ENGINEERING A FUNGAL L-ARABINOSE PATHWAY TOWARDS THE CO-UTILIZATION OF HEMICELLULOSIC SUGARS FOR PRODUCTION OF XYLITOL

BY

RYAN PATRICK SULLIVAN

DISSEPTION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2009

Urbana, Illinois

Doctoral Committee:

Professor Huimin Zhao, Chair
Professor Daniel Pack
Assistant Professor Nathan Price
Professor Hans Blaschek
ABSTRACT

The biosynthesis of value-added products has shown great promise in recent years due to the advances in molecular biology and protein engineering. Many advantages over chemical synthesis include high selectivity and specificity, increased yield, and function under milder conditions without the need for toxic metals, organic solvents, strong acids or bases, or high pressures and temperatures. However, naturally occurring enzymatic pathways are often times not particularly well suited towards obtaining high product yield due to the organism’s evolution not dependent on the production of such high levels of desired product. Xylitol is an attractive pentose sugar alcohol that has many applications in the food, pharmaceutical, and high-value based biochemical product industries. It is still, however, relatively expensive to produce, which prevents its economical integration into the marketplace. To expand the possible routes of xylitol synthesis, and provide a process which can utilize both hemicellulosic pentose sugars D-xylose and L-arabinose from renewable plant biomass as feed substrates, my thesis research focused on utilizing a fungal-derived biosynthetic pathway for production of xylitol. The pathway consists of two NADPH-dependent reductases, xylose reductase (XR) and L-xylulose reductase (LXR), and an NAD$^+$-dependent L-arabinitol 4-dehydrogenase (LAD). The XR enzyme converts D-xylose directly to xylitol, while the three enzyme in tandem convert L-arabinose to xylitol. However, the cofactor imbalance presented with the pathway potentially makes nicotinamide regeneration difficult for production in vivo. After cloning and characterization of a highly active and stable NAD$^+$-dependent L-arabinitol 4-dehydrogenase (LAD) from *Neurospora crassa*, subsequent engineering via rational design and directed evolution resulted in the isolation of the first known NADP$^+$-dependent LAD enzyme. This novel engineered LAD was then introduced into the fungal xylitol pathway and expressed in a
model organism, *E. coli*, and the effect of the cofactor specificity alteration was evaluated in the conversion of L-arabinose to xylitol. Further investigation of the cofactor balancing benefits were applied to xylitol dehydrogenase (XDH) for full redox balancing of the initial steps of the L-arabinose pathway, and investigations of the limiting steps in L-arabinose utilization conducted in *S. cerevisiae* and *E. coli* led to further proposed engineering of LAD.
To my loving family
ACKNOWLEDGEMENTS

First and foremost, many thanks to my advisor, Huimin Zhao, for his constant guidance, his uncanny availability, and his great patience throughout the completion of my graduate career. Thanks to my committee members, Daniel Pack, Nathan Price, and Hans Blaschek, for their valued support. Thanks to every fellow lab member who was a part of the Zhao laboratory during my graduate career. In particular, I would like to thank Michael Simurdiak and Tyler Johannes for providing helpful instruction and training when I first joined the lab. Additional thanks to Nikhil Nair, Fei Wen, Sheryl Rubin-Pitel, Matthew Desieno, Michael McLauchlan, Zengyi Shao, and Hua Zhao for their intellectual insight and valued contributions. Special thanks goes to the members of the IGB/EBI team of Jing Du, Byoungjin Kim, Amit Ghosh, Sijin Li, Tae-Hee Lee, and Dawn Eriksen, who helped me tremendously during my final year of study. Thanks to Brian Bae, Stacy Keller, and Satish Nair for their crystallization collaborative efforts. Thanks to my undergraduate assistants Anu Biswas, Mark Laurenz, and Saroj Saha. Thanks to the many friends I have gained and whose friendships I hope to sustain past my stay in Illinois, particularly (and in no particular order) Matthew Cole and Jennifer Younker, Lon and Jessica Chubiz, Matthew and Kristen Willis, John and Jennifer Schmidt, Mohan Karulkar, Eric Mock, Matthew Langer, Fikhil Brushett and Karen Guralnick, Nicolas Londono, Esther Jeng, Tasha Desai, Rachel Graff, Alice Hollister, Grant Pitel, Nate Gabrielson, Jonathan Silvestre and Lily Wong, Doug Viehman, and Everett Scheer. To any names I may have missed throughout, please receive my sincerest apology and know that you’re in my heart always if not in my typing hands at this particular moment. Finally, a special thanks to my family – my father Kevin, my mother Caroline, and my sister Corrine – for always providing me with the strength and determination to follow through with my aspirations, and for being my true inspiration.
# TABLE OF CONTENTS

## CHAPTER 1 – Biosynthesis of Value-Added Products and Biofuels

1.1. Introduction ................................................................................................................. 1

1.2. Chemical Synthesis of Xylitol .................................................................................. 2

1.3. Biotechnological Synthesis of Xylitol from Xylose .................................................. 3

  1.3.1. Bacterial Production of Xylitol ........................................................................... 3

  1.3.2. Fungal Production of Xylitol .............................................................................. 3

  1.3.3. Yeast Production of Xylitol ............................................................................... 4

1.4. L-Arabinose as a Substrate for Xylitol Production ..................................................... 5

1.5. Protein Engineering .................................................................................................... 8

1.6. Project Overview ......................................................................................................... 9

References ............................................................................................................................ 12

## CHAPTER 2 – Cloning, Expression, Purification, and Characterization of L-Arabinitol 4-Dehydrogenase from *Neurospora crassa*

2.1. Introduction .................................................................................................................. 18

2.2. Results .......................................................................................................................... 19

  2.2.1. *N. crassa* LAD Gene Identification .................................................................... 19

  2.2.2. Cloning, Expression, and Purification of *N. crassa* LAD ................................. 21

  2.2.3. Steady-state Kinetics ......................................................................................... 23

  2.2.4. Temperature and pH Dependence ..................................................................... 26

  2.2.5. Determination of Protein Mass and Quaternary Structure ................................ 28

  2.2.6. Cofactor Specificity ........................................................................................... 29

  2.2.7. Homology Modeling ......................................................................................... 30

  2.2.8. F59 Mutational Analysis ..................................................................................... 32

2.3. Discussion .................................................................................................................... 32

2.4. Conclusions and Future Work ..................................................................................... 34

2.5. Materials and Methods ............................................................................................... 35

  2.5.1. Materials ............................................................................................................. 35

  2.5.2. RT-PCR and Cloning (Notebook #1, p. 42; Notebook #2, p. 51) ...................... 36
2.5.3. Lysate Assay ........................................................................................................... 37
2.5.4. GST-tag Removal (Notebook #2, p. 67) ................................................................. 38
2.5.5. Enzyme Kinetics (Notebook #2, pp. 79-102) ......................................................... 38
2.5.6. Thermal Dependence (Notebook #2, pp. 79-102) .................................................. 39
2.5.7. pH Rate Profile (Notebook #2, pp. 79-102) ........................................................... 39
2.5.8. Determination of Protein Mass and Quaternary Structure (Notebook #2, p. 202) ...... 40
2.5.9. HPLC Analysis (Notebook #1, p. 186) ................................................................. 40
2.5.10. Homology Modeling ............................................................................................ 40
References ....................................................................................................................... 42

CHAPTER 3 – Engineering an NADP\(^+\)-Dependent L-Arabinitol 4-Dehydrogenase for Xylitol Production ................................................................................................................. 45

3.1. Introduction .................................................................................................................. 45
3.2. Results ......................................................................................................................... 50
  3.2.1. Reversal of Cofactor Specificity of \(N.\ crassa\) LAD by Rational Design .................. 50
  3.2.2. Screening Procedure Validation and EP-PCR Results .......................................... 53
  3.2.3. Kinetic Data of the Engineered ncLAD Mutants ..................................................... 54
3.3. Discussion .................................................................................................................... 57
  3.3.1. Protein Engineering Strategy .................................................................................. 57
  3.3.2. Properties of the ncLAD-3x Mutant ....................................................................... 60
  3.3.3. Structural Analysis of the Mutations ..................................................................... 60
3.4. Conclusions ..................................................................................................................... 65
3.5. Materials and Methods ............................................................................................... 66
  3.5.1. Materials ............................................................................................................... 66
  3.5.2. Site-Directed and Saturation Mutagenesis ............................................................ 66
  3.5.3. Error-Prone PCR Mutagenesis and Library Creation (Notebook #3, p. 130) ........ 68
  3.5.4. Activity Assay Screening Procedure (Notebook #3, p. 136) .................................. 70
  3.5.5. DNA Sequencing and Analysis .............................................................................. 71
  3.5.6. LAD Over-Expression and Purification (Notebook #4, p. 230-231) ..................... 72
  3.5.7. HPLC Analysis of Products .................................................................................... 73
  3.5.8 Enzyme Kinetics (Notebook #4, p. 232) ............................................................... 73
References ......................................................................................................................... 75
CHAPTER 4 – Application of an Engineered Fungal Pathway for Xylitol Production from Hemicellulosic Sugars L-Arabinose and D-Xylose ........ 79

4.1. Introduction ...................................................................................................................................................... 79

4.2. Results .............................................................................................................................................................. 80

  4.2.1. Strain Development .................................................................................................................................. 80

  4.2.2. Operon Design ......................................................................................................................................... 82

  4.2.3. Shake Flask Studies ................................................................................................................................... 85

    4.2.3.1. Effect of Using Glucose as Co-Substrate and pH on Xylitol Production ........................................... 85

    4.2.3.2. Wild-type vs. Engineered Xylitol Pathway ....................................................................................... 86

  4.2.4. Bioreactor Studies ...................................................................................................................................... 88

    4.2.4.1. Wild-Type vs. Engineered Pathway ................................................................................................. 88

    4.2.4.2. Engineered Xylitol Pathway – pH control vs. non-pH control ............................................................ 92

    4.2.4.3. WT vs. ENGR Pathway – Co-utilization ......................................................................................... 94

4.3. Discussion ......................................................................................................................................................... 95

4.4. Conclusions and Outlook ................................................................................................................................. 98

4.5. Materials and Methods .................................................................................................................................. 99

  4.5.1. Materials .................................................................................................................................................. 99

  4.5.2. Gene Deletion (Notebook #3, p. 127) .................................................................................................... 99

  4.5.3. Mutant crp* Integration ......................................................................................................................... 100

  4.5.4. Operon Design and Construction (Notebook #3, p. 231-234) ............................................................. 101

  4.5.5. HPLC Analysis ....................................................................................................................................... 103

  4.5.6. Shake Flask Studies (Notebook #4, p. 279) .......................................................................................... 103

  4.5.7. Bioreactor Studies (Notebook #4, p. 158) ............................................................................................ 104

References ............................................................................................................................................................... 105

CHAPTER 5 – Fungal Arabinose Pathway Complementation Towards Determining Limitations in Arabinose Utilization ................. 107

5.1. Introduction ....................................................................................................................................................... 107

5.2. Results .............................................................................................................................................................. 110

  5.2.1. Cloning, Characterization, and Engineering of an NADP⁺-Dependent N. crassa Xylitol Dehydrogenase ......................................................... 110

  5.2.2. Introduction of L-Arabinose Pathway into S. cerevisiae ....................................................................... 114
5.2.3. *E. coli* Complementation Experiments .......................................................... 116
5.3. Discussion ............................................................................................................. 119
5.4. Conclusions ......................................................................................................... 121
5.5. Materials and Methods ..................................................................................... 121
  5.5.1. Materials .......................................................................................................... 121
  5.5.2. RT-PCR and Cloning ..................................................................................... 122
  5.5.3. Cell Lysate Assay .......................................................................................... 123
  5.5.4. Enzyme Kinetics ............................................................................................ 124
  5.5.5. pH Profile ....................................................................................................... 124
  5.5.6. Pathway Construction (*S. cerevisiae*) (Notebook #5, p. 101) ..................... 124
  5.5.7. Yeast Transformation (*S. cerevisiae*) ............................................................ 125
  5.5.8. Pathway Construction (*E. coli*) (Notebook #4, p. 151) .............................. 125
References .................................................................................................................... 127
CHAPTER 1 – Biosynthesis of Value-Added Products and Biofuels

1.1. Introduction

Xylitol is a five-carbon sugar alcohol with a growing market as a sweetener that can be produced by either chemical or biotechnological methods. The early research and development work of xylitol began in Finland during the 1970’s, and has been thoroughly reviewed by Mäkinen and Peldyak\(^1\). In 1974, the Finnish Sugar Company Ltd. began the first large-scale production of xylitol in Southern Finland. In 1975, the first sugar-free dental product was launched through the introduction of a xylitol chewing gum (first in Finland, then the United States). Over the next 25 years, global awareness of the significant advantages that xylitol offers to great-tasting confectionery and gum products has steadily increased, especially the unique health and dental benefits it provides. When continuously supplied in the diet, xylitol limits the tendency to obesity, and its incorporation into food formulations results in improvement of the color and taste of preparations without causing undesired changes in properties during storage. Along with fructose, xylitol is the sugar recommended for diabetic patients, due to its human metabolism not being insulin-mediated. In addition, the anticariogenic property of xylitol inhibits the growth of oral bacteria so as to reduce plaque formation\(^2\), and it has also been reported that xylitol helps to prevent acute otitis media (ear infection) in small children\(^3\). Recently the Department of Energy (DOE) reported that xylitol was considered one of the top twelve sugar-derived building blocks that can be subsequently converted to a number of high-value bio-based chemicals or materials\(^4\). However, there is currently limited commercial production of xylitol, due mainly in part to its comparatively high production cost (about 10 times that of alternatives sucrose or sorbitol) by chemical reduction of D-xylose\(^5\).
1.2. Chemical Synthesis of Xylitol

Industrially, xylitol is produced by the chemical reduction of pure D-xylose, obtained from hardwood hydrolysates, in the presence of a Raney nickel catalyst under high hydrogen pressure\(^6\). The pure D-xylose is obtained using an industrial scale chromatographic method for separating the different wood hemicellulosic sugars developed in Finland, where at the same time the beneficial effects of xylitol in preventing dental caries were found. Pure and defined D-xylose is the preferred starting material, which is difficult to obtain from the plant biomass. Pretreatment and hydrolysis of lignocellulose from biomass yield a multi-component product mixture that, in addition to the xylan (or D-xylose) fraction, contains other reducing sugars (such as D-glucose from starch, L-arabinose from arabinan, etc.), acids, and various products of decomposition of lignin and sugars\(^7\). The recovery of xylitol from the xylan fraction is about 50–60\%, or roughly 8–15\% of the raw material employed\(^8\). The value depends on the xylan content of the raw material. There have been a number of patents filed on methods for obtaining pure xylose and production of xylitol from xylan-containing materials\(^6,9-13\). Drawbacks of the chemical route of synthesis are the requirements of high pressure (up to 50 atm) and temperature (80–140 °C), use of an expensive catalyst, and use of extensive multi-step separation and purification techniques, including mechanical filtration and chromatographic separation, to both obtain pure initial D-xylose and to remove the other polyols and by-products after chemical reduction.
1.3. Biotechnological Synthesis of Xylitol from Xylose

1.3.1. Bacterial Production of Xylitol

Xylitol has been reported to be produced by bacteria such as *Enterobacter liquefaciens*\(^\text{14}\), *Corynebacterium sp.*\(^\text{15}\), and *Myobacterium smegmatis*\(^\text{16}\). In experiments leading to xylitol production, an *E. liquefaciens* strain was reported to reach 33.3 g xylitol L\(^{-1}\) in a fermentation medium containing 100 g initial xylose L\(^{-1}\), with a volumetric productivity of 0.35 g L\(^{-1}\) h\(^{-1}\)\(^\text{14}\). Fermentations with *M. smegmatis* (either as washed cells or as an immobilized form) transformed D-xylose to xylitol at the same rate, giving 70% yield, and xylitol production from D-xylose using commercial xylose isomerase from *Bacillus coagulans* and immobilized *M. smegmatis* achieved a yield of 4 g xylitol from 10 g of D-xylose\(^\text{16}\). Recombinant *Lactococcus lactis* carrying the xylose reductase from *Pichia stipitis* and the xylose transporter from *Lactobacillus brevis* gave xylitol productivity of 2.72 g L\(^{-1}\) h\(^{-1}\) during 20 h fermentations\(^\text{17}\). Some advantages of using bacterial hosts are a nearly quantitative yield of xylitol from xylose, competitive volumetric productivity, and no requirement for aeration. However, conversions typically do no proceed to completion at high initial xylose concentrations, requiring the implementation of separation and recycling of D-xylose remaining in the product medium.

1.3.2. Fungal Production of Xylitol

Conversion of D-xylose to xylitol in a fermentation system using a fungal culture *Petromyces albertensis* has been reported. In culture mediums containing 100 g L\(^{-1}\) initial D-xylose, cultures reached 39.8 g L\(^{-1}\) xylitol after 10 days\(^\text{18}\). After 160 h of anaerobic fermentation, *Fusarium oxysporum* utilized almost 35 g L\(^{-1}\) of xylose and the maximum xylitol concentration (11.5 g L\(^{-1}\)) was observed using ammonium as nitrogen source\(^\text{19}\). Small concentrations of xylitol have also been reported from cultured filamentous fungi including *Penicillium, Aspergillus, Rhizopus,*
Gliocladium, Byssochlamys, Myrothecium, and Neurospora when grown on D-xylose-based media\textsuperscript{20}.

1.3.3. Yeast Production of Xylitol

Xylitol can be produced by natural xylose-assimilating yeasts, and varies a great deal with species and aeration conditions. A wide range of yeast species belonging to the genus Candida are well-known for their potential industrial applications, and include Candida tropicalis\textsuperscript{21-25}, Candida guilliermondii\textsuperscript{26}, Candida boidini\textsuperscript{27,28}, Candida parapsilosis\textsuperscript{29}, Candida maltosa\textsuperscript{30}, and Candida peltata\textsuperscript{21,32}. Candida tropicalis produced 8.5 g L\textsuperscript{-1} h\textsuperscript{-1} xylitol from D-xylose (yield of 0.85 mol mol\textsuperscript{-1} D-xylose, final concentration 180 g L\textsuperscript{-1} xylitol) in a cell recycling process in a submerged membrane bioreactor with suction pressure and air sparging\textsuperscript{23}. When attempting to ferment the processed hydrolysate without the added step of purifying substrate D-xylose, C. tropicalis xylitol production was found to be 0.43 g/g and 0.45 g/g of D-xylose utilized with corn fiber and sugarcane bagasse hydrolysate respectively, and after strain improvement by sub-culturing, the production jumped to levels of 0.58 g/g and 0.65 g/g of D-xylose with corn fiber hydrolysate and sugarcane bagasse hydrolysate respectively\textsuperscript{33}. The productivity for an engineered C. tropicalis with an integrated NADH-prefering XR from Candida parapsilosis was 5.1 g L\textsuperscript{-1} h\textsuperscript{-1}\textsuperscript{34}. Debaryomyces hansenii (Candida famata) produced 4.6 g L\textsuperscript{-1} h\textsuperscript{-1} xylitol with a xylitol yield of 0.78 mol mol\textsuperscript{-1} from D-xylose and a final xylitol concentration of 221 g L\textsuperscript{-1}\textsuperscript{35}. The ability of Candida peltata to ferment D-xylose to xylitol was evaluated and a maximum xylitol yield of 0.56 g g\textsuperscript{-1} D-xylose was obtained when the yeast was cultivated in optimized conditions. The yeast also produced ethanol (0.41 g g\textsuperscript{-1} in 40 h) from glucose (50 g L\textsuperscript{-1}) and arabinitol (0.55 g g\textsuperscript{-1} in 87 h) from L-arabinose (50 g L\textsuperscript{-1}). Interestingly, moderate levels of glucose (10 g L\textsuperscript{-1}), ethanol (7.5 g L\textsuperscript{-1}) and acetate (5 g L\textsuperscript{-1}) inhibited xylitol production by 61, 84
and 68%, respectively, whereas L-arabinose (10 g L\(^{-1}\)) had no inhibitory effect on xylitol production\(^{31}\). Another reoccurring phenomenon in most fermentation experiments was that yeasts preferentially utilize glucose > D-xylose > L-arabinose from mixed substrates.

Xylitol production from D-glucose was also demonstrated using \(S. \text{cerevisiae}^{36}\). A transketolase-deficient strain (W303-1B \(t\)1 \(k\)l \(k\)2) growing on glucose accumulated D-xylulose 5-phosphate intracellularly and released ribitol and pentose sugars D-ribose, D-ribulose, and D-xylulose into the growth medium. Over-expressing the xylitol dehydrogenase from \(P. \text{stipitis}\), which catalyzes the reduction of D-xylulose to xylitol, and deletion of the xylulokinase-encoding gene (\(XKS1\)) resulted in the production of xylitol and ribitol from D-xylulose and D-ribulose, respectively. The yield of xylitol and ribitol formation was considerably lower than other methods, although the authors did not attempt to optimize production conditions or metabolically engineer their strain for improved PPP capacity.

### 1.4. L-Arabinose as a Substrate for Xylitol Production

Hemicellulose is the second most common polysaccharide in nature, representing 20–40% of lignocellulosic (plant material) biomass. Hemicelluloses are heterogeneous polymers comprised mostly of pentoses (D-xylose, L-arabinose), with minor contributions from hexoses (mannose, glucose, galactose), and sugar acids\(^{37}\). The utilization of the hemicellulose fraction for xylitol production has been the subject of numerous research papers and patents\(^9,33,37-41\), but only the D-xylose component is converted to xylitol, with the other major component of hemicellulose, L-arabinose, being converted to undesired byproduct L-arabinitol. However, D-xylose and L-arabinose metabolism in yeasts and filamentous fungi conveniently share xylitol as an intermediate (Figure 1.1). If the L-arabinose component of hemicellulose could also be utilized for xylitol production, the economic feasibility by fermentative means could increase
dramatically. However, L-arabinose utilization by yeasts and filamentous fungi is poorly characterized to date\textsuperscript{42}, and strains that do utilize L-arabinose typically produce large amounts of L-arabinitol, instead of xylitol, along with biomass, suggesting a limitation earlier in the utilization pathway\textsuperscript{43}.

One major issue that has yet to be resolved in utilization of L-arabinose as a substrate for xylitol or ethanol is the subject of redox cofactor imbalance. The fungal pathway consists of two oxidations and two reductions, which is redox neutral. However, the oxidations are NAD(H)-linked and the reductions are NADP(H)-linked, so that there is an imbalance of the cofactors involved in the process. In the view of xylitol production, the imbalance cannot be properly regenerated by the cell’s metabolism, leading to accumulation of L-arabinitol, the product of the first enzyme in the pathway. In the view of fermentation, it remains an open question how fungal micro-organisms cope with this cofactor imbalance. It has been suggested that NADPH is mainly regenerated through the oxidative part of the pentose phosphate pathway\textsuperscript{44}. The filamentous fungus \textit{Aspergillus niger} exhibited higher activities of glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase when growing on pentoses\textsuperscript{44}. In the oxidative part of the pentose phosphate pathway, however, the reduction of NADP\textsuperscript{+} is coupled to CO\textsubscript{2} production. In this case the anaerobic conversion of L-arabinose to CO\textsubscript{2} and ethanol is no longer redox neutral, i.e. the fermentation of L-arabinose is inhibited by an accumulation of reduced redox cofactors. Some attempts at engineering cofactor regeneration into these processes have met with limited success, and those pertaining to xylose utilization are mentioned below.

One proposed strategy of dealing with the cofactor imbalance was to integrate transhydrogenase activity to facilitate the equilibrium between the coenzymes involved. Transhydrogenases
catalyze the reduction of NADP$^+$ to NADPH with concomitant oxidation of NADH to NAD$^+$. Yeasts are not believed to have such endogenous activity$^{45}$, so the expression of a cytosolic transhydrogenase gene from *Azotobacter vinlandii* in *S. cerevisiae* was attempted and the intracellular concentrations of the NAD(P) and NAD(P)H cofactors were measured$^{46}$. It was hypothesized that since NADH could be consumed and NADPH produced by the enzyme, expression of the gene encoding transhydrogenase could result in a decrease in glycerol formation and the carbon flux through the pentose phosphate pathway, where there is a loss of carbon in the form of carbon dioxide. This reduction in the carbon flux towards ‘waste’ components could be redirected towards formation of ethanol, leading to higher yield. However, this was not found to be the case, as there was a resulting decrease in ethanol production and increase in production of 2-oxoglutarate and glycerol. Unfortunately, no mention was made of the effects on pentose consumption or xylitol production.

Another strategy was to induce changes in the pentose phosphate pathway itself in order to try and uncouple NADPH regeneration and CO$_2$ production. In one study, the disruption of the GND1 gene, one of the isogenes of 6-phosphogluconate dehydrogenase, or disruption of the ZWF1 gene, which encodes glucose 6-phosphate dehydrogenase, blocked the pentose phosphate pathway. These modifications resulted in a lower xylitol yield and higher ethanol yield than in control strains. Also, xylitol production was shown to be strongly connected to the flux through the oxidative part of the pentose phosphate pathway using metabolic flux analysis$^{47}$. A follow-up to this approach expressed GDP1, which encodes for a fungal NADP$^+$-dependent D-glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH), in a ΔZWF1 *S. cerevisiae* strain, resulting in further stimulated xylose fermentation with respect to rate and yield, with increased ethanol production under anaerobic conditions$^{48}$.
There was a recent publication during the course of the present thesis preparation by Bettiga et al.\textsuperscript{49} that attempted at resolving the redox balance for L-arabinose utilization, which sought to introduce a pathway that depends solely on NAD(H) using an engineered NADH-preferring xylose reductase and a native NADH-preferring L-xylulose reductase along with NAD\textsuperscript{+}-dependent dehydrogenases for a cofactor balanced pathway. While the authors were focused on improving pentose fermentation to ethanol, the attempt at improving the flux past L-arabinitol using a cofactor-balancing strategy was shown to be effective, and will be discussed in Chapter 5. The goal of my work was to engineer NADP\textsuperscript{+}-dependent L-arabinose pathway dehydrogenases by means of rational design and directed evolution for application in xylitol or biofuel production from hemicellulosic pentose sugars.

1.5. Protein Engineering

The tailoring of enzymes can be accomplished through two experimental routes. The first is rational design, which targets specific residues of a protein for mutagenesis to predetermined amino acid mutations, and is only applicable when there is detailed knowledge of the relationships between the enzyme’s structure and mechanism/function. And while an increasing number of enzymes are being characterized, the majority do not have this depth of information readily available, as it requires considerable effort to obtain. In the absence of this information, the tailoring of an enzyme can still be accomplished through the second route - directed evolution.

Directed evolution is the general term applied to the combined techniques of generation of a library of protein mutants (or variants) and selection of a protein with desirable function from within that library\textsuperscript{50}. It is an iterative Darwinian optimization process, whereby the fittest variants are selected from an ensemble of mutants\textsuperscript{51}. Directed evolution can be used to target a
number of enzymatic characteristics, including activity, substrate specificity, thermal and oxidative stability, enantioselectivity or enantiospecificity, pH optima or range, and tolerance to solvent\textsuperscript{52}. While a typical directed evolution experiment focuses on a single enzymatic trait, there are some examples of improving several traits simultaneously.

Choosing the appropriate methods of library generation and screening or selection is paramount to the success of any directed evolution experiment. Library diversity can be created through either mutagenesis (random or semi-rational) or gene recombination, and which of these methods is chosen depends on many factors, such as the availabilities of homologous genes, structural knowledge, and characteristic data of the enzyme of interest. The library size created is typically very large ($>10^{4-6}$), and close evaluation of each variant is not feasible. The need for a method to find the improved “needle in a haystack” enzyme becomes evident, with several strategies including selection, enrichment, and high-throughput screening offering ways to sift through the library clutter and find a variant with the desired enzymatic trait. However, once an evolved enzyme has been found that exhibits improved characteristics, the artificial conditions in which the selection method was carried out may result in an enzyme whose properties may not carry over to the real biocatalytic process. Therefore, the more similar a screening system is to the actual application process, the more likely it is to find an improved enzyme that will be complementary to the application.

1.6. Project Overview

My research focused on two main projects: the engineering of an NADP\textsuperscript{+}-dependent L-arabinitol 4-dehydrogenase and its application in xylitol biosynthesis from pentose sugars L-arabinose and D-xylose, and engineering a cofactor balanced fungal L-arabinose pathway and analyzing its effects on biofuel production. The first project focuses on the utilization of protein engineering
tools of rational and semi-rational design combined with directed evolution to reverse the cofactor specificity of a single enzyme of a three-enzyme pathway in an attempt to improve flux from L-arabinose substrate to product xylitol. The effect of this cofactor balancing is investigated in a model organism, *E. coli*, and shown to enhance xylitol production from L-arabinose and D-xylose. The engineering of an NADP⁺-dependent xylitol dehydrogenase and introduction of the full L-arabinose assimilation pathway is established, with investigation into the limiting factors of growth complementation in *S. cerevisiae* and *E. coli* when using the cofactor dependent fungal pathway, with insights leading to further projects past the scope of this thesis.

Chapter 2 describes the cloning and characterization of a novel L-arabinitol 4-dehydrogenase (LAD) from *Neurospora crassa*, which is one of the most active and stable LADs reported. Similar to all other characterized LADs, the enzyme is strictly NAD⁺-dependent. The high stability is attributed to the structural zinc ion, and the substrate specificity of the LAD is analyzed further by mutational analysis by comparison of active site residues to homologous xylitol and sorbitol dehydrogenases. The enzyme is highly active and stable, acts on several five carbon sugar alcohol substrates, and operates over a wide pH range, although the activity in acidic pHs drops considerably, which is an issue that is addressed in its usage in the xylitol production strategy in Chapter 4.

Chapter 3 describes the application of rational and semi-rational design followed by directed evolution to reverse the cofactor specificity of L-arabinitol 4-dehydrogenase from NAD⁺ to NADP⁺. Using the engineered enzyme along with NADPH-dependent xylose reductase and L-xylulose reductase was hypothesized to alleviate the cofactor imbalance of the pathway such that conversion of L-arabinose and/or D-xylose to xylitol along with a cosubstrate would produce
more desired product xylitol than the wild-type pathway. By using bioinformatics tools for BLAST search and sequence alignment, along with homology modeling of the enzyme of interest, key residues involved in the cofactor specificity were identified and subjected to site-directed mutagenesis and saturation mutagenesis and screened for improved NADP\(^+\) activity. When rational design sites were exhausted, the enzyme kinetics were further improved by implementation of a directed evolution strategy following the validation of a proper screening procedure. This resulted in the first known NADP\(^+\)-dependent LAD enzyme.

Chapter 4 describes the application of expressing xylose reductase, wild-type or engineered L-arabinitol 4-dehydrogenase, and L-xylulose reductase for xylitol biosynthesis from L-arabinose and D-xylose. The effects of pH-control as well as substrate sugar ratios are also addressed.

Chapter 5 describes the cloning, characterization of xylitol dehydrogenase (XDH) from \textit{N. crassa}, and the engineering of an NADP\(^+\)-dependent XDH by rational design. With the engineered LAD created in Chapter 4, the cofactor imbalance associated with the reductases and dehydrogenases of the initial L-arabinose pathway is relieved, and the subsequent introduction of the pathway into \textit{S. cerevisiae} and \textit{E. coli} paves the way for investigation into the limiting factors of L-arabinose utilization with the fungal-based pathway.
References


(33) Rao, R. S.; Jyothi Ch, P.; Prakasham, R. S.; Sarma, P. N.; Rao, L. V. Xylitol production from corn fiber and sugarcane bagasse hydrolysates by *Candida tropicalis* Bioresour Technol 2006, 97, 1974-8.


(49) Bettiga, M.; Bengtsson, O.; Hahn-Hagerdal, B.; Gorwa-Grauslund, M. F. Arabinose and xylose fermentation by recombinant Saccharomyces cerevisiae expressing a fungal pentose utilization pathway Microb Cell Fact 2009, 8, 40.


CHAPTER 2 – Cloning, Expression, Purification, and Characterization of L-Arabinitol 4-Dehydrogenase from Neurospora crassa

2.1. Introduction

As detailed in Chapter 1, the route through which L-arabinose is converted to xylitol in the fungal pathway passes through three key enzymes: xylose reductase (XR), L-arabinitol 4-dehydrogenase (LAD), and L-xylulose reductase (LXR). There are many examples of XRs that have been characterized from various source organisms, including a highly active XR characterized by Ryan Woodyer and Michael Simurdiak in Professor Zhao’s laboratory from Neurospora crassa, a filamentous fungus known to ferment xylose\(^1\). With the first enzyme of the process already available, the focus turned to the remaining two enzymes that would complete the process. However, there are relatively few cases of the other two enzymes being isolated and identified. The cloning and characterization of the LXR from N. crassa was done by Nikhil Nair in the Zhao laboratory. However, this enzyme did not possess high affinity towards substrate L-xylulose, with an unusually high \(K_{m,L-xylulose}\) value of > 275 mM. It was also recently challenged that the homologous LXR1 from Hypocrea jecorina was instead a D-mannitol 2-dehydrogenase (MDH) based on phylogenetic analysis and deletion experiments showing deletion of \(lxr1\) did not affect growth on L-arabinose, whereas MDH activity levels dropped\(^2\). Due to the inconclusive identity of N. crassa LXR and less than desirable kinetic parameters, the reported LXR from Aspergillus niger was selected for the xylitol production process due to its higher affinity towards L-xylulose (\(K_{m,L-xylulose}\) of 17 mM)\(^3\).

The final enzyme remaining, L-arabinitol 4-dehydrogenase (LAD, EC 1.1.1.12), is commonly found in yeasts and filamentous fungi and catalyzes the second step of the fungal L-arabinose
metabolic pathway by oxidizing L-arabinol to L-xylulose with concomitant NAD⁺ reduction. LAD is purportedly a fungal orthologue of the eukaryotic sorbitol dehydrogenase (SDH)⁴ and belongs to the family of zinc-containing alcohol dehydrogenases. Several LADs have successfully been cloned and expressed³,⁵,⁶. However, they are not optimal for in vivo enzymatic production of xylitol due to their poor stability and/or activity. This chapter will focus on the cloning, expression, purification, and characterization of a novel LAD from N. crassa with high activity and stability for use in xylitol production from L-arabinose and D-xylose. This enzyme serves as the template for engineering altered coenzyme specificity that will be detailed further in Chapter 3.

2.2. Results

2.2.1. N. crassa LAD Gene Identification

LADs from Hypocrea jecorina (GenBank accession number AF355628.1) and Aspergillus oryzae (AB116938.2) were used as templates for a protein BLAST search (www.ncbi.nlm.nih.gov). Utilizing the whole-genome sequence of N. crassa⁷, a postulated LAD-encoding gene, hypothetical protein NCU00643.1 (EAA36547.1), was discovered which had the greatest sequence identity (~80%). This protein (referred to as N. crassa LAD hereafter) had significant homology (72 to 80% identity) with other LADs (Fig. 2.1). Among the conserved residues were those that formed the active site and the structural Zn²⁺-binding site⁸ and the glycine fingerprint⁹ found in polyol dehydrogenases, as well as the majority of those shown to bind substrate in the SDH homologues⁴. 
Neurospora crassa MAASS---------------------------KT1N1GVTPNPHDLIWASEPSLSVQKEE
Hypocrea jecorina HPTVSSLQ1GIDRVAIEPNILNAEPFLGNYKEVEFLSTPFVPGGL
Aspergillus niger AGLGILTLSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL
Aspergillus oryzae AGLGILITLASSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL
Aspergillus fumigatus AGLGILITLASSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL

Neurospora crassa LKEGEVTVAVRSTGLGSVDHFHKBCIGPMVECDHVLGHSEAEVIS
Hypocrea jecorina LKPGEVTIAVRSTGLGSVDHFHKBCIGPMVECDHVLGHSEAEVIS
Aspergillus niger LQGEVTIEVRSTGLGSVDHFHKBCIGPMVECDHVLGHSEAEVIS
Aspergillus oryzae LKPGEVTQVRSTGLGSVDHFHKBCIGPMVECDHVLGHSEAEVIS
Aspergillus fumigatus LKPGEVTQVRSTGLGSVDHFHKBCIGPMVECDHVLGHSEAEVIS

*: ****: **: ::::: *: : :

Neurospora crassa HPSVIKKVDRAVIEPNILNAEPFLGNYKEVEFLSTPFVPGGL
Hypocrea jecorina HPTVSSLQ1GIDRVAIEPNILNAEPFLGNYKEVEFLSTPFVPGGL
Aspergillus niger AGLGILTLSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL
Aspergillus oryzae AGLGILITLASSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL
Aspergillus fumigatus AGLGILITLASSARAGAAGSPITAVIDGDEGLFAKSLVPDVTYKQFLNL

*: ****: **: ::::: *: : :

Neurospora crassa RRYNHVAPWCHIKIGHNSYEGALLPLSVVALAGQRKAKVQGDLPCVTG
Hypocrea jecorina RRYNHVAPWCHIKIGHNSYEGALLPLSVVALAGQRKAKVQGDLPCVTG
Aspergillus niger RRYNHVAPWCHIKIGHNSYEGALLPLSVVALAGQRKAKVQGDLPCVTG
Aspergillus oryzae RRYNHVAPWCHIKIGHNSYEGALLPLSVVALAGQRKAKVQGDLPCVTG
Aspergillus fumigatus RRYNHVAPWCHIKIGHNSYEGALLPLSVVALAGQRKAKVQGDLPCVTG

*: ****: **: ::::: *: : :

Neurospora crassa AIEAKKIVESPFO----GIEPAVALECTGVESS1AAIAAYVFKPGK
Hypocrea jecorina AIEAKKIVESPFO----GIEPAVALECTGVESS1AAIAAYVFKPGK
Aspergillus niger AIEQAGNN1NIFNDGSGPGLAIRPRIMGCTAESVASAIAVFKPGK
Aspergillus oryzae AIEQAGNN1NIFNDGSGPGLAIRPRIMGCTAESVASAIAVFKPGK
Aspergillus fumigatus AIEQAGNN1NIFNDGSGPGLAIRPRIMGCTAESVASAIAVFKPGK

*: ****: **: ::::: *: : :

Neurospora crassa VFIYGVWKNEIQIPFRASVREVLQFQYRCYNTWPRAIRLVEVLDT
Hypocrea jecorina VFIYGVWKNEIQIPFRASVREVLQFQYRCYNTWPRAIRLVEVLDT
Aspergillus niger VFIYGVWKNEIQIPFRASVREVLQFQYRCYNTWPRAIRLVEVLDT
Aspergillus oryzae VFIYGVWKNEIQIPFRASVREVLQFQYRCYNTWPRAIRLVEVLDT
Aspergillus fumigatus VFIYGVWKNEIQIPFRASVREVLQFQYRCYNTWPRAIRLVEVLDT

*: ****: **: ::::: *: : :

Neurospora crassa RLVTWRFPLLEDALKAFETADPKTGAIMVQSL-----------------
Hypocrea jecorina KPTVHFPLEDALKAFETADPKTGAIMVQSL-----------------
Aspergillus niger KLVTWRFPLLEDALKAFETADPKTGAIMVQSL-----------------
Aspergillus oryzae KLVTWRFPLLEDALKAFETADPKTGAIMVQSL-----------------
Aspergillus fumigatus RLVTWRFPLLEDALKAFETADPKTGAIMVQSL-----------------

*: ****: **: ::::: *: : :

Figure 2.1. Protein sequence alignment of N. crassa LAD with four other closely related LAD sequences from filamentous fungi and yeast. Residues highlighted in gray represent the four conserved residues that make up the catalytic zinc binding tetrad. Residues highlighted in red represent the four conserved cysteine residues that make up the proposed structural zinc binding tetrad.
2.2.2. Cloning, Expression, and Purification of *N. crassa* LAD

RT-PCR performed on total RNA isolated from L-arabinose-induced *N. crassa* 10333 showed the expected size of gene product (~1.1 kb). The RT-PCR product was cloned into the pGEX-4T-3 vector using *EcoRI* and *NotI* restriction sites and was transformed into *E. coli* BL21 (DE3). This construct (pGEX-lad1) expressed *N. crassa* LAD as an N-terminal GST-tagged fusion with a thrombin cleavage site. Cell lysates of IPTG-induced cultures of these cells were prepared, analyzed by SDS-PAGE, and assayed for LAD activities. The construct produced soluble GST-tagged *N. crassa* LAD at ~16% of the total soluble cellular proteins (Fig. 2.2), which was then purified in a single step with a GST-Bind kit according to manufacturer’s protocol. The purified protein was desalted by ultrafiltration with several washes of 50 mM morpholinepropanesulfonic (MOPS) buffer (pH 7.25). After digesting with biotinylated thrombin, the enzyme was incubated with streptavidin agarose to remove the thrombin and then passed through GST-Bind resin again to remove the cleaved GST-tag. GST-tagged LAD cleaved with thrombin was used for characterization purposes, as it had about 65% greater specific activity than the tagged LAD enzyme. LAD stocks were stored frozen with 10% (v/v) glycerol at −80 °C. Protein concentrations were determined by the Bradford method\(^\text{10}\) by using an estimated extinction coefficient (San Diego Supercomputer Center Biology Workbench [http://workbench.sdsc.edu]) of 35.3 mM\(^{-1}\) cm\(^{-1}\) at 280 nm with similar results. The purity of the protein was analyzed by an SDS-PAGE gel stained with Coomassie brilliant blue (Fig. 2.2). The final yield of protein was 30 mg/L of culture (~9 mg/g of *E. coli*) of greater than 95% pure LAD with a molecular mass of ~39 kDa, consistent with the predicted value of 39.6 kDa.
Figure 2.2. Overexpression and purification of recombinant *N. crassa* LAD. Lane 1, the molecular weight marker proteins (size in kDa is shown); lane 2, cell-free crude extract; lane 3, purified LAD enzyme with N-terminal GST-tag; lane 4, purified LAD enzyme with GST-tag removed by thrombin cleavage.
2.2.3. Steady-state Kinetics

Purified *N. crassa* LAD displayed activity with NAD$^+$ as the preferred cofactor (Table 2.1), although there was small yet detectable activity with NADP$^+$, which was verified later by high performance liquid chromatography (HPLC) as shown in Figure 2.5. This is the first reported detection of NADP$^+$ utilization by an LAD, although it is still considered a strongly NAD$^+$-dependent enzyme.

Table 2.2 displays the kinetic constants of several other sugar substrates accepted by *N. crassa* LAD. D-Arabinitol, adonitol, xylitol, D-sorbitol, and D-mannitol were all examined as alternative substrates for *N. crassa* LAD with NAD$^+$ as the cofactor. Of those, only the five carbon sugar alcohols acted as substrates, with $K_m$ values of 80 mM and 280 mM for adonitol and xylitol, respectively. This pattern of substrate promiscuity is similar to those of LADs isolated from other sources$^{5,6}$. Kinetic parameters of characterized LADs from *Hypocrea jecorina*, *Aspergillus niger*, and *Aspergillus oryzae* are displayed in Table 2.3.$^{3-6}$.
TABLE 2.1. Kinetic parameters for *N. crassa* LAD<sup>a</sup>.

<table>
<thead>
<tr>
<th><em>N. crassa</em> LAD with indicated coenzyme</th>
<th>$K_m$ for NAD(P) (mean ± SD) (µM)</th>
<th>$k_{cat}$ (mean ± SD) (min&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>$k_{cat}/K_m$ for NAD(P) (µM&lt;sup&gt;-1&lt;/sup&gt; min&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>$K_m$ for L-arabinitol (mean ± SD) (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAD</td>
<td>174 ± 24</td>
<td>1,206 ± 54</td>
<td>6.9</td>
<td>16 ± 3</td>
</tr>
<tr>
<td>NADP</td>
<td>-</td>
<td>-</td>
<td>4.3 x 10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> All assays were performed at 25 °C in 50 mM Tris, pH 8.0

TABLE 2.2. Kinetic parameters for *N. crassa* LAD with other substrates<sup>a</sup>.

<table>
<thead>
<tr>
<th><em>N. crassa</em> LAD with indicated substrate</th>
<th>$k_{cat}$ (mean ± SD) (min&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>$K_m$ (mean ± SD) (mM)</th>
<th>$k_{cat}/K_m$ (mM&lt;sup&gt;-1&lt;/sup&gt; min&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>% Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Arabinol</td>
<td>1,210 ± 30</td>
<td>18 ± 2</td>
<td>67</td>
<td>100</td>
</tr>
<tr>
<td>Xylitol</td>
<td>970 ± 40</td>
<td>290 ± 27</td>
<td>3.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Adonitol</td>
<td>1,080 ± 30</td>
<td>35 ± 3</td>
<td>31</td>
<td>46</td>
</tr>
<tr>
<td>D-Arabinol</td>
<td>-</td>
<td>-</td>
<td>ND&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0</td>
</tr>
<tr>
<td>D-Sorbitol</td>
<td>-</td>
<td>-</td>
<td>ND&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>D-Mannitol</td>
<td>-</td>
<td>-</td>
<td>ND</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> All assays were performed at 25°C in 50 mM Tris, pH 8.0, at saturated NAD<sup>+</sup> concentration

<sup>b</sup> ND, not detected

<sup>c</sup> trace activity at 2 M D-sorbitol concentration, possibly due to substrate contamination
### TABLE 2.3. Kinetic parameters of LAD from various source organisms.

<table>
<thead>
<tr>
<th>Organism (reference)</th>
<th>Specific activity (U/mg)</th>
<th>$k_{\text{cat}}$ (min$^{-1}$)</th>
<th>$K_m$, L-arabinitol (mM)</th>
<th>$k_{\text{cat}}/K_m$, L-arabinitol (mM$^{-1}$ min$^{-1}$)</th>
<th>$K_m$, NAD$^+$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. crassa</em> (this work)</td>
<td>31</td>
<td>1,206</td>
<td>16</td>
<td>75</td>
<td>174</td>
</tr>
<tr>
<td><em>H. jecorina</em></td>
<td>1.6</td>
<td>N/A$^a$</td>
<td>40</td>
<td>N/A</td>
<td>180</td>
</tr>
<tr>
<td><em>H. jecorina</em></td>
<td>0.013</td>
<td>51</td>
<td>4.5</td>
<td>11</td>
<td>N/A</td>
</tr>
<tr>
<td><em>A. niger</em></td>
<td>96</td>
<td>N/A</td>
<td>89</td>
<td>N/A</td>
<td>50</td>
</tr>
<tr>
<td><em>A. oryzae</em></td>
<td>0.04</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

$^a$N/A, not determined

### TABLE 2.4. Kinetic parameters of F59 mutants$^d$.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Enzyme</th>
<th>$k_{\text{cat}}$ (mean ± SD) (min$^{-1}$)</th>
<th>$K_m$ (mean ± SD) (mM)</th>
<th>$k_{\text{cat}}/K_m$ (mM$^{-1}$ min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Arabinol</td>
<td>WT</td>
<td>1,210 ± 30</td>
<td>18 ± 2</td>
<td>67</td>
</tr>
<tr>
<td>F59Y</td>
<td>840 ± 30</td>
<td>42 ± 5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>F59S</td>
<td>60 ± 3</td>
<td>62 ± 9</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td>F59A</td>
<td>&gt; 400</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xylitol</td>
<td>WT</td>
<td>970 ± 40</td>
<td>290 ± 27</td>
<td>3.3</td>
</tr>
<tr>
<td>F59Y</td>
<td>&gt; 880</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F59S</td>
<td>&gt; 1,400</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F59A</td>
<td>&gt; 1,850</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adonitol</td>
<td>WT</td>
<td>1,080 ± 30</td>
<td>35 ± 3</td>
<td>31</td>
</tr>
<tr>
<td>F59Y</td>
<td>1,420 ± 50</td>
<td>193 ± 11</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>F59S</td>
<td>120 ± 5</td>
<td>430 ± 48</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>F59A</td>
<td>&gt; 1,110</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^d$All assays were performed at 25 °C in 50 mM Tris, pH 8.0.

$^b$Saturation of substrate was not reached.
2.2.4. Temperature and pH Dependence

The data show the optimum temperature to be between 45 and 55 °C (Fig. 2.3A). At higher temperatures, the enzyme inactivates rapidly, and at lowered temperatures, the rate increases with temperature according to the Arrhenius equation. Utilizing the Arrhenius equation to fit the data from 12 to 30 °C, the energy of activation for L-arabinitol oxidation by *N. crassa* LAD was determined to be 47 kJ/mol. The stability for *N. crassa* LAD was relatively high, as it retained activity at room temperature for longer than one month and at 4 °C for several months. Figure 2.3B shows the results of thermal inactivation of *N. crassa* LAD at 50 °C, which followed a first-order exponential decay with a half-life of 36 min. Interestingly, when tested at a slightly lower temperature of 45 °C, the enzyme was able to retain ~70% of its activity after 3 hr. The high stability can potentially be attributed to the conserved structural zinc binding residues. A similar thermostabilizing mechanism was also found in alcohol dehydrogenase\textsuperscript{11} and engineered xylitol dehydrogenase\textsuperscript{12}.

The pH range for *N. crassa* LAD activity was large, with >25% of the activity occurring with pH values of 7.0 to 10.5 (Fig. 2.3C). The pH optimum was around pH 9.5, and >60% of the activity remained in the 2 pH-unit span from 8.0 to 10.0.
Figure 2.3. A) $k_{\text{cat}}$ dependence on temperature. *N. crassa* LAD was assayed at different temperatures from 12 to 65 °C at saturating concentrations of 200 mM L-arabinitol and 2 mM NAD$^+$. B) Thermal inactivation of LAD at 50 °C. The heat inactivation at 50 °C was irreversible and followed first-order kinetics with a half-life of 45 min. C) pH rate profile. Saturating concentrations of 200 mM L-arabinitol and 2 mM NAD$^+$ were used to measure the activity in a universal buffer at various pH values from 7.0 to 11.0.
2.2.5. Determination of Protein Mass and Quaternary Structure

Based on the standardized retention times of a Bio-Rad molecular mass standard, the molecular mass of the native *N. crassa* LAD was calculated from its retention time to be ~129 kDa (Fig. 2.4). Monomerization was induced in the presence of 15% SDS, causing *N. crassa* LAD to elute as a single peak with a retention time corresponding to a molecular mass of ~39 kDa. The data suggest that the native LAD is a noncovalently linked tetramer, which is typical for fungal derived zinc-containing alcohol dehydrogenases\textsuperscript{13}.

![Figure 2.4. HPLC size exclusion chromatography. A size exclusion standard was used to calibrate a Bio-Sil SEC-250, 300 x 7.8 mm column with a mobile phase of 0.1 M Na\textsubscript{2}HPO\textsubscript{4}, 0.15 M NaCl, 0.01 M NaN\textsubscript{3}, pH 6.8 at a flow rate of 1 ml/min. The standard proteins are represented by closed squares. LAD samples (represented by open squares) with and without 15% SDS were injected separately and fitted to the standard curve.](image-url)
2.2.6. Cofactor Specificity

The cofactor specificity of *N. crassa* LAD was examined by HPLC. The separation of NAD⁺, NADP⁺, NADH, and NADPH (Figs. 2.5A and 2.5B) was carried out as previously described. No discernible cross-contamination of oxidized cofactors was observed (data not shown). 20 µL reaction mixtures consisting of equal parts of 1 mM NAD(P)⁺ and 25 mM L-arabinitol in 50 mM Tris-HCl (pH 8.0) were set up, and following addition of approximately 1 µg of enzyme, the reaction was allowed to proceed for 20 min at 37 °C. When NAD⁺ was used as the cofactor, the products were analyzed by HPLC, and a single peak (UV 340 nm) was observed that had the same retention time as authentic NADH (Fig. 2.5C). The same process was carried out for NADP⁺ as the cofactor, and a small yet detectable peak was observed with a retention time corresponding to an authentic sample of NADPH (Fig. 2.5D). This indicated the strong preference for NAD⁺ as the cofactor of *N. crassa* LAD.
2.2.7. Homology Modeling

The structural model of *N. crassa* LAD was verified for consistency with known protein folds and allowed $\phi$ and $\psi$ angles. The Profiles3-D check resulted in a self-compatibility score of 139.94, which compares well to the scores of 150.53 and 145.49 for the coordinates from 1PL8 and 1E3J, respectively. The ProStat check of $\phi$ and $\psi$ angles were determined to be 81.2% within their core expected values, comparing well to the 83.3 and 82.4% for the same analysis of PDB structures 1PL8 and 1E3J, respectively. This model was very similar to the human SDH crystal structure in overall fold and binding of coenzyme. The only major deviation between the backbone of these two structures is the N-terminal region of amino acids 1 through 8. However,
this may be due to the different conformations of the N-terminus between being in solution and forming dimerization contacts found to be present in SDH. The conserved catalytic zinc binding residues C53, H78, E79, and E163 have similar orientations and locations in the *N. crassa* LAD model (Fig. 2.6). When comparing proposed substrate binding residues from SDH to *N. crassa* LAD, the majority—S55, F127, T130, E163, R308, Y309—are strictly conserved and configured in similar orientations. However, one substrate-binding residue, F59, was different from the homologous tyrosine residue in SDH, making the *N. crassa* LAD-binding pocket slightly more hydrophobic (Fig 2.6B).

Figure 2.6. A) Homology model of *N. crassa* LAD with bound NAD$^+$ and catalytic zinc ion, built using the Insight II and MOE programs. The catalytic zinc ion (Zn$^{2+}$), four catalytic zinc binding residues, and NAD$^+$ cofactor are colored by atom type. B) Active site of *N. crassa* LAD with docked catalytic zinc (shown in space-fill), cofactor NAD$^+$, and substrate L-arabinitol (colored green) coordinated with the active zinc. Key residues located within the substrate binding pocket are labeled.
2.2.8. F59 Mutational Analysis

Enzyme activity assays were carried out for three mutants of *N. crassa* LAD (F59A, F59S, and F59Y) to study the effect of mutation of this active site, putative substrate-binding residue homologous to tyrosine in other SDHs. All assays were carried out similar to substrate specificity profile for the wild-type enzyme (see “Steady-state kinetics” in “Materials and Methods”), with the cofactor NAD$^+$ held at saturating concentration of 2 mM for all assays. The mutants were still found to have activity with L-arabinitol, xylitol, and adonitol, and their kinetic parameters are displayed in Table 2.4. D-Sorbitol was also tested but showed no significant activity over the wild-type *N. crassa* LAD.

2.3. Discussion

Hypothetical protein NCU00643.1 (EAA36547.1) from *N. crassa* was found to encode an LAD of 363 amino acid residues with a calculated Mr of 39,245. Sequence alignment with other reported LADs shows high homology (~70–80%), with conserved regions for Zn$^{2+}$-binding, cofactor binding, and active site residues. Comparison with several mammalian SDHs (mouse, rat, bovine, sheep, and human) showed ~40% sequence identity, whereas comparison with xylitol dehydrogenases (*Hypocrea jecorina*, *Aspergillus oryzae*, *Candida tropicalis*, *Pichia stipitis*, and *Aspergillus fumigatus* Af293) was ~30% sequence identity.

Kinetic parameters of characterized LADs from *H. jecorina*, *Aspergillus niger*, and *A. oryzae* are displayed in Table 2.3. It should be noted that only the LAD enzyme from *H. jecorina* by Pail et al.$^4$ was purified to homogeneity, whereas the others were characterized either as partially purified enzymes$^{3,5}$ or as cell-free extracts.$^6$ The $K_m$ value of *N. crassa* LAD was 16 mM for L-arabinitol, which when compared to LADs from *H. jecorina* and *A. niger*, is one of the lower
values reported of characterized LADs. The values of 45 mM$^5$ and 4.5 mM$^4$ were reported for the same *H. jecorina* LAD, although the enzymes were purified from 15 different heterologous hosts (*S. cerevisiae* and *E. coli*, respectively), with the former LAD being obtained by cleavage of a GST-fusion protein. In addition, it is difficult to determine the effect of the larger $K_{m,NAD^+}$ of *N. crassa* LAD compared to other LADs, as no other characterizations reported both $k_{cat}$ and $K_m$ values for the cofactor. With a specific activity of the purified *N. crassa* LAD equal to about 31 U/mg, it is almost 20-fold greater than *H. jecorina* LAD purified from *S. cerevisiae* heterologous expression and orders of magnitude higher than other LADs except for *A. niger* LAD, which shows about three-fold greater specific activity than that of *N. crassa* LAD. However, it was also reported that the purified *A. niger* LAD was highly unstable, with rapidly diminishing activity at 4 °C and complete loss of activity after freeze-thawing of the enzyme. In contrast, *N. crassa* LAD is quite stable and does not markedly lose activity when frozen repeatedly.

There has been no in-depth study of the substrate binding residues for LAD, but the enzyme has been postulated to be a fungal orthologue of the eukaryotic D-SDHs$^4$, which have been investigated more thoroughly. Based on these reports, the active site substrate-binding residues are all strictly conserved in all LADs characterized to date (see Figure 2.6). When comparing these residues in *N. crassa* LAD to SDHs, only F59 was not conserved, which instead was a tyrosine residue in all of the SDHs examined. Mutational studies of this position were examined for F59A, F59S, and F59Y, to determine what effects this residue has on substrate specificity alteration, and are shown in Table 2.4. Replacement of the native F59 residue with the homologous tyrosine found in SDH decreased the catalytic efficiency toward all active substrates. The ability of LAD to bind each substrate markedly decreased as the size of the amino acid at position 59 was decreased. Although these results suggest this residue is important
for binding and catalysis for the active substrates of the wild-type LAD, it did not confer the enzyme the ability to accept D-sorbitol as a substrate. This confirms previously reported hypotheses that the amino acids flanking the active site cleft may be responsible for the activity and affinity patterns between LAD and SDH\textsuperscript{4}. The results were also recently confirmed in mutational analysis studies of \textit{A. niger} LAD\textsuperscript{15}. An important finding of this study was the discovery of mutation F59Y, which abolished xylitol activity but did not eliminate L-arabinitol activity, although the \( k_{\text{cat}} \) was lowered by \( \sim 33\% \). It is possible this mutation could be necessary for the production of xylitol from L-arabinose from the combination of XR, LAD, and LXR, as LAD could continue to convert xylitol to D-xylulose. However, as will be discussed in Chapter 4, the low LAD activity with xylitol does not prevent xylitol from being detected from the three enzyme pathway.

\textbf{2.4. Conclusions and Future Work}

In conclusion, a gene from \textit{N. crassa} encoding an LAD was cloned and purified in high yields. The enzyme is highly active and stable, acts on several five carbon sugar alcohol substrates, and operates over a wide pH range. This enzyme should prove to be a useful starting point for the ultimate goal of the production of xylitol and ethanol from L-arabinose derived from renewable resources. However, the newly characterized \textit{N. crassa} LAD still has a strict preference for NAD\(^{+}\) over NADP\(^{+}\) as the native coenzyme. In order to alleviate some of the cofactor regeneration issues that were mentioned in Chapter 1, engineering of the cofactor specificity of \textit{N. crassa} LAD will need to be implemented to generate a successful strategy towards \textit{in vitro} xylitol production and possibly \textit{in vivo} ethanol fermentation. The next chapter will focus on efforts to engineer the crucial cofactor-binding residues to introduce NADP\(^{+}\) activity into \textit{N. crassa} LAD.
2.5. Materials and Methods

2.5.1. Materials

The *Neurospora crassa* genomic sequence and LAD protein sequences were accessed from the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). *Neurospora crassa* 10333 was obtained from the American Type Culture Collection (ATCC). *Escherichia coli* BL21(DE3), GST-Bind resin, biotinylated thrombin, and streptavidin agarose were purchased from Novagen (Madison, WI). *E. coli* WM1788 was kindly provided by Professor William Metcalf at the University of Illinois (Urbana, IL). GST-fusion expression vector pGEX-4T-3 was purchased from Amersham Biosciences. SuperScript™ One-Step RT-PCR with Platinum® *Taq* kit was obtained from Invitrogen (Carlsbad, CA). Shrimp alkaline phosphatase, and PCR grade dNTPs were obtained from Roche Applied Sciences (Indianapolis, IN). Phusion High-Fidelity DNA Polymerase and DNA-modifying enzymes DNase I, *Nde*I, *Sac*I, and T4 DNA ligase and their appropriate buffers were purchased from New England Biolabs (NEB) (Beverly, MA). L-arabinitol, D-arabinitol, adonitol, xylitol, D-sorbitol, D-mannitol, ampicillin, kanamycin, isopropyl β-D-thiogalactopyranoside (IPTG), NADH, NADP⁺, and NADPH were purchased from Sigma (St. Louis, MO). NAD⁺ was a gift from Jülich Fine Chemicals (Jülich, Germany). Other required salts and reagents were purchased from Fisher (Pittsburgh, PA) or Sigma-Aldrich. The QIAprep spin plasmid mini-prep kit, QIAquick gel purification kit, RNeasy midiprep kit, and QIAquick PCR purification kit were purchased from Qiagen (Valencia, CA). Various oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). SDS-PAGE gel materials, electrophoresis equipment, protein size markers, size exclusion standards (catalog number 151-1901) and Bio-Sil SEC-250, 300 x7.8 mm column were purchased from Bio-Rad (Hercules, CA).
**2.5.2. RT-PCR and Cloning (Notebook #1, p. 42; Notebook #2, p. 51)**

*N. crassa* 10333 was grown on rich potato media at 30 °C for 24 h and induced with 150 mM L-arabinose for 2 hr. Because the predicted gene contained one intron, reverse transcription PCR (RT-PCR) was utilized to isolate the processed gene. Total ribonucleic acid (RNA) was purified from collected cells (RNeasy purification kit, Qiagen) and treated with DNase I to remove residual genomic DNA. RT-PCR was performed using SuperScript™ One-Step RT-PCR with Platinum® *Taq* (Invitrogen) following the manufacturer’s guidelines. A control reaction consisted of the same protocol except for the SuperScript™ enzyme mix was heated to 95 °C for ten minutes to thermally inactivate the reverse transcriptase enzyme and the reverse transcription thermocycler step was omitted. The primers used for the RT-PCR were: forward 5′-GTA GCT ACG TCA GAA TTC CAT GGC TTC TAG CGC TTC C-3′ and reverse 5′-GCT GAT TCT GCG GCC GCT TAC TCC AGA CTC TGG ATC-3′. The forward primer contained an *EcoRI* restriction site (shown in bold), while the reverse primer contained a *NotI* restriction site (shown in bold) and stop codon (italicized). The resulting RT-PCR product was isolated by a QIAquick agarose gel purification kit and amplified by an additional 20 cycles of PCR. The product was digested with *EcoRI* and *NotI* restriction enzymes and purified again by agarose gel electrophoresis. It was then ligated into pGEX-4T-3 which had been previously prepared by *EcoRI* and *NotI* digestion, dephosphorylation by shrimp alkaline phosphatase, and gel purification. The ligation mixture was precipitated with *n*-butanol and resuspended in water. The new construct was used to transform *E. coli* WM1788 by electroporation. Positive clones were selected on Luria–Bertani (LB) solid media with ampicillin at 37 °C overnight. All colonies were then removed from the plates and grown to saturation in 5 mL liquid LB from which the plasmids were purified using a QIAprep spin plasmid miniprep kit, which were used to transform
*E. coli* BL21(DE3) by heat shock. Positive clones were selected on LB solid media with ampicillin, picked individually, and assayed for LAD activity by the cell lysate assay described below. Plasmids were sequenced using the BigDye® Terminator sequencing method and an ABI PRISM 3700 sequencer (Applied Biosystems, Foster City, CA). Sequencing results determined at the Biotechnology Center of the University of Illinois showed that the product had four silent mutations compared with the predicted sequence from the NCBI database.

### 2.5.3. Lysate Assay

*E. coli* BL21 (DE3) harboring the pGEX-4T-3-derived vector was grown to a maximum optical density at 600 nm (OD$_{600}$) at 37 °C with shaking at 250 rpm. Fifty microliters was used to inoculate a new culture, which was grown at 37 °C with shaking at 250 rpm until an OD$_{600}$ of ~0.6 was reached. The cultures were then induced with 0.3 mM IPTG and shaken at 250 rpm at 25°C for 4 h. One milliliter of cells was harvested by centrifugation and lysed by resuspension in 1 mL of 1 mg/mL lysozyme/50 mM Tris-HCl (pH 8.0). The cells were frozen at −80 °C and thawed at room temperature. The resulting lysate was vortexed thoroughly and centrifuged to remove cell debris. Ten microliters of the lysate was used in a kinetic assay described below in the Enzyme Kinetics Section with 200 mM L-arabinitol and 2 mM NAD$^+$ as the substrates in 50 mM Tris (pH 8.0). To determine soluble and insoluble expression, lysozyme was utilized as the lysis reagent for the induced and normalized cells following the manufacturer’s recommendations, and samples were subsequently analyzed by SDS-PAGE.
2.5.4. GST-tag Removal (Notebook #2, p. 67)

The N-terminal GST-tag was removed by incubation with biotinylated thrombin overnight at 4 °C, incubation with streptavidin agarose for 30 min at 4 °C to remove thrombin, and passing mixture through GST-Bind resin to remove GST-tag, leaving five residues (GlySerProAsnSer) attached to the N terminus of the N. crassa LAD sequence. To determine the effect of removal of the GST-tag, the purified LAD was incubated with and without thrombin at 4 °C overnight. Complete cleavage of the 25.7 kDa tag was verified by SDS-PAGE. The specific activities of the cleaved and noncleaved samples were compared. It was determined that removal of the GST-tag enhanced activity by about 65%. Because of this significant difference in activity, the cleaved enzyme was used in all subsequent assays.

2.5.5. Enzyme Kinetics (Notebook #2, pp. 79-102)

Initial rates were determined by measuring the initial rate of reaction as observed by the increase of absorbance at 340 nm using a Varian Cary 100 Bio UV-visible spectrophotometer (Varian, Palo Alto, CA) at 25 °C in 50 mM Tris-HCl (pH 8.0). N. crassa LAD kinetic measurements with substrate L-arabinitol and cofactor NAD⁺ were taken in a five-by-five matrix format, with substrate and cofactor concentrations varied from below their $K_m$ to tenfold higher than their $K_m$. The kinetic data were analyzed with a modified version of Cleland’s program\(^\text{16}\). $V_{\text{max}}$ and $K_m$ for both L-arabinitol and NAD⁺ were obtained by fitting the data to a sequential ordered mechanism with NAD⁺ binding first, based on the proposed mechanism for sorbitol dehydrogenase\(^\text{17}\):

$$v = \frac{V_{\text{max}}AB}{(K_{ia}K_B + K_A B + K_B A + AB)}$$  \hspace{1cm} (Eq. 1)
where \( v \) is the initial velocity, \( V_{\text{max}} \) is the maximum velocity, \( K_A \) and \( K_B \) are the Michaelis-Menten constants for NAD\(^+\) and L-arabinitol, respectively, \( A \) and \( B \) are the concentrations of NAD\(^+\) and L-arabinitol, respectively, and \( K_{\text{ia}} \) is the dissociation constant for NAD\(^+\).

*N. crassa* LAD kinetic measurements with alternate substrates were taken with varied concentrations of substrate sugar alcohol while keep NAD\(^+\) concentration saturating at 2 mM. The data were used to calculate the kinetic constants by fitting the Michaelis-Menten equation using Origin 5.0. *N. crassa* LAD displayed typical Michaelis-Menten type kinetics with respect to all substrates tested except D-sorbitol. The data represent averages of assays performed in duplicate or triplicate on two separate occasions.

**2.5.6. Thermal Dependence (Notebook #2, pp. 79-102)**

The optimal temperature of activity was determined by assaying *N. crassa* LAD activities at temperatures ranging from 12 to 65 °C in saturating substrate conditions. Thermal inactivation of *N. crassa* LAD was studied by incubating at varied elevated temperatures in 50 mM Tris-HCl (pH 8.0), with samples removed at various times and assayed for activity in saturating substrate conditions.

**2.5.7. pH Rate Profile (Notebook #2, pp. 79-102)**

Activity was measured at pH values between 7.0 and 11.0 under saturating concentrations of NAD\(^+\) (2 mM) and L-arabinitol (200 mM) in a universal buffer (50 mM morpholineethanesulfonic acid (MES) / 50 mM Tris / 50 mM glycine). Ionic strength of the universal buffer was taken into consideration according to previously established buffer compositions.\(^{18}\)
2.5.8. Determination of Protein Mass and Quaternary Structure (Notebook #2, p. 202)

The quaternary structure of native *N. crassa* LAD was determined using an Agilent 1100 series HPLC system with a Bio-Sil SEC-250 column (300 by 7.8 mm) and a mobile phase of 0.1 M Na$_2$HPO$_4$, 0.15 M NaCl, and 0.01 M NaN$_3$, pH 6.8. Monomerization was induced in the presence of 15% SDS and loaded onto the HPLC in similar fashion to the native enzyme to determine the *N. crassa* LAD monomer size.

2.5.9. HPLC Analysis (Notebook #1, p. 186)

The enzyme (1 µg) was incubated in a mixture of buffer and 25 mM L-arabinitol with 1 mM NAD(P)$^+$ at 37 °C for 20 min. The sample was eluted on a Zorbax 150 mm x 3.0 mm C-18 (3.5 µm) column with a UV detector (Agilent 1100 series). The eluent consisted of two elements: 0.1 M KH$_2$PO$_4$ containing 5 mM tetrabutylammonium hydrogen sulfate (pH 5.5) (buffer A) and 100% methanol (buffer B). The most suitable gradient was an initial isocratic step for 6 min at 93% buffer A, a gradient for 5 min from 7 to 30% buffer B, and a final isocratic step for 5 min at 30% buffer B.

2.5.10. Homology Modeling

Insight II was used to prepare the model (Insight II, version 2000; Accelrys Inc., San Diego, CA) and MOE (Chemical Computing Group Inc., Montreal, Canada) was used for optimization. To verify the model, the overall fold was checked using Profiles3-D (Insight II), and the allowed states for $\phi$ and $\psi$ angles and bond distances were checked using ProStat (Insight II), both with default settings. The Profiles3-D (Insight II, default parameters) check resulted in a self-compatibility score of 139.94, which compares well to the scores of 150.53 and 145.49 for the coordinates from 1PL8 and 1E3J, respectively. The ProStat check of $\phi$ and $\psi$ angles were
determined to be 81.2% within their core expected values, comparing well to the 83.3% and 82.4% for the same analysis of PDB structures 1PL8 and 1E3J, respectively.
References


(15) Rutten, L.; Ribot, C.; Trejo-Aguilar, B.; Wosten, H. A.; de Vries, R. P. A single amino acid change (Y318F) in the L-arabitol dehydrogenase (LadA) from Aspergillus niger results in a significant increase in affinity for D-sorbitol BMC Microbiol 2009, 9, 166.


CHAPTER 3 – Engineering an NADP⁺-Dependent L-Arabinitol 4-Dehydrogenase for Xylitol Production

3.1. Introduction

The goal of my research project is to introduce the first three enzymes of L-arabinose metabolism into a model organism, *Escherichia coli*, and address the bottlenecks in the conversion of L-arabinose and D-xylose to xylitol, with one of the main issues being the nicotinamide cofactor imbalance involved. Xylose reductase (XR) and L-xylulose reductase (LXR) are preferably NADPH-dependent, whereas L-arabinitol 4-dehydrogenase (LAD) is NAD⁺-dependent. This results in the formation of byproduct nicotinamide cofactors NADP⁺ and NADH, which must be regenerated by an organism’s metabolism (Figure 3.1A). In an attempt to reduce the burden placed on the cell to regenerate these accumulating cofactors, the use of an NADP⁺-dependent LAD would alleviate the imbalance partially (Figure 3.1B). The resultant pathway would then require one equivalent of NADP⁺ to be regenerated, which has been shown to be possible in xylitol production from D-xylose alone in *E. coli*.\textsuperscript{1,2} However, there has been no report to date of a naturally occurring NADP⁺-dependent LAD. It was then proposed to attempt to engineer the NAD⁺-dependent ncLAD (described in Chapter 2) into an NADP⁺-dependent ncLAD. With the newly engineered enzyme, it was hypothesized that the partial alleviation of the cofactor imbalance could potentially lead to improved xylitol production from D-xylose and L-arabinose in a microbial host.
Figure 3.1. Xylitol production scheme using D-xylose and/or L-arabinose as starting substrate. (A) Wild-type pathway shows inherent difficulties associated with redox cofactor imbalance. (B) Engineered pathway is designed with an NADP⁺-dependent LAD (in red) to regenerate NADPH and eliminate the need for NAD⁺ regeneration.
There exists a large body of literature describing the alteration of nicotinamide cofactor specificity\textsuperscript{3-10}, including two review articles outlining the typical determinants and evolution of nicotinamide binding sites\textsuperscript{11,12}. Yet despite the prevalence of these attempts, altering cofactor specificity remains a challenge, as very few examples exist where catalytic efficiency for the initially disfavored cofactor has been improved such that it reaches the same level of efficiency as with the physiological cofactor. Comparison of the strategies required to achieve efficient use of the non-physiological cofactor in these enzymes indicates that there is no clear formula for success, as mutations found in some enzymes do not necessarily translate to the same activity effect in another template enzyme. As shown in Figure 3.2, NAD(H) and NADP(H) differ only in the phosphate group esterified at the 2’-position of adenosine ribose, and therefore there are a limited number of amino acid residues interacting with this characteristic moiety that are suitable as the first candidates for protein engineering using site-directed mutagenesis. The ability of polyol dehydrogenases to discriminate between NAD\textsuperscript{+} and NADP\textsuperscript{+} has been established to lie in the amino acid sequence of the \( \beta\)-\( \alpha\)-\( \beta\) motif of the coenzyme binding domain. The primary determinant of NAD\textsuperscript{+} specificity is the presence of an aspartate residue, which forms double hydrogen bonds to both the 2’- and 3’-hydroxyl groups in the ribosyl moiety of NAD\textsuperscript{+} and induces negative electrostatic potential to the binding site. Commonly, this residue in NADP\textsuperscript{+}-dependent dehydrogenases is replaced by a smaller and uncharged residue such as Gly, Ala, and Ser, accompanied by the concurrent presence of an arginine residue that forms a positive binding pocket for the 2’-phosphate group of NADP\textsuperscript{+}. Despite the prevalence of this information, the full reversal of coenzyme specificity, in terms of having a mutant enzyme catalytically efficient as the wild-type, has rarely been achieved.
Figure 3.2. Structural comparison of nicotinamide cofactors. NAD\(^+\) (top) and NADP\(^+\) (bottom) are shown colored by atom. Although these two cofactors tend to adapt slightly different bound conformations, the only structural difference is the 2’-phosphate group found on the adenine ribose of NADP\(^+\).

Sorbitol dehydrogenases (SDHs) are typically NAD\(^+\)-dependent, but there is one reported case\(^{13}\) from silverleaf whitefly *Bemisia argentifolii* that is the only known enzyme showing a strict preference to NADP\(^+\), and a crystal structure of the enzyme is available (Protein Data Bank [PDB] ID:1E3J)\(^{14}\). Again, the homologous residues of the motif binding the signature phosphate moiety of NADP\(^+\) has the characteristics of a smaller residue (serine) and positively charged...
residue (arginine) in the cofactor binding pocket. *N. crassa* LAD (ncLAD) shares homology with NADP⁺-dependent *B. argentifolii* SDH (43% identity, 61% positive), so it was expected that similar mutations would result in reversal of cofactor specificity.

Xylitol dehydrogenases are also strongly NAD⁺-dependent as well, but there have been several reports of successful reversal of cofactor specificity towards NADP⁺. Watanabe *et al.* described the engineering of NAD⁺-dependent XDH from yeast *Pichia stipitis* (psXDH), which resulted in completely reversing the coenzyme specificity to generate a novel NADP⁺-dependent XDH mutant by multiple site-directed mutagenesis⁴. However, the mutations produced decreased thermostability compared with the wild-type enzyme, so the authors focused on a structural zinc atom that is not contained in native PsXDH. Introduction of specific cysteine ligands into the enzyme gave an additional zinc-binding site and improved thermostability. Furthermore, the catalytic efficiency of NADP⁺-dependent PsXDH was also improved by the introduction of an additional zinc atom. Their best NADP⁺-dependent mutant XDH had three mutations (D207A/I208R/F209S) in the cofactor binding pocket near the phosphate moiety of NADP⁺, with three additional mutations (S96C/S99C/Y102C) used to introduce a second structural zinc later to stabilize their mutant enzyme. The ncLAD enzyme already possesses the structural zinc motif and has been shown to contain two zinc ions per monomer (catalytic zinc and structural zinc), so it was promising that reversal of cofactor specificity in ncLAD could be attained without loss of stability.

A double mutant (D38S/M39R) of the NAD⁺-dependent short-chain xylitol dehydrogenase from *Gluconobacter oxydans* (goXDH) was constructed that was able to use NADP⁺ exclusively. In fact, the affinity for NADP⁺ was higher (*Kₘ = 206 \( \mu \)M) than the wild-type enzyme for NAD⁺ (*Kₘ = 348 \( \mu \)M), with only a slight drop in *kₐₚ*. While the goXDH is a short-chain xylitol
dehydrogenase with a different fold and mechanism for catalysis than medium-chain sugar alcohol dehydrogenases (including ncLAD), the mutations conferring cofactor specificity are of the same type, with amino acids aspartate and methionine mutated to a smaller residue serine and positively-charged arginine, respectively.

To the best of our knowledge, there has been no report of a naturally-occurring or engineered NADP⁺-dependent LAD enzyme in literature. Therefore, initial attempts at reversing the cofactor specificity of *N. crassa* LAD were focused on determining the important residues involved in differentiating between NAD⁺ and NADP⁺ based on results from homologous enzymes sorbitol dehydrogenase (XDH) and xylitol dehydrogenase (SDH). This work describes the successful engineering of a novel NADP⁺-dependent LAD enzyme using both rational design and directed evolution.

3.2. **Results**

3.2.1. **Reversal of Cofactor Specificity of *N. crassa* LAD by Rational Design**

*N. crassa* LAD (ncLAD) shares homology with *Pichia stipitis* XDH (35% identity, 52% positive), and it was hypothesized that mutating the homologous residues to those that were reported to be important for cofactor specificity alteration from NAD⁺ to NADP⁺ would result in an efficient way to generate an NADP⁺-dependent ncLAD enzyme for later use in xylitol production or other value-added products from L-arabinose. Therefore, a triple mutant ncLAD (D211A/I212R/D213S) was constructed and tested for cofactor activity. However, there was no detectable activity with either NAD⁺ or NADP⁺ in the cell lysate activity assay, and later it was determined that the mutant was not found to be expressed in the soluble fraction following SDS-PAGE analysis (data not shown). A more systematic approach was deemed necessary to analyze
each residue in order to gain insight into which mutations would be important in reversing the cofactor specificity. Single mutants of LAD were generated as described in the Materials and Methods section, and consisted of mutating D211 to smaller amino acid residues (Ala, Gly, Ser, Thr, or Val), I212 to a basic residue (Lys or Arg), and D213 to serine. All of these were expressed in *E. coli* and SDS-PAGE analysis was performed to determine the solubility of these mutant LADs. However, the solubility of these mutants was variable, with D211S, I212R, I212K and D213S being the most soluble, followed by D211A and D211G at about half of the wild-type enzyme level. The solubilities of D211T and D211V were dramatically reduced compared to that of the wild type, although purification of these LADs from cell lysate did result in a small amount of recovered enzyme (~0.1-0.2 mg /g cells). All mutant enzymes were purified to homogeneity and then analyzed by HPLC to determine the effects of the mutations, and three single mutants D211A, D211G, and D211S all showed improvements in NADP⁺ activity.

Based on the single mutant studies, engineering of ncLAD for reversal of cofactor specificity from NAD⁺ to NADP⁺ was continued, starting from the single mutants D211A, D211G, and D211S. Position 212 in engineered NADP⁺-dependent xylitol dehydrogenase is an arginine residue, and has been shown to be arginine in both sorbitol dehydrogenase¹⁴, engineered phosphite dehydrogenase⁹, and engineered formate dehydrogenase¹⁵ to confer NADP⁺ specificity/activity. The double mutants ncLAD-AR, -GR, and -SR were created, expressed, and purified. HPLC analysis showed further improved activity towards NADP⁺, with the largest gains in activity being with ncLAD-AR and ncLAD-SR.

The last rational design attempt at incorporating homologous residues from other NADP⁺-dependent enzymes was D213S. Introduction of this mutation in the background of ncLAD-AR
was attempted to begin with, which resulted in an enzyme that was not expressed in the soluble fraction, so the mutation was placed into the remaining two double mutants, ncLAD-SR and ncLAD-GR, and also resulted in significantly decreased expression. Although no increased activity was found, the decrease in activity suggested this position was important for binding NADP⁺.

Saturation mutagenesis libraries at residue D213X on both the ncLAD-AR and ncLAD-SR templates were created and screened for activity as detailed in the Materials and Methods Section. Positive mutants with >25% activity than the parent from the ncLAD-SR screen were found, with none from the ncLAD-AR parent library, and were selected and verified by tube assay to have improved activity. All clones verified were sequenced and found to have an asparagine residue in position 213. This triple mutant was named ncLAD-SRN. As shown in Table 3.1, ncLAD-SRN displayed almost 1.4-fold higher $k_{cat}$ than parent ncLAD-SR, but suffered a 3-fold loss in affinity as the $K_{m,NADP}$ rose from 0.48 mM for ncLAD-SR to 1.45 mM for ncLAD-SRN. The overall catalytic efficiency decreased 2-fold. Unfortunately, the ncLAD-SRN mutant was also significantly less stable than the ncLAD-SR parent, with high amount of precipitated enzyme occurring during the buffer exchange and concentration steps of the purification protocol (Materials and Methods Section). This could explain why the measured catalytic efficiency of ncLAD-SRN was lower than ncLAD-SR, because the screen should have resulted in improvement of this property. Despite having generated a successful LAD mutant (ncLAD-SR) with specificity now towards NADP⁺ and minimal activity towards NAD⁺, the low affinity for NADP⁺ was still of major concern. The intracellular concentrations of NADP⁺ in *E. coli* range from 0.1 mM to 0.4 mM, based on converting reported values with estimated intracellular *E. coli* cell volumes. It would be desired to have a mutant LAD with $K_m$ values
at or below these levels for optimal performance. Saturation mutagenesis libraries of residues surrounding the rational design targets (210X, 214X, 215X) were also generated and screened, without any positive hits found. Therefore attention was shifted to attempt directed evolution on ncLAD-SR template.

3.2.2. Screening Procedure Validation and EP-PCR Results

Before attempting to screen a mutant library, particularly a larger-size library generated from error-prone PCR (EP-PCR), it is particularly important to verify the reproducibility of the screening procedure\textsuperscript{21,22}. Variability inherent in every screen arises from various factors such as varying cell growth rates, expression of enzyme from multiple-copy vectors, cell lysis efficiency, pipetting errors, etc. These variations must be minimized in order to eliminate false positives and simplify data interpretation.

The production of NAD(P)H at 340 nm is commonly used to measure the activity of dehydrogenases. Therefore, a 96-well plate screening assay was developed for high-throughput analysis of a library of error-prone PCR mutants of ncLAD-SR. \textit{E. coli} BL21(DE3) cells harboring plasmid pET-28a ncLAD-SR were used to inoculate 96-well plates and following growth, induction, and lysing steps, the reproducibility of NADP\textsuperscript{+} activity measurement was tested by calculating the coefficient of variance (CV).\textsuperscript{21} The assay was confirmed to be sufficient with a coefficient of variance (CV) of approximately 15%.

After the screening procedure was validated, the optimization of the diversity of the error-prone library was addressed. The diversity of a mutant library is dependent on the level of mutagenesis, which can be modified by varying the concentration of Mn\textsuperscript{2+} in the PCR reaction\textsuperscript{21}. Three different libraries of ncLAD-SR, created using 0.10, 0.15, and 0.20 mM Mn\textsuperscript{2+} were tested
with the validated screening procedure. Based on these results, an optimal concentration of 0.15 mM Mn$^{2+}$ resulted in 40-50% of the clones being inactive, which correlates to approximately 1-2 amino acid substitutions per gene. ncLAD has 364 amino acids, so a library of this size can have 7,280 single amino acid substitutions. However, due to codon degeneracy and that in most cases only a single nucleotide within a single codon is altered, only about 6 amino acids can be accessed by EP-PCR. This results in a library of about 2,200 potentially different clones. A library size of 3-5 times this size must be screened in order to have a 95% confidence level of screening all possible variants, which can easily be screen with the 96-well plate assay. Therefore, after screening a little more than 6,000 clones, a final mutant LAD, ncLAD-3x, was discovered and characterized.

### 3.2.3. Kinetic Data of the Engineered ncLAD Mutants

The kinetic parameters of the parent wild-type and engineered mutant ncLAD enzymes towards substrates NAD$^+$, NADP$^+$, and L-arabinitol are listed in Table 3.1. The best first round rational design generated mutant, D211S, showed a dramatic decrease in activity towards NAD$^+$, with minimal yet detectable activity increase towards NADP$^+$. The best second round rational design mutant, D211S/I212R, displayed a significant reversal in cofactor specificity, although still had ~5-fold lower $k_{cat}/K_m$ than the wild-type had with NAD$^+$. With rational design targets exhausted, a round of EP-PCR and screening via 96-well plate format resulted in the discovery of ncLAD-3x, which contained a third mutation S348T. In addition, activity with NAD$^+$ was also kept low, only slightly above background activity. This result was promising, as the goal of the engineering effort was to not only improve activity towards NADP$^+$, but to also minimize activity with NAD$^+$, such that the cofactor imbalance in the wild-type xylitol pathway could be handled by inclusion of a suitable co-substrate to regenerate the excess NADP$^+$. However, as
shown in Table 3.2, an unfortunate artifact of the screening procedure was the subsequent loss of affinity for substrate L-arabinitol of the mutant ncLAD-3x ($K_{m,L}$-arabinitol of ~290 mM) compared to the wild-type ($K_{m,L}$-arabinitol of ~12 mM). As a result, the catalytic efficiency is also substantially reduced from 64 mM$^{-1}$ min$^{-1}$ for the wild-type to 4 mM$^{-1}$ min$^{-1}$ ncLAD-3x. This loss of affinity and catalytic efficiency came about despite keeping the L-arabinitol concentration of the screening assay at 20 mM, which was near the wild-type $K_{m,L}$-arabinitol value. Intermediate mutants ncLAD-S and ncLAD-SR could not be saturated for sufficient kinetic parameter analysis.
Table 3.1. Kinetic parameters of ncLAD mutants for nicotinamide cofactors.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$</th>
<th>$K_m$</th>
<th>$k_{cat}/K_m$</th>
<th>$k_{cat}$</th>
<th>$K_m$</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min$^{-1}$</td>
<td>mM</td>
<td>mM$^{-1}$min$^{-1}$</td>
<td>min$^{-1}$</td>
<td>mM</td>
<td>mM$^{-1}$min$^{-1}$</td>
</tr>
<tr>
<td>ncLAD-wt</td>
<td>650</td>
<td>0.14</td>
<td>4740</td>
<td>-</td>
<td>&gt;8</td>
<td>&lt;30</td>
</tr>
<tr>
<td></td>
<td>± 25</td>
<td>± 0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ncLAD-S</td>
<td>515</td>
<td>2.9</td>
<td>175</td>
<td>-</td>
<td>&gt;5</td>
<td>&lt;40</td>
</tr>
<tr>
<td></td>
<td>± 21</td>
<td>± 0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ncLAD-SR</td>
<td>191</td>
<td>3.6</td>
<td>53</td>
<td>494</td>
<td>0.48</td>
<td>1030</td>
</tr>
<tr>
<td></td>
<td>± 17</td>
<td>± 0.7</td>
<td></td>
<td>± 26</td>
<td>± 0.08</td>
<td></td>
</tr>
<tr>
<td>ncLAD-SRN</td>
<td>-</td>
<td>&gt; 5</td>
<td>&lt;100</td>
<td>714</td>
<td>1.45</td>
<td>496</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>± 8</td>
<td>± 0.05</td>
<td></td>
</tr>
<tr>
<td>ncLAD-3x</td>
<td>-</td>
<td>&gt; 5</td>
<td>&lt;100</td>
<td>1210</td>
<td>0.55</td>
<td>2190</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>± 44</td>
<td>± 0.09</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Not determined, saturation of cofactor not reached
All assays run at 25 °C, 50 mM Tris-HCl pH 8.0
All enzymes were purified and characterized with N-His$_6$Tag

Table 3.2. Kinetic parameters of ncLAD mutants for substrate L-arabinitol.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$k_{cat}$</th>
<th>$K_m$</th>
<th>$k_{cat}/K_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min$^{-1}$</td>
<td>mM</td>
<td>mM$^{-1}$min$^{-1}$</td>
</tr>
<tr>
<td>ncLAD-wt</td>
<td>736 ± 110</td>
<td>11.6 ± 0.9</td>
<td>64</td>
</tr>
<tr>
<td>ncLAD-S</td>
<td>- $^a$</td>
<td>&gt; 1000</td>
<td>- $^a$</td>
</tr>
<tr>
<td>ncLAD-SR</td>
<td>- $^a$</td>
<td>&gt; 500</td>
<td>- $^a$</td>
</tr>
<tr>
<td>ncLAD-3x</td>
<td>1190 ± 390</td>
<td>292 ± 110</td>
<td>4</td>
</tr>
</tbody>
</table>

$^a$Not determined, saturation of cofactor not reached
All assays run at 25 °C, 50 mM Tris-HCl pH 8.0
All enzymes were purified and characterized with N-His$_6$Tag
3.3. Discussion

3.3.1. Protein Engineering Strategy

Initial attempts to alter the cofactor specificity of *N. crassa* LAD were conducted in similar fashion to the successful cofactor reversal of xylitol dehydrogenase\(^4\). However, when introducing similar mutations and generating a triple mutant (D211A/I212R/D213S), a cell lysate assay resulted in no significant activity increase from the background in converting NADP\(^+\). One possible reason behind this result was that the mutations introduced were not suitable for cofactor specificity change in the context of *N. crassa* LAD. Another possibility was that the cell lysate assay was not sensitive enough to detect the subtle increase in activity that could result from mutagenesis, as the wild type enzyme did not show any detectable cell lysate activity on its own towards NADP\(^+\). SDS-PAGE analysis also showed that the triple mutant was not expressed in the soluble fraction (data not shown), which could have also been a reason for no increased activity being detected. A more systematic approach was then taken (Figure 3.3) to engineer an NADP\(^+\)-dependent LAD enzyme.
Figure 3.3. Lineage and protein level mutations of the engineered LAD variants. Mutant enzymes outlined were purified and kinetics characterized in Table 3.1.
Another attempt was to develop a screening method to determine beneficial mutations incorporated by site-directed saturation mutagenesis of the proposed important residues that conferred cofactor specificity. However, the lack of sensitivity in the high-throughput 96-well plate assay resulted in no positive mutants being identified with increased affinity for NADP$^+$. A third attempt at improving NADP$^+$ affinity was based on HPLC analysis in which the generation of NADPH could be detected with much higher sensitivity when compared to the 96-well plate or test tube based activity assays. As mentioned in Chapter 2, HPLC analysis was able to show that the wild type *N. crassa* LAD did indeed possess activity with NADP$^+$, although with an exceedingly low $k_{cat}/K_m$ value of $\sim 6.5 \times 10^{-7} \mu M^{-1} min^{-1}$. Single mutants of LAD were generated as described in the Materials and Methods section, and consisted of mutating D211 to smaller amino acid residues (Ala, Gly, Ser, Thr, or Val), I212 to a basic residue (Lys or Arg), and D213 to serine. The best mutants (D211A, D211G, and D211S) were then subjected to site-directed mutagenesis for introduction of arginine at position 212, with improvements found in ncLAD-AR and ncLAD-SR. Targeting position 213 with replacement with serine did not improve activity and resulted in insoluble protein. However, the double mutants provided a reproducible positive control in a 96-well plate formatted screening procedure, and saturation mutagenesis at position 213 was performed. The best mutant found had a D213N mutation, but this enzyme was found to be less stable than previous rounds, with noticeable amounts of white precipitate forming during purification and concentration, leading to kinetic parameter analysis that did not match the screening results. Therefore, the best double mutant, ncLAD-SR, was used in an error-prone PCR library screen in an attempt to enhance NADP$^+$-activity and improve solubility and stability. The final mutant, ncLAD-3x, contained the two rational design mutations (D211A/I212R) and a third mutation S348T. It should be noted that a second round of EP-PCR
was attempted with ncLAD-3x as the parent, but no improvement in NADP⁺-dependent activity was able to be found.

3.3.2. Properties of the ncLAD-3x Mutant

This work resulted in the engineering of an NADP⁺-dependent LAD, ncLAD-3x. From the cofactor kinetics, the $k_{cat}$ value of 1210 min⁻¹ was close to 2-fold higher than the wild type value of 650 min⁻¹. And while the $K_{m,NADP⁺}$ value of 0.55 mM was slightly higher than the template ncLAD-SR $K_{m,NADP⁺}$ value of 0.48 mM, the catalytic efficiency for NADP⁺ experienced an almost 2-fold increase from 1030 to 2190 mM⁻¹ min⁻¹. The decrease in affinity for substrate L-arabinitol, going from $K_{m,arabinitol}$ of 20 mM for the wild-type to $K_{m,arabinitol}$ of 200 mM for ncLAD-3x, was an unfortunate artifact of the screen. Although the concentration of L-arabinitol was kept at a low level (25 mM) in the screening procedure, the improvement in the NADP⁺-activity apparently masked this effect.

3.3.3. Structural Analysis of the Mutations

There is currently not a crystal structure of L-arabinitol 4-dehydrogenase available from any source organism. Therefore, a homology model of wild-type ncLAD was built using Insight II software (Accelrys, San Diego, CA) during the engineering efforts. The NAD⁺-dependent human sorbitol dehydrogenase (SDH, PDB ID:1PL8) and NADP⁺-dependent SDH from silverleaf whitefly, *Bemisia argentifolii* (PDB ID:1E3J) were used as templates, sharing 44% and 43% homology with ncLAD, respectively. The resulting model was docked with NAD⁺ and the catalytic zinc ion and subjected to energy minimization by using the Molecular Operating Environment (MOE, Chemical Computing Group, Montreal, Canada) program. Based on this structural model, the molecular basis for reversal of cofactor specificity was further probed.
Figure 3.4. Homology model illustration of cofactor binding interactions of ncLAD-wt. Only the adenine portion of NAD(P) is shown for clarity. A) LAD bound with NAD\(^+\) in the active site, showing favorable interaction through hydrogen bonding between D211 and the 2\(^{'}\)-hydroxyl of the adenine ribose. B) With NADP\(^+\) bound, however, D211 and the adenine ribose 2\(^{'}\)-phosphate of NADP\(^+\) are electrostatically repulsed, resulting in poor binding and significantly reduced activity due to the strong steric hindrance.

Firstly, the docked NAD\(^+\) cofactor in the wild-type molecule was changed to NADP\(^+\) using the builder tool in MOE and then subjected to energy minimization. When the phosphate moiety of NADP\(^+\) is present in the wild-type model, there are both unfavorable electrostatic repulsions and steric interactions that prevent proper binding of the NADP\(^+\) cofactor (Figure 3.4). As in most dehydrogenase domains that bind NAD\(^+\), the space that would be taken up by the extra phosphate of NADP\(^+\) is occupied by the conserved hydrogen bond between the aspartate in the binding site (D211 in \textit{N. crassa} LAD) and the adenosine ribose of the NAD\(^+\)\textsuperscript{24}.

During the course of the engineering efforts, the crystal structure of ncLAD-wt was obtained in collaboration with Brian Bae in Dr. Satish Nair's group (manuscript in preparation). Figure 3.5A shows the comparison of the monomeric structure of ncLAD in the homology model versus the
crystal structure. The catalytic domain (left-hand side of the figure) shows slightly more deviation, particularly among the loop structures, between the model and the crystal structure than with the cofactor binding domain (right hand side of the figure). This is to be expected as the template structures used for the modeling were either with NAD$^+$ bound (1PL8) or with no cofactor ligand bound (1E3J), and the enzyme is known to have a hinge region between the two domains that closes when ligand is bound. However, as shown in Figure 3.5B, the residues in the cofactor binding site as well as the docked NAD$^+$ ligand in the homology model were very closely arranged as in the crystal structure, with only a slight turn in the I212 residue being the most noticeable difference. Regardless, the model was deemed suitable enough for the earlier cofactor binding analysis, but the S348 residue was pointed in the opposite direction entirely, so the crystal structure was used to analyze the mutations found in ncLAD-3x.
Figure 3.5. (A) Comparison of the monomeric ncLAD subunit in the homology model (shown in red) and the crystal structure (shown in blue). The structural zinc ion is only shown for the crystal structure as this was not docked into the homology due to the engineering efforts focusing on the cofactor binding site. (B) Cofactor binding domain of homology model (shown as lines) and the crystal structure (shown as sticks).
The three mutations in ncLAD-3x were introduced into the crystal structure using MOE and energy minimized, and the engineered cofactor binding site is shown in Figure 3.6. The D211A substitution served to open up the cofactor binding pocket for the phosphate moiety to enter, as well as remove the negatively charged electrostatic repulsion of aspartate with phosphate. The second rational design target residue, I212R, introduced a positively charged residue to balance the negative charge of the phosphate moiety of NADP⁺. The final mutation found from directed evolution screening, S348T, is located between two alpha helices approximately 6.9 angstroms from the phosphate moiety of the docked NADP⁺ molecule, and does not appear to form any direct contact with the cofactor. The side chain atoms are actually positioned away from the phosphate moiety. However, it is possible this mutation alters the positioning of surrounding helix structures to enable more space for the extra phosphate moiety to position itself in the cofactor binding pocket, along with making the pocket too large for the wild-type’s preferential cofactor, NAD⁺, to be held securely. It should be noted that the improvement of ncLAD-3x over the double mutant was mostly attributed to an increase in \( k_{cat} \). The mutation S348T also is located on the opposite end of the alpha helix in close proximity to the catalytic zinc. Therefore, it is entirely possible that the S348T mutation served to alter the secondary structure of the enzyme such that the catalytic site was modified to improve the turnover number and/or reduce the affinity for substrate L-arabinitol. Saturation mutagenesis and subsequent screening of the library did not result in finding a mutant with improved NADP⁺ activity, so no efforts were made to generate and characterize other mutants at this site to determine the true nature of the mutation’s effect.
3.4. Conclusions

The cofactor specificity of L-arabinitol 4-dehydrogenase from *Neurospora crassa* was successfully reversed from NAD$^+$ to NADP$^+$ preference. Two mutations, D211S and I212R, are similar to those found in other NADP$^+$-dependent dehydrogenase enzymes, and a third mutation, S348T, is the first of its position to be described to influence LAD activity. Modeling the third mutation showed the difficulty in choosing this position to mutate rationally, showing that directed evolution can aid in cases where rational design targets have been exhausted. The
combined rational design and directed evolution approaches resulted in the isolation of ncLAD-3x with 2-fold improvement in $k_{\text{cat}}$, and a catalytic efficiency with NADP$^+$ close to the wild-type with NAD$^+$. With a novel NADP$^+$-dependent LAD developed, attention can be shifted towards designing and implementing an engineered fungal pathway for production of xylitol from hemicellulosic sugars, which will be addressed in Chapter 4.

3.5. Materials and Methods

3.5.1. Materials

*Escherichia coli* BL21 (DE3) and pET-28a were purchased from Novagen (Madison, WI). Ampicillin, kanamycin, isopropyl-β-D-thiogalactopyranoside (IPTG), NAD, NADP, NADH, and NADPH were obtained from Sigma-Aldrich (St. Louis, MO). Other salts and reagents were purchased from either Fisher (Pittsburg, PA) or Sigma-Aldrich. Co$^{2+}$ Talon™ immobilized metal affinity resin was purchased from Clontech BD biosciences (San Jose, CA). Amicon® Ultra-15 filter devices were purchased from Millipore. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). Kits for plasmid purification, gel and column purification of DNA fragments were obtained from Qiagen (Valencia, CA). DNA oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA).

3.5.2. Site-Directed and Saturation Mutagenesis

A megaprimer PCR method for site-directed mutagenesis$^{25}$ was utilized to introduce site-specific mutations using purified plasmid pET-28a ncLAD-wt as the template. Two oligonucleotide primers flanking the LAD gene were used in combination with the mutagenic primers listed in Table 3.3. For the construction of each mutant, the first step was the creation of the megaprimer piece, using the forward mutagenic primer and reverse flanking primer. After PCR purification,
a second PCR reaction was run, this time with the forward flanking primer and the megaprimer piece from the previous reaction utilized as primers for amplification. The correctly sized gene products (~1.1 kb) were purified from the agarose gel after DNA electrophoresis and another round of amplification was run with the flanking primers to increase product yield when necessary. Products of the correct size were isolated by gel electrophoresis, digested with NdeI and SacI, and ligated into the NdeI-SacI digested, dephosphorylated pET-28a(+) vector as N-terminal His6-Tag fusions. Chemically competent E. coli BL21(DE3) were then transformed with the ligation mixtures by heat shock treatment and grown on agar plates containing 50 µg/mL kanamycin. Several colonies were picked and clones were first analyzed by colony PCR to ensure the gene insert was incorporated into the pET-28a vector. Cultures of the clones with correct sized inserts were grown again and the subsequently isolated plasmids were sequenced in both directions at the Biotechnology Center of the University of Illinois using the Big Dye™ Terminator sequencing method and an ABI PRISM® 3700 sequencer (Applied Biosystems, Foster City, CA). Sequencing of the mutant genes was conducted to confirm correct mutations were introduced and to eliminate the chance of PCR-introduced random mutations being incorporated into the final DNA construct. Colonies with plasmids containing the correct mutant genes were used for protein expression and purification.
Table 3.3. Primers sequences used in site-directed mutagenesis and saturation mutagenesis of ncLAD. Restriction sites are shown in bold. Codons underlined encode the mutations incorporated.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>DNA Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fwd ncLAD NdeI</td>
<td>5’-GTA GCT ACG TCA CAT <strong>ATG</strong> GCT TCT AGC GCT TCC AAG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211A</td>
<td>5’-CTT GTC ATT ACC <strong>GCC</strong> ATT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211G</td>
<td>5’-CTT GTC ATT ACC <strong>GGN</strong> ATT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211S</td>
<td>5’-CTT GTC ATT ACC <strong>TCN</strong> ATT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211T</td>
<td>5’-CTT GTC ATT ACC <strong>ACN</strong> ATT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211V</td>
<td>5’-CTT GTC ATT ACC <strong>GTN</strong> ATT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD I212R</td>
<td>5’-CTT GTC ATT ACC <strong>CGT</strong> GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D213S</td>
<td>5’-CTT GTC ATT ACC <strong>TCC</strong> GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211A/I212R</td>
<td>5’-CTT GTC ATT ACC <strong>GCC</strong> CGT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211G/I212R</td>
<td>5’-CTT GTC ATT ACC <strong>GNN</strong> CGT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211S/I212R</td>
<td>5’-CTT GTC ATT ACC <strong>TCN</strong> CGT GAC GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211A/I212R/D213S</td>
<td>5’-CTT GTC ATT ACC <strong>GCC</strong> CGT TCN GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211G/I212R/D213S</td>
<td>5’-CTT GTC ATT ACC <strong>GNN</strong> CGT TCN GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211S/I212R/D213S</td>
<td>5’-CTT GTC ATT ACC <strong>TCN</strong> CGT TCN GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Fwd ncLAD D211S/I212R/D213X</td>
<td>5’-CTT GTC ATT ACC <strong>TCN</strong> CGT NNS GAA GGC CGC TTG-3’</td>
</tr>
<tr>
<td>Rev ncLAD SacI</td>
<td>5’-AGC TGA TAG CGA <strong>GCT</strong> CTT ACT CCA GAC TCT GGA TC-3’</td>
</tr>
</tbody>
</table>

3.5.3. Error-Prone PCR Mutagenesis and Library Creation (Notebook #3, p. 130)

Random mutagenesis was carried out using error-prone PCR\(^{26}\). The following stock solutions were prepared for the PCR protocol. A stock solution of 10X Mutagenic Buffer was prepared containing 70 mM MgCl\(_2\), 500 mM KCl, 100 mM Tris-HCl (pH 8.3 @ 25 °C), and 0.1%
(wt/vol) gelatin. A stock solution of 5 mM MnCl\textsubscript{2} was prepared by dilution of a stock of 250 mM MnCl\textsubscript{2}. For a mutagenesis rate of \sim 1-2 amino acid substitutions per gene product, it was established by an initial experiment with varied MnCl\textsubscript{2} concentrations and subsequent sequencing of a number of the resultant clones that a final concentration of 0.15 mM MnCl\textsubscript{2} was to be included in the PCR mixture. A 10X stock solution of unbalanced dNTPs was prepared containing 2 mM dATP, 2 mM dGTP, 10 mM dCTP, and 10 mM dTTP.

In a 0.6 mL thin-wall PCR tube, 10 \mu L of 10X Mutagenic Buffer, 3 \mu L of 5 mM MnCl\textsubscript{2}, 50 ng parent template plasmid DNA, 50 pmole of T7-fwd primer (5’-TAATACGACTCACTATAGGG-3’), 50 pmole of T7-rev primer (5’-GCTAGTTATTGCTCAGCGG-3’), 10 \mu L of 10X unbalanced dNTPs, and an amount of water to bring the mixture volume to 99 \mu L. 1 \mu L of Taq DNA Polymerase (5 U/mL, Roche Applied Science, Indianapolis, IN) was added and mixtures were gently mixed by pipetting. The error-prone PCR reaction was carried out in an MJ Research PTC-200 thermal cycler (Watertown, MA) with the following program:

- Step 1 = 94 °C for 4:00
- Step 2 = 94 °C for 0:45
- Step 3 = 55 °C for 0:45
- Step 4 = 72 °C for 1:45
- Step 5 = Goto 2, 18 times
- Step 6 = 72 °C for 5:00
- Step 7 = 4 °C for ever
- Step 8 = End
PCR products were run on a 1% Tris Acetate EDTA gel and purified by using a QiaQuick® Gel Extraction Kit (Qiagen, Valencia, CA). PCR products were then digested with NeoI and SacI and ligated into pET-28a vector digested with the same two restriction enzymes. Digested pET-28a was also dephosphorylated with alkaline phosphatase (New England Biolabs, Ipswich, MI) prior to ligation to lower background of subsequent transformation. Ligation reactions of 10 µL final volume contained 100 ng vector, 60 ng insert (3:1 insert to vector molar ratio), 1X T4 DNA Ligase Buffer, and 0.5 U T4 DNA Ligase, and were incubated at either room temperature for 1 h or overnight at 16 °C. The ligated library was transformed into freshly prepared electrocomtent E. coli BL21(DE3) cells, which were then plated onto Luria-Bertani (LB) agar plates containing 50 µg/mL kanamycin. Viable colonies each contained one mutant gene, and were used to inoculate media-containing 96-well plates for activity screening, as described below in Section 3.5.4.

3.5.4. Activity Assay Screening Procedure (Notebook #3, p. 136)

To identify LAD mutants that showed improved NADP⁺ activity while keeping NAD⁺ activity limited, a 96-well plate based screening procedure was developed. Library colonies were selected with sterile toothpicks and placed into single wells to inoculate 96-well plates containing 100 µL LB media supplemented with 50 µg/mL kanamycin. The plates were secured in place by placement in plastic pipette tip boxes and incubated at 37 °C for five hours with shaking (250 rpm). The cell cultures were then induced with IPTG (final concentration of 0.3 mM IPTG) followed by incubation at 30 °C for 12-16 hours overnight. Plates were wrapped with wetted towels to prevent too much evaporation from occurring to the outside wells of the 96-well plates. After protein expression, the plates were centrifuged at 4,000 rpm for 15 minutes and the media was decanted. Cell pellets were lysed by resuspension in 100 uL lysis buffer
consisting of 100 mM phosphate buffer (pH 7.0) with 1 mg/mL lysozyme. Adhesive labels were placed on the 96-well plates and vortexed until cell pellets were uniformly resuspended, and then followed by a freeze-thawing step. 96-well plates containing cell lysates were then centrifuged at 4,000 rpm at 4 °C for 15 minutes, and 20 mL of cell lysates was transferred to a fresh 96-well plate. Addition of 100 μL of activity assay mix comprised of 100 mM KH₂PO₄ buffer (pH 7.0), 25 mM L-arabinitol, and 0.125 mM NADP⁺ was immediately followed by monitoring the change in absorbance at 340 nm for 5 minutes in a SpectraMax 340PC microplate reader (Molecular Devices, Sunnyvale, CA) to determine the initial activities of each library mutant. Mutants with increased activity towards NADP⁺ were found by comparison to the parent enzyme, which was inoculated in triplet for each 96-well plate screened. One well was also reserved for the wild-type enzyme to serve as a negative control for the assay. Positive mutants with the improved activity towards NADP⁺ were verified by scale up of the 96-well screening procedure to a 14 mL tube assay variation.

3.5.5. DNA Sequencing and Analysis

Plasmid DNA from *E. coli* BL21(DE3) was isolated using QIAprep® spin Miniprep kit (Qiagen, Valencia, CA). Sequencing reactions contained 50 ng of template plasmid DNA, 10 pmole of primer, 1X BigDye sequencing buffer, and the BigDye reagent (Applied Biosystems, Foster City, CA). Reactions were carried out in an MJ Research PTC-200 thermal cycler (Watertown, MA). Samples were submitted to the Biotechnology Center at the University of Illinois for sequencing on an ABI PRISM® 3700 sequencer (Applied Biosystems, Foster City, CA). The sequencing reaction program was as follows:

Step 1 = 96 °C for 2:00
Step 2 = 96 °C for 0:20
Step 3 = 50 °C for 0:10
Step 4 = 60 °C for 3:50
Step 5 = Goto 2, 30 times
Step 6 = 4 °C for ever
Step 7 = End

3.5.6. LAD Over-Expression and Purification (Notebook #4, p. 230-231)

Protein purification of wild-type or mutant LAD enzymes was carried out in a protocol described elsewhere with some modifications.\(^{27}\) \textit{E. coli} BL21(DE3) pET-28a ncLAD cells were grown in 2 L baffled flasks containing 600 mL Luria-Bertani (LB) medium supplemented with kanamycin (50 µg/mL). Cells were grown at 37 °C to OD\(_{600}\) ~0.6 and then induced with IPTG to a final concentration of 0.3 mM. The culture was incubated at 30 °C for an additional 6 hours, after which the cells were harvested by centrifugation (15 min at 10,000 rpm) and resuspended in 20 mL Buffer A, composed of 20 mM Tris-HCl (pH 7.2 at 4 C), 15% glycerol, and 0.5 M NaCl. Lysozyme (1 mg/mL) was added to the resuspended cells and the sample was incubated at room temperature for 15 minutes and then freeze-thawed. The cells were further lysed by French press, cell debris was removed by centrifugation at 13,000 rpm for 15 minutes (twice), and cell lysate was filtered with 0.45 µm syringe filters prior to affinity column loading. The N-His\(_6\)-tagged enzymes were purified using immobilized metal affinity chromatography (IMAC). TALON\(^{\text{TM}}\) Superflow Co\(^{2+}\) resin (Clontech, Mountain View, CA) was charged and equilibrated with buffer A according to manufacturer’s instructions. The clarified supernatant was loaded onto the resin and washed with Buffer A. Non-specifically binding proteins were removed by washing the column with Buffer A with 10 mM imidazole (Buffer B). His\(_6\)-tagged proteins were eluted from the column with Buffer A with 250 mM imidazole (elution buffer). Proteins were concentrated using a Millipore Amicon Ultra-15 centrifugal filter device (MWCO = 10 kDa) at
4,000 rpm at 4 °C, washed three times with 50 mM 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer (pH 7.2) containing 15% glycerol, and concentrated. Concentrated proteins were stored at -80 °C. Protein concentration was determined using Bio-Rad Protein Assay with BSA as standard.

3.5.7. HPLC Analysis of Products

The mutant enzyme (10 μg) was incubated in a mixture of buffer and 250 mM L-arabinitol with 2.5 mM NAD(P)⁺ at 37 °C for 30 min. The sample was eluted on a Zorbax 150 mm x 3.0 mm C-18 (3.5 μm) column with a UV detector (Agilent 1100 series). The eluent consisted of two elements: 0.1 M KH₂PO₄ containing 5 mM tetrabutylammonium hydrogen sulfate (pH 5.5) (buffer A) and 100% methanol (buffer B). The most suitable gradient was an initial isocratic step for 6 min at 93% buffer A, a gradient for 5 min from 7 to 30% buffer B, and a final isocratic step for 5 min at 30% buffer B.

3.5.8 Enzyme Kinetics (Notebook #4, p.232)

The kinetic rate constants for the purified N-His₆-tagged wild-type and mutant LAD enzymes were determined as described elsewhere²⁸. Briefly, the initial rates of reaction were determined by monitoring the increase in absorbance of NAD(P)H (ε = 6.22 mM⁻¹ cm⁻¹) at 340 nm. Initial rates were measured at 25 °C using a Varian Cary UV 100 Bio UV-visspectrophotometer. Reactions were initiated by addition of ~1-20 μg purified LAD enzyme with N-terminal His₆-tag. Enzyme concentrations were determined with the Bio-Rad Protein Assay (Bio-Rad, Hercules, CA) with bovine serum albumin (Sigma, St. Louis, MO) as standard. For determining kinetics, L-arabinitol or NAD(P)⁺ concentrations were varied from below the K_m value to at least 5-times higher than the K_m value, with saturating concentrations of the other substrate/cofactor present.
The Michaelis-Menten equation was fitted to the data using the non-linear curve fitting with least squares regression analysis in OriginPro 8 (OriginLab Corporation, Northampton, MA) to determine $k_{\text{cat}}$ and $K_m$ values.
References


CHAPTER 4 – Application of an Engineered Fungal Pathway for Xylitol Production from Hemicellulosic Sugars L-Arabinose and D-Xylose

4.1. Introduction

While there are numerous strategies that have been developed to convert D-xylose to xylitol (see Chapter 1), there have been very limited number of reports on converting L-arabinose to xylitol in literature, as the major focus tends to be towards conversion of both D-xylose and L-arabinose to biofuels and other alternative value-added chemicals. To the best of our knowledge, only one report by Sakakibara et al. described the development of a novel pathway for bioconversion of L-arabinose to xylitol. The three-enzyme pathway consisted of three reactions: isomerization of L-arabinose to L-ribulose by L-arabinose isomerase from *E. coli* (AraA), epimerization of L-ribulose to L-xylulose by D-psicose 3-epimerase from *Rhizobium radiobacter* (DPE), and reduction of L-xylulose to xylitol by an NADH-dependent L-xylulose reductase from *Ambrosiozyma monospora* (LXR). When this pathway was expressed under the control of P$_{BAD}$ promoter in an *E. coli* strain deficient in L-arabinose degradation, the xylitol yields were 0.72 g/g initial arabinose, and 0.95 g/g consumed L-arabinose. Low amounts of ribitol (1.0 g/L) and L-arabinitol (0.1 g/L) were also produced from initial concentration of 20.1 g/L L-arabinose. However, in order for the pathway to efficiently produce xylitol, a co-substrate glycerol was necessary to regenerate the NADH used by the LXR enzyme. In addition, the pathway could not incorporate D-xylose conversion to xylitol, although the authors hinted that they were attempting to incorporate a separate pathway involving xylose isomerase (XylA), to convert D-xylose to D-xylulose, followed by xylitol dehydrogenase, to convert D-xylulose to xylitol, while deleting endogenous *xylB* to prevent xylose assimilation in their pathway expressing strain.
In this chapter, we discuss how the application of an NADP⁺-dependent L-arabinitol 4-dehydrogenase (LAD) in the bioconversion of xylose and arabinose to xylitol improves the flux towards desired product over the wild-type pathway using an NAD⁺-dependent LAD. In addition, the effects of altering pentose sugar and co-substrate concentrations and the detrimental effect of acidic pH are discussed. This work represents the first example of an engineered, entirely fungal-derived system for xylitol production from both xylose and arabinose sugars present in hemicellulosic hydrolysates. Further optimization is necessary to improve productivity, and these issues will be addressed.

4.2. Results

4.2.1. Strain Development

In order to effectively compare the wild-type vs. the engineered xylitol pathway, a couple of necessary characteristics of the expression strain were: (1) the ability to uptake pentose sugars as well as co-substrate simultaneously, and (2) the inability to assimilate xylose or arabinose to prevent shuttling pentose substrates to biomass or undesired products.

The co-utilization of both pentose sugars along with co-substrate sugar, in this case D-glucose, to regenerate NADPH cofactor was first addressed. As a result of cyclic AMP receptor protein (CRP)-dependent control of xyl genes, E. coli exhibits diauxic growth characteristics such that glucose is preferentially assimilated before xylose. Catabolite repression refers to the phenomenon in which the transcription of sensitive operons is reduced by certain carbon sources in the medium, most prominently by glucose (‘glucose effect’). E. coli exhibits diauxic growth in sugar mixtures due to CRP-mediated catabolite repression and inducer exclusion related to phosphotransferase system enzyme activity. Replacement of the native crp gene with a
catabolite repression mutant (referred to as crp*) enables co-utilization of glucose and other sugars in *E. coli*⁶⁻⁸. With these desired properties satisfied, the mutant crp* was incorporated into the xylitol expression strain presented in this work.

In xylose metabolism by *E. coli*, the bacterial xylAB operon consists of two genes; xylA (D-xylose isomerase) and xylB (D-xylulokinase). These genes encode for enzymes that convert D-xylose to D-xylulose and D-xylulose to D-xylulose-5-phosphate, respectively. The deletion of the xylA gene was performed using the λ red system⁹, such that the resultant xylitol pathway expression strain could not use D-xylose as a carbon source for growth, leaving any internalized D-xylose substrate available for xylitol production. This strain was constructed by Nikhil Nair using the crp* mutant strain as the parent, and was labeled as HZ1450. The presence of the xylB gene was later determined not be to a major issue, as the deletion of xylA resulted in a polar effect that dramatically lowered XylB activity to levels such that expression of xylA on a multi-copy plasmid could not complement the xylA deletion and restore cell growth on D-xylose (see Chapter 5).

The bacterial araBAD operon consists of three genes; araB (L-ribulokinase), araA (L-arabinose isomerase), and araD (L-ribulose-5-phosphate 4-epimerase). These genes encode for enzymes that convert L-arabinose directly to D-xylulose 5-phosphate through L-ribulose and L-ribulose 5-phosphate.¹⁰ The deletion of the araBAD operon was performed using the λ red system⁹, such that the resultant xylitol pathway expression strain would not be able to use L-arabinose as a carbon source for cell growth, leaving any internalized L-arabinose substrate available solely for xylitol production. The final strain used in the expression of all xylitol constructs (C600 crp* ΔxylA ΔaraBAD) was named HZ1173.
4.2.2. Operon Design

The xylitol pathway that is required to perform this bioconversion is also composed of three enzymes; xylose reductase (XR), L-arabinitol dehydrogenase (LAD), and L-xylulose reductase (LXR). Therefore, it was strategized to design the xylitol operon after the araBAD operon structure, with the same promoter region, intergenic regions between the 1st and 2nd and between the 2nd and 3rd genes, and neglect the terminator region which would already be present in the expression vector chosen as the backbone. However, one difference from the araBAD structure was implemented in the xylitol operon, and that was the order of the genes involved. Typically, the first gene of an operon is translated and expressed the most of all the genes in the operon. There are concerns that ncLAD-wt (or ncLAD-3x for the engineered pathway) is the rate limiting enzymatic step due to the accumulation of L-arabinitol in most L-arabinose utilizing species, which may be due to enzyme expression, kinetics compared to the other pathway enzymes, its pH optimum falling in the basic pH range when intracellular pH is typically neutral to acidic, etc.. Therefore, LAD was placed first in the operon, followed by LXR, and lastly XR, which had the highest activity towards its substrates than the other enzymes.

The final design of the xylitol operon is illustrated in Figure 4.1a, and consisted of the following placed in order; (1) the araBp promoter, (2) ncLAD-wt gene for the wild-type pathway operon or ncLAD-3x gene for the engineered pathway operon, (3) intergenic region 1 (linkBA), (4) the anLXR gene, (5) intergenic region 2 (linkAD), and finally, (6) the XR gene. The operons were digested with NsiI and EcoRI sites that had been designed into the flanking primers, and ligated into pTrc99a digested with the same restriction enzymes (Figure 4.1b). The constructs were verified to be correct by restriction digest and colony PCR amplification of the genes of the
operon, and wild-type and engineered versions containing NAD$^+$-dependent LAD and NADP$^+$-dependent LAD respectively, were transformed into HZ1173.

Figure 4.1. (a) Operon map of the xylitol pathway. (b) Vector map of the expression plasmid containing the xylitol operon. Two versions were created, with ncLAD represented by either the AD$^+$-dependent wild-type *N. crassa* LAD or engineered NADP$^+$-dependent *N. crassa* LAD described in Chapter 3.
The resulting strains were designated as HZE2 (with wild-type xylitol synthesis operon 231-wt) and HZE3 (with engineered xylitol synthesis operon 231-3x). When the strain were grown and induced with 1% L-arabinose for four hours at 30 °C and analyzed by SDS-PAGE, the levels of LAD were highest, followed by LXR, and little detection of XR was visible (Figure 4.2). While the absence of XR expression from the gel staining was alarming at first, the activity check with cell lysates verified that XR activity was present over a negative control lysate from a strain with only empty vector.

Figure 4.2. SDS-PAGE analysis of xylitol synthesis pathway expressing strains. HZE1 = HZ1173 + pTrc99a empty vector. HZE2 = wild-type pathway expression strain. HZE3 = engineered pathway expression strain. Only soluble fractions of cell lysates are shown.
4.2.3. Shake Flask Studies

4.2.3.1. Effect of Using Glucose as Co-Substrate and pH on Xylitol Production

In using glucose as the co-substrate for xylitol production, the glycolysis rate at high sugar concentrations often exceeds respiratory capacity, leading to build-up of intermediate metabolites. In *E. coli*, this results in the accumulation of acetate, leading to an acidic environment that results in decreased growth rates as well as lowered recombinant protein production. Previous work in a similar *crp*\(^*\) strain showed that at 18 g/L glucose concentration, the acetate production is significant, accumulating up to 70 mM.\(^{11}\) When HZE3 was grown in 30 g/L glucose and 10 g/L arabinose, the pH dropped to 5.1 within 24 hours, and growth was inhibited. Further initial studies attempting xylitol production resulted in minimal xylitol production by the wild-type strain (HZE2) and engineered strain (HZE3) (data not shown).

As discussed in Chapter 2, the three enzymes involved in the pathway for conversion of L-arabinose to xylitol have very different pH profiles. This raises the concern that the optimal activity of LAD is around pH 9.0-9.5, with a significant decrease in activity found at pH 6.8 (~20% maximum activity), which is the reported intracellular pH of *E. coli*\(^{12}\). While it was reported that there exists an intracellular pH homeostasis in *E. coli* that serves to keep the intracellular pH at a relatively constant level regardless of external pH\(^{13}\), a shift in so much as half of a pH unit around physiological pH could drastically affect the LAD activity.

Bioreactors are typically pH-controlled in order to allow for high cell densities to be reached. However, pH-control by acid or base addition during shake flask experimentation is not possible aseptically. Therefore, flask studies were then run which incorporated the addition of a buffering component (MOPS) to the media in an attempt to offset the acetate buildup and allow for
improved cell growth and potentially xylitol production. As shown in Figure 4.3, the extracellular pH dropped dramatically in cultures containing 2% or more glucose, while $pH_{\text{ext}}$ remained relatively constant at pH 7.2 for 1% glucose. The final cell densities also increased with increasing buffering capacity. Therefore, 1% glucose was chosen as the amount used in further shake flask studies.

![Figure 4.3. Effect of glucose concentration on extracellular pH and OD$_{600}$ of cultures of HZE3. 3% glucose, squares; 2% glucose, circles; 1% glucose, triangles. $pH_{\text{ext}}$ values shown in filled symbols, OD$_{600}$ shown in empty symbols.](image)

**4.2.3.2. Wild-type vs. Engineered Xylitol Pathway**

Cultures of HZE2 and HZE3 were compared in M9 + tryptone media with 100 mM MOPS added as a buffering component. Glucose was added in excess of 3:1 (w/v). As shown in Figure 4.4, both pathways consumed all glucose and arabinose in 20 hours. However, the wild-type pathway showed xylitol production up to ~10 mM after 33 hours (approximately 50% molar yield from L-arabinose) whereas the engineered pathway showed minimal production with less
than 1 mM (<5% molar yield) xylitol produced after 56 hours. One possible explanation for the slower xylitol production in the engineered pathway is the inherent 10-fold reduced affinity for ncLAD-3x for substrate L-arabinitol (see Chapter 3). The maximum amount of L-arabinitol that can be formed is equal to the starting L-arabinose concentration (~20 mM), which is close to the $K_{m,arabinitol}$ value for ncLAD-wt (16 mM), but much less than ncLAD-3x ($K_{m,arabinitol} = 292$ mM). Unfortunately, the amount of starting arabinose is limited in flask studies due to the fixed 1% co-substrate glucose level restraint in order to keep the media properly buffered and prevent acetate accumulation and resultant pH drop. It was found that adding more arabinose while keeping a constant 1% glucose concentration results in insufficient conversion of arabinose (data not shown). In order to test whether L-arabinose levels were limiting, the pathways were compared in pH-controlled bioreactors.

Figure 4.4. Xylitol production profiles from substrate L-arabinose with glucose co-substrate in shake flasks. (A) Wild-type xylitol pathway and (B) Engineered xylitol pathway.
4.2.4. Bioreactor Studies

4.2.4.1. Wild-Type vs. Engineered Pathway

Figure 4.5 shows the xylitol production profiles from 0.4 L batch controlled bioreactor cultures in M9 media + tryptone supplemented with 20 g/L glucose and 20 g/L L-arabinose using strain HZE2 (Fig. 4.5A) and HZE3 (Fig. 4.5B). Media was maintained at pH 7.0 by automatic addition of 5 N NaOH. After depletion of the initial glucose, additional glucose was added from a sterile 60% stock. Although residual arabinose was still left in the media, the wild-type pathway showed generation of L-arabinitol (up to 25 mM) and xylitol (up to 13 mM) only until ~50 hours and then the conversion is stalled. The engineered pathway containing the NADP⁺-dependent LAD, however, shows slow yet steady conversion of L-arabinose to L-arabinitol (up to 50 mM) and finally to xylitol (up to 30 mM) over the course of the time prior to glucose addition at 96 hours. After the glucose spiking at 96 hours, both strains continued to produce L-arabinitol and xylitol, although the wild-type pathway did so at a modest rate compared to the engineered pathway.
Figure 4.5. Comparison of 0.4 L (working volume) fermentation profiles for (A) HZE2 and (B) HZE3. The initial culture contained 2% glucose and 2% arabinose in minimal media + tryptone. Additional glucose was added at the 96 hour mark. pH was controlled at 7.0.
Figure 4.6 shows the results of comparing the wild-type xylitol pathway vs. engineered xylitol pathway in M9 + tryptone media supplemented with 3% glucose:1% arabinose. The higher ratio of cosubstrate:substrate resulted in full conversion of arabinose after two days, and also resulted in different product profiles than when using an equal weight percent ratio. The wild-type xylitol pathway (Figure 4.6A) showed xylitol production for two days, but then xylitol concentration began to decrease afterwards while the arabinitol concentration continues to increase. The engineered pathway (Figure 4.6B) showed conversion of arabinose to arabinitol within two days, and slowly afterwards the xylitol accumulation began to increase with decreasing arabinitol.

Figure 4.6. Comparison of 0.4 L (working volume) fermentation profiles for (A) HZE2 and (B) HZE3. The initial culture contained 3% glucose and 1% arabinose in minimal media + tryptone. pH was controlled at 7.0.

Further analysis of the samples is necessary to determine whether the arabinitol peak detected, particularly after xylitol production begins, is either the L-arabinitol enantiomer or possibly the D-arabinitol enantiomer. LAD has low activity with xylitol (see Chapter 2) for conversion to D-xylulose, and LXR or another endogenous reductase could conceivably convert D-xylulose to D-arabinitol, although this activity has not been measured with LXR in vitro. In case the xylitol
was converted to D-xylulose and then to D-xylulose-5-phosphate by xylB, which was not deleted in the expression strain, it would be taken up into the pentose phosphate pathway and used for biomass production. However, the OD_{600} of the cultures did not deviate after glucose was consumed (data not shown), and it was verified by testing growth of the strains on minimal media plates supplemented with arabinose that neither HZE2 nor HZE3 could grow, which indicated that the endogenous pathways were indeed removed and that the xylitol concentration decrease seen in the wild-type pathway (Figure 4.6A) is not due to conversion to D-xylulose and subsequent conversion to D-xylulose-5-phosphate by xylB. It is worth noting that negligible xylulose was found in the cultures during the experimental run in both cases, indicating LAD is still a limiting conversion step in the xylitol production process from L-arabinose. Also unfortunate is that xylitol production in this case was faster and of higher yield in the wild-type pathway than in the engineered pathway. It is difficult to ascertain why this result occurred. One possible explanation is the decrease in expression levels of ncLAD-3x in HZE3 compared to ncLAD-wt in HZE2 (Figure 4.2). Another possibility is the apparent decrease in affinity of ncLAD-3x for substrate arabinitol compared to the wild-type enzyme (see Chapter 3). To test whether a higher arabinose concentration would overcome the reduced affinity, an attempt at xylitol production in a bioreactor setting with 4% arabinose was attempted, but results were inconclusive due to cultures stalling production in both strains after 24 hours prior to sugar addition, and another run could not be completed by the submission deadline for this dissertation. Further optimization of this process is apparent, utilizing both buffered flasks studies combined with pH controlled bioreactor runs, to determine the proper conditions for optimal xylitol production and the effects of the cofactor balancing.
4.2.4.2. Engineered Xylitol Pathway – pH control vs. non-pH control

Figure 4.7 shows the profile of HZE3 when run in a bioreactor without pH control (Fig. 4.7A) and with pH control (Fig. 4.7B). Starting levels of glucose were 2% and arabinose was decreased to 1% in an attempt to fully convert L-arabinose at least to L-arabinitol, and this purpose was served in the pH controlled run. As seen from the uncontrolled profile, within 24 hours cells cease to grow and L-arabinose conversion as well as glucose consumption is halted. L-arabinitol accumulates to ~16 mM, but no xylitol production is seen during the course of the fermentation. When pH control is applied, cell density reaches a much higher level (OD$_{600}$ ~13), and L-arabinose if fully converted to L-arabinitol within 36 hours, where xylitol production is starting to be noticed. L-arabinitol is slowly converted to xylitol (up to 24 mM) until the end of the reactor run, but is never fully converted nor do levels drop below that of xylitol. While optimization of the xylitol pathway is still an issue in terms of productivity, the effect of the pH control is dramatic and the engineered pathway continued to show steady xylitol production from L-arabinose.
Figure 4.7. Comparison of 0.4 L (working volume) fermentation profiles for (A) uncontrolled and (B) pH-stat controlled HZE3 @ pH 7.0. The initial culture contained 2% glucose and 1% arabinose in minimal media + tryptone.
4.2.4.3. *WT* vs. *ENGR* Pathway – Co-utilization

As shown in Figure 4.8, the extension of the fungal xylitol pathway for co-utilization of pentose sugars xylose and arabinose resulted in a similar effect seen in the arabinose-only fermentations. Similar to the effect seen in the flask experiments and arabinose-only fermentation, the wild-type pathway (Figure 4.8A) again showed a halting of xylitol production after 48 hours, while the engineered pathway (Figure 4.8B) again showed slow yet steady production of xylitol from L-arabinitol.

![Figure 4.8](image-url)

Figure 4.8. Comparison of 0.4 L (working volume) fermentation profiles for co-utilization of xylose and arabinose by (A) HZE2 and (B) HZE3. The initial culture contained 3% glucose, 0.5% xylose, and 0.5% arabinose in minimal media + tryptone. Cultures were run at 30 °C, pH 7.0.
4.3. Discussion

In this work, we have studied the possibility of developing a more efficient pathway for xylitol production from L-arabinose and D-xylose, the two major components of hemicellulosic biomass hydrolysates. The pathway consists of three reactions, the first of which, xylose reductase, catalyzes the reduction of D-xylose directly to the desired product xylitol, and also converts L-arabinose to L-arabinitol. L-arabinitol is subsequently converted to L-xylulose by L-arabinitol 4-dehydrogenase and finally to xylitol by L-xylulose reductase. However, the reductases catalyze preferentially NADPH-dependent reactions, whereas the dehydrogenase is strictly NAD\(^+\)-dependent. Therefore, the xylitol pathway is redox imbalanced and the cellular metabolism is relied on to regenerate the cofactors in order to enhance flux towards product xylitol. In an attempt to alleviate the burden placed upon the cell to regenerate NADP\(^+\) and NADH, it was proposed to engineer the LAD enzyme to be NADP\(^+\)-dependent (as discussed in Chapter 3), which consolidates the xylitol pathway to a single cofactor pair, NADP(H). This effectively reduces the cell metabolism’s workload to only be needed to regenerate NADPH, which can be accomplished through either the pentose phosphate pathway or citric acid cycle with the addition of an appropriate co-substrate, in this case D-glucose.

In order to test the effects of utilizing an NADP\(^+\)-dependent LAD in an engineered xylitol pathway, *E. coli* was chosen as a suitable host organism. *E. coli* is an ideal organism for industrial production of chemicals due to its ability to assimilate both hexose and pentose sugars, rapid growth rates, ease of manipulation, and inexpensive growth medium requirements\(^14\). Most *E. coli* strains (including K-12) can uptake D-xylose and L-arabinose, but do not naturally synthesize or metabolize xylitol. However, *E. coli* will grow readily on pentose sugar substrates D-xylose and L-arabinose, and it has been previously established that xylitol production by
introduction of heterologously expressed enzymes that convert D-xylose to xylitol is severely limited when endogenous pathways for D-xylose utilization are present. In order to prevent the assimilation of pentose sugars by the pathway-expressing strain, the endogenous pathway genes for D-xylose (xylAB) and L-arabinose (araBAD) utilization were deleted.

For the engineered pathway, even though the NADP⁺-dependent LAD enzyme partially relieves the cofactor imbalance, there are still an odd number of enzymes involved in the conversion of either D-xylose or L-arabinose to xylitol, so regeneration of NADPH is still necessary by the cell’s metabolism. NADPH can be regenerated by the pentose phosphate pathway and the citric acid cycle when using co-substrate D-glucose. However, due to catabolite repression effects of using glucose with other sugars, the expression strain needed to be tailored to utilization these sugars simultaneously. Different strategies of catabolite de-repression for the co-utilization of the glucose, xylose, and arabinose were tested by Nikhil Nair in the Zhao laboratory. He found that incorporation of a mutant cyclic AMP-independent receptor protein (CRP*) showed efficient simultaneous assimilation of all three sugars, although xylose was shown to be the slowest\cite{15}. For all experiments run for conversion of L-arabinose to xylitol in this chapter, the simultaneously consumption of glucose, arabinose, and/or xylose was confirmed using a crp* expression strain.

The effect of engineering and NADP-dependent LAD for use in the xylitol production pathway varied depending on the reaction conditions. In bioreactor runs involving L-arabinose conversion alone with equal weight percent co-substrate glucose (Fig. 4.4) as well as co-utilization of D-xylose and L-arabinose (Fig. 4.6), it appeared the engineered pathway was better suited for xylitol production. Whereas the wild-type pathway experienced stalling of xylitol production, the engineered pathway with improved cofactor balancing showed slow yet continual
production of xylitol from the accumulated L-arabinitol. However, when monitoring the conversion more closely with flask studies, there appears to be an initial generation of xylitol in the wild-type pathway followed by conversion of xylitol to an unknown by-product. The xylitol conversion was not noticeable in the engineered pathway experiments, but the reactions were not carried out past five or six days and it is possible that the reaction is simply slower and the xylitol conversion does not occur until later in the reaction. Further analysis of the samples is currently being conducted to determine if the by-product is D-arabinitol, which could not be separated from L-arabinitol by the HPLC method used to detect all sugars in the culture broth. There were no detectable amounts of accumulating L-xylulose in the fermentation broths, which suggests that the LXR reaction is not limiting in xylitol formation and that LAD is still likely the bottleneck in the process. Fermentation with both D-xylose and L-arabinose as substrates showed full conversion of xylose to xylitol and arabinose to arabinitol, with slow steady conversion of arabinitol to xylitol in the engineered pathway compared to conversion of xylitol to an unknown byproduct in the wild-type pathway, as seen with buffered flask studies and arabinose-only fermentation.

The effect of pH-stat control is illustrated strongly in Figure 4.5. Without addition of base to counter the acetate production, cell growth halted after 24 hours and stayed around OD$_{600}$ levels of ~5.5 to 6.0. Low levels of L-arabinitol (~16 mM) were produced in the first 24 hours, but all xylitol pathway production shut down after this point and no xylitol was produced. In contrast, utilizing a pH-stat to keep levels at pH 7.0 resulted in enhanced growth of the expression strain, resulting in more than twice the amount of biomass produced. L-arabinose was fully converted the first enzyme, XR, to L-arabinitol within 24 hours and xylitol production was noticed after 48
hours. L-arabininitol was then slowly converted to xylitol for the remainder of the bioreactor run, although the process did not go to completion within 162 hours.

4.4. Conclusions and Outlook

This chapter illustrated the application of a fungal L-arabinose pathway for co-utilization of D-xylose and L-arabinose for xylitol production. While this represents a novel route of xylitol biosynthesis, the potential added benefit of introducing an engineered NADP⁺-dependent LAD from Chapter 3 to attempt to alleviate the cofactor imbalance inherent in the pathway is not yet established. There remain unsolved issues with low productivity that requires further optimization of the pathway. Also, the depletion of xylitol in the wild-type pathway has yet to be elucidated, and work is currently investigating whether byproduct D-arabininitol is accumulating instead, which could lead to a novel biosynthetic route to a different sugar alcohol than what was originally intended. Regardless, this groundwork represents an excellent strategy for improving utilization of both pentose sugars present in hemicellulosic materials for synthesis of value-added product xylitol. Chapter 5 will discuss work dedicated towards extending this engineering strategy for further downstream products attainable by fermentation of pentose sugars such as bioethanol.
4.5. Materials and Methods

4.5.1. Materials

*Escherichia coli* C600 was kindly provided by the Coli Genetics Stock Center (CGSC, Yale University). Plasmid pTrc99a was obtained from Amersham Biosciences. Restriction enzymes, *NsiI* and *EcoRI*, T4 DNA Ligase, and Phusion DNA polymerase were purchased from New England Biolabs (NEB, Beverly, MA). All media components were purchased from Becton-Dickinson (BD, Sparks, MD). Oligonucleotide primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA). All DNA purification kits were obtained from Qiagen (Valencia, CA). Wizard Genomic DNA Purification Kit was purchased from Promega (Madison, WI). Cells were maintained on Luria-Bertani (LB) plates containing 1.5% agar and the appropriate antibiotic.

4.5.2. Gene Deletion (Notebook #3, p. 127)

For xylitol production, *E. coli* K-12 C600 derivative strains were used. Deletions were performed using the λ red system. Briefly, PCR product containing *cat* flanked by FRT (Flp recognition target) and 45-50 nucleotides of sequence identical to the target locus (using forward primer “H1-P1 fwd araBp” with sequence 5’-AGC GGA TCC TAC CTG ACG CTT TTT ATC GCA ACT CTC TAC TGT GTA GGC TGG AGC TGC TTC-3’ and reverse primer “H2-P2 rev araD” with sequence 5’-TTC GCC GTT GAT TTC TGC GTC GGT CAT TTT GCG GGT GCA GCA GTG GAA TAT GGA TAT CCT CCT TAG-3’ in the amplification reaction) was transformed into cells expressing λ red recombinase proteins (encoded on pKD46). Gene replacement was selected on chloramphenicol plates and verified by functional assay and PCR.
Table 4.1. Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains/plasmids</th>
<th>Relevant Characteristics</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C600</td>
<td><em>F</em> tonA21 thi-1 thr-1 leuB6 lacY1 glnV44 rfbC1 fhuA1 λ-</td>
<td>CGSC, Yale University</td>
</tr>
<tr>
<td>ET23</td>
<td>Source of crp*::Tn10</td>
<td>Eppler and Boos, 1999</td>
</tr>
<tr>
<td>HZ1302</td>
<td>C600 crp*::Tn10</td>
<td>Nair and Zhao, 2009</td>
</tr>
<tr>
<td>HZ1450</td>
<td>C600 crp*:: Tn10 ΔxylA</td>
<td>Nair and Zhao, 2009</td>
</tr>
<tr>
<td>HZ1173</td>
<td>C600 crp*:: Tn10 ΔxylA ΔaraBAD</td>
<td>This work</td>
</tr>
<tr>
<td>HZE2</td>
<td>HZ1173 with pRPS231-wt</td>
<td>This work</td>
</tr>
<tr>
<td>HZE3</td>
<td>HZ1173 with pRPS231-3x</td>
<td>This work</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pTrc99a-wt</td>
<td>Amp, pBR322-derived plasmid</td>
<td>Amersham Pharmacia</td>
</tr>
<tr>
<td>pRPS231-wt</td>
<td>pTrc99a with LAD-wt, XR, LXR operon under araBp promoter</td>
<td>This work</td>
</tr>
<tr>
<td>pRPS231-3x</td>
<td>pTrc99a with LAD-3x, XR, LXR operon under araBp promoter</td>
<td>This work</td>
</tr>
</tbody>
</table>

The resistance marker was then removed by the expression of Flp recombinases from a thermo-inducible promoter on a temperature sensitive plasmid (pCP20). Flp recombinase plasmid loss and *cat* loss occurred simultaneously and was verified by the sensitivity to ampicillin and chloramphenicol. Deletion of *xylA* was previously performed by Nikhil Nair, and deletion of *araBAD* was verified by PCR using cell lysates as template and appropriate flanking primers “Fwd minus 275 araC” (5’-AAA CCA ATT GTC CAT ATT GCA TCA GA-3’) and “Rev araD stop” (5’-TTA CTG CCC GTA ATA TGC CTG CGC GC-3’). It should be noted that *araB* and *araA* were fully deleted, while *araD* was truncated to avoid any possible downstream conflicts of the *araBAD* operon.

4.5.3. Mutant crp* Integration

The *crp* mutant was generated by Nikhil Nair in the Zhao laboratory by P1 transduction from *crp*-containing plasmid ET23 (obtained from Patrick C. Cirino, Penn State University) and selecting for TetR integrants. Verification of glucose de-repression was first done by blue/white screening on LB plates containing 10 g/L glucose. Strong induction of *lacZ* in presence of
glucose indicated de-repressed phenotype. Finally, monitoring of sugar co-utilization in shake flasks was used to verify de-repression.

4.5.4. Operon Design and Construction (Notebook #3, p. 231-234)

Overlap extension PCR (OE-PCR)\(^{16}\) was used to construct the xylitol synthesis operons. Operon 231-wt consisted of the following in order: an arabinose-inducible promoter from \(E.\ coli\) C600 genomic DNA (araBp), wild-type \(N.\ crassa\) L-arabinitol 4-dehydrogenase (ncLAD-wt), \(A.\ niger\) L-xylulose reductase (anLXR), and \(N.\ crassa\) xylose reductase (ncXR). The linker region between \(E.\ coli\) genes araB and araA (10 bp) was used as the linker region between ncLAD-wt and anLXR, and was designed into the primers used for amplification. The linker region between \(E.\ coli\) genes araA and araD (284 bp) was used as the linker region between anLXR and ncXR, and was amplified from \(E.\ coli\) C600 genomic DNA, with sequences listed in Table 4.3. All primers used for operon construction are listed in Table 4.2. Operon 231-3x was constructed in a similar fashion to operon 231-wt, except that engineered \(N.\ crassa\) L-arabinitol 4-dehydrogenase (ncLAD-3x) was used in place of ncLAD-wt. Operons 231-wt and 231-3x were digested with \(NsiI\) and \(EcoRI\) and ligated into identical sites of vector pTrc99a with T4 DNA Ligase. Ligation mixtures were transformed into DH5\(\alpha\) and selected on Luria Broth (LB) agar plates supplemented with 100 \(\mu\)g/mL ampicillin. Colonies were inoculated into 5 mL LB medium supplemented with 100 \(\mu\)g/mL ampicillin, and plasmids were isolated from the overnight culture using the Plasmid Miniprep Kit from Qiagen (Valencia, CA). Plasmids isolated were then subjected to restriction digestion by \(NsiI\) and \(EcoRI\), after which the reaction mixtures were loaded onto 1% agarose gels and checked for correct restriction digestion pattern by DNA electrophoresis.
Table 4.2. Primers sequences used in construction of the xylitol synthesis operons. Forward primer names consist of directionality, restriction site or overhang, and the name of the region being amplified. For example, “Fwd NsiI araBp” stands for the forward primer used to amplify *araBp* with an *Nsi*I cut site in front of the promoter (5’-end). Reverse primer names consist of directionality, the region being amplified, and the restriction site or overhang. For example, “Rev araBp ncLAD” stands for the reverse primer used to amplify *araBp*, with a 3’-end overhang sequence homologous to *ncLAD*. Restriction sites shown with underlines.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>DNA Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fwd NsiI araBp</td>
<td>TTGATCGGTTATGCATCCGCCATTCAGAGAAAGAAC</td>
</tr>
<tr>
<td>Rev araBp ncLAD</td>
<td>GAAGCGCTAGAAGCCATCGTTCCACTCCATCC</td>
</tr>
<tr>
<td>Fwd araBp ncLAD</td>
<td>GGATGGAGTGAAACGATGCGTTCTAGCGCTTC</td>
</tr>
<tr>
<td>Rev ncLAD-linkBA-anLXR</td>
<td>GAGGGAATGGAGATAGGCATTATCGTGCTCTTACTCCAGACTCTGGATC</td>
</tr>
<tr>
<td>Fwd ncLAD-linkBA-anLXR</td>
<td>GATCCAGAGTCTGGAGTAAGGACACGATAATGCCTATCTCCATTCCCTC</td>
</tr>
<tr>
<td>Rev anLXR-linkAD</td>
<td>GTTACATACCGGATCGGCTACTTACCAGCACGCTGTATCC</td>
</tr>
<tr>
<td>Fwd anLXR-linkAD</td>
<td>GGATACACCGTGGTGTAAGGCGCATCGGTATGTAAC</td>
</tr>
<tr>
<td>Rev linkAD-ncXR</td>
<td>GAGCTTGATAGCAGGAAACCATGTGGACTTCCCTGGTGCCGGATG</td>
</tr>
<tr>
<td>Fwd linkAD-ncXR</td>
<td>CATCCGGCAAGGAGTCAACATGGTTCTGCTATCAAGCTC</td>
</tr>
<tr>
<td>Rev ncXR EcoRI</td>
<td>AAGTCGGATCGAATTCCTAACCAGAAAATCCAGAGGTTT</td>
</tr>
</tbody>
</table>
### Table 4.3. Sequences of non-gene components of xylitol synthesis operons.

<table>
<thead>
<tr>
<th>Component</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>araBp promoter</td>
<td>5’-cgccattaacagaagaaacaattgtcatatgacacattgcgactgcgctgttttaactgctt ttttttttgatggga-3’</td>
</tr>
<tr>
<td>linkBA</td>
<td>5’-ggacacagata-3’</td>
</tr>
<tr>
<td>linkAD</td>
<td>5’-gtagccgcatccggtatgcagaagctgtacgctgttactgtggacacctgcgatgtttagggc gataaagccgcacggacctgcgctgatcagctggttgctttttgacgctgcgatttttactgct</td>
</tr>
<tr>
<td></td>
<td>gcctgcctcaacgcgctgatggtttttaagcggataagccgcacggacctgcgctgatcagctggttgctttttgacgctgcgatttttactgct</td>
</tr>
</tbody>
</table>

#### 4.5.5. HPLC Analysis

Glucose, xylose, arabinose, xylitol, and arabinitol concentrations were quantified using Shimadzu LC-20AT high performance liquid chromatography (HPLC) equipped with a low temperature evaporative light scattering detector (ELSD-LT). Products were separated with a Bio-Rad HPX-87C column (20 μL injection) with water as the mobile phase (0.2 mL/min, 85 °C).

#### 4.5.6. Shake Flask Studies (Notebook #4, p. 279)

For shake flask studies, overnight cultures were grown at 37 °C in LB media + 100 mM MOPS (pH 7.4) and appropriate antibiotics. 0.5 mL of overnight cultures were inoculated into 50 mL of
LB media + 100 mM MOPS (pH 7.4) containing sugars (glucose, xylose, and/or arabinose) and appropriate antibiotics in 250 mL baffled flasks and grown at 30 °C. Samples were withdrawn at various time points and measured for cell densities at OD$_{600}$ using a Cary Bio 100 UV Vis spectrometer. Samples were also filtered using 0.2 µm syringe filters (Millipore) and analyzed by HPLC for sugar analysis.

4.5.7. Bioreactor Studies (Notebook #4, p. 158)

For bioreactor studies, the seed culture was prepared by growing cells first on a fresh LB plate supplemented with 10 µg/mL tetracycline and 100 µg/mL ampicillin and inoculating a single colony into 4 mL LB media and appropriate antibiotics at 37 °C. 1 mL of saturated overnight culture was centrifuged at 4,000 rpm for 4 minutes and media were decanted. Cells were resuspended in 5 mL sterile water and centrifuged again at 4,000 rpm for 4 minutes and water was decanted. Cells were resuspended in M9 medium supplemented with 2 mM MgSO$_4$, 0.1 mM CaCl$_2$, 30 g/L glucose, 10 g/L tryptone, and appropriate antibiotic(s) and grown for another 4-5 hours at 37 °C. The log phase cultures were then inoculated into a Multifors multiple benchtop bioreactor (Infors HT, Switzerland) with 0.5 L working volume containing 400 mL of the same M9 + tryptone medium with additional 10 g/L each of D-xylose and/or L-arabinose, as well as antifoam. Bioreactors were run at 30 °C with an agitation rate of 800 rpm by dual Rushton impellers and a sparging rate of 0.8 L/min with sterile air. pH was maintained at 7.0 ± 0.1 with 5 N NaOH and 2 N H$_2$SO$_4$. Samples were withdrawn at various time points and measured for cell densities at OD$_{600}$ using a Cary Bio 100 UV Vis spectrometer. Samples were also filtered using 0.2 µm syringe filters (Millipore) and analyzed by HPLC for sugar analysis.
References


(11) Akinterinwa, O.; Cirino, P. C. Heterologous expression of D-xylulokinase from Pichia stipitis enables high levels of xylitol production by engineered Escherichia coli growing on xylose Metab Eng 2009, 11, 48-55.


CHAPTER 5 – Fungal Arabinose Pathway Complementation Towards Determining Limitations in Arabinose Utilization

5.1. Introduction

Lignocellulosic biomass is regarded as a highly promising feedstock for a rapidly expanding biofuels industry in response to a pressing energy problem\(^1\). Lignocellulose refers to plant biomass that is composed of cellulose (long-chain homopolymers of D-glucose), hemicellulose (branched polysaccharides mainly consisting of pentose sugars D-xylose and L-arabinose), and lignin (complex cross-linked macromolecules). The majority of lignocellulosic materials consists of D-glucose (~30-50%) and D-xylose (~20-25%), but L-arabinose is still a significant third portion in some forms of lignocellulose (~ 5-10%)\(^2,3\). Utilization of all sugars in the raw material is favored for an economically competitive process.

Yeast cells are particularly attractive for cellulosic ethanol processes because they have been used in biotechnology for hundreds of years, are well-characterized physiologically and genetically, have ample genetic tools, are tolerant to high ethanol and inhibitor concentrations, and can grow at low pH values which avoids bacterial contaminations. Unfortunately, the widely used industrial yeast *Saccharomyces cerevisiae* cannot metabolize the pentose sugars D-xylose or L-arabinose found in hemicellulose. *S. cerevisiae* naturally harbors genes for xylose utilization, but they are expressed at such low levels that they do not support growth on D-xylose\(^4\). A large number of yeast species metabolize D-xylose and L-arabinose and display fermentative ability, but only approximately 1% of them are capable of fermenting D-xylose to ethanol\(^5\). Various yeasts and fungi were initially screened to evaluate their ability to ferment L-arabinose under oxygen-limited conditions when grown in defined minimal media containing
mixtures of L-arabinose, D-xylose, and D-glucose\textsuperscript{6}. Although all of the yeasts and some of the fungi consumed L-arabinose, L-arabinose was not fermented to ethanol by any of the strains tested. L-Arabinitol was the only major product other than cell mass formed from L-arabinose. The inability to ferment L-arabinose was postulated to be a consequence of inefficient or incomplete assimilation pathways for this pentose sugar. A subsequent study identified four yeast species able to ferment L-arabinose to ethanol\textsuperscript{7}, but the discrepancy was likely due to the use of complex (YP) medium that could have contained compounds that may act as electron acceptors and thus aid conversion of arabinose to ethanol\textsuperscript{5}.

Pathways for D-xylose and L-arabinose metabolism share a common intermediate in the form of D-xylulose-5-phosphate, but are distinctly different in bacteria and fungi. Bacterial catabolism of D-xylose and L-arabinose begins with isomerization to D-xylulose and L-ribulose, respectively, by an isomerase. D-xylulose and L-ribulose are phosphorylated by a xylulokinase to D-xylulose-5-phosphate and to L-ribulose-5-phosphate, respectively, with the latter being converted by an epimerase to D-xylulose-5-phosphate and connecting the initial pentose metabolism to the central metabolism through the pentose phosphate pathway (PPP). The yeast pathway has been previously described (Figure 1), and has been more successful in its implementation due to troubles getting functional bacterial enzymes in \textit{S. cerevisiae}.

Many metabolic engineering strategies have been attempted to improve fermentation of D-xylose to ethanol in recombinant \textit{S. cerevisiae} harboring D-xylose-utilization enzymes. These include over-expression of endogenous xylulokinase\textsuperscript{8}, glyceraldehydes-3-phosphate dehydrogenase\textsuperscript{9}, transaldolase and transketolase, or deletion of glucose 6-phosphate dehydrogenase\textsuperscript{10}. Alternative protein engineering strategies aimed at the cofactor imbalance of the initial pathway enzymes has also been pursued, by introducing NADH-preferring XR mutants\textsuperscript{11,12}, or NADP-dependent XDH
mutants\textsuperscript{13,14}. However, these cofactor balancing strategies have only been dedicated towards xylose fermentation.

The fermentation of L-arabinose to ethanol has only recently been established in \textit{S. cerevisiae}, with reports of the introduction of the bacterial pathway\textsuperscript{15,16}, or the fungal pathway\textsuperscript{17,18} showing promising progress in the field of L-arabinose utilization by \textit{S. cerevisiae}, but yields of ethanol using either strategy have been low. Co-utilization of D-xylose and L-arabinose has also been shown to occur with fungal D-xylose and bacterial L-arabinose pathways introduced into \textit{S. cerevisiae} \textsuperscript{5,19}, but most of the L-arabinose was converted to byproduct L-arabinitol by the fungal xylose reductase. During the course of the work presented in this thesis, Bettiga \textit{et al.} expressed an engineered fungal L-arabinose pathway into \textit{S. cerevisiae} and achieved anaerobic biomass production and ethanol production. The pathway combined i) an engineered mutant XR with increased NADH preference; ii) a NAD$^+$-dependent LAD from \textit{Trichoderma reesei} (Hypocreajecorina); iii) a NADH-dependent LXR from \textit{Ambrosiozyma monospora} and iv) a NAD$^+$-dependent XDH from \textit{Pichia stipitis}. The resultant pathway was cofactor balanced by consolidating the initial pathway enzymes to the alternative cofactor pairing, NAD(H), whereas my research focused on using a pathway designed for NADP(H). Expression of the Bettiga pathway in their strain TMB3043 was chosen as background since it has been endowed with genetic modifications allowing for improved pentose fermentation, such as overexpression of PPP and XKS. In addition, TMB3043 harbors the deletion of GRE3, coding for an unspecific NADPH aldose reductase. The removal of the Gre3p NADPH-dependent activity contributes to a more cofactor balanced pentose utilizing pathway. The final ethanol concentration was 9.3 g/l, with a yield of 0.21 g/(g sugar) based on total sugars present at the beginning of the fermentation, 0.42 g/(g sugar) based on consumed sugars and 0.35 g/(g sugar) based on
consumed L-arabinose. While this strategy represented an almost 10-fold improvement over the previous reported values\textsuperscript{17}, clearly much progress is still needed in this area to make the process more efficient and economically feasible.

The proposed pentose utilization project deals with utilizing L-arabinose in combination with D-xylose for the production of ethanol. The initial sequence of four alternating reduction and oxidation steps in L-arabinose catabolism is potentially a rate-limiting factor in the catabolic sequence\textsuperscript{20}, as natural L-arabinose metabolizing yeasts, as well as engineered \textit{Saccharomyces cerevisiae}, produce large amounts of L-arabinitol. The goal of this project is to determine the effect of reversing the cofactor specificity for the dehydrogenases (LAD and XDH) involved in the fungal L-arabinose utilization pathway. This chapter will discuss progress made in cloning and engineering an NADP\textsuperscript{+}-dependent \textit{N. crassa} XDH, the introduction of the full initial L-arabinose pathway into \textit{S. cerevisiae}, and the development of a selection method using \textit{E. coli} to set the foundation for future pathway engineering.

5.2. Results

5.2.1. Cloning, Characterization, and Engineering of an NADP\textsuperscript{+}-Dependent \textit{N. crassa} Xylitol Dehydrogenase

A putative \textit{Neurospora crassa} xylitol dehydrogenase sequence was found using a protein BLAST search on the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov) using the \textit{Pichia stipitis} xylitol dehydrogenase enzyme as a query sequence. The two enzymes were aligned fully using a ClustalW algorithm and found to share 44\% identity and 60\% similarity (Figure 5.1). The whole-genome sequence of \textit{Neurospora}
*Neurospora crassa* has been published\(^1\), and it was utilized to design primers for cloning of the putative xylitol dehydrogenase gene.

Score = 299 bits (765), Expect = 3e-79, Method: Compositional matrix adjust.
Identities = 169/365 (46%), Positives = 222/365 (60%), Gaps = 22/365 (6%)

<table>
<thead>
<tr>
<th>psXDH</th>
<th>ANPSLVNLKIDISFETYDAPEISEPTDVQLVQVKTGICGSDIH+FYAHRGIGNFVLTKPM 62</th>
</tr>
</thead>
<tbody>
<tr>
<td>ncXDH</td>
<td>SNLSFVLNKLPLDVCPQDKPVKINSHPDHLVAVNYTGICGDVHYVLHGAIGHFFVVDPM 66</td>
</tr>
<tr>
<td>psXDH</td>
<td>VLGHESA GTTVQVGKVTSLKVGDNVAIEPGISPSEPSEYKSGHYNLCPHMFAAPNSK 122</td>
</tr>
<tr>
<td>ncXDH</td>
<td>VLGHESA GT+V VG +L VG VA+EPG P R SGHYNLCP M FAATP</td>
</tr>
<tr>
<td>psXDH</td>
<td>GAGPVGLLA AAVAKTFGAKGIVVDIFDNKLMKADIGAATHTFNSTGSEE-------LI 237</td>
</tr>
<tr>
<td>ncXDH</td>
<td>GAGPVGLLAAVAKTFGAKGIVVDIFDNKLMKADIGAATHTFNSTGSEE-------LI 237</td>
</tr>
<tr>
<td>psXDH</td>
<td>KAFGGNVPSMVECTGAEPCIKLGDVAPGGRFQVQGAAPGVSVFPITVFAMKEITLFG 297</td>
</tr>
<tr>
<td>ncXDH</td>
<td>AAGPVGLLCAAVAKTFGAKGIVVDIFDNKLMKADIGAATHTFNSTGSEE-------LI 237</td>
</tr>
<tr>
<td>psXDH</td>
<td>SFRYGDFTDKVGTFNYQNGRNAPIDEQGTLTHYKFDAEAYDLVPAEGGAVKC 357</td>
</tr>
<tr>
<td>ncXDH</td>
<td>SFRYG D+ A+ +S +G +VQG +++FP +K +F +A EA+ VG +G +K</td>
</tr>
<tr>
<td>psXDH</td>
<td>LIAGP 354</td>
</tr>
</tbody>
</table>

**Figure 5.1.** BLAST Results for putative XDH in *Neurospora crassa*.

RT-PCR performed on total RNA isolated from D-xylose-induced *N. crassa* 10333 showed the expected size of gene product (~1.1 kb). The RT-PCR product was cloned into the pET-28a vector using *Nde*I and *Sac*I restriction sites and was transformed into *E. coli* BL21 (DE3). This construct (pET-28a ncXDH) expressed *N. crassa* XDH as an N-terminal His\(_6\)-tagged fusion with a thrombin cleavage site. Cell lysates of IPTG-induced cultures of these cells were prepared, analyzed by SDS-PAGE, and assayed for XDH activities. The XDH was then purified by...
immobilized metal ion affinity chromatography (IMAC) using Talon® Co\textsuperscript{2+} Superflow resin (Clontech, Mountain View, CA) according to manufacturer’s protocol. The purified protein was desalted by ultrafiltration with several washes of 50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) buffer (pH 7.25) + 15% glycerol and stored frozen at -80 °C. Protein concentrations were determined by the Bradford method\textsuperscript{22}.

ncXDH is a strictly NAD\textsuperscript{+}-preferring enzyme. ncXDH also displays high stability (half-life of ~200 min at 50 °C) and expression. Previous work by Watanabe \emph{et al.} was aimed at reversing the cofactor specificity of \textit{Pichia stipitis} xylitol dehydrogenase (psXDH).\textsuperscript{14} Through sequence alignment, residues D204, I205, and V206 were targeted for site-directed mutagenesis to alanine, arginine, and serine, respectively, to create ncXDH-ARS. Table 5.1 shows that ncXDH-ARS has completely reversed cofactor specificity, now preferring NADP\textsuperscript{+}, with little change in the \( k_{\text{cat}} \) value and only a 2.5-fold higher \( K_{m, \text{NADP}^+} \) for ncXDH-ARS compared to the \( K_{m, \text{NAD}^+} \) for the wild-type. Table 5.2 shows that the affinity for substrate xylitol did not suffer the loss that was previously seen in engineering an NADP\textsuperscript{+}-dependent ncLAD (see Chapter 4). It should be noted that the same mutations were attempted in the ncLAD engineering, but the improvement was not as drastic, necessitating the library screening that was implemented in Chapter 3.
Table 5.1. Kinetic parameters for \textit{N. crassa} and \textit{P. stipitis} XDH and XDH-ARS with nicotinamide cofactors NAD$^+$ and NADP$^+$. \\

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>( k_{\text{cat}} ) (min$^{-1}$)</th>
<th>( K_m ) (mM)</th>
<th>( k_{\text{cat}}/K_m ) (min$^{-1}$)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ncXDH-wt</td>
<td>2160</td>
<td>0.127</td>
<td>17000</td>
<td>This work</td>
</tr>
<tr>
<td>ncXDH-ARS</td>
<td>~3.5</td>
<td>~165</td>
<td>~68</td>
<td>This work</td>
</tr>
<tr>
<td>psXDH</td>
<td>1050</td>
<td>0.381</td>
<td>2760</td>
<td>Watanabe et al.</td>
</tr>
<tr>
<td>psXDH-ARS</td>
<td>240</td>
<td>1.3</td>
<td>181</td>
<td>Watanabe et al.</td>
</tr>
</tbody>
</table>

\( ^a \)Not determined, cofactor saturation not reached

All assays were performed at 25 °C in 50 mM Tris, pH 8.0

Table 5.2. Kinetic parameters of ncXDH mutants for substrate xylitol. \\

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>( k_{\text{cat}} ) (min$^{-1}$)</th>
<th>( K_m ) (mM)</th>
<th>( k_{\text{cat}}/K_m ) (min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ncXDH-wt</td>
<td>2170 ± 135</td>
<td>6.6 ± 2.0</td>
<td>330</td>
</tr>
<tr>
<td>ncXDH-ARS</td>
<td>2090 ± 35</td>
<td>4.3 ± 0.3</td>
<td>490</td>
</tr>
</tbody>
</table>

\( ^a \)Not determined, saturation of cofactor not reached

All assays run at 25 °C, 50 mM Tris-HCl pH 8.0
All enzymes were purified and characterized with N-His$_6$Tag

As discussed in Chapter 4, the pH rate profile of \textit{N. crassa} LAD in the L-arabinose pathway favor basic conditions, whereas it is known that the intracellular pH of \textit{E. coli}\textsuperscript{23} or \textit{S. cerevisiae}\textsuperscript{24} is neutral to acidic. The same holds true for \textit{N. crassa} XDH, as shown in Figure 5.2, but XDH activity exhibits a higher tolerance to more acidic conditions with activity extending down to pH 4.0, whereas LAD activity is abolished at pH 5.0 in the \textit{in vitro} activity assay.
Figure 5.2. Comparison of pH rate profiles of *N. crassa* LAD and XDH. Data taken from the characterization of LAD from Chapter 2 (open circles) was performed in universal buffer MES/Tris/glycine, and overlapped with data for ncXDH (closed triangles) and ncLAD (closed circles) performed in universal buffer acetic acid/MES/Tris for lower pH values.

### 5.2.2. Introduction of L-Arabinose Pathway into *S. cerevisiae*

The initial five enzymes of the fungal L-arabinose pathway that convert L-arabinose to D-xylulose-5-phosphate leading into the pentose phosphate pathway were constitutively expressed in *S. cerevisiae*. The wild-type pathway was assembled with NAD\(^+\)-dependent ncLAD-wt and ncXDH-wt, while the engineered pathway was assembled with NADP\(^+\)-dependent ncLAD-3x (see Chapter 3) and ncXDH-ARS (see Section 5.2.1) and correctly assembled constructs were verified by restriction digestion with *Bam*HI (Fig. 5.3a) and by PCR amplification of the five gene cassettes (Fig. 5.3b). Strains were then grown on synthetic complete media lacking uracil (SC-Ura) supplemented with either glucose, xylose, or arabinose as a carbon source. Figure 5.4 shows that strains containing either the wild-type or engineered pathway grew readily in SC-Ura + xylose, while the parent strain YSG50 could not utilize xylose. However, neither strain was
found to grow in liquid SC-Ura + 2% arabinose. There was minimal growth in solid agar plates with the same media over the course of 14 days.

Figure 5.3. Example of yeast plasmid verification. (a) Restriction digest with *Bam*HI, where sL = simulated ladder, sD = simulated digest pattern, L = DNA ladder, and 1-7 digests of plasmids prepared from seven different colonies with 100% efficiency of correctness. (b) PCR amplification of gene cassettes from verified plasmid. L = DNA ladder, 1 = XR cassette, 2 = XDH cassette, 3 = XKS cassette, 4 = LAD cassette, 5 = LXR cassette, (-) = negative control with respective cassette primers run with empty plasmid.

Figure 5.4. Growth of L-arabinose pathway expressing *S. cerevisiae* strains in SC-Ura + 2% xylose. The WT pathway contains all wild-type L-arabinose pathway genes, while the ENGR pathway replaces both NAD⁺-dependent ncLAD-wt and ncXDH-wt with NADP⁺-dependent ncLAD-3x and ncXDH-ARS. The control strain is YSG50 with empty plasmid pRS426.
5.2.3. *E. coli* Complementation Experiments

Due to the inability to get a functional L-arabinose pathway in *S. cerevisiae*, *E. coli* was chosen as a suitable host for the development of a selection method to determine (1) whether the fungal pathway would be able to complement the bacterial pathway and circumvent the issues associated with *S. cerevisiae* such as pentose transport, gene expression and protein solubility, potential acidic intracellular pH conditions, and transcriptional regulation, and (2) whether the cofactor balanced engineered pathway would result in improved growth and/or ethanol production. Table 5.3 lists the strains used for xylose complementation experiments in *E. coli* MG1655. The deletion of *xylA* by *cat* insertion was confirmed to abolish the ability of HZE60 to grow in minimal media supplemented with xylose. In order to confirm that complementation was possible by expression of the endogenous genes under the xylAp promoter in a plasmid, constructs with either *xylA* or *xylA* and *xylB* together were transformed into HZE60 and selected in xylose minimal media. As shown in Figure 5.5, over-expression of *xylA* alone was able to complement growth in xylose minimal media. When both *xylA* and *xylB* were over-expressed, the strain grew faster, which suggests that the deletion of XylA activity with the *cat* insertion resulted in a polar effect that subsequently decreased XylB expression as well in HZE60.

| Table 5.3: Strains constructed for D-xylose growth complementation experiments |
|---------------------------------|---------------------------------|
| MG1655                          | F λ  ilvG rfb-50 rph-1 (ATCC 700926) |
| HZE60                           | MG1655 ΔxylA::cat (obtained from Tasha Desai) |
| HZE66                           | MG1655 ΔxylA::cat pTrc99a |
| HZE67                           | MG1655 ΔxylA::cat pTrc99a-xylAp-xylA |
| HZE68                           | MG1655 ΔxylA::cat pTrc99a-xylAp-xylAB |
| HZE101                          | MG1655 ΔxylA::cat pTrc99a-xylAp-ncXDH(wt)-ncXR-xylB |
| HZE102                          | MG1655 ΔxylA::cat pTrc99a-xylAp-ncXDH(ARS)-ncXR-xylB |
| HZE109                          | MG1655 ΔxylA::cat pTrc99a-xylAp-ncXDH(wt)-ncXR-psXKS |
| HZE110                          | MG1655 ΔxylA::cat pTrc99a-xylAp-ncXDH(ARS)-ncXR-psXKS |
| HZE123                          | MG1655 ΔxylA::cat pBLMA-xylAp-ncXDH(wt)-ncXR-xylB |
| HZE124                          | MG1655 ΔxylA::cat pBLMA-xylAp-ncXDH(ARS)-ncXR-xylB |
To determine whether XR and XDH together could complement XylA activity, two constructs were tested, containing XR and either wild-type NAD⁺-dependent XDH or engineered NADP⁺-dependent XDH. In Figure 5.5, the wild-type pathway under the control of xylAp with either xylB (HZE101) or psXKS (HZE109) was confirmed to complement growth, whereas the engineered pathways (HZE102 and HZE110) did not complement growth in minimal media with xylose. This is contradictory to what was seen in *S. cerevisiae*. However, when plated on M9 media with 5% xylose, slow grow was noticable with the engineered pathway under the control of xylAp (HZE110, Figure 5.6).

Figure 5.5. Growth complementation on M9 + 0.4% xylose at 37 °C (See Table 5.3 for legend).
Figure 5.6. Growth complementation on M9 + 5% arabinose at 30 °C. (See Table 5.3 for legend).

Table 5.4 lists the strains used for arabinose complementation experiments in *E. coli* MG1655. Again, the same control strategy was applied by knocking out endogenous *araBAD*, which resulted in the lack of cell growth in minimal media supplemented with L-arabinose, and expression of *araBAD* in the knockout strain showing complementation was achieved.

Table 5.4. Strains constructed for L-arabinose growth complementation experiments

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HZE62</td>
<td>MG1655 Δ<em>araBAD</em> (obtained from Tasha Desai)</td>
</tr>
<tr>
<td>HZE71</td>
<td>MG1655 Δ<em>araBAD</em> pTrc99a</td>
</tr>
<tr>
<td>HZE72</td>
<td>MG1655 Δ<em>araBAD</em> pTrc99a araBp-araBAD</td>
</tr>
<tr>
<td>HZE103</td>
<td>MG1655 Δ<em>araBAD</em> pTrc99a + pRSFDuet</td>
</tr>
<tr>
<td>HZE104</td>
<td>MG1655 Δ<em>araBAD</em> pTrc99a araBp-ncLAD(wt)-anLXR-ncXR + pRSFDuet araBp-ncXDH(wt)-xylB</td>
</tr>
<tr>
<td>HZE105</td>
<td>MG1655 Δ<em>araBAD</em> pTrc99a araBp-ncLAD(3x)-anLXR-ncXR + pRSFDuet araBp-ncXDH(ARS)-xylB</td>
</tr>
</tbody>
</table>

However, when the wild-type (HZE104) or engineered (HZE105) fungal arabinose pathways were expressed in the *araBAD*-deletion strain, neither construct could complement cell growth on L-arabinose media (data not shown).
5.3. Discussion

As mentioned earlier, the wild-type L-arabinose pathway has been introduced into *S. cerevisiae* previously by Richard *et al.* although with little resultant ethanol production from L-arabinose.\(^{17}\) With the first four enzymes of the L-arabinose pathway already cloned, the final enzyme, D-xylulokinase from *Pichia stipitis* (psXKS) was first cloned from *P. stipitis* cDNA into *E. coli* prior to introduction into *S. cerevisiae*. Over-expression of xylulokinase in *S. cerevisiae* has been reported to enhance specific ethanol production when grown on D-xylose\(^{25}\). Introduction of the L-arabinose pathway into *S. cerevisiae* was accomplished by a new homologous recombination-based DNA assembly method developed by Zengyi Shao and Hua Zhao in the Zhao laboratory.\(^{26}\)

The full-length pathway placed in pRS426 was comprised of the following gene cassettes (“p” indicates promoter, “t” indicates terminator sequence): ADH1p-ncXR-ADH1t, GPDp-ncLAD-PYK1t, HXT7p-anLXR-HXT7t, PGK1p-ncXDH-CYC1t, and PYK1p-psXKS-ADH2t. Colonies were selected on synthetic complete media lacking uracil (SC-Ura) with either D-xylose or L-arabinose as the sole carbon source to confirm the pathway was functional. Colonies formed readily on D-xylose after about four days, albeit slower than when compared to the typical two day incubation for D-glucose containing media. The engineered pathway strain (containing the NADP-dependent dehydrogenases ncLAD-SRN and ncXDH-ARS) also grew on D-xylose, and had similar growth rate compared to the wild-type pathway. An interesting, although unfortunate, observation was that while colonies were able to form on L-arabinose plates after ~10-14 days, subsequent inoculation of these colonies into liquid SC-Ura with L-arabinose as the sole carbon source were unable to grow for reasons yet to be determined. Colonies that grew up
on both pentose sugar plates were checked by colony PCR and restriction digestion analysis to confirm the pathway had been assembled properly.

It remains unknown why S. cerevisiae grew so poorly in L-arabinose media. Cofactor imbalance, one of the key areas of research in the proposed project, is thought to be one reason. However, transport of pentose sugars has also been brought to attention27-29, and this is currently being investigated by Jing Du and Sijin Li in the Zhao group. Recently, strong evidence has shown support for the need of improvement of the LAD enzyme. Two natural L-arabinose-utilizing yeasts, Candida arabinofermentans PYCC 5603T and Pichia guilliermondii PYCC 3012 were compared to gain more insights into the limitations of L-arabinose metabolism. 13C NMR was used to trace the fate of L-[2-13C]arabinose through branching metabolic pathways, and intracellular accumulation of arabinitol appeared in both strains, indicating a bottleneck at the level of LAD which was also NAD+-dependent in these strains30. It is hypothesized that an improved ncLAD enzyme, along with cofactor reversal for the dehydrogenase enzymes for NADP+-dependence, can shed light on the utilization of pentose sugars for ethanol production. However, the results from expressing the fungal L-arabinose pathway in S. cerevisiae suggest that the cofactor imbalance alone is not the answer to obtaining a strain that can grow readily on L-arabinose.

As shown in Section 5.2.4, the L-arabinose pathways were also expressed in E. coli as a model organism. E. coli can grow readily in either xylose or arabinose containing media, which suggests it does not suffer from the possible limitations that S. cerevisiae has in pentose utilization. Preliminary results have shown that the wild-type and, to a lesser extent, the engineered fungal xylose pathway under the control of xylAp can complement growth in M9 + 5% xylose media at 30 °C.
5.4. Conclusions

The development of a selection method based on the complementation of *E. coli* pentose utilization genes with those of fungal origin can help to determine the bottlenecks associated with the initial pathway enzymes and conclude whether engineering a cofactor balanced pathway helps in pentose utilization. The XDH from *N. crassa* was cloned, characterized, and engineered for NADP⁺-dependence, and the five genes that convert L-arabinose to D-xylulose-5-phosphate for entrance into the pentose phosphate pathway were introduced into *S. cerevisiae*. Unfortunately, only growth on xylose was attained, with minimal growth on solid media containing arabinose only. Due to potential issues involved with transport, expression, etc., the pathways were tested in *E. coli* with resultant growth only achievable on xylose and not arabinose. Further optimization of the pathway expression is still required in order for the arabinose pathway to be functional.

5.5. Materials and Methods

5.5.1. Materials

The *Neurospora crassa* genomic sequence and LAD protein sequences were accessed from the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). *Neurospora crassa* 10333 was obtained from the American Type Culture Collection (ATCC). *Escherichia coli* BL21(DE3) and pET-28a(+) expression vector were obtained from Novagen (Madison, WI). SuperScript™ One-Step RT-PCR with Platinum® Taq kit was obtained from Invitrogen (Carlsbad, CA). PCR grade dNTPs were obtained from Roche Applied Sciences (Indianapolis, IN). Phusion High-Fidelity DNA Polymerase, shrimp alkaline phosphatase, and DNA-modifying enzymes DNase I, *NdeI*, *SacI*, and T4 DNA ligase and their appropriate buffers were
purchased from New England Biolabs (NEB) (Beverly, MA). Xylitol, β-D-thiogalactopyranoside (IPTG), and NADP⁺ were purchased from Sigma (St. Louis, MO). NAD⁺ was a gift from Jülich Fine Chemicals (Jülich, Germany). Other required salts and reagents were purchased from Fisher (Pittsburgh, PA) or Sigma-Aldrich. The QIAprep spin plasmid mini-prep kit, QIAquick gel purification kit, RNeasy midiprep kit, and QIAquick PCR purification kit were purchased from Qiagen (Valencia, CA). Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA). SDS-PAGE gel materials, electrophoresis equipment, and protein size markers were from Bio-Rad (Hercules, CA).

5.5.2. RT-PCR and Cloning

*N. crassa* 10333 was grown on rich potato media at 30 °C for 24 h and induced with 150 mM D-xylose for 2 hr. Total ribonucleic acid (RNA) was purified from collected cells (RNeasy purification kit, Qiagen) and treated with DNase I to remove residual genomic DNA. RT-PCR was performed using SuperScript™ One-Step RT-PCR with Platinum® Taq (Invitrogen) following the manufacturer’s guidelines. A control reaction consisted of the same protocol except for the SuperScript™ enzyme mix was heated to 95 °C for ten minutes to thermally inactivate the reverse transcriptase enzyme and the reverse transcription thermocycler step was omitted. The primers used for the RT-PCR were: forward 5′- GTA GCT ACG TCA CAT ATG CTA CCG ACG GCA AGT CT -3′ and reverse 5′- AGC TGA TAG CGA GCT C7T AGC AGC AGC CAG AAC C -3′. The forward primer contained an *NdeI* restriction site (shown in bold), while the reverse primer contained a *SacI* restriction site (shown in bold) and stop codon (italicized). The resulting RT-PCR product was isolated by a QIAquick agarose gel purification kit. The product was digested with *NdeI* and *SacI* restriction enzymes and purified again by agarose gel electrophoresis. It was then ligated into pET-28a(+) which had been previously
prepared by NdeI and SacI digestion, dephosphorylation by shrimp alkaline phosphatase, and gel purification. The ligation mixture was used to transform *E. coli* DH5α by electroporation. Positive clones were selected on Luria–Bertani (LB) solid media with ampicillin at 37 °C overnight. All colonies were then removed from the plates and grown to saturation in 5 mL liquid LB from which the plasmids were purified using a QIAGEN spin plasmid miniprep kit, which were used to transform *E. coli* BL21(DE3) by heat shock. Positive clones were selected on LB solid media with ampicillin, picked individually, and assayed for XDH activity by the cell lysate assay described below. Plasmids were sequenced using the BigDye® Terminator sequencing method and an ABI PRISM 3700 sequencer (Applied Biosystems, Foster City, CA). Sequencing results determined at the Biotechnology Center of the University of Illinois.

5.5.3. Cell Lysate Assay

*E. coli* BL21 (DE3) harboring the pGEX-4T-3-derived vector was grown to a maximum optical density at 600 nm (OD$_{600}$) at 37 °C with shaking at 250 rpm. Fifty microliters was used to inoculate a new culture, which was grown at 37 °C with shaking at 250 rpm until an OD$_{600}$ of ~0.6 was reached. The cultures were then induced with 0.3 mM IPTG and shaken at 250 rpm at 25 °C for 4 h. One milliliter of cells was harvested by centrifugation and lysed by resuspension in 1 mL of 1 mg/mL lysozyme/50 mM Tris-HCl (pH 8.0). The cells were frozen at −80 °C and thawed at room temperature. The resulting lysate was vortexed thoroughly and centrifuged to remove cell debris. Ten microliters of the lysate was used in a kinetic assay described below in the Enzyme Kinetics Section with 200 mM xylitol and 2 mM NAD$^+$ as the substrates in 50 mM Tris (pH 8.0).
5.5.4. Enzyme Kinetics

Initial rates were determined by measuring the initial rate of reaction as observed by the increase of absorbance at 340 nm using a Varian Cary 100 Bio UV-visible spectrophotometer (Varian, Palo Alto, CA) at 25 °C in 50 mM Tris-HCl (pH 8.0). *N. crassa* XDH and XDH-ARS kinetic measurements were taken with saturation of xylitol and varying cofactor NAD$^+$ or NADP$^+$ concentrations or vice versa. The data were used to calculate the kinetic constants by fitting the Michaelis-Menten equation using Origin 8.0. The data represent averages of assays performed in duplicate or triplicate on two separate occasions.

5.5.5. pH Profile

Activity was measured at pH values between 3.5 and 8.0 under saturating concentrations of NAD$^+$ (2 mM) and xylitol (200 mM) in a universal buffer (50 mM morpholineethanesulfonic acid (MES) / 50 mM Tris / 50 mM glycine). Ionic strength of the universal buffer was taken into consideration according to previously established buffer compositions$^{31}$.

5.5.6. Pathway Construction (*S. cerevisiae*) (Notebook #5, p. 101)

*S. cerevisiae* YSG50 (*MATα ade2-1 ade3Δ22 ura3-1 his3-11,15 trp1-1 leu2-3,112 and can1-100*) was used as the host for DNA assembly. pRS426 plasmid (New England Biolabs, Beverly, MA) was linearized with *Bam*HI and *Xho*I restriction enzymes. To prepare individual gene expression cassettes, yeast promoters (ADH1p, 1500 bp; PGK1p, 750 bp; PYK1p, 1000 bp; GPDp, 655 bp, and HXT7p, 395 bp), genes (ncXR, ncXDH, psXKS, ncLAD, and anLXR), and yeast terminators (ADH1t, 327 bp; CYC1t, 250 bp, ADH2t, PYK1t, 400 bp; HXT7t, 400 bp) were amplified from corresponding plasmids in the Zhao laboratory stock library. Each individual gene cassette was assembled by overlap extension PCR (OE-PCR)$^{32}$. Following electrophoresis, the OE-PCR product was gel purified from a 1.0% agarose gel using Qiagen Gel
Purification Kit (Valencia, CA). Each individual gene expression cassette (~300 ng) was mixed with linearized pRS426 and transformed into *S. cerevisiae*.

### 5.5.7. Yeast Transformation (*S. cerevisiae*)

Yeast transformation was performed by electroporation. YPAD medium (50 mL) was inoculated with a 0.5 mL overnight culture of *S. cerevisiae* YSG50 and incubated at 30 °C with shaking at 250 rpm for 4-5 hours until OD$_{600}$ reached 0.8-1.0. Yeast cells were harvested by centrifugation at 4 °C and 4000 rpm for 5 min. The supernatant was decanted and the cell pellet was washed with 50 mL cold sterile water, followed by another wash with cold 1 M sorbitol and finally resuspended in 250 µL cold 1 M sorbitol. An aliquot of 50 µL of yeast cells together with DNA mixture containing linearized pRS426 vector and five gene cassettes, and electroporated in a 0.2 cm cuvette at 1.5 kV. The transformed cells were immediately mixed with 1 mL room temperature YPAD medium and shaken at 30 °C for 1 hour. Cells were harvested by centrifugation, washed with room temperature 1 M sorbitol several times to remove remaining media, and finally resuspended in 1 mL 1 M sorbitol. Aliquots of 30-50 µL were spread on SC-Ura plates for the selection of the transformants, and the plates were incubated at 30 °C for 2-4 days until colonies appeared. Colonies were randomly picked into SC-Ura liquid media and grown for 1 day, after which the yeast plasmid was isolated using Zymoprep II Yeast plasmid Miniprep kit (Zymo Research, CA), and confirmed by restriction digestion analysis and PCR.

### 5.5.8. Pathway Construction (*E. coli*) (Notebook #4, p. 151)

Xylose and arabinose pathway pathway operons were amplified by overlap extension PCR (OE-PCR). Xylose pathway operons consisted of XDH (wild-type XDH-wt or engineered XDH-ARS), linker BA (see Chapter 4), XR, linker AD (see Chapter 4), and xylB under the xylose-inducible xylAp promoter. Following electrophoresis, xylose pathway operons were digested
with \textit{NsiI} and \textit{EcoRI} and ligated into pTrc99a vector and transformed into DH5\(\alpha\). Plasmids were purified and verified by plasmid purification and digestion with respective flanking restriction enzymes and PCR amplification of genes, yielding plasmids pRPS415-wt and pRPS415-eng that were transformed into \textit{xylA}-mutant \textit{E. coli} HZE60. L-Arabinose pathways consisted of two operons. The first operon contained the xylose pathway (wild-type or engineered) as described above except that \textit{xylAp} was replaced with the arabinose-inducible \textit{araBp} promoter. Operons were digested with \textit{NsiI} and \textit{EcoRI}, and ligated into pTrc99a. The second operon contained LAD (wild-type LAD-wt or engineered LAD-3x), linker BA (see Chapter 4), and LXR under the \textit{araBp} promoter. These operons were digested with \textit{XbaI} and \textit{EcoRI} and ligated into pRSFDuet. Ligations were transformed into DH5\(\alpha\), constructs were verified and plasmids were co-transformed into \textit{araBAD}-mutant \textit{E. coli} HZE62.
References


(12) Petschacher, B.; Nidetzky, B. Altering the coenzyme preference of xylose reductase to favor utilization of NADH enhances ethanol yield from xylose in a metabolically engineered strain of *Saccharomyces cerevisiae* *Microb Cell Fact* **2008**, *7*, 9.


