~1.4 g/L) and the operating temperature is 32 ^{o}C. The rate could be accelerated by increasing reaction temperatures from 30 to 80 ^{o}C, increasing the use of enzymes responsible for rate-limited reactions, increasing substrate concentrations by 50-fold or higher, increasing overall enzyme concentrations by 10-fold or more, accelerating the reaction rates by metabolite channeling, and increasing the catalytic efficiency of enzymes (Zhang 2010b). By a conservative estimation, the above strategies would result in a 3000-fold increase of hydrogen production rate to 23.6 g H_2/L/h. It would be sufficient for applications such as sugar battery and sugar fuel cell vehicles (namely sugar car) (Zhang 2010b).

Now the production of low-cost biofules is still too costly by SyPaB. It is mainly attributed to costly unstable enzymes and labile cofactors, for example NAD(P)H. Enzyme costs can be

decreased by use of (hyper)-thermostable enzymes, enzyme immobilization, simple enzyme purification, large-scale production of recombinant protein, and so on. A number of (hyper-)thermostable enzymes have been cloned and expressed in *E. coli*, e.g. three phosphorylases studied here, *C. thermocellum* PGM (Wang and Zhang 2010), *T. martima* 6PGDH (Wang and Zhang 2009b) and FBP (Myung et al. 2010). Both rational and random approaches have been applied to construct highly active and highly stable enzymes. A combination of enzyme immobilization and thermostable enzymes can further increase the lifetime of enzymes. An impressive example is immobilized thermostable glucose isomerase that was the capablity to be active at 55 ^{o}C for more than two years (Vasic-Racki 2006). In order to decrease enzyme purification costs, a few economical and scalable protein purification methods have been developed, simply relying on adsorption/desorption (Hong et al. 2008a; Hong et al. 2008b) or heat precipitation for thermostable enzymes (Zhang and Mielenz 2011). With respect to costly labile cofactors, NAD(P)H has been efficiently recycled via immobilization on polymers and regenerated in high yields through SyPaB (Wang et al. 2011). It is also expected that high-cost of cofactors will be solved by using less costly biomimetic NAD analogues (Lo and Fish 2002). Therefore, it is achievable to decrease enzyme costs to tens of dollars per kg per dry protein weight and to increase the total turn-over number (TTN) of NAD to ~ 1,000,000. When sugar (carbohydrates) prices are \$0.18/kg, the above achievement would possibly decrease the hydrogen production cost as low as \$1.50 per kg (Zhang 2010b). In this case, hydrogen production from carbohydrates by SyPaB would be very economically viable in the future.

References

Adams MWW, Stiefel EI. 1998. Biological hydrogen production: Not so elementary. Science 282(5395):1842-1843.

Ahn JO, Choi ES, Lee HW, Hwang SH, Kim CS, Jang HW, Haam SJ, Jung JK. 2004. Enhanced secretion of *Bacillus stearothermophilus* L1 lipase in *Saccharomyces cerevisiae* by translational fusion to cellulose-binding domain. Appl. Microbiol. Biotechnol. 64:833-839.

Akanuma S, Yamagishi A, Oshima T, Tanaka N. 1998. Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from *Bacillus subtilis* by experimental evolution. Protein Sci. 7(3):698-705.

Akanuma S, Yamagishi A, Tanaka N, Oshima T. 1999. Further improvement of the thermal stability of a partially stabilized *Bacillus subtilis* 3-isopropylmalate dehydrogenase variant by random and site-directed mutagenesis. Eur. J. Biochem. 260(2):499-504.

Akkerman I, Janssen M, Rocha J, Wijffels RH. 2002. Photobiological hydrogen production: photochemical efficiency and bioreactor design. Int. J. Hydrogen Energy 27(11-12):1195-1208.

Alexander JK. 1961. Characteristics of cellobiose phosphorylase. J. Bacteriol. 81:903-910.

Alexander JK. 1986. Purification and specificity of cellobiose phosphorylase from *Clostridium thermocellum*. J. Biol. Chem. 243:2899-2904.

Altschul SF, Madden TL, Schaffer 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.

Arai M, Tanaka K, Kawaguchi T. 1994. Purification and properties of cellodextrin phosphorylase from *Clostridium thermocellum*. J. Ferment. Bioeng. 77(3):239-242.

Arnold FH, Georgiou G, editors. 2003. Directed Evolution Library Creation: Methods and Protocols Humana Press.

Asada Y, Miyake J. 1999. Photobiological hydrogen production. J. Biosci. Bioeng. 88(1):1-6.

Barbir F. 2005. PEM electrolysis for production of hydrogen from renewable energy sources. Solar Energy 78(5):661-669.

Benemann J. 1996. Hydrogen biotechnology: Progress and prospects. Nat. Biotech. 14:1101-1103.

Benemann JR. 1998. Process analysis and economics of biophotolysis of water. CA, USA: DOE Scientific and technical information.

Berg IA, Kockelkorn D, Buckel W, Fuchs G. 2007. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. Science 318(5857):1782-1786.

Bhosale S, Rao M, Deshpande V. 1996. Molecular and industrial aspects of glucose isomerase. Microbiol. Rev. 60(2):280-300.

Bhuiyan SH, Rus'd AA, Kitaoka M, Hayashi K. 2003. Characterization of a hyperthermostable glycogen phosphorylase from *Aquifex aeolicus* expressed in *Escherichia coli*. J. Mol. Catal. B: Enzym. 22:173-180.

Bibel M, Brettl C, Gosslar U, Kriegshäuser G, Liebl W. 1998. Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. FEMS Microbiol. Lett. 158:9-15.

Blake AW, McCartney L, Flint JE, Bolam DN, Boraston AB, Gilbert HJ, Knox JP. 2006. Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. J. Biol. Chem. 281:29321-29329.

Bloom JD, Arnold FH. 2009. In the light of directed evolution: Pathways of adaptive protein evolution. Proc. Nat. Acad. Sci. USA 106(Supplement 1):9995-10000.

Bockris JO. 2002. The origin of ideas on a Hydrogen Economy and its solution to the decay of the environment. Int. J. Hydrogen Energy 27(7-8):731-740.

Boeck B, Schinzel R. 1996. Purification and characterisation of an α–glucan phosphorylase from the thermophilic bacterium *Thermus thermophilus*. Eur. J. Biochem. 239:150-155.

Boerio-Goates J. 1991. Heat-capacity measurements and thermodynamic functions of crystalline α-D-glucose at temperatures from 10 K to 340 K. J. Chem. Thermodynamics 23:403-409.

Bolton JR. 1996. Solar photoproduction of hydrogen: A review. Solar Energy 57(1):37-50.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. Biochem. J. 382(3):769-781.

Boraston AB, Creagh AL, Alam MM, Kormos JM, Tomme P, Haynes CA, Warren RAJ, Kilburn DG. 2001. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry 40(21):6240-6247.

Bornscheuer UT, Pohl M. 2001. Improved biocatalysts by directed evolution and rational protein design. Curr. Opin. Chem. Biol. 5(2):137-143.

Böttcher D, Bornscheuer UT. 2010. Protein engineering of microbial enzymes. Curr. Opin. Microbiol. 13(3):274-282.

Bourne Y, Henrissat B. 2001. Glycoside hydrolases and glycosyltransferases: Families and functional modules. Curr. Opin. Struct. Biol. 11(5):593-600.

Boyer ME, Stapleton JA, Wang C-w, Swartz JR. 2005. A program for the directed evolution of oxygen-tolerant hydrogenase enzymes for use in photobiological hydrogen generation. AICHE annual meeting, 2005. Cincinnati, OH.

Bryant FO, Adams MW. 1989. Characterization of hydrogenase from the hyperthermophilic archaebacterium *Pyrococcus furiosus* J. Biol. Chem. 264:5070-5079.

Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C, Arnold FH. 2003. Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. Appl. Environ. Microbiol. 69(2):987-995.

Burgdorf T, De Lacey AL, Friedrich B. 2002. Functional analysis by site-directed mutagenesis of the NAD^+-reducing hydrogenase from *Ralstonia eutropha*. J. Bacteriol. 184(22):6280-6288.

Cannon WR, MacDonald G. 2009. The effect of high concentration salt on the structure, stability, and aggregation of RecA. Biophys. J. 96:81-82.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucl. Acids Res. 37:D233-238.

Carrard G, Koivula A, Söderlund H, Béguin P. 2000. Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. Proc. Nat. Acad. Sci. USA 97:10342-10347.

Champine JE, Uffen RL. 1987. Membrane topography of anaerobic carbon monoxide oxidation in *Rhodocyclus gelatinosus*. J. Bacteriol. 169(10):4784-4789.

Chase M. 1998. NIST-JANAF thermochemical tables. In J. Phys. Chem. Ref. Data, Monograph 9. p 1-1951.

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol. Bioeng. 79(1):53-73.
Chen G, Andries J, Luo Z, Spliethoff H. 2003. Biomass pyrolysis/gasification for product gas production: the overall investigation of parametric effects. Energy Convers. Manage. 44(11):1875-1884.
Chen G, Segel I. 1968. Purification and properties of glycogen phosphorylase from *Escherichia coli*. Arch. Biochem. Biophys. 127:175-186.
Chen R. 2001. Enzyme engineering: rational redesign versus directed evolution. Trends Biotechnol. 19(1):13-14.
Cheng S, Logan BE. 2007a. Ammonia treatment of carbon cloth anodes to enhance power generation of microbial fuel cells. Electrochem. Commun. 9(3):492-496.
Cheng S, Logan BE. 2007b. Sustainable and efficient biohydrogen production via electrohydrogenesis. Proc. Natl. Acad. Sci. USA 104(47):18871-18873.
Cherry JR, Fidantsef AL. 2003. Directed evolution of industrial enzymes: an update. Curr. Opin. Biotechnol. 14(4):438-443.
Cherry JR, Lamsa MH, Schneider P, Vind J, Svendsen A, Jones A, Pedersen AH. 1999. Directed evolution of a fungal peroxidase. Nat. Biotech. 17(4):379-384.
Chheda Juben N, Huber George W, Dumesic James A. 2007. Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. Angew. Chem. Int. Ed. 46(38):7164-7183.
Cho A. 2004. Renewable energy - Hydrogen from ethanol goes portable. Science 303:942-943.
Claassen PAM, van Lier JB, Lopez Contreras AM, van Niel EWJ, Sijtsma L, Stams AJM, de Vries SS, Weusthuis RA. 1999. Utilisation of biomass for the supply of energy carriers. Appl. Microbiol. Biotechnol. 52(6):741-755.
Clune J, Misevic D, Ofria C, Lenski RE, Elena SF, Sanjuán R. 2008. Natural selection fails to optimize mutation rates for long-term adaptation on rugged fitness landscapes. PLoS Comput. Biol. 4(9):e1000187.
Conrado RJ, Varner JD, DeLisa MP. 2008. Engineering the spatial organization of metabolic enzymes: Mimicking nature's synergy. Curr. Opin. Biotechnol. 19(5):492-499.
Cooney MJ, Svoboda V, Lau C, Martin G, Minteer SD. 2008. Enzyme catalysed biofuel cells. Energy Environ. Sci. 1:320-337.
Corella J, Aznar M-P, Gil J, Caballero MA. 1999. Biomass gasification in fluidized bed: Where to locate the dolomite to improve gasification? Energ. Fuel. 13(6):1122-1127.
Corella J, Toledo JM, Padilla R. 2004. Olivine or dolomite as in-bed additive in biomass gasification with air in a fluidized bed: Which is better? Energ. Fuel. 18(3):713-720.
Cortright RD, Davda RR, Dumesic JA. 2002. Hydrogen from catalytic reforming of biomass-derived hydrocarbons in liquid water. Nature 418:964-967.
Cournac L, Guedeney G, Peltier G, Vignais PM. 2004. Sustained photoevolution of molecular hydrogen in a mutant of *Synechocystis sp*. strain PCC 6803 deficient in the type I NADPH-dehydrogenase complex. J. Bacteriol. 186(6):1737-1746.
Cummer KR, Brown RC. 2002. Ancillary equipment for biomass gasification. Biomass Bioenerg. 23(2):113-128.
Dabrock B, Bahl H, Gottschalk G. 1992. Parameters affecting solvent production by *Clostridium pasteurianum*. Appl. Environ. Microbiol. 58(4):1233-1239.
Das D, Veziroglu TN. 2001. Hydrogen production by biological processes: a survey of literature. Int. J. Hydrogen Energy 26(1):13-28.

Daugherty PS, Chen G, Iverson BL, Georgiou G. 2000. Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. Proc. Nat. Acad. Sci. USA 97(5):2029-2034.

Davda RR, Shabaker JW, Huber GW, Cortright RD, Dumesic JA. 2005. A review of catalytic issues and process conditions for renewable hydrogen and alkanes by aqueous-phase reforming of oxygenated hydrocarbons over supported metal catalysts. Appl. Catal., B-Environ 56(1-2):171-186.

De Groeve MRM, De Baere M, Hoflack L, Desmet T, Vandamme EJ, Soetaert W. 2009. Creating lactose phosphorylase enzymes by directed evolution of cellobiose phosphorylase. Protein Eng. Des. Sel. 22(7):393-399.

Deluga GA, Salge JR, Schmidt LD, Verykios XE. 2004. Renewable hydrogen from ethanol by autothermal reforming. Science 303:993-997.

Demirbas A. 2002. Hydrogen production from biomass by the gasification process. Energ. Source. 24(1):59-68.

DOE. 2003. Hydrogen production and delivery research. . Washington, DC. Report nr Solicitation DE-PS36-03GO93007.

Donahue JL, Bownas JL, Niehaus WG, Larson TJ. 2000. Purification and characterization of glpX-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*. J. Bacteriol. 182(19):5624-5627.

Dosztányi Z, Fiser A, Simon I. 1997. Stabilization centers in proteins:Identification, characterization and predictions. J. Mol. Biol. 272(4):597-612.

Dosztányi Z, Magyar C, Tusnády G, Simon I. 2003. SCide: Identification of stabilization centers in proteins. Bioinformatics 19(7):899-900.

Douglas T, Ball A, Torgesen J. 1951. Heat capacity of crystalline dextrose between 25 to 95 ^{o}C. J. Am. Chem. Soc. 73:1360-1361.

Drueckes P, Boeck B, Palm D, Schinzel R. 1996. Mutational analysis of the oligosaccharide recognition site at the active site of *Escherichia coli* maltodextrin phosphorylase. Biochemistry 35:6727-6734.

Drummond DA, Iverson BL, Georgiou G, Arnold FH. 2005a. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. J. Mol. Biol. 350(4):806-816.

Drummond DA, Silberg JJ, Meyer MM, Wilke CO, Arnold FH. 2005b. On the conservative nature of intragenic recombination. Proc. Nat. Acad. Sci. USA 102(15):5380-5385.

Dunn S. 2002. Hydrogen futures: Toward a sustainable energy system. Int. J. Hydrogen Energy 27(3):235-264.

El-Shishtawy RMA, Kawasaki S, Morimoto M. 1997. Biological H_2 production using a novel light-induced and diffused photoreactor. Biotechnol. Tech. 11(6):403-407.

Elliott D, Werpy T, Wang Y, Frye JJ; 2001. Ruthenium on rutile catalyst, catalytic system, and method for aqueous phase hydrogenations patent US Patent #6,235,797.

Elliott DC, Sealock LJ. 1983. Aqueous catalyst systems for the water-gas shift reaction. 1. Comparative catalyst studies. Ind. Eng. Chem. Prod. Res. Dev. 22(3):426-431.

Elliott DC, Sealock LJ, Baker EG. 1993. Chemical processing in high-pressure aqueous environments. 2. Development of catalysts for gasification. Ind. Eng. Chem. Res. 32(8):1542-1548.

Eroglu E, Eroglu I, Gündüz U, Türker L, Yücel M. 2006. Biological hydrogen production from olive mill wastewater with two-stage processes. Int. J. Hydrogen Energy 31(11):1527-1535.

Eroglu I, Aslan K, Gündüz U, Yücel M, Türker L. 1999. Substrate consumption rates for hydrogen production by *Rhodobacter sphaeroides* in a column photobioreactor. J. Biotechnol. 70(1-3):103-113.

Eronina TB, Chebotareva NA, Bazhina SG, Makeeva VF, Kleymenov SY, Kurganov BI. 2009. Effect of proline on thermal inactivation, denaturation and aggregation of glycogen phosphorylase b from rabbit skeletal muscle. Biophys. Chem. 141:66-74.

Esvelt KM, Carlson JC, Liu DR. 2011. A system for the continuous directed evolution of biomolecules. Nature 472(7344):499-503.

Fang HHP, Liu H, Zhang T. 2005. Phototrophic hydrogen production from acetate and butyrate in wastewater. Int. J. Hydrogen Energy 30(7):785-793.

Fascetti E, D'Addario E, Todini O, Robertiello A. 1998. Photosynthetic hydrogen evolution with volatile organic acids derived from the fermentation of source selected municipal solid wastes. Int. J. Hydrogen Energy 23(9):753-760.

Fascetti E, Todini O. 1995. *Rhodobacter sphaeroides* RV cultivation and hydrogen production in a one- and two-stage chemostat. Appl. Microbiol. Biotechnol. 44:300-305.

Fedorov AS, Tsygankov AA, Rao KK, Hall DO. 1998. Hydrogen photoproduction by *Rhodobacter sphaeroides* immobilised on polyurethane foam. Biotechnol. Lett. 20(11):1007-1009.

Fletterick RJ, Madsen NB. 1980. The structures and related functions of phosphorylase a. Annu. Rev. Biochem. 49:31-61.

Forster AC, Church GM. 2007. Synthetic biology projects *in vitro*. Genome Res. 17(1):1-6.

Frare E, de Laureto PP, Scaramella E, Tonello F, Marin O, Deana R, Fontana A. 2005. Chemical synthesis of the RGD-protein decorsin: Pro → Ala replacement reduces protein thermostability. Protein Eng. Des. Sel. 18(10):487-495.

Friedrich W, Dietmar L, Kurt W. 1977. 1,4-alpha-Glucan phosphorylase from *Klebsiella pneumoniae* covalently coupled on porous glass. Biotechnol. Bioeng. 19:1387-1403.

Fukui T, Shimomura S, Nakano K. 1982. Potato and rabbit muscle phosphorylases: Comparative studies on the structure, function and regulation of regulatory and nonregulatory enzymes. Mol. Cell. Biochem. 42:129-144.

Ginnings D, Douglas T, Ball A. 1950. Heat capacity of sodium between 0 ^{o}C and 900 ^{o}C. J. Res. Nat. Bureau Standards 45:23.

Giver L, Gershenson A, Freskgard P-O, Arnold FH. 1998. Directed evolution of a thermostable esterase. Proc. Nat. Acad. Sci. USA 95(22):12809-12813.

Goldstein RA. 2007. Amino-acid interactions in psychrophiles, mesophiles, thermophiles, and hyperthermophiles: Insights from the quasi-chemical approximation. Protein Sci. 16(9):1887-1895.

Gomez E, Rani DA, Cheeseman CR, Deegan D, Wise M, Boccaccini AR. 2009. Thermal plasma technology for the treatment of wastes: A critical review. J. Hazard. Mater. 161(2):614-626.

Greenberg MM, Yen TTT, Bobbitt JL. 1972. Identification and characterization of pituitary triose-phosphate isomerase. Eur. J. Biochem. 24(3):416-420.

Griessler R, D'Auria S, Schinzel R, Tanfani F, Nidetzky B. 2000. Mechanism of thermal denaturation of maltodextrin phosphorylase from *Escherichia coli*. Biochem. J 346:255-263.

Hachem MA, Nordberg Karlsson E, Bartonek-Roxâ E, Raghothama S, Simpson PJ, Gilbert HJ, Williamson MP, Holst O. 2000. Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochem. J. 345(1):53-60.

Hallenbeck PC, Benemann JR. 2002. Biological hydrogen production: Fundamentals and limiting processes. Int. J. Hydrogen Energy 27(11-12):1185-1193.

Hallgren AL. 1995. Hot gas cleanup in biomass gasification-review of activities within EC sponsored R&D programs. ACS fuels: symposium on direct coal liquefaction. Chicago.

Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus species*. Proc. Nat. Acad. Sci. USA 96(7):3578-3583.

Hansel A, Lindblad P. 1998. Towards optimization of cyanobacteria as biotechnologically relevant producers of molecular hydrogen, a clean and renewable energy source Appl. Microbiol. Biotechnol. 50(2):153-160.

Happe T, Schutz K, Bohme H. 2000. Transcriptional and mutational analysis of the uptake hydrogenase of the filamentous cyanobacterium *Anabaena variabilis* ATCC 29413. J. Bacteriol. 182(6):1624-1631.

Haryanto A, Fernando S, Murali N, Adhikari S. 2005. Current status of hydrogen production techniques by steam reforming of ethanol: A review. Energ. Fuel. 19(5):2098-2106.

Hawkes FR, Dinsdale R, Hawkes DL, Hussy I. 2002. Sustainable fermentative hydrogen production: Challenges for process optimisation. Int. J. Hydrogen Energy 27(11-12):1339-1347.

He D, Bultel Y, Magnin J-P, Roux C, Willison JC. 2005. Hydrogen photosynthesis by *Rhodobacter capsulatus* and its coupling to a PEM fuel cell. J. Power Sources 141(1):19-23.

Henshaw JL, Bolam DN, Pires VMR, Czjzek M, Henrissat B, Ferreira LMA, Fontes CMGA, Gilbert HJ. 2004. The family 6 carbohydrate binding module *Cm*CBM6-2 contains two ligand-binding sites with distinct specificities. J. Biol. Chem. 279(20):21552-21559.

Hermann WA. 2006. Quantifying global exergy resources. Energy 31(12):1685-1702.

Hidaka M, Honda Y, Kitaoka M, Nirasawa S, Hayashi K, Wakagi T, Shoun H, Fushinobu S. 2004. Chitobiose phosphorylase from *Vibrio proteolyticus*, a member of glycosyl transferase family 36, has a clan GH-L-like $(\alpha/\alpha)_6$ barrel fold. Structure 12(6):937-947.

Hidaka M, Kitaoka M, Hayashi K, Wakagi T, Shoun H, Fushinobu S. 2006. Structural dissection of the reaction mechanism of cellobiose phosphorylase. Biochem. J 398(1):37-43.

Hiraishi M, Igarashi K, Kimura S, Wada M, Kitaoka M, Samejima M. 2009. Synthesis of highly ordered cellulose II in vitro using cellodextrin phosphorylase. Carbohydr. Res. 344(18):2468-2473.

Holladay JD, Hu J, King DL, Wang Y. 2009. An overview of hydrogen production technologies. Catal. Today 139(4):244-260.

Hong J, Wang Y, Ye X, Zhang YHP. 2008a. Simple protein purification through affinity adsorption on regenerated amorphous cellulose followed by intein self-cleavage. J. Chromatogr. A 1194(2):150 - 154.

Hong J, Ye X, Wang Y, Zhang YHP. 2008b. Bioseparation of recombinant cellulose-binding module-proteins by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. Anal. Chim. Acta. 621(2):193-199.

Hong J, Ye X, Zhang YHP. 2007. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. Langmuir 23(25):12535-12540.

Hong Y, Wu L, Liu B, Peng C, Sheng D, Ni J, Shen Y. 2008c. Characterization of a glucan phosphorylase from the thermophilic archaeon *Sulfolobus tokodaii* strain 7. J. Mol. Catal. B: Enzym. 54:27-34.

Horne RN, Anderson WB, Nordlie RC. 1970. Glucose dehydrogenase activity of yeast glucose 6-phosphate dehydrogenase: Inhibition by adenosine 5'-triphosphate and other nucleoside 5'-triphosphates and diphosphates. Biochemistry 9(3):610-616.

Hough DW, Danson MJ. 1999. Extremozymes. Curr. Opin. Chem. Biol. 3(1):39-46.

Huang H, Tang L. 2007. Treatment of organic waste using thermal plasma pyrolysis technology. Energy Convers. Manage. 48(4):1331-1337.

Huang WD, Zhang YHP. 2011. Analysis of biofuels production from sugar based on three criteria: Thermodynamics, bioenergetics, and product separation. Energy Environ. Sci. 4:784-782.

Huber GW, Chheda JN, Barrett CJ, Dumesic JA. 2005. Production of liquid alkanes by aqueous-phase processing of biomass-derived carbohydrates. Science 308(5727):1446-1450.

Huber GW, Iborra S, Corma A. 2006. Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. Chem. Rev. 106(9):4044-4098.

Huber GW, Shabaker JW, Dumesic JA. 2003. Raney Ni-Sn catalyst for H_2 production from biomass-derived hydrocarbons. Science 300(5628):2075-2077.

Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. J. Mol. Graphics 14(1):33-38.

Jacobson MZ, Colella WG, Golden DM. 2005. Cleaning the air and improving health with hydrogen fuel-cell vehicles. Science 308:1901-1905.

Jaenicke R, Schurig H, Beaucamp N, Ostendorp R. 1996. Structure and stability of hyperstable proteins: Glycolytic enzymes from hyperthermophilic bacterium *Thermotoga maritima*. In: Frederic M. Richards DSE, Peter SK, editors. Advances in Protein Chemistry: Academic Press. p 181-269.

Janecek S. 1993. Strategies for obtaining stable enzymes. Process Biochem. 28:435-445.

Jenney FE, Adams MW. 2008. Hydrogenases of the model hyperthermophiles. Ann. N. Y. Acad. Sci. 1125:252-266.

Ji P, Feng W, Chen B. 2009. Production of ultrapure hydrogen from biomass gasification with air. Chem. Eng. Sci. 64(3):582-592.

Jiang X, Kowalski J, Kelly JW. 2001. Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. Protein Sci. 10(7):1454-65.

Johnson PE, Joshi MD, Tomme P, Kilburn DG, McIntosh LP. 1996. Structure of the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC determined by nuclear magnetic resonance spectroscopy. Biochemistry 35:14381-14394.

Jürgen OM. 2006. Production of liquid hydrocarbons from biomass. Angew. Chem. Int. Ed. 45(5):696-698.

Kapdan IK, Kargi F. 2006. Bio-hydrogen production from waste materials. Enzyme Microb. Technol. 38(5):569-582.
Karita S, Sakka K, Ohmiya K. 1996. Cellulose-binding domains confer an enhanced activity against insoluble cellulose to *Ruminococcus albus* endoglucanase IV. J. Ferment. Bioeng. 81:553-556.
Kawaguchi T, Ikeuchi Y, Tsutsumi N, Kan A, Sumitani J-I, Arai M. 1998. Cloning, nucleotide sequence, and expression of the *Clostridium thermocellum* cellodextrin phosphorylase gene and its application to synthesis of cellulase inhibitors. J. Ferment. Bioeng. 85(2):144-149.
Kerby RL, Ludden PW, Roberts GP. 1995. Carbon monoxide-dependent growth of *Rhodospirillum rubrum*. J. Bacteriol. 177(8):2241-2244.
Kim M-S, Baek J-S, Lee JK. 2006. Comparison of H_2 accumulation by *Rhodobacter sphaeroides* KD131 and its uptake hydrogenase and PHB synthase deficient mutant. Int. J. Hydrogen Energy 31(1):121-127.
Kimura M. 1960. Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. J. Genet. 57(1):21-34.
Kittur FS, Mangala SL, Rus'd AA, Kitaoka M, Tsujibo H, Hayashi K. 2003. Fusion of family 2b carbohydrate-binding module increases the catalytic activity of a xylanase from *Thermotoga maritima* to soluble xylan. FEBS Lett. 549(1-3):147-151.
Klass DL. 1998. Biomass for renewable energy, fuels, and chemicals San Diego, CA, USA: Academic Press.
Klass DL, Emert GH. 1981. Fuels from biomass and wastes. Ann Arbor, MI, USA: Ann Arbor Science Publishers, Inc.
Kleerebezem R, van Loosdrecht MCM. 2007. Mixed culture biotechnology for bioenergy production. Curr. Opin. Biotechnol. 18(3):207-212.
Koku H, Eroglu I, Gündüz U, Yücel M, Türker L. 2002. Aspects of the metabolism of hydrogen production by *Rhodobacter sphaeroides*. Int. J. Hydrogen Energy 27(11-12):1315-1329.
Koku H, Eroglu I, Gündüz U, Yücel M, Türker L. 2003. Kinetics of biological hydrogen production by the photosynthetic bacterium *Rhodobacter sphaeroides* O.U. 001. Int. J. Hydrogen Energy 28(4):381-388.
Kondo T, Arakawa M, Hirai T, Wakayama T, Hara M, Miyake J. 2002. Enhancement of hydrogen production by a photosynthetic bacterium mutant with reduced pigment. J. Biosci. Bioeng. 93(2):145-150.
Kotay SM, Das D. 2008. Biohydrogen as a renewable energy resource--Prospects and potentials. Int. J. Hydrogen Energy 33(1):258-263.
Krishnareddy M, Kim YK, Kitaoka M, Mori Y, Hayashi K. 2002. Cellodextrin phosphorylase from *Clostridium thermocellum* YM4 strain expressed in *Escherichia coli*. J. Appl. Glycosci. 49:1-8.
Kumar N, Das D. 2001. Continuous hydrogen production by immobilized *Enterobacter cloacae* IIT-BT 08 using lignocellulosic materials as solid matrices. Enzyme Microb. Technol. 29(4-5):280-287.
Kumar S, Nussinov R. 2001. How do thermophilic proteins deal with heat? Cell. Mol. Life Sci. 58(9):1216-1233.
Kumazawa S, Mitsui A. 1981. Characterization and optimization of hydrogen photoproduction by a saltwater blue-green alga, *Oscillatoria sp.* Miami BG7. I. Enhancement through limiting the supply of nitrogen nutrients. Int. J. Hydrogen Energy 6(4):339-348.

Kurland CG. 1991. Codon bias and gene expression. FEBS Lett. 285:165-169.
Lambert C, Léonard N, De Bolle X, Depiereux E. 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18(9):1250-1256.
Lay J-J. 2001. Biohydrogen generation by mesophilic anaerobic fermentation of microcrystalline cellulose. Biotechnol. Bioeng. 74(4):280-287.
Lee C-M, Chen P-C, Wang C-C, Tung Y-C. 2002. Photohydrogen production using purple nonsulfur bacteria with hydrogen fermentation reactor effluent. Int. J. Hydrogen Energy 27(11-12):1309-1313.
Lehmann M, Kostrewa D, Wyss M, Brugger R, D'Arcy A, Pasamontes L, van Loon APGM. 2000. From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. Protein Eng. 13(1):49-57.
Lehmann M, Wyss M. 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr. Opin. Biotechnol. 12(4):371-375.
Lehrer GM, Barker R. 1970. Conformational changes in rabbit muscle aldolase: Kinetic studies. Biochemistry 9(7):1533-1540.
Levin DB, Pitt L, Love M. 2004. Biohydrogen production: prospects and limitations to practical application. Int. J. Hydrogen Energy 29(2):173-185.
Li Y, Irwin DC, Wilson DB. 2010. Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of *Thermobifida fusca* Cel9A. Appl. Environ. Microbiol. 76:2582-2588.
Lin S, Harada M, Suzuki Y, Hatano H. 2005. Process analysis for hydrogen production by reaction integrated novel gasification (HyPr-RING). Energy Convers. Manage. 46(6):869-880.
Liu A, Tomita H, Li H, Miyaki H, Aoyagi C, Kaneko S, Hayashi K. 1998. Cloning, sequencing and expression of the cellobiose phosphorylase gene of *Cellvibrio gilvus*. J. Ferment. Bioeng. 85(5):511-513.
Liu H, Grot S, Logan BE. 2005. Electrochemically assisted microbial production of hydrogen from acetate. Environ. Sci. Technol. 39(11):4317-4320.
Liu W, Hong J, Bevan DR, Zhang YHP. 2009. Fast identification of thermostable β-glucosidase mutants on cellobiose by a novel combinatorial selection/screening approach. Biotechnol. Bioeng. 103:1087-1094.
Liu W, Zhang A, Cheng Y, Zhou H, Yan Y. 2007. Effect of magnesium ions on the thermal stability of human poly(A)-specific ribonuclease. FEBS Lett. 581:1047-1052.
Liu W, Zhang X, Zhang Z, Zhang YHP. 2010. Engineering of *Clostridium phytofermentans* endoglucanase Cel5A for improved thermostability. Appl. Environ. Microbiol. 76:4914-4917.
Lo HC, Fish RH. 2002. Biomimetic NAD+ models for tandem cofactor regeneration, horse liver alcohol dehydrogenase recognition of 1,4-NADH derivatives, and chiral synthesis. Angew. Chem. Int. Ed. 41(3):478-481.
Logan BE. 2004. Extracting hydrogen and electricity from renewable resources. Environ. Sci. Technol. 38(9):160A-167A.
Logan BE, Regan J. 2006. Electricity-producing bacterial communities in microbial fuel cells. Trends Microbiol. 14(12):512-518.

Lou J, Dawson K, Strobel H. 1996. Role of phosphorolytic cleavage in cellobiose and cellodextrin metabolism by the ruminal bacterium *Prevotella ruminicola*. Appl. Environ. Microbiol. 62(5):1770-1773.

Lowry OH, Passonneau JV. 1969. Phosphoglucomutase kinetics with the phosphates of fructose, glucose, mannose, ribose, and galactose. J. Biol. Chem. 244(4):910-916.

Lutz S. 2010. Beyond directed evolution--semi-rational protein engineering and design. Curr. Opin. Biotechnol. 21(6):734-743.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. 66:506-577.

Ma K, Weiss R, Adams MWW. 2000. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. J. Bacteriol. 182(7):1864-1871.

Ma K, Zhou ZH, Adams MWW. 1994. Hydrogen production from pyruvate by enzymes purified from the hyperthermophilic archaeon, *Pyrococcus furiosus*: A key role for NADPH. FEMS Microbiol. Lett. 122:245-250.

Maddaiah V, Madsen N. 1966. Kinetics of purified liver phosphorylase. J. Biol. Chem. 241:3873-3881.

Maeda T, Sanchez-Torres V, Wood T. 2008a. Protein engineering of hydrogenase 3 to enhance hydrogen production. Appl. Microbiol. Biotechnol. 79(1):77-86.

Maeda T, Sanchez-Torres V, Wood TK. 2008b. Metabolic engineering to enhance bacterial hydrogen production. Microbial Biotechnol. 1(1):30-39.

Maglione G, Matsushita O, Russell JB, Wilson DB. 1992. Properties of a genetically reconstructed *Prevotella ruminicola* endoglucanase. Appl. Environ. Microbiol. 58(11):3593-3597.

Manfred Rizzi MBUTMR. 1997. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. Biotechnol. Bioeng. 55(4):592-608.

Mangala SL, Kittur FS, Nishimoto M, Sakka K, Ohmiya K, Kitaoka M, Hayashi K. 2003. Fusion of family VI cellulose binding domains to *Bacillus halodurans* xylanase increases its catalytic activity and substrate-binding capacity to insoluble xylan. J. Mol. Catal. B: Enzym. 21(4-6):221-230.

Manish S, Banerjee R. 2008. Comparison of biohydrogen production processes. Int. J. Hydrogen Energy 33(1):279-286.

Marchand M, Kooystra U, Wierenga RK, Lambeir A-M, Van Beeumen J, Opperdoes FR, Michels PAM. 1989. Glucosephosphate isomerase from *Trypanosoma brucei*: Cloning and characterization of the gene and analysis of the enzyme. Eur. J. Biochem. 184(2):455-464.

Matsumura Y, Minowa T, Potic B, Kersten SRA, Prins W, van Swaaij WPM, van de Beld B, Elliott DC, Neuenschwander GG, Kruse A and others. 2005. Biomass gasification in near- and super-critical water: Status and prospects. Biomass Bioenerg. 29(4):269-292.

McCartney L, Blake AW, Flint J, Bolam DN, Boraston AB, Gilbert HJ, Knox JP. 2006. Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. Proc. Nat. Acad. Sci. USA 103:4765-4770.

McLean BW, Boraston AB, Brouwer D, Sanaie N, Fyfe CA, Warren RAJ, Kilburn DG, Haynes CA. 2002. Carbohydrate-binding modules recognize fine substructures of cellulose. J. Biol. Chem. 277:50245-50254.

McTavish H, Sayavedra-Soto LA, Arp DJ. 1995. Substitution of *Azotobacter vinelandii* hydrogenase small-subunit cysteines by serines can create insensitivity to inhibition by O_2 and preferentially damages H_2 oxidation over H_2 evolution. J. Bacteriol. 177(14):3960-3964.

Meng M, Bagdasarian M, Zeikus JG. 1993. Thermal stabilization of xylose isomerase from *Thermoanaerobacterium thermosulfurigenes*. Bio/Technology 11:1157-1161.

Milne TA, Elam, C. C. & Evans, R. J. 2002. Hydrogen from biomass: State of the art and research challenges. National Renewable Energy Laboratory, Golden, CO. 1-82 p.

Minowa T, Fang Z. 1998. Hydrogen production from cellulose in hot compressed water using reduced nickel catalyst: Product distribution at different reaction temperatures. J. Chem. Eng. Jpn 31(3):488-491.

Minowa T, Zhen F, Ogi T. 1998. Cellulose decomposition in hot-compressed water with alkali or nickel catalyst. J. Supercrit. Fluids 13(1-3):253-259.

Miura T, Sonigo P. 2001. A mathematical model for experimental gene evolution. J. Theor. Biol. 209(4):497-502.

Miura Y. 1995. Hydrogen production by biophotolysis based on microalgal photosynthesis. Process Biochem. 30(1):1-7.

Miyake J, Miyake M, Asada Y. 1999. Biotechnological hydrogen production: research for efficient light energy conversion. J. Biotechnol. 70(1-3):89-101.

Mizuno O, Dinsdale R, Hawkes FR, Hawkes DL, Noike T. 2000. Enhancement of hydrogen production from glucose by nitrogen gas sparging. Bioresource Technol. 73(1):59-65.

Moehlenbrock M, Minteer S. 2008. Extended lifetime biofuel cells. Chem. Soc. Rev. 37:1188-1196.

Morse DE, Horecker BL. 1968. The mechanism of action of aldolases. In: Nord FF, editor. Advances in Enzymology and Related Areas of Molecular Biology. New York: Interscience Publishers. p 125-181.

Mosi R, Withers SG. 1999. Synthesis and kinetic evaluation of 4-deoxymaltopentaose and 4-deoxymaltohexaose as inhibitors of muscle and potato a-glucan phosphorylases. Biochem. J. 338:251-256.

Myung S, Wang Y, Zhang YHP. 2010. Fructose-1,6-bisphosphatase from a hyper-thermophilic bacterium *Thermotoga maritima*: Characterization, metabolite stability, and its implications. Process Biochem. 45(12):1882-1887.

Nagamine Y, Kawasugi T, Miyake M, Asada Y, Miyake J. 1996. Characterization of photosynthetic bacterium *Rhodobacter sphaeroides* RV for hydrogen production. J. Mar. Biotechnol. 4:34-37.

Narvaez I, Corella J, Orio A. 1997. Fresh tar (from a biomass gasifier) elimination over a commercial steam-reforming catalyst: Kinetics and effect of different variables of operation. Ind. Eng. Chem. Res. 36(2):317-327.

Nath K, Das D. 2004. Improvement of fermentative hydrogen production: Various approaches. Appl. Microbiol. Biotechnol. 65(5):520-529.

Navarro RM, Pena MA, Fierro JLG. 2007. Hydrogen production reactions from carbon feedstocks: Fossil fuels and biomass. Chem. Rev. 108:3952-3991.

Newgard CB, Hwang PK, Fletterick RJ. 1989. The family of glycogen phosphorylases: Structure and function. Crit. Rev. Biochem. Mol. Biol. 24:69-99.

Ni M, Leung DYC, Leung MKH, Sumathy K. 2006. An overview of hydrogen production from biomass. Fuel Process. Technol. 87(5):461-472.

Nidetzky B, Weinhäusel A, Haltrich D, Kulbe KD, Schinzel R. 1996. Maltodextrin phosphorylase from *Escherichia coli*: Production and application for the synthesis of α-glucose-1-phosphate. Ann. N.Y. Acad. Sci. 782:208-218.

NIST. http://webbook.nist.gov/chemistry/form-ser.html.

Noltmann EA. 1972. Aldose-ketose isomerases. In: Boyer PD, editor. The Enzymes. 3rd ed. New York: Academic Press. p 271-354.

Notenboom V, Boraston AB, Kilburn DG, Rose DR. 2001. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry 40:6248-6256.

O'Reilly M, Watson K, Schinzel R, Palm D, Johnson L. 1997. Oligosaccharide substrate binding in *Escherichia coli* maltodextrin phosphorylase. Nat. Struct. Biol. 4:405-412.

Ohdan K, Fujii K, Yanase M, Takaha T, Kuriki T. 2007. Phosphorylase coupling as a tool to convert cellobiose into amylose. J. Biotechnol. 127:496-502.

Pacala S, Socolow R. 2004. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. Science 305(5686):968-972.

Penner SS. 2006. Steps toward the hydrogen economy. Energy 31(1):33-43.

Percy A, Ono H, Hayashi K. 1998. Acceptor specificity of cellobiose phosphorylase from *Cellvibrio gilvus*: Synthesis of three branched trisaccharides. Carbohydr. Res. 308(3-4):423-429.

Perl D, Mueller U, Heinemann U, Schmid FX. 2000. Two exposed amino acid residues confer thermostability on a cold shock protein. Nat. Struct. Biol. 7:380-383.

Polizzi KM, Chaparro-Riggers JF, Vazquez-Figueroa E, Bommarius AS. 2006. Structure-guided consensus approach to create a more thermostable penicillin G acylase. Biotechnol. J. 1(5):531-536.

Polle JEW, Kanakagiri S, Jin E, Masuda T, Melis A. 2002. Truncated chlorophyll antenna size of the photosystems--a practical method to improve microalgal productivity and hydrogen production in mass culture. Int. J. Hydrogen Energy 27(11-12):1257-1264.

Pritchard L, Corne D, Kell D, Rowland J, Winson M. 2005. A general model of error-prone PCR. J. Theor. Biol. 234(4):497-509.

Probstein RF, Hicks RE. 2006. Synthetic fuels. New York, USA: Dover Publications.

Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucl. Acids Res. 35:W126-W131.

Rath VL, Ammirati M, Danley DE, Ekstrom JL, Gibbs EM, Hynes TR, Mathiowetz AM, McPherson RK, Olson TV, Treadway JL and others. 2000a. Human liver glycogen phosphorylase inhibitors bind at a new allosteric site. Chem. Biol. 7(9):677-682.

Rath VL, Ammirati M, LeMotte PK, Fennell KF, Mansour MN, Danley DE, Hynes TR, Schulte GK, Wasilko DJ, Pandit J. 2000b. Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core. Mol. Cell 6:139-148.

Rath VL, Hwang PK, Fletterick RJ. 1992. Purification and crystallization of glycogen phosphorylase from *Saccharomyces cerevisiae*. J. Mol. Biol. 225:1027-1034.

Ravalason H, Herpoël-Gimbert I, Record E, Bertaud F, Grisel S, de Weert S, van den Hondel CAMJJ, Asther M, Petit-Conil M, Sigoillot J-C. 2009. Fusion of a family 1 carbohydrate binding module of *Aspergillus niger* to the *Pycnoporus cinnabarinus* laccase for efficient softwood kraft pulp biobleaching. J. Biotechnol. 142(3-4):220-226.

Reichenbecher M, Lottspeich F, Bronnenmeier K. 1997. Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium Stercorarium*. Eur. J. Biochem. 247(1):262-267.

Rendina AR, Hermes JD, Cleland WW. 1984. Use of multiple isotope effects to study the mechanism of 6-phosphogluconate dehydrogenase. Biochemistry 23(25):6257-6262.

Rezaiyan J, Cheremisinoff NP. 2005. Gasification technologies: A primer for engineers and scientist. New York, USA: CRC press. 119-145 p.

Ribbe M, Gadkari D, Meyer O. 1997. N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. J. Biol. Chem. 272(42):26627-26633.

Rogers PV, Luo S, Sucic JF, Rutherford CL. 1992. Characterization and cloning of glycogen phosphorylase 1 from *Dictyostelium discoideum*. BBA - Gene Struct. Expr. 1129:262-272.

Rollin JA, Zhu Z, Sathitsuksanoh N, Zhang YHP. 2010. Increasing cellulose accessibility is more important than removing lignin: A comparison of cellulose solvent-based lignocellulose fractionation and soaking in aqueous ammonia. Biotechnol. Bioeng. 108:22-30.

Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 10(12):866-876.

Saheki S, Takeda A, Shimazu T. 1985. Assay of inorganic phosphate in the mild pH range, suitable for measurement of glycogen phosphorylase activity. Anal. Biochem. 148:277-281.

Salge JR, Dreyer BJ, Dauenhauer PJ, Schmidt LD. 2006. Renewable hydrogen from nonvolatile fuels by reactive flash volatilization. Science 314(5800):801-804.

Samain E, Lancelon-Pin C, Férigo F, Moreau V, Chanzy H, Heyraud A, Driguez H. 1995. Phosphorolytic synthesis of cellodextrins. Carbohydr. Res. 271(2):217-226.

Sasaki M, Fang Z, Fukushima Y, Adschiri T, Arai K. 2000. Dissolution and hydrolysis of cellulose in subcritical and supercritical water. Ind. Eng. Chem. Res. 39(8):2883-2890.

Schinzel R, Nidetzky B. 1999. Bacterial α-glucan phosphorylases. FEMS Microbiol. Lett. 171:73-79.

Schlapbach L, Zuttel A. 2001. Hydrogen-storage materials for mobile applications. Nature 414:353-358.

Schmidt-Dannert C, Arnold FH. 1999. Directed evolution of industrial enzymes. Trends Biotechnol. 17(4):135-136.

Schomburg D, Schomburg I, Chang A. 2009. Class 2 : Transferases IV. In: Schomburg D, Schomburg I, editors. Springer handbook of enzymes: Springer, New York. p 434-438.

Shabaker JW, Huber GW, Davda RR, Cortright RD, Dumesic JA. 2003. Aqueous-phase reforming of ethylene glycol over supported platinum catalysts. Catal. Lett. 88(1):1-8.

Shafikhani S, Siegel R, Ferrari E, Schellenberger V. 1997. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. BioTechniques 23(2):304-310.

Sheth K, Alexander JK. 1967. Cellodextrin phosphorylase from *Clostridium thermocellum*. Biochim. Biophys. Acta, Gen. Subj. 148(3):808-810.

Shi X-Y, Yu H-Q. 2004. Hydrogen production from propionate by *Rhodopseudomonas capsulata*. Appl. Biochem. Biotechnol. 117(3):143-154.

Shi X-Y, Yu H-Q. 2005. Response surface analysis on the effect of cell concentration and light intensity on hydrogen production by *Rhodopseudomonas capsulata*. Process Biochem. 40(7):2475-2481.

Shoseyov O, Shani Z, Levy I. 2006. Carbohydrate binding modules: Biochemical properties and novel applications. Microbiol. Mol. Biol. Rev. 70(2):283-295.

Simpson PJ, Jamieson SJ, Abou-Hachem M, Karlsson EN, Gilbert HJ, Holst O, Williamson MP. 2002. The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. Biochemistry 41:5712-5719.

Solomon BO, Zeng AP, Biebl H, Schlieker H, Posten C, Deckwer WD. 1995. Comparison of the energetic efficiencies of hydrogen and oxychemicals formation in *Klebsiella pneumoniae* and *Clostridium butyricum* during anaerobic growth on glycerol. J. Biotechnol. 39(2):107-117.

Srivastava DK, Bernhard SA. 1986. Metabolite transfer via enzyme-enzyme complexes. Science 234(4780):1081-1086.

Steipe B, Schiller B, Plückthun A, Steinbacher S. 1994. Sequence statistics reliably predict stabilizing mutations in a protein domain. J. Mol. Biol. 240(3):188-192.

Stevens DJ. 2001. Hot gas conditioning: Recent progress with larger-scale biomass gasification systems. National Renewable Energy Laboratory.

Sun F. 1995. The polymerase chain reaction and branching processes. J. Comput. Biol. 2(1):63-86.

Suzuki T, Yasugi M, Arisaka F, Yamagishi A, Oshima T. 2001. Adaptation of a thermophilic enzyme, 3-isopropylmalate dehydrogenase, to low temperatures. Protein Eng. 14:85-91.

Takabatake H, Suzuki K, Ko I-B, Noike T. 2004. Characteristics of anaerobic ammonia removal by a mixed culture of hydrogen producing photosynthetic bacteria. Bioresource Technol. 95(2):151-158.

Takaha T, Yanase M, Takata H, Okada S. 2001. Structure and properties of *Thermus aquaticus* α-glucan phosphorylase expressed in *Escherichia coli*. J. Appl. Glycosci. 48:71-78.

Takata H, Takaha T, Okada S, Takagi M, Imanaka T. 1998. Purification and characterization of *α*-glucan phosphorylase from *Bacillus stearothermophilus*. J. Ferment. Bioeng. 85:156-161.

Tamagnini P, Axelsson R, Lindberg P, Oxelfelt F, Wunschiers R, Lindblad P. 2002. Hydrogenases and hydrogen metabolism of cyanobacteria. Microbiol. Mol. Biol. Rev. 66(1):1-20.

Tanaka K, Kawaguchi T, Imada Y, Ooi T, Arai M. 1995. Purification and properties of cellobiose phosphorylase from *Clostridium thermocellum*. J. Ferment. Bioeng. 79(3):212-216.

Tang L, Huang H. 2005. Plasma pyrolysis of biomass for production of syngas and carbon adsorbent. Energ. Fuel. 19(3):1174-1178.

Tanisho S, Kuromoto M, Kadokura N. 1998. Effect of CO_2 removal on hydrogen production by fermentation. Int. J. Hydrogen Energy 23(7):559-563.

Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV and others. 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur. J. Biochem. 267(17):5313-5329.

Tewari Y, Goldberg R. 1989. Thermodynamics of hydrolysis of disaccharides: Cellobiose, gentiobiose, isomaltose, and maltose. J. Biol. Chem. 264:3966-3971.

Tracewell CA, Arnold FH. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr. Opin. Chem. Biol. 13(1):3-9.
Tsygankov AA, Fedorov AS, Laurinavichene TV, Gogotov IN, Rao KK, Hall DO. 1998. Actual and potential rates of hydrogen photoproduction by continuous culture of the purple non-sulphur bacterium *Rhodobacter capsulatus*. Appl. Microbiol. Biotechnol. 49(1):102-107.
Turner JA. 2004. Sustainable hydrogen production. Science 305(5686):972-974.
Ueno Y, Tatara M, Fukui H, Makiuchi T, Goto M, Sode K. 2007. Production of hydrogen and methane from organic solid wastes by phase-separation of anaerobic process. Bioresource Technol. 98(9):1861-1865.
Uffen RL. 1983. Metabolism of carbon monoxide by *Rhodopseudomonas gelatinosa*: Cell growth and properties of the oxidation system. J. Bacteriol. 155(3):956-965.
Valenzuela MB, Jones CW, Agrawal PK. 2006. Batch aqueous-phase reforming of woody biomass. Energ. Fuel. 20(4):1744-1752.
van Bueren AL, Morland C, Gilbert HJ, Boraston AB. 2005. Family 6 carbohydrate binding modules recognize the non-reducing end of *β*-1,3-linked glucans by presenting a unique ligand binding surface. J. Biol. Chem. 280(1):530-537.
Van den Burg B, Vriend G, Veltman OR, Venema G, Eijsink VGH. 1998. Engineering an enzyme to resist boiling. Proc. Nat. Acad. Sci. USA 95(5):2056-2060.
van Groenestijn JW, Hazewinkel JHO, Nienoord M, Bussmann PJT. 2002. Energy aspects of biological hydrogen production in high rate bioreactors operated in the thermophilic temperature range. Int. J. Hydrogen Energy 27(11-12):1141-1147.
Vaseghi S, Baumeister A, Rizzi M, Reuss M. 1999. *In Vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. Metab. Eng. 1(2):128-140.
Vasic-Racki D. 2006. History of industrial biotransformations - Dreams and realities KGaA. In: A. Liese SSaCW, editor. Industrial biotransformations: Wiley-VCH, Weinheim. p 1-37.
Vavilin VA, Rytow SV, Lokshina LY. 1995. Modelling hydrogen partial pressure change as a result of competition between the butyric and propionic groups of acidogenic bacteria. Bioresource Technol. 54(2):171-177.
Vázquez-Figueroa E, Chaparro-Riggers J, Bommarius AS. 2007. Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. ChemBioChem 8(18):2295-2301.
Vignais PM, Magnin J-P, Willison JC. 2006. Increasing biohydrogen production by metabolic engineering. Int. J. Hydrogen Energy 31(11):1478-1483.
Wagner U, Geiger B, Schaefer H. 1998. Energy life cycle analysis of hydrogen systems. Int. J. Hydrogen Energy 23(1):1-6.
Wang Y, Huang W, Sathitsuksanoh N, Zhu Z, Zhang YHP. 2011. Biohydrogenation from Biomass Sugar Mediated by In Vitro Synthetic Enzymatic Pathways. Chem. Biol. 18(3):372-380.
Wang Y, Zhang YHP. 2009a. Cell-free protein synthesis energized by slowly-metabolized maltodextrin. BMC Biotechnol. 9:59.
Wang Y, Zhang YHP. 2009b. Overexpression and simple purification of the *Thermotoga maritima* 6-phosphogluconate dehydrogenase in *Escherichia coli* and its application for NADPH regeneration. Microb. Cell Fact. 8:30.
Wang Y, Zhang YHP. 2010. A highly active phosphoglucomutase from *Clostridium thermocellum*: Cloning, purification, characterization and enhanced thermostability. J. Appl. Microbiol. 108(1):39-46.

Watson KA, McCleverty C, Geremia S, Cottaz S, Driguez H, Johnson LN. 1999. Phosphorylase recognition and phosphorolysis of its oligosaccharide substrate: Answers to a long outstanding question. EMBO J. 18:4619-4632.

Weerachanchai P, Horio M, Tangsathitkulchai C. 2009. Effects of gasifying conditions and bed materials on fluidized bed steam gasification of wood biomass. Bioresource Technol. 100(3):1419-1427.

Weinhausel A, Griessler R, Krebs A, Zipper P, Haltrich D, Kulbe KD, Nidetzky B. 1997. alpha-1,4-D-glucan phosphorylase of gram-positive *Corynebacterium callunae*: isolation, biochemical properties and molecular shape of the enzyme from solution X-ray scattering. Biochem. J 326:773-783.

Wells J, Russell J, Shi Y, Weimer P. 1995. Cellodextrin efflux by the cellulolytic ruminal bacterium *Fibrobacter succinogenes* and its potential role in the growth of nonadherent bacteria. Appl. Environ. Microbiol. 61(5):1757-1762.

Wen F, Nair NU, Zhao HM. 2009. Protein engineering in designing tailored enzymes and microorganisms for biofuels production. Curr. Opin. Biotechnol. 20:412-419.

Wong K-B, Lee C-F, Chan S-H, Leung T-Y, Chen YW, Bycroft M. 2003. Solution structure and thermal stability of ribosomal protein L30e from hyperthermophilic archaeon *Thermococcus celer*. Protein Sci. 12(7):1483-1495.

Woodruff WW, 3rd, Wolfenden R. 1979. Inhibition of ribose-5-phosphate isomerase by 4-phosphoerythronate. J. Biol. Chem. 254(13):5866-5867.

Woodward J, Orr M, Cordray K, Greenbaum E. 2000. Enzymatic production of biohydrogen. Nature 405(6790):1014-1015.

Xu X, Matsumura Y, Stenberg J, Antal MJ. 1996. Carbon-catalyzed gasification of organic feedstocks in supercritical water. Ind. Eng. Chem. Res. 35(8):2522-2530.

Yanase M, Takaha T, Kuriki T. 2006. alpha-Glucan phosphorylase and its use in carbohydrate engineering. J. Sci. Food Agric. 86:1631-1635.

Yanase M, Takata H, Fujii K, Takaha T, Kuriki T. 2005. Cumulative effect of amino acid replacements results in enhanced thermostability of potato type L α-glucan phosphorylase. Appl. Environ. Microbiol. 71:5433-5439.

Ye X, Chu J, Zhuang Y, Zhang S. 2005. Multi-scale methodology: A key to deciphering systems biology. Front. Biosci. 10:961-965.

Ye X, Rollin J, Zhang YHP. 2010. Thermophilic α-glucan phosphorylase from *Clostridium thermocellum*: Cloning, characterization and enhanced thermostability. J. Mol. Catal. B: Enzym. 65:110-116.

Ye X, Wang Y, Hopkins Robert C, Adams Michael WW, Evans Barbara R, Mielenz Jonathan R, Zhang YHP. 2009. Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. ChemSusChem 2(2):149-152.

Yokoi H, Maki R, Hirose J, Hayashi S. 2002. Microbial production of hydrogen from starch-manufacturing wastes. Biomass Bioenerg. 22(5):389-395.

Yokoi H, Mori S, Hirose J, Hayashi S, Takasaki Y. 1998. H_2 production from starch by a mixed culture of *Clostridium butyricum* and *Rhodobacter* sp. M-19. Biotechnol. Lett. 20(9):895-899.

Yokoi H, Saitsu A, Uchida H, Hirose J, Hayashi S, Takasaki Y. 2001. Microbial hydrogen production from sweet potato starch residue. J. Biosci. Bioeng. 91(1):58-63.

Yoshida A, Nishimura T, Kawaguchi H, Inui M, Yukawa H. 2005. Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. Appl. Environ. Microbiol. 71(11):6762-6768.

Yu D, Aihara M, Antal MJ. 1993. Hydrogen production by steam reforming glucose in supercritical water. Energ. Fuel. 7(5):574-577.

Zaccolo M, Gherardi E. 1999. The effect of high-frequency random mutagenesis on *in vitro* protein evolution: a study on TEM-1 β-lactamase. J. Mol. Biol. 285(2):775-783.

Zaccolo M, Williams DM, Brown DM, Gherardi E. 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J. Mol. Biol. 255(4):589-603.

Zanzi R, Sjostrom K, Bjornbom E. 2002. Rapid pyrolysis of agricultural residues at high temperature. Biomass Bioenerg. 23(5):357-366.

Zhang XZ, Zhang YHP. 2011. Simple, fast and high-efficiency transformation system for directed evolution of cellulase in *Bacillus subtilis*. Microb. Biotechnol. 4:98-105.

Zhang YHP. 2008a. Reviving the carbohydrate economy via multi-product biorefineries. J. Ind. Microbiol. Biotechnol. 35(5):367-375.

Zhang YHP. 2008b. A sweet solution. Public Service Review: European Union. Science & Technology. p 126.

Zhang YHP. 2009. A sweet out-of-the-box solution to the hydrogen economy: Is the sugar-powered car science fiction? Energy Environ. Sci. 2(3):272-282.

Zhang YHP. 2010a. Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: Challenges and opportunities. Biotechnol. Bioeng. 105(4):663-677.

Zhang YHP. 2010b. Renewable carbohydrates are a potential high-density hydrogen carrier. Int. J. Hydrogen Energy 35(19):10334-10342.

Zhang YHP, Cui JB, Lynd LR, Kuang LR. 2006a. A transition from cellulose swelling to cellulose dissolution by *o*-phosphoric acid: Evidences from enzymatic hydrolysis and supramolecular structure. Biomacromolecules 7:644-648.

Zhang YHP, Evans BR, Mielenz JR, Hopkins RC, Adams MWW. 2007. High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. PLoS One 2(5):e456.

Zhang YHP, Himmel ME, Mielenz JR. 2006b. Outlook for cellulase improvement: Screening and selection strategies. Biotechnol Adv 24:452-481.

Zhang YHP, Lynd LR. 2003. Cellodextrin preparation by mixed-acid hydrolysis and chromatographic separation. Anal. Biochem. 322:225-232.

Zhang YHP, Lynd LR. 2004a. Kinetics and relative importance of phosphorolytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of *Clostridium thermocellum*. Appl. Environ. Microbiol. 70:1563-1569.

Zhang YHP, Lynd LR. 2004b. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol. Bioeng. 88:797-824.

Zhang YHP, Lynd LR. 2005a. Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation. Proc. Natl. Acad. Sci. USA 102:7321-7325.

Zhang YHP, Lynd LR. 2005b. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. Biomacromolecules 6:1510-1515.

Zhang YHP, Lynd LR. 2006. A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnol. Bioeng. 94:888-898.

Zhang YHP, Mielenz JR. 2011. Renewable hydrogen carrier - carbohydrate: Constructing the carbon-neutral carbohydrate economy. Energies 4(2):254-275.
Zhang YHP, Ye X, Wang Y. 2008. Biofuels production by cell free synthetic enzymatic technology. In: Richter F, editor. Biotechnology: Research, Technology and Applications. Hauppauge, NY, US: Nova Science Publishers.
Zhao H, Chockalingam K, Chen Z. 2002. Directed evolution of enzymes and pathways for industrial biocatalysis. Curr. Opin. Biotechnol. 13(2):104-110.
Zhu H, Wakayama T, Asada Y, Miyake J. 2001. Hydrogen production by four cultures with participation by anoxygenic phototrophic bacterium and anaerobic bacterium in the presence of NH_4^+. Int. J. Hydrogen Energy 26(11):1149-1154.
Zurrer H, Bachofen R. 1979. Hydrogen production by the photosynthetic bacterium *Rhodospirillum rubrum*. Appl. Environ. Microbiol. 37(5):789-793.
Zürrer H, Bachofen R. 1982. Aspects of growth and hydrogen production of the photosynthetic bacterium *Rhodospirillum rubrum* in continuous culture. Biomass 2(3):165-174.