8-2013

Bioengineering the Expression of Active Recombinant Human Cathepsin G, Enteropeptidase, Neutrophil Elastase, and C-Reactive Protein in Yeast

Eliot T. Smith

East Tennessee State University

Follow this and additional works at: https://dc.etsu.edu/etd

Part of the Biochemistry Commons, Immune System Diseases Commons, Medical Biochemistry Commons, Respiratory Tract Diseases Commons, and the Structural Biology Commons

Recommended Citation


This Dissertation - unrestricted is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.
Bioengineering the Expression of Active Recombinant Human Cathepsin G, Enteropeptidase, Neutrophil Elastase, and C-Reactive Protein in Yeast

A dissertation
presented to
the faculty of the Department of Biomedical Sciences
East Tennessee State University
In partial fulfillment
of the requirements for the degree
Doctor of Philosophy in Biomedical Sciences

by
Eliot Thomas Smith
August 2013

David A. Johnson, Ph.D., Chair
Michelle M. Duffourc, Ph.D.
J. Russell Hayman, Ph.D.
Antonio E. Rusiñol, Ph.D.
Douglas P. Thewke, Ph.D.

Keywords: Cathepsin G, C-Reactive Protein, Cytochrome B5, Elastase, Enteropeptidase, Neutrophil, Pichia pastoris, Kluyveromyces lactis, Polymorphonuclear Leukocyte, Serine Protease
ABSTRACT

Bioengineering the Expression of Active Recombinant Human Cathepsin G, Enteropeptidase, Neutrophil Elastase, and C-Reactive Protein in Yeast

by

Eliot T. Smith

The yeasts *Pichia pastoris* and *Kluyveromyces lactis* were used to express several recombinant human proteins for further biochemical characterization. Two substitution variants of recombinant human enteropeptidase light chain (rhEP\textsubscript{L}) were engineered to modify the extended substrate specificity of this serine protease. Both were secreted as active enzymes in excess of 1.7 mg/L in *P. pastoris* fermentation broth. The substitution variant rhEP\textsubscript{L} R96Q showed significantly reduced specificities for the preferred substrate sequences DDDDK and DDDDR; however, the rhEP\textsubscript{L} Y174R variant displayed improved specificities for these substrate sequences relative to all other reported variants of this enzyme. The neutrophil serine proteases human cathepsin G (hCatG) and human neutrophil elastase (HNE) were expressed in *P. pastoris* and HNE was also expressed in *K. lactis*. The recombinant variants rhCatG and rHNE, with intact C-terminal extensions, were expressed as fusion proteins with the soluble heme-binding domain of cytochrome B5 (CytB5) and an N-terminal hexahistidine (6xHis) tag for purification. The CytB5 domain was linked to the native N-termini of active rhCatG and rHNE by the EP\textsubscript{L}-cleavable substrate sequence DDDDK~I, where ~ is the sessile bond. These fusion proteins were directed for secretion. The yeast *P. pastoris* expressed up to 3.5 mg/L of EP\textsubscript{L}-activable rHNE in fermentation broth; however, only 200 µg/L of rhCatG could be produced by this method. Recombinant expression in *K. lactis* never surpassed 100 µg/L of activable rHNE. The CytB5 fusion domain was present in the heme-bound form, conferring a red color and 410 nm
absorbance peak to solutions containing the fusion proteins. This absorbance pattern was most readily visible during the purification of CytB5-rHNE from *P. pastoris*. Human C-reactive protein (hCRP) and the substitution variant CRP E42Q were expressed in recombinant form and secreted by *P. pastoris*. Both products were found to bind phosphocholine (PCh) in the same manner as native hCRP. Difficulties encountered during purification revealed that wild type recombinant CRP (rCRP) was produced at 2 different molecular masses. The *P. pastoris* recombinant expression system yielded better results than *K. lactis*. Bioreactor-scale fermentation in a 5 L vessel facilitated expression and characterization of these recombinant proteins.
DEDICATION

For my mother and father,
Because of their eternal love and support.

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

-T.S. Eliot

Do not go where the path may lead,
Go instead where there is no path
And leave a trail.

-R.W. Emerson
ACKNOWLEDGEMENTS

Thanks be to God for all of the gifts and wonders of this life.

To Dave Johnson: Thank you for your wisdom, your knowledge, your patience, and your guidance. You are a wonderful mentor and true friend. I have learned more from you and with you than words can convey.

To Michelle Duffourc, Russ Hayman, Antonio Rusiñol, and Doug Thewke: Thank you for your guidance and advice as members of my graduate committee and as friends.

To lab mates, past and present: Thank you all for your hard work and support on these projects. Thanks for singing along…and for all the clean glassware….

To Beverly Sherwood and Angela Thompson: Thank you both for your guidance and support.

This work was supported by NIH grant R15HL091770.
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>4</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>5</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>11</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>14</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>15</td>
</tr>
</tbody>
</table>

### Chapter

1. INTRODUCTION ........................................................................................................ 17
   - Recombinant Protein Engineering ........................................................................ 17
     - Recombinant Expression Systems ......................................................................... 18
       - Bacterial Systems ............................................................................................. 18
       - Animal Systems ................................................................................................. 19
       - Yeast Systems ................................................................................................... 20
     - *P. pastoris* and *K. lactis* as Recombinant Expression Systems ...................... 21
       - *P. pastoris* Expression System .......................................................................... 21
       - *K. lactis* Expression System ............................................................................ 22
     - Custom Tailored Gene Synthesis and Codon Usage Optimization ............................ 23
   - Designing Recombinant Proteins ............................................................................ 27
     - Amino Acid Substitutions and Deletions ................................................................ 27
     - Fusion Proteins and Tags ..................................................................................... 28
     - Designing Internal Proteolytic Sites .................................................................... 31
   - Serine Proteases .................................................................................................... 33
Neutrophil Serine Proteases of Primary (Azurophil) Granules ........................................ 34
Human Proteinase 3 ............................................................................................................. 38
Human Neutrophil Elastase ................................................................................................. 39
Human Cathepsin G ............................................................................................................. 41
Expression of Recombinant Human NSPs ......................................................................... 42
C-reactive Protein ................................................................................................................ 44
Specific Aims ........................................................................................................................ 50
2. HUMAN ENTEROPEPTIDASE LIGHT CHAIN: BIOENGINEERING OF RECOMBINANTS AND KINETIC INVESTIGATIONS OF STRUCTURE AND FUNCTION ........................................................................................................................................... 51
Abstract ............................................................................................................................... 52
Introduction ............................................................................................................................ 53
Results ................................................................................................................................. 55
Protein Engineering and Molecular Cloning ....................................................................... 55
Purification ............................................................................................................................. 56
Identity Confirmation and Stability Analyses ....................................................................... 57
Kinetic Analyses .................................................................................................................... 59
Discussion .............................................................................................................................. 63
Materials and Methods ........................................................................................................ 65
Molecular Cloning ............................................................................................................... 67
Fermentation .......................................................................................................................... 67
Purification ............................................................................................................................. 68
SDS-PAGE and Western Blotting ......................................................................................... 68
Cleavage of CCP ................................................................................................................... 69
Stability Assay ...................................................................................................................... 69
Kinetic Assays ....................................................................................................................... 69
Molecular Modeling .............................................................................................................. 70
3. ENGINEERING THE EXPRESSION OF HUMAN NEUTROPHIL ELASTASE AND CATHEPSIN G IN *PICHIA PASTORIS* AND *KLUYVEROMYCES LACTIS* ................................................................. 75

Abstract ........................................................................................................................................ 76

Introduction ...................................................................................................................................... 77

Results and Discussion ..................................................................................................................... 79

   Protein Modifications and Fusion Protein Design .......................................................................... 79

      NSP Domains and Modifications ................................................................................................ 79

      Cytochrome B5 Recognition and Purification Domain ............................................................... 81

      Fusion Protein Zymogens ............................................................................................................ 81

   Transformation and Screening ....................................................................................................... 83

   Protein Expression ......................................................................................................................... 85

   Partial Purification and Activation: rHNE and rhCatG in *P. pastoris* ........................................ 86

   Partial Purification and Activation: rHNE in *K. lactis* ............................................................. 91

The Utility of Cytochrome B5 Fusion Domains in Recombinant Protein Expression ......................... 93

Conclusions ....................................................................................................................................... 95

Materials and Methods .................................................................................................................... 96

   Molecular Cloning ........................................................................................................................ 96

   Transformation of *P. pastoris* ....................................................................................................... 97

   Transformation of *K. lactis* ........................................................................................................... 97

Expression Screening ....................................................................................................................... 98

      Culture Conditions ..................................................................................................................... 98

      Enzyme Activation ...................................................................................................................... 98

      Activity Assays .......................................................................................................................... 98

      PCR Testing for Multicopy Transformants .............................................................................. 99
4. THE EXPRESSION OF NATIVE HUMAN C-REACTIVE PROTEIN AND HUMAN C-REACTIVE PROTEIN E42Q IN PICHIA PASTORIS

Abstract

Introduction

Results

Protein Engineering and Molecular Cloning

Transformant Screening

Protein Expression

PCh Binding Assays

Initial Purification and SDS-PAGE

Further Purification

Discussion

Materials and Methods

Molecular Cloning and Yeast Transformation

Transformant Screening

ELISA

Protein Expression

Purification

SDS-PAGE and Western Blotting

Acknowledgements

References
5. DISCUSSION AND CONCLUSION .......................................................................................... 134
  Yeast Recombinant Expression Systems .............................................................................. 134
  Promoter Selection, Codon Optimization, and Custom Gene Synthesis .............................. 134
  Glycosylation and Cleavage Site Disruptions ....................................................................... 135
  Thoughts on Arginine Content and Recombinant Protein Expression ................................. 136
  Enteropeptidase: Variations on a Theme .............................................................................. 137
  The Untold Story: Other Expression Attempts with rhCatG and rHNE .............................. 139
  Improving the Purification of CytB5-rhCatG and CytB5-rHNE ............................................ 141
  Learning to Fly: Lessons in Recombinant Yeast Fermentation ........................................... 143
REFERENCES ................................................................................................................................ 144
APPENDIX: ON BIOREACTOR BASED FERMENTATION OF YEAST ...................................... 168
VITA ............................................................................................................................................... 188
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α1-AT</td>
<td>α1-antitrypsin</td>
</tr>
<tr>
<td>ANCA</td>
<td>Antineutrophil Cytoplasmic Antibody</td>
</tr>
<tr>
<td>AOX1</td>
<td>Alcohol Oxidase 1 Gene</td>
</tr>
<tr>
<td>AP3</td>
<td>Adaptor Protein 3</td>
</tr>
<tr>
<td>BBS</td>
<td>Borate Buffered Saline</td>
</tr>
<tr>
<td>bEP</td>
<td>Bovine Enteropeptidase</td>
</tr>
<tr>
<td>bEP_L</td>
<td>Bovine Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>BMGY</td>
<td>Buffered Minimal Glycerol Medium</td>
</tr>
<tr>
<td>BMMY</td>
<td>Buffered Minimal Methanol Medium</td>
</tr>
<tr>
<td>BPTI</td>
<td>Basic Pancreatic Trypsin Inhibitor</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>BSM</td>
<td>Basal Salts Medium</td>
</tr>
<tr>
<td>CatG</td>
<td>Cathepsin G</td>
</tr>
<tr>
<td>CCP</td>
<td>Cleavage Control Protein</td>
</tr>
<tr>
<td>CN</td>
<td>Cyclic Neutropenia</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disorder</td>
</tr>
<tr>
<td>CRP</td>
<td>C-Reactive Protein</td>
</tr>
<tr>
<td>CytB5</td>
<td>Cytochrome B5</td>
</tr>
<tr>
<td>DPP I</td>
<td>Dipeptidyl Peptidase I</td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5'-dithiobis-(2-nitrobenzoic acid)</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>eGFP</td>
<td>Enhanced Green Fluorescent Protein</td>
</tr>
<tr>
<td>ELANE</td>
<td>Neutrophil Elastase Gene</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>EP</td>
<td>Enteropeptidase</td>
</tr>
<tr>
<td>EP_L</td>
<td>Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>ERAD</td>
<td>Endoplasmic Reticulum Associated Degradation</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>FMN</td>
<td>Flavin Mononucleotide</td>
</tr>
<tr>
<td>GD4K-na</td>
<td>Gly-Asp-Asp-Asp-Asp-Lys-naphthylamide</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>hCatG</td>
<td>Human Cathepsin G</td>
</tr>
<tr>
<td>hEP</td>
<td>Human Enteropeptidase</td>
</tr>
<tr>
<td>hEP_L</td>
<td>Human Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>HNE</td>
<td>Human Neutrophil Elastase</td>
</tr>
<tr>
<td>HSA</td>
<td>Human Serum Albumin</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>ICC</td>
<td>Immunocytochemistry</td>
</tr>
<tr>
<td>IHC</td>
<td>Immunohistochemistry</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized Metal Affinity Chromatography</td>
</tr>
<tr>
<td>LAC4</td>
<td>β-Galactosidase Gene</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny Broth</td>
</tr>
<tr>
<td>LDL</td>
<td>Low Density Lipoprotein</td>
</tr>
<tr>
<td>MAC</td>
<td>Membrane Attack Complex</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple Cloning Site</td>
</tr>
<tr>
<td>MeO-Suc-AAPV-pNA</td>
<td>Methoxysuccinyl-Ala-Ala-Pro-Val-para-nitroanilide</td>
</tr>
<tr>
<td>MPO</td>
<td>Myeloperoxidase</td>
</tr>
<tr>
<td>MSG</td>
<td>Monosodium Glutamate</td>
</tr>
<tr>
<td>MUGB</td>
<td>4-Methylumbelliferyl-para-guanidinobenzoate</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular Weight Cut Off</td>
</tr>
<tr>
<td>NET</td>
<td>Neutrophil Extracellular Trap</td>
</tr>
<tr>
<td>NHS</td>
<td>N-Hydroxysuccinamid</td>
</tr>
<tr>
<td>NSP</td>
<td>Neutrophil Serine Protease</td>
</tr>
<tr>
<td>oxLDL</td>
<td>Oxidized Low Density Lipoprotein</td>
</tr>
<tr>
<td>PAR</td>
<td>Protease Activated Receptor</td>
</tr>
<tr>
<td>PCh</td>
<td>Phosphocholine</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>pI</td>
<td>Isoelectric Point</td>
</tr>
<tr>
<td>PMN</td>
<td>Polymorphonuclear</td>
</tr>
<tr>
<td>Pr3</td>
<td>Proteinase 3</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene Fluoride</td>
</tr>
<tr>
<td>RA</td>
<td>Relative Adaptiveness</td>
</tr>
<tr>
<td>rbEP_L</td>
<td>Recombinant Bovine Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>rCRP</td>
<td>Recombinant Human C-Reactive Protein</td>
</tr>
<tr>
<td>RF</td>
<td>Relative Frequency</td>
</tr>
<tr>
<td>rhCatG</td>
<td>Recombinant Human Cathepsin G</td>
</tr>
<tr>
<td>rhEP_L</td>
<td>Recombinant Human Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>rhEP_L,sc</td>
<td>Supercharged Recombinant Human Enteropeptidase Light Chain</td>
</tr>
<tr>
<td>rHNE</td>
<td>Recombinant Human Neutrophil Elastase</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcription Polymerase Chain Reaction</td>
</tr>
<tr>
<td>SCN</td>
<td>Severe Congenital Neutropenia</td>
</tr>
<tr>
<td>SHMP</td>
<td>Sodium Hexametaphosphate</td>
</tr>
<tr>
<td>SLE</td>
<td>Systemic Lupus Erythematosis</td>
</tr>
<tr>
<td>SMM</td>
<td>Synthetic Minimal Medium</td>
</tr>
<tr>
<td>STI</td>
<td>Soybean Trypsin Inhibitor</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Suc-AAPF-SBzl</td>
<td>Succinyl-Ala-Ala-Pro-Phe Thiobenzyl Ester</td>
</tr>
<tr>
<td>SUMO</td>
<td>Small Ubiquitin-Like Modifier Protein</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-Buffered Saline</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumor Necrosis Factor α</td>
</tr>
<tr>
<td>WCW</td>
<td>Wet Cell Weight</td>
</tr>
<tr>
<td>WG</td>
<td>Wegener’s Granulomatosis</td>
</tr>
<tr>
<td>YCB</td>
<td>Yeast Carbon Base</td>
</tr>
<tr>
<td>YNB</td>
<td>Yeast Nitrogen Base</td>
</tr>
<tr>
<td>Z-Lys-SBzl</td>
<td>N-α-carbobenzyloxy-L-Lysine Thiobenzyl Ester</td>
</tr>
</tbody>
</table>


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>STI-Agarose Purification of rhEP\textsubscript{L} Variants</td>
<td>57</td>
</tr>
<tr>
<td>2.2</td>
<td>Kinetic Parameters of rhEP\textsubscript{L} Variants</td>
<td>60</td>
</tr>
<tr>
<td>2.3</td>
<td>Inhibition of rhEP\textsubscript{L} by BPTI</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Three Step Purification of rCRP</td>
<td>121</td>
</tr>
<tr>
<td>5.1</td>
<td>Other Expression Attempts with rhCatG and rHNE</td>
<td>139</td>
</tr>
<tr>
<td>A1</td>
<td>Fermentation Media Components</td>
<td>169</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Codon Usage Frequency</td>
<td>25</td>
</tr>
<tr>
<td>1.2</td>
<td>Codon Optimization</td>
<td>26</td>
</tr>
<tr>
<td>1.3</td>
<td>NSP Sequence Alignment</td>
<td>37</td>
</tr>
<tr>
<td>1.4</td>
<td>CRP Monomer</td>
<td>45</td>
</tr>
<tr>
<td>1.5</td>
<td>CRP Pentamer</td>
<td>46</td>
</tr>
<tr>
<td>2.1</td>
<td>EP&lt;sub&gt;L&lt;/sub&gt; Sequence Alignment</td>
<td>56</td>
</tr>
<tr>
<td>2.2</td>
<td>Molecular Characterization of Purified EP&lt;sub&gt;L&lt;/sub&gt; Variants</td>
<td>58</td>
</tr>
<tr>
<td>2.3</td>
<td>Enzyme Stability Assay</td>
<td>59</td>
</tr>
<tr>
<td>2.4</td>
<td>Specificity Constants</td>
<td>61</td>
</tr>
<tr>
<td>2.5</td>
<td>Electrostatic Models of EP&lt;sub&gt;L&lt;/sub&gt; Extended Binding Pockets</td>
<td>64</td>
</tr>
<tr>
<td>3.1</td>
<td>Amino Acid Sequence Alignments</td>
<td>80</td>
</tr>
<tr>
<td>3.2</td>
<td>Fusion Construct Schematics</td>
<td>82</td>
</tr>
<tr>
<td>3.3</td>
<td>Screening Colonies for rhCatG</td>
<td>84</td>
</tr>
<tr>
<td>3.4</td>
<td>Absorbance Spectra for CytB5 Fusion Protease Zymogens After</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Partial Purification by IMAC</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>IMAC Elution Profiles</td>
<td>88</td>
</tr>
<tr>
<td>3.6</td>
<td>SDS-PAGE and Western Blot for Partially Purified CytB5-rHNE</td>
<td>89</td>
</tr>
<tr>
<td>3.7</td>
<td>Activation of CytB5-rHNE Partially Purified by IMAC from</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K. lactis Fermentation Medium</td>
<td>92</td>
</tr>
<tr>
<td>4.1</td>
<td>Primary Structure of rCRP Construct with α-Mating Factor</td>
<td>114</td>
</tr>
<tr>
<td>4.2</td>
<td>PCh Binding Assays</td>
<td>118</td>
</tr>
<tr>
<td>4.3</td>
<td>SDS-PAGE of PCh-Affinity Purified rCRP</td>
<td>119</td>
</tr>
<tr>
<td>4.4</td>
<td>Three Step Purification of rCRP</td>
<td>120</td>
</tr>
<tr>
<td>4.5</td>
<td>SDS-PAGE of Three Step rCRP Purification</td>
<td>121</td>
</tr>
<tr>
<td>4.6</td>
<td>Western Blot of hCRP and rCRP</td>
<td>122</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>A1</td>
<td>Catabolism of Methanol</td>
<td>172</td>
</tr>
<tr>
<td>A2</td>
<td>Bioreactor Vessel Head Plate</td>
<td>177</td>
</tr>
<tr>
<td>A3</td>
<td>Bioreactor Vessel</td>
<td>179</td>
</tr>
<tr>
<td>A4</td>
<td>Bioreactor Control Tower</td>
<td>180</td>
</tr>
<tr>
<td>A5</td>
<td>Bioreactor Digital Interface</td>
<td>181</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Recombinant Protein Engineering

Recombinant protein expression has emerged as a powerful tool for the biomedical sciences during the past several decades. Researchers began using recombinant DNA technologies to reprogram *Escherichia coli* for protein expression in the 1970s (Marians et al. 1976; Wood and Lee 1976; Itakura et al. 1977; Meagher et al. 1977; Miller et al. 1977a; Miller et al. 1977b). Recombinant human insulin produced in *E. coli* was first approved by the FDA for therapeutic use in 1983 (Johnson 1983) and has since supplanted native porcine and bovine insulin as the preferred source of the drug. In the decades since this innovation, the pharmaceutical industry has heavily pursued recombinant proteins as novel approaches to the treatment of many diseases.

Recombinant protein expression technologies now enable a plethora of otherwise impossible experimental approaches. Recombinant human proteins are often produced in large quantities for further experimentation. Amino acid sequences are modified to study the contributions of individual residues to protein structures and functions. Multiple protein domains with distinct properties can be coupled into fusion constructs that exhibit combined traits of the individual component proteins, a strategy that can be used to facilitate detection and purification. Novel proteins can even be invented *de novo*, a promising science that is still in its infancy (Koga et al. 2012). These pursuits, among many others that have evolved out of recombinant protein expression technologies, continue to provide new methods for biomedical research and yield novel medical treatments.
The general approach to recombinant protein expression begins with the molecular cloning of protein-coding DNA into a genetic vector and delivery of this programing into a suitable host organism (Marians et al. 1976; Wood and Lee 1976; Itakura et al. 1977; Meagher et al. 1977; Miller et al. 1977a; Miller et al. 1977b). The transformed host can then express the encoded protein; however, the specific nature of any recombinant expression depends on the interplay of several factors including the traits of the chosen host organism, the nature of the genetic program, and the properties of the recombinant protein.

Recombinant Expression Systems

Many organisms are useful as hosts for recombinant protein expression, including especially bacterial, animal, and yeast cell types. With distinct biological properties, each of these expression systems behaves differently and each is best suited to different project goals. Successful expression requires consideration of many factors in selecting an appropriate host organism as an expression platform.

Bacterial Systems. *E. coli* strains are the most commonly used host organisms because they are prokaryotes and possess a simple genetic system. The native genetics of these bacteria have been well-characterized and the genome sequences of many *E. coli* strains are available through the National Center for Biotechnology Information (NCBI) database. Recombinant DNA is typically introduced using plasmid vectors that are easily manipulated with standard molecular cloning techniques. *E. coli* can be transformed by electroporation or chemical means and transformants are routinely screened by overnight growth with antibiotic resistance selection on lysogeny broth (LB) agar plates. Liquid cultures of *E. coli* grow quickly, can grow to high
cell densities, and generate large quantities of protein under proper conditions, thus these cultures are suitable for batch fermentation on an industrial scale.

*E. coli* do not contain the machinery required for eukaryotic post-translational processing steps like chaperone-aided folding, disulfide bond formation, glycosylation, and trafficking; consequently, large and complex proteins often express poorly in *E. coli*, commonly forming inclusion bodies within the cytosol. These aggregates of insoluble, misfolded protein must be isolated from the lysed cells, solubilized, and refolded. Such complicated processing can be difficult and time consuming, can require harsh chemical treatments and can significantly reduce the amount of functional protein that is recovered. Despite these drawbacks, *E. coli* strains remain the first choice for many recombinant protein engineers due to organism simplicity, ease of manipulation, rapid growth, and industrial scalability.

**Animal Systems.** In contrast to the simplicity of *E. coli*, animal cell lines, which are equipped with eukaryotic machinery, provide co- and post-translational processing that may be necessary for successful expression of many human proteins. The variety of available mammalian cell types offers more options for customizing the expression system to the specific requirements of the target protein and the project. This can be necessary for studies that require proteins to be produced in their native environs, such as experiments on trafficking, multienzyme pathways, and membrane-associated proteins.

Despite these advantages, expression systems based on animal cell lines suffer from many drawbacks. Stable, permanent genetic manipulations do not occur as readily in animal cells as they do in many microorganisms, making transformants more difficult to generate. Many animal cell lines do not grow in suspension and require an adherent surface; furthermore, such cell lines grow slowly and are poorly suited for industrial scalability. Even cell lines that do
grow in suspension and scale well for industrial purposes typically need complex growth media, often containing fetal bovine serum (FBS) and other specialized components that microorganisms do not require. Total protein yields are often low because of the complexity of animal host cells. Ultimately, the use of animal cells incurs higher costs, requires more time, and may still result in low yield protein expression.

Yeast Systems. Yeast species such as *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Kluyveromyces lactis* provide many of the advantages of both bacterial and animal expression systems while eliminating many drawbacks because they are single-cell fungi. Yeast genetics are well-characterized and easily manipulated. *E. coli* retain their recombinant genetic programming as plasmids that can be lost with time and lack of selection pressure and stable transformants of animal cells can be difficult to generate and screen. Many expression systems for yeast, however, rely on homologous recombination to insert the new genes directly into the host genome. This yields many stable transformants that grow rapidly and are easy to screen, with almost no risk of the cells losing their programing over time. Yeasts grow rapidly in suspension to high cell densities and require simple media formulations and environmental conditions. They are also very well suited to large batch fermentations; furthermore, they often yield milligrams or grams of protein per liter of fermentation broth, making them useful for industrial scale production.

Yeasts contain machinery for co- and post-translational modifications that are often required for proper expression of human proteins; furthermore, lack of inclusion body formation, a problem with *E. coli*, results in high yield recovery of properly folded proteins without the need for harsh resolubilizing and refolding procedures. Although some yeast strains are known to hyperglycosylate recombinant proteins (Niles et al. 1998; van Oort et al. 2002; van Oort et al.
Pichia GlycoSwitch®; http://www.rctech.com/licensing/gxt-pichia-glycoswitch.php). Through the use of yeast secretion signals like α-mating factor, recombinant proteins can be directed for secretion into the growth media; consequently, this can eliminate the need for cell lysis.

**P. pastoris and K. lactis as Recombinant Expression Systems**

Recombinant expression systems based on the yeast species *P. pastoris* and *K. lactis* have become well established and new improvements continue to emerge. Many expression plasmids have been developed that apply various strategies to enable successful expression in both species. Genetic programs available in these plasmids include options for inducible or constitutive promoters, amino- (N-) and carboxyl (C-) terminal fusion tags, multiple reading frames, and secreted or internal expression. Existing plasmids can, furthermore, be modified through standard molecular cloning procedures to completely customize any aspect of the vector. Homologous recombination is generally exploited to insert copies of linearized plasmid DNA, called insertion cassettes, into the host genome. This creates a permanent stable transformation of the cells into a new strain that expresses the recombinant protein construct of interest.

**P. pastoris Expression System.** Plasmids frequently used for expression in the methylotrophic yeast *P. pastoris* include variants of the pPIC family (Life Technologies) that use the yeast’s methanol-inducible AOX1 gene transcriptional promoter. This promoter provides tight regulation of recombinant protein expression based on the choice of carbon source because it is repressed by glucose and glycerol and induced by methanol, which the yeast can metabolize as the sole carbon source; furthermore, such strong control of induction can benefit the
expression of toxic proteins. Clones containing multiple repeat tandem copies of the insertion cassette occur with high frequency and can improve protein expression levels.

The pPICza plasmids encode the secretion signal propeptide sequence for yeast α-mating factor that directs recombinant proteins for secretion into the growth medium, whereas pPICz plasmids lack this sequence and result in intracellular protein accumulation. The α-mating factor propeptide is removed from the N-terminus during secretion by including the sequence of a cleavage target site for the furin-like protease kexin, resulting in production of recombinant protein with a native N-terminus that is present in the culture supernatant.

The pPICz and pPICza plasmids contain a gene for resistance to the antibiotic Zeocin (Life Technologies) that enables selection of transformed *P. pastoris* and *E. coli*. Zeocin selection produces virtually no background of untransformed clones and serves to prevent later contamination of transformant stocks with untransformed cells.

*K. lactis* Expression System. The plasmid pKLAC1 (New England Biolabs) is designed for homologous recombination of multiple repeat tandem copies of the cassette into the genome of *K. lactis* under control of the galactose-inducible LAC4 promoter, also repressed by glucose and glycerol. Galactose has the advantage of being nonflammable, unlike the methanol used for induction of AOX1. As with pPICza, pKLAC1 encodes the yeast α-mating factor signal propeptide sequence that is removed by kexin in the Golgi apparatus and results in secretion of mature proteins with native N-termini.

To isolate transformants, pKLAC1 uses ampicillin antibiotic resistance selection for *E. coli* and acetamide metabolism selection for *K. lactis* (Read et al. 2007). The plasmid transforms *K. lactis* to express the enzyme acetamidase that is required for the metabolism of acetamide as a nitrogen source; consequently, plating on nitrogen-free, yeast carbon base (YCB) agar containing
acetamide selects for growth of transformed clones. This selection process improves the selection of multicopy transformants; however, it can generate a background of colonies that slowly metabolize acetamide and can be difficult to distinguish from true transformants.

**Custom Tailored Gene Synthesis and Codon Usage Optimization**

Traditional methods for cloning typically use reverse transcription polymerase chain reaction (RT-PCR) to create and amplify complementary DNA (cDNA) copies of a naturally occurring, intron-free, mRNA that encodes the protein of interest. PCR primers can be designed to incorporate suitable restriction endonuclease cleavage sites that then facilitate cloning into a chosen vector. In contrast to this approach, custom gene synthesis has many advantages that simplify the cloning process, enhance the malleability of the gene and protein sequences, and improve expression levels of recombinant proteins. Commercially synthesized genes of any nucleotide sequence can be customized with any features such as restriction endonuclease recognition sequences for molecular cloning into vectors; similarly, undesirable restriction endonuclease sites and mRNA secondary structure can be eliminated from the sequence through the use of alternative codons. Custom synthesis facilitates the engineering of multiple-domain fusion proteins, especially those whose component sequences deviate significantly from their naturally occurring mRNA counterparts.

Because not all species express the same levels of the same tRNAs, different organisms translate individual codons with varying efficiencies. This phenomenon of codon usage preference can have considerable impact on protein production in foreign hosts because the codons most prevalent in the recombinant gene may not correspond to the codons used most efficiently by the expression system. This can cause ribosomal stalling and disassembly that
result in incomplete, partially translated proteins and a reduction in the efficiency of protein production.

DNA sequence databases have been analyzed extensively to determine the statistical frequencies with which codons occur within the total population of characterized protein-coding genes for many species (Nakamura et al. 2000). These data are available publicly (http://www.kazusa.or.jp/codon/) (Nakamura et al. 2000) and an internet website (http://gcua.schoedl.de/index.html) (Fuhrmann et al. 2004) facilitates comparisons of codon usage for gene sequences between organisms.

A comparison of the codon usage tables for the relative frequencies (RFs) of several amino acids in *Homo sapiens*, *P. pastoris*, and *K. lactis* reveals some important differences (Figure 1.1). For example, among the 6 available codons for the amino acid Arg, the codons AGA, AGG, CGC, and CGG each appear approximately 20% of the time in human genes (Figure 1.1a); however, the codon AGA is used in 48% and 51% of total Arg codon occurrences in *P. pastoris* and *K. lactis*, respectively, (Figure 1.1b). Also, the sequences CGC and CGG each only occur as 5% of the Arg codons in *P. pastoris* and *K. lactis* although they appear more frequently in human genes. Gln provides another good example in that the codon CAG is preferred, with RF = 73%, in human genes, whereas CAA is the preferred codon, with RF = 61% and 67% in *P. pastoris* and *K. lactis*, respectively. Frequency of codon occurrence may correlate with protein translation efficiency; however, data are lacking that directly characterize the relative frequencies of tRNA expression in different organisms.
**Figure 1.1: Codon Usage Frequency.** Partial comparison shows differences in relative codon usage frequencies between *H. sapiens* (blue), *P. pastoris* (black), and *K. lactis* (red). Data represent the relative frequency (RF) as a percentage of total codon occurrences for the given amino acid in the genome of the species indicated. **A.** *H. sapiens* and *P. pastoris*. **B.** *K. lactis* and *P. pastoris*. Images adapted from [http://gcua.schoedl.de/index.html](http://gcua.schoedl.de/index.html).

Custom DNA synthesis facilitates the optimization of codon usage within a recombinant gene to match the codon preferences of the protein expression system (Figure 1.2). For example, Figure 1.2 shows a codon optimization of the first 50 residues of active human neutrophil elastase (HNE) for *P. pastoris* codon usage frequency tables. Figure 1.2a shows the native human cDNA sequence, by codon, along with the relative adaptiveness (RA) of that codon for expression in *P. pastoris*. Unlike the RF (Figure 1.1), the RA designates a value of 100% for the most frequently used codon for a given amino acid and the RA values for the other analogous
codons are presented as a percentage of the most frequently used codon. The RA measurement scale is preferred over RF for comparing cDNA sequences because it facilitates rapid visual interpretation of the data.

Figure 1.2: Codon Optimization. Codon usage analysis is shown for expression in *P. pastoris* based on relative adaptiveness (RA). The codons are shown for the amino acids Ile16-Ala66 of (A) Native HNE cDNA and (B) Recombinant HNE cDNA that is codon optimized for *P. pastoris*. RA represents the percentage of codon usage relative to the codon most frequently used for the given amino acid by *P. pastoris*. Codons of low occurrence with RA > 15% appear in gray and those with RA > 10% appear in red (A). Codon optimization replaces these codons with sequences more prevalent in the target genome (B). Images adapted from [http://gcua.schoedl.de/index.html](http://gcua.schoedl.de/index.html).

Figure 1.2a shows that the codons used in the native human cDNA are not the codons most preferred by *P. pastoris*, especially in the cases of 4 Arg residues with RA = 10% and six
Ala residues with RA = 13%. A synthetic cDNA for the same amino acid sequence codon-optimized for expression in *P. pastoris* (Figure 1.2b) incorporates codons with much higher RF in the yeast that results in higher RA values in the cDNA. The most preferred codon is not always chosen, however, in order to avoid mRNA self-complementarity that forms stable secondary structures, intron splicing sites, and unwanted internal restriction sites. This process of codon optimization can have robust positive effects on expression levels, although such effects are not always experienced.

**Designing Recombinant Proteins**

Recombinant protein expression technologies facilitate endeavors in protein experimentation and design. Amino acid substitutions by site directed mutagenesis are useful for studying the contributions of individual residues to the overall structural and functional characteristics of a given protein. Fusion proteins can be created by linking multiple distinct protein domains to combine their individual properties into a single molecular chimera, a process whose benefits are limited only by the folding and functionality of the selected domains. Amino acid sequences with known molecular recognition and binding properties can be exploited to affect the functionality of the product.

**Amino Acid Substitutions and Deletions.** Custom gene synthesis and site-directed mutagenesis can be used to selectively replace individual amino acid residues within a recombinant protein. Such substitutions prove advantageous in recombinant protein expression for many reasons. Residues with key roles in enzyme specificity, activity, and other functions can be manipulated individually to determine the effects of their replacement and alter the behaviors of proteins. Similarly, entire regions of native proteins can be deleted to evaluate their
contributions to processes such as protein trafficking, DNA and RNA binding, signaling, and enzymatic activity. Although alterations of amino acid sequence can have unintended consequences, even these results can provide insights into the roles of individual residues in protein behavior.

It has been shown that many proteins of high isoelectric point (pI) do not express well in the yeast *P. pastoris* (Boettner et al. 2007). Often, highly charged proteins are susceptible to proteolysis due to the high content of arginine and lysine residues, which can be substrates for many proteases. Of particular concern for secreted expression in yeast, a high frequency of basic residues increases the likelihood that the dibasic sequences RR~X and KR~X, substrates for the yeast protease kexin, will occur that may lead to unwanted proteolysis during protein secretion. These issues can be ameliorated by the selective substitution of amino acids to eliminate such sites.

Known and potential sites of co- and post-translational modification can also be manipulated or disrupted to evaluate their roles in normal biological processes. This can aid in the evaluation of potential sites of glycosylation, phosphorylation, proteolysis, or other modifications to determine how these processes affect the behavior of a protein. As previously mentioned, many yeast species are prone to over glycosylate recombinant proteins (Niles et al. 1998; van Oort et al. 2002; van Oort et al. 2004; Lockhart et al. 2005; Pepeliaev et al. 2011); furthermore, many biochemical studies require the enzymatic removal of glycosylation. Conservative amino acid substitution to eliminate glycosylation sites can prevent these issues and simplify research efforts.

**Fusion Proteins and Tags.** By combining sequences for different protein domains into a single polypeptide chain, a fusion protein can be expressed that exhibits the properties of its
individual, modular components. Several types of domains are frequently exploited to improve the expression, detection, and purification of recombinant proteins. With some creativity, novel approaches can also be engineered that take advantage of protein properties in other ways through their inclusion in a fusion design.

As mentioned above, the yeast α-mating factor signal sequence is often exploited to direct recombinant proteins for secretion by both *P. pastoris* and *K. lactis* and this can be used to improve expression and recovery of recombinant proteins. Secreted proteins do not require cell lysis prior to purification, nor are they as prone to rapid degradation as proteins expressed inside the cell, although some extracellular proteases are produced. In *E. coli*, recombinant proteins commonly accumulate as insoluble inclusion bodies that require harsh chemical treatments to solubilize and refold, often resulting in poor yields; consequently, highly soluble proteins such as thioredoxin, serum albumin, or small ubiquitin-like modifier (SUMO) protein are sometimes tethered to desired recombinants in efforts to thwart inclusion body formation, thus improving expression and recovery.

Numerous fusion domains are available to facilitate detection of recombinant proteins as well. Several fluorogenic proteins based on the *Aequorea victoria* green fluorescent protein (GFP) (Prasher et al. 1992; Chalfie et al. 1994; Inouye and Tsuji 1994) have been engineered that exhibit fluorescent properties across a spectrum of colors including cyan (Heim and Tsien 1996; Tsien 1998), blue (Heim et al. 1994; Heim and Tsien 1996), green (Yang et al. 1996; Tsien 1998; Zapata-Hommer and Griesbeck 2003), and yellow (Ormö et al. 1996; Wachtler et al. 1998). Fluorescent proteins based on the red fluorescent protein of *Discosoma* species have also been developed in a range of fluorescent colors spanning cyan to far-red (Matz et al. 1999; Campbell et al. 2002; Shaner et al. 2004; Wang et al. 2004; Shaner et al. 2005; Shu et al. 2006; Shaner et
Protein fluorescence is often used to visualize protein localization and transport by fluorescence microscopy; furthermore, studies of transcriptional promoters as well as protein-DNA and protein-protein interactions can be facilitated through the use of fluorescent proteins as reporter genes for experiments in promoter alteration or yeast two-hybridization.

Limited work has also been done regarding the use of chromogenic proteins as fusion partners, although few such proteins are known. Rubredoxin, which binds iron, and the heme-binding protein cytochrome B5 (CytB5) both display a characteristic absorbance of light at 410 nm accompanied by a red coloration and have been used to detect the expression of their recombinant fusions (Kohli and Ostermeier 2003; Finn et al. 2005; Mitra et al. 2005). With a higher extinction coefficient at 410 nm, CytB5 provides a stronger, more easily detectable signal than rubredoxin (Mitra et al. 2005). The flavin mononucleotide (FMN)-binding domain of human cytochrome P450 reductase has also been used for its characteristic yellow color and absorbance at 450 nm (Finn et al. 2005). While the detection of fluorescent proteins requires specialized equipment, chromogenic proteins can be detected visually and by simple spectrophotometry and this makes them useful for following recombinant protein expression and purification.

Antibody epitope short peptide tags like the myc-tag (EQKLISEEDL), the FLAG® tag (DYKDDDDK) (Sigma), and polyhistidine tags that can be added to the N- or C- terminus of a protein see considerable use for detection and they can also be useful for purification. Antibodies are available that recognize each of these sequences and this is especially useful for Western blotting, ELISA, immunohistochemistry (IHC), and immunocytochemistry (ICC). Myc-tag and FLAG® tag monoclonal antibodies can also be used to purify proteins by antibody
affinity chromatography. Polyhistidine tags, most commonly the hexahistidine (6xHis) tag, will bind divalent metal cations like Co$^{2+}$, Cu$^{2+}$, and Ni$^{2+}$ by sharing electrons and this can be exploited to purify polyhistidine tagged proteins by immobilized metal affinity chromatography (IMAC) (Hochuli 1988).

**Designing Internal Proteolytic Sites.** Because many recombinant fusion proteins require the separation of fusion partners by proteolysis for proper function or further study, proteolytic sites can be incorporated; however, few proteases possess the high level of substrate specificity required to prevent unintended proteolytic events. Among the most commonly used proteases for this purpose are coagulation factor Xa, thrombin, tobacco etch virus (TEV) protease, and enteropeptidase. The commonly used thrombin cleavage site is LVPR~GS and the usual TEV protease cleavage sites are ENLYFQ~G and ENLYFQ~S (where “~” represents the cleavage site); however, these approaches leave G or S (TEV protease) or GS (thrombin) at the N-terminus of the C-terminal domain, which can be problematic, especially for the activation of enzymes, such as serine proteases, whose functions require an exposed native N-terminus.

Factor Xa, which turns prothrombin into active thrombin and prefers the sequences IEGR~X and IDGR~X (where X can be any amino acid other than proline), and enteropeptidase (EP, originally called enterokinase), which converts inactive trypsinogen into active trypsin and prefers the sequences DDDDK~X (LaVallie et al. 1993; Lu et al. 1997) and DDDDR~X substrates (Mikhailova et al. 2007; Gasarian et al. 2011; Smith and Johnson 2013), have been nicknamed “restriction proteases” for their extreme substrate specificities, similar to the well-known restriction endonucleases from which the moniker is derived. Unlike thrombin and TEV protease, factor Xa and EP exhibit no preference for amino acids on the P’ side (nomenclature of Schechter and Berger (Schechter and Berger 1967)) of the target peptide bond; consequently, the
specificities of these 2 enzymes can be used to expose the native N-termini of proteins upon the removal of fusion domains.

Although EP has high preference for the DDDDK~X (LaVallie et al. 1993; Lu et al. 1997) and DDDDR~X (Mikhailova et al. 2007; Gasparian et al. 2011) substrate sequences, this protease will cleave other sequences with significantly lower specificity in the absence of a sufficient concentration of the intended target enzyme (Light et al. 1980; Likhareva et al. 2002; Liew et al. 2005; Shahravan et al. 2008; Chen et al. 2012). The enzyme’s preference comes from charge-based attractions to the substrate. The positively charged Lys (K) or Arg (R) in the P1 position is drawn into the negatively charged S1 specificity pocket of EP; furthermore, the 4 negatively charged Asp (D) residues coordinate with positively charged, adjacent residues on the enzyme surface. While EP from all species use a negatively charged Asp 189 (chymotrypsinogen numbering system) at the base of the S1 pocket, species-specific differences in positive charge distribution in the S2 – S4 sites affect the extended substrate specificity. It is possible to use recombinant protein engineering to create modified variants of EP that manipulate this positive charge distribution to alter or improve the enzyme’s substrate specificity (Chun et al. 2011; Ostapchenko et al. 2012; Smith and Johnson 2013).

Chapter 2 presents the recombinant expression of 2 variants of the light chain of human EP with modified substrate specificities. Both variants were secreted by P. pastoris as functional enzymes and purified to homogeneity. One variant, R96Q, reduced the specificity of EP for the extended substrates DDDDK~X and DDDDR~X by reducing the amount of positive charge in the extended binding pocket. The other EP substitution variant, Y174R, improved the enzyme’s specificity significantly for both substrates through the addition of positive charge to improve charge based interactions between the substrates and the extended binding site.
Serine Proteases

Enteropeptidase, factor Xa, and thrombin are members of the chymotrypsin-like S1 clan of serine proteases, characterized by the high similarity of their tertiary structures and by their identical active site architecture. All serine proteases are related through the use of an electron relay mechanism that activates a nucleophilic oxygen atom on the side chain of the eponymous serine residue in the active site. Among the chymotrypsin-like serine proteases, the active site consists of 3 residues: His 57, Asp 102, and Ser 195 (chymotrypsinogen numbering system) (Ressler 1985; Craik et al. 1987; Sprang et al. 1987; Carter and Wells 1988; Polgár 2005; Fuhrmann et al. 2006) that are arranged so that His 57, stabilized by Asp 102, withdraws the H\(^+\) atom from the side chain oxygen of Ser 195. The topology and charge distribution of the substrate binding grooves of serine proteases position selected protein substrates so that the activated serine nucleophile attacks the carbonyl carbon of a target peptide bond to form a tetrahedral intermediate that then undergoes hydrolysis as the enzyme catalyst releases the resulting product fragments and returns to its original nucleophile form.

The roles of serine proteases in numerous biological processes and diseases emphasize their importance in biomedical research. Blood coagulation is initiated by an activation cascade of either 3 (extrinsic pathway) or 6 (intrinsic pathway) serine proteases, resulting in the generation of active thrombin that converts inert fibrinogen into clot-forming fibrin (Davie and Ratnoff 1964; MacFarlane 1964); furthermore, disruptions in this protease activation cascade cause hemophilia. In the duodenum (Hermon-Taylor et al. 1977), enteropeptidase converts trypsinogen into active trypsin, a serine protease with broad specificity for positively charged P1 residues (Schechter-Berger nomenclature) (Schechter and Berger 1967). This initiates a cascade of digestive enzyme activation including the three archetypal S1 clan serine proteases: trypsin,
chymotrypsin, and pancreatic elastase. These 3 proteases digest proteins into single amino acids and dipeptides for intestinal absorption. Deficiencies in enteropeptidase (Haworth et al. 1971; Hadorn et al. 1975) and trypsin (Townes 1965) lead to digestive disorders with symptoms including intestinal malabsorption, hypoproteinemia, and growth failure. These are only 2 examples of the innumerable functions of serine proteases in human physiology and disease.

Neutrophil Serine Proteases of Primary (Azurophil) Granules

Neutrophils are the most abundant type of white blood cell and the human body’s first line of defense against invasion by pathogenic bacteria and fungi (Borregaard and Cowland 1997; Borregaard et al. 2005). Also known as polymophonuclear (PMN) leukocytes, neutrophils engulf and destroy infiltrators by phagocytosis, at which time primary (azurophil) granules fuse with the phagolysosome and release their cytotoxic contents upon the pathogen. The primary granules of neutrophils store large amounts of the enzyme myeloperoxidase (MPO) and the chymotrypsin-like neutrophil serine proteases (NSPs) cathepsin G (CatG), human neutrophil elastase (HNE), and proteinase 3 (Pr3, also called myeloblastin), in addition to several other antimicrobial proteins (Faurschou and Borregaard 2003; Garwicz et al. 2005; Pham 2006).

While the oxidative stress of MPO-derived hypochlorous acid was originally considered the primary means of microbial killing, research has since shown that NSPs of the azurophil granules play important, nonredundant roles in neutrophil lethality (Belaauaj et al. 1998; Belaaouaj et al. 2000; Tkalcevic et al. 2000; Belaaouaj 2002; Reeves et al. 2002; Hirche et al. 2004; López-Boado et al. 2004; Urban et al. 2006); furthermore, the killing functions of these NSPs are not limited to their proteolytic activity, but also a result of abundant positively charged residues. Antimicrobial peptides of strong positive charge are released by the degradation of
NSPs and both the positively charged enzymes and peptides can associate with the negatively charged surfaces of microorganisms and exert cytotoxic effects independent of proteolytic activity (Bangalore et al. 1990; Shafer et al. 1990; Shafer et al. 1991; Miyasaki et al. 1993; Shafer et al. 1993; Shafer et al. 1996; Shafer et al. 2002; Korkmaz et al. 2010). In a recently discovered extracellular killing mechanism, the neutrophil extracellular trap (NET), neutrophils also release a network of enzyme-laden chromatin to ensnare and destroy microbes in the extracellular milieu (Brinkmann et al. 2004; Urban et al. 2006; Wartha et al. 2007; Papayannopoulos et al. 2010; Amulic and Hayes 2011).

CatG, HNE, and Pr3 not only function to directly destroy pathogens but also to affect the inflammatory response through signaling pathways. These enzymes process cytokines (Scuderi et al. 1991; Bank et al. 1999; Benabid et al. 2012), chemokines (Padrines et al. 1994; Nufer et al. 1999; Rao et al. 2004; Berahovich et al. 2005; Ryu et al. 2005; Wittamer et al. 2005), and protease activated receptors (PARs) (Sambrano et al. 2000; Cumashi et al. 2001) with both proinflammatory and antiinflammatory consequences (Pham 2006; Korkmaz et al. 2008; Korkmaz et al. 2010). Although Pr3 and CatG cleave chemokines that promote monocyte chemotaxis (Padrines et al. 1994; Nufer et al. 1999; Berahovich et al. 2005) and HNE and CatG action can cause recruitment of antigen-presenting macrophages and dendritic cells (Wittamer et al. 2005), the net effects of NSPs on inflammation and immune cell recruitment remain unclear because these enzymes also process several antiinflammatory chemokines and cytokines as well. For example, both CatG and HNE inactivate TNFα (Scuderi et al. 1991), whereas Pr3 activates this proinflammatory cytokine (Robache-Gallea et al. 1995); however, CatG, HNE, and Pr3 all inactivate the proinflammatory cytokine IL-6 (Bank et al. 1999).
The NSPs of azurophil granules have also been implicated in several human diseases associated with inflammation and tissue destruction (Korkmaz et al. 2008; Korkmaz et al. 2010). These enzymes have strong proteolytic activities and are present in abundance at sites of inflammation; furthermore, they can destroy host tissues if improperly regulated. They can all cleave collagens, elastin, fibronectin, laminin, and proteoglycans (Starkey 1977; Pipoly and Crouch 1987; Rao et al. 1991; Korkmaz et al. 2008), making them particularly damaging to extracellular matrices and connective tissues, especially in inflammatory lung diseases like acute lung injury, acute respiratory distress syndrome, chronic obstructive pulmonary disorder (COPD), cystic fibrosis, and emphysema (Lee and Downey 2001a; 2001b; Shapiro 2002; Moraes et al. 2003; Shapiro et al. 2003; Moraes et al. 2006; Korkmaz et al. 2008; Owen 2008; Korkmaz et al. 2010).

CatG, HNE, and Pr3 are all synthesized in the promyelocytic stage of early neutrophil maturation as inactive preproproteins and directed to the primary granules (Gullberg et al. 1997; Garwicz et al. 1998; Garwicz et al. 2005). Although the key component of NSP sorting into these granules is the timing of their biosynthesis, the molecular means by which the enzymes are directed to their eventual destinations is unclear (Korkmaz et al. 2008; Korkmaz et al. 2010). All 3 inactive proteins contain N-terminal leader sequences that direct them into the rough endoplasmic reticulum and are removed during posttranslational processing (Salvesen and Enghild 1991; McGuire et al. 1993; Sköld et al. 2002; Korkmaz et al. 2008). The resulting NSP zymogens contain N-terminal prodipeptides that are removed by dipeptidyl peptidase I (DPP I, also called Cathepsin C) inside the granules to expose the native N-terminal Ile 16, required for the activity of all chymotrypsin-like serine proteases (Salvesen and Enghild 1991; McGuire et al. 1993; Sköld et al. 2002; Korkmaz et al. 2008).
In addition to this type of zymogen activation, which is common to all members of the S1 clan, CatG, HNE, and Pr3 also contain unusual C-terminal propeptide extensions (“tails”) of varying lengths and sequences (Gullberg et al. 1995; Gullberg et al. 1997; Capizzi et al. 2003). In an evolutionarily conserved process, these C-terminal tails are removed by an unidentified protease (or proteases); furthermore, the tails are dissimilar and have no known functions. They do not appear to affect the activities of the enzymes, nor are they required for correct intracellular trafficking (Gullberg et al. 1995; Källquist et al. 2010; Korkmaz et al. 2010). An amino acid sequence alignment of CatG, HNE, and Pr3 (Figure 3.1) shows the relative positions of Ile 16, His 57, Asp 102, Ser 195, and the C-terminal extensions.

**Figure 1.3: NSP Sequence Alignment.** Native sequences for human CatG, HNE, and Pr3 are aligned, illustrating similarities and differences. All residues required for catalytic activity appear in green. The N-terminal Ile 16 (▲) must be exposed for catalytic activity. The residues that comprise the catalytic triad (△) include His 57, Asp 102, and Ser 195. C-terminal propeptide extensions appear in **bold with red underline.**
**Human Proteinase 3.** Pr3 can be found not only in the primary granules and in NETs but also in secondary (specific) granules, secretory vesicles, and in association with the neutrophil cell surface (Korkmaz et al. 2008; Korkmaz et al. 2010). This serine protease demonstrates a preference for short hydrophobic amino acid residues, especially Val, Ala, and Thr, in the P1 position; furthermore, endogenous inhibitors of Pr3 include α2-macroglobulin, α1-antitrypsin (α1-AT, also called α1-proteinase inhibitor), and elafin (Korkmaz et al. 2008; Korkmaz et al. 2010). Mature Pr3 contains 222 amino acids (Campanelli et al. 1990), has an isoelectric point (pI) of ~9.5 and it bears a positive charge at physiologically relevant pH due to its high content of arginine (13 residues); consequently, Pr3 has bactericidal properties independent of its proteolytic function, as do HNE and CatG (Bangalore et al. 1990; Shafer et al. 1990; Shafer et al. 1991; Miyasaki et al. 1993; Shafer et al. 1993; Shafer et al. 1996; Shafer et al. 2002; Korkmaz et al. 2010). The C-terminal tail of Pr3, 8 amino acids in length, is the shortest among the azurophil granule NSPs and little is known about its function (Capizzi et al. 2003).

Antineutrophil cytoplasmic antibodies (ANCAs) are responsible for several diseases characterized by autoantibodies against neutrophils accompanied by blood vessel inflammation. Pr3 is the main antigen recognized by cytoplasmic ANCAs (cANCAs) and this phenomenon corresponds with the relatively uncommon disease Wegener’s Granulomatosis (WG), for which the cause is not fully understood (Campanelli et al. 1990; Rao et al. 1991; Specks et al. 1997; Capizzi et al. 2003; Caughey 2006; Korkmaz et al. 2010). The primary symptoms of WG are vasculitis and necrotizing, granulomatous inflammation of small blood vessels (Fauci and Wolff 1973; Sarraf and Sneller 2005); furthermore, it affects approximately 1 in 20,000 people, most between 40 – 60 years of age (Cotch et al. 1996). Secondary symptoms can include cold or sinusitis that is not cured by normal means, inflammation of the ears or eyes, nasal membrane
crusting and ulceration, or saddle-nose deformity (Fauci and Wolff 1973; Sarraf and Sneller 2005; Korkmaz et al. 2010). According to the National Institute of Allergy and Infectious Disease, WG is typically treated with glucocorticoids and cytotoxic compounds (http://www.niaid.nih.gov/topics/wegeners/pages/treatment.aspx).

**Human Neutrophil Elastase.** HNE shares 54% sequence identity with Pr3; however, unlike Pr3, HNE is localized in NETs and the azurophil granules but not in other granules or vesicles, although some HNE is also associated with the neutrophil cell surface and in the nucleus (Pham 2006; Korkmaz et al. 2008; Korkmaz et al. 2010). Similar to Pr3, HNE also prefers small hydrophobic residues at the P1 position, especially Val, Ile, Thr, Ala, and Leu (Bode et al. 1986). Still, HNE and Pr3 do not share extended substrate specificity for residues in other positions; consequently, they process different proteins and play separate roles in neutrophil function. Mature HNE contains 218 amino acids (Bode et al. 1986; Sinha et al. 1987) and has a pI of ~10.8, making it more positively charged than Pr3 under physiological conditions. This strong positive charge results from 19 arginine residues and also gives HNE antibacterial properties independent of catalytic activity (Korkmaz et al. 2010). HNE is inhibited physiologically by the same inhibitors as PR3: α2-macroglobulin, α1-AT, and elafin; however, HNE is also inhibited by secretory leukocyte proteinase inhibitor (SLPI), which does not inhibit Pr3 (Korkmaz et al. 2008; Korkmaz et al. 2010).

The C-terminal tail of HNE is 20 amino acids long, the longest among the 3 azurophil granule NSPs; furthermore, it has been studied more extensively that the tails of Pr3 or CatG because of its potential role in severe congenital neutropenia (SCN), discussed below. Although the C-terminal extension is not required for the granular localization of HNE, the presence of the tail does direct at least a portion of HNE to the cell membrane prior to reuptake of the enzyme by
endocytosis (Tapper et al. 2006); furthermore, blocking the tail of HNE with a fusion extension prevents this reuptake (Källquist et al. 2010). Trafficking of Pr3 does not occur in the same manner (Källquist et al. 2010). The removal of the C-terminal tail exposes a recognition site for the µ3a subunit of the adaptor protein 3 (AP3) complex (Benson et al. 2003). It was previously suggested that the µ3a subunit may bind HNE at this site to direct granular localization (Benson et al. 2003); however, a more recent model suggests that the tetraspanin CD63 associates with HNE to mediate AP3-dependent trafficking (Källquist et al. 2008).

Cyclic neutropenia (CN, also called cyclic hematopoiesis), and SCN are diseases of neutrophil maturation caused primarily by mutations in the HNE gene, ELANE (Horwitz et al. 1999; Dale et al. 2000; Li and Horwitz 2001; Benson et al. 2003; Horwitz et al. 2004; Horwitz et al. 2007; Korkmaz et al. 2010; Horwitz et al. 2013). CN is characterized by neutrophil counts that oscillate between normal and almost zero with a 21-day cycling time (Krance et al. 1982; Lange 1983). In a genomic study of 13 families with a history CN, 7 heterozygous mutations of ELANE are found in all cases (Horwitz et al. 1999). SCN is a disease of neutrophil maturation arrest, with symptoms including chronic low circulating neutrophil counts, structurally abnormal neutrophils with functional deficiencies, and increased susceptibility to infections (Kostmann 1956; 1975; Carlsson and Fasth 2001; Horwitz et al. 2003; Horwitz et al. 2007;). Although other gene mutations have been implicated in some, 21 out of 27 independent cases of SCN show 1 of 15 different heterozygous mutations of ELANE (that sometimes also overlap with mutations that cause CN) (Dale et al. 2000). All chain-terminating mutations of ELANE that result in production of C-terminally truncated HNE, which accompany many cases of SCN but not CN, occur in the fifth (final) exon (Benson et al. 2003; Horwitz et al. 2013).
Two hypotheses have been proposed for how HNE can cause CN and SCN. The mislocalization hypothesis is based on canine cyclic neutropenia, or gray collie syndrome, in which mutations of the beta subunit of AP3 lead to mislocalization of HNE (Benson et al. 2003); however, the human disease most analogous to canine cyclic neutropenia is Hermansky-Pudlak syndrome and not CN or SCN (Horwitz et al. 2013). The second hypothesis suggests that the offending ELANE mutations cause misfolding of HNE and induction of an unfolded protein response that leads to neutrophil apoptosis (Köllner et al. 2006). Several neutropenia-causing mutations of ELANE lead to the cytoplasmic accumulation of unfolded HNE causing apoptosis in transformed U937 cells (Köllner et al. 2006).

**Human Cathepsin G.** CatG appears only in the azurophil granules, on the neutrophil cell surface, and in NETs (Pham 2006; Korkmaz et al. 2008; Korkmaz et al. 2010). The broad, hydrophobic S1 binding pocket of CatG confers a chymotrypsin-like P1 specificity for the hydrophobic aromatic amino acid residues Tyr and Phe; however, a positively charged Glu 226 side chain resides at the base of the specificity pocket that also gives CatG trypsin-like P1 specificity for the positively charged amino acids Lys and Arg (Tanaka et al. 1985; Hof et al. 1996). This dual specificity is unusual among serine proteases and warrants further biochemical research. Mature CatG has an uncommonly high pI of ~12.0 because this protein, 224 amino acids in length (Salvesen et al. 1987; Hof et al. 1996), contains 34 Arg residues, 15.2 % of its total amino acid content. Proteolytically inactive CatG and synthetic peptides derived from positively charged portions of the CatG amino acid sequence have strong antimicrobial properties (Bangalore et al. 1990; Shafer et al. 1990; Shafer et al. 1991; Miyasaki et al. 1993; Shafer et al. 1993; Shafer et al. 1996; Shafer et al. 2002). CatG is endogenously inhibited by α2-
macroglobulin, α1-AT, α1-antichymotrypsin, and SLPI, but not elafin, which inhibits HNE and Pr3 (Korkmaz et al. 2008; Korkmaz et al. 2010).

The C-terminal tail of CatG is 11 – 13 amino acids long and the exact site of its removal is unclear. Among the available crystallographic models of mature native CatG, 3 show the presence of Ser 244 at the C-terminus (PDB ID: 1CGH, 1AU8, and 1T32) while 1 ends at Arg 243 (PDB ID: 1KYN) (Hof et al. 1996; Greco et al. 2002; de Garavilla et al. 2005). It is possible that CatG may remove its own tail by limited autolysis at Arg 243; however, CatG cannot remove the tail by cleaving at Ser 244, which appears to be the more common cleavage site.

Expression of Recombinant Human NSPs. Recombinant human Pr3 has been expressed successfully in several animal cell lines including: U937 cells (Rao et al. 1996; Dublet et al. 2005), Sf9 insect cells (Jenne and Kuhl 2006), RBL cells (Pederzoli et al. 2005), HMC-1 cells (Pederzoli et al. 2005; Jenne and Kuhl 2006), HEK239 cells (Jenne and Kuhl 2006), and COS cells (Källquist et al. 2010). Much of this work has been spurred by the role of Pr3 as the key autoantigen in Wegener’s Granulomatosis (see above). By contrast, HNE has been expressed only in COS cells (Okano et al. 1990; Källquist et al. 2008; Källquist et al. 2010) and in RBL-1 cells (Gullberg et al. 1995; Tapper et al. 2006; Averhoff et al. 2008). For CatG, the recombinant human enzyme has only been expressed in RBL cells or the murine myeloid cell line 32D (Gullberg et al. 1994; Garwicz et al. 1995; Gullberg et al. 1995), while mouse CatG (lacking trypsin-like substrate affinity found in primate and human CatG) and “humanized” mouse CatG (with the mutation A226E that bestows the trypsin-like substrate affinity) have been expressed in adherent Sf9 insect cells (Raymond et al. 2010).

While recombinant expression in mammalian cells has provided much information regarding azurophil granule NSPs, these systems do not produce large (milligram-to-gram)
quantities of protein efficiently, making them impractical for many in vitro biochemical studies.

Active, full-length human Pr3, including the C-terminal tail with an added 6xHis tag for purification, was previously expressed in *P. pastoris* as a secreted protein with a yield of ~17 mg/L (Harmsen et al. 1997); however, similar attempts with HNE only generated 100 µg/L of active enzyme and, separately, 1 mg/L of an enzymatically inactive H57A mutant, with low (5%) recovery yields (Dall'Acqua et al. 1999). Cathepsin G has never been expressed as an active enzyme in any unicellular microorganism and the need persists for robust expression platforms for recombinant HNE and CatG.

Chapter 3 shows the successful expression and partial purification of recombinant HNE and human CatG, both with intact C-terminal tails, by secretion using *P. pastoris* and the AOX1 methanol-inducible promoter. Expression of HNE in *K. lactis*, with limited success, is also reported. Both enzymes are produced as inactive protease zymogens with an N-terminal fusion domain connected by the EP cleavage site DDDDK-I; furthermore, these fusion proteins exhibit proteolytic activities after treatment with EP to expose their native Ile 16 N-termini. The fusion partner linked to each of these enzymes is the soluble heme-binding domain of the red chromogenic protein cytochrome B5 (CytB5) with an N-terminal 6xHis tag for detection and purification. Expression levels of active HNE in *P. pastoris* are improved 10–35 times compared to the previous attempts of others (Dall'Acqua et al. 1999), likely by producing an activable zymogen rather than the potentially toxic active enzyme. Although less than 200 µg of active human CatG was expressed per liter of *P. pastoris* fermentation broth, this is the first reported successful recombinant expression of active human CatG using cells that grow in suspension in minimal growth medium to high culture density. The results of this study suggest that the expression levels of HNE and human CatG in *P. pastoris*, as well as their purification,
can be improved with further efforts and that this yeast makes a suitable platform for the expression of useful quantities of modified NSPs.

**C-Reactive Protein**

Human C-reactive protein (CRP) is a positive acute-phase reactant produced by hepatocytes and the circulating concentration of CRP in the bloodstream increases significantly in response to upregulation of cytokine interleukin-6 (IL-6) as part of the nonspecific acute-phase response (Pepys and Hirschfield 2003; Black et al. 2004). Discovered in 1930 and named for its ability to precipitate C-lipopolysaccharide (Tillett and Francis 1930), which is a bacterial endotoxin of the polysaccharide capsule of *Streptococcus pneumonia*, CRP binds phosphocholine (PCh) that is exposed on the surface of pathogenic bacteria and damaged host cells. The plasma half-life of CRP is about 19 hours (Vigushin et al. 1993); furthermore, long-term elevated levels of CRP reflect an increased risk of atherosclerosis (Chang et al. 2002; Venugopal et al. 2002; Danenberg et al. 2003; Jialal et al. 2004; Paul et al. 2004; Agrawal et al. 2010), cardiovascular disease (Verma and Yeh 2003; Clearfield 2005; Yaron et al. 2006; Yosef-Levi et al. 2007), and diabetes (Pradhan et al. 2001; Dehghan et al. 2007). The American Heart Association and the Centers for Disease Control and Prevention recommend its use as a clinical marker for inflammation and for cardiovascular disease (Kimberly et al. 2003; Pearson et al. 2003; Kilpatrick and Bunk 2009).

Soluble, globular CRP monomers (Figure 1.4) associate via electrostatic interactions into a pentamer (Figure 1.5) with a planar pentagonal quaternary ring structure and a central pore; furthermore, this makes CRP a member of the pentraxin family of acute phase pattern recognition receptors. The CRP pentamer has 2 opposing faces of distinct function. The
recognition face (Figure 1.5a) binds PCh in a Ca\(^{2+}\) dependent manner and each of the 5 subunits contains binding sites for 2 Ca\(^{2+}\) ions that complete another binding site by conferring high affinity for the negative head group of 1 PCh molecule (Thompson et al. 1999). Other possible ligands include chromatin, fibronectin, histones, laminin, phosphoethanolamine, ribonuclear proteins, oxidized LDL (oxLDL) and the oxLDL receptor LOX-1 (Szalai et al. 1999; Black et al. 2003; Fujita et al. 2009; Shih et al. 2009; Singh et al. 2012). The globular head of C1q, a complement component, binds to the central pore of CRP on the effector face (Figure 1.5b) (Gaboriaud et al. 2003).

![Recognition Face and Side View](image)

**Figure 1.4: CRP Monomer.** The tertiary backbone structure of the globular CRP monomer is shown from (A) the recognition face or (B) a side view. On the backbone ribbon, beta sheet structures appear in aquamarine and alpha helices appear in red. Ca\(^{2+}\) ions and PCh molecules appear as space-filling models colored by atom where calcium = green, carbon = gray, nitrogen = blue (obscured), phosphate = orange, and oxygen = red. Images were generated with Accelrys Discovery Studio 3.1 using the crystal structure of pentameric CRP (PDB ID: 1B09).
Figure 1.5: CRP Pentamer. The pentameric form of CRP is shown from (A) the recognition face and (B) the effector face. Backbone ribbon structures appear inside space-filling models that illustrate the distribution of charge, where positive charge appears in blue and negative charge appears in red. Ca$^{2+}$ ions appear in yellow and PCh molecules appear in green, with the phosphate head groups of PCh oriented toward the Ca$^{2+}$ ions. Images were generated with Accelrys Discovery Studio 3.1 using the crystal structure of pentameric CRP (PDB ID: 1B09).

Pentameric CRP forms 2-dimensional arrays bound to PCh on the cell surface and these arrays bind the globular head domains of C1q in a fashion similar to C1q cross-linking to Fc domains of immune complexes; furthermore, C1q cross-linking with either Fc or CRP can activate the complement cascade. CRP-mediated C1q recruitment and complement activation depend only on the exposure of PCh instead of antigen recognition, although only complement components C1-C4, but not C5-C9, are activated in response to CRP (Thompson et al. 1999; Black et al. 2004). Complement components C1-C4 form the classical pathway of complement activation and result in macrophage phagocytosis of opsonized target cells. Because CRP-
mediated complement activation does not activate C5-C9, membrane attack complexes (MACs) do not assemble or cause target cell lysis.

CRP has other immune-related functions in addition to the activation of complement, with both pro- and anti-inflammatory results; however, the net in vivo effects of CRP on inflammation are unclear and may be context dependent (Tilg et al. 1993; Heuertz et al. 1994; Ahmed et al. 1996; Szalai et al. 1999; Mold et al. 2002; Black et al. 2004). CRP can be cleaved by HNE into protein fragments that induce neutrophil apoptosis in vitro (Kakuta et al. 2006) and this may help to limit neutrophil longevity and activity at sites of inflammation. Human CRP, either administered or expressed in recombinant form, shows protective effects in experiments using murine models of lung inflammatory diseases (modeled by induced alveolitis) (Heuertz et al. 1994; Ahmed et al. 1996), multiple sclerosis (modeled by mouse allergic encephalomyelitis) (Szalai et al. 2002), and systemic lupus erythematosus (SLE) (Du Clos et al. 1994; Szalai et al. 2003; Russell et al. 2004).

Much research suggests that CRP, which is known to localize in atherosclerotic lesions, may have functions related to atherosclerosis; however, the net effects on lesion and plaque formation are uncertain. While CRP does not normally bind native LDL (Chang et al. 2002; Taskinen et al. 2002; van Tits et al. 2005), it has been shown, in vitro, to bind immobile aggregates of LDL (de Beer et al. 1982; Fu and Borensztajn 2002; Agrawal et al. 2010) as well as LDL that has been either enzymatically modified or chemically oxidized into proatherosclerotic forms (Bhakdi et al. 1999; Agrawal et al. 2010; Chang et al. 2002). CRP binds enzymatically modified LDL at neutral pH and this affinity increases with increasing acidity (Bhakdi et al. 1999; Taskinen et al. 2002; Biró et al. 2007; Singh et al. 2008a). Controversy exists regarding the affinity of CRP for oxidized LDL (oxLDL). While it has been
shown that CRP binds oxLDL under acidic conditions (Singh et al. 2009) and some studies
demonstrate the same at physiological pH (Chang et al. 2002; van Tits et al. 2005), other work
indicates that CRP and oxLDL do not readily associate in a physiologically relevant environment
(Taskinen et al. 2002; Ji et al. 2006).

Some evidence indicates that CRP may opsonize LDL particles, resulting in the
phagocytosis of LDL by macrophages, which turns them into plaque-forming foam cells (Bhakdi
et al. 1999; Zwaka et al. 2001). Several other studies also support the hypothesis that CRP
enhances LDL uptake by macrophages and foam cell formation (Fu and Borensztajn 2002;
Verma et al. 2002; van Tits et al. 2005; Singh et al. 2008b). In other studies whose results
suggest that CRP may actually prevent foam cell formation, CRP-LDL complexes did not
convert macrophages into foam cells (Singh et al. 2008a) and the presence of CRP increased
LDL degradation by macrophages (Mookerjea et al. 1994). Unfortunately, the effects of CRP on
atherosclerosis have not been elucidated in vivo because CRP displays inconsistent effects on
disease progression in studies of animal models of atherosclerosis (Paul et al. 2004; Hirschfield
Torzewski et al. 2008; Agrawal et al. 2010).

It is proposed that increased acidity causes conformational changes in the CRP pentamer
that enable it to bind oxLDL and that this may have in vivo significance in the microenvironment
of atherosclerotic lesions (Singh et al. 2012). Similarly, researchers also suggest that the failure
of animal models of atherosclerosis to yield significant results in response to human CRP is due
to species-specific differences in this microenvironment (Singh et al. 2012). Recombinant CRP
E42Q, a single amino acid substitution variant, has been expressed in Chinese hamster ovary
(CHO) cells and found to bind oxLDL with improved affinity at higher pH, likely by increasing
pentamer flexibility due to the disruption of the salt bridge between Glu 42 of each CRP subunit and Lys 119 of the adjacent subunit (Singh et al. 2012). Because of the additional flexibility and improved oxLDL affinity, the E42Q variant may provide a useful tool to overcome the current issues with in vivo studies of the effects of CRP-LDL complex formation on atherosclerosis in animal models. This could help determine the net effects of CRP on atherosclerotic progression. If net protective roles against plaque formation are discovered, then recombinant CRP variants may even have therapeutic value in the treatment of atherosclerosis.

Because CHO cells suffer from many of the drawbacks of other mammalian recombinant expression platforms, production times and total yields of CRP E42Q may be insufficient to provide enough protein for studies in animal models. By producing CRP and variants like CRP E42Q in a recombinant protein expression system, larger quantities of these proteins may be generated more rapidly and with less cost, thus facilitating further studies of CRP binding interactions with LDL and other clinically relevant substrates. Human CRP has been expressed in large quantities in E. coli and found to be functionally indistinguishable from native CRP (Tanaka et al. 2002); however, purified recombinant CRP from E. coli can retain endotoxin contaminants unsuitable for in vivo animal studies. The human protein was recently also expressed in P. pastoris for mass spectrometry studies; however, it was found to contain the additional N-terminal amino acid sequence EAEA, a result of incomplete processing to remove α-mating factor at the kexin cleavage site (Kilpatrick et al. 2012). While these additional amino acids did not adversely impact their efforts to improve standardization of CRP assays, the use of recombinant CRP for animal studies of atherosclerosis requires proteins without such inadvertent differences.
Chapter 4 shows the use of *P. pastoris* for secreted expression of human wild type CRP as a correctly folded, functional pentamer with yields in excess of 3.5 mg/L in growth media. The substitution variant CRP E42Q was also expressed, at over 1.5 mg/L. Recombinant CRP and CRP E42Q were identified and quantified by ELISA using anti-CRP antibodies. Wild type CRP and E42Q were PCh affinity purified for comparison to native human CRP. Purified wild type CRP was characterized by SDS-PAGE under reducing conditions and found to exist in 2 distinct monomeric forms of slightly different mass. These could not be separated from each other despite the use of PCh affinity chromatography, anion exchange, and gel filtration and both forms were labeled by CRP-specific antibodies on a Western blot. This result is possibly an anomaly of the specific yeast clone that was selected during the screening process and that other transformants containing the same insertion cassette may express CRP in only the correct form. Further investigation is warranted to solve these problems and generate recombinant CRP suitable for *in vivo* animal studies.

**Specific Aims**

**Aim 1:** Express functional recombinant human enteropeptidase light chain (hEP*L*) with amino acid substitutions to characterize and modify the contributions of residues 96 and 174 to the extended substrate specificity of the enzyme.

**Aim 2:** Express the functional recombinant neutrophil serine proteases (NSPs) human neutrophil elastase (HNE) and human neutrophil cathepsin G (CatG) with intact C-terminal peptide tails.

**Aim 3:** Express the functional pentamer of native human C-reactive protein (CRP) and the substitution variant CRP E42Q.
CHAPTER 2

HUMAN ENTEROPEPTIDASE LIGHT CHAIN:
BIOENGINEERING OF RECOMBINANTS AND KINETIC INVESTIGATIONS OF
STRUCTURE AND FUNCTION

Eliot T. Smith and David A. Johnson*

Department of Biomedical Sciences, James H. Quillen College of Medicine, East Tennessee
State University, Johnson City TN 37614

* Correspondence to: David A. Johnson, Department of Biomedical Sciences, James H. Quillen
College of Medicine, ETSU, Box 70582, Johnson City, TN 37614.
Phone: (423) 439-2027. Fax: (423) 439-2030. E-mail: davidj@etsu.edu

Keywords: Enteropeptidase, Enterokinase, Recombinant Protein, Kinetics, Extended Peptide
Substrates; Pichia pastoris.

Protein Science 22 (2013) 577-585
Abstract

The serine protease enteropeptidase exhibits a high level of substrate specificity for the cleavage sequence DDDDK~X, making this enzyme a useful tool for the separation of recombinant protein fusion domains. In an effort to improve the utility of enteropeptidase for processing fusion proteins and to better understand its structure and function, two substitution variants of human enteropeptidase, designated R96Q and Y174R, were created and produced as active (>92%) enzymes secreted by *P. pastoris* with yields in excess of 1.7 mg/Liter. The Y174R variant showed improved specificities for substrates containing the sequences DDDDK (\(k_{\text{cat}}/K_M = 6.83 \times 10^6 \text{ M}^{-1} \text{ sec}^{-1}\)) and DDDDR (\(k_{\text{cat}}/K_M = 1.89 \times 10^7 \text{ M}^{-1} \text{ sec}^{-1}\)) relative to all other enteropeptidase variants reported to date. BPTI inhibition of Y174R was significantly decreased. Kinetic data demonstrate the important contribution of the positively charged residue 96 to extended substrate specificity in human enteropeptidase. Modeling shows the importance of the charge-charge interactions in the extended substrate binding pocket.
Introduction

Enteropeptidase (EC 3.4.21.9), originally named enterokinase, is a trypsin-like serine protease that activates pancreatic digestive enzymes in the duodenum\(^1-7\). The proteolytically active light chain has high selectivity for the sequence DDDDK (D\(\text{K}\))\(^8,9\). Physiologically, enteropeptidase removes the N-terminal pro-peptide of trypsinogen by cleaving the sequence D\(\text{K}\)-IVGG to activate trypsin. Genetic deficiencies of enteropeptidase have been reported in two families with symptoms of intestinal malabsorption, diarrhea, and growth failure, revealing the importance of human enteropeptidase (hEP) in the digestive process\(^10,11\). Inhibitors of enteropeptidase have been designed and tested in a murine model to evaluate their efficacy as anti-obesity drugs\(^12\).

Although some non-specific cleavages have been reported for this enzyme\(^6,13-16\), the high selectivity of enteropeptidase light chain (EP\(_L\)) for the sequence D\(\text{K}\)-X makes it useful as a biotechnological tool for site-specific separation of recombinant fusion protein domains\(^17,18\). Recombinant bovine enteropeptidase light chain (rbEP\(_L\)) has been expressed in Pichia pastoris\(^19,20\), Saccharomyces cerevisiae\(^21\), and Escherichia coli\(^18,22,23\). The monomeric rbEP\(_L\) from E. coli proved superior to the native heterodimeric bEP at cleaving fusion proteins\(^18\).

Human EP\(_L\) has greater catalytic efficiency compared to bovine EP\(_L\) for D\(\text{K}\)-R-X and D\(\text{K}\)-X substrates\(^24,25\) and recombinant human enteropeptidase light chain (rhEP\(_L\)) was shown to have a 10-fold higher specificity constant (k\(_{\text{cat}}/K_c\)) for the D\(\text{K}\)-X sequence than bEP\(_L\)\(^26\).

The x-ray crystal structure (PDB code: 1EKB) for bEP\(_L\) in complex with the inhibitor VD\(\text{K}\)-chloromethane was used in conjunction with rbEP\(_L\) variants from a baculovirus expression system to reveal the critical role of K99 (chymotrypsinogen numbering system) in the S2-S4 (Schechter-Berger system)\(^27\) extended binding site that interacts with Asp P2-P4 substrate
residues\textsuperscript{27,28}. A recent study used the human-to-bovine substitutions R96K and K219Q to shift the catalytic properties of rhEP\textsubscript{L} to more closely parallel bEP\textsubscript{L}, demonstrating additive contributions of positive residues R96 and K219 to improved extended specificity\textsuperscript{29}. They constructed in silico models of hEP\textsubscript{L} based on the bEP\textsubscript{L} x-ray structure\textsuperscript{28} and proposed that R96 may shield the P2 residue from the solvent better than K96 and that K219 may form salt bridges with Asp-P3 and Asp-P5\textsuperscript{29}. Others created, in E. coli, a Y174R variant of bEP\textsubscript{L} that showed an approximate four-fold improvement in specificity toward the GD\textsubscript{4}K~IVGG substrate due to proposed salt bridges by R174 with Asp-P3 and Asp-P4\textsuperscript{22}. Recently, a supercharged rhEP\textsubscript{L} (rhEP\textsubscript{L,sc}) was created in E. coli to improve enzyme solubility and refolding yields\textsuperscript{30,31} and an x-ray crystal structure of this mutant was obtained (PDB code: 4DGJ)\textsuperscript{28,30,31}.

In an effort to improve the utility of rhEP\textsubscript{L} for processing fusion proteins and to better understand the structure and function of hEP\textsubscript{L}, two rhEP\textsubscript{L} variants were created and produced as active enzymes secreted by P. pastoris. Kinetic analyses employed synthetic D\textsubscript{4}K and D\textsubscript{4}R p-nitroanilide (D\textsubscript{4}K-pNA and D\textsubscript{4}R-pNA) substrates and the N-\alpha-carbobenzyloxy-L-lysine-thiobenzyl ester (Z-Lys-SBzl) substrate for comparison with other reports. Potential Asn-linked glycosylation sites were mutated to Gln to yield non-glycosylated proteins. In one variant, arginine was replaced with glutamine at position 96 to create R96Q; furthermore, the kinetic properties of this variant in comparison with others suggests that the human R96 residue influences extended substrate affinity through charge-dependent interaction with Asp-P2. In the second variant (Y174R), tyrosine at residue 174 was replaced with arginine, which has been shown to improve the specificity in rbEP\textsubscript{L}\textsuperscript{22}. The human Y174R enzyme shows the highest specificity for the Asp\textsubscript{4} motif of any EP\textsubscript{L} variant reported to date.
Results

Protein Engineering and Molecular Cloning

The amino acid sequences for the variants R96Q and Y174R were designed based on the sequence of native hEP$_L$ (Figure 2.1). Due to the tendency of $P$. pastoris to hyper-glycosylate secreted proteins$^{32-36}$, the conservative amino acid substitutions N75Q, N113Q, N135Q, and N175Q were used to disrupt four potential Asn-linked glycosylation sites in both R96Q and Y174R (Figure 2.1). The Furin-like protease kexin (Kex2) of the Golgi apparatus in $P$. pastoris can cleave dibasic sequences such as KR~X$^{37}$. The substitution R98Q was incorporated into both R96Q and Y174R variants to eliminate a potential internal kexin cleavage site (Figure 2.1), supported by data that an R98A substitution does not reduce the activity of bEP$_L$$^{28}$. DNA coding for the human R96Q and Y174R mutants was commercially codon-optimized for expression in $P$. pastoris, synthesized, and cloned into the pPICza A plasmid for secreted expression.

Plasmids encoding R96Q or Y174R were used to transform the X-33 strain of $P$. pastoris and the most productive clones were identified in a screening protocol. During secretion, the Kexin protease removes $\alpha$-mating factor via cleavage of the sequence KR~IVGG, resulting in secretion of active R96Q or Y174R variants. Approximately 1.7 mg/L of active R96Q and 2.2 mg/L of active Y174R, as estimated by Z-Lys-SBzl activity, were expressed in 5 L batches of bioreactor fermentation. Untransformed X-33 $P$. pastoris do not secrete enzymatic activity for the Z-Lys-SBzl substrate.
**Figure 2.1: EP\textsubscript{L} Sequence Alignment.** Human and bovine EP\textsubscript{L} aligned with human R96Q and Y174R variants. All engineered substitutions appear in bold and residues 96 and 174 are labeled with stars. Asn-to-Gln substitutions (bent arrows) disrupt four potential Asn-linked glycosylation sites. The substitution R98Q (dot) eliminates one possible kexin cleavage site. Serine protease active site residues (triangles) are also indicated.

**Purification**

Both enzymes were recovered from culture media by affinity chromatography using the reversible, Kunitz-type protease inhibitor, Soybean Trypsin Inhibitor (STI). The R96Q and Y174R variants were purified 1352-fold and 969-fold, respectively, from fermentation media (Table 2.1). Although the total protein recoveries of purified R96Q and Y174R were similar, the total and specific activities were unexpectedly low for Y174R (Table 2.1). Assays during
purification were in the absence of CaCl$_2$ or Triton X-100, which were subsequently found to enhance the activity of Y174R consistent with a report on solubility problems with recombinant hEP$^{30}$.

<table>
<thead>
<tr>
<th>Purification Step</th>
<th>Total Protein (mg)</th>
<th>Total Activity (U)</th>
<th>Specific Activity (U/mg)</th>
<th>Yield</th>
<th>Fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>R96Q Medium</td>
<td>2294</td>
<td>3.32 x 10$^7$</td>
<td>1.45 x 10$^4$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>R96Q Eluate</td>
<td>1.21</td>
<td>2.35 x 10$^7$</td>
<td>1.96 x 10$^7$</td>
<td>71%</td>
<td>1352</td>
</tr>
<tr>
<td>Y174R Medium</td>
<td>2157</td>
<td>6.98 x 10$^6$</td>
<td>3.24 x 10$^3$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Y174R Eluate</td>
<td>1.18</td>
<td>3.70 x 10$^6$</td>
<td>3.14 x 10$^6$</td>
<td>53%</td>
<td>969</td>
</tr>
</tbody>
</table>

Total Protein (mg) estimates based on $\varepsilon_{0.1\%} = 2.11$ A280.
Units (U) are mA410/min from assays in 300 μM Z-Lys-SBzl, 100 mM Tris HCl, 1 mM DTNB, pH = 8.0.

Subsequent kinetic analyses including 0.02 mM CaCl$_2$ and 0.01% Triton X-100 show that both variants have similar $K_M$ and $k_{cat}$ values for Z-Lys-SBzl (Table 2.2), suggesting that Y174R is less soluble than R96Q. The single-step STI affinity purification yielded highly purified enzymes (Figure 2.2a) and the A280-based concentration estimates were confirmed to within 8% by active site titration$^{38}$.

Identity Confirmation and Stability Analyses

The purified recombinant enzymes were analyzed by SDS-PAGE (Figure 2.2a), confirming high levels of purity at the predicted molecular weight of 26 kDa for non-glycosylated rhEP$_L$. Western blot analysis (Figure 2.2b) further confirms the identities of the rhEP$_L$ variants using a monoclonal antibody against bEP$_L$ with non-glycosylated rbEP$_L$ as a positive control. Both substitution variants selectively cut an enteropeptidase cleavage control protein (CCP; EMD Millipore) with the same cleavage pattern as observed for Tag*off™
positive control enzyme (Figure 2.2c). Tag*off™ is a non-glycosylated recombinant hEP\textsubscript{L} (EMD Millipore).

Over the course of five days, Tag*off™, R96Q, and Y174R were evaluated for stability at 37 °C in 100 mM Tris, 0.02 mM CaCl\textsubscript{2}, 0.01% Triton X-100, pH = 8.0. Tag*off™ and R96Q had half-lives longer than one week (16 days and 9 days, respectively); however, the Y174R substitution reduced the half-life to 48 hours (Figure 2.3). Residue R174 may provide a target for EP\textsubscript{L} autolysis. All of the enzymes showed long-term stability in low pH storage buffers at +4 °C and -20 °C.

**Figure 2.2: Molecular Characterization of Purified EP\textsubscript{L} Variants.** (A) SDS-PAGE shows high purity and correct mass (~26 kDa) for non-glycosylated enteropeptidase variants R96Q and Y174R. (B) Monoclonal antibody against bEPL also labels human variants R96Q and Y174R by Western blot. (C) Cleavage assay shows that R96Q and Y174R variants cleave the CCP like Tag*Off™.
**Figure 2.3: Enzyme Stability Assay.** Tag*Off™ (diamonds) has an estimated half-life of 16 days. R96Q (squares) has an estimated half-life of 9 days. Y174R (circles) has an estimated half-life of 48 hours. Error bars show standard deviation of triplicate data. Regression analyses were fit to the equation $\frac{1}{2} = e^{-kt}$.

**Kinetic Analyses**

The Michaelis Constant $K_M$, the catalytic rate constant $k_{cat}$, and the specificity constant $k_{cat}/K_M$ were evaluated for Tag*off™, R96Q, and Y174R using the synthetic peptide substrates Z-Lys-SBzl, GD₄K-pNA, and GD₄R-pNA (Table 2.2). These assays were conducted in 100mM Tris, 0.02 mM CaCl₂, 0.01% Triton X-100, 10% DMSO, pH = 8.0. CaCl₂ enhances EP₄ activity at the 0.02 mM concentration. NaCl was excluded because it has been shown to reduce EP₄ activity for GD₄K-na. Due to the limited supply of Tag*off™, the stock concentration of active Tag*off™ could not be determined by active site titration; consequently, all Tag*off™ $k_{cat}$ values reported here are estimates based on the averages of the values for the R96Q and Y174R variants, which are very similar (Table 2.2); furthermore, the estimate reported here for
Tag*off™ with GD4K-pNA is similar to the $k_{\text{cat}}$ reported elsewhere for hEP_L with the substrate GD$_4$K-nan$_2$\textsuperscript{26}.

<table>
<thead>
<tr>
<th>Table 2.2: Kinetic Parameters of rhEP_L Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (M)</td>
</tr>
<tr>
<td>Z-Lys-SBzl</td>
</tr>
<tr>
<td>Tag*off™</td>
</tr>
<tr>
<td>R96Q</td>
</tr>
<tr>
<td>Y174R</td>
</tr>
<tr>
<td>GD$_4$K-pNA</td>
</tr>
<tr>
<td>Tag*off™</td>
</tr>
<tr>
<td>R96Q</td>
</tr>
<tr>
<td>Y174R</td>
</tr>
<tr>
<td>GD$_4$R-pNA</td>
</tr>
<tr>
<td>Tag*off™</td>
</tr>
<tr>
<td>R96Q</td>
</tr>
<tr>
<td>Y174R</td>
</tr>
</tbody>
</table>

Values are expressed as mean ± S.D. of three independent replicates.
Assays conducted in 100 mM Tris, 0.02 mM CaCl$_2$, 0.01% Triton X-100, 10% DMSO, pH = 8.0.
* Active site concentrations of Tag*off™ could not be determined by titration due to limited supply. The parenthetical $k_{\text{cat}}$ values are taken from the average of the $k_{\text{cat}}$ for R96Q and Y174.

For the single residue substrate Z-Lys-SBzl, no statistically significant differences were observed among the $K_M$ values; likewise, the $k_{\text{cat}}$ value for R96Q is only 1.1 times that for Y174R ($P<0.05$) (Table 2.2). The comparison of $k_{\text{cat}}/K_M$ ratios reveals that R96Q cleaves Z-Lys-SBzl about 1.2 times as efficiently as either Tag*off™ ($P<0.05$, estimated) or Y174R ($P<0.05$) (Table 2.2). The Y174R variant and Tag*off™ do not have statistically different $k_{\text{cat}}/K_M$ values.

For the substrate GD$_4$K-pNA, the enzymes behave very differently. The Y174R variant shows the smallest $K_M$ for this substrate (Table 2.2). The $K_M$ for Tag*off™ is 1.6 times larger than for Y174R ($P<0.01$) (Table 2.2). With the weakest affinity for GD$_4$K-pNA, R96Q has a $K_M$ that is 9.3-fold greater than Tag*off™ ($P<0.0005$) and 15.5-fold greater than the Y174R variant.
The $k_{\text{cat}}$ of R96Q was 1.2-fold that of Y174R (P<0.005) (Table 2.2). The single amino acid substitution Y174R improved the catalytic efficiency for the GD$_4$K-pNA substrate by 1.5-fold compared to Tag*$\text{off}^{\text{TM}}$ (P<0.01, estimated) (Table 2.2, Figure 2.4). The specificity of R96Q for D$_4$K-pNA showed only 12% of the value for Tag*$\text{off}^{\text{TM}}$ (P<0.0001, estimated) and only 8% of the Y174R result (P<0.0001) (Table 2.2, Figure 2.4). Compared to 100 mM Tris HCl, pH = 8.0, the addition of 0.02 mM CaCl$_2$ and 0.01% Triton X-100 improved the specificity constants of R96Q and Y174R by 6 times and 20 times, respectively, for GD$_4$K-pNA.

![Figure 2.4: Specificity Constants.](image)

Figure 2.4: Specificity Constants. Specificity constants of Tag*$\text{off}^{\text{TM}}$, R96Q, and Y174R for the extended substrates GD$_4$K-pNA and GD$_4$R-pNA are shown. All values are mean average ± standard deviation, as listed in Table 2.2. P < 0.005 for all * and ** comparisons.

$K_M$ values for GD$_4$R-pNA were lowest for Y174R, followed by Tag*$\text{off}^{\text{TM}}$ at 2.4 times Y174R (P<0.005) and then R96Q at 6.9 times Y174R (P<0.001) (Table 2.2). As with the other two synthetic substrates, the $k_{\text{cat}}$ values of R96Q and Y174R are very similar. The $k_{\text{cat}}/K_M$
specificity constant of Y174R for the GD4R-pNA substrate is 2.6 times that of Tag*off™ (P<0.005, estimated) and 8-fold that of R96Q (P<0.005) (Table 2.2, Figure 2.4). The kcat/KM for Tag*off™ is 3 times that of R96Q (P<0.005, estimated) (Table 2.2, Figure 2.4). Compared to 100 mM Tris HCl, pH = 8.0, the addition of 0.02 mM CaCl2 and 0.01% Triton X-100 improved the specificity constants of R96Q and Y174R for GD4R-pNA by 6.8 times and 28 times, respectively.

Comparison of the relative efficiencies of each enzyme for the extended peptide synthetic substrates reveals that the GD4R-pNA sequence is preferred over GD4K-pNA by all variants, especially for Y174R, by 2.8 times (P<0.01) (Table 2.2, Figure 2.4). The Tag*off™ enzyme only showed 1.3 times greater selectivity for GD4R-pNA (P<0.005) compared with GD4K-pNA, although the KM was reduced by 42% (Table 2.2, Figure 2.4). While R96Q had the lowest specificity constants with either of the extended peptide substrates, its preference for GD4R-pNA was 4.3 times that of GD4K-pNA (P<0.005) (Table 2.2, Figure 2.4).

The substitution Y174R significantly decreases the binding affinity for the inhibitor BPTI (Table 2.3). The inhibition equilibrium constant, Ki, is 4 times larger for Y174R (P<0.0001) than for Tag*off™ and 3.6 times larger (P<0.0001) than for R96Q (Table 2.3). The R96Q substitution did not significantly change the Ki from that of Tag*off™ (Table 2.3).

**Table 2.3: Inhibition of rhEP_L by BPTI**

<table>
<thead>
<tr>
<th></th>
<th>K_i (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag*off™</td>
<td>4.61 x 10^-8 ± 1.66 x 10^-8</td>
</tr>
<tr>
<td>R96Q</td>
<td>6.32 x 10^-8 ± 1.84 x 10^-9</td>
</tr>
<tr>
<td>Y174R</td>
<td>2.28 x 10^-7 ± 7.02 x 10^-9</td>
</tr>
</tbody>
</table>

Values are expressed as mean ± S.D. of three independent replicates. Assays conducted in 300 µM Z-Lys-SBzl, 100 mM Tris, 0.02 mM CaCl2, 0.01% Triton X-100, 10% DMSO, 1 mM DTNB, pH = 8.0.
Discussion

The extended substrate binding pocket of EP_L can be thought of as a five-walled cage where residues 96, 99, 174, and 219 form walls 1-4 (Figure 2.5), with the active site (S195) and the S1 specificity pocket (D189) forming the floor. The open fifth wall and the open ceiling of the cage allow the substrate to enter. In all variants, “wall 2” (K99) and the “floor” are required for D_4K specificity. Added specificity comes from the properties of the other walls. In bEP_L, “wall 1” is K96, “wall 3” is Y174, and “wall 4” is Q219 (Figure 2.5a). The positive charge of K96 has been shown to improve specificity; however, Y174 contributes only a weak hydrogen bond between the tyrosine hydroxyl group and the carboxyl side chain of Asp-P3. In the study of the E. coli variant rbEP_L Y174R (Figure 2.5b), the presence of the ε-guanidino group of R174 on “wall 3” of the cage was shown to improve extended substrate specificity and was proposed to form a salt bridge with Asp-P3. According to the x-ray crystal structure of bEP_L (Figure 2.5a), Q219 side chain does not interact with the substrate in any way, extending away from the binding site instead.
Figure 2.5: Electrostatic Models of EP<sub>L</sub> Extended Binding Pockets. Proposed electrostatic models show EPL binding pockets with bound D3K substrate. Positively charged (blue) residues (green labels) improve binding affinity for Asp P2-P4. Negative charges are red. (A) Bovine EPL (PDB Code: 1EKB). (B) Bovine EPL with Y174R substitution. (C) Human EPL (by substitution of PDB model 4DGJ). (D) Human EPL with Y174R substitution. (E) Supercharged human EPL (PDB code: 4DGJ). (F) Human EPL with R96Q substitution.
The improved specificity of hEP\textsubscript{L} over bEP\textsubscript{L} can be attributed in part to differences in two of the cage walls (Figure 2.5a and 2.5c) and in part to species-specific differences in the stability of the active site and specificity pocket\textsuperscript{29}. First, amino acid substitution by others has shown that R96 stabilizes Asp-P2 better than K96\textsuperscript{29}. Second, another substitution by the same group has also shown that K219 improves the substrate specificity of rhEP\textsubscript{L} compared to Q219, especially in conjunction with R96\textsuperscript{29}. The positively charged K219 side chain amino group may by attracted toward the binding site in the presence of Asp-P3 in hEP\textsubscript{L}\textsuperscript{29} to form “wall 4” (Figure 2.5c, 2.5d, and 2.5f) whereas Q219 and D219 side chains extend into the solvent in the crystal structures of bEP\textsubscript{L}\textsuperscript{28} (Figure 2.5a and 2.5b) and rhEP\textsubscript{L,sc}\textsuperscript{34} (Figure 2.5e).

The human Y174R variant presented here combines the improved specificity of the bovine Y174R substitution\textsuperscript{22} with the positive influences of residues R96 and K219\textsuperscript{29} in the human enzyme. The result is the novel human Y174R (Figure 2.5d) with improved extended substrate specificity relative to previously reported rEP\textsubscript{L} variants. This Y174R enzyme may prove advantageous for the separation of recombinant protein fusion constructs. Although Y174R has only a slightly better specificity than Tag*off\textsuperscript{TM} for the D\textsubscript{4}K sequence, Y174R displays a very high specificity constant for the D\textsubscript{4}R sequence (Table 2.2, Figure 2.4). This reduces nonspecific cleavage by Y174R in the presence of excess fusion protein if the D\textsubscript{4}R linker is employed. The shorter enzyme half-life of Y174R may be due to autolysis at R174. Although poor stability is typically undesirable in enzymes, the limited half-life of Y174R under catalytic conditions (pH = 8.0, 37 °C) may actually contribute to a reduction of non-specific substrate cleavage relative to native EP\textsubscript{L}.

The R96Q variant generated during this project reveals that the absence of positive charge at position 96 reduces the catalytic efficiency by 94% relative to Tag*off\textsuperscript{TM} for D\textsubscript{4}K and
by 68% relative to Tag*off™ for D₄R (Table 2.2, Figure 2.4). Others have shown that the K96A substitution in rbEPₜ reduces its specificity by approximately 80% of native bEPₜ for GD₄K- na²⁸. It has also been previously shown that the human-to-bovine substitution rhEPₜ R96K reduces its specificity by only about 6% relative to rhEPₜ²⁹. These data suggest that the positive charge of residue 96 contributes significantly to the substrate specificities of all EPₜ variants, despite crystallographic evidence suggesting that K96 in bEPₜ does not coordinate with Asp-P₂²⁸.

Examination of an x-ray crystal structure of the β-trypsin + BPTI complex (PDB code: 3OTJ)³⁹ aligned with the structure of EPₜ reveals that BPTI residue Arg39 lies in close proximity to EPₜ residue 174. As a result, an electrostatic repulsion likely occurs, accounting for the approximately 5-fold increase in the Kᵢ of Y174R for BPTI.

Non-glycosylated recombinant human enteropeptidase variants Y174R and R96Q were easily expressed into the growth media of P. pastoris as properly folded active enzymes. Both enzymes were highly purified by chromatography on immobilized STI. The Y174R variant displayed increased specificity for the GD₄R-pNA substrate relative to a commercial sample of recombinant human enteropeptidase (Tag*off™). The potential utility of a D₄R linker between fusion proteins²⁴,²⁵ is thus amplified by the use of Y174R. The kinetic properties of R96Q demonstrate the importance of R96 to the binding of extended peptide substrates.

**Materials and Methods**

The hEPₜ amino acid sequence was taken from NCBI Ref. Seq. NP_002763.2. DNAs were codon-optimized and synthesized by different companies; R96Q by GenScript and Y174R
by DNA 2.0. The recombinant human EP$_L$ named Tag*off™ and a cleavage control protein (CCP) were products of EMD Millipore.

**Molecular Cloning**

The DNA sequences for R96Q and Y174R were cloned into the pPICza A plasmid (Life Technologies). The linearized vectors were used to transform X-33 strain *P. pastoris* (Life Technologies) by electroporation. Transformants were selected on YPDS agar plates with 100 µg/mL Zeocin and 100 µg/mL Ampicillin. Selected colonies were grown in 12-well microtiter plates as 2 mL cultures in synthetic minimal medium (SMM) (10 g/L MSG, 2.5 g/L (NH$_4$)$_2$SO$_4$, 100 mM KH$_2$PO$_4$ at pH = 5.0, 13.4 g/L YNB, and 0.4 mg/L biotin) containing 0.5% methanol$^{40}$. Cultures were grown for four days at room temperature with daily addition of 0.5% methanol. Culture supernatants were tested for enteropeptidase activity using the synthetic substrate Z-Lys-SBzl and activities were normalized to wet cell pellet weights (WCW) for selection of the best expressers.

**Fermentation**

Cultures were grown using the strategy in the Life Technologies *Pichia* Fermentation Process Guidelines, with some modification. Basal salts medium (BSM) contained 26.7 mL/L H$_3$PO$_4$, 18.2 g/L K$_2$SO$_4$, 14.9 g/L MgSO$_4$, 4.13 g/L KOH, 40 g/L glycerol and was modified by adding 10 g/L MSG and 2.5 g/L (NH$_4$)$_2$SO$_4$$^{40}$. CaSO$_4$ was omitted to reduce the formation of insoluble calcium phosphates. BSM was supplemented with PTM4 trace elements$^{41}$. Antifoam 204 (0.5 mL/L) was used to prevent foaming. Fermentation occurred at pH at 5.0 and pure O$_2$.
was used to maintain the dissolved O\textsubscript{2} content (dO\textsubscript{2}) at 35\% by regulating the impeller speed. Culture density and protein accumulation were monitored daily.

**Purification**

Culture supernatants, adjusted to pH = 8.0, were centrifuged and filtered through glass fiber pre-filters and 0.45 µm PE filters. Enzymes were affinity purified at pH = 8.0 on an agarose column with immobilized Soybean Trypsin Inhibitor (STI). The wash buffer contained 50mM Tris, 0.5 M NaCl, pH = 8.0 while 50 mM acetic acid, 0.5 M NaCl, pH = 3.5 was used for protein elution. Active fractions (Z-Lys-SBzl) were pooled and concentrated by ultrafiltration on a 10K MWCO membrane (Millipore) and stored in 100 mM glycine, pH = 3.0. Active site titrations of R96Q and Y174R were performed with 4-methylumbelliferyl-p-guanidinobenzoate (MUGB)\textsuperscript{38}.

**SDS-PAGE and Western Blotting**

SDS-PAGE was performed with 12\% (Figure 2.3a and 2.3b) or 4-12\% (Figure 2.3c) NuPAGE gels using MOPS SDS running buffer (Life Technologies) and stained with NuBlu Coomassie stain (NuSep). Western blot on PVDF was blocked for 1 hour in TBST + 1\% BSA at 20 °C, then reacted with 0.1 µg/mL of a mouse monoclonal antibody against bEP\textsubscript{L} (GenScript) in TBST. After washing with TBST, the blot was then incubated for 1 hour with HRP-conjugated goat anti-mouse antibody (0.08 µg/mL) in TBST (Pierce), washed in TBST, and visualized using Super Signal West Pico ECL (Pierce). Images were captured and processed using a GBox (Syngene).
Cleavage of CCP

CCP was incubated with Tag*off™, R96Q, or Y174R enzyme at 37 °C in 20 mM Tris-HCl, 50 mM NaCl, 2 mM CaCl2, pH = 7.4 (provided with Tag*off™). At 2 hours and 6 hours, samples were collected and frozen (-20 °C) prior to SDS-PAGE analysis.

Stability Assay

Enzymes were incubated for five days at 37 °C in 20 mM Tris-HCl, 50 mM NaCl, 2 mM CaCl2, pH = 8.0 and daily samples were stored at -20 °C prior to assays in 300 µM Z-Lys-SBzl, 100 mM Tris, 0.02 mM CaCl2, 0.01% Triton X-100, 10% DMSO, 1 mM DTNB, pH = 8.0.

Kinetic Assays

All enzyme activity assays were conducted in 100 µl format on a 96-well microtiter plate reader at 25 °C ± 0.5 °C. All substrate stocks were freshly dissolved in DMSO. The assay conditions were 100 mM Tris, 0.02% CaCl2, 0.01% Triton X-100, 10% DMSO, pH = 8.0. Assays with 0.05-1.0 mM Z-Lys-SBzl (Sigma or Bachem) also included 1 mM DTNB. The extended peptide substrates GD₄K-pNA and GD₄R-pNA (LifeTein) were used at concentrations between 0.02-1.0 mM and 5-100 µM, respectively. Reactions were initiated by adding enzyme to a 2nM final concentration into premixed buffer and substrate. Absorbances at 410 nm were followed kinetically for 10 minutes. Product extinction coefficients were 13,600 M⁻¹ cm⁻¹ for the thiobenzyl assays and 8,800 M⁻¹ cm⁻¹ for p-nitroaniline. The kinetic parameters Kₘ and kcat were determined by hyperbolic regression analysis of data using Hyper32 (J. Easterby, ©2003).

BPTI inhibition assays were conducted in the same format and with 100 mM Tris, 0.02% CaCl₂, 0.01% Triton X-100, 10% DMSO, pH = 8.0, containing 1mM DTNB and using 300-600
μM Z-Lys-SBzl as the substrate. Here, 2 nM enzyme concentrations were premixed with 50-600 nM BPTI and incubated at room temperature for 15 minutes prior to the addition of Z-Lys-SBzl to initiate the reactions. Dixon plots were generated from the reaction velocity data to determine the K_i values^{44}.

**Molecular Modeling**

Modeling was performed in Discovery Studio 3.1 (Accelrys), using bEP_L (PDB code: 1EKB) (Figure 2.5a) to create the proposed model of rbEP_L Y174R (Figure 2.5b) by single amino acid substitution. This side chain was repositioned by manual bond rotation, compared to the previously proposed position of R174\textsuperscript{22}, and its geometry was cleaned using the software’s Dreiding-like force field. The structure of rhEP_Lsc (PDB code: 4DGJ) provided the template for modeling all human EP_L variants in complex with the D\textsubscript{3}K substrate (taken from 1EKB). The key residues R96, R174, and K219 were repositioned, as above, in comparison with previous reports\textsuperscript{22,29} to reflect possible interactions with the D\textsubscript{3}K substrate (Figure 2.5c, 2.5d, 2.5e, and 2.5f). Residue Q96 (Figure 2.5f) positioning could not be predicted based on currently available data.

**Acknowledgements**

This work was supported by NIH grant R15HL091770. We thank Dr. Michelle Duffourc of the ETSU Molecular Biology Core Facility for her helpful advice and for DNA sequencing.

The authors have no competing interests affecting the objectivity or integrity of the publication.
References


CHAPTER 3

ENGINEERING THE EXPRESSION OF HUMAN NEUTROPHIL ELASTASE AND CATHEPSIN G IN PICHIA PASTORIS AND KLUYVEROMYCES LACTIS

Eliot T. Smith and David A. Johnson*

Department of Biomedical Sciences, James H. Quillen College of Medicine, East Tennessee State University, Johnson City TN 37614

* Correspondence to: David A. Johnson, Department of Biomedical Sciences, James H. Quillen College of Medicine, ETSU, Box 70582, Johnson City, TN 37614.
Phone: (423) 439-2027. Fax: (423) 439-2030. E-mail: davidj@etsu.edu

Keywords: Neutrophil Serine Proteases, Human Neutrophil Elastase, Cathepsin G, Cytochrome B5, Recombinant Protein Expression, Pichia pastoris, Kluyveromyces lactis.
Abstract

Recombinant human neutrophil elastase (rHNE) and human Cathepsin G (rhCatG) were expressed with intact C-terminal pro-peptide extensions in the yeast *Pichia pastoris* and, with limited success, in the yeast *Kluyveromyces lactis*. Both enzymes were engineered as fusion proteins linked to the heme-binding domain of cytochrome B5 (CytB5) and hexahistidine (6xHis) tag by an enteropeptidase-cleavable linker sequence, forming protease zymogens. Both full-length fusion proteins were secreted into the growth media of *P. pastoris* and partially purified by immobilized metal affinity chromatography for the 6xHis tag, furthermore, the partially purified enzymes displayed absorbance at 410 nm that characterizes heme-bound recombinant CytB5 fusion proteins. Both rHNE and rhCatG were activated from their zymogens upon treatment with enteropeptidase, and the concentrations of both activities corresponded to the intensities of absorbance at 410 nm. This is the first report of successful recombinant expression of proteolytically active HNE and hCatG in any microorganism, as well as the first report of the use of CytB5 as a recombinant fusion protein domain in any eukaryotic recombinant expression system. Neither active enzyme was successfully purified to homogeneity, despite repeated attempts using multiple ion exchange chromatography and affinity chromatography for the specific, reversible serine protease inhibitors basic pancreatic trypsin inhibitor (BPTI, Trasylol®) and secretory leukocyte proteinase inhibitor (SLPI).
Introduction

The neutrophil serine proteases (NSPs) human neutrophil elastase (HNE; EC 3.4.21.37) and cathepsin G (CatG; EC 3.4.21.20) act in multiple ways during acute inflammatory and immune responses\(^1\)\(-\)\(^3\). Both HNE and CatG, along with a third NSP, Proteinase 3 (Pr3, Myeloblastin), reside in primary (azurophil) granules of neutrophils (polymorphonuclear leukocytes, PMNs)\(^4\),\(^5\) and act against pathogenic bacteria\(^6\)\(-\)\(^10\) and fungi\(^11\)\(-\)\(^13\), by intracellular\(^6\),\(^7\),\(^11\),\(^14\) and extracellular\(^13\),\(^15\)\(-\)\(^18\) means. In addition to proteolytic and non-proteolytic\(^19\)\(-\)\(^25\) antimicrobial properties, HNE and CatG also process various cytokines\(^26\)\(-\)\(^28\), chemokines\(^29\)\(-\)\(^34\), and protease activated receptors (PARs)\(^35\),\(^36\) to participate in inflammatory regulation and cell signaling\(^1\)\(-\)\(^3\). In the BRENDA enzyme database, more information and links to research articles can be found for HNE (http://www.brenda-enzymes.org/php/result_flat.php4?ecno=3.4.21.37) and for CatG (http://www.brenda-enzymes.org/php/result_flat.php4?ecno=3.4.21.20).

Many mutations of the HNE-coding gene (ELANE) result in cyclic\(^37\)\(-\)\(^40\) or severe congenital\(^39\),\(^41\)\(-\)\(^45\) forms of hereditary neutropenia, diseases of neutrophil maturation characterized by reduced numbers of neutrophils in circulation and increased susceptibility to infections, and these mutations affect HNE activity in different ways\(^46\). Chain-terminating mutations in ELANE have been identified, localized to the final (fifth) exon, that lead to translation of HNE lacking a complete C-terminus\(^47\) and these mutations are closely associated with severe congenital neutropenia (SCN)\(^46\); however, the roles of the C-terminus of HNE are unclear.

All chymotrypsin-like serine proteases, including NSPs\(^3\),\(^48\)\(-\)\(^50\), are initially translated as inactive precursors containing N-terminal signal sequences and activation dipeptides; however, the NSP precursors also possess unusual C-terminal propeptide tails of varying length that do not resemble each other and whose functions are not known\(^47\),\(^51\),\(^52\). These tails are removed post
translationally by an unidentified protease and they are not required for trafficking of NSPs to neutrophil primary granules\textsuperscript{1,47,53}. HNE with an intact C-terminal tail, but not Pr3, is directed, at least in part, through a detour to the cell membrane, from where it is re-internalized by endocytosis\textsuperscript{53}; however, it is unclear if this is also true for CatG. HNE contains a binding site for the μ3a subunit of AP3 that is only exposed after removal of the C-terminal tail\textsuperscript{54} but it appears that this is not how AP3 and HNE interact during trafficking\textsuperscript{53}. Instead, the tetraspanin CD63 has been shown to associate with HNE to mediate AP3-dependent granule delivery\textsuperscript{55}, although the CD63 binding site on HNE has not been identified. The roles of the C-terminal tail of CatG have not been investigated as thoroughly as HNE and the function of the tail is still unknown.

Although others have previously expressed variants of HNE and CatG in mammalian cells\textsuperscript{51,56-58}, active variants of these enzymes have not been expressed in sufficient quantities for most biochemical studies. Recombinant variants of active HNE and active CatG containing intact C-terminal tails have also not been expressed and may facilitate investigations into the function(s) of NSP C-termini and the mechanism(s) responsible for their removal. CatG also exhibits a broad, duel specificity for both trypsin- and chymotrypsin- like substrates\textsuperscript{59,60} that is attributed to the negative charge of Glu 226 (chymotrypsinogen numbering system) at the base of a deep, relatively hydrophobic S1 binding pocket (Schechter-Berger numbering system)\textsuperscript{61}. Recombinant expression of CatG in milligram or gram quantities would enable studies of amino acid substitutions to affect changes in CatG substrate specificity.

HNE and CatG were expressed successfully as secreted, recombinant fusion proteins in the yeast \textit{P. pastoris} and secreted expression attempts were also made, with limited success, in a similar expression system based on the yeast \textit{K. lactis}. Both enzymes were produced with intact C-terminal pro-domain tails as protease zymogens whose enzymatic functions were activated by
the enteropeptidase (EP) -mediated removal of an N-terminal heme-binding cytochrome B5 (CytB5) domain. The fusion proteins also had the heme-binding properties of CytB5; exhibiting absorbance maxima at 410 nm that facilitated the partial purification by immobilized metal (Ni$^{2+}$) affinity chromatography (IMAC) affinity for an N-terminal hexahistidine (6xHis) tag. Attempts to further purify the activated recombinant enzymes to homogeneity were unsuccessful and require further investigation.

Results and Discussion

Protein Modifications and Fusion Protein Design

**NSP Domains and Modifications.** The recombinant HNE (rHNE) and recombinant human CatG (rhCatG) amino acid sequences (Figure 3.1), starting at residue Ile16 (chymotrypsinogen numbering) and including intact C-terminal extensions, were designed from the known sequences of native HNE (NCBI Ref. Seq. NP_001963.1) (Figure 3.1a) and hCatG (NCBI Ref. Seq. NP_001902.1) (Figure 3.1b). An N-terminal EP-cleavable linker sequence, DYKDDDDK, was added to each (Figure 3.1). Conservative substitutions were made to disrupt potential glycosylation sites because *P. pastoris* often hyper-glycosylates secreted proteins$^{62-66}$. In rHNE, three potential N-linked glycosylation sites were disrupted by the substitutions N72Q, N109Q, and N159Q; similarly, one predicted N-linked glycosylation site was disrupted in rhCatG by the substitution N65Q. One potential O-linked glycosylation site was disrupted in rHNE by the substitution S260N; however, this created a potential N-linked glycosylation site. Alternatives to this substitution were considered, and this issue was resolved by the additional substitution T262A.
**Figure 3.1: Amino acid sequence alignments.** Alignments of (A) native HNE and engineered rHNE, with N-terminal EP cleavage site extension DYKDDDK, or (B) native hCatG and engineered rhCatG, also with DYKDDDK, highlight key features and differences. The N-terminal Ile 16 residue (▲) and the catalytic triad residues (△) that are necessary for proteolytic activity, as well as the Glu 226 residue (□) that confers trypsin-like substrate specificity, appear in **green**. Substitution sites to disrupt potential N-linked (●) and O-linked (〇) glycosylation appear in **orange** and **purple**, respectively. Arg-to-Gln substitutions (★) to reduce the pI and eliminate potential kexin cleavage sites appear in **blue**. The C-terminal pro-peptide tails that are removed during neutrophil maturation appear in **bold with red underline**.
Both *P. pastoris* and *K. lactis* express similar furin-like kexin proteases that can cleave dibasic sequences such as KR~X and RR~X\(^{67,68}\). In rHNE, the substitutions R21Q, R76Q, R128Q, and R178Q were used to disrupt four potential kexin cleavage sites (Figure 3.1a) and, in rhCatG, the substitutions R76Q, R86Q, R111Q, R114Q, R148Q, R178Q, R186Q, and R188(a)Q were similarly used (Figure 3.1a). The replacement of highly basic Arg residues with Gln at these sites also reduced the calculated isoelectric points of both rHNE, from 9.7 to 8.7, and of rhCatG, from 11.9 to 10.9. This is thought to be beneficial because it has been shown that a comparatively high protein isoelectric point correlates with poor expression in *P. pastoris*\(^{69}\).

**Cytochrome B5 Recognition and Purification Domain.** Separately, the amino acid sequence of a fusion domain was designed for addition to the N-terminus of the rHNE and rhCatG proteins. This fusion domain consists of the soluble heme-binding domain of cytochrome B5 (CytB5) from *Rattus norvegicus* (NCBI Ref. Seq. NP_071581.1) with an N-terminal hexahistidine (6xHis) tag for IMAC purification and antibody-based detection\(^{70-72}\). The CytB5 domain has been used previously to produce, in *E. coli*, heme-binding recombinant proteins with a distinctive spectrophotometric absorbance peak at the wavelength \(\lambda = 410\) nm\(^{71,72}\). Additionally, the CytB5 fusion decreased the pI of the secreted protein and prevented premature activation of the enzymes which could have been damaging to the yeast cells.

**Fusion Protein Zymogens.** Based on the three modified amino acid sequences, three protein-coding genes were codon-optimized for recombinant expression in *P. pastoris* and *K. lactis* and synthesized as cDNAs with suitable restriction sites for subsequent plasmid construction. Expression plasmids were assembled in two stages and used to successfully transform *P. pastoris* and *K. lactis*. 
**Figure 3.2: Fusion Construct Schematics.** The multi-domain structure of the rhCatG and rHNE fusion constructs, starting with the N-termini on the left. The α-mating factor yeast secretion signal pre-domain (yellow) directs both polypeptide chains to the secretory pathway. In the Golgi, the protease kexin (Kex2) cleaves the Kex2 site (blue) KR→H to remove the pre-domain. This exposes the N-terminal 6xHis tag (orange) of the purification and activation pro-domain of the secreted, inactive, mature zymogen. The cytochrome B5 domain (red) of the mature zymogen exists in a heme-bound state. The treatment of these constructs with the serine protease enteropeptidase (EP) specifically cleaves the D_{4}K→I peptide bond of the EP site to activate rhCatG and rHNE (green) by removing the pro-domain to expose the N-terminal Ile16 of either enzyme.

The fusion proteins encoded by these plasmids contain multiple domains intended to facilitate the expression, secretion, purification, identification, and activation of the enzymes rHNE and rhCatG. The initial intracellular products of gene translation include, at their N-termini, the yeast α-mating factor signal sequence that directs the protein for secretion and is, subsequently, removed by kexin in the Golgi apparatus (Figure 3.2). The final secreted protein contains an N-terminal 6xHis tag and CytB5 heme-binding domain connected by the EP-cleavable linker sequence to the rHNE or rhCatG protease domain (Figure 3.2). The constructs are protease zymogens that can only be activated through cleavage of the sequence DDDDK→I by enteropeptidase to separate the protease domains from their fusion partners and expose the
native Ile16 N-termini that are necessary for the activity of all S1 clan serine proteases (Figure 3.2)\textsuperscript{73}.

**Transformation and Screening**

*P. pastoris* and *K. lactis* cells were transformed with linearized insertion cassettes designed to integrate into the yeast genomes by site-specific homologous recombination into the AOX1 transcriptional promoter of *P. pastoris* or the LAC4 transcriptional promoter of *K. lactis*. Positive transformed colonies from agar plates were grown in 2 mL or 10 mL cultures, depending on the number of transformants, under inducing conditions for four days. After harvesting, culture supernatants were adjusted to pH = 8.0 and treated overnight with EP and then tested for rHNE or rhCatG enzymatic activity (units/mL). The wet cell pellet mass (g/mL) was also determined for each culture, after which the ratio of enzymatic activity to wet cell pellet mass (units/g) was calculated. Figure 3.3a shows the results for CytB5-rhCatG screening of 44 *P. pastoris* transformants as an example. The four colonies with the highest ratios were selected for growth in baffled flasks with equal starting culture densities, to confirm the results obtained in smaller volumes. In the sample case of rhCatG, the four best colonies were #3, #27, #36, and #43. Of these, #43 expressed more enzyme than #36 (not shown), which had given the best results in the smaller volume (Figure 3.3a).
Figure 3.3: Screening Colonies for rhCatG. (A) Activity (units) per gram of wet cell mass was calculated for 44 P. pastoris clones grown for 4 days in 2 mL cultures. Clarified culture supernatants were activated with EP and tested for rhCatG activity (units/mL) using the chromogenic synthetic peptide substrate Suc-AAPF-SBzl. The weight (g/mL) was determined for each cell pellet. The four best clones (orange and red bars) were grown in 100 mL cultures in induction medium, starting at an OD600 = 1, and screened for activity (not shown), finding that colony #43 (red bar) expressed more activity per cell density than the others. (B) PCR on gDNA extracted from yeast cell pellets confirms the presence of multiple repeat tandem copies of the insertion cassette in colony #43 by the appearance of the ~1 kb PCR amplification product. Colonies #3 and #27 also tested positive, although the bands appear very faint. The negative control and #36 tested negative. Band intensity does not correlate to the number of genomic inserts, due to variation in the concentration and integrity of the gDNA samples used as PCR templates.

Because multiple recombination events can occur within a single cell of either species during transformation to create a multi-copy integrant, the most productive clones were screened by PCR to identify those containing multiple genomic copies of the insertion cassette. Genomic DNA (gDNA) samples were extracted from cell pellets using glass bead disruption, phenol/chloroform extraction, and ethanol precipitation; furthermore, the integrity of each gDNA
sample was evaluated by gel electrophoresis. Both PCR tests used primers that only amplified their product if two adjacent copies of the insertion cassette occurred in tandem in the gDNA. Results for the CytB5-rhCatG clones #3, #27, #36, and #43 are provided as an example (Figure 3.3b). In this instance, the clone #43 showed a clean PCR product of the correct size (~1 kb) that confirms clone #43 as a multi-copy integrant. For each yeast species and each fusion protein, a single clone was identified by this process as the best for continued use.

**Protein Expression**

The most successful and heavily pursued of the CytB5-rHNE and CytB5-rhCatG fusion protein expression attempts occurred by fermentation of *P. pastoris* transformants. Expression attempts in *K. lactis* produced only limited success, thus they were abandoned in favor of the *P. pastoris* platform. CytB5-rHNE *K. lactis* fermentation produced less than 100 µg/L of EP-activable rHNE activity against MeO-Suc-AAPV-pNA (less than 150 µg/L CytB5-rHNE); furthermore, neither the cells nor the culture supernatant displayed 410 nm absorbance peaks and no red color was visible prior to purification (below). Bioreactor scale fermentation of CytB5-rhCatG *K. lactis* was not pursued due to poor results on a small scale and the limited success of CytB5-rHNE.

Clarified culture supernatants of both CytB5-rHNE and CytB5-rhCatG strains contained their respective activities, which were only detectable after activation with EP, but the media did not exhibit 410 nm absorbance peaks. The CytB5-rHNE supernatants from successful fermentations of *P. pastoris* are estimated to have produced between 1.5 and 3.5 mg/L of EP-activated rHNE (between 2.3 and 5.3 mg/L of CytB5-rHNE recombinant protein), as determined by catalytic activity for the chromogenic synthetic peptide substrate MeO-Suc-AAPV-pNA in
comparison to known concentrations of purified native HNE. The CytB5-rhCatG *P. pastoris* never produced more than 200 µg/L of EP-activated enzyme (~300 µg/L of CytB5-rhCatG recombinant protein) in fermentation media, compared to purified native hCatG for the substrate Suc-AAPF-SBzl.

During the methanol induction phase of growth, *P. pastoris* cultures expressing both CytB5 constructs developed a dark pink color distinct from the light tan normally seen in *P. pastoris* cultures. The majority of this color was retained by cell pellets and only recoverable upon cell lysis. Clarified cell lysates exhibited the dark red color and 410 nm absorbance peak indicative of CytB5 protein expression; however, these lysates did not display HNE or CatG activity before or after treatment with EP.

**Partial Purification and Activation: rHNE and rhCatG in *P. pastoris***

Both CytB5-rHNE and CytB5-rhCatG fusion proteins were partially purified from *P. pastoris* fermentation media using immobilized metal (Ni²⁺) affinity chromatography (IMAC) for the 6xHis tag; however, efficient recoveries were dependent on buffer exchange by ultrafiltration of fermentation media prior to purification. Due to the heme-binding properties of the CytB5 domain, IMAC elution fractions containing purified fusion constructs could be identified by the presence of 410 nm peaks in their absorbance spectra (Figure 3.4). Purified fractions of CytB5-rHNE had a distinct absorbance peak at 410 nm in addition to the expected 280 nm absorbance peak (Figure 3.4a); furthermore, fractions with strong 410 nm absorbance peaks often appeared dark red-brown in color. Purified fractions of CytB5-rhCatG contained an additional, stronger absorbance peak at 350 nm (Figure 3.4b) and these fractions displayed a dark green color believed to be attributable to a co-purified contaminant that was not successfully removed.
Figure 3.4: Absorbance Spectra for CytB5 Fusion Protease Zymogens After Partial Purification by IMAC. Partially purified (A) CytB5-rHNE exhibits the characteristic spectrum for purified CytB5 fusion proteins, with clear absorbance maxima at 280 nm (○) and 410 nm (●) accompanied by a visible dark red color. (B) CytB5-rhCatG recovered by IMAC was contaminated with an unidentified substance that made the eluate visibly dark green instead of dark red, as confirmed by the additional absorbance maximum at 350 nm (●) and the increased A280/A410 ratio relative to (A) CytB5-rHNE.

IMAC elution fractions that contained detectable 410 nm absorbance peaks were activated by overnight treatment with EP. Levels of subsequent rHNE and rhCatG activities increased to their maxima as functions of treatment time (data not shown) and correlated by fraction to the intensity of absorbance at 410 nm, but not 280 nm, suggesting the identity and partial purification of both enzymes, as in the examples shown (Figure 3.5), where approximately 600 µg of CytB5-rHNE (Figure 3.5a) and 270 µg of CytB5-rhCatG (Figure 3.5b) were recovered, as estimated by activity. About 1.2 mg of CytB5-rHNE was recovered from 1 L of fermentation medium containing approximately 1.5 mg, an 80% recovery yield, during the most successful purification attempt. Due to limited CytB5-rhCatG expression, 3.5 L of fermentation medium, concentrated more than 50-fold and containing about 370 µg activable
enzyme, was necessary to acquire the estimated 270 µg indicated above, a 73% recovery yield. The ratios of A410/A280 reflect the relative amount of purification by fraction.

**Figure 3.5: IMAC elution profiles.** Profiles are shown for the partial purification of (A) CytB5-rHNE and (B) CytB5-rhCatG. Absorbance at 280 nm (○) shows total protein content and 410 nm (■) shows the presence of heme-bound CytB5 fusion domain. Bar graphs show the enzymatic activity in 25 µL of EP-activated fractions of (A) rHNE for MeO-Suc-AAPV-pNA or (B) rhCatG for Suc-AAPF-SBzl. Enzymatic activity is proportional to absorbance at 410 nm, but not 280 nm. The higher ratio of A410/A280 for CytB5-rHNE indicates greater purity relative to CytB5-rhCatG.
Figure 3.6: SDS-PAGE and Western Blot for Partially Purified CytB5-rHNE. CytB5-rHNE partially purified by IMAC and activated by EP in comparison to native HNE. (A) SDS-PAGE under reducing conditions shows partial purification of CytB5-rHNE and activation by EP. Native HNE is a mixture of three isoforms appearing between approximately 25 – 30 kDa that vary by glycosylation pattern. Prominent bands appearing between 10 – 15 kDa in the native HNE sample correspond to fragments of autolysis and also appear in commercial preparations of purified native HNE. (B) Activated rHNE, but not CytB5-rHNE, detection by Western blot confirms the identity of active rHNE. The primary antibody was polyclonal rabbit anti-HNE IgG and the HRP-conjugated secondary antibody was polyclonal goat anti-rabbit IgG. The 10 kDa band in CytB5-rHNE is unidentified, but may be due to non-specific binding. Two bands are labeled in the native HNE sample, showing the mature enzyme (~25 kDa) and a fragment of limited autolysis (< 15 kDa) as in (A).

IMAC purified CytB5-rHNE from P. pastoris, with and without EP pre-treatment, were analyzed by SDS-PAGE and Western Blot to verify the purity, identity, and activation of the recombinant construct (Figure 3.6). SDS-PAGE under reducing conditions shows the presence of a protein above 35 kDa that corresponds to the calculated molecular mass of 38 kDa for
CytB5-rHNE (Figure 3.6a); however, many other proteins of lower molecular mass also appear in the sample, likely products of incomplete translation or protein degradation that co-purify with the activable enzyme as a consequence of the N-terminal location of the 6xHis tag (Figure 3.6a). After overnight treatment with EP, the same sample displays a prominent band at 25 kDa (Figure 3.6a), the calculated molecular mass of activated rHNE. Faint bands of varying molecular mass in the activated rHNE sample likely correspond to inactive fragments cut by EP or HNE (Figure 3.6a). Purified, native HNE is a mixture of three isoforms with different glycosylation patterns that migrate between about 25 – 30 kDa (Figure 3.5a)\(^{74,75}\). The native HNE sample also displays prominent bands between 10 – 15 kDa due to limited autolysis, similar to commercial preparations from whole blood (http://www.athensresearch.com/products/enzymes-and-inhibitors/elastase-human-neutrophil).

Western blotting with a rabbit anti-HNE primary polyclonal antibody and HRP-conjugated goat anti-rabbit IgG secondary antibody labeled both native HNE and activated rHNE at ~25 kDa (Figure 3.6b); however, the CytB5-rHNE zymogen was not detected. The 10 kDa band that appears in the CytB5-rHNE sample on the blot (Figure 3.6b) is likely due to non-specific binding. In the native HNE sample, an additional band below 15 kDa is also labeled on the Western blot (Figure 3.5b) and corresponds to an autolytic fragment that is also seen in commercial preparations from whole blood (http://www.abcam.com/Neutrophil-Elastase-protein-Active-ab91099.html).

Attempts were made (not shown) to further purify the enzymes by ion exchange chromatography as well as affinity chromatography on columns with immobilized inhibitors basic pancreatic trypsin inhibitor (BPTI, Trasylol®) and secretory leukocyte proteinase inhibitor (SLPI); however, these pursuits met with limited success. In the case of BPTI, a small amount of
activated rHNE was recovered but most activity did not adsorb and the results were not reproducible. HNE affinity for the SLPI affinity column was very strong and resulted in complete adsorption of rHNE activity. This rHNE activity was never recovered successfully, despite the use of various denaturants and chaotropic agents. Of these, 8 M urea released some measurable activity from the column yet only about 20% of the original activity could be accounted for; furthermore, 8 M urea did not reduce the activity of HNE. The further purification of rhCatG has also been unsuccessful thus far.

**Partial Purification and Activation: rHNE in *K. lactis***

Despite poor production with the *K. lactis* expression platform, CytB5-rHNE was partially purified from *K. lactis* fermentation medium, but without the use of buffer exchange or ultrafiltration. Recombinant protein was detectable in IMAC elution fractions by the presence of 410 nm peaks in the absorbance spectra, as in Figure 3.4a; however, absorbance at 280 nm was obscured by excess imidazole. The intensity of 410 nm absorbance was very low (<0.06 units) and indicative of poor recovery. Samples activated with varying amounts of EP (volumes listed, concentration unknown) revealed the activation of rHNE as a function of time and the amount of EP used (Figure 3.7). Based on the final maximum activity (Figure 3.7), only 100 ng of active rHNE was recovered in the most active fraction and the total recovery was less than 1 µg.
Figure 3.7: Activation of CytB5-rHNE Partially Purified by IMAC from K. lactis Fermentation Medium. Identical 50 µL samples were treated with the indicated volumes of an EP sample (concentration unknown) in a total reaction volume of 100 µL. After 4 hours (yellow) and 24 hours (blue), samples were measured for activity with chromogenic MeO-Suc-AAPV-pNA. The results show that IMAC purified CytB5-rHNE from K. lactis is activated by EP and that increasing amounts of EP reduced the amount of time needed for the fusion protein to become fully active. Based on the maximum activity (~ 0.5 mA410/min), this 2 mL fraction contains only about 100 ng of active rHNE (150 ng of activable CytB5-rHNE).

Extremely low levels of expression and inefficient IMAC recovery hindered the expression of rHNE and rhCatG in K. lactis. Work with this yeast occurred prior to the more successful efforts with P. pastoris; furthermore, limited experience with yeast fermentation during the work with K. lactis may have contributed to the poor results of this expression platform. The low level expression and purification of CytB5-rHNE from K. lactis, as well as the accompanying 410 nm absorbance properties of the protein, encouraged the further pursuit of this fusion construct design; however, the decision was made to change to the P. pastoris platform due to the volume of research available, as well as the prior familiarity of the investigators, for this system.
The Utility of Cytochrome B5 Fusion Domains in Recombinant Protein Expression

The heme-bound CytB5 fusion protein expression in *P. pastoris* has several implications that extend well beyond the specific expression of these two recombinant serine proteases.

First, the red coloration of the cells, concentrated media, adsorbed samples on resins, and eluted fractions, significantly improves the processes of fermentation and purification through the use of visual confirmation and characteristic absorbance spectrum patterns to follow protein location and concentration. It is hypothetically possible, although as yet experimentally unconfirmed, that, if used in conjunction with the constitutively active GAP transcriptional promoter, transformant cells might even express sufficient CytB5-based coloration on agar selection plates to facilitate the identification of “super-producing” clones through the use of simple visual observation during the transformant selection process.

Second, the ratio of the intensities of 410 nm and 280 nm absorbance peaks, A410/A280, can be used to evaluate the relative purity and concentration of a sample. In the case of the NSPs expressed here, the A410 reflected not only full length protein but also heme-binding, partially translated products and this is a consequence of the use of an N-terminal 6xHis tag. With the use of a C-terminal tag or other purification strategy designed to recover only full-length recombinant proteins, the A410 would more accurately reflect the concentration. The molar extinction coefficient of heme-bound CytB5 at 413 nm absorbance is 114,000 M\(^{-1}\) cm\(^{-1}\) for the oxidized form and 185,000 M\(^{-1}\) cm\(^{-1}\) for the reduced form\(^{71}\). It is, however, important to note that only a portion of the CytB5 fusion protein is in the heme-bound form\(^{71}\), which is necessary for the 410 nm absorbance; consequently, the A410 cannot be used to directly quantify the total concentration of fusion protein, only the heme-bound fraction.
The ratio of A410/A280 does correlate with the relative level of purity of the sample. In the case of CytB5-rHNE and CytB5-rhCatG, the most active fraction of rHNE (Figure 3.4a, Fraction 6) has an A410/A280 ratio = 0.68 whereas the most active fraction of rhCatG (Figure 3.4b, Fraction 36) has an A410/A280 ratio = 0.13. This means that a larger percentage of the protein in the rHNE fraction is heme-bound fusion protein than the percentage of rhCatG in its equivalent fraction.

Third, the accumulation of dark red color in cells and culture media post-induction relative to uninduced and untransformed cultures confirms that the two yeast species manufacture hemin endogenously and that the induced overexpression of CytB5 fusion constructs creates a “heme-sink” in the yeasts that is similar to results shown previously in E. coli. Centrifuged cell pellets from harvested P. pastoris fermentation cultures retained a dark pink color that, even after extensive rinsing with buffers of high ionic strength, was only extractable in a supernatant upon yeast cell lysis by French pressure cell disruption, sonication, or manual glass-bead disruption methods; however, treatment of these cell extracts with EP did not result in active rHNE or active rhCatG.

This suggests that the majority of CytB5-rHNE and CytB5-rhCatG proteins being produced were retained by the cells and either misfolded or degraded. Recently, it has been shown that P. pastoris expressing recombinant proteins under control of the methanol inducible AOX1 promoter, but not the constitutively active GAP promoter, also exhibited secretory limitations that resulted in accumulation of product in the ER and initiation of the ERAD pathway. Expression of the CytB5 NSP fusion proteins may prove more successful with the use of the constitutive GAP promoter. If these proteins were more efficiently secreted, the locations and intensities of the CytB5-based 410 nm absorbance peaks might reflect this.
Conclusions

Further efforts are warranted in the recombinant expression of HNE and CatG based on their successful secretion and partial purification from *P. pastoris*. Strains of this yeast have recently been made available that express proteins with human glycosylation patterns (Pichia GlycoSwitch®; http://www.rctech.com/licensing/gxt-pichia-glycoswitch.php); consequently, variants of rHNE and rhCatG should be attempted in which the native glycosylation sites have not been disrupted. It is likely that the use of the constitutively active GAP transcriptional promoter instead of AOX1 would improve the secretory capacity of the cells\textsuperscript{77} and this is especially important in the case of CatG due to its low levels of expression. The procedures and materials for purifying active rHNE and rhCatG must be revisited and optimized to improve both the yield and purity of the recovered products.

The results of CytB5 fusion protein expression are encouraging with regard to the use of CytB5 for the detection of protein production and purification using yeast expression platforms, as has previously been demonstrated for *E. coli*\textsuperscript{71,72}, and this method is worthy of further application. The expression of rHNE and rhCatG would have been much more difficult without observation of the 410 nm peak in the absorbance spectra because these enzymes require enzymatic activation prior to detection and quantification based on activity. Absorbance maxima at 280 nm are sufficient to follow total protein recovery during the purification process; however, confirmation of protein location by the presence of a red color enabled the visual verification of protein expression and optimization of protein purification.
Materials and Methods

Molecular Cloning

The NCBI database was used as a source for the native amino acid sequences of HNE (NCBI Ref. Seq. NP_001963.1) and CatG (NCBI Ref. Seq. NP_001902.1). The amino acid sequence of the soluble heme-binding domain of R. norvegicus CytB5 (from NCBI Ref. Seq. NP_071581.1) was identical to that previously used for recombinant expression in E. coli. All amino acid and gene sequences were analyzed and modified using the software Geneious® 6.1.4 (Biomatters, http://www.geneious.com). Custom genes were codon-optimized for expression in P. pastoris and K. lactis and synthesized by GenScript, who provided the genes in pUC57 plasmids for subcloning. The CytB5 gene was flanked by 5’ XhoI and 3’ KpnI restriction sites whereas both rHNE and rhCatG included 5’ KpnI and 3’ NotI sites.

Standard techniques were used throughout the molecular cloning process. Briefly, plasmids were treated with appropriate restriction enzymes (Thermo Scientific, Fermentas FastDigest®) and digestion fragments were recovered from agarose gels using the Zymoclean™ gel DNA recovery kit (Zymo Research). The recovered inserts were ligated into the multiple cloning sites (MCSs) of the target plasmids using Quick-Stick Ligase (Bioline) and transformed into Z-Competent™ chemically competent DH5α E. coli (Zymo Research). Plasmid assembly occurred in two stages. First, the CytB5 gene was inserted separately into both pPICza A and pKLAC1 expression plasmids using XhoI and KpnI restriction sites to position the inserted sequence in the correct reading frame at the 3’ end of the sequence encoding α-mating factor and the two plasmids were designated as pPICza-CytB5 and pKLAC1-CytB5. Second, the rHNE and rhCatG genes were inserted similarly into the MCSs of pPICza-CytB5 and pKLAC1-CytB5 with the restriction sites KpnI and NotI. The sizes of all plasmid inserts were verified by PCR to...
amplify the MCS contents using primers provided with the yeast expression kits, GoTAQ® All-In-One Master Mix DNA Polymerase (Promega) and its protocols.

Transformation of *P. pastoris*

The *P. pastoris* transformation protocol was adapted from instructions for the EastSelect™ Pichia Expression Kit (Life Technologies). The plasmids, designated pPICza-CytB5-rHNE and pPICza-CytB5-rhCatG, were treated with the restriction enzyme BstXI to generate linearized insertion cassettes for electroporation. These cassettes were recovered with a DNA cleanup kit (Zymo) and further concentrated by ethanol precipitation. X-33 *P. pastoris* cultures were prepared for and transformed by electroporation as prescribed by Life Technologies, using a cuvette with a 2 mm gap and a BTX ECM 600 electroporation system with settings at 1.5 kV and 186 Ω to result in a 7.5 msec pulse length. Transformants were selected for resistance by growth at 30 °C on YPDS agar plates with ampicillin and Zeocin™ (100 µg/mL, each) and positive colonies were re-plated to fresh YPD agar plates, also containing ampicillin and Zeocin™, and numbered for further screening.

Transformation of *K. lactis*

The chemical transformation of *K. lactis* used reagents and followed instructions provided with the *K. lactis* Protein Expression Kit (New England Biolabs, NEB). The plasmids, pKLAC1-CytB5-rHNE and pKLAC1-CytB5-rhCatG, were both linearized with the BstXI restriction enzyme and recovered with a DNA clean up kit (Zymo). Chemically competent GG799 *K. lactis* were thawed, treated with linearized DNA and a transformation reagent, and subjected to treatment as per instructions. Transformed cells expressing the enzyme acetamidase
were selected for growth at 30 °C on nitrogen-free yeast carbon base (YCB) agar plates containing 5 mM acetamide. Positive colonies were re-plated to YPD agar plates and numbered for further screening.

Expression Screening

**Culture Conditions.** Numbered clones were grown under sterile conditions as either 2 mL liquid cultures in 12-well plates or 10 mL cultures in 50 ml conical tubes, depending on the number of transformants available. *P. pastoris* were grown in synthetic minimal medium (SMM) comprised of 10 g/L monosodium glutamate (MSG), 2.5 mg/L (NH₄)₂(SO₄), 1.34% yeast nitrogen base, >4 x 10⁻⁵% biotin, and 100 mm potassium phosphate at pH = 5.0 and containing 100 µg/mL of both ampicillin and Zeocin® (Life Technologies); furthermore, they were fed 1% methanol every 24 hours and grown with orbital shaking at 28.5 °C for 4 days prior to harvest. *K. lactis* were grown in SMM at pH = 7.0 containing 2% casamino acids and 100 µg/mL of ampicillin. *K. lactis* were fed 1% galactose every 24 hours and grown with orbital shaking at 28.5 °C for 4 days prior to harvest. Culture supernatants were clarified by centrifugation and retained, along with the separate cell pellet, for further analyses.

**Enzyme Activation.** Supernatants from screening cultures were adjusted to pH = 8.0 by titration with 1 M Tris, pH = 9.0 and centrifuged to remove insoluble material. Samples were treated with EP overnight at 37 °C prior to activity testing.

**Activity Assays.** All reported activity assays for HNE and CatG activity, regardless of source, purity, or stage of analysis, follow this protocol. EP-treated samples were tested for the enzymatic cleavage of chromogenic, synthetic peptide substrates. Methoxysuccinyl-AAPV-para-nitroanilide (MeO-Suc-AAPV-pNA; EMD Millipore Cat#454454) and the similar
thiobenzyl ester substrate Suc-AAPV-SBzl (LifeTein custom synthesis) were used to detect HNE activity; similarly, CatG activity was detected with Suc-AAPF-pNA(Sigma #S-7388) and Suc-AAPF-SBzl (LifeTein custom synthesis). Assays were conducted in 100 µL reaction volumes in 96-well microtiter plates and monitored over the course of 10 minutes for the accumulation of 410 nm absorbance indicating substrate cleavage. Each reaction was composed initially of 5 to 25 µL of activated sample with sufficient assay buffer containing 100 mM HEPES, 1 M NaCl, 10 % glycerol, 0.01 % Triton X-100, and 0.01 % NaN₃ at pH = 8.0 to bring the total volume to 50 µL. Immediately prior to initiating the absorbance assay, 50 µL of substrate buffer made from the assay buffer with 300 µM of the appropriate substrate were added to the reaction mixture. Assays using the thiobenzyl ester substrates also contained 100 mM DTNB.

PCR Testing for Multicopy Transformants. Genomic DNA (gDNA) samples were recovered from cell pellets by phenol/chloroform extraction and ethanol precipitation using standard procedures. The integrity of all gDNA samples was confirmed by gel electrophoresis using the FlashGel cartridge system (Lonza) per its instructions. PCR for single- and multi-copy integrations were conducted using the procedure and integration primers provided with the EasySelect™ Pichia Expression Kit (Life Technologies). PCR screening of K. lactis used the procedure and primers provided with the K. lactis Protein Expression Kit (NEB). All PCR used GoTAQ® All-In-One Master Mix DNA Polymerase (Promega) and its protocols.

Fermentation and Media Processing

All fermentations were conducted in a BioFlo 110 bioreactor (New Brunswick), following its user’s manual, with a dissolved oxygen sensor (Mettler-Toledo) and a pH probe to monitor culture conditions. Cultures were grown following the guidelines of the Pichia
Fermentation Process Guidelines by Life Technologies and modifying feed and media formulations as needed. Several growth media were tried for *P. pastoris* and *K. lactis* based on the formulae for basic salts medium (BSM) recommended by the Life Technologies guidelines and FM2282 and using their respective trace element supplements, PTM1 and PTM4. The most successful medium formulation was found to be a variation of BSM supplemented with PTM4 to reduce the formation of insoluble salts. This BSM variant included the addition of 10 g/L MSG and a supplementary vitamins mix that was made as a 100x concentrate containing 0.1 g/L L-His-HCl, 0.2 g/L DL-Met, 0.2 g/L DL-Trp, 0.0002 g/L biotin, 0.0002 g/L folic acid, 0.2 g/L inositol, 0.04 g/L niacin, 0.02 g/L p-aminobenzoic acid, 0.04 g/L pyridoxine HCl, 0.02 g/L riboflavin, 0.04 g/L thiamine HCl, 0.04 g/L calcium pantothenate, 10 g/L CaCl2.

Upon harvest, fermentation media supernatants were clarified by centrifugation followed by filtration through glass fiber pre-filters (Millipore Cat#AP2004200) and 0.45 µm filters (Pall P/N 60173). Clarified *P. pastoris* broth volumes were reduced by ultrafiltration with 10 kDa molecular weight cut-off (MWCO) tangential flow filtration cartridges (Sartorius Vivaflow50 P/N VF05P0). During ultrafiltration, the media were flushed extensively with IMAC binding buffer containing 25 mM sodium phosphate and 300 mM NaCl at pH = 8.0. *K. lactis* fermentation supernatants were not subject to ultrafiltration.

**Immobilized Metal Affinity Chromatography**

IMAC used either His-Bind® Resin (EMD Millipore) or Profinity™ IMAC Resin (Bio-Rad). All stripping, equilibration, loading, washing, and elution buffers contained 25 mM sodium phosphate and 300 mM NaCl at pH = 8.0. The metal ion stripping buffer contained 100
mM EDTA; the equilibration and loading buffers contained no additional components; the washing buffer contained 5 mM imidazole; the elution buffer contained 20 mM imidazole.

**SDS-PAGE and Western Blotting**

NuPAGE® Bis-Tris polyacrylamide gels (Life Technologies) were run using MOPS-SDS running buffer and stained with Coomassie blue stain (SimplyBlue™; Life Technologies). For Western blotting, gels were run as above and blotted to PVDF by semi-dry transfer. Membranes were blocked with 1% BSA in TBST. The primary antibody was polyclonal rabbit anti-HNE (Athens Research) and the secondary antibody was polyclonal goat anti-rabbit HRP-conjugate (Pierce).

**Spectrophotometry**

All protein absorbance values and spectra were read on a Biotek PowerWave™ XS2 plate reader using Corning Half Area 96 well plates P/N 3679. For DNA samples, concentrations were determined by measuring absorbance in a NanoVue™ spectrophotometer (GE).

**Acknowledgements**

This work was supported by NIH grant R15HL091770. We thank Dr. Michelle Duffourc of the ETSU Molecular Biology Core Facility for her helpful advice and for DNA sequencing. We also thank Michael Bradfield, Allyn Heath, Karie Hodges, Haley Klimecki, Branson Mauck, Evan Perry, Jessica Pugh, Junny “Junior” Tayou, Tim Wesley, and Dustin Wood for their technical assistance.
References


CHAPTER 4

THE EXPRESSION OF NATIVE HUMAN C-REACTIVE PROTEIN AND HUMAN C-REACTIVE PROTEIN E42Q IN *PICHIA PASTORIS*

Eliot T. Smith¹, Avinash Thirumalai², Sanjay K. Singh², Alok Agrawal², and David A. Johnson¹*  

Department of Biomedical Sciences, James H. Quillen College of Medicine, East Tennessee State University, Johnson City TN 37614

¹ Recombinant Protein Engineering/Expression

² Quantification, Purification, Characterization of Variants

* Correspondence to: David A. Johnson, Department of Biomedical Sciences, James H. Quillen College of Medicine, ETSU, Box 70582, Johnson City, TN 37614. Phone: (423) 439-2027. Fax: (423) 439-2030. E-mail: davidj@etsu.edu

Keywords: C-Reactive Protein, Pentaxin, Blood Plasma Protein, Acute Phase Reactant, Recombinant Expression, *Pichia pastoris*
Abstract

C-reactive protein (CRP) is an acute phase protein that binds phosphocholine (PCh) and the complement protein C1q. CRP is known to bind oxidized LDL (oxLDL) under acidic conditions and the E42Q amino acid substitution improves oxLDL binding at physiological pH. Recombinant wild-type human CRP (rCRP) and the rCRP E42Q substitution variant were expressed in the yeast *Pichia pastoris* under methanol induction, yielding correctly folded and functional pentameric proteins in the culture media. For rCRP, 3.7 mg/L were expressed when grown in synthetic minimal medium in baffled shaker flasks while the rCRP E42Q variant expressed 1.7 mg/L and these proteins were purified by PCh affinity chromatography. Both proteins were shown to bind to immobilized PCh like native human CRP (hCRP); however, rCRP was expressed in two forms, one with the correct monomeric molecular mass of ~25 kDa and a second, slightly larger monomeric form, as shown by SDS-PAGE and Western blot with a CRP-specific antibody. A later SDS-PAGE only showed a single band for rCRP purified by PCh affinity. The two forms could not be separated by PCh affinity chromatography, ion-exchange chromatography, or gel-filtration chromatography. The source of the extra mass has not been determined.
Introduction

C-reactive protein (CRP) expression by hepatocytes increases rapidly early in the acute phase response of innate immunity, with plasma levels increasing as much as 1000-fold in response to acute inflammatory stimuli. CRP was discovered in 1930 for its ability to precipitate C-lipopolysaccharide present on the cell wall of *Streptococcus pneumoniae*. Later, phosphocholine (PCh) was identified as the component of C-lipopolysaccharide that serves as the ligand for CRP. PCh-associated CRP binds complement component C1q and allows CRP to activate the classical complement activation cascade, although only C1-C4 are activated, without activation of the later complement components C5-C9. CRP has also been shown to have other potential binding partners, including chromatin, fibronectin, histones, laminin, phosphoethanolamine, ribonuclear proteins, oxidized LDL (oxLDL), and the oxLDL receptor LOX-1.

CRP is a member of the pentraxin family of acute phase pattern recognition receptors, characterized by a homopentameric functional form. Identical monomers form a ring primarily via electrostatic surface interactions between subunits. One side of the ring, the recognition face, contains five calcium-dependent PCh binding sites, one on each subunit. The other face, the effector face, binds the globular head of complement component C1q, which spans the central pore of CRP; furthermore, the binding of CRP with immunoglobulin Fcγ receptors is also thought to occur on the effector face.

Patients with chronically elevated CRP levels have greater risk of diabetes, atherosclerosis, and cardiovascular disease. CRP is frequently used as a clinical marker of inflammation and cardiovascular disease, as recommended by the American Heart Association.
and the Centers for Disease Control and Prevention\textsuperscript{25-27}. CRP displays both pro-inflammatory and anti-inflammatory\textsuperscript{8,11,28-30} activities whose net \textit{in vivo} effects may depend on the context.

CRP has been shown to bind oxLDL under acidic conditions\textsuperscript{20,31,32} and the single residue variant CRP E42Q also binds oxLDL at physiological pH\textsuperscript{9}, likely by relaxing the interface between the CRP subunits to improve the flexibility of the pentamer. Wild type and substitution variants of CRP have been expressed as recombinant proteins in several expression systems including \textit{E. coli}, \textit{Leishmania tarentolae}, \textit{Trichoplusia ni}, and the yeast \textit{Pichia pastoris}\textsuperscript{33-36}. CRP recently expressed in \textit{P. pastoris} was found to contain the additional amino acid sequence EAEA at the N-terminus\textsuperscript{36} because these amino acids were used to separate the kexin cleavage site (KR) from the CRP N-terminus. These extra amino acids are not always completely removed by the Ste13 dipeptidyl peptidase in \textit{P. pastoris} following cleavage by kexin. While the additional N-terminal amino acids did not seem to affect its function, the expression of wild-type CRP, without added amino acids, in \textit{P. pastoris} would provide a robust source of recombinant protein that is structurally and functionally identical to the native human pentamer and an easier method for producing variants.

For further study regarding the binding of CRP to oxLDL, both recombinant wild type CRP (rCRP) and the single amino acid variant rCRP E42Q were expressed as secreted proteins in \textit{P. pastoris}, quantified by ELISA, and tested for PCh affinity. Both proteins formed pentamers, as evinced by the ability to bind PCh like the native human CRP (hCRP) pentamer. Wild type rCRP was further analyzed by SDS-PAGE and Western blot and it was found in two monomeric forms of distinct molecular mass. Despite further purification efforts with anion exchange chromatography and gel filtration, the two forms of wild type rCRP could not be separated and this indicates that these two rCRP monomer types associated randomly into
pentamers. The problem could be unique to the transformant clone that was selected for expression or due to the culture conditions during expression; however, further studies are required to resolve this issue.

Results

Protein Engineering and Molecular Cloning

Amino acid sequences for mature hCRP (NCBI Ref. Seq. NP_000558.2) and the CRP E42Q single residue substitution variant were used, without further modification, to generate synthetic cDNAs codon optimized for expression in the yeast *P. pastoris*. The yeast were transformed with pPICza A plasmids under transcriptional regulation of the methanol-inducible AOX1 promoter and directed for secretion by the α-mating factor secretion signal sequence that is removed by kexin during secretion. The kexin cleavage sequence, KR, is adjacent to the N-terminus of each CRP sequence, with no intervening residues (Figure 4.1)

**Figure 4.1: Primary Structure of rCRP Construct with α-Mating Factor.** The amino acid sequence for rCRP (Black) with the α-mating factor secretion signal (Red) connected to the N-terminus of mature CRP by the kexin cleavage sequence KR~ (Blue), where “~” represents the site of kexin mediated proteolysis. The substitution of Glu 42 (Green) with Gln 42 is the only difference in the sequence of rCRP E42Q and these proteins contain no other modifications. In the genes for both constructs, two stop codons (not shown) immediately follow the codon for the C-terminal Pro 206.
The codon-optimized sequences for both constructs were cloned into the pPICzα A expression plasmid for *P. pastoris*, under control of the repressible, methanol-inducible transcriptional promoter AOX1. These plasmids were used to transform the X-33 strain of *P. pastoris* and positive transformants were selected by Zeocin® resistance for further screening.

**Transformant Screening**

Transformant clones of both protein variants were screened for best production using 2 mL cultures grown in synthetic minimal medium (SMM) in 12-well microtiter plates. These cultures were grown for four days with methanol induction prior to harvest. Upon harvest by centrifugation, the wet cell pellet weights (WCW, g/L) were measured for all cultures. CRP production levels in culture supernatants were quantified (µg/L) by sandwich ELISA with a polyclonal rabbit anti-human CRP IgG as the capture antibody, the mouse monoclonal anti-CRP antibody HD2.4 for detection, and an HRP-conjugated goat anti-mouse secondary antibody for chemiluminescence. The ELISA results were normalized to the WCW measurements to determine ratios of production/cell mass, in units of µg protein/g WCW. These values were analyzed to find the cultures expressing the most protein/cell mass. Of 47 clones screened for each construct, four were selected for further analyses.

For the narrower pool of eight transformants (two sets of four clones), 10 mL cultures were grown, inoculated to equal starting optical densities, OD₆₀₀ = 1.0, in 50 mL conical tubes in SMM with 1% methanol and fed 1% methanol each day for four days. These cultures were analyzed again, as above, to isolate the most productive clone for each product.
Protein Expression

The best producing clones were scaled up to 300 mL volumes of media in baffled flasks. Various growth media based on recipes for BMMY and SMM were employed in efforts to improve expression and secretion. The formulae experimented with these parameters: temperature, pH, Ca\(^{2+}\) content, ionic strength, phosphate source (potassium orthophosphate or sodium hexametaphosphate), culture volume (up to a 5 L scale in bioreactor fermentations), induction phase length, methanol concentration, and induction co-feeding with glycerol or sorbitol. Both BMMY-based and SMM-based media performed similarly; however, SMM media were preferred for the improved regulation of culture conditions that they afford. The best expression results were 3.7 mg/L of rCRP and 1.7 mg/L of rCRP E42Q, both achieved in normal SMM at pH = 5.0 and buffered with potassium phosphate. These cultures were fed 1% methanol each day for four days in 1 L baffled shaking flasks no more than 1/3 full with broth at 28.5 °C. Glycerol and sorbitol co-feed attempts did not improve yields of either CRP construct; furthermore, production was markedly reduced by glycerol (a known repressor of the AOX1 promoter).

PCh Binding Assays

Expression media for both recombinant variants were tested for binding to immobilized PCh. These assays were performed similar to the sandwich ELISA; however, the polyclonal rabbit anti-CRP antibody was replaced by immobilized PCh-BSA for CRP capture. Serial dilutions of rCRP culture growth media displayed binding to PCh-BSA that was dependent on Ca\(^{2+}\) and inhibited with EDTA (Figure 4.2a). PCh-BSA affinity was also dependent on either the rCRP or rCRP E42Q concentration with results matching the behavior of the native human
pentamer across the same concentration range (Figure 4.2b). These assays show that both recombinant variants were secreted as stable pentamers because monomeric CRP has a significantly lower affinity for PCh-BSA \(^{32}\). Both variants could also be recovered from culture media by PCh-agarose affinity chromatography; however, problems emerged during the purification process. Low levels of protein expression and the heavy formation of insoluble phosphate salts in culture supernatants during purification led to reformulation of the growth conditions and media. The best results were achieved with SMM at pH = 6 with 100 mM potassium phosphate and fed 1% methanol daily for four days. At lower pH, CRP is more likely to stick to other things and many previous cultures had been grown at pH = 5. Increasing the pH caused the formation of insoluble phosphates; consequently, the clarified culture supernatants were adjusted to pH = 8 followed by centrifugation and filtration to precipitate and remove as much insoluble phosphate as possible. The E42Q variant was not further characterized and all subsequent work focused on the production and purification of wild type rCRP, as shown below.
Figure 4.2: PCh Binding Assays. For PCh binding assays, microtiter plate wells with immobilized PCh-BSA were used to capture CRP from culture medium. Adsorbed CRP was then detected with the monoclonal mouse anti-CRP antibody HD2.4 and an HRP-conjugated goat anti-mouse secondary antibody. (A) Serial dilutions of culture medium showed binding of rCRP to PCh-BSA in the presence of Ca$^{2+}$ and the sequestration of Ca$^{2+}$ with EDTA significantly reduced this affinity, consistent with the known behavior of native hCRP. (B) The binding of rCRP and rCRP E42Q to PCh occurred in a concentration-dependent manner identical to the behavior of native hCRP. This confirms that rCRP and rCRP E42Q are expressed and secreted as functional pentamers because the PCh binding affinity of monomeric CRP is significantly weaker.$^{32}$

Initial Purification and SDS-PAGE

Wild type rCRP was recovered from culture medium by immobilized PCh affinity chromatography. Following this single-step purification, samples were analyzed by SDS-PAGE (4% - 20%) under reducing conditions (Figure 4.3a). Although hCRP migrated as a single band at the correct molecular weight, rCRP migrated as two bands (Figure 4.3a). The more prominent band, of lower molecular mass, migrated at the same molecular weight as native protein; however, the less prominent band has an apparent molecular mass of ~1 kDa greater than hCRP.
A later, separate SDS-PAGE (12%) under reducing conditions (Figure 4.3b) did not reveal the contaminant band in rCRP, although this may be due to overloading.

**Figure 4.3: SDS-PAGE of PCh-Affinity Purified rCRP.** (A) A 4% - 20% SDS-PAGE, under reducing conditions, compares hCRP with rCRP upon single-step purification by affinity for PCh. The rCRP migrates as two bands of distinct molecular mass, unlike the expected single band of ~25 kDa, as shown for hCRP. The band of lower mass migrates like hCRP; however, a second band of higher mass also appears approximately 1 kDa above. The band of larger mass represents approximately 20% of the total protein in the sample, estimated by densitometry. (B) In a separate reducing gel (12% SDS-PAGE), rCRP and hCRP both migrate at ~25 kDa, although the second band appearing in (A) may be obscured in the broader band in (B).
Further Purification

Because a second band was initially found to co-purify with rCRP based on PCh affinity, further attempts were made to separate these two proteins by a three step process. First, rCRP was purified by PCh (Figure 4.4a), as above. Second, PCh-purified rCRP was further purified by anion exchange (Figure 4.4b). Third, the dual-purified sample was separated by gel filtration (Figure 4.4c). Sample recovery data chronicles each stage of purification (Table 4.1).

Figure 4.4: Three Step Purification of rCRP. (A) An elution profile for purification of rCRP by PCh affinity chromatography. Fraction volumes were 500 µL and fractions 13 – 28 were pooled (5.5 mL) for subsequent steps. (B) This pooled sample was further purified by anion exchange chromatography, which eluted as a single peak and focused the protein into a smaller elution volume (2 mL pooled). (C) The anion exchange elution pool was separated by gel filtration and eluted as a single major peak.
At each step of the purification, samples were collected for analysis by SDS-PAGE, with each sample loaded in 1 µg and 5 µg quantities (Figure 4.5). PCh-purified rCRP, both with and without EDTA, migrated with the same pattern as before; furthermore, the two bands failed to separate at any stage of purification (Figure 4.5).

<table>
<thead>
<tr>
<th>Step</th>
<th>Total rCRP (mg)</th>
<th>rCRP Concentration (µg/mL)</th>
<th>Yield (% of Input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM (PCh Input)</td>
<td>2.14</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>PCh Recovery</td>
<td>1</td>
<td>200</td>
<td>46.7</td>
</tr>
<tr>
<td>Anion Exchange Input</td>
<td>0.885</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>Anion Exchange Recovery</td>
<td>0.573</td>
<td>287</td>
<td>64.7</td>
</tr>
<tr>
<td>Gel Filtration Input</td>
<td>0.48</td>
<td>1370</td>
<td></td>
</tr>
<tr>
<td>Gel Filtration Recovery</td>
<td>0.147</td>
<td>70</td>
<td>30.6</td>
</tr>
</tbody>
</table>

**Table 4.1: Three Step Purification of rCRP**

**Figure 4.5: SDS-PAGE of Three Step rCRP Purification.** Samples from each step of the three-stage purification of rCRP are shown in comparison to hCRP. In each step, the two bands fail to separate and the purity was never improved beyond the single-step PCh affinity purification. The band of larger mass represents approximately 20% of the total protein in the sample.
Purified hCRP and rCRP were analyzed by Western blotting (Figure 4.6). Both were labeled by the polyclonal rabbit anti-CRP primary antibody and heavy sample loading shows some CRP degradation products in both sample types (Figure 4.6). Of particular importance, both bands of the purified rCRP are labeled by the antibodies, indicating that the band with ~1 kDa more mass is likely an modified form of the protein (Figure 4.6).

**Figure 4.6:** Western Blot of hCRP and rCRP. Purified hCRP and rCRP samples of 75 ng, 150 ng, and 300 ng were loaded, as indicated. The PVDF membrane was probed with the polyclonal rabbit anti-CRP primary antibody and an HRP-conjugated donkey anti-rabbit secondary antibody. For both hCRP and rCRP, some degradation products can be seen. Both of the bands that appear on the SDS-PAGE for the purified rCRP samples are labeled by the anti-CRP antibodies on the Western blot, indicating that the band of larger molecular mass may be a modified version of the correctly expressed protein.
Discussion

Both recombinant CRP variants were present as functional pentamers in the culture supernatant. This was confirmed by the binding of the proteins to immobilized PCh-BSA in a manner dependent on the concentrations of Ca$^{2+}$ and CRP (Figure 4.2). The successful expression of functional rCRP and rCRP E42Q in P. pastoris provides a useful source for producing these proteins efficiently in large quantities. This expression platform may provide further use for the production of other CRP variants as well. The presence of a second band of protein that co-purified with rCRP provides a puzzle that has not yet been solved. Because later SDS-PAGE of samples purified by PCh (Figure 4.3b) did not reveal the presence of this band, it may be due to an unknown experimental effect or changes in growth conditions during protein production. Examination of other transformed clones may reveal that the second band is anomalous exclusively to the selected clone.

If the second band truly does co-purify with rCRP and is not an artifact, then further analyses are required to determine the source of this modification. It is possible that a portion of the protein was post-translationally modified in some unforeseen manner; however, no glycosylation sites are present. An SDS-PAGE (not shown) was treated with a stain specific for phosphorylated proteins and found that both hCRP and rCRP stained to the same extent. Mass spectrometry may illuminate this issue.

Although much exploration of culture media and growth conditions failed to improve the expression of these proteins beyond 4 mg/L, significant experience has been gained during this process regarding the production of recombinant proteins in the yeast P. pastoris. With further attempts, it is anticipated that bioreactor-based yeast fermentation may yield improved results relative to those previously seen for the expression of CRP. Finally, although the synthetic genes
used for rCRP and rCRP E42Q were produced by two different companies using different optimization algorithms the expression levels did not differ significantly.

**Materials and Methods**

**Molecular Cloning and Yeast Transformation**

The amino acid sequence for hCRP was taken from the NCBI database (NCBI Ref. Seq. NP_000558.2). Codon-optimized gene sequences were optimized and synthesized by DNA2.0 (rCRP E42Q) and by GenScript (rCRP) and provided with 5’ XhoI and 3’ NotI in the pUC57 subcloning vector. The coding sequence was cloned into the pPICza A *P. pastoris* expression plasmid (Life Technologies) using standard molecular cloning techniques, as described in the EasySelect™ *Pichia* Expression Kit (Life Technologies). The restriction enzymes were FastDigest® enzymes (Thermo Scientific). Digestion fragments were separated by agarose gel electrophoresis and gel extraction (Zymo). Protein-coding inserts were ligated into pPICza A with Quick-Stick Ligase (Bioline) and the assembled plasmids were propagated in Z-Competent™ DH5α *E. coli* (Zymo) and transformed *E. coli* were screened for Zeocin™ resistance.

X-33 *P. pastoris* were transformed with linearized expression plasmids by electroporation using the protocol in the EasySelect™ *Pichia* Expression Kit (Life Technologies). Electroporation was conducted using cuvettes with a 2 mm gap in a BTX ECM 600 electroporation system set at 1.5 kV and 186 Ω with an approximate pulse length of 7.5 msec. Transformants were grown at 30 °C on YPDS agar plates containing 100 µg/mL of Zeocin™ for transformant selection and 100 µg/mL of ampicillin to prevent bacterial
contamination. After three days, individual colonies were plated to fresh YPD agar plates (with the same antibiotics) and numbered.

**Transformant Screening**

Numbered clones were grown in 2 mL cultures in 12-well plates in synthetic minimal medium (SMM) comprised of 10 g/L monosodium glutamate (MSG), 2.5 mg/L (NH$_4$)$_2$(SO$_4$), 1.34% yeast nitrogen base, >4 x 10$^{-5}$% biotin, and 100 mm potassium phosphate at pH = 5.0 and containing 100 µg/mL of both ampicillin and Zeocin® (Life Technologies). The medium was supplemented with 1% methanol and an additional 0.5% methanol was added each day for four days. After four days, the cultures were harvested and the wet cell pellet weights (WCW) were determined. Culture supernatants were clarified by centrifugation and screened by ELISA (see below) for the presence of CRP. The ELISA results for CRP concentration (µg/mL) were normalized by dividing by the WCW (mg/mL) to determine production efficiency (µg protein/mg WCW). The four most productive clones of rCRP and of rCRP E42Q (eight total clones) were identified and further screened for production in larger volumes.

In the second phase of colony screening, the most productive clones were all growth from the same starting density (OD$_{600}$ = 1.0) in SMM containing 1% methanol and fed 1% methanol every day for four days. These cultures were grown in 10 ml volumes in 50 ml sterile conical tubes with shaking at 30 °C. These clones were analyzed as described above for the 2 mL cultures.
ELISA

Sandwich-type ELISA analyses quantified recombinant protein samples relative to a standard curve following the protocol previously published for quantification of CRP\textsuperscript{37}. Microtiter plate wells were treated with 100 µL of polyclonal rabbit anti-human CRP IgG (Sigma) that was diluted 1:1000 in TBS, pH = 7.2. After 2 hours of incubation at room temperature, the antibody was aspirated and the wells were blocked with TBS + 5 g/L gelatin at room temperature for 45 minutes. After the blocking solution was removed, diluted samples were added to the wells, with a series of samples of known concentration of hCRP to create a standard curve. After 2 hours of incubation at 37 °C, the samples were removed and the wells were washed three times with ELISA buffer (TBS + 0.1% gelatin + 0.02% Tween-20). A 100 µL aliquot of the monoclonal mouse anti-CRP antibody HD2.4 (0.5 µg/mL) was then added to each well, followed by incubation for 1 hour at 37 °C. After the wells were washed again three times with ELISA buffer, 100 µL of 1:1000 diluted HRP-labeled goat anti-mouse IgG (GE Healthcare) was added and left to incubate for 1 hour at 37 °C. After a final round of three washes with ELISA buffer, the HRP substrate was added. After a 10 minute incubation, the absorbance values were measured at \( \lambda = 405 \) nm.

Protein Expression

Expression cultures were grown with methanol in 300 mL broth volumes in 1 L baffled flasks with orbital shaking at 30 °C; similarly, starter cultures were grown with glycerol in 50-100 mL volumes in 250 mL baffled flasks. The growth media BMGY and BMMY, described in the EasySelect™ Pichia Expression Kit, contain 1% yeast extract, 2% peptone, 100 mM potassium phosphate (pH = 6.0), 1.34% YNB, 4 x 10\textsuperscript{-5}% biotin, and either 1% - 2% glycerol.
(BMGY) or 0.5% - 1% methanol (BMMY). Synthetic minimal media (SMM) contained 10 g/L monosodium glutamate, 2.5 g/L (NH₄)₂SO₄, 100 mM KH₂PO₄ (pH = 5.0), 4 x 10⁻⁵% biotin, 1.34% YNB, and either 2% glycerol or 0.5% - 1% methanol. Multiple modifications of the SMM formula were attempted.

In some media formulations, potassium phosphate was replaced with sodium hexametaphosphate (SHMP) to reduce precipitation; although, SHMP was found to be a poor pH buffer. Some cultures received 0.1% - 0.5% glycerol or 0.5% - 1% sorbitol co-feeds with methanol during the induction phase; however, these co-feeds did not improve expression. In many cultures, solutions of vitamins and minerals were prepared in the lab as a substitute for commercial YNB; however, these solutions were compositionally identical to YNB and did not improve culture growth or protein expression in comparison. In addition to the pH values listed above, cultures were also grown under other conditions between 4.5 ≤ pH ≤ 7.0. Some formulae were supplemented with 2% Casamino acids to limit extracellular proteolysis of the protein products. Despite multiple attempts with a range of media formulations, the standard SMM medium was preferred due to composition control (relative to BMMG and BMGY) and formula simplicity (relative to other SMM formulations).

**Purification**

All purification processes follow procedures previously described³¹,³⁸. Affinity chromatography with PCh-conjugated Sepharose (Pierce) used a loading buffer of borate buffered saline (BBS) (100 mM boric acid, 25 mM tetra-sodium borate, 225 mM NaCl, pH = 8.35) with 3 mM CaCl₂ and an elution buffer of BBS with 10 mM EDTA. Eluted samples were
dialyzed overnight against Mono Q buffer A (20 mM Tris, 150 mM NaCl, 0.1 mM EDTA, pH = 7.8).

Anion exchange chromatography used a Mono Q column (5/50 GL Pharmacia #17-5166-01) equilibrated with Mono Q buffer A and connected to a BioRad Biologic Duo Flow Protein Purification System. After loading at 0.5 mL/min, proteins were eluted with a buffer gradient from 150 mM NaCl (Mono Q buffer A) to 1 M NaCl (Mono Q buffer B).

HPLC gel filtration was performed using Superose 12 (10/300 GL Pharmacia #17-5173-01) with a bed volume of 24 mL in a 10 x 300 mm column connected to a BioRad Biologic Duo Flow Protein Purification System. The sample was prepared by concentrating 1.8 mL of Mono Q purified CRP to a final volume of 250 µL containing 480 µg of protein that was injected into the column. Purified proteins were stored in TBS with 0.1 mM CaCl₂.

**SDS-PAGE and Western Blotting**

The 4%-20% SDS-PAGE gels were manually cast and run in Tris-Glycine buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH = 8.3). These gels were run at 80 – 100 V until the samples entered the stacking gel and then at 180 V for 4 hours. The 12% SDS-PAGE used a precast NuPAGE® Bis-Tris acrylamide gel (Life Technologies) in MOPS-SDS running buffer. Gels were stained with Coomassie blue stain.

Western blotting was performed by semi-dry transfer for 35 minutes at 25 V with PVDF in a Tris-glycine western transfer buffer (39 mM glycine, 48 mM Tris, 0.03% SDS, 20% methanol, pH = 8.9). The membrane was blocked with 5% powdered milk in TBS overnight at 4 °C. The primary antibody was rabbit polyclonal anti-CRP and the secondary antibody was HRP-conjugated donkey anti-rabbit IgG. Results were visualized by ECL chemiluminescence.
Acknowledgements

This work was supported by NIH grants R15HL091770 and R01HL071233.

The authors have no competing interests affecting the objectivity or integrity of the publication.
References


CHAPTER 5
DISCUSSION AND CONCLUSION

Yeast Recombinant Expression Systems

The yeast *Pichia pastoris* served as a suitable host for the transgenic expression of recombinant variants of the human proteins enteropeptidase, cathepsin G, neutrophil elastase, and C-reactive protein. Although the results of protein expression in *Kluyveromyces lactis* were less successful, many of the issues at that time resulted from lack of user experience and limited availability of literature on the platform. These recombinant protein expression studies yielded much experience regarding the fermentation of yeasts; however, the *K. lactis* constructs were abandoned in favor of the more familiar *P. pastoris* before much success was achieved. The results of this work should not be taken to imply that the *K. lactis* system suffers from any inferiority as a recombinant expression system; contrarily, *K. lactis* shares many of the advantages of *P. pastoris*, as already discussed.

Promoter Selection, Codon Optimization, and Custom Gene Synthesis

While the AOX1 methanol-inducible promoter provided strict control over protein production in *P. pastoris*, future work to express variants of these enzymes should use the constitutively active GAP promoter, based on recent research that indicates improved secretory capacity with GAP (Love et al. 2012). Levels of recombinant protein expression for EP, CatG, HNE, and CRP never exceeded 5 mg/L; however, Life Technologies reports much higher production yields for *P. pastoris*. The relatively low yields that were obtained may have been the result of poor secretion and not poor production. The finding that red color accumulated
inside *P. pastoris* during the expression of CytB5-rhCatG and CytB5-rHNE suggests that much of the product was not secreted and may have led to intracellular protein degradation, as has been shown elsewhere to occur with use of the AOX1 promoter in *P. pastoris* (Love et al. 2012). Expression of these proteins under control of the GAP promoter may alleviate this problem.

It is impossible to say if codon optimization of the gene sequences had a significant impact on levels of protein expression without making a comparison to expression results for nonoptimized genes encoding the same protein. Even then, confounding factors such as insertion cassette copy number, variations in secretory capacity, and other clone-to-clone variations have significant effects on protein expression levels that cannot always be predicted and can significantly affect the evaluation of codon optimization. Indeed, screening of over 50 transformant clones often yields only 1 or 2 with significantly higher expression levels for which the cause of improvement is unknown.

Custom gene synthesis was found to have many advantages for simplifying the process of engineering recombinant proteins regardless of the usefulness of codon optimization. This process made it easier to create the multisubunit CytB5 fusion constructs by eliminating the need for multiple PCR steps. The disruption of numerous potential sites of glycosylation and kexin cleavage was also performed without the need for PCR-based site-directed mutagenesis and subcloning procedures.

**Glycosylation and Cleavage Site Disruptions**

The disruption of glycosylation sites in rEP, rhCatG, and rHNE did not appear to adversely affect the enzymatic properties of any of these serine proteases; however, future studies may require serine proteases with native human glycosylation. Recently, a new *P.*
*pastoris* expression system has been developed that allows users to select from several patterns of human glycosylation (Pichia GlycoSwitch®; [http://www.rctech.com/licensing/gxt-pichia-glycoswitch.php](http://www.rctech.com/licensing/gxt-pichia-glycoswitch.php)). Use of this system not only eliminates the need to disrupt glycosylation sites but also allows recombinant production of proteins that more closely resemble their natural counterparts. Future attempts to express recombinant human serine proteases in *P. pastoris* may benefit from a humanized glycosylation pattern.

The decision to disrupt potential sites of kexin proteolysis may or may not have improved protein expression. Because previous attempts to express CatG from the human cDNA had met with only limited success, kexin proteolysis was proposed as a potential factor reducing expression. It has been previously shown that high pI proteins like CatG do not express well in *P. pastoris* (Boettner et al. 2007) and the high frequency of dibasic sequences certainly indicates that kexin proteolysis may be a factor. With some genetic manipulation, a strain of *P. pastoris* could be engineered by disruption of the KEX2 gene that would significantly reduce the vulnerability of dibasic sequences to proteolysis during secreted expression. Although the expression products would retain the α-mating factor after secretion, the use of an EP cleavage site would enable removal of this N-terminal fusion domain without the need for kexin protease.

**Thoughts on Arginine Content and Recombinant Protein Expression**

Although it is possible that a prevalence of dibasic sequences predisposes a yeast-secreted recombinant protein to cleavage by kexin, at least 2 other possibilities may also be responsible for the poor expression of high pI proteins, as proved to be the case with rhCatG. First, proteins and peptides with high arginine content can have antimicrobial properties, as has previously been shown for CatG (Bangalore et al. 1990; Shafer et al. 1990; Shafer et al. 1991;
Miyasaki et al. 1993; Shafer et al. 1993; Shafer et al. 1996; Shafer et al. 2002). While fermentation cultures of *P. pastoris* expressing CytB5-rhCatG did not display any signs of impaired growth rates or increased cell death, it is possible that highly basic proteins may interfere with intracellular function in other ways that could reduce production efficiency. Second, an excess demand for Arg could surpass the ability of the yeast to supply charged Arg-tRNAs, resulting in a higher incidence of premature ribosome disassembly and accumulation of truncated protein products. This potential problem could be exacerbated by an overabundance of one Arg codon in a codon-optimized sequence; however, careful regulation of protein production rates by limiting methanol availability and temperature could also counteract this issue by allowing the cells time to regenerate their supply of charged tRNAs. As mentioned previously, absolute levels of tRNA abundance have not been characterized for *P. pastoris* and this information would greatly enhance the current work on codon optimization for the improvement of recombinant protein expression.

**Enteropeptidase: Variations on a Theme**

The recombinant expression of variants of hEP$_L$ proved very successful regarding the manipulation of this enzyme’s high level of substrate specificity. While the reduced specificity of the R96Q substitution variant was of academic value for confirming the role of this residue in the extended substrate binding preferences of hEP$_L$, the improved specificity conferred by the Y174R substitution has practical value for reducing nonspecific cleavages by this highly specific protease that is used in biotechnology to separate recombinant fusion protein domains. Although it was not mentioned and no data was presented, studies with CatG indicated that the CatG substrate Suc-AAPF-SBzl was cleaved by EP$_L$, likely due to the geometry of the EP$_L$ S1
specificity pocket, which appears to be large enough to accommodate the Phe residue of this synthetic substrate. In an attempt to eliminate this undesirable activity, an EP₅ variant was attempted in which the residue Gly 226 was replaced with Asp, in addition to the Y174R substitution; however, this completely abolished the enzymatic activity of the protein, likely by limiting the size of the S1 specificity pocket.

Despite the failure of the G226D substitution to improve, or even retain, substrate specificity, further efforts to modify the extended binding site of EP₅ may yield further improvements on specificity or even facilitate the custom design, based on other similar serine proteases, of restriction proteases with novel substrate preferences. The chymotrypsin-like serine protease granzyme B, for example, has a relatively strong preference for P1 Asp residues due to the presence of Arg 226 in the base of its S1 specificity pocket. Because granzyme B and EP₅ both share the same chymotrypsin-like tertiary structure, residues with the same number share very similar positions regardless of the identity of the amino acid at the given position. Although the substitution G226R in EP₅ would likely limit the size of the S1 specificity pocket as was seen for G226D, the fundamental concept of altering the charge of this binding site could completely change the substrate specificity.

In one particularly promising design, either substitution D189R or D189K could completely convert the P1 preference of EP₅ from positively charged Lys or Arg to negatively charged Asp or Glu. By itself, this change in substrate specificity would confer specificity for any poly-Asp sequence; however, the exact site of proteolytic cleavage might be difficult to control. By adding the substitutions R96D and K99D, the substrate specificity of hEP₅ could theoretically be completely reversed from DDDDK~X to, for example, KKKKD~X. Although
this may seem like an esoteric exercise, experimentation along these lines could result in novel approaches to engineer customized substrate specificity for biotechnological purposes.

The Untold Story: Other Expression Attempts with rhCatG and rHNE

The recombinant expression of rhCatG and rHNE was subject to many difficulties and multiple fusion partner domains were attempted prior to the successful results obtained with the CytB5 soluble heme-binding domain. A table is provided that chronologically summarizes the attempted constructs (Table 5.1), all of which failed to yield the desired results.

Table 5.1: Other Expression Attempts with rhCatG and rHNE

<table>
<thead>
<tr>
<th>Fusion Domain(s)</th>
<th>Purification Strategy</th>
<th>Detection Strategies</th>
<th>Activation Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Serum Albumin (HSA)</td>
<td>HSA-Cibacron Blue Affinity, Inhibitor Affinity</td>
<td>A280, Activity Assays</td>
<td>EP&lt;sub&gt;L&lt;/sub&gt; Cleavage Site</td>
</tr>
<tr>
<td>6xHis, Enhanced Green Fluorescent Protein (eGFP)</td>
<td>6xHis-IMAC, Inhibitor Affinity</td>
<td>A280, eGFP Fluorescence, Activity Assays</td>
<td>EP&lt;sub&gt;L&lt;/sub&gt; Cleavage Site</td>
</tr>
<tr>
<td>6xHis, Rubredoxin</td>
<td>6xHis-IMAC, Inhibitor Affinity</td>
<td>A280, A410, Visible Red Color, Activity Assays</td>
<td>EP&lt;sub&gt;L&lt;/sub&gt; Cleavage Site</td>
</tr>
<tr>
<td>6xHis, SUMO</td>
<td>6xHis-IMAC, Inhibitor Affinity</td>
<td>A280, Activity Assays</td>
<td>SUMO-specific protease or EP&lt;sub&gt;L&lt;/sub&gt; Cleavage Site</td>
</tr>
</tbody>
</table>

The first attempted fusion was human serum albumin (HSA), which is the most abundant protein in blood plasma and displays a known affinity for the blue dye Cibacron Blue. Although as much as 1.4 g/L of HSA has been expressed in P. pastoris (Kobayashi et al. 2000), this fusion partner did not provide a suitable “carrier protein” to facilitate production of rhCatG or rHNE. Because this was the first construct that was attempted, lack of positive results may have been caused by limited user experience; however, many other issues have been encountered since then that may have also contributed to the failure of the HSA fusion protein design.
In an effort to improve the detection of fusion protein production, enhanced green fluorescent protein (eGFP) was employed for its ease of detection, with an N-terminal 6xHis tag for purification. All 6xHis tags employed in these designs were located at the N-terminus of the secreted protein and this leads to the purification of truncated protein products and partial degradation fragments. Because the nature of this study required the production of enzymes with C-termini free of amino acid extensions, the more desirable C-terminal 6xHis tag could not be used. Initial results with eGFP suggested that some protein was being produced by this method; however, very little activity and virtually no green fluorescence could be detected in the culture supernatant. Later analysis of yeast cells using fluorescence microscopy suggested that this recombinant fusion protein may have been accumulating inside the cells, results that have since been reinforced by the intracellular accumulation of red color in P. pastoris cells expressing CytB5 fusion proteins.

The third recombinant fusion domain, rubredoxin, was the first attempt by this lab to exploit a protein’s chromogenic properties for the detection of fusion constructs. Rubredoxin binds iron atoms and displays an absorbance peak at \( \lambda = 410 \text{ nm} \); however, the extinction coefficient of rubredoxin at 410 nm is only 5.5% that of CytB5 (Mitra et al. 2005). The recombinant expression of rubredoxin chimeras, much like eGFP, only resulted in low yields of enzymatic activity. Curiously, the culture supernatant did display a rusty brown color; however, this color was a precipitate that settled upon centrifugation (15,000 – 21,000 x g).

The presence of the rusty brown precipitate was tantalizing because it had not been observed with the previous fusion constructs. At this time, other chromogenic proteins were evaluated as candidates and CytB5 was selected based on previous success with this fusion domain in E. coli (Finn et al. 2005; Mitra et al. 2005). While this later proved to be the most
successful attempt at expressing rhCatG and rHNE in \textit{P. pastoris}, the CytB5 fusion construct was at one time discarded due to difficulties with the 6xHis-IMAC purification process. Initial results showed acceptable levels of protein expression for both enzymes; however, the proteins did not adsorb well to IMAC resin charged with Ni$^{2+}$, Co$^{2+}$, or Cu$^{2+}$.

After many attempts to improve IMAC purification yields for CytB5 fusion proteins, a new approach was devised using the small ubiquitin-like modifier (SUMO) protein as a fusion tag. Due to continued success with recombinant expression of EP$_L$, 2 versions of the SUMO fusion were constructed apiece for rhCatG and rHNE, one with the EP$_L$ cleavage linker sequence DDDDK and another lacking this sequence, being activated instead by a protease specific to SUMO. Neither of these constructs yielded significantly better production results than the CytB5 fusion proteins. Activation with the SUMO protease was costly because, unlike EP$_L$, SUMO protease was not available from recombinant expression system in the lab.

\textbf{Improving the Purification of CytB5-rhCatG and CytB5-rHNE}

Fortunately, further research proved that the issues with 6xHis-IMAC purification of the CytB5 were surmountable. Of particular note, many manufacturers’ protocols indicate that Tris can be used as a suitable buffer, but that it can also impair 6xHis binding to IMAC resin. For the purification of CytB5-rhCatG and CytB5-rHNE, Tris was found to interfere with this stage of the purification process. In contrast, phosphate buffers were found to significantly improve protein binding to IMAC. Also, as-yet unidentified components of spent \textit{P. pastoris} fermentation media seem to interfere with 6xHis-IMAC purification. While buffer exchange by membrane ultrafiltration improved IMAC purification yields, these proteins were eluted under very low concentrations of imidazole (20 – 50 mM) and were difficult to separate from other contaminants.
that dissociated from the column under the same imidazole concentrations. Future attempts to improve the expression of rhCatG and rHNE by these methods should employ a deca-histidine (10xHis) tag (or longer) for improved Ni\(^{2+}\) affinity because this may increase the concentration of imidazole required to elute the protein of interest and could improve efforts to eliminate contaminants.

The more significant difficulty that was encountered in the purification of rhCatG and rHNE was the failure to recover these activated enzymes by affinity chromatography for basic pancreatic trypsin inhibitor (BPTI, Trasylol\®). Attempts to use secretory leukocyte proteinase inhibitor (SLPI) in place of BPTI succeeded in adsorbing over 99% of active rHNE and native HNE; however, the extremely tight binding interactions between HNE and SLPI proved very difficult to break. Because BPTI has been used routinely in the purification of HNE and CatG in the past (Baugh and Travis 1976), the efficiency of the BPTI column that was used is called into question.

The BPTI columns used in these experiments were created by coupling BPTI to N-hydroxysuccinimide (NHS) reactive groups on the surface of an agarose support matrix (BioRad Affi-Gel® 10). These NHS groups react with the side chain amino groups of Lys residues. BPTI only contains 4 Lys residues; furthermore, 1 of these, Lys 15, is the P1 residue of the active site of this reversible protease inhibitor. By coupling BPTI to NHS-agarose without first blocking the active site residue Lys 15, a significant portion of the total BPTI bound to the resin will be inactivated by the reaction of Lys 15 with NHS. To ensure that 100% of the BPTI coupled to the column is active, the BPTI should first be equilibrated with excess acetylated trypsin (Ac-trypsin) to block Lys 15. Ac-trypsin has no free amines available to react with NHS-agarose; consequently, it will not couple to the resin. Because the BPTI-trypsin interaction is
pH-dependent, BPTI can be coupled to NHS-agarose using the other Lys residues while Lys 15 is blocked with Ac-trypsin. This protease can then be eluted from the column by lowering the pH.

It is proposed that the BPTI columns used during this work did not contain a high enough concentration of functionally active BPTI to efficiently adsorb the low concentrations of rhCatG and rHNE present in samples. BPTI affinity matrix prepared using Ac-trypsin, in the manner described above, could circumvent this problem and drastically improve the purification of these enzymes by the well-established method of BPTI affinity chromatography.

**Learning to Fly: Lessons in Recombinant Yeast Fermentation**

The single largest hurdle in the successful expression of large amounts of recombinant proteins by yeast is the control and optimization of growth conditions. Bioreactor-based fermentation methods are strongly recommended over methods that use baffled flasks because of the increased level of control that bioreactors afford. Careful regulation of carbon, nitrogen, oxygen, phosphate, pH, and temperature significantly improve efforts to optimize culture growth conditions. When working with the methanol-inducible AOX1 promoter, it is very important to properly regulate the concentration of methanol in solution to prevent its accumulation to toxic levels. High-yield protein production also requires adequate supplies of oxygen that cannot always be supplied by normal air. The pH of a yeast culture does not remain static throughout its growth and cannot be held constant in baffled flasks. Bioreactors enable the careful regulation and incremental modification of culture growth conditions and facilitate optimization of growth media. Bioreactor-based fermentation of yeast, especially *P. pastoris*, is detailed in the appendix.
REFERENCES


Amberg DC, Burke DJ, Strathern JN. 2006. Preparation of genomic DNA from yeast using glass beads. CSH Protoc 2006(1).


Collins-Racie LA, McColgan JM, Grant KL, DiBlasio-Smith EA, McCoy JM, LaVallie ER. 1995. Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. Biotechnology (N Y) 13(9):982-7.


APPENDIX
ON BIOREACTOR BASED FERMENTATION OF YEAST

Recombinant yeast can be grown to high density in large volumes through the use of bioreactor fermentation; furthermore, bioreactors facilitate monitoring and regulation of culture conditions to provide an optimal growth environment for maximized protein production. Factors such as temperature, acidity, and the bioavailability of several compounds can be controlled with more accuracy and consistency in bioreactors than in the baffled flasks routinely used for culture screening. Despite the additional factors to which the operator must attend, bioreactor fermentation occurs on a time scale similar to that of the baffled flask platform, does not require much additional effort, and results in higher yield of recombinant protein. Numerous review and methods papers have been published (Stratton et al. 1998; Cereghino and Cregg 2000; Cereghino et al. 2002; Zhang et al. 2002; Zhang et al. 2003; Plantz et al. 2006; Li et al. 2007; Bollok et al. 2009; Murasugi 2010; Ramón and Marín 2011; Bawa and Darby 2012; Bora 2012; Damasceno et al. 2012).

Considerations for Media Formulation

Although complex media based on yeast extract and peptone are frequently used for recombinant protein production in yeast, synthetic media provide many advantages, regardless of the scale on which production will occur. Because they are derived from biological sources, yeast extract and peptone are complex mixtures of poorly defined components; however, synthetic media assembled from pure compounds provide a well-defined environment of
specified composition and often result in improved culture growth rates and recombinant protein production.

For the bioreactor fermentation of *P. pastoris*, the formulation prescribed by the *Pichia* Fermentation Process Guidelines (Life Technologies), designated Basal Salts Medium (BSM) and supplemented with PTM1 Trace Salts, is used most frequently and is based on the formulae used during the early development of this recombinant expression platform (Siegel and Brierley 1989; Brierley et al. 1990), although variations on this recipe such as FM22 (Stratton et al. 1998; Wang et al. 2005) with PTM4 (Stratton et al. 1998) are often used (Table A1). All synthetic media must include bioavailable sources of carbon, oxygen, nitrogen, and phosphate in addition to several other ions including especially calcium, magnesium, potassium, and sulfate (Siegel and Brierley 1989). The sources of these components, their concentrations, and the method of their assembly can have significant impact on fermentation results and must be considered in detail.

**Table A1: Fermentation Media Components**

<table>
<thead>
<tr>
<th>Component</th>
<th>BSM</th>
<th>FM22</th>
<th>FM22</th>
<th>PTM1</th>
<th>PTM4</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₃PO₄ (ACS = 85%)</td>
<td>26.7 mL/L</td>
<td>-</td>
<td>-</td>
<td>6 g/L</td>
<td>2 g/L</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>-</td>
<td>42.9 g/L</td>
<td>5 g/L</td>
<td>0.08 g/L</td>
<td>0.08 g/L</td>
</tr>
<tr>
<td>(NH₄)₂SO₄</td>
<td>-</td>
<td>-</td>
<td></td>
<td>NaI</td>
<td>0.08 g/L</td>
</tr>
<tr>
<td>CaSO₄ * 2H₂O</td>
<td>0.93 g/L</td>
<td>1 g/L</td>
<td>1 g/L</td>
<td>MuSO₄ * H₂O</td>
<td>3.0 g/L</td>
</tr>
<tr>
<td>K₂SO₄</td>
<td>18.2 g/L</td>
<td>14.3 g/L</td>
<td>0.2 g/L</td>
<td>Na₂MoO₄ * 2H₂O</td>
<td>0.2 g/L</td>
</tr>
<tr>
<td>MgSO₄ * 7H₂O</td>
<td>14.9 g/L</td>
<td>11.7 g/L</td>
<td>0.02 g/L</td>
<td>H₃BO₃</td>
<td>0.5 g/L</td>
</tr>
<tr>
<td>KOH</td>
<td>4.13 g/L</td>
<td>-</td>
<td>To Adjust pH = 4.5</td>
<td>CoCl₂</td>
<td>0.5 g/L</td>
</tr>
<tr>
<td>Glycerol</td>
<td>40 g/L</td>
<td>40 g/L</td>
<td></td>
<td>ZnCl₂</td>
<td>7 g/L</td>
</tr>
<tr>
<td>Trace Elements</td>
<td>4.35 mL/L PTM1</td>
<td>1 mL/L PTM4</td>
<td></td>
<td>Fe(II)SO₄ * 7H₂O</td>
<td>22 g/L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Biotin</td>
<td>0.2 g/L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H₂SO₄</td>
<td>5 mL/L</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CaSO₄ * 2H₂O</td>
<td>0.5 g/L</td>
</tr>
</tbody>
</table>
Carbon Sources

The carbon sources made available to yeast during fermentation may or may not impact the rate of protein production depending on the chosen transcriptional promoter. For *P. pastoris*, the two promoters most frequently used to direct recombinant expression are the alcohol oxidase 1 (AOX1) promoter and the glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter. The GAP promoter is constitutively active and not sensitive to environmental conditions; however, the AOX1 promoter is repressed by several carbon sources and induced only in the presence of methanol. The GAP promoter is useful for expression of foreign proteins that do not adversely affect culture growth because it does not require any special induction strategy. Such cultures can be grown to high density in glucose or glycerol and they express the gene product throughout the fermentation. Protein expression based on the *K. lactis* platform often uses the galactose-inducible β-galactosidase gene promoter LAC4, which, like AOX1, is repressed by other carbon sources.

Use of an inducible promoter provides tight control over recombinant protein expression, which can be useful for the production of toxic proteins; however, this strategy requires careful control of carbon compounds that are available. The AOX1 promoter is only induced by methanol and is repressed to varying degrees by many sugars and alcohols including glucose, ethanol (by conversion to acetate), and glycerol. Typically, fermentation of yeast using an inducible promoter begins with rapid culture expansion using glycerol as the carbon source followed by induction in the absence of glycerol. Supplementation of culture media with monosodium glutamate (MSG) can improve culture growth without repressing recombinant protein expression. Reducing sugars such as glucose and galactose must be autoclaved separately from BSM containing primary amines to prevent Amadori or Maillard reactions;
however, non-reducing carbon sources like the sugar alcohols glycerol and sorbitol do not react
with primary amines and can be included in the BSM prior to autoclave sterilization. Sorbitol
has been shown to improve protein expression when used as a cofeed with methanol during the
induction phase of growth (Holmes et al. 2009; Wang et al. 2010).

Oxygen Sources

Although comparative investigations of the effects of various gas mixtures on
recombinant expression have not been reported and the effects of excess O₂ are unclear, the
availability of dissolved O₂ gas is fundamental to yeast culture growth. Because all of the
potential pre-induction carbon sources are catabolized through the central glycolytic pathway,
the yeast cells require sufficient oxygen for aerobic respiration at all times to prevent the
accumulation of anaerobic ethanol and acetate that will inhibit recombinant expression driven by
the AOX1 promoter (Inan and Meagher 2001). For Mut⁺ (including strain KM71H) or Mut⁺
(including strains X-33, GS115 or SMD1168H) P. pastoris as well as all strains of K. lactis, the
same rationale applies to the induction phase.

In the case of Mut⁺ P. pastoris fermentation, the cells metabolize methanol like the wild-
type and consume large amounts of methanol during the induction phase; furthermore, methanol
catabolism requires oxygen during two phases (Veenhuis et al. 1983; Cereghino and Cregg
2000). First, the conversion of methanol to formaldehyde by alcohol oxidase is directly
dependent on oxygen and methanol will accumulate to lethal levels in the prolonged absence of
sufficient oxygen. Second, the subsequent conversion of formaldehyde to CO₂ is a 2-step
oxidation accompanied by the total reduction of 2 NAD⁺ to 2 NADH that then drive aerobic
respiration. In the event of oxygen depletion, NADH cannot be oxidized to replenish the NAD\(^+\) pool sufficiently to prevent the toxic accumulation of intracellular formaldehyde.

![Diagram of catabolism of methanol](image)

**Figure A1: Catabolism of Methanol.** The catabolism of methanol is a three step process that requires oxygen, both directly in the oxidation of methanol to formaldehyde and indirectly in the oxidation of NADH by the electron transport chain (e\(-\) Transfer) to replenish intracellular NAD\(^+\). Mut\(^+\) strains, such as KM71H, grow slowly on methanol because the AOX1 gene has been disrupted and the cells must use the weaker AOX2 gene for methanol metabolism.

**Nitrogen Sources**

Bioavailable nitrogen is provided as ammonium (NH\(_4^+\)), either in the form of ammonium hydroxide or ammonium sulfate. The choice of nitrogen source for the growth media depends on the acidity of the initial solution. For media formulated with phosphoric acid, such as BSM, the amount of ammonium hydroxide required to increase the pH supplies sufficient initial nitrogen content for the yeast. Other media may use less acidic phosphate sources like potassium phosphate or sodium hexametaphosphate (SHMP) that do not require neutralization by ammonium hydroxide. In these cases, the media must be supplemented with ammonium sulfate as it provides a nitrogen source without affecting the pH. Regardless of the initial formulation, metabolically active *P. pastoris* continuously acidify the culture media and consume nitrogen during fermentation; consequently, the bioreactor monitors the culture pH and can be set to
automatically add ammonium hydroxide to maintain the desired pH while feeding the cells nitrogen.

**Phosphate Sources**

The BSM recipe includes phosphoric acid as the source of bioavailable phosphate whereas the FM22 medium uses monobasic potassium phosphate. While these compounds affect the pH differently and determine which nitrogen source to include, they both contain inorganic orthophosphate, $\text{PO}_4^{3-}$. The $\text{PO}_4^{3-}$ ion is required by all organisms and must be abundant for successful fermentation; furthermore, orthophosphate provides a pH buffer for the growth media. Unfortunately, many salts of orthophosphate exhibit poor solubility, including calcium, iron, and zinc salts. Fermentation media made with orthophosphate typically contain some insoluble phosphate salts and the amount depends on the temperature and pH of the solution as well as the calcium, iron, zinc, and phosphate concentrations. Because the PTM1 trace salts solution contains more ferrous sulfate and zinc chloride than PTM4, fermentation media containing PTM1 typically develop more insoluble precipitate than those containing PTM4.

Sodium hexametaphosphate (SHMP, or “phosphate glass”), $(\text{NaPO}_3)_6$, can be substituted for orthophosphate to prevent the formation of insoluble salts; however, SHMP does not buffer pH effectively and decomposes into orthophosphate at elevated temperature. In media formulations that use SHMP instead of orthophosphate, a suitable pH buffer such as MES is suggested in order to facilitate bioreactor regulation of pH; however, acetate must be avoided as a pH buffer because it represses the AOX1 promoter (Inan and Meagher 2001). SHMP must be
sterilized separately by ultrafiltration and added aseptically after the sterile fermentation vessel has cooled to prevent SHMP decomposition into orthophosphate.

**Trace Elements and Other Supplements**

The final fermentation media require a number of trace elements that are combined into a filter-sterilized, concentrated supplement and added to cool, sterile BSM in the fermentation vessel. The two most commonly used trace element supplements are PTM1 and PTM4 and these formulations vary primarily in the relative concentrations of several components. PTM1 was found to produce more insoluble precipitate at neutral to alkaline pH than PTM4, despite additional calcium, likely because of the additional masses of ferrous sulfate and zinc chloride in PTM1 that form insoluble orthophosphate salts in the media. It has been shown that increased levels of biotin supplementation improve recombinant protein expression in *P. pastoris* (Jungo et al. 2007).

While both PTM1 and PTM4 supplements contain biotin, it is suggested here that additional biotin be provided. Similarly, fermentations of *P. pastoris* or *K. lactis* may or may not benefit from a supplementary vitamins mix, made as a 100x concentrate containing 0.1 g/L L-His-HCl, 0.2 g/L DL-Met, 0.2 g/L DL-Trp, 0.0002 g/L biotin, 0.0002 g/L folic acid, 0.2 g/L inositol, 0.04 g/L niacin, 0.02 g/L p-aminobenzoic acid, 0.04 g/L pyridoxine HCl, 0.02 g/L riboflavin, 0.04 g/L thiamine HCl, 0.04 g/L calcium pantothenate, 10 g/L CaCl₂. These optional supplements do not impair but may improve culture growth and protein expression; furthermore, they are components of the Yeast Nitrogen Base (YNB) generally used in many yeast media.

Bioavailability of free amino acids in the fermentation media may improve secreted recombinant protein expression by reducing the demand for amino acid biosynthesis and
alleviating the drain placed on the cellular amino acid pool by the overexpression of recombinant proteins. Proteins directed for secretion into the growth media may be susceptible to inactivation or degradation through the actions of secreted neutral proteases and amino acid supplementation of the media may reduce the expression of these yeast proteases. Casamino acids are one option, but they are a complex mixture of amino acids and peptide fragments from the acid hydrolysis of casein. As an alternative, a mixture of pure amino acids can be assembled with a known, controlled composition.

**Fermentation Medium and Bioreactor Setup and Assembly**

Prior to culture growth, the sterilization and assembly of the fermentation medium and the bioreactor vessel must occur. This process consists of 3 sequential phases of assembly, based on sterility and temperature, where each phase includes a prescribed set of steps required for correct fermentation conditions and bioreactor operations. The procedures described here demonstrate the use of the BioFlo® 110 Bioreactor (New Brunswick) but also apply to other similar apparatus.

**Setup Phase 1: Presterile Assembly**

**Initial Medium Preparation.** To begin, the initial components of the fermentation medium are solubilized into deionized water and added into the fermentation vessel. The initial volume for a 7.5 L vessel should be approximately 2.5-3.0 L of medium containing enough of each ingredient per liter for 3.5 L of total medium.

The synthetic fermentation media formulations tried during the production of recombinant enzymes described in Chapters 2, 3, and 4 are variations based on the BSM recipe
recommended by Life Technologies and the FM22 recipe created by Stratton et al. (Stratton et al. 1998). The formulae of these media are provided in Table A1, above. Some formulations are supplemented with 10 g/L monosodium glutamate (MSG) (Boettner et al. 2002).

Provided that the phosphate source is an orthophosphate like phosphoric acid or potassium phosphate, it can be included at this time. SHMP must be filter sterilized and added during setup phase 3 because it decomposes into orthophosphate at elevated temperatures. A nonreducing carbon source like glycerol can also be included here; however, reducing sugars such as glucose and galactose must be autoclaved separately and added during setup phase 3. If the phosphate source is phosphoric acid, then no nitrogen source is required in the initial solution; however, 5 – 10 g/L ammonium sulfate should be included if the phosphate source is a potassium salt.

**Bioreactor Setup and Sterilization.** The fermentation vessel containing the initial medium preparation is assembled with the head plate and rubber seal seated securely. During heat sterilization, all head plate openings into the container must be sealed with the exception of the exhaust tube, which should be stoppered with a gauze plug. For the external opening of the metal oxygen sparging pipe, polymer tubing such as Tygon® should be attached, with a slip knot tied tightly in the free end; otherwise, pressure from the heated gas in the vessel will push liquid out through this opening. All other metal ports that open into the interior of the vessel should be similarly sealed with knotted tubing. An image of the head plate, prepared for autoclave sterilization, is shown (Figure A2).
Figure A2: Bioreactor Vessel Head Plate. The bioreactor head plate is shown, from above, immediately prior to autoclave sterilization. All important components of the head plate are labeled. All ports that expose the vessel interior are “capped” or “plugged” and all ports that do not expose the vessel interior are “open.”

The bioreactor makes use of a dissolved oxygen (dO₂) sensor to monitor the percentage of dissolved oxygen in the media, relative to a calibrated range of dO₂ content. This metal probe must be handled very carefully to avoid damaging the exposed membrane at its tip; however, it can be mounted in the head plate prior to the subsequent sterilization of the vessel. The calibration of this probe occurs during setup phase 3. The probe contains an electrolyte solution that may require infrequent replenishment and a bottle of this is provided by the manufacturer.
The pH probe can be mounted in the head plate prior to autoclave treatment; however, this probe must be calibrated using standardized pH buffers prior to sterilization. As an alternative, the pH probe can be sterilized with ACS methanol immediately prior to insertion into a sterile vessel. The calibration of this probe requires standardization to two set points, “zero” at pH = 7.0 and “span” at pH = 4.0.

Autoclave sterilization proceeds on a liquid cycle lasting 30 minutes at 121 °C.

**Setup Phase 2: High Temperature Sterile Assembly**

Once the fermentation vessel and medium have been heat sterilized, the vessel’s closed-circuit coolant piping is connected to an external liquid cooling unit containing antifreeze. This cooling unit is turned on and set to 25 °C; furthermore, care must be taken to ensure that the coolant reservoir is full and that all air is bled from the coolant lines because this can impede proper temperature regulation. A temperature sensor connected to the bioreactor control module (“the tower”) is inserted into a metal pipe that extends into but does not open on the interior of the vessel. Until the vessel and medium have completely cooled, the oxygen sparger tubing should remain tied and the exhaust tube should remain plugged with gauze. The electrical impeller motor can be mounted on the impeller driveshaft and connected to the tower, ensuring that the rubber seal of the driveshaft is always well-lubricated with vacuum grease. The impeller can be set at low speed (≤ 350 rpm) to reduce vessel cooling time or the vessel can be allowed to cool overnight. If the pH probe was not calibrated and inserted prior to vessel sterilization, this should be done now. The dO2 probe must be connected to the tower and allowed to polarize for a minimum of six hours prior to calibration in setup phase 3.
Setup Phase 3: Final Assembly

After the vessel has cooled to between 25 °C and 30 °C, the oxygen sparger tubing can be connected to the oxygen supply tank via a regulator, an airflow meter, and a 0.22 µm in-line filter. The exhaust tube can be connected to the liquid cooled exhaust condenser that is attached to tubing that bubbles exhaust gas through dilute chlorine bleach in an Erlenmeyer flask. The sampling apparatus is mounted on the head plate and connected by tubing to the metal sampling pipe. The impeller should be set at 350 rpm. Images of the connected bioreactor vessel (Figure A3) and bioreactor control tower (Figure A4) are shown.

**Figure A3. Bioreactor Vessel.** The bioreactor vessel is shown that illustrates the connection of the assembled fermenter. The impeller motor has been mounted on the impeller driveshaft. The impeller motor, pH probe, dO2 probe, and temperature sensor are connected to the bioreactor control tower (see Figure A4). The air inlet hose, the exhaust hose, and the main coolant lines are obscured in this image. Refer to Figure A2 for locations of these ports.
**Figure A4. Bioreactor Control Tower.** The bioreactor control tower is shown with all connections to the bioreactor vessel (left). The pump head contains four peristaltic pumps regulated by the control tower. These pumps can be programmed to feed acid, base, antifoam, or other agents either at constant rates or in response to pH and foam sensors. A separate pump is used for delivering the carbon feed for improved volumetric regulation of the feed rate. The bioreactor control tower is modular and can be arranged in other ways. The digital interface unit (computer) for the control tower is stacked above the dO2/pH probe inputs and cannot be seen in this image (see Figure A5).

**Final Medium Assembly and pH adjustment.** The remaining components of the fermentation medium are added aseptically at this time into the stirring fermentation vessel. The trace elements solution (e.g. PTM1 or PTM4) is added to the appropriate final dilution. Extra sterile components such as biotin, a vitamins mix, amino acid supplements, SHMP, and reducing sugars can also be added at this time. Once these components have been added, the bioreactor is setup for pH control.
Figure A5. Bioreactor Digital Interface. The digital interface of the control tower is shown. This computer monitors and displays sensor data in real-time. The context-specific buttons make selections as indicated adjacently on the digital screen.

The tower can be programmed through the digital interface to activate acid and/or base feed pumps in response to deviations of culture pH from the set point. An image of the digital interface (Figure A5) is shown for reference, see the bioreactor user’s manual for details. In the case of *P. pastoris*, the growing culture acidifies the media; consequently, the base pump is setup to provide a regulated drip feed from a reservoir of ACS ammonium hydroxide diluted 1:1 into deionized water. The tubing from the base pump is connected to the head plate of the fermentation vessel. Only then is the tower set to regulate the pH and the pump turned on. If the pH of the broth is below the desired set point, then the bioreactor will add ammonium hydroxide at this time to adjust the pH. Culture pH affects the activity of extracellular proteases that can degrade recombinant protein secreted into the medium. The unwanted activity of neutral proteases can be prevented by the addition of broad spectrum protease inhibitors or reduction of
the culture pH when inhibitors are undesirable; consequently, *P. pastoris* fermentations are frequently conducted at pH < 5.0, especially during induction.

**dO2 Probe Calibration.** To ensure proper monitoring of the dissolved oxygen content of the culture, the dO2 probe must be calibrated. First, the “zero” set point is established. This can be accomplished by disconnecting the probe from the tower just long enough for the raw reading to stabilize (about 2 seconds) or by sparging the medium with pure nitrogen gas until the raw reading stabilizes. After the minimum raw reading has been reached, the “zero” set point is set to 0 %. Second, the “span” set point must be established. For this, the oxygen flow rate is set to 1 vvm (volume of gas per volume of culture per minute), e.g. 3.5 LPM (liters per minute) of oxygen for a 3.5 L liquid volume, and the impeller motor is set to its highest speed (1200 rpm for a New Brunswick BioFlo-10 bioreactor). The fermenter must then be left undisturbed until the raw reading reaches a stable maximum, approached asymptotically, at which point the “span” set point is entered as 100.0 %, representing the maximum amount of oxygen driven into solution at equilibrium by the given gas supply under maximum stirring conditions. Physical chemistry dictates that the equilibrium of dissolved O\(_2\) at constant temperature and pressure is directly proportional to both the partial pressure of O\(_2\) in the gas and the surface area of the gas/liquid interface; consequently, pure oxygen and maximum stir rates maximize the absolute amount of O\(_2\) dissolved at equilibrium. Once the probe is calibrated, the O2 supply can be turned off until the culture is inoculated.

Throughout the fermentation of yeast, assessments of culture health and metabolism rely on the careful observation and accurate interpretation of changes in dO2 content and impeller behavior because the respiration reflects carbon consumption, both critical components for high yield culture expansion and protein production.
Yeast Fermentation

Bioreactor based fermentation of yeast using a repressible/derepressible transcriptional promoter like AOX1 and LAC4 consists of four culture growth stages: inoculum growth, pre-induction batch phase, pre-induction fed-batch phase, and induction fed-batch phase. The culture inoculum expands while consuming the available carbon during the preinduction batch phase. Upon carbon depletion, the pre-induction fed-batch phase further increases the culture density with the drip-fed addition of concentrated pre-induction carbon source at a regulated rate. A culture of sufficient density is then transitioned to a drip-feed of the carbon source required for the induction fed-batch phase. Yeast transformants using a constitutive promoter only require a batch phase followed by a fed-batch phase, with no distinction based on induction; however, the general procedures remain the same.

Inoculum Growth

Two days prior to fermenter inoculation, a 10 mL liquid seed culture of the yeast is set to grow with shaking at 30 °C overnight in YPD with antibiotics (100 µg/mL Ampicillin and, for P. pastoris, 100 µg/mL Zeocin) to which the strain is resistant. On the day before fermenter inoculation, the starter culture in a nutrient-rich synthetic minimal medium (SMM) is seeded from the YPD culture. This starter culture is grown for 24 – 36 hours with shaking at 30 °C in a 1 L baffled shaker flask that is filled approximately 1/3 with SMM containing the pre-induction carbon source, often glycerol. The starter culture should be grown to high density and the volume of inoculum should not exceed approximately 10 % of the initial fermentation volume.

Inoculation of the fermentation vessel proceeds only after the completion of bioreactor setup and assembly. The entire starter culture volume is added aseptically to the fermentation
vessel and a 25 ml sample is immediately withdrawn and centrifuged to determine the wet cell pellet weight (WCW) in grams per liter of medium. Throughout the fermentation process, such 25 ml samples will be collected at least once per day and at the start of every phase of fermentation; furthermore, the WCW data should be graphed over time to monitor the health and growth rate of the culture, which should continue to expand throughout the fermentation.

**Preinduction Batch Phase**

Once the fermentation vessel has been inoculated, the culture enters the pre-induction batch phase, during which it consumes the supply of bioavailable carbon in the broth. At the time of inoculation, the bioreactor control tower should be set up to automatically regulate the dO$_2$ by adjusting the impeller speed. The dO$_2$ set point should be no less than 35% to ensure that the actual dO$_2$ of the culture does not drop below 20%; furthermore, the impeller speed range should be set between a minimum of 350 rpm and the maximum allowable speed (1200 rpm for a New Brunswick BioFlo-10 bioreactor) for automatic speed control.

The starting dO$_2$ will likely be above 95% due to saturation from probe calibration. As the culture grows, the dO$_2$ will gradually fall until it drops below the dO$_2$ set point, at which time the impeller speed will be automatically increased to compensate for the increased oxygen demand. Changes in the dO$_2$ content of the medium reflect the metabolic health of the culture and the bioavailability of carbon; consequently, careful dO$_2$ monitoring throughout all stages is critical to successful fermentation.

The preinduction batch phase lasts until the culture has depleted the bioavailable carbon in the growth medium. The end of this phase is marked by a reduced oxygen demand of the
yeast in response to the lack of carbon and a significant increase of the dO₂ above the programmed set point, despite a reduction in impeller speed to its minimum.

**Preinduction Fed-Batch Phase**

At this time, a steady pump-fed, drop wise addition of the pre-induction carbon source is initiated. A sterile solution of 50 % glycerol in deionized water is recommended with a feed rate of 1 mL/min (Bio-Rad Econo peristaltic pump) for a 3.5 L culture volume (17 – 18 mL/hr per liter of initial volume). The culture will respond to the addition of fresh bioavailable carbon by increasing the oxygen demand, as evinced by a steady reduction of dO₂ and an increased impeller speed. This phase lasts until the culture density has reached a wet cell pellet weight (WCW) of 180 – 200 g/L, approximately 4 – 6 hours.

By the end of the preinduction fed-batch phase and throughout the rest of the fermentation, the culture should also be carbon limited, with a carbon consumption rate equal to the carbon addition rate so that no excess feed accumulates in the medium. A carbon limited culture runs out of carbon source very shortly after the drip feed is turned off and this results in a rapid increase in dO₂ (a dO₂ spike), usually within two minutes. Carbon limitation at the end of the pre-induction fed-batch phase ensures that no promoter repressing carbon compounds interfere with the induction phase.

**Induction Fed-Batch Phase**

A carbon limited yeast culture of WCW = 180 – 220 g/L is ready for induction, which is initiated simply by changing the carbon feed. In the case of *P. pastoris* controlled by AOX1, the transition to methanol must occur in stages to allow the yeast time to adapt. The initial feed rate
of pure methanol, 3.6 mL/hr per liter of initial fermentation volume, is maintained for 3–4 hours, until the dO\textsubscript{2} holds steady and dO\textsubscript{2} spikes occur quickly. At this time, the methanol feed rate is doubled to 7.2 mL/hr per liter of initial volume and maintained for 2-3 hours, after which the culture is tested again for carbon limitation by the dO\textsubscript{2} spike method. If a dO\textsubscript{2} spike occurs shortly after terminating the methanol feed, then the culture is carbon limited and the feed rate is increased to 10.9 mL/hr per liter, where it remains until the end of fermentation. The induction fed-batch phase usually lasts for about 72 hours, with an average feed rate of 10.4 mL/hr per liter. These feed rates follow the recommendations of the Life Technologies *Pichia* Fermentation Process Guidelines. Not all transformants will produce proteins optimally under these conditions; consequently, the feed rates and fermentation time may require adjustment to maximize protein production.

It has been shown that supplementation of the methanol feed with sorbitol improves culture growth and protein yields (Holmes et al. 2009; Wang et al. 2010). Because it does not dissolve in pure methanol, sorbitol should be fed through a separate line rather than dissolved into methanol diluted with water. Diluted methanol requires a faster feed rate than pure methanol to provide the same amount of carbon, which dilutes the culture and reduces the growth rate; furthermore, the volume of water added to the fermentation during fed-batch phases should be minimized to prevent the culture volume from exceeding the capacity of the vessel.

Carbon limitation should be evaluated periodically throughout the induction fed-batch phase to ensure that methanol does not accumulate in the vessel over time, as excess methanol is toxic. To verify that the culture is carbon limited, the drip feed is stopped and the time to the dO\textsubscript{2} spike is observed by monitoring both the dO\textsubscript{2} reading and changes in the impeller speed. It is important to confirm that an increase in dO\textsubscript{2} is actually a spike and not just in response to an
increase in the impeller speed. For a true spike, the dO\textsubscript{2} will continue to rise even as the impeller speed falls.

**Culture Harvest**

After completion of the induction fed-batch phase, the culture is harvested from the fermentation vessel. Although several approaches are available for recovering the culture volume, the use of oxygen pressure is suggested to drive the broth out through the sample collection tube. This can be achieved by clamping the exhaust tubing, unclamping the sample collection tubing, and collecting the stream of liquid directly into centrifuge bottles, which minimizes the amount of clean up that is required. The yeast should be centrifuged with a relative centrifugal force of 5,000 – 10,000 x g for 10 – 20 minutes or until all cells have formed a pellet. For proteins that are retained by the yeast, the cells can be lysed by a variety of methods including mechanical disruption (e.g. glass beads), enzymatic treatment (e.g. lyticase), repeated freeze/thaw cycles, sonication, or the use of a French pressure cell. For proteins secreted by the yeast, the culture supernatant is further clarified after separation from the cell pellet. The supernatant can be passed through glass fiber prefilters as well as submicron filter membranes made of low protein binding materials like polyethersulfone (PES).
VITA

ELIOT THOMAS SMITH

Personal Data: Date of Birth: December 29, 1980
Place of Birth: Johnson City, Tennessee
Marital Status: Single

Education: B.S. Biological Sciences, University of Tennessee, Knoxville, Tennessee, 2006
Ph.D. Biomedical Sciences, East Tennessee State University, Johnson City, Tennessee, 2013

Professional Experience: Graduate Assistant, East Tennessee State University, College of Medicine, 2007 – 2013


Poster Presentations: GRC/GRS, Proteolytic Enzymes and Their Inhibitors, 2012
“Expression of Recombinant Human Neutrophil Elastase in \textit{Pichia pastoris}”

EB/FASEB/ASBMB Conference, 2012
“Recombinant Expression of Human Enteropeptidase Light Chain in \textit{Pichia pastoris}”

EB/FASEB/ASBMB Conference, 2011
“Expression of Recombinant Human Neutrophil Elastase in \textit{Kluyveromyces lactis}”

Appalachian Research Forum, 2011
“Expression of Recombinant Human Neutrophil Elastase in \textit{Kluyveromyces lactis}”

Appalachian Research Forum, 2010
“Recombinant Human Cathepsin G Expression in \textit{Kluyveromyces lactis}”
Protein Society Conference, 2009
“Expression of Recombinant Human Enterokinase Light Chain in Kluyveromyces lactis”

Appalachian Research Forum, 2009
Expression of Human Neutrophil Cathepsin G in K. lactis with eGFP Fusion for Detection

Oral Presentations:
Appalachian Research Forum, 2013
“Human enteropeptidase light chain: Bioengineering of recombinants and kinetic investigations of Structure and function”

Appalachian Research Forum, 2012
“Expression of Recombinant Human Neutrophil Elastase in Pichia pastoris”

Honors and Awards:
ASBMB Graduate and Postdoctoral Travel Award, 2012

Appalachian Research Forum, 2012
1st Place Oral Presentation: Biomedical & Health Sciences and Natural Sciences

Appalachian Research Forum, 2011
2nd Place Poster Presentation, Division III: Biomedical Sciences

Service and Committees:
Biomedical Sciences Graduate Student Association
President, 2010-2012

Liaison Committee on Medical Education Task Force Committee
Graduate Student Representative 2010-2011