

**THE CBM9 FUSION TAG – A NEW TECHNOLOGY
FOR INEXPENSIVE PRODUCTION AND AFFINITY
PURIFICATION OF RECOMBINANT PROTEINS**

by

Mojgan Kavooosi

B.Sc., The University of British Columbia, 1993

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF**

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Chemical and Biological Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

October 2007

© Mojgan Kavooosi, 2007

Abstract

Downstream processing of proteins and other biological products has long been dominated by packed-bed chromatography (Rankin 2003). Despite the generally high cost of the technique, chromatography remains widely used because it offers extraordinarily high resolution under conditions that do not denature or alter the chemistry of the product, an imperative for therapeutic proteins since purity and activity are strict requirements. However, the inability of patients and governments to meet the rising costs of healthcare, particularly the cost of recombinant protein therapeutics, and the sharp increase in competition over the past decade for market share of recombinant-protein based treatments of major illnesses have led to intense downward pressure on the cost of goods, especially for high volume products such as monoclonal antibodies and other recombinant proteins (Morrow 2002). Industry is therefore seeking to develop more cost effective downstream processes, including cheaper and more selective forms of chromatography.

Generic affinity chromatography based on affinity-tag technology has the potential to simplify downstream processing by achieving higher yields and purities than conventional modes of chromatography. However, the high cost of current affinity tag technologies, due mainly to the expense of their associated affinity chromatography media, limits their application at production scales. This thesis addresses this problem by reporting on a novel affinity chromatography platform utilizing the family 9 carbohydrate-binding module (CBM9) of xylanase 10A from *T. maritima*, a new affinity tag that binds to both soluble sugars and insoluble cellulose to permit the highly efficient capture and purification of CBM9-tagged fusion proteins on a very inexpensive cellulose-based affinity media. Development of this technology has required (i) design of a generic CBM9 expression vector for production of chimeric fusions containing an N-terminal CBM9, a linker region containing a suitable processing site at its C-terminus for efficient removal of the affinity tag following affinity purification, and a C-terminal target protein, (ii) development of an effective strategy to design a linker sequence to stably connect the CBM9 tag to the target protein and to permit efficient tag removal through enzyme-catalyzed cleavage, (iii) derivation and validation of a mathematical model to predict

binding and elution behavior of CBM9 fusion proteins on a high-capacity cellulose column, (iv) solutions to certain technology scale-up issues, including the synthesis of a mechanically stable stationary phase, and finally, (v) validation of the performance of the technology in terms of product yield, purity and concentration factor.

Two bioinformatics-based strategies were developed to successfully identify a linker with improved resistance to endogeneous proteases of the host when compared against the popular poly-glycine based linker. A simple and effective assay was developed to identify the optimal conditions for efficient tag removal post-purification. The technique, based on Luminescence Resonance Energy Transfer (LRET) prescreens a library of linkers and processing enzymes to identify a CBM9-target protein fusion with enhanced processing efficiency. A novel two-zone model (TZM) of pore diffusion is presented to describe the rate of uptake of CBM9 fusion proteins within the stationary phase of the associated affinity chromatography column and thereby provide improved predictions of product breakthrough, including elution behavior from a bacterial lysate feed. Finally, a mechanically stable cellulose-based chromatography media was synthesized to allow preparative-scale purification of recombinant proteins using CBM9. A fixed-effect two-way response surface methodology was used to optimize the concentrations of the two primary reactants, epichlorohydrin and dimethyl sulfoxide (DMSO), required to cross-link the starting material, Perloza™ MT100, a compressible cellulose-based chromatography resin. This resulted in a cross-linked affinity chromatography media capable of operating at an order-of-magnitude higher linear velocity than permitted by unmodified MT100. In sharp contrast to MT100, the mechanical stability and purification performance of the cross-linked media are not diminished by scale-up or repeated column use. The results of this thesis thereby provide industry with a ready-made expression vector that can be used to express any target protein as a CBM9 fusion protein and to then inexpensively purify the target recombinant protein at an overall level of performance that is either superior or comparable to current commercially available fusion-tag technologies.

Rankin P: Perilous economics of the industry. *The Source* 2003, Sept issue, pp. 5-10.
J. Morrow: Economics of antibody production, *Genet Eng News* 22 (2002), pp. 1-39

Table of Contents

Abstract.....	ii
Table of Contents.....	iv
List of Tables.....	ix
List of Figures.....	xi
Nomenclature.....	xviii
Dimensionless numbers.....	xxi
Greek letters.....	xxii
Acknowledgements.....	xxiv
Co-Authorship Statement.....	xxvi
1 Introduction, Background and Thesis Objectives.....	1
1.1 Overview.....	1
1.2 Literature Review.....	3
1.2.1 Downstream Processing in Biotechnology.....	3
1.2.2 Affinity Chromatography in Preparative-Scale Protein Purification.....	4
1.2.3 “Natural” Affinity Chromatography.....	5
1.2.4 Affinity Fusion-Tag Technology.....	6
1.2.5 Factors to Consider in Designing the Affinity Tag System.....	7
1.2.5.1 The Choice of Peptide Versus Protein Tags.....	7
1.2.5.2 Position of the Affinity Tag.....	8
1.2.5.3 Expression of Fusion Proteins in Host Cells.....	9
1.2.5.4 Tag Removal Post-Purification.....	10
1.2.5.5 Affinity Tag-Ligand Interaction.....	11
1.2.6 Affinity Tag Systems.....	12
1.2.7 Carbohydrate Binding Modules.....	15
1.2.8 The Family 9 Carbohydrate-Binding Module (CBM9).....	17
1.2.9 Cellulose-Based Chromatography Resins.....	18
1.3 Thesis Objectives.....	18
1.4 Tables.....	20
1.5 Figures.....	24
1.6 References.....	26

2	Inexpensive One-Step Purification of Polypeptides Expressed in <i>E. Coli</i> as Fusions With the Family 9 Carbohydrate-Binding Module of Xylanase 10A from <i>T. Maritima</i>	44
2.1	Introduction.....	44
2.2	Materials and Methods.....	47
2.2.1	Reagents.....	47
2.2.2	Cloning of CBM9-GFP Fusion Protein	47
2.2.3	Protein Production	48
2.2.4	Affinity Chromatography.....	49
2.2.5	Tag Cleavage by Factor X _a	49
2.2.6	Fluorescence Calibration Curves	50
2.2.7	Measurement of Binding Isotherms.....	50
2.3	Results and Discussion	51
2.3.1	Binding Isotherms and Thermodynamics	52
2.3.2	Fusion Protein Expression and Stability	53
2.3.3	Affinity Purification on Perloza™ MT100 Column.....	54
2.3.4	Column Reusability	55
2.3.5	Removal of the CBM9-S ₃ N ₁₀ -IEGR Affinity Tag Using an Immobilized Factor X _a Column	55
2.4	Conclusions.....	56
2.5	Tables.....	58
2.6	Figures.....	60
2.7	References.....	65
3	Strategy for Selecting and Characterizing Linker Peptides for CBM9-Tagged Fusion Proteins Expressed in <i>E. Coli</i>	68
3.1	Introduction.....	68
3.2	Materials and Methods.....	71
3.2.1	Reagents.....	71
3.2.2	Cloning of CBM9-Linker-FX _a -GFP Fusion Proteins	72
3.2.3	Fusion Protein Production and Purification.....	72
3.2.4	Binding Isotherm Measurement and Analysis.....	73
3.2.5	Linker Stability Analysis	73
3.2.6	Factor X _a Processing Analysis.....	74
3.2.7	Differential Scanning Calorimetry Studies.....	75

3.2.8	LRET Studies.....	75
3.3	Results and Discussion	78
3.3.1	Linker Selection and Screening Using MEROPST TM	78
3.3.2	Impact of the Linker on Fusion-Tag Performance.....	80
3.3.3	Determination of Characteristic Distances Using Luminescence Resonance Energy Transfer	82
3.3.4	Influence of the Linker on the Thermodynamic Stability of CBM9-Linker-FX _a -GFP Fusion Proteins.....	83
3.4	Conclusions.....	84
3.5	Tables.....	86
3.6	Figures.....	92
3.7	References.....	98
4	Direct Measurement of the Kinetics of CBM9 Fusion-Tag Bioprocessing Using Luminescence Resonance Energy Transfer (LRET)	107
4.1	Introduction.....	107
4.2	Methods and Materials.....	111
4.2.1	Materials	111
4.2.2	Cloning and purification of fusion protein substrates.....	111
4.2.3	Preparation of apo-fusion protein	112
4.2.4	Isothermal titration calorimetry studies	112
4.2.5	LRET-based assay of fusion-tag cleavage kinetics	112
4.2.6	Gel-based analysis of EKmax TM cleavage reaction.....	115
4.3	Results and Discussion	115
4.3.1	Characterization of energy transfer from bound Tb(III) to GFP	116
4.3.2	Fusion-tag bioprocessing kinetics.....	117
4.4	Tables.....	120
4.5	Figures.....	121
4.6	References.....	129
5	A Novel Two-Zone Protein Uptake Model for Affinity Chromatography and Its Application to the Description of Elution Band Profiles of Proteins Fused to a Family 9 Cellulose Binding Module Affinity Tag.....	133
5.1	Introduction.....	133
5.2	A Proposed Generalized Two-Zone Model of Affinity Chromatography.....	137
5.3	Two-Zone Model Solution Algorithm.....	140

5.4	Materials and Methods.....	141
5.4.1	Chromatographic Media and Reagents.....	141
5.4.2	Scanning Electron Micrographs.....	141
5.4.3	Protein Production.....	142
5.4.4	Equilibrium binding isotherms.....	142
5.4.5	Confocal Laser Scanning Microscopy (CLSM).....	143
5.4.6	Characterization and Application of Perloza™ MT100/G15 Composite Media Column.....	144
5.5	Results and Discussion.....	144
5.5.1	Geometric and Sorption Properties of Perloza™ MT100/Sephadex G15 Composite Media.....	144
5.5.2	Purification of CBM9-GFP on Perloza™ MT100/G15 Composite Media Column.....	146
5.5.3	Characterization of Solute Mass Transfer Within Perloza™ MT100/G15 Composite Columns.....	146
5.5.4	CLSM-Derived Rates of CBM9-GFP Uptake.....	150
5.5.5	Simulation of Breakthrough Curves.....	151
5.5.6	CBM9-GFP Breakthrough from a Clarified Cell Extract Feed.....	152
5.6	Conclusions.....	153
5.7	Tables.....	154
5.8	Figures.....	157
5.9	References.....	172
6	A Mechanically Stable Porous Cellulose Media for Affinity Purification of CBM9-Tagged Fusion Proteins.....	178
6.1	Introduction.....	178
6.2	Materials and Methods.....	181
6.2.1	Reagents.....	181
6.2.2	Cross-Linking of MT100.....	181
6.2.3	Hydrodynamic Characterization.....	182
6.2.4	Measurement of Binding Isotherms.....	182
6.2.5	Affinity Purification of CBM9-GFP.....	183
6.3	Results and Discussion.....	183
6.3.1	Response Surface Methodology to Improve Mechanical Properties.....	183
6.3.2	Comparison of Sorption Properties.....	187

6.3.3	Comparison of Column Parameters and Properties	187
6.3.4	Preparative Scale Affinity Purification of CBM9-GFP on CRL100-7... ..	189
6.3.5	Column Reusability	190
6.4	Conclusions.....	191
6.5	Tables.....	192
6.6	Figures.....	197
6.7	References.....	208
7	Conclusions and Recommendations	212
7.1	References.....	216

List of Tables

Table 1.1	Affinity tag systems described in the literature. Tag size and complementary ligand immobilized on the stationary phase are given. Tags that are currently offered commercially are denoted by *.	20
Table 1.2	Estimated costs of several commercially available affinity tag technologies. Prices provided by respective vendors.	23
Table 2.1	Langmuir adsorption parameters (equilibrium association constant K_a and binding capacity q_i^{\max}) for binding of CBM9 and CBM9-GFP to Perloza™ MT100 at 4°C. Solvent contains pure protein in high-salt buffer.	58
Table 2.2	Binding affinity and capacity of CBM9-GFP on various cellulosic resins.	58
Table 2.3	Summary of purification of CBM9-GFP on Perloza™ MT100 at 4°C.	59
Table 2.4	CBM9-GFP yield and purity for consecutive purification runs through the same Perloza™ MT100 column.	59
Table 3.1	Oligonucleotides used in the construction of CBM9-Linker-IEGR-GFP fusion proteins. Restriction sites are underlined and the Factor X_a recognition sequence is in bold.	86
Table 3.2	Linker lengths, sequences, and susceptibility to proteolytic cleavage as predicted by MEROPS™ and confirmed experimentally by LC/MS/MS.	87
Table 3.3	Regressed Langmuir isotherm parameters for binding of CBM9-Linker-IEGR-GFP on Perloza™ MT100 at pH 7 (4°C). Standard deviations (σ) computed from triplicate measurements.	88
Table 3.4	Half-lives for Factor X_a cleavage of CBM9-Linker-IEGR-GFP fusion proteins. All experiments were conducted in 50 mM potassium phosphate buffer (pH 7, 21 °C) at a fusion protein to Factor X_a concentration ratio of [1000] to [1].	89
Table 3.5	Average relative distance of separation R/R_p between the bound terbiums on CBM9 and the fluorescent chromophore of GFP determined by LRET at 21°C.	90
Table 3.6	Melting temperatures (T_m) and denaturation enthalpies (ΔH_{cal}) for CBM9 on its own and as part of various CBM9-linker-IEGR-GFP fusion proteins.	91
Table 4.1	Michaelis Menten kinetic constants and reaction half-lives determined by the LRET-based assay.	120

Table 5.1	Measured Langmuir isotherm parameters for binding of CBM9-GFP and each of its fusion partners to Perloza™ MT100, Sephadex G15, and the composite MT100/G15 media at 4 °C. The solvent consisted of 50 mM potassium phosphate, 150 mM NaCl, pH 7. NB indicates that no binding was observed.....	154
Table 5.2	Yield, purity and concentration factor for the affinity purification of CBM9-GFP on a MT100/G15 composite column and on a pure MT100 column. Clarified <i>E. coli</i> cell lysate containing CBM9-GFP was loaded onto each column at a superficial velocity of $4.25 \times 10^{-3} \text{ cm s}^{-1}$. Fractions were collected and analyzed both by absorbance at 280 nm and by fluorescence intensity to obtain reported data. The total lysate volume loaded was 65 mL and contained a CBM9-GFP concentration of 5.7 μM	155
Table 5.3	Measured mass-transfer and column-geometry parameters for CBM9-GFP transport in a Perloza™ MT100/G15 composite media column.	156
Table 6.1	Cross-linked resins and reaction conditions	192
Table 6.2	Results of fitting quadratic response surface models to the data set reported in Table 6.1	193
Table 6.3	Equilibrium adsorption parameters for MT100 and CRL100-7 affinity media determined from two independent experiments.....	194
Table 6.4	Column properties and parameters	195
Table 6.5	Yield and purity for consecutive purification cycles of CBM9-GFP on a CRL100-7 column	196

List of Figures

- Figure 1.1 Schematic representation of the domain structure of *Cellulomonas fimi* cellulases and xylanases. Note: the domain structure of xylanase 10A of *T. maritima* is homologous to that of Xyn10A of *C. fimi* with the family 2a CBM9 substituted by CBM9. 24
- Figure 1.2 Backbone secondary structure of CBM9 from *T. maritima*, presented in ribbon diagram form, determined from the x-ray crystallographic data of Notenboom (Notenboom et al. 2001). Trp71 and Trp 175 (solid arrow) involved in ligand binding, and Glu77, Gln96, Arg98, Gln151, Arg161, and Asn172 (dashed arrow), involved in hydrogen bonding, with the ligand are shown. 25
- Figure 2.1 Schematic representation of gene fragment coding for the CBM9-S3N10-IEGR-GFP fusion protein 60
- Figure 2.2 Equilibrium adsorption isotherms for binding of CBM9 and CBM9-GFP to Perloza™ MT100 at 4°C. CBM9-GFP binding to Perloza™ MT100 at 4°C in high salt buffer (solid circle). CBM9 binding at 4°C in high-salt buffer (open circle), where q_i is the bound protein concentration and C_i is the equilibrium concentration of protein free in solution. 60
- Figure 2.3 Proteolytic stability of the S3N10 linker in CBM9-GFP. 12% SDS-PAGE of CBM9-GFP purified with Perloza™ MT100 in a small batch system. PMSF treated cell extract containing CBM9-GFP was mixed end-over-end, washed with buffer and desorbed with 1-M glucose in TBS8. Despite its molecular weight of 22 kDa, CBM9 runs as a 26-31 kDa protein depending on the length of the linker fragment attached to it (Wassenberg et al. 1997) 61
- Figure 2.4 Chromatogram of CBM9-GFP purification from an *E. coli* BL21 clarified cell lysate on Perloza™ MT100 at 4°C. 50 mL of clarified cell extract was loaded at 0.2 mL min⁻¹ on a 17 mL column packed with Perloza™ MT100 resin, and then washed with 10 column volumes (CV) high salt buffer and 5 CV low salt buffer. Bound fusion protein was desorbed with 1 M glucose in TBS8. 10 mL fractions were collected and analyze by fluorescence (509 nm) (solid circle) and absorbance at 280 nm (line).... 62
- Figure 2.5 SDS-PAGE documentation of the affinity purification of CBM9-GFP. 12% SDS-PAGE of CBM9-GFP purified on a 17 mL Perloza™ MT100 column. All samples dissolved in sample buffer containing 10% SDS. Lane M: molecular mass markers in kg mol⁻¹. Lane 1: clarified cell extract prior to column loading. Lane 2: column flow-through. Lane 3: high salt wash. Lane 4: low salt wash. Lane 5: pure CBM9-GFP eluted in TBS8

	containing 1-M glucose. Lane 6: purified GFP after affinity-tag removal by immobilized Factor X _a	63
Figure 2.6	Time course of CBM9-GFP cleavage by Factor X _a at 23°C as shown on a 12% SDS-PAGE. A fusion protein to Factor X _a concentration ratio of 1000:1 was used.	64
Figure 3.1	Schematic representation of the CBM9-Linker-FX _a -GFP fusion protein. FX _a indicates the recognition site (IEGR) for Factor X _a processing.	92
Figure 3.2	Intracellular Expression of CBM9-Linker-FX _a -GFP fusion proteins in <i>E. coli</i> . Cells expressing a CBM9-Linker-FX _a -GFP fusion protein were grown at 30°C to an OD ₆₂₀ of ~0.9 and protein expression was induced with 0.1 mM IPTG. CBM9-Linker-FX _a -GFP expression was continuously monitored at 510 nm with excitation at 400 nm.	92
Figure 3.3	Long-term proteolytic stability of the linkers in CBM9-Linker-FX _a -GFP fusion proteins. Clarified cell lysates were incubated at 23°C for 4 hours and then the intact and degraded CBM9-Linker-FX _a -GFP fusion proteins were purified on Perloza™ MT100 in the absence of protease inhibitors. Each lane represents proteins eluted off the Perloza™ MT100 resin with elution buffer containing 3 M glucose. Lane M contains the molecular weight markers.	93
Figure 3.4	Equilibrium adsorption isotherm for binding of CBM9-G ₃ -IEGR-GFP to Perloza™ MT100. Isotherm measured at 4°C in high salt buffer (pH7; 50 mM potassium phosphate, 1 M NaCl).	94
Figure 3.5	FX _a cleavage kinetics. CBM9-P-IEGR-GFP fusion protein was incubated at 23°C with Factor X _a at a fusion protein/FX _a concentration ratio of 1000:1. Samples were taken at each time point indicated and analyzed by SDS-PAGE.	95
Figure 3.6	Sensitized fluorescence emission lifetime data for CBM9-(PT) ₂ P-IEGR-GFP at 23°C. Solid line indicates fit to the first-order exponential decay equation ($r^2 = 0.999$).	96
Figure 3.7	Differential scanning calorimetry thermograms for CBM9, GFP, CBM9-P-IEGR-GFP and CBM9-(PT) ₇ P-IEGR-GFP in 50 mM potassium phosphate buffer (pH 7). Area under the denaturation peak(s) provides the denaturation enthalpy through the relationship $\Delta H_{cal} = \int C_p(T)dT$, where C_p is the heat capacity.	97
Figure 4.1	Schematic representation of the terbium bound fusion protein used for the LRET-based peptidase assay (not to scale). Excitation of bound Tb(III) at 235 nm results in the transfer of energy at 490 nm to GFP, leading to the	

	emission of a signal at 510 nm. D ₄ K indicates the recognition site for the serine peptidase, enterokinase.....	121
Figure 4.2	Individual spectrum of donor and acceptor. Solid line is the emission spectrum of CBM9-bound terbium (donor) obtained after excitation at 235 nm and a post-excitation delay of 200 μ s. Dashed line is the absorbance spectrum of GFP (acceptor). The spectral overlap (480 nm to 505 nm) between CBM9-bound terbium and GFP allows for energy transfer to take place.	122
Figure 4.3	Emission spectra of CBM9-bound terbium (donor domain only) (solid line) and Tb ³⁺ -CBM9-(PT) ₂ PID4K-GFP fusion protein (dashed line). Samples were excited at 235 nm followed by a 200 μ s delay prior to measurement of emission signal at each wavelength. Scans were performed at a rate of 1 nm/s.	123
Figure 4.4	Real-time LRET-based signal from EKmax TM mediated hydrolysis reaction. 32 μ M of Tb ³⁺ -CBM9-PID ₄ K-GFP fusion protein was incubated at 30°C with 9.3 nM EKmax TM (solid squares). The control reaction has buffer in place of enzyme (open squares). Sample was excited at 235 nm and emission was measured at 520 nm following a 200 μ s delay. Each measurement is an average of 10 excitations.....	124
Figure 4.5	Baseline corrected logarithmic decay in $\ln(\phi)$ due to EKmax TM mediated hydrolysis. Reaction conditions are the same as in Figure 4.4. The solid line represents the linear least-squares fit to the initial data.	125
Figure 4.6	Michaelis-Menten analysis. Initial rates of EKmax TM -catalyzed hydrolysis measured as a function of initial substrate concentration. Tb ³⁺ -CBM9-PID ₄ K-GFP was incubated with EKmax TM (9.3 nM) at 30°C....	126
Figure 4.7	Lineweaver-Burk representation of initial rates of EKmax TM -catalyzed hydrolysis of Tb ³⁺ -CBM9-PID ₄ K-GFP measured as a function of initial substrate concentration. Reaction conditions are the same as reported in Figure 4.4.	127
Figure 4.8	SDS-PAGE documentation of EKmax TM -catalyzed hydrolysis of Tb ³⁺ -CBM9-PID ₄ K-GFP. Reaction conditions are the same as reported in Figure 4.4.	128
Figure 5.1	Scanning electron micrographs of Perloza TM MT100 beaded media. Figures A, B and C show an MT100 particle at 1.0k, 13.0k and 80.0k magnification, respectively. The pore structure shown at the center of Figure B is further magnified and shown in Figure C.	157

- Figure 5.2 Equilibrium adsorption isotherm for batch binding of CBM9-GFP to Perloza™ MT100/G15 composite media at 4°C. The solvent consisted of 50 mM potassium phosphate, 150 mM NaCl, pH 7. The solid curve represents the best fit of the experimental data to the Langmuir isotherm equation..... 158
- Figure 5.3 Chromatogram for CBM9-GFP purification on Perloza™ MT100/G15 composite media. Clarified *E. coli* cell lysate containing CBM9-GFP was loaded at a superficial velocity of $4.25 \times 10^{-3} \text{ cm s}^{-1}$ onto a 10 mL column. Fractions were collected and analyzed both by absorbance at 280 nm (line) and by fluorescence intensity (solid diamond) as shown..... 159
- Figure 5.4 SDS-PAGE documentation of CBM9-GFP purification on a 10 mL Perloza™ MT100/G15 column. All samples were dissolved in sample buffer containing 10% SDS. Lane M: molecular mass markers; Lane 1: clarified cell lysate prior to column loading; Lane 2: column flow through; Lane 3: high salt wash; Lane 4: low salt wash; Lane 5: CBM9-GFP eluted in low salt buffer containing 1 M glucose. 160
- Figure 5.5 First moment (μ_1) analysis for a 10 mL column packed with Perloza™ MT100/G15. Integrated μ_1 values are reported for pulse injections of a 50 μL solution of blue dextran (MW 2000 kDa) over the interstitial velocity range $4.25 \times 10^{-3} \text{ cm s}^{-1}$ to $2.02 \times 10^{-2} \text{ cm s}^{-1}$. The mobile phase consisted of 50 mM phosphate buffer, 150 mM NaCl (pH 7, 4°C)..... 161
- Figure 5.6 Measured average porosity ε_p of composite media as a function of molecular weight of standard proteins. Pulse injections of standard molecular weight protein markers were used over the interstitial velocity range $2.1 \times 10^{-3} \text{ cm s}^{-1}$ to $8.5 \times 10^{-3} \text{ cm s}^{-1}$. The porosity was determined from equation 5.15 using the measured void fraction of 0.425. 162
- Figure 5.7 Determination of axial dispersion coefficient D_L and overall solute mass-transfer coefficient K_M for injection of CBM-GFP under nonbinding conditions onto a 10 mL column packed with Perloza™ MT100/G15. Pulse injections of a 50 μL solution of CBM9-GFP was used over the interstitial velocity range $4.25 \times 10^{-3} \text{ cm s}^{-1}$ to $1.38 \times 10^{-2} \text{ cm s}^{-1}$. The mobile phase consisted of 50 mM potassium phosphate, 150 mM NaCl (pH 7, 4°C) with 2 M glucose added to achieve nonbinding conditions. 163
- Figure 5.8 Time course fluorescent intensity profile of CBM9-GFP uptake into Perloza™ MT100 particle. Protein uptake monitored by an inverted Zeiss LSM 510 confocal laser scanning microscope with the center of the particle used as the focal plane. Excitation and emission wavelengths were 488 nm and 505 nm, respectively. The initial CBM9-GFP concentration C_i^0 outside the particle was 5.4 μM 164

- Figure 5.9 Time dependent radial profiles of CBM9-GFP uptake into a Perloza™ MT100 particle for a feed concentration of 5.4 μM . Predicted uptake rates using (Figure 9A) pore-diffusion model and (Figure 9B) two-zone model are compared with experiment. The rapid drop in fluorescence intensity at radial positions above *ca.* 39 μm indicates the position of the outer radius of the bead..... 165
- Figure 5.10 Comparison of TZM model predictions of $r_c(t)$ with values computed from CLSM data. CLSM determined core radius reported as the radius at which the measured fluorescence intensity falls below 3X the standard deviation of the background fluorescence. Load conditions same as stated in Figure 5.9..... 166
- Figure 5.11 Time dependent radial profiles of CBM9-GFP uptake into a Perloza™ MT100 particle for a feed concentration of 49 μM . Predicted uptake rates using (Figure 11A) pore-diffusion model and (Figure 11B) two-zone model are compared with experiment..... 167
- Figure 5.12 Comparison of TZM (solid curve) and PDM (dashed curve) predictions with experimental (points) breakthrough curves. Pure CBM9-GFP loaded at a superficial velocity of $8.5 \times 10^{-3} \text{ cm s}^{-1}$ onto a Perloza™ MT100/G15 composite media column: (A) frontal load of $3.4 \times 10^{-3} \text{ mol m}^{-3}$ CBM9-GFP, (B) frontal load of $1.85 \times 10^{-2} \text{ mol m}^{-3}$ CBM9-GFP..... 168
- Figure 5.13 TZM predicted (line) and experimental (points) breakthrough curves for pure CBM9-GFP loaded onto a Perloza™ MT100/G15 composite media column at three different feed concentrations: $C_i^o = 4.77 \times 10^{-2} \text{ mol m}^{-3}$ (triangles), $2.2 \times 10^{-2} \text{ mol m}^{-3}$ (squares), and $8.0 \times 10^{-3} \text{ mol m}^{-3}$ (circles). Mobile phase loaded at a flow rate of 0.4 mL min^{-1} 169
- Figure 5.14 TZM predicted (line) and experimental (points) breakthrough curves as a function of interstitial velocity. Pure CBM9-GFP loaded onto a Perloza™ MT100/G15 composite media column: $u = 1.7 \times 10^{-2} \text{ cm s}^{-1}$ (triangles), $8.5 \times 10^{-3} \text{ cm s}^{-1}$ (circles), and $4.2 \times 10^{-3} \text{ cm s}^{-1}$ (squares). 170
- Figure 5.15 Predicted (line) and experimental (points) breakthrough curves for frontal loading of a clarified cell extract onto a Perloza™ MT100/G15 composite media column. The clarified cell extract contained 25 μM CBM9-GFP and was loaded at a superficial velocity of $8.5 \times 10^{-3} \text{ cm s}^{-1}$. Eluent absorbance data at 280 nm (open squares) are also shown to indicate total protein as a function of time. 171
- Figure 6.1 Picture of an MT100 column pre- and post-compression. Picture shows bed compression at a superficial velocity of $1.06 \times 10^{-4} \text{ m/s}$ 197

Figure 6.2	Schematic representation of the proposed epoxide-based reaction for cross-linking cellulose with epichlorohydrin.....	198
Figure 6.3	Imageplot (A) and wireframe surface (B) for a locally weighted polynomial regression fit to the measured u_{crit} (in $m\ s^{-1} \times 10^4$) data set reported in Table 1. Created with the loess, levelplot, and wireframe functions in R (Young et al. 1980) and the add-on package Lattice (Lattice Graphics R package, version 0.14-16), with a span of 0.65 and a degree of 2 (i.e. local fits are quadratic). The location of the maximum is indicated by (\otimes).....	199
Figure 6.4	Wireframe surface of the results obtained from the standard quadratic response surface model (6.2) when applied to the measured u_{crit} (in $m\ s^{-1} \times 10^4$) data set reported in Table 1.	200
Figure 6.5	Pressure curve for solvent flow through an MT100 (open circles) or CRL100-7 (filled circles) column monitored over a range of superficial velocities. The pressure drop curve for CRL100-7 was compiled from two independent experiments.	201
Figure 6.6	Equilibrium adsorption isotherms for binding of CBM9-GFP to MT100 (filled circles) and CRL100-7 (open squares) at 21°C. Control experiment shows ovalbumin binding to CRL100-7 (filled squares) at 21°C. The mobile phase buffer consisted of 50 mM phosphate buffer, 100 mM NaCl, pH 7. The curve represents the best fit of the experimental data to the Langmuir adsorption isotherm equation where q_i is the protein concentration bound to the media surface and c_i is the equilibrium concentration of protein free in solution.....	202
Figure 6.7	Measured first central moment (μ_1) for blue dextran (filled squares) and CBM9-GFP (filled circles) on a column (I.D. 1.0 cm) packed with CRL100-7.	203
Figure 6.8	Second moment analysis under non-binding conditions for a column (I.D. 1.0 cm) packed with CRL100-7 media. Pulse injections of CBM9-GFP were used over the superficial velocity range 4.25×10^{-5} m/s to 4.25×10^{-4} m/s. The mobile phase consisted of 2 M glucose in 50 mM potassium phosphate, 100 mM NaCl (pH 7, 21°C).....	204
Figure 6.9	Van Deemter plot describing CBM9-GFP transport within the CRL100-7 column (I.D. 1.0 cm X 11.4 cm); the reduced plate height $h = \left(\sigma^2 L \right) / \left(\mu_1^2 d_p \right)$ is plotted as a function of the reduced linear velocity, given by $(ud_p)/D_M$	205

Figure 6.10 Chromatogram for purification of CBM9-GFP on a 60-mL CRL100-7 column at 21°C. Clarified cell extract from *E. coli* BL21 were loaded onto a CRL100-7 column (2.6 cm I.D. X 11.8 cm) at a superficial velocity of 1.5×10^{-2} m/s, washed with *ca* 3 column volumes (CV) of high salt buffer, *ca* 2.5 CV of low salt buffer and bound CBM9-GFP was eluted with 1M glucose in low salt buffer. 10 ml fractions were collected and analyzed by both absorbance at 280 nm (solid line) and fluorescence emission at 509 nm (filled circles)..... 206

Figure 6.11 12% SDS-PAGE documentation for preparative-scale affinity purification of CBM9-GFP on a CRL100-7 column (I.D. 2.6 cm). All samples were dissolved in sample buffer containing 10% SDS. Lane M: molecular weight markers in kg/mole; Lane 1: clarified cell extract prior to column loading; Lane 2: column flow through; Lane 3: high salt wash; Lane 4: low salt wash; Lane 5: pure CBM9-GFP eluted in low salt buffer containing 1 M glucose..... 207

Nomenclature

A_s	peak asymmetry factor	
c_i	total concentration of free solute i in the pore liquid	(mol m ⁻³)
c_i^*	equilibrium concentration of free solute i in the pore liquid	(mol m ⁻³)
C_i	total concentration of solute i in the interstitial mobile phase liquid	(mol m ⁻³)
C_i^o	total concentration of solute i in the interstitial liquid at time t=0	(mol m ⁻³)
C_i^{feed}	total concentration of solute i in the injected sample	(mol m ⁻³)
C_P	heat capacity	(kcal°C mol ⁻¹)
d_c	column diameter	(cm)
d_p	particle diameter	(μm)
D_L	axial dispersion coefficient	(m ² s ⁻¹)
D_M	molecular coefficient in bulk liquid	(m ² s ⁻¹)
D_p	intraparticle diffusivity	(m ² s ⁻¹)
D_s	surface diffusivity	(m ² s ⁻¹)
E	efficiency of energy transfer	
E_T	total enzyme concentration	(nM)
f_D	corrected fluorescence intensity of donor	(arbitrary units)
$HETP$	height equivalent to a theoretical plate	(m)
ΔH_{cal}	calorimetric enthalpy of denaturation	(J mol ⁻¹)
I	fluorescence emission intensity at specific wavelength	(arbitrary units)

J	spectral overlap between donor fluorescence and acceptor absorbance	$(\text{nm}^4 \text{M}^{-1} \text{cm}^{-1})$
k	aspect factor	
k_{ads}	sorption rate constant	$(\text{m}^3 \text{mol}^{-1} \text{s}^{-1})$
k_{ai}	Langmuir binding constant	(M^{-1})
k_f	fluid film mass transfer coefficient	(m s^{-1})
k_{cat}	catalytic constant	(s^{-1})
k_{obs}	observed rate constant	(s^{-1})
K_M	Michaelis Menten kinetic constant	(μM)
κ_M	overall solute mass transfer coefficient	(s^{-1})
L	column length	(m)
M	protein molecular weight	(g mol^{-1})
n	refractive index of medium	
NTU	number of theoretical units	
P	pressure	(Pa)
q_i^*	equilibrium concentration of bound solute I	$(\mu\text{mol g}^{-1} \text{resin})$
q_i	concentration of bound solute i in pore liquid	(mol m^{-3})
q_i^{max}	maximum capacity of sorbent to bind solute i	$(\mu\text{mol g}^{-1} \text{resin})$
q_i^{sat}	sorbent saturation capacity (ρq_i^{max})	(mol m^{-3})
Q_D	quantum yield of donor chromophore	
r_c	core radius	(m)

r_p	particle radius	(m)
R	average Forster distance	(Å)
R_o	Forster distance at E = 50%	(Å)
R_M	overall resistance to solute mass transfer	(s)
\bar{s}_i	average solute concentration within sorbent particle	(mol m ⁻³)
s_i	solute concentration within sorbent particle	(mol m ⁻³)
$[s]$	substrate concentration	(μM)
$[s]_o$	initial substrate concentration	(μM)
t	time	(s)
$t_{1/2}$	half-life of substrate	(h)
T	temperature	(°C)
T_m	melting temperature	(K)
u	superficial fluid velocity	(m s ⁻¹)
u_o	interstitial fluid velocity	(m s ⁻¹)
u_{crit}	critical superficial velocity for column compression	(m s ⁻¹)
v	rate of enzyme cleavage	(μM h ⁻¹)
v_o	initial reaction rate	(μM h ⁻¹)
v_{max}	maximum reaction rate	(μM h ⁻¹)
z	axial coordinate	(m)

Dimensionless numbers

- Bi* Biot Number ($k_f d_p / 6D_p$)
- Da* Damkohler Number ($C_i^0 k_{ads} d_p^2 / 4D_p$)
- h* reduced plate height ($\sigma^2 L / (\mu_1^2 d_p)$)
- Re* Reynolds number ($d_p u \rho / \mu$)
- Sc* Schmidt number ($\mu / \rho D_M$)
- Sh* Sherwood number ($k_f d_p / D_M$)
- Pe* Peclet Number ($u d_p / D_p$)
- u_{red}* Reduced Linear Velocity ($u d_p / D_M$)

Greek letters

β_0	regression coefficient for total error in fit of equation	(m s ⁻¹)
β_1	linear regression coefficient for epichlorohydrin concentration	(g m ⁻² s ⁻¹)
β_2	linear regression coefficient for DMSO concentration	(g m ⁻² s ⁻¹)
β_{11}	quadratic regression coefficient for epichlorohydrin concentration	(g ² m ⁻⁵ s ⁻¹)
β_{12}	regression coefficient for linear interaction between ξ_1 and ξ_2	(g ² m ⁻⁵ s ⁻¹)
α, γ, δ	fitted parameters (Equation 5.17)	
ξ_1	input variable for epichlorohydrin concentration	(ml g ⁻¹)
ξ_2	input variable for DMSO concentration	(ml g ⁻¹)
Ψ	random variable accounting for total error in measurement	(m s ⁻¹)
ε	interstitial void fraction of column	
ε_a	extinction coefficient	(M ⁻¹ cm ⁻¹)
ε_p	average stationary phase porosity	
κ^2	orientation factor	
η	fluid viscosity	(g.m ⁻¹ s ⁻¹)
ϕ	fraction of substrate remaining	
ρ	fluid density	(g m ⁻³)
σ^2	second central moment of an eluent peak	(s ²)
τ_{da}	lifetime of donor signal in presence of acceptor chromophore	(ms)
τ_d	lifetime of donor signal in absence of acceptor chromophore	(ms)

τ_{Tb}	lifetime of free terbium (III) signal	(ms)
μ	fluid viscosity	(g m ⁻¹ s ⁻¹)
μ_1	first central moment of an eluent peak	(s)

Acknowledgements

First and foremost, I would like to thank my research supervisor, Dr. Charles Haynes for his invaluable guidance, for allowing me the opportunity to gain experience in a variety of research areas and for teaching me to see the forest beyond the trees.

I would also like to thank Drs. Doug Kilburn and Tony Warren for giving me the opportunity to become a member of their group. Their kindness and guidance will always be remembered and appreciated.

I'm grateful to Drs. Steve Withers and Chris Overall for allowing me use of their resources. Without their generosity, certain aspects of my research would have been harder to fulfill. A special thanks goes to Dr. Richard Fellmen for his advice on enzyme kinetics.

I would like to thank all the members of the Michael Smith Laboratories and especially my colleagues in both the Haynes' Lab and the Cellulase Group for providing a fun and friendly environment to work in. To Dr. Louise Creagh, thank you for all your advice.

Last but not least, I would like to thank my parents for their love and support. This thesis is dedicated to them.

To my parents, Houshang and Parizad Kavooosi

Co-Authorship Statement

This work was done under the guidance of Dr. Charles A. Haynes and Dr. Douglas G. Kilburn. Dr. Louise A. Creagh helped with the analysis of the differential scanning calorimetry results reported in Chapter 3 and performed the isothermal titration calorimetry experiments reported in Chapter 4. Dr. Jenny Bryan assisted with the statistical analysis reported in Chapter 6. Nooshafarin Sanaie wrote the computer programs used to predict the breakthrough curves reported in Chapter 5. Florian Dimer from Dr. Jürgen Hubbuch's group did the confocal laser scanning microscopy experiments reported in Chapter 5. Emily Kwan provided the immobilized CBM2a-Factor X_a fusion protein used for experiments reported in Chapter 2. Julia Meijer and Dexter Lam were undergraduate students under my supervision. I performed all other experiments and data analysis reported in this thesis. I am also the principle author on all the publications reported in this thesis and had primary responsibility for both their conceptual and practical aspects.

1 Introduction, Background and Thesis Objectives

1.1 Overview

In 2004, the global biotechnology industry reported annual revenues of \$54.6 billion and accounted for over 20% of all venture capital investments. This included 365 products in the pipeline from Phase I to Phase III clinical trials and 55 new drug applications submitted for review by the Food and Drug Administration (FDA) (Source: Ernst and Young). On average, the cost of product development is now estimated at \$800 million per new drug. This rapidly rising cost and the increasing number of new protein therapeutics reaching the market are placing significant pressure on a global healthcare industry already challenged, particularly in North America, by an increasingly aging population. Sustainable growth of the biotechnology industry is therefore dependent on discovery of new ways to increase efficiency and productivity in order to keep both development and manufacturing costs in check.

The process of manufacturing a biologic involves a number of unit operations, including (i) cell-line development and clone selection, (ii) fermentation/cell culture, (iii) cell harvest/removal, (iv) cell disruption (if needed), (v) product concentration, (vi) product purification, and (vii) final product formulation. Each of the final six processing steps generally involves one or more unit operations. Due to technological advancements over the past few decades, the cost of producing a biologic has decreased considerably. Refinements in expression system design and in fed-batch and perfusion-culture protocols now enable the production of the target biomolecule at g/L quantities at relatively little cost (\$100-\$500/g total cost of goods). As a result, the overall economics of therapeutic protein production now tends to be dominated by the cost of product purification (Lowe et al. 2001).

New strategies to purify recombinant proteins from complex high-density cultures must therefore be developed with the aim of maximizing yield and purity while minimizing costs. A key objective is to streamline downstream operations into a smaller

number of unit operations, with each step employing a separation strategy highly selective for the target protein. The bioselective adsorption process known as affinity chromatography is one such unit operation. Affinity chromatography takes advantage of highly selective interactions found in nature, such as the tight and specific binding ($K_a \sim 10^6$ to 10^9 M^{-1}) between an antibody and its antigen. For example immobilized IgG can be used to purify protein A from *Staphylococcus aureus* (Moks et al. 1987; Yarnall and Boyle 1986). Conversely, protein A covalently coupled to activated-Sepharose media is now widely used to remove and isolate immunoglobulins from serum of different species (Chen et al. 2006; Jungbauer and Hahn 2004) and in the purification of monoclonal antibodies (mAbs), the fastest growing class of recombinant protein therapeutics (Follman and Fahrner 2004; Hahn et al. 2006).

Advances in recombinant DNA technology have allowed the power of high-affinity interactions to be generically applied to the concentration and purification of recombinant proteins. Recombinant hybrids containing an N- or C-terminally fused affinity polypeptide tag that selectively binds to a complementary ligand immobilized onto a suitable chromatographic matrix are now widely used to facilitate the purification of target proteins or peptides. Numerous affinity tags have been developed over the years (Hearn and Acosta 2001) with many of them successfully commercialized for laboratory scale applications (Nilsson et al. 1997; Terpe 2003). However, their application at the preparative scale is sparse, due in large part to the high cost of the associated affinity media (Ladisich 2001). This high cost can arise from the complex chemical modifications needed to immobilize the ligand to the resin surface. In addition, immobilized ligands can experience fouling and structural damage, particularly at the ligand-resin junction, resulting in a reduced column capacity and a short column lifetime. This latter sensitivity often leads to a low tolerance of many affinity media to repeated processing and sanitization cycles, which limits their use in production-scale purifications.

Thus, although affinity chromatography has the potential to simplify downstream processing by achieving higher yields and purities than conventional methods, it is not without problems. This doctoral thesis aims to alleviate ligand stability and cost limitations associated with conventional affinity chromatography by introducing an

affinity tag that binds to a very inexpensive and stable cellulose-based matrix. The research focuses on the development of a generic and inexpensive affinity purification technology based on high-level recombinant expression in *E. coli* of fusion proteins containing the family 9 carbohydrate-binding module (CBM9) attached to either the N- or C-terminus of the target protein or polypeptide. A linker sequence containing a specific processing site is designed to facilitate rapid and quantitative removal of the tag, and to thereby recover the desired target sequence following affinity purification. The development, modeling and commercialization of this technology are the central objectives of this thesis.

1.2 Literature Review

1.2.1 Downstream Processing in Biotechnology

Recombinant DNA technology has provided effective protein therapeutics for many of our deadliest and most debilitating diseases, such as various forms of cancer, asthma, arthritis, heart disease, diabetes and AIDS, to name just a few. Yet despite the great benefit of recombinant proteins to human health, the \$800 million dollar price tag to bring each drug to market (Source: Ernst and Young) is serving to limit the technology largely to the production of blockbuster therapeutics - *i.e.* protein therapeutics expected to achieve greater than \$100 million in sales per annum. Many important but less profitable indications are therefore receiving little attention. In order to overcome this ethical problem and to meet the challenges created by rising healthcare costs, industry must focus on reducing development and manufacturing costs of novel and generic biopharmaceutical proteins. Lowering the number of processing steps and increasing the yield at each step would serve this objective (Ladisich 2001). Many manufacturers are therefore looking towards redesigning and improving the efficiency and yields of existing production processes to lower costs (Narayanan 1994; Thomson 1996).

Manufacturing of biopharmaceuticals must comply with quality assurance and cGMP (current good manufacturing practice). The FDA and other regulatory agencies such as the EMEA (European Medicines Agency) require that all finished biologics be manufactured under full cGMP compliance. This means that the final therapeutic

product must be a well-characterized biologic with defined purity, efficacy, potency, stability, pharmacokinetics, pharmacodynamics, toxicity and immunogenicity. The product must also be exhaustively tested for contaminants such as nucleic acids, viruses, pyrogens, residual host cell proteins, cell culture media, and leachates from the separation media. In addition, many therapeutic proteins are naturally and recombinantly produced in different isoforms originating from post-translational modifications such as glycosylation, sulphation, oxidation, end termination modifications, misfolding, aggregation, misaligned disulphide bridges and nicked or truncated variants. The process by which a therapeutic protein is produced is known to influence all of these product quality characteristics. Thus, the FDA views the process as the product, and great care must be taken in the selection and operation of the purification process.

1.2.2 Affinity Chromatography in Preparative-Scale Protein Purification

Regulatory compliance, process economics and strict quality assurance have instigated a rethinking of the design and operation of purification processes, with effectiveness, robustness and economics largely determining the choice of the purification strategy. Traditional, multi-step purification protocols based on non-specific physio-chemical properties such as pH, temperature, and solubility, are being replaced with highly selective and more sophisticated strategies, including those based on affinity chromatography (Bonnerjea et al. 1986). Affinity chromatography exploits natural biological binding processes for the selective separation of the target protein. Affinity separation techniques have the power to reduce the total number of processing steps, thereby increasing yields and downsizing capital equipment leading to improved process economics. However, these techniques do suffer from problems, particularly with respect to regulatory and other issues discussed below (Stevenson 1997).

Most affinity ligands used in research and industry originate from natural sources. The library of affinity ligands commonly used today includes immobilized sugars and polysaccharides, nucleic acids, dyes, chelating agents, proteins, protein fragments, monoclonal antibodies, peptides and amino acids (Hearn and Acosta 2001; Hopp 1988). While these ligands often show good selectivity, their use in affinity chromatography can

be challenged by a number of factors. For example, they must be produced and purified, may be contaminated with host DNA and viruses, tend to be fragile, show lot-to-lot variation and are often costly to produce. In addition, production of therapeutics requires stringent column sterilization typically involving clean-in-place protocols that utilize strong caustic solutions which can degrade the immobilized ligand, shortening column life and contaminating the end product with potentially toxic or immunogenic leachates. Biological ligands can also suffer from low binding capacities, limited life cycles and low scale-up potential.

1.2.3 “Natural” Affinity Chromatography

In natural affinity chromatography the binding motif is endogeneous to the target protein, allowing for direct and specific purification through selective binding to its cognate ligand immobilized onto an appropriate chromatography media. Examples of natural affinity chromatography are abundant in the literature (Jones 1990). Many glycoproteins are currently purified using immobilized lectins (Cummings 1997; Yang and Hancock 2004), while conversely many lectins have been efficiently isolated by specific adsorption on immobilized carbohydrates (Chen and Billingsley 1999; Lee et al. 1995). Other examples include the industrial purification of human plasminogen from blood plasma using immobilized lysine (Deutsch and Mertz 1970), the purification of ATP-dependent kinases and NAD⁺-dependent dehydrogenases using immobilized 5'-AMP (Mulcahy et al. 2002), and the use of the sulfated polysaccharide heparin to affinity purify coagulation factors (Levine 1976).

Aromatic compounds, particularly natural and synthetic textile dyes, have also been used as ligands in natural affinity chromatography. Cibacron blue is the most popular of these dyes and has been widely used in the research setting to purify albumin and other blood proteins, including oxido-reductases, carboxylases, glycolytic enzymes, nucleases, hydrolases, lyases, synthetases and transferases (Clonis et al. 2000; Denizli and Piskin 2001). The application of synthetic aromatics and dyes to the purification of therapeutic proteins has however been limited by concerns over the purity, selectivity, leakage and toxicity of these capture agents. Finally, for immunogenic proteins that do

not possess a known binding motif, protein-specific chromatography can be carried out using an immobilized monoclonal antibody (Ostermann 1990; Wells et al. 1993). However, the exceptionally high cost of monoclonal antibody production effectively limits the large-scale application of the technology to purification of rare and very high value proteins.

1.2.4 Affinity Fusion-Tag Technology

Affinity fusion-tag technology takes advantage of an existing and well-characterized interaction between a unique and infrequent binding motif and its corresponding ligand to purify a target protein. Recombinant DNA technology is used to genetically fuse the binding motif (henceforth called the affinity tag) to either the N- or C-terminus of a target protein. The recombinant fusion protein can then be expressed and purified using the selective interaction between the tag and the complementary ligand immobilized onto a suitable chromatographic support. Affinity tag chromatography therefore allows for good selectivity and greater predictability in the separation process, and the general ability to be applied to a range of target proteins without the need for developing a new protocol for each new protein of interest (Uhlen and Moks 1990).

Affinity tags range in complexity and molecular size from short, unstructured peptides to large binding domains. In addition to their application in the detection and purification of the protein of interest, affinity tags can offer a number of advantages. In cases where the target protein is conformationally unstable or requires complex disulfide bond formation, some affinity tags can stabilize the target protein, allowing it to fold more efficiently into the correct “near native” conformation *in vivo* by functioning as pseudo-chaperones or secondary structure nucleation centers (Bardwell et al. 1991; Hammarberg et al. 1989; Murby et al. 1991; Nygren et al. 1994; Tucker and Grisshammer 1996). Others can act as reporters, allowing for easy detection and quantification of the fusion protein using high-sensitivity detection assays such as an enzyme-linked immunosorbent assay (ELISA). Positioning the tag at the C-terminus ensures the production and recovery of a fully intact fusion protein. Truncated variants brought about by errors in translation will not have the affinity tag at the C-terminus and

therefore will not bind to the column containing the corresponding immobilized ligand (Jones et al. 1995). Some fusion tags have also been shown to extend the half-life and increase the pharmacokinetic activity of therapeutic proteins. For example, CD4 fused to the Fc region of IgG has an activity 200X greater than native CD4 (Capon et al. 1989). In addition, fusion tags, due to their highly selective interaction, can be used to purify low abundance proteins present in a complex feedstock (Brewer et al. 1991). Finally, certain affinity tags are an attractive tool for structure-function studies, functioning as specific reporters for use in systems such as *in vitro* two-hybrid protein-protein binding systems (Hoffmann and Roeder 1991; Jonasson et al. 1996; Lu et al. 1993; Nieba et al. 1997; Nilsson et al. 1997). Therefore, additional benefits may be realized through selection of an appropriate affinity tag based on the characteristics and intended application of the target protein.

1.2.5 Factors to Consider in Designing the Affinity Tag System

1.2.5.1 The Choice of Peptide Versus Protein Tags

Affinity tags can be grouped into either peptide tags or protein tags. Tags less than 25 amino acids in size are typically classified as peptide tags whereas small globular modules with unique binding properties are considered protein tags. Each class of affinity tag offers unique properties that must be considered during the selection process. For example, peptide affinity tags, due to their small size, are less likely to interfere with the structure or alter the catalytic activity of the target protein, and hence, their removal may not be necessary after purification (Tucker and Grisshammer 1996). For cases where tag removal is required, their simple structure generally permits enhanced accessibility of the processing enzyme to any introduced cleavage site (Jonasson et al. 1996).

On the other hand, protein affinity tags have been known to help enhance structural stability and solubility by promoting the correct three-dimensional folding of the target protein. For example, both insulin-like growth factor II (IGF-II) (Hammarberg et al. 1989) and T-cell receptor (Murby et al. 1991) have been shown to be more stable when expressed as a fusion protein. Protein tags such as protein A (Abrahmsen et al.

1986) and maltose-binding protein (MBP) (di Guan et al. 1988) not only stabilize the protein, but are generally more compatible with leader peptide insertion to direct the fusion protein to the periplasm where the oxidizing environment can promote disulfide bond formation. In addition, protein tags can facilitate protein folding by acting as covalently linked pseudochaperones (LaVallie 1995) or “protein enhancers” (Nygren et al. 1994; Stahl et al. 1997a; Stahl et al. 1997b). For example, the ZZ domain protein derived from *Staphylococcal* protein A has been shown to dramatically improve the correct folding of recombinant IGF-I (Samuelsson et al. 1996; Samuelsson et al. 1994).

1.2.5.2 Position of the Affinity Tag

The positioning of an affinity tag can influence the expression of the resulting fusion protein. For example, a hexahistidine tag is seldom placed at the N-terminus of proteins destined for cellular export since the multiple positively charged imidazolyl side chain groups can interfere with signal sequence processing, trafficking and export biorecognition, resulting in arrest of protein secretion or incorrect cleavage of the signal peptide (Seidler 1994). In general, N-terminal tags offer the advantage of yielding a target protein with its authentic N-terminal sequence, provided an appropriate specific protease recognition site is inserted between the tag and the target protein (Jones et al. 1995). Disadvantages of N-terminal tagging include the potential to copurify truncated products with the full-length protein, as well as the aforementioned potential for the tag to interfere with translation initiation. A C-terminal tag, on the other hand, does select only full-length protein, but generally adds unwanted amino acids to the end of the target protein upon protease cleavage (Jones et al. 1995).

Finally, in certain circumstances, regardless of their position, fusion tags have been found to be buried within the internal structure of the folded protein (Kronina et al. 1999; LaVallie 1995; Lindner et al. 1992; Sharrocks 1994; Uhlen and Moks 1990) or to partially cover a binding/catalytic site on the target protein (Casey et al. 1995). For example, a hexahistidine tag placed at the C-terminus of a tumor-associated single chain Fv sequence was found to partially cover the antigen binding site of the protein (Goel et al. 2000). Therefore, care must be taken in tag selection and positioning to ensure that it

functions independently and does not affect the activity or cause any perturbation in the conformation of the target protein (Hearn and Anspach 1990; Hearn and Gomme 2000).

1.2.5.3 Expression of Fusion Proteins in Host Cells

A large number of organisms have now been specifically engineered to express recombinant proteins (Makrides 1996). They include *E. coli* and other prokaryotic cells (Ford et al. 1991; Nilsson et al. 1997), *Saccharomyces cerevisiae* (Hopp 1988; Shimabuku et al. 1991), *Pichia pastoris* (Andrade et al. 2000; Higgins and Cregg 1998; Rajamohan et al. 2000) and other yeast strains, mammalian cells (Chubet and Brizzard 1996; Lo et al. 1998; Witzgall et al. 1994), insect cells based on the baculovirus expression system (Allet et al. 1997; Kuusinen et al. 1995; Yet et al. 1995), egg sack yolk (Fassina et al. 1998; Hansen et al. 1998), and even transgenic plants and animals (Hennig and Schafer 1998; Hermanson and Turchi 2000; Kleiner et al. 1999). Higher eukaryotic organisms possess the machinery to perform post-translational modifications and as such, are often used to express proteins that require complex modifications including O- or N-linked glycosylations, phosphorylation/methylation and complex disulfide bond formation (Marino 1989). These systems can however, suffer from low expression of recombinant proteins (Ford et al. 1991), hyperglycosylation or alternative sites of glycosylation (Higgins and Cregg 1998). Bacterial expression systems can provide higher expression levels but often at the expense of increased protein loss during the subsequent refolding (if necessary) and purification steps (Brewer et al. 1991). New to the field of recombinant protein expression, the egg yolks and the milk of lactating transgenic animals are gaining much interest for offering a fast, simple and inexpensive method for manufacturing biologics.

Recombinant proteins can sometimes be expressed in the form of inclusion bodies (Kane and Hartley 1988; Krueger et al. 1989). Within these inclusion bodies, proteins exist in a biologically inactive, partially unfolded and reduced state. The recovery protocol needed to refold the protein back to its native state is unique for each individual protein and must be determined experimentally. This adds to the cost and complexity of the bioprocess since each protocol must be optimized with respect to temperature, pH and

buffer composition (Flaschel and Friehs 1993; Rudolph 1996). However, with a robust renaturation protocol predetermined for a given protein, incorporating inclusion body formation as part of a bioprocess offers the advantage of ease of recovery from the cell homogenates (Thatcher 1990) and protection from cellular proteolysis (Sassenfeld 1990).

For problematic cases, inclusion body formation can be curbed by directing the fusion protein into an oxidative environment such as the periplasm of *E. coli* or through the rough endoplasmic reticulum and golgi apparatus and into the culture medium in eukaryotic cells (Abrahmsen et al. 1986; Carter et al. 1992). Trafficking the fusion protein to a specific compartment of the cell or secretion into the culture supernatant offers a number of benefits, including (i) limited exposure to cytoplasmic proteases, which could degrade the target protein (Murby et al. 1991), and (ii) a simplified purification process by secreting the fusion protein into an environment with less contamination from other cellular components (Hansson et al. 1995; Hockney 1994).

1.2.5.4 Tag Removal Post-Purification

Recombinant proteins destined for pharmaceutical/therapeutic application typically must have their tag removed to minimize adverse side effects, including immunogenic responses, and to meet regulatory standards (Pedersen et al. 1999). This is achieved by introducing, at the cloning stage, a site-specific cleavage site between the tag and the target protein. Once the fusion protein is purified, cleavage at this specific processing site can be accomplished by either enzymatic or chemical means (Brewer et al. 1991; Casey et al. 1995). Chemical cleavage is most often achieved with cyanogen bromide (Itakura et al. 1977) which cleaves only after methionine residues. Although inexpensive, it is toxic and the harsh reaction conditions can cause denaturation and degradation of the target protein. Cleavage using site specific endopeptidases such as enterokinase (Hosfield and Lu 1999; Prickett et al. 1989), thrombin (Hakes and Dixon 1992), Factor X_a (Mcperson et al. 1992), and the two relatively new processing enzymes, tobacco etch virus (TEV) (Shih et al. 2005) and human rhinovirus 3C protease (HRV 3C) (Libby et al. 1988), offers a gentler alternative. Although enzymatic methods are safer and more specific, they are generally less efficient since their rate of cleavage is

affected by the accessibility of the cleavage site and the chemistry of the adjacent amino acids, particularly those lying in the linker sequence immediately upstream of the recognition site. Some endopeptidases such as thrombin, trypsin and HRV 3C also leave behind extra amino acids on the termini of the target protein upon cleavage (Wyborski et al. 1999).

Self-cleavable systems have attracted much interest in recent years. Intein-based systems are based on natural protein splicing sequences fused to affinity tags such as the chitin-binding domain (Derbyshire et al. 1997; Wood et al. 1999). Cleavage of the intein can be achieved by addition of thiols, shifting pH, or temperature. Although replacing specific processing recognition sites with inteins would simplify the cost and complexity of a bioprocess, inteins do have some significant limitations: (i) the large size of the intein and its associated catalytic machinery places a greater metabolic burden on the cells; (ii) the processing efficiency is dependent on the amino acid sequences at the junctions; (iii) inteins generally have a very slow autoprocessing rate; and (iv) inteins do not generally enhance the solubility or facilitate the purification of the fusion partner.

1.2.5.5 Affinity Tag-Ligand Interaction

To provide for a robust and selective separation process, the interaction between the fusion tag and the ligand should be a reversible through a mild change in solution conditions. That is, the association should be sufficiently strong and specific, yet under the appropriate condition, allow for an efficient, yet gentle, dissociation. Ideally, the association constant (K_a) should be in the range of 10^4 to 10^8 M^{-1} (Chao et al. 1998). With a K_a less than 10^4 M^{-1} , the interaction will be too weak, resulting in product loss, while with a K_a greater than 10^8 M^{-1} the complex is sufficiently strong to make elution difficult (Lindner et al. 1992).

While identifying the ideal adsorption conditions is often relatively straightforward, determining the elution conditions that result in a high yield with a large purification factor can require extensive trials (Thiele and Fahrenholz 1994). The difficulty lies in identifying a condition that disrupts the tag-ligand interaction without causing degradation or conformational changes in the target protein. Low pH buffers

containing high concentrations of salts or chaotropic chemicals are often used to elute bound fusion proteins, but their application can be detrimental to the target protein (Hofmann et al. 1980; Schmidt and Skerra 1993; Yarmush et al. 1992). The use of excess ligand is a milder alternative, but in some cases, it is inefficient, expensive and provides low yields (Thiele and Fahrenholz 1994). Other protocols include elution based on irradiation with light (Olejnik et al. 1998; Thiele and Fahrenholz 1994), Ca^{+2} dependent binding (Hentz et al. 1996; Schauer-Vukasinovic and Daunert 1999) and even on-column use of proteases (Graslund et al. 1997; Nilsson et al. 1997).

1.2.6 Affinity Tag Systems

The number of binding motifs designed for use as affinity tags has increased significantly over the past three decades. They include small antigenic epitope tags that bind to antibodies immobilized on columns, polyionic tags that adsorb onto ion exchange resins, tags that chelate with immobilized metal ions, biotinylated and non-biotinylated tags, tags derived from protein-protein or enzyme-substrate interactions, and finally, tags that bind to carbohydrate moieties (Table 1.1). All have been used at the laboratory scale, but only a few have succeeded beyond the laboratory that pioneered the given tag and become part of a commercial product (marked with asterisk in Table 1.1).

The wide availability and relative low cost of polysaccharide matrixes might explain the commercially available affinity systems derived from protein tags that bind to carbohydrate based affinity resins. These tags vary in size from ~40 kDa for the maltose binding protein (MBP) (Blondel and Bedouelle 1990; di Guan et al. 1988; Riggs 2000) to about 119 amino acids for the starch binding domain (SBD) (Chen et al. 1991b). Currently, the MBP tag is the best studied and most popular of these carbohydrate-based affinity tags. Fusion proteins containing the MBP affinity tag can be purified on columns containing cross-linked amylose and the bound protein can be desorbed by competitive elution with 10 mM maltose in physiological buffer (Kellermann and Ferenci 1982). The MBP tag has been shown to act as a solubility enhancer (Kapust 1999) but only when fused to the N-terminus of the target protein (Sachdev and Chirgwin 1999).

A number of commercially available affinity systems also take advantage of peptide tags. These tags have flexible structures that minimize steric interference with the correct folding of the target protein and allow for enhanced accessibility of processing enzymes to any introduced cleavage sites. For example, the Flag tag (Hopp 1988) is a hydrophilic, 8 amino acid peptide that binds to immobilized monoclonal antibodies. Elution can be achieved either by the addition of chelating agents such as EDTA or by a decrease in the effective pH of the buffer. The Flag tag offers the unique advantage of encoding for a specific protease processing site at its C-terminus that can subsequently be used to remove the Flag peptide tag and obtain the purified target protein.

Of the tag technologies listed, the glutathione S-transferase and polyhistidine tag technologies have found the greatest use in preparative-scale purifications. Both of these affinity tags were among the first to become commercially available; therefore, their timing, more than performance, might account for their more frequent use. Immobilized metal affinity chromatography (IMAC) is generally used to purify recombinant proteins containing a polyhistidine tag. First described in 1975 (Porath et al. 1975), IMAC is based on the interaction between an immobilized divalent transition-metal ion (Ni^{+2} , Co^{+2} , Cu^{+2} , Zn^{+2}) and solvent exposed aromatic residues on proteins or peptide tails. Histidine, or more specifically the electron-donating nitrogen of the acylimidazole ring of histidine, binds particularly tight through favorable coordination with the electron-accepting transition metal. A number of expression vectors are available that fuse a polyhistidine peptide, typically containing between 5 and 10 histidine residues, to the N- or C-terminus of any recombinant protein of interest (Hochuli et al. 1987). The resulting chimeric protein can then be adsorbed onto traditional IMAC columns containing an immobilized divalent metal ion, usually Ni^{+2} , and eluted either by adjusting the pH of the buffer or by adding free imidazole. The use of imidazole can be problematic, as it can influence NMR experiments, competition studies and crystallographic trials, and its presence can result in protein aggregation (Hefti et al. 2001). In addition, the polyhistidine tag is not recommended for purification of metal containing proteins as the bound metal ion can dissociate, resulting in elution of an apo-protein with potential for denaturation and/or protein aggregation. The IMAC columns designed to bind target proteins tagged with the polyhistidine tail are particularly susceptible to this problem

because the nickel cation that serves as the polyhistidine coordination site is physically adsorbed to the column through binding to immobilized ligands with high affinity for metal ions. Finally, the potential for dissociation and leakage of the divalent metal ion, even at ppm levels, typically prohibits its use in the purification of FDA approved proteins.

The GST affinity tag system, developed by Smith and Johnson (Smith and Johnson 1988), takes advantage of the natural affinity between the enzyme glutathione S-transferase (GST) and its substrate, glutathione. Fusion proteins containing the GST affinity tag can therefore be purified on columns displaying immobilized glutathione. Being the first protein-based affinity tag system commercially available for protein purification, the GST tag technology has gained widespread use and an almost exclusive market share of large-scale purifications of therapeutic proteins that cannot be purified by conventional methods. The GST tag is, however, a homodimer (Kaplan et al. 1997), making it unsuitable for the purification of oligomeric proteins.

All of the affinity tags listed in Table 1.1 exhibit one or more of a number of characteristics that have proven disadvantageous for production-scale purification of therapeutic proteins. Most require a ligand that is expensive to produce and/or to immobilize. For example, the coiled-coil affinity tag system requires synthesis and immobilization of a long peptide sequence (35 amino acids), composed of five consecutive repeats of a specific heptad sequence. Robust chemistries for coupling these ligands to the stationary phase are often not available, resulting in gradual leaching of the ligand into the mobile phase during column processing. The potential for dissociation and leakage of the ligand, even at ppm levels, typically prohibits their use in the purification of FDA approved proteins.

Other, more tag specific problems have also been reported. For example, the GST tag has the propensity to produce fusion proteins in inclusion bodies (denatured protein aggregates), possibly due to the four solvent exposed cysteine residues present in each subunit (Kaplan et al. 1997). Thus, recovery of the purified protein in its native, fully active state requires a robust refolding strategy. Difficulties in eluting chitin-binding

protein tagged fusion proteins have been reported (Vaillancourt et al. 2000), and low binding efficiencies have been observed with MBP fusions (Pryor 1997).

However, the dominant impediment to the use of these affinity tags in large-scale bioprocessing is cost. Table 1.2 lists the costs associated with using some of the common commercial affinity tags. Production-scale affinity chromatography typically requires column volumes greater than 10 L and up to 5000 L. For the systems reported in Table 1.2, this represents columns where the resin cost alone is in excess of \$30,000 and can be as high as \$11,000,000. A new generic affinity tag that binds to an inexpensive, chemically and hydrodynamically robust chromatography media is therefore highly desirable.

1.2.7 Carbohydrate Binding Modules

Carbohydrate binding modules (CBMs) are discrete protein modules found in a large number of hydrolases involved in the degradation of biomass (Gilkes et al. 1991; Tomme et al. 1995b). Within these carbohydrases, the CBMs function to bind the enzyme to its substrate and in so doing, increase the effective substrate concentration proximal to the enzyme (Warren 1996). CBMs are also found in a few non-hydrolytic proteins (Goldstein et al. 1993; Shoseyov and Doi 1990; Tomme et al. 1995b) where they function as part of a scaffolding subunit that organizes synergistic enzymes into a cohesive multienzyme complex called a cellulosome, which then binds strongly to cellulose to catalyze its degradation (Beguin and Aubert 1994).

More than 200 CBMs are known and can be classified into forty eight unique families, seventeen of which are known to contain members that bind cellulose [Carbohydrate Active Enzymes database: (<http://www.cazy.org/>) (Coutinho and Henrissat 1999)]. CBMs from different families have different binding affinities and binding specificities (Tomme et al. 1998). For example, CfXyn10A-CBM2a (hereafter referred to as CBM2a), a family 2 CBM from xylanase 10A of *Cellulomonas fimi*, binds crystalline regions of insoluble cellulose with a dissociation constant in the low micromolar range (McLean 2000). In contrast, the family 4 CBM, CfCel9B-CBM4-1 from endoglucanase 9B of *C. fimi*, binds amorphous cellulose and water-soluble cello-

oligosaccharides, but shows no affinity for crystalline cellulose (Abou-Hachem et al. 2000).

CBMs come in a wide range of sizes from larger domains of up to 180 amino acids (family III) to small (~35 amino acids) fungal domains (family 1 CBMs) and are found in various positions within the parent carbohydrase. As shown in Figure 1.1, Cel9A from *Cellulomonas fimi* has an internal family 3 CBM and a C-terminal family 2a CBM, while Cel9B, also from *Cellulomonas fimi*, has two family 4 CBMs in tandem located on the N-terminal side of its catalytic module. CfCel6A has an N-terminal family 2a CBM binding module while CfXyn10A has a family 2a CBM at the C-terminus. In addition, naturally occurring N-terminal CBMs have been translocated to the C-terminus and used as C-terminal affinity tags in purification of recombinant protein without loss of binding affinity (Tomme et al. 1994). The ability to use a CBM as either a N- or C-terminal tag offers flexibility in the purification of recombinant proteins, especially in cases where the protein of interest can only tolerate foreign sequences at one end to remain active.

CBMs have been used as tags for immobilization or purification of a number of recombinant proteins. For example, CBMs have been fused to a number of bioprocessing enzymes such as *E. coli* alkaline phosphatase (Greenwood et al. 1992), *Agrobacterium* β -glucosidase (Ong 1991) and human factor X (Assouline et al. 1993), as well as to medically relevant enzymes such as heparinase (Shpigel 1999) and interleukin-2 (Ong 1995). Capture agents such as *Streptomyces* streptavidin (Le 1994) and *Staphylococcus* protein A (Ramirez et al. 1993) have been fused to a CBM to permit their immobilization on cellulose for binding of the target ligand. CBMs have also been expressed in a number of different hosts, including *E. coli* (Greenwood et al. 1992; Le 1994; Ong et al. 1989; Ramirez et al. 1993; Shpigel 1999; Shpigel 2000), the yeast *P. pastoris* (Doheny et al. 1999; Ong 1995), mammalian cells (Assouline et al. 1993; Guarna et al. 2000; Ong 1995), and insect cells (Pfeifer and Theilmann 2001). Although not well-understood, periplasmic expression of fusion proteins with a CBM from *C. fimi* often result in their non-specific leakage into the culture supernatant (Hasenwinkle et al. 1997; Ong et al. 1993; Tomme et al. 1995a). This property offers great potential not only for expression

of proteins with complex disulfide bond formation but would also simplify any purification process by directing the protein into an environment with less protein contaminant.

1.2.8 The Family 9 Carbohydrate-Binding Module (CBM9)

While CBMs from some families can and have served as affinity tags for recombinant protein purification at the laboratory scale (Ong 1995; Sakka 1998; Shpigel 1999), the family 9 carbohydrate-binding module (CBM9) of xylanase 10A from the thermophilic bacterium *Thermotoga maritima* (Winterhalter et al. 1995) has not; yet it offers some unique functional characteristics that appear to make it a particularly useful tag for preparative-scale affinity purification applications. In particular, CBM9 is able to bind to both soluble sugars and insoluble cellulose (amorphous cellulose) (Boraston et al. 2001). As a result, CBM9 bound to insoluble preparations of cellulose can be dissociated from the solid substrate by the addition of a soluble sugar, such as glucose. The binding specificity of CBM9 has recently been characterized (Boraston et al. 2001) and its crystal structure determined up to 1.9 Å resolution (Notenboom et al. 2001), providing a sound fundamental basis for designing an efficient purification technology based on the use of CBM9 as an affinity fusion tag.

CBM9 is an extremely thermostable, 22 kDa globular protein (Wassenberg et al. 1997) with the unique ability to bind to the reducing ends of cellulose and soluble polysaccharides. It binds three Ca^{+2} ions, thought to contribute to the unusually high thermal stability of this protein (Notenboom et al. 2001). Its structure consists of two twisted anti-parallel β -sheets, each made up of five β -strands that form a β -sandwich fold (Figure 1.2). The crystal structure of cellobiose bound CBM9 indicates that two tryptophan residues function to sandwich the substrate within the binding pocket (Figure 1.2). Thermodynamic (Boraston et al. 2001) and structural analyses (Notenboom et al. 2001) confirm that the disaccharide cellobiose is sufficient to fully occupy the binding site.

1.2.9 Cellulose-Based Chromatography Resins

Beaded cellulose and various derivatives of cellulose are currently used as stationary phase media for a number of commercial liquid chromatography applications. Size-exclusion chromatography media comprised of cellulose include Whatman CC31, which is a pure microgranular cellulose powder, Whatman CF1, a particulate microcrystalline cellulose optimized for batch chromatography, Perloza™ MT, a beaded cellulose gel filtration media, Perloza™ ST, a dry regenerated cellulose that can be used as initial material for the preparation of all basic types of ion exchange and for a number of further modified derivatives of spherical cellulose, Perloza™ SF, a pure dry beaded cellulose that has been sterilised by radiation, and Prolinx Versalinx™, a cross-linked cellulose network. Most of these media are available in a range of bead sizes and porosities to facilitate optimization of the chromatographic separation for a given feedstock. More importantly, these cellulose-based media are generally very inexpensive, can withstand stringent clean-in-place (CIP) protocols, and have good mechanical properties. For example, the bulk cost of Perloza™ MT100 is \$35 USD per liter, or about 1/100 the bulk cost of an equivalent amount of the glutathione affinity resin used in conjunction with the GST tag system.

1.3 Thesis Objectives

Generic affinity chromatography based on affinity tag technology has the potential to simplify downstream processing by achieving higher yields and purities than conventional methods. However, the high cost of current affinity tag technologies, due mainly to the expense of their associated affinity chromatography media, limits their application at production scales. This thesis aims to alleviate this cost limitation by introducing an affinity tag that binds to a very inexpensive cellulose-based matrix. Development of this technology will involve the following research objectives: (i) design of a generic CBM9 expression vector for production of chimeric fusions containing an N-terminal CBM9, a linker region containing a suitable processing site at its C-terminus for efficient removal of the affinity tag following affinity purification, and a C-terminal target protein; (ii) validate the performance of the affinity system in terms of product

yield, purity and concentration factor; (iii) develop a useful strategy to design an effective linker sequence to stably connect the CBM9 tag to the target protein and to permit efficient tag removal through enzyme-catalyzed cleavage; (iv) derive and validate a mathematical model to predict binding and elution behavior of CBM9 fusion proteins on a Perloza™ MT100 column; and finally, (v) briefly address technology scale-up issues, including the synthesis of a mechanically stable stationary phase.

The first objective will provide industry with a ready-made expression vector that can be used to express any target protein as a CBM9 fusion protein. The second objective will provide evidence that my proposed technology provides for efficient production and purification of a target recombinant protein with an overall level of performance that is either superior or comparable to current commercially available fusion-tag technologies. Although often overlooked, the linker is an important component of any affinity tag technology. The third objective will look at a number of strategies for selecting an optimal linker that will provide for *in vivo* stability, efficient processing with the serine protease rhFactor X_a or enterokinase, and independent functioning of the two domains. The fourth and fifth objectives will help industry with the application of the CBM9 fusion tag technology.

1.4 Tables

Table 1.1 Affinity tag systems described in the literature. Tag size and complementary ligand immobilized on the stationary phase are given. Tags that are currently offered commercially are denoted by *.

Affinity tag	Size	Immobilized Ligand or Matrix	Reference
<i>Polypeptide-binding proteins</i>			
Staphylococcal Protein* A (SpA)	14-31 kDa	IgG	(Moks et al. 1987; Nilsson and Abrahmsen 1990; Uhlen et al. 1983)
ZZ domains	7 kDa	IgG	(Hansson et al. 1995; Nilsson et al. 1996)
Albumin-binding protein	5-25 kDa	HSA	(Nygren et al. 1988)
Phosphate-binding domain	200-263 kDa	Hydroxyapatite	(Anba et al. 1987)
<i>Carbohydrate-binding domains</i>			
Maltose-binding protein*(MBP)	40 kDa	Cross-linked amylose	(Bedouelle and Duplay 1988; di Guan et al. 1988)
Starch-binding domain	119 aa	Starch	(Chen et al. 1991a; Chen et al. 1991b)
Chitin-binding domain*	51 aa	Chitin	(Chong et al. 1997; Mathys et al. 1999)
Cellulose-binding domain (families 1, 2a* and 3*)	27- 189 aa	Cellulose	(Greenwood et al. 1989; Ong 1995; Tomme et al. 1998)
<i>Enzymes</i>			
β -Galactosidase	116 kDa	APTG	(Germino and Bastia 1984; Oliaro et al. 2000)
Glutathione-S- transferase* (GST)	26 kDa	Glutathione	(Guan and Dixon 1991; Smith and Johnson 1988)
Chloramphenicol acetyltransferase	24 kDa	chloramphenicol	(Dekeyzer et al. 1994; Dykes et al. 1988)

Affinity tag	Size	Immobilized Ligand or Matrix	Reference
<i>Biotin-binding domains</i>			
Biotin-binding domain	8 kDa	Streptavidin	(Cronan 1990)
Avidin	63 kDa	2-iminobiotin-agarose	(Airenne et al. 1999)
<i>Metal-affinity tags</i>			
Poly-Histidine*	6-10 aa	Ni ⁺² -NTA; Co ⁺² -CMA	(Crowe et al. 1994; Porath et al. 1975)
ProHis ₅ Pro	7 aa	Ni ⁺² -NTA; Zn ⁺² -NTA	(Skerra et al. 1991)
HAT*	19 aa	Ni ⁺² -NTA	(Chaga et al. 1999)
<i>Non-biotinylated affinity tags</i>			
Strep-tag II*	8 aa	streptavidin	(Schmidt 1996; Voss and Skerra 1997)
Streptavidin-binding peptide* (SBP)	38 aa	Streptavidin	(Keefe 2001; Wilson et al. 2001)
<i>Antigenic epitopes (Ca⁺² dependent)</i>			
FLAG*	8 aa	Anti-FLAG M1 mAb	(Einhauer and Jungbauer 2001; Hopp 1988)
FLAG II*	4 aa	Anti-FLAG M2 mAb	(Brizzard et al. 1994)
HPC4-binding peptide	12 aa	Anti-KT3 mAb	(Martin et al. 1990)
<i>Antigenic epitopes (Ca⁺² independent)</i>			
Myc	10 aa	CA5 mAb	(Chen et al. 1993; Munro and Pelham 1986)
T7	11 aa	Anti-T7 mAb	(Borjigin and Nathans 1994)
AU5	6 aa	Anti-AU5 mAb	(Lutzfreyermuth et al. 1990)
Btag	6 aa	Anti-Btag mAb	(Rubinfeld et al. 1991; Wang et al. 1996)
ε-tag	12 aa	Anti-ε-tag mAb	(Olah et al. 1994)

Affinity tag	Size	Immobilized Ligand or Matrix	Reference
VSV tag	11 aa	Anti-VSV mAb	(Green et al. 1996)
RecA	144 aa	Anti-recA mAb	(Krivi et al. 1985)
<i>Charged amino acid tags</i>			
Poly-Arginine	5-15 aa	Anionic resins	(Sassenfeld 1984; Smith 1984)
Poly-Phenylalanine	11 aa	Phenyl-superose	(Persson et al. 1988)
<i>Other tags</i>			
Calmodulin-binding peptide* (CBP)	2.96	Calmodulin	(Stofko-Hahn et al. 1992; Vaillancourt et al. 1997; Zheng et al. 1997)
S-peptide	15 aa	S-protein	(Kim and Raines 1993; Raines et al. 2000)
Heterodimeric coiled-coil tags	35 aa	Immobilized hexapeptide	(Tripet et al. 2000; Tripet et al. 1996)

Table 1.2 Estimated costs of several commercially available affinity tag technologies. Prices provided by respective vendors.

System	Estimated Cost (USD)	
	Resin (100 mL)	Eluent (\$/L)
CBP	\$2200	\$0.86 (2 mM EGTA)
GST	\$1067	\$10.07 (10 mM reduced glutathione)
Polyhistidine	\$656	\$2.87 (250 mM imidazole)
MBP	\$300	\$0.71 (10 mM maltose)

1.5 Figures

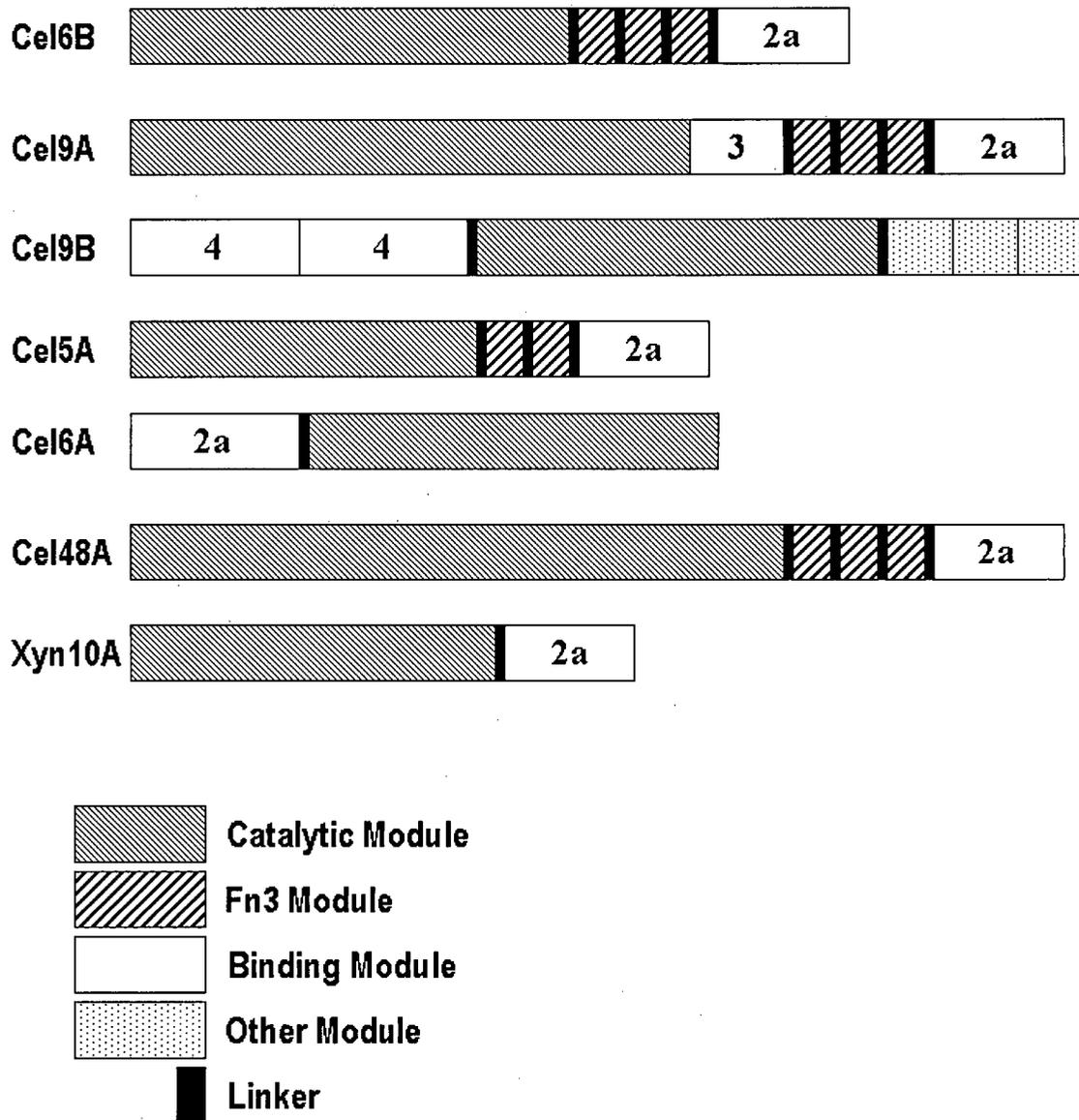


Figure 1.1 Schematic representation of the domain structure of *Cellulomonas fimi* cellulases and xylanases. Note: the domain structure of xylanase 10A of *T. maritima* is homologous to that of Xyn10A of *C. fimi* with the family 2a CBM9 substituted by CBM9.

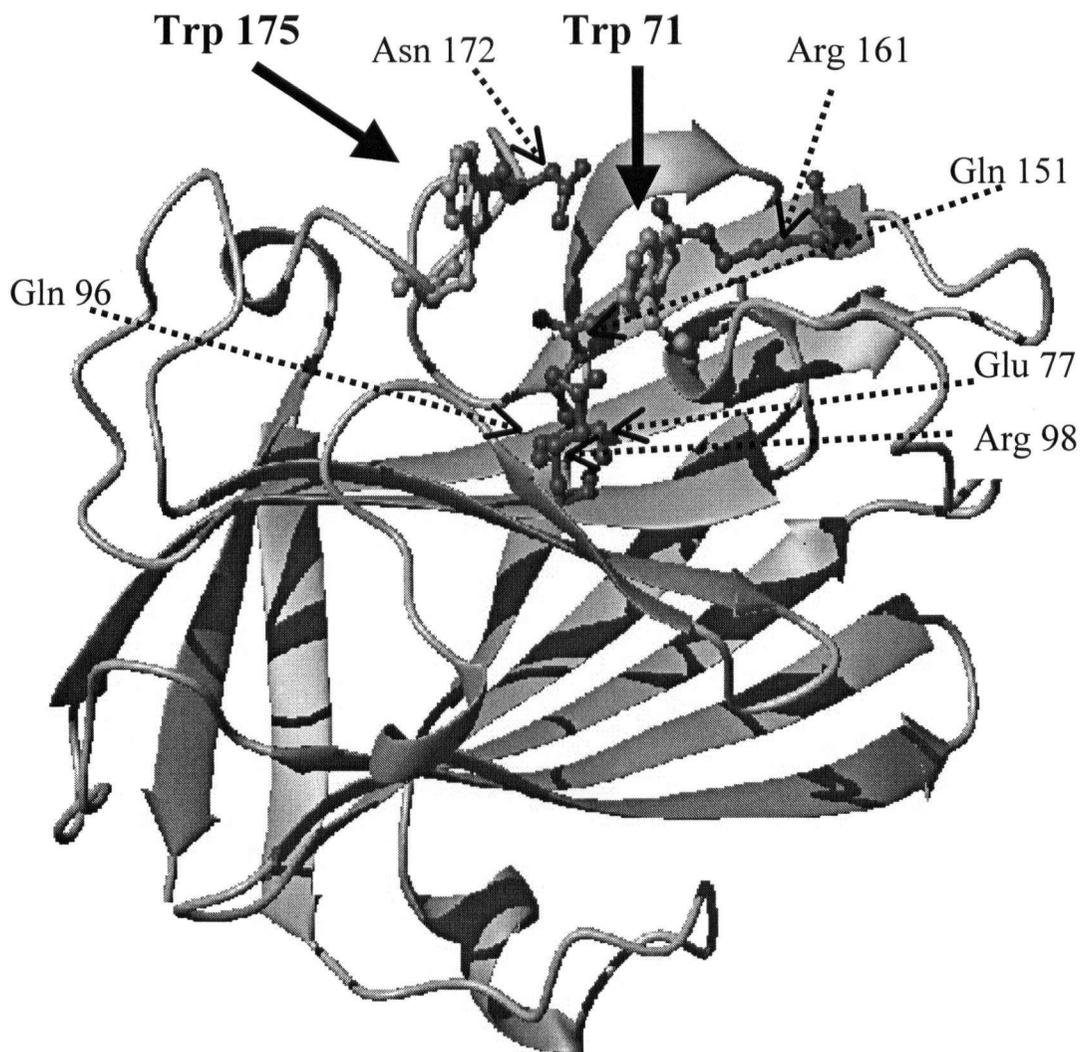


Figure 1.2 Backbone secondary structure of CBM9 from *T. maritima*, presented in ribbon diagram form, determined from the x-ray crystallographic data of Notenboom (Notenboom et al. 2001). Trp71 and Trp 175 (solid arrow) involved in ligand binding, and Glu77, Gln96, Arg98, Gln151, Arg161, and Asn172 (dashed arrow), involved in hydrogen bonding, with the ligand are shown.

1.6 References

- Abou-Hachem, M.; Nordberg-Karlsson, E.; Bartonek-Roxa, E.; Raghothama, S.; Simpson, P. J.; Gilbert, H. J.; Williamson, M. P.; Holst, O. (2000). "Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies". *Biochemical Journal* 345(1):53-60.
- Abrahmsen, L.; Moks, T.; Nilsson, B.; Uhlen, M. (1986). "Secretion of Heterologous Gene-Products to the Culture-Medium of *Escherichia-Coli*". *Nucleic Acids Research* 14(18):7487-7500.
- Airenne, K. J.; Marjomaki, V. S.; Kulomaa, M. S. (1999). "Recombinant avidin and avidin-fusion proteins". *Biomolecular Engineering* 16(1-4):87-92.
- Allet, B.; Bernard, A. R.; Hochmann, A.; Rohrbach, E.; Graber, P.; Magnenat, E.; Mazzei, G. J.; Bernasconi, L. (1997). "A bacterial signal peptide directs efficient secretion of eukaryotic proteins in the baculovirus expression system". *Protein Expression and Purification* 9(1):61-68.
- Anba, J.; Baty, D.; Llobes, R.; Pages, J. M.; Joseph-Liauzun, E.; Shire, D.; Roskam, W.; Lazdunski, C. (1987). "Expression vector promoting the synthesis and export of the human growth-hormone-releasing factor in *Escherichia coli*". *Gene* 53(2-3):219-226.
- Andrade, E. V.; Albuquerque, F. C.; Moraes, L. M.; Brigido, M. M.; Santos-Silva, M. A. (2000). "Single-chain Fv with Fc fragment of the human IgG1 tag: construction, *Pichia pastoris* expression and antigen binding characterization". *Journal of Biochemistry* 128(6):891-895.
- Assouline, Z.; Shen, H.; Kilburn, D. G.; Warren, R. A. J. (1993). "Production and Properties of a Factor-X Cellulose-Binding Domain Fusion Protein". *Protein Engineering* 6(7):787-792.
- Bardwell, J. C. A.; McGovern, K.; Beckwith, J. (1991). "Identification of a Protein Required for Disulfide Bond Formation *In vivo*". *Cell* 67(3):581-589.
- Bedouelle, H.; Duplay, P. (1988). "Production in *Escherichia coli* and one-step purification of bifunctional hybrid proteins which bind maltose. Export of the Klenow polymerase into the periplasmic space". *European Journal of Biochemistry* 171(3):541-549.
- Beguín, P.; Aubert, J. P. (1994). "The Biological Degradation of Cellulose". *FEMS Microbiology Reviews* 13(1):25-58.

- Blondel, A.; Bedouelle, H. (1990). "Export and Purification of a Cytoplasmic Dimeric Protein by Fusion to the Maltose-Binding Protein of Escherichia-Coli". *European Journal of Biochemistry* 193(2):325-330.
- Bonnerjea, J.; Oh, S.; Hoare, M.; Dunnill, P. (1986). "Protein purification: the right step at the right time". *Bio/Technology* 4(11):954, 956, 958.
- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Borjigin, J.; Nathans, J. (1994). "Insertional Mutagenesis as a Probe of Rhodopsins Topography, Stability, and Activity". *Journal of Biological Chemistry* 269(20):14715-14722.
- Brewer, S. J.; Haymore, B. L.; Hopp, T. P.; Sassenfeld, H. M. (1991). "Engineering proteins to enable their isolation in a biologically active form". *Bioprocess Technology* 12:239-266.
- Brizzard, B. L.; Chubet, R. G.; Vizard, D. L. (1994). "Immunoaffinity purification of FLAG epitope-tagged bacterial alkaline phosphatase using a novel monoclonal antibody and peptide elution". *Biotechniques* 16(4):730-732, 734-735.
- Capon, D. J.; Chamow, S. M.; Mordenti, J.; Marsters, S. A.; Gregory, T.; Mitsuya, H.; Byrn, R. A.; Lucas, C.; Wurm, F. M.; Groopman, J. E.; Broder, S.; Smith, D. H. (1989). "Designing Cd4 Immuno adhesins for Aids Therapy". *Nature* 337(6207):525-531.
- Carter, P.; Kelley, R. F.; Rodrigues, M. L.; Snedecor, B.; Covarrubias, M.; Velligan, M. D.; Wong, W. L. T.; Rowland, A. M.; Kotts, C. E.; Carver, M. E.; Yang, M.; Bourell, J. H.; Shepard, H. M.; Henner, D. (1992). "High-Level Escherichia-Coli Expression and Production of a Bivalent Humanized Antibody Fragment". *Bio/Technology* 10(2):163-167.
- Casey, J. L.; Keep, P. A.; Chester, K. A.; Robson, L.; Hawkins, R. E.; Begent, R. H. J. (1995). "Purification of Bacterially Expressed Single-Chain Fv Antibodies for Clinical-Applications Using Metal Chelate Chromatography". *Journal of Immunological Methods* 179(1):105-116.
- Chaga, G.; Bochkariov, D. E.; Jokhadze, G. G.; Hopp, J.; Nelson, P. (1999). "Natural poly-histidine affinity tag for purification of recombinant proteins on cobalt(II)-carboxymethylaspartate crosslinked agarose". *Journal of Chromatography A* 864(2):247-256.

- Chao, H. M.; Bautista, D. L.; Litowski, J.; Irvin, R. T.; Hodges, R. S. (1998). "Use of a heterodimeric coiled-coil system for biosensor application and affinity purification". *Journal of Chromatography B* 715(1):307-329.
- Chen, C.; Billingsley, P. F. (1999). "Detection and characterization of a mannan-binding lectin from the mosquito, *Anopheles stephensi* (Liston)". *European Journal of Biochemistry* 263(2):360-366.
- Chen, C.; Huang, Q.-L.; Jiang, S.-H.; Pan, X.; Hua, Z.-C. (2006). "Immobilized protein ZZ, an affinity tool for immunoglobulin isolation and immunological experimentation". *Biotechnology and Applied Biochemistry* 45(2):87-92.
- Chen, L. J.; Ford, C.; Kusnadi, A.; Nikolov, Z. L. (1991a). "Improved Adsorption to Starch of a Beta-Galactosidase Fusion Protein Containing the Starch-Binding Domain from *Aspergillus Glucoamylase*". *Biotechnology Progress* 7(3):225-229.
- Chen, L. J.; Ford, C.; Nikolov, Z. (1991b). "Adsorption to Starch of a Beta-Galactosidase Fusion Protein Containing the Starch-Binding Region of *Aspergillus Glucoamylase*". *Gene* 99(1):121-126.
- Chen, Y. T.; Holcomb, C.; Moore, H. P. H. (1993). "Expression and localization of two molecular weight GTP-binding proteins, Rab8 and Rab10, by epitope tag". *Proceedings of the National Academy of Sciences of the United States of America* 90(14):6508-6512.
- Chong, S.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M. Q. (1997). "Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element". *Gene* 192(2):271-281.
- Chubet, R. G.; Brizzard, B. L. (1996). "Vectors for expression and secretion of FLAG epitope-tagged proteins in mammalian cells". *Biotechniques* 20(1):136-141.
- Clonis, Y. D.; Labrou, N. E.; Kotsira, V. P.; Mazitsos, C.; Melissis, S.; Gogolas, G. (2000). "Biomimetic dyes as affinity chromatography tools in enzyme purification". *Journal of Chromatography, A* 891(1):33-44.
- Coutinho, P. M.; Henrissat, B. (1999). "Carbohydrate-active enzymes: an integrated database approach". *Special Publication - Royal Society of Chemistry* 246(Recent Advances in Carbohydrate Bioengineering):3-12.
- Cronan, J. E., Jr. (1990). "Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins". *Journal of Biological Chemistry* 265(18):10327-10333.

- Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stuber, D.; Henco, K. (1994). "6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification". *Methods in Molecular Biology* 31:371-387.
- Cummings, R. D. (1997). "Lectins as tools for glycoconjugate purification and characterization". *Glycosciences*:191-199.
- Dekeyzer, N.; Engelborghs, Y.; Volckaert, G. (1994). "Cloning, expression and purification of a sarcoplasmic calcium-binding protein from the sandworm *Nereis diversicolor* via a fusion product with chloramphenicol acetyltransferase". *Protein Engineering* 7(1):125-130.
- Denizli, A.; Piskin, E. (2001). "Dye-ligand affinity systems". *Journal of Biochemical and Biophysical Methods* 49(1-3):391-416.
- Derbyshire, V.; Wood, D. W.; Wu, W.; Dansereau, J. T.; Dalgaard, J. Z.; Belfort, M. (1997). "Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution". *Proceedings of the National Academy of Sciences of the United States of America* 94(21):11466-11471.
- Deutsch, D. G.; Mertz, E. T. (1970). "Plasminogen: purification from human plasma by affinity chromatography". *Science* 170(962):1095-1096.
- di Guan, C.; Li, P.; Riggs, P. D.; Inouye, H. (1988). "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein". *Gene* 67(1):21-30.
- Doheny, J. G.; Jervis, E. J.; Guarna, M. M.; Humphries, R. K.; Warren, R. A. J.; Kilburn, D. G. (1999). "Cellulose as an inert matrix for presenting cytokines to target cells: production and properties of a stem cell factor-cellulose-binding domain fusion protein". *Biochemical Journal* 339(2):429-434.
- Dykes, C. W.; Bookless, A. B.; Coomber, B. A.; Noble, S. A.; Humber, D. C.; Hobden, A. N. (1988). "Expression of atrial natriuretic factor as a cleavable fusion protein with chloramphenicol acetyltransferase in *Escherichia coli*". *European Journal of Biochemistry* 174(2):411-416.
- Einhauer, A.; Jungbauer, A. (2001). "The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins". *Journal of Biochemical and Biophysical Methods* 49(1-3):455-465.
- Fassina, G.; Verdoliva, A.; Palombo, G.; Ruvo, M.; Cassani, G. (1998). "Immunoglobulin specificity of TG19318: a novel synthetic ligand for antibody affinity purification". *Journal of Molecular Recognition* 11(1-6):128-133.

- Flaschel, E.; Friehs, K. (1993). "Improvement of Downstream Processing of Recombinant Proteins by Means of Genetic-Engineering Methods". *Biotechnology Advances* 11(1):31-77.
- Follman, D. K.; Fahrner, R. L. (2004). "Factorial screening of antibody purification processes using three chromatography steps without protein A". *Journal of Chromatography, A* 1024(1-2):79-85.
- Ford, C. F.; Suominen, I.; Glatz, C. E. (1991). "Fusion tails for the recovery and purification of recombinant proteins". *Protein Expression and Purification* 2(2-3):95-107.
- Germino, J.; Bastia, D. (1984). "Rapid Purification of a Cloned Gene-Product by Genetic Fusion and Site-Specific Proteolysis". *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 81(15):4692-4696.
- Gilkes, N. R.; Henrissat, B.; Kilburn, D. G.; Miller, R. C., Jr.; Warren, R. A. J. (1991). "Domains in microbial beta-1, 4-glycanases: sequence conservation, function, and enzyme families". *Microbiological Reviews* 55(2):303-315.
- Goel, A.; Colcher, D.; Koo, J. S.; Booth, B. J. M.; Pavlinkova, G.; Batra, S. K. (2000). "Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct". *Biochimica Et Biophysica Acta-General Subjects* 1523(1):13-20.
- Goldstein, M. A.; Takagi, M.; Hashida, S.; Shoseyov, O.; Doi, R. H.; Segel, I. H. (1993). "Characterization of the Cellulose-Binding Domain of the Clostridium-Cellulovorans Cellulose-Binding Protein-A". *Journal of Bacteriology* 175(18):5762-5768.
- Graslund, T.; Nilsson, J.; Lindberg, A. M.; Uhlen, M.; Nygren, P. A. (1997). "Production of a thermostable DNA polymerase by sites-specific cleavage of a heat-eluted affinity fusion protein". *Protein Expression and Purification* 9(1):125-132.
- Green, E. A.; Botting, C.; Webb, H. M.; Hirst, T. R.; Randall, R. E. (1996). "Construction, purification and immunogenicity of antigen-antibody-LTB complexes". *Vaccine* 14(10):949-958.
- Greenwood, J. M.; Gilkes, N. R.; Kilburn, D. G.; Miller, R. C., Jr.; Warren, R. A. J. (1989). "Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose". *FEBS Letters* 244(1):127-131.
- Greenwood, J. M.; Ong, E.; Gilkes, N. R.; Warren, R. A. J.; Miller, R. C.; Kilburn, D. G. (1992). "Cellulose-Binding Domains - Potential for Purification of Complex Proteins". *Protein Engineering* 5(4):361-365.

- Guan, K. L.; Dixon, J. E. (1991). "Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Guarna, M. M.; Cote, H. C.; Kwan, E. M.; Rintoul, G. L.; Meyhack, B.; Heim, J.; MacGillivray, R. T.; Warren, R. A. J.; Kilburn, D. G. (2000). "Factor X fusion proteins: improved production and use in the release in vitro of biologically active hirudin from an inactive alpha-factor-hirudin fusion protein". *Protein Expression and Purification* 20(2):133-141.
- Hahn, R.; Shimahara, K.; Steindl, F.; Jungbauer, A. (2006). "Comparison of protein A affinity sorbents III. Life time study". *Journal of Chromatography, A* 1102(1-2):224-231.
- Hakes, D. J.; Dixon, J. E. (1992). "New Vectors for High-Level Expression of Recombinant Proteins in Bacteria". *Analytical biochemistry* 202(2):293-298.
- Hammarberg, B.; Nygren, P. A.; Holmgren, E.; Elmlad, A.; Tally, M.; Hellman, U.; Moks, T.; Uhlen, M. (1989). "Dual affinity fusion approach and its use to express recombinant human insulin-like growth factor II". *Proceedings of the National Academy of Sciences of the United States of America* 86(12):4367-4371.
- Hansen, M. B.; Lihme, A.; Spitali, M.; King, D. (1998). "Capture of human Fab fragments by expanded bed adsorption with a mixed mode adsorbent". *Bioseparation* 8(1-5):189-193.
- Hansson, L.; Noppa, L.; Nilsson, A. K.; Stromqvist, M.; Bergstrom, S. (1995). "Expression of Truncated and Full-Length Forms of the Lyme-Disease Borrelia Outer Surface Protein-a in Escherichia-Coli". *Protein Expression and Purification* 6(1):15-24.
- Hasenwinkle, D.; Jarvis, E.; Kops, O.; Liu, C.; Lesnicki, G.; Haynes, C. A.; Kilburn, D. G. (1997). "Very high-level production and export in Escherichia coli of a cellulose binding domain for use in a generic secretion-affinity fusion system". *Biotechnology and Bioengineering* 55(6):854-863.
- Hearn, M. T.; Acosta, D. (2001). "Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins". *Journal of Molecular Recognition* 14(6):323-369.
- Hearn, M. T. W.; Anspach, B. (1990). "Chemical, physical, and biochemical concepts in isolation and purification of proteins". *Bioprocess Technology 9(Separation Processes Biotechnology)*:17-64.

- Hearn, M. T. W.;Gomme, P. T. (2000). "Molecular architecture and biorecognition processes of the cystine knot protein superfamily: Part I. The glycoprotein hormones". *Journal of Molecular Recognition* 13(5):223-278.
- Hefti, M. H.; Van Vugt-Van der Toorn, C. J. G.; Dixon, R.;Vervoort, J. (2001). "A novel purification method for histidine-tagged proteins containing a thrombin cleavage site". *Analytical Biochemistry* 295(2):180-185.
- Hennig, L.;Schafer, E. (1998). "Protein purification with C-terminal fusion of maltose binding protein". *Protein Expression and Purification* 14(3):367-370.
- Hentz, N. G.; Vukasinovic, V.;Daunert, S. (1996). "Affinity chromatography of recombinant peptides proteins based on a calmodulin fusion tail". *Analytical Chemistry* 68(9):1550-1555.
- Hermanson, I. L.;Turchi, J. J. (2000). "Overexpression and purification of human XPA using a Baculovirus expression system". *Protein Expression and Purification* 19(1):1-11.
- Higgins, D. R.;Cregg, J. M. (1998). "Introduction to *Pichia pastoris*". *Methods in Molecular Biology* 103(*Pichia* Protocols):1-15.
- Hochuli, E.; Dobeli, H.;Schacher, A. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues". *Journal of Chromatography* 411:177-184.
- Hockney, R. C. (1994). "Recent Developments in Heterologous Protein-Production in *Escherichia-Coli*". *Trends in Biotechnology* 12(11):456-463.
- Hoffmann, A.; Roeder, R. G. (1991). "Purification of His-Tagged Proteins in Nondenaturing Conditions Suggests a Convenient Method for Protein-Interaction Studies". *Nucleic Acids Research* 19(22):6337-6338.
- Hofmann, K.; Wood, S. W.; Brinton, C. C.; Montibeller, J. A.;Finn, F. M. (1980). "Iminobiotin affinity columns and their application to retrieval of streptavidin". *Proceedings of the National Academy of Sciences of the United States of America* 77(8):4666-4668.
- Hopp, T., Pricket, KS, Price, VL, Libby, RT, March, CJ, Ceretti, DP, Urdal, DL, Conlon, PJ. (1988). "A short polypeptide marker sequence useful for recombinant protein identification and purification". *Bio/Technology* 6:1204-1210.
- Hosfield, T.; Lu, Q. (1999). "Influence of the amino acid residue downstream of (Asp)(4)Lys on enterokinase cleavage of a fusion protein". *Analytical Biochemistry* 269(1):10-16.

- Itakura, K.; Hirose, T.; Crea, R.; Riggs, A. D.; Heyneker, H. L.; Bolivar, F.; Boyer, H. W. (1977). "Expression in Escherichia-Coli of a Chemically Synthesized Gene for Hormone Somatostatin". *Science* 198(4321):1056-1063.
- Jonasson, P.; Nilsson, J.; Samuelsson, E.; Moks, T.; Stahl, S.; Uhlen, M. (1996). "Single-step trypsin cleavage of a fusion protein to obtain human insulin and its C peptide". *European Journal of Biochemistry* 236(2):656-661.
- Jones, C.; Patel, A.; Griffin, S.; Martin, J.; Young, P.; Odonnell, K.; Silverman, C.; Porter, T.; Chaiken, I. (1995). "Current Trends in Molecular Recognition and Bioseparation". *Journal of Chromatography A* 707(1):3-22.
- Jones, K. (1990). "Affinity chromatography: a technology update". *American Biotechnology Laboratory* 8(13):26-30.
- Jungbauer, A.; Hahn, R. (2004). "Engineering protein A affinity chromatography". *Current Opinion in Drug Discovery & Development* 7(2):248-256.
- Kane, J. F.; Hartley, D. L. (1988). "Formation of Recombinant Protein Inclusion-Bodies in Escherichia-Coli". *Trends in Biotechnology* 6(5):95-101.
- Kaplan, W.; Husler, P.; Klump, H.; Erhardt, J.; Sluis-Cremer, N.; Dirr, H. (1997). "Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: A detoxification enzyme and fusion-protein affinity tag". *Protein Science* 6(2):399-406.
- Kapust, R. B., Waugh, D. S. (1999). "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused". *Protein Science* 8(8):1668-1674.
- Keefe, A. D., Wilson, D. S., Seelig, B., Szostak, J. W. (2001). "One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag". *Protein Expression and Purification* 23(3):440-446.
- Kellermann, O. K.; Ferenci, T. (1982). "Maltose-Binding Protein from Escherichia-Coli". *Methods in Enzymology* 90:459-463.
- Kim, J. S.; Raines, R. T. (1993). "Ribonuclease S-Peptide as a Carrier in Fusion Proteins". *Protein Science* 2(3):348-356.
- Kleiner, O.; Butenandt, L.; Carell, T.; Batschauer, A. (1999). "Class II DNA photolyase from Arabidopsis thaliana contains FAD as a cofactor". *European Journal of Biochemistry* 264(1):161-167.
- Krivi, G. G.; Bittner, M. L.; Rowold, E., Jr.; Wong, E. Y.; Glenn, K. C.; Rose, K. S.; Tiemeier, D. C. (1985). "Purification of recA-based fusion proteins by

- immunoabsorbent chromatography. Characterization of a major antigenic determinant of *Escherichia coli* recA protein". *Journal of Biological Chemistry* 260(18):10263-10267.
- Kronina, V. V.; Wirth, H. J.; Hearn, M. T. W. (1999). "Characterisation by immobilised metal ion affinity chromatographic procedures of the binding behaviour of several synthetic peptides designed to have high affinity for Cu(II) ions". *Journal of Chromatography A* 852(1):261-272.
- Krueger, J. K.; Kulke, M. H.; Schutt, C.; Stock, J. (1989). "Protein inclusion body formation and purification". *BioPharm Journal* 2(3):40, 42-45.
- Kuusinen, A.; Arvola, M.; Okerblom, C.; Keinanen, K. (1995). "Purification of Recombinant Glur-D Glutamate-Receptor Produced in Sf21 Insect Cells". *European Journal of Biochemistry* 233(3):720-726.
- Ladisch, M. R. (2001). *Bioseparations engineering : principles, practice, and economics*. New York: Wiley.
- LaVallie, E. R., McCoy, J. M. (1995). "Gene fusion expression systems in *Escherichia coli*". *Current Opinion in Biotechnology* 6(5):501-506.
- Le, K. D., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr., Saddler, J. N., Warren, R. A. (1994). "A streptavidin-cellulose-binding domain fusion protein that binds biotinylated proteins to cellulose". *Enzyme and Microbial Technology* 16(6):496-500.
- Lee, W.-C.; Hsiao, C.-C.; Ruaan, R.-C. (1995). "Affinity chromatography of glucose-specific lectin using silica-based support". *Journal of Chemical Technology and Biotechnology* 64(1):66-72.
- Levine, S. P., Wohl, H. (1976). "Human platelet factor 4: Purification and characterization by affinity chromatography. Purification of human platelet factor 4". *Journal of Biological Chemistry* 251(2):324-328.
- Libby, R. T.; Cosman, D.; Cooney, M. K.; Merriam, J. E.; March, C. J.; Hopp, T. P. (1988). "Human Rhinovirus 3C Protease - Cloning and Expression of an Active Form in *Escherichia-Coli*". *Biochemistry* 27(17):6262-6268.
- Lindner, P.; Guth, B.; Wuelfing, C.; Krebber, C.; Steipe, B.; Mueller, F.; Plueckthun, A. (1992). "Purification of native proteins from the cytoplasm and periplasm of *Escherichia coli* using IMAC and histidine tails: A comparison of proteins and protocols". *Methods* 4(1):41-56.

- Lo, K. M.; Sudo, Y.; Chen, J.; Li, Y.; Lan, Y.; Kong, S. M.; Chen, L. L.; An, Q.; Gillies, S. D. (1998). "High level expression and secretion of Fc-X fusion proteins in mammalian cells". *Protein Engineering* 11(6):495-500.
- Lowe, C. R.; Lowe, A. R.; Gupta, G. (2001). "New developments in affinity chromatography with potential application in the production of biopharmaceuticals". *Journal of Biochemical and Biophysical Methods* 49(1-3):561-574.
- Lu, T.; Vandyke, M.; Sawadogo, M. (1993). "Protein-Protein Interaction Studies Using Immobilized Oligohistidine Fusion Proteins". *Analytical Biochemistry* 213(2):318-322.
- Lutfreyermuth, C.; Query, C. C.; Keene, J. D. (1990). "Quantitative-Determination That One of 2 Potential Rna-Binding Domains of the α -Protein Component of the U1 Small Nuclear Ribonucleoprotein Complex Binds with High-Affinity to Stem Loop-II of U1 Rna". *Proceedings of the National Academy of Sciences of the United States of America* 87(16):6393-6397.
- Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in *Escherichia coli*". *Microbiological Reviews* 60(3):512-538.
- Marino, M. H. (1989). "Expression systems for heterologous protein production". *BioPharm Journal* 2(7):18-33.
- Martin, G. A.; Viskochil, D.; Bollag, G.; McCabe, P. C.; Crosier, W. J.; Haubruck, H.; Conroy, L.; Clark, R.; O'Connell, P.; et al. (1990). "The GAP-related domain of the neurofibromatosis type 1 gene product interacts with ras p21". *Cell* 63(4):843-849.
- Mathys, S.; Evans, T. C., Jr.; Chute, I. C.; Wu, H.; Chong, S.; Benner, J.; Liu, X.-Q.; Xu, M.-Q. (1999). "Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation". *Gene* 231(1-2):1-13.
- McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., Kilburn, D. G. (2000). "Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues". *Protein Engineering* 13(11):801-809.
- Mcpherson, D. T.; Morrow, C.; Minehan, D. S.; Wu, J. G.; Hunter, E.; Urry, D. W. (1992). "Production and Purification of a Recombinant Elastomeric Polypeptide, G-(Vpvg)₁₉-Vpvg, from *Escherichia-Coli*". *Biotechnology Progress* 8(4):347-352.

- Moks, T.; Abrahmsen, L.; Holmgren, E.; Bilich, M.; Olsson, A.; Uhlen, M.; Pohl, G.; Sterky, C.; Hultberg, H.; Josephson, S.; Holmgren, A.; Jornvall, H.; Nilsson, B. (1987). "Expression of Human Insulin-Like Growth Factor-I in Bacteria - Use of Optimized Gene Fusion Vectors to Facilitate Protein-Purification". *Biochemistry* 26(17):5239-5244.
- Mulcahy, P.; O'Flaherty, M.; Jennings, L.; Griffin, T. (2002). "Application of kinetic-based biospecific affinity chromatographic systems to ATP-dependent enzymes: studies with yeast hexokinase". *Analytical Biochemistry* 309(2):279-292.
- Munro, S.; Pelham, H. R. B. (1986). "An Hsp70-Like Protein in the Er - Identity with the 78 Kd Glucose-Regulated Protein and Immunoglobulin Heavy-Chain Binding-Protein". *Cell* 46(2):291-300.
- Murby, M.; Cedergren, L.; Nilsson, J.; Nygren, P. A.; Hammarberg, B.; Nilsson, B.; Enfors, S. O.; Uhlen, M. (1991). "Stabilization of recombinant proteins from proteolytic degradation in Escherichia coli using a dual affinity fusion strategy". *Biotechnology and Applied Biochemistry* 14(3):336-346.
- Narayanan, S. R. (1994). "Preparative Affinity-Chromatography of Proteins". *Journal of Chromatography A* 658(2):237-258.
- Nieba, L.; NiebaAxmann, S. E.; Persson, A.; Hamalainen, M.; Edebratt, F.; Hansson, A.; Lidholm, J.; Magnusson, K.; Karlsson, A. F.; Pluckthun, A. (1997). "BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip". *Analytical Biochemistry* 252(2):217-228.
- Nilsson, B.; Abrahmsen, L. (1990). "Fusions to staphylococcal protein A". *Methods in Enzymology* 185(Gene Expression Technology):144-161.
- Nilsson, J.; Jonasson, P.; Samuelsson, E.; Stahl, S.; Uhlen, M. (1996). "Integrated production of human insulin and its C-peptide". *Journal of Biotechnology* 48(3):241-250.
- Nilsson, J.; Stahl, S.; Lundeberg, J.; Uhlen, M.; Nygren, P.-A. (1997). "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins". *Protein Expression and Purification* 11(1):1-16.
- Notenboom, V.; Boraston, A. B.; Kilburn, D. G.; Rose, D. R. (2001). "Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms". *Biochemistry* 40(21):6248-6256.
- Nygren, P. A.; Eliasson, M.; Abrahmsen, L.; Uhlen, M.; Palmcrantz, E. (1988). "Analysis and use of the serum albumin binding domains of streptococcal protein G". *Journal of Molecular Recognition* 1(2):69-74.

- Nygren, P. A.; Stahl, S.; Uhlen, M. (1994). "Engineering Proteins to Facilitate Bioprocessing". *Trends in Biotechnology* 12(5):184-188.
- Olah, Z.; Lehel, C.; Jakab, G.; Anderson, W. B. (1994). "A Cloning and Epsilon-Epitope-Tagging Insert for the Expression of Polymerase Chain Reaction-Generated Cdna Fragments in Escherichia-Coli and Mammalian-Cells". *Analytical Biochemistry* 221(1):94-102.
- Olejnik, J.; Krzymanska-Olejnik, E.; Rothschild, K. J. (1998). "Photocleavable affinity tags for isolation and detection of biomolecules". *Methods in Enzymology* 291(Caged Compounds):135-154.
- Oliaro, J.; Johnson, R. D.; Chen, W. X.; Chadwick, V. S.; Murray, A. (2000). "Identification of an immunogenic 18-kDa protein of Helicobacter pylori by alkaline phosphatase gene fusions". *Journal of Medical Microbiology* 49(7):643-650.
- Ong, E., Alimonti, J. B., Greenwood, J. M., Miller, R. C., Jr., Warren, R. A., Kilburn, D. G. (1995). "Purification of human interleukin-2 using the cellulose-binding domain of a prokaryotic cellulase". *Bioseparation* 5(2):95-104.
- Ong, E.; Gilkes, N. R.; Miller, R. C.; Warren, R. A. J.; Kilburn, D. G. (1993). "The Cellulose-Binding Domain (Cbdcx) of an Exoglucanase from Cellulomonas-Fimi - Production in Escherichia-Coli and Characterization of the Polypeptide". *Biotechnology and Bioengineering* 42(4):401-409.
- Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, A. J., Kilburn, D. G. (1991). "Enzyme immobilization using a cellulose-binding domain: properties of a beta-glucosidase fusion protein". *Enzyme and Microbial Technology* 13(1):59-65.
- Ong, E.; Greenwood, J. M.; Gilkes, N. R.; Kilburn, D. G.; Miller, R. C.; Warren, R. A. (1989). "The Cellulose-Binding Domains of Cellulases - Tools for Biotechnology". *Trends in Biotechnology* 7(9):239-243.
- Ostermann, C., Dickmann, U., Muley, T., Mader, M. (1990). "Large-scale purification of choline acetyltransferase and production of highly specific antisera". *European Journal of Biochemistry* 192(1):215-218.
- Pedersen, J.; Lauritzen, C.; Madsen, M. T.; Dahl, S. W. (1999). "Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases". *Protein Expression and Purification* 15(3):389-400.
- Persson, M.; Bergstrand, M. G.; Bulow, L.; Mosbach, K. (1988). "Enzyme purification by genetically attached polycysteine and polyphenylalanine affinity tails". *Analytical Biochemistry* 172(2):330-337.

- Pfeifer, T. A., Guarna, M. M., Kwan, E. M., Lesnicki, , .;Theilmann, D. A., Grigliatti, T. A., Kilburn, D. G. (2001). "Expression analysis of a modified factor X in stably transformed insect cell lines". *Protein Expression and Purification* 23(2):233-241.
- Porath, J.; Carlsson, J.; Olsson, I.;Belfrage, G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation". *Nature* 258(5536):598-599.
- Prickett, K. S.; Amberg, D. C.;Hopp, T. P. (1989). "A calcium-dependent antibody for identification and purification of recombinant proteins". *Biotechniques* 7(6):580-589.
- Pryor, K. D., Leiting, B. (1997). "High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system". *Protein Expression and Purification* 10(3):309-319.
- Raines, R. T.; McCormick, M.; Van Oosbree, T. R.;Mierendorf, R. C. (2000). "The S tag fusion system for protein purification". *Applications of Chimeric Genes and Hybrid Proteins, Pt A* 326:362-376.
- Rajamohan, F.; Doumbia, S. O.; Engstrom, C. R.; Pendergras, S. L.; Maher, D. L.;Uckun, F. M. (2000). "Expression of biologically active recombinant pokeweed antiviral protein in methylotrophic yeast *Pichia pastoris*". *Protein Expression and Purification* 18(2):193-201.
- Ramirez, C.; Fung, J.; Miller, R. C.; Antony, R.; Warren, J.;Kilburn, D. G. (1993). "A Bifunctional Affinity Linker to Couple Antibodies to Cellulose". *Bio/Technology* 11(13):1570-1573.
- Riggs, P. (2000). "Expression and purification of recombinant proteins by fusion to maltose-binding protein". *Molecular Biotechnology* 15(1):51-63.
- Rubinfeld, B.; Munemitsu, S.; Clark, R.; Conroy, L.; Watt, K.; Crosier, W. J.; McCormick, F.;Polakis, P. (1991). "Molecular cloning of a GTPase activating protein specific for the Krev-1 protein p21rap1". *Cell* 65(6):1033-1042.
- Rudolph, R. (1996). "Successful protein folding on an industrial scale." *Protein Engineering*:283-298.
- Sachdev, D.;Chirgwin, J. M. (1999). "Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein". *Journal of Protein Chemistry* 18(1):127-136.
- Sakka, K., Karita, S., Kimura, T., Ohmiya, K. (1998). "Purification of a fusion protein using the family VI cellulose-binding domain of *Clostridium stercoarium* XynA". *Annals of the New York Academy of Sciences* 864:485-488.

- Samuelsson, E.; Jonasson, P.; Viklund, F.; Nilsson, B.; Uhlen, M. (1996). "Affinity-assisted in vivo folding of a secreted human peptide hormone in *Escherichia coli*". *Nature Biotechnology* 14(6):751-755.
- Samuelsson, E.; Moks, T.; Nilsson, B.; Uhlen, M. (1994). "Enhanced in-Vitro Refolding of Insulin-Like Growth-Factor-I Using a Solubilizing Fusion Partner". *Biochemistry* 33(14):4207-4211.
- Sassenfeld, H. M. (1990). "Engineering Proteins for Purification". *Trends in Biotechnology* 8(4):88-93.
- Sassenfeld, H. M., Brewer, S. J. (1984). "A polypeptide fusion designed for purification of recombinant proteins". *Bio/Technology* 2:76-81.
- Schauer-Vukasinovic, V.; Daunert, S. (1999). "Purification of recombinant proteins based on the interaction between a phenothiazine-derivatized column and a calmodulin fusion tail". *Biotechnology Progress* 15(3):513-516.
- Schmidt, T. G., Koepke, J., Frank, R., Skerra, A. (1996). "Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin". *Journal of Molecular Biology* 255(5):753-766.
- Schmidt, T. G. M.; Skerra, A. (1993). "The Random Peptide Library-Assisted Engineering of a C-Terminal Affinity Peptide, Useful for the Detection and Purification of a Functional Ig Fv Fragment". *Protein Engineering* 6(1):109-122.
- Seidler, A. (1994). "Introduction of a Histidine Tail at the N-Terminus of a Secretory Protein Expressed in *Escherichia-Coli*". *Protein Engineering* 7(10):1277-1280.
- Sharrocks, A. D. (1994). "A T7 Expression Vector for Producing N-Terminal and C-Terminal Fusion Proteins with Glutathione-S-Transferase". *Gene* 138(1-2):105-108.
- Shih, Y. P.; Wu, H. C.; Hu, S. M.; Wang, T. F.; Wang, A. H. J. (2005). "Self-cleavage of fusion protein in vivo using TEV protease to yield native protein". *Protein Science* 14(4):936-941.
- Shimabuku, A. M.; Saeki, T.; Ueda, K.; Komano, T. (1991). "Production of a Site Specifically Cleavable P-Glycoprotein-Beta-Galactosidase Fusion Protein". *Agricultural and Biological Chemistry* 55(4):1075-1080.
- Shoseyov, O.; Doi, R. H. (1990). "Essential 170-Kda Subunit for Degradation of Crystalline Cellulose by *Clostridium-Cellulovorans* Cellulase". *Proceedings of the National Academy of Sciences of the United States of America* 87(6):2192-2195.

- Shpigel, E., Goldlust, A., Efroni, G., Avraham, A., Eshel, A., Dekel, M., Shoseyov, O. (1999). "Immobilization of recombinant heparinase I fused to cellulose-binding domain". *Biotechnology and Bioengineering* 65(1):17-23.
- Shpigel, E., Goldlust, A., Eshel, A., Ber, I. K., Efroni, G., Singer, Y., Levy, I., Dekel, M., Shoseyov, O. (2000). "Expression, purification and applications of staphylococcal protein A fused to cellulose-binding domain". *Biotechnology and Applied Biochemistry* 31 (Pt 3):197-203.
- Skerra, A.; Pfitzinger, I.; Plueckthun, A. (1991). "The functional expression of antibody Fv fragments in *Escherichia coli*: improved vectors and a generally applicable purification technique". *Bio/Technology* 9(3):273-278.
- Smith, D. B.; Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase". *Gene* 67(1):31-40.
- Smith, J. C., Derbyshire, R. B., Cook, E., Dunthorne, L., Viney, J., Brewer, S. J., Sassenfeld, H. M., Bell, L. D. (1984). "Chemical synthesis and cloning of a poly(arginine)-coding gene fragment designed to aid polypeptide purification". *Gene* 32(3):321-327.
- Stahl, S.; Nygren, P. A.; Uhlen, M. (1997a). "Detection and isolation of recombinant proteins based on binding affinity of reporter: protein A". *Methods in Molecular Biology* 63:103-118.
- Stahl, S.; Nygren, P. A.; Uhlen, M. (1997b). "Strategies for gene fusions". *Methods in Molecular Biology* 62:37-54.
- Stevenson, R. (1997). "Affinity interactions: it's more than affinity chromatography". *American Biotechnology Laboratory* 15(10):24, 26, 28.
- Stofko-Hahn, R. E.; Carr, D. W.; Scott, J. D. (1992). "A single step purification for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase". *FEBS Letters* 302(3):274-278.
- Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology* 60(5):523-533.
- Thatcher, D. R. (1990). "Recovery of Therapeutic Proteins from Inclusion-Bodies - Problems and Process Strategies". *Biochemical Society Transactions* 18(2):234-235.
- Thiele, C.; Fahrenholz, F. (1994). "Photocleavable Biotinylated Ligands for Affinity-Chromatography". *Analytical Biochemistry* 218(2):330-337.

- Thomson, A. R. (1996). "Product modification to assist isolation (facilitated processing)". *Downstream Processing of Natural Products*:105-121.
- Tomme, P.; Boraston, A.; McLean, B.; Kormos, J.; Creagh, A. L.; Sturch, K.; Gilkes, N. R.; Haynes, C. A.; Warren, R. A. J.; Kilburn, D. G. (1998). "Characterization and affinity applications of cellulose-binding domains". *Journal of Chromatography B* 715(1):283-296.
- Tomme, P.; Driver, D. P.; Amandoron, E. A.; Miller, R. C.; Antony, R.; Warren, J.; Kilburn, D. G. (1995a). "Comparison of a Fungal (Family-I) and Bacterial (Family-II) Cellulose-Binding Domain". *Journal of Bacteriology* 177(15):4356-4363.
- Tomme, P.; Gilkes, N. R.; Miller, R. C., Jr.; Warren, R. A. J.; Kilburn, D. G. (1994). "An internal cellulose-binding domain mediates adsorption of an engineered bifunctional xylanase/cellulase". *Protein Engineering* 7(1):117-123.
- Tomme, P.; Warren, R. A. J.; Gilkes, N. R. (1995b). "Cellulose hydrolysis by bacteria and fungi". *Advances in Microbial Physiology* 37:1-81.
- Tripet, B.; Wagschal, K.; Lavigne, P.; Mant, C. T.; Hodges, R. S. (2000). "Effects of Side-chain Characteristics on Stability and Oligomerization State of a de Novo-designed Model Coiled-coil: 20 Amino Acid Substitutions in Position 'd'". *Journal of Molecular Biology* 300(2):377-402.
- Tripet, B.; Yu, L.; Bautista, D. L.; Wong, W. Y.; Irvin, R., T.; Hodges, R. S. (1996). "Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins". *Protein Engineering* 9(11):1029-1042.
- Tucker, J.; Grisshammer, R. (1996). "Purification of a rat neurotensin receptor expressed in *Escherichia coli*". *Biochemical Journal* 317:891-899.
- Uhlen, M.; Moks, T. (1990). "Gene fusions for purpose of expression: an introduction". *Methods in Enzymology* 185(Gene Expression Technology):129-143.
- Uhlen, M.; Nilsson, B.; Guss, B.; Lindberg, M.; Gatenbeck, S.; Philipson, L. (1983). "Gene fusion vector based on the gene for staphylococcal protein A". *Gene* 23(3):369-378.
- Vaillancourt, P.; Simcox, T. G.; Zheng, C. F. (1997). "Recovery of polypeptides cleaved from purified calmodulin-binding peptide fusion proteins". *Biotechniques* 22(3):451-453.

- Vaillancourt, P.; Zheng, C. F.; Hoang, D. Q.; Breister, L. (2000). "Affinity purification of recombinant proteins fused to calmodulin or to calmodulin-binding peptides". *Methods in Enzymology* 326:340-362.
- Voss, S.; Skerra, A. (1997). "Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification". *Protein Engineering* 10(8):975-982.
- Wang, L. F.; Yu, M.; White, J. R.; Eaton, B. T. (1996). "BTag: A novel six-residue epitope tag for surveillance and purification of recombinant proteins". *Gene* 169(1):53-58.
- Warren, R. A. J. (1996). "Microbial hydrolysis of polysaccharides". *Annual Review of Microbiology* 50:183-212.
- Wassenberg, D.; Schurig, H.; Liebl, W.; Jaenicke, R. (1997). "Xylanase XynA from the hyperthermophilic bacterium *Thermotoga maritima*: Structure and stability of the recombinant enzyme and its isolated cellulose-binding domain". *Protein Science* 6(8):1718-1726.
- Wells, P. A.; Beiderman, B.; Garlick, R. L.; Lyle, S. B.; Martin, J. P., Jr.; Herberg, J. T.; Meyer, H. F.; Henderson, S. L.; Eckenrode, F. M. (1993). "Large-scale immunoaffinity purification of recombinant soluble human antigen CD4 from *Escherichia coli* cells". *Biotechnology and Applied Biochemistry* 18 (Pt 3):341-357.
- Wilson, D. S.; Keefe, A. D.; Szostak, J. W. (2001). "The use of mRNA display to select high-affinity protein-binding peptides". *Proceedings of the National Academy of Sciences of the United States of America* 98(7):3750-3755.
- Winterhalter, C.; Heinrich, P.; Candussio, A.; Wich, G.; Liebl, W. (1995). "Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*". *Molecular Microbiology* 15(3):431-444.
- Witzgall, R.; O'Leary, E.; Bonventre, J. V. (1994). "A Mammalian Expression Vector for the Expression of Gal4 Fusion Proteins with an Epitope Tag and Histidine Tail". *Analytical Biochemistry* 223(2):291-298.
- Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. (1999). "A genetic system yields self-cleaving inteins for bioseparations". *Nature Biotechnology* 17(9):889-892.
- Wyborski, D. L.; Bauer, J. C.; Zheng, C. F.; Felts, K.; Vaillancourt, P. (1999). "An *Escherichia coli* expression vector that allows recovery of proteins with native N-

termini from purified calmodulin-binding peptide fusions". *Protein Expression and Purification* 16(1):1-10.

Yang, Z.; Hancock, W. S. (2004). "Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column". *Journal of Chromatography, A* 1053(1-2):79-88.

Yarmush, M. L.; Antonsen, K. P.; Sundaram, S.; Yarmush, D. M. (1992). "Immunoabsorption - Strategies for Antigen Elution and Production of Reusable Adsorbents". *Biotechnology Progress* 8(3):168-178.

Yarnall, M.; Boyle, M. D. P. (1986). "Isolation and partial characterization of a type II Fc receptor from a group A Streptococcus". *Molecular and Cellular Biochemistry* 70(1):57-66.

Yet, S. F.; Moon, Y. K.; Sul, H. S. (1995). "Purification and Reconstitution of Murine Mitochondrial Glycerol-3-Phosphate Acyltransferase - Functional Expression in Baculovirus-Infected Insect Cells". *Biochemistry* 34(22):7303-7310.

Zheng, C. F.; Simcox, T.; Xu, L.; Vaillancourt, P. (1997). "A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins". *Gene* 186(1):55-60.

2 Inexpensive One-Step Purification of Polypeptides Expressed in *E. Coli* as Fusions With the Family 9 Carbohydrate-Binding Module of Xylanase 10A from *T. Maritima*

2.1 Introduction

The continued maturation of the pharmaceutical and biotechnology industries has created an increasing need for practical and economical large-scale processing techniques. Production methods such as fed-batch fermentations of recombinant microbes have advanced to a level where target biomolecules can be produced in g/L concentrations at relatively modest cost. As a result, downstream processing often accounts for more than 60% of the total operating cost, and as much as 70% of the capital cost of current biochemical production processes (Ladisich 2001). Purification of a target protein during manufacturing usually requires several chromatographic steps in series due to the relatively non-specific physico-chemical interactions that drive separations in these columns. Although product purities are often quite high, overall yields from multi-step chromatographic procedures are generally low due to the accumulated loss of product (Chase 1984; Hearn and Acosta 2001). The challenge therefore is to reduce costs and increase overall yields by process simplification through elimination or combination of purification steps.

* A version of this chapter is published in the *Journal of Chromatography B*. [Reference: Mojgan Kavooosi, Julia Meijer, Emily Kwan, A. Louise Creagh, Douglas G. Kilburn, Charles A. Haynes, Inexpensive one-step purification of polypeptides expressed in *E. coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritime*. *J. Chromatography B*. 807(1); 87-94; (2004)]

Toward this goal, a number of affinity separation systems have been developed in the past two decades to replace difficult multi-step chromatographic procedures with a highly selective binding step that serves to both purify and concentrate the product (Lowe et al. 2001; Wilchek and Chaiken 2000). Polypeptide fusion tags that selectively bind a complementary ligand immobilized onto a suitable chromatographic matrix are now widely used at the laboratory scale to facilitate recombinant expression and purification of target proteins (Nilsson et al. 1997; Terpe 2003). In addition to allowing rapid purification, affinity fusion tags have been shown in certain cases to increase *in vivo* proteolytic stability of the target protein, improve product solubility, and control product localization in or secretion from the expression host (Jonasson et al. 2002). Skillful engineering of the fusion-tag/immobilized-ligand pair can therefore provide a robust and generic method for efficient production and high-resolution affinity purification of recombinant protein targets.

Commercially available affinity tag systems include the calmodulin binding peptide (Vaillancourt et al. 1997; Zheng et al. 1997), the glutathione S-transferase (GST) tag from *Schistosoma japonicum* (Guan and Dixon 1991; Smith and Johnson 1988), and various polyamino-acid affinity tags such as the polyhistidine tag (Crowe et al. 1994; Hochuli et al. 1987; Porath et al. 1975) and its associated immobilized metal-ion affinity chromatography (IMAC) capture column. Each of these affinity tag systems has been used extensively at the laboratory scale, but only the GST and polyhistidine tag systems have found any appreciable use in manufacturing. More extensive use of affinity tags systems by industry has been thwarted in part by the complex chemical modifications required to cross-link the solid support or to graft the affinity-tag receptor to the resin surface, and by the relatively low tolerance of many of these affinity resins to repeated processing and sanitization-in-place cycles. However, the dominant impediment to the use of current commercially available generic affinity tags in large-scale bioprocessing is cost. A new affinity tag that binds to an inexpensive, chemically and hydrodynamically robust chromatographic resin is therefore highly desirable.

Here, we present a generic and inexpensive affinity purification technology based on high-titer recombinant expression in *E. coli* of fusion proteins containing the

carbohydrate-binding module CBM9 attached to the N-terminus of the target protein or polypeptide. TmXyn10ACBM9-2 (henceforth referred to as CBM9), the C-terminal family 9 carbohydrate-binding module of xylanase 10A from *Thermotoga maritima* (Winterhalter et al. 1995), binds specifically to the reducing ends of cellulose and soluble polysaccharides, a property that is currently unique to this CBM. Measured association constants (K_a) for adsorption of CBM9 to insoluble allomorphs of cellulose are between 2×10^5 to $2 \times 10^6 \text{ M}^{-1}$. CBM9 also binds a range of soluble sugars (Boraston et al. 2001), including glucose, such that a 1-M glucose solution is effective in quantitatively eluting CBM9 and CBM9-tagged fusion proteins from a cellulose-based capture column. The presence of the CBM9 tag therefore allows for affinity capture and purification of a fusion protein on an inexpensive cellulose-based chromatography resin.

A unique processing site is encoded at the C-terminus of the tag to facilitate rapid and quantitative removal of the tag by Factor X_a to recover the desired target protein sequence following affinity purification (Nagai and Thogersen 1984). Validation of the technology is provided by fusing the CBM9 affinity tag to the N-terminus of green fluorescent protein (GFP) from the jellyfish, *Aequorea victoria* (Cramer et al. 1996; Shimomura et al. 1962). The use of GFP as the target protein has the advantage that the natural fluorescence of GFP measured at 509 nm (excitation at 395 nm) offers a direct and convenient means of tracking the target fusion protein throughout the production and affinity purification process.

The generic CBM9 affinity-tag technology proposed here involves five distinct processing steps: (i) recombinant production (cytoplasmic) of the properly folded fusion protein in recombinant *E. coli* BL21 (DE3) cells, (ii) cell lysis and lysate resuspension, (iii) affinity purification (including elution) of the CBM9-tagged fusion protein on a suitable commercial cellulose-based chromatography resin, (iv) cleavage of the CBM9-linker-IEGR affinity tag sequence using immobilized recombinant Factor X_a , and finally, (v) removal of CBM9 to obtain the purified target (GFP). Each of these essential processing steps is evaluated in terms of product yield, purity and concentration factor to provide a measure of the overall performance of the technology.

2.2 Materials and Methods

2.2.1 Reagents

Kanamycin, glucose, and all other chemicals were purchased from Sigma Chemicals (St. Louis, MO, USA). All reagents were analytical grade unless stated otherwise. Restriction enzymes were purchased from New England Biolabs (Beverly, MA). T4-DNA ligase was obtained from Roche Molecular Biochemicals (Laval, Quebec). Perloza™ MT100 and MT500 chromatography resins having a nominal particle diameter distribution of 50-80 µm and 100-250 µm, respectively, were purchased from Iontosorb Inc. (Czech Republic). *E. coli* BL21 (DE3) and Ni²⁺-Sepharose resin were obtained from Novagen (Milwaukee, MI).

2.2.2 Cloning of CBM9-GFP Fusion Protein

All cloning procedures were performed using standard molecular biology techniques (Sambrook et al. 1989). The GFP and CBM9 coding regions were amplified from the vectors pGFPuv (Clontech, Palo Alto, CA) and pETCBM9, respectively. A Bsp HI restriction endonuclease site (underlined) was introduced at the 5' end of CBM9 gene fragment, using the oligonucleotide 5'-TTGCTAGCTTCATGACTAGCGGAATAATG GTAGC-3' as primer. The sequence encoding for the S₃N₁₀LA linker (*italic*) (henceforth referred to as S₃N₁₀ linker) and a Pvu I site (underlined) were introduced at the 3' end of the CBM9 coding region using the oligonucleotide 5'-TCCCTCGATCGCGAGGTTGTTGTTATTGTTATTGTTGTTGTTGT TCGAGCTCGAA GCTTGATGAGCCTGAGGTTACC-3' as primer. For the GFP gene fragment, the sequence encoding for the Factor X_a recognition site (*IEGR*) (*italic*) and a Pvu I restriction endonuclease site (underlined) were placed at the 5' end, using the oligonucleotide 5'-CCGATCGAGGGTCGTATCATGAGTAAAGGAGA-3' as primer. For the 3' end, a Not I site (underlined) was introduced using the oligonucleotide 5'-TGCGGCCGCTTTGTAGAGCTCATCCATGCCATGTGTAATCCC-3' as primer. Each PCR mixture (50 µL total volume) contained 50 ng of template, 30 pmol of each primer, 5% DMSO, 0.4 mM 2'-deoxy-nucleoside 5'-triphosphates, and 1 unit of *PWO* DNA

polymerase in buffer (Roche Molecular Biochemicals, Laval, Quebec). The following protocol for 25 successive PCR cycles was followed: denaturation at 94°C for 30 s, annealing for 2 min by linearly increasing the temperature from 55°C to 72°C, and primer extension at 72°C for 45 s. The resulting CBM9-S₃N₁₀ and FXa-GFP coding regions were digested with Bsp HI/Pvu I and Pvu I/Not I, respectively, and ligated (16°C, 16 h) into the pET28b vector (Novagen) previously digested with Nco I and Not I to give the appropriate pET28-CBM9-S₃N₁₀-IEGR-GFP construct (hereafter referred to as pET28-CBM9-GFP). DNA sequencing was then completed to verify the construct (NAPS Unit, Biotechnology Laboratory, University of British Columbia).

2.2.3 Protein Production

Overnight cultures of *E. coli* strain BL21/pET28-CBM9-GFP were diluted 100-fold in tryptone-yeast extract-phosphate medium (TYP) supplemented with 50 µg/mL of kanamycin and grown at 37°C to a cell density (OD_{600 nm}) of ~ 1.0 absorbance units. Isopropyl-1-thio-β-D-galactoside (IPTG) was added to a final concentration of 0.3 mM. Incubation was then continued at 30°C for a further 10-12 h. The cells were harvested by centrifugation (8500g) at 4°C for 20 min and then resuspended in high salt buffer (1 M NaCl, 50 mM potassium phosphate, pH 7.0) by gentle mixing. Cells were ruptured by two passages through a French pressure cell (21000 lb/in²) and the cell debris removed by centrifugation for 30 min at 27000 g and 4°C. CBM9-GFP fusion protein was purified by affinity chromatography as described below.

The stability of the fusion protein against proteolysis was assayed as follows. At 18 hours post-induction, the culture was divided into two equal volumes and cells were harvested as described above. The cells in one container were resuspended in high salt buffer while the cells in the second container were resuspended in high salt buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma). Cells were disrupted (in the presence of fresh 1 mM PMSF or buffer) and the clarified cell extract (1.5 mL) incubated with 200 µL of Perloza™ MT100 (94.5 mg dry weight/mL). The mixture was mixed at room temperature for 3.5 hours by rotating end-over-end. The resin was collected by centrifugation at 8500g for 8 min, washed three times with 1 mL high salt

buffer, 2X with low salt buffer (50 mM potassium phosphate pH 7.0) and 1X with TBS8 (15 mM NaCl, 10mM TrisHCl, pH 8.0). The protein was desorbed with 400 μ L of 1-M glucose in TBS8 and analyzed by SDS-PAGE.

2.2.4 Affinity Chromatography

A Pharmacia XK-16 column (8.5 cm X 1.6 cm I.D.) was packed by standard inclined pouring with Perloza™ MT100 resin to give a final bed volume of *ca.* 17 mL. All purification chromatograms were completed on a Pharmacia P-500 FPLC system (Amersham Biosciences) at 4°C and with a flow rate of 0.5 ml/min. The Perloza™ MT100 affinity column was equilibrated with ~ 10 column volumes (CV) of high salt buffer. Clarified cell extract (50 mL) was loaded onto the column and unbound protein was removed by washing the column with 10 CV of high salt buffer, 5 CV of low salt buffer and 4 CV of TBS8 buffer. CBM9-GFP was then eluted from the column with 5 CV of 1-M glucose in TBS8. Eluted protein fractions were analyzed for purity by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 20% SDS sample buffer. Column regeneration was completed using 10 CV water followed by 10 CV of high salt buffer for equilibration.

2.2.5 Tag Cleavage by Factor X_a

In this experiment carried out at 21°C and pH 7, a purified chimeric protein comprised of Factor X_a fused to CBM2a, the family 2a carbohydrate binding module (CBM2a) of xylanase 10A of the soil bacterium *Cellulomonas fimi*, was immobilized onto a Perloza™ MT500 column (henceforth called CBM2a-FX_{a_{im}}) (Kwan et al. 2002) and used to enzymatically remove the CBM9 affinity tag. Pure CBM9-GFP fractions were pooled and incubated with CBM2a-FX_{a_{im}} at 21°C, rotating end-over-end. After 108 hours, CBM2a-FX_{a_{im}} was removed by centrifugation (8500g, 15 min) and washed extensively to collect all cleaved product. The cleaved products were buffer exchanged into low salt buffer and concentrated in a stirred ultrafiltration (UF) unit (Amicon, Beverly, MA) on a 1 KDa cutoff filter (Filtron, Northborough, MA). The concentrated protein solution was applied to a column (24 cm X 0.9cm I.D.) packed with Perloza™

MT100 and washed with 15 CV of low salt buffer. Free GFP was collected in the flow through and concentrated by UF.

The processing time required for Factor X_a cleavage of the IEGR-terminal affinity tag was also assayed. 3 mg of purified CBM9-GFP fusion protein was incubated with 3 μ L CBM2a-FX_{a_{im}} (Kwan et al. 2002) at 21°C (final [Factor X_a] to [fusion protein] ratio of 1:1000). The control experiments contained buffer in place of CBM2a-FX_{a_{im}}. Samples were taken at the following time points (0, 0.5, 1, 2, 4, 6, 8, 10, 12, 14, 25.5, and 35 hours post-incubation) and analyzed by SDS-PAGE.

2.2.6 Fluorescence Calibration Curves

Fluorescence measurements were used to quantify target protein (CBM9-GFP) concentrations in buffer and in complex cell lysate feed stocks. Highly pure CBM9-GFP was obtained by sequentially purifying CBM9-GFP on the Perloza™ MT100 affinity column followed by immobilized-metal affinity chromatography (IMAC) on a Ni⁺²-Sephacrose IMAC resin (according to manufacturer's instructions). This highly pure protein was buffer exchanged into low salt buffer and concentrated as described above. Concentrations of the purified protein were determined by UV absorbance (280 nm) using a calculated molar extinction coefficient of 62870 cm⁻¹ M⁻¹ (Mach et al. 1992).

Varying concentrations of highly pure protein were mixed with either loading buffer (TBS8), elution buffer (1 M glucose in TBS8) or BL21 cell extract ($A_{280} = 6$ or $A_{280} = 0.6$) and the fluorescence measured (395 nm for excitation; 509 nm for emission) using a Cary Eclipse fluorescence spectrophotometer (Varian, Palo Alto, CA). Linear calibration curves (0 to 0.35 μ M CBM9-GFP) for measured fluorescence as a function of CBM9-GFP concentration in buffer and in BL21 cell extract were constructed from each data set.

2.2.7 Measurement of Binding Isotherms

Samples containing purified fusion protein at concentrations ranging from 1 to 30 μ M were incubated with resin (1 mg (dry weight) of Perloza™ MT100, 5 mg CF31, 5 mg

CF1 and 5 mg Avicel) in high salt buffer to a final volume of 1 ml. The samples were then incubated for 30 hours at 4°C (25°C for Avicel samples) while mixing end-over-end. The cellulose was removed by centrifugation at 27000g for 16 min at 4°C. The supernatant was collected and the concentration of unbound protein was determined by UV absorbance (280 nm) using a Cary 100 Spectrophotometer (Varian).

An isotherm was generated by plotting the concentration of bound protein ($\mu\text{mol/g}$ of resin) against the concentration of unbound protein (μM). The binding parameters were determined by a non-linear fitting of the Langmuir-type adsorption isotherm equation to the experimental data using GraphPad Prism 3.0 software.

Binding isotherms were also measured for CBM9-GFP in the presence of bacterial cell extract (A_{280} of 5.3). Samples were incubated for 16 hours at 4°C while mixing end-over-end. The cellulose was removed by centrifugation as described above and the fluorescence of the supernatant was measured. The concentration of unbound protein in the supernatant was determined from the calibration curve. Bound CBM9-GFP concentrations were then computed by mass balance.

2.3 Results and Discussion

Figure 2.1 shows a block diagram for the CBM9-GFP fusion protein construct used in this work to validate the utility and performance of our CBM9 tag technology for inexpensive affinity purification of recombinant proteins and peptides in *E. coli*. In the pET28-CBM9-GFP vector, the coding sequence for the N-terminal CBM9 is followed by the gene fragment encoding an S_3N_{10} linker that serves to separate the CBM9 fusion tag from the target protein. The synthetic S_3N_{10} linker was used in this study because it has proven useful in our laboratory in the stable expression of a range of fusion proteins. The combined CBM9- S_3N_{10} fusion tag is separated from the n-terminal amino acid of GFP by the four amino acid IEGR processing site for the endoprotease recombinant human Factor X_a . The presence of the IEGR processing site allows Factor X_a catalyzed removal of the affinity tag following fusion protein purification to recover the pure target protein with its natural N-terminus.

2.3.1 Binding Isotherms and Thermodynamics

Equilibrium adsorption isotherms at 4°C for CBM9-GFP binding to the porous cellulose-based chromatography resin Perloza™ MT100 are shown in Figure 2.2. As shown in Table 2.1, Perloza™ MT100 stationary phase resin binds pure CBM9-GFP with a capacity of 9.93(±0.31) μmol/g MT100. For the 53 kg mol⁻¹ CBM9-GFP fusion protein, this equates to a saturation loading capacity of 527 mg protein bound per gram of resin, or *ca.* 115 mg/mL of column. In pure buffer at 4°C, CBM9-GFP binds Perloza™ MT100 with an affinity of 1.5(±0.21) x 10⁶ M⁻¹.

As shown in Table 2.2, CBM9-GFP also binds to a number of other commercially available cellulose-based resins. However, in each case, the resin capacity (and to a lesser extent the binding affinity) is significantly lower than observed for binding to Perloza™ MT100, indicating that the Perloza™ resin offers a relatively high concentration of entropically unhindered reducing ends for CBM9 binding. CBM9 binding to cellulose is exothermic (Boraston et al. 2001). Thus, the affinity characterizing binding to Avicel will increase with decreasing temperature. However, q_i^{\max} for this resin is appreciably lower than for Perloza™ MT100, making it a less desirable matrix for affinity chromatography applications.

The equilibrium adsorption isotherm at 4°C for binding of pure CBM9 to Perloza™ MT100 is also shown (Figure 2.2, Table 2.1). The binding properties (K_a and q_i^{\max}) of the isolated fusion tag (CBM9) are similar to those of the fusion protein (CBM9-GFP), indicating that the presence of the target protein does not significantly affect the performance of the CBM9 affinity tag.

The intrinsic fluorescence of GFP allowed us to also measure the binding isotherm for the CBM9-GFP fusion protein in the presence of the *E. coli* cell lysate from which it is purified. Although errors in CBM9-GFP fluorescence measurements are large when cell lysate components are present in the solution phase, the data suggest that neither the binding capacity of the resin nor the affinity of the CBM9-GFP fusion protein for the resin is significantly altered by the presence of a large concentration of

contaminant proteins (data not shown), indicating the specificity of the Perloza™ MT100 cellulose-based resin for CBM-tagged proteins.

2.3.2 Fusion Protein Expression and Stability

Unoptimized batch fermentation yields of soluble CBM9-GFP in recombinant *E. coli* BL21 cultures were typically around 210 mg/L of culture, which represents a 40% increase in GFP yield over more standard expression systems (Chalfie and Kain 1998; Chalfie et al. 1994). The tendency for CBM fusion tags, including the more commonly used maltose binding protein, to increase soluble expression of otherwise low expressing proteins is well documented (Fox et al. 2001; Fox and Waugh 2003). This ability to enhance titers of soluble protein is likely due, at least in part, to the relatively high solubility of CBMs, which allows them to serve as effective solubilizing agents for aggregation-prone polypeptides. In certain cases, fusion to a CBM can also promote the proper folding of the attached protein into its biologically active conformation. This chaperone-like quality distinguishes CBMs such as CBM9 and MBP from other affinity tags and greatly enhances their value as a fusion partner.

The performance of a fusion tag technology depends not only on the properties of the tag, but also on the stability of the amino-acid sequence that links the tag to the target protein. Spiking and incubation of purified CBM9 in an *E. coli* BL21 culture lysate resulted in no detectable degradation of the CBM during its purification as measured by SDS PAGE. The stability of the S₃N₁₀IEGR linker against degradation by endogenous *E. coli* proteases present in the cytoplasm and cell lysate was therefore analyzed by SDS PAGE following cell disruption and lysate clarification, either in the presence or absence of the protease inhibitor PMSF. As the CBM is not degraded significantly by endogenous proteases, proteolytic degradation within the linker results in the appearance of a band on an SDS PAGE gel corresponding to (or close to) the molecular mass of CBM9. As shown in Figure 2.3, when PMSF is added to the washed cell suspension, a very small amount of proteolytic degradation of the S₃N₁₀ linker occurs, either *in vivo* or during the cell processing and affinity purification steps. In the absence of a protease inhibitor, a slightly larger fraction of the recombinantly expressed CBM9-GFP fusion

protein is lost due to degradation within the linker region. Under both processing conditions, however, the vast majority of expressed fusion protein remains intact through the induction, cell lysis and affinity chromatography steps.

2.3.3 Affinity Purification on Perloza™ MT100 Column

A typical chromatogram for affinity purification on a Perloza™ MT100 capture column of CBM9-GFP from an *E. coli* BL21 clarified cell lysate is shown in Figure 2.4. No protease inhibitor (PMSF) was added to the cell suspension or lysate. The corresponding SDS PAGE gel documentation of the purification process is shown in Figure 2.5, and a summary of the fusion protein yield, purity, and concentration factor following elution from the Perloza™ MT100 column is provided in Table 2.3.

The intrinsic fluorescence of GFP allows us to monitor simultaneously the elution of contaminating proteins (UV absorbance @ 280 nm) and the concentration (fluorescence intensity @ 509 nm) of CBM9-GFP and its degradation products in each elution fraction. A small amount of CBM9-GFP or GFP within the clarified lysate load is lost in the column flowthrough. It is likely that most if not all of this fluorescent material represents the small amount of fusion protein that is degraded within the S₃N₁₀ linker region, as shown in Figure 2.3. Weakly bound contaminating proteins are sequentially removed in the column flowthrough and the two column wash steps. No loss of CBM9-GFP is detected in either wash step (Figure 2.5, lanes 3 and 4).

A 1-M glucose solution (in TBS8) is effective in quantitatively eluting all specifically bound fusion protein (Figure 2.5, lane 5). CBM9-GFP elutes from the column in a single sharp peak, as is evident from the overlapping A₂₈₀ and fluorescence intensity peaks in the chromatogram. The purity of CBM9-GFP in the pooled fractions of the elution peak was greater than 95% at an average yield of 86%. Both values are competitive with (in fact superior to) the published performance of other commercially available affinity tag systems, including the GST and poly-His fusion-tag technologies (Appa Rao et al. 1997; Baek et al. 1997; Chang et al. 1999; Littlejohn et al. 2000; Modesti et al. 1995).

In these experiments, the Perloza™ MT100 column was loaded to less than half saturation capacity to guarantee capture of all CBM9-GFP from the clarified cell lysate. Despite operating the column at under-loaded conditions, a remarkably high concentration factor of *ca.* 46 was achieved, indicating that the fusion protein loads, binds and elutes from the column in a reasonably tight band.

2.3.4 Column Reusability

As the cost of any affinity chromatography technology is largely determined by the purchase price and reusability of the capture resin, we investigated the ability of the Perloza™ MT100 resin to provide acceptable and predictable purification performance with repeated column use. Six consecutive purifications were performed on a single Perloza™ MT100 column to identify any changes in column performance with increasing number of purification cycles. Very high product purity (>95%) was achieved in all six purification cycles. As shown in Table 2.4, product yield and concentration factor, however, were affected by repeated column use. An average yield of 86(±3.6)% was observed for the affinity purification of CBM9-GFP from clarified cell lysate on a clean, freshly poured Perloza™ MT100 column. Slightly lower yields of *ca.* 79% were then consistently observed for each purification cycle thereafter. The product concentration factor followed the same trend, with a measured concentration factor of 46(±9.9) for the first column cycle falling to a consistent value of *ca.* 28 for each subsequent cycle. The source of these modest changes is unclear. However, the repeatable good performance (> 95% purity, 79% yield, concentration factor of 28) of the column following the first column cycle suggests that our CBM9 fusion tag technology can provide a robust platform for affinity purification of recombinant proteins.

2.3.5 Removal of the CBM9-S₃N₁₀-IEGR Affinity Tag Using an Immobilized Factor X_a Column

In certain cases, such as in the production of a human therapeutic protein, removal of the fusion tag following purification is required to recover the desired target protein with its natural N-terminus. We therefore have incorporated a Factor X_a processing site adjacent to the N-terminus of the target protein to facilitate tag removal by specific

enzymatic cleavage. Figure 2.6 is an SDS PAGE gel showing the kinetics of tag cleavage when purified CBM9-GFP is processed at 21 °C and pH 8 with CBM2a-FXa_{im} at a fusion-protein to Factor X_a concentration ratio of 1000 to 1. To avoid stagnant settling of the CBM2a-FXa_{im} Perloza™ MT500 resin, the reaction mixture, which also contained 1-M glucose in the liquid phase, was mixed end-over-end in an orbital mixer. In the presence of 1-M glucose, CBM9 does not bind to Perloza™ MT500, while binding of CBM2a-FXa is irreversible at these conditions. Complete cleavage of the CBM9-S₃N₁₀-IEGR fusion tag was observed after 28 h.

The Factor X_a treated solution was then diafiltered on a 1 K cut-off filter to remove the 1-M glucose and loaded onto a second Perloza™ MT100 column to capture the cleaved CBM9 tag. Pure, N-terminally correct GFP was collected in the flow through with a yield of 98% and a purity of greater than 95%. This resulted in an overall yield of the purified target protein (GFP) of 84% when a fresh Perloza™ MT100 column was used, or 77% when the same column was used for multiple purification cycles.

2.4 Conclusions

We have shown that proteins expressed in *E. coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima* can be affinity purified on a cellulose-based Perloza™ MT100 column. The performance of our technology is competitive with all commercial fusion tag systems, and may offer advantages with respect to improving the expression of the target protein in a soluble form.

Acceptance and use of affinity tag systems in manufacturing of recombinant proteins have been slowed, at least in part, by the associated costs of the technology, particularly the cost of the resin. Perloza™ MT100 is a simple, highly porous regenerated cellulose/cellulose xanthate of uniform particle size and flow characteristics. The polymer bead structure is stabilized by hydrogen bonds only; there are no covalent cross-links within the resin. As a result, it is a durable and surprisingly inexpensive resin. When bought in bulk quantities, the cost of Perloza™ MT100 is *ca.* \$35 U.S. per liter of resin, which, for example, is close to 1/100 the cost of an equivalent volume of

glutathione affinity resin used to purify GST tagged proteins. The cost of Perloza™ MT100 resin also compares very favorably with the costs of those resins designed to capture fusion proteins tagged with GST or calmodulin binding protein. Direct capture on a packed column of Perloza™ MT100, which binds CBM9-tagged proteins with extraordinarily high capacity (in excess of 500 mg/g resin), therefore appears to offer a robust and inexpensive strategy for affinity purification of proteins expressed in soluble form as fusions with the CBM9 tag.

2.5 Tables

Table 2.1 Langmuir adsorption parameters (equilibrium association constant K_a and binding capacity q_i^{\max}) for binding of CBM9 and CBM9-GFP to Perloza™ MT100 at 4°C. Solvent contains pure protein in high-salt buffer.

Protein	K_a (M^{-1})	q_i^{\max} ($\mu\text{mol protein/g resin}$)
CBM9-GFP	$1.5 (\pm 0.21)^* \times 10^6$	9.93 (± 0.31)
CBM9	$1.2 (\pm 0.06) \times 10^6$	11.20 (± 0.15)

* Reported errors represent 2σ (i.e. 95% confidence interval)

Table 2.2 Binding affinity and capacity of CBM9-GFP on various cellulosic resins

Resin	K_a (M^{-1})	q_i^{\max} ($\mu\text{mol protein/g resin}$)
Perloza™ MT100 ^a	$1.5 (\pm 0.21)^* \times 10^6$	9.93 (± 0.31)
CF1 ^a	$2.5 (\pm 0.11) \times 10^5$	0.12 (± 0.01)
CF31 ^a	$4.5 (\pm 0.19) \times 10^5$	0.30 (± 0.03)
Avicel ^b	$4.1 (\pm 0.68) \times 10^5$	0.52 (± 0.02)

Reported errors as in Table 2.1

a: Binding performed at 4°C in high-salt buffer

b: Binding performed at 25°C in high-salt buffer

Table 2.3 Summary of purification of CBM9-GFP on Perloza™ MT100 at 4°C

	Protein ¹ (mg)	Yield	Purity ²	Concentration factor
Cell extract	71.5	100%		
Elution	61.5	86 (±3.6)%	> 95%	45.7 (±9.9)
Free GFP (after tag removal)	60.2	84%	> 95%	

1 Protein concentration was quantified by fluorescence ($\lambda_{\text{ex}} = 395 \text{ nm}$, $\lambda_{\text{em}} = 510 \text{ nm}$)

2 Purity determined by SDS-PAGE

Table 2.4 CBM9-GFP yield and purity for consecutive purification runs through the same Perloza™ MT100 column

Column cycle ¹	Yield ²	Concentration factor
1	86 (±1.7) %	45.7
2	78.4 (±1.6) %	28.8
3	81.2 (±1.6) %	29.2
4	79 (±1.6) %	28.8
5	78.8 (±1.6) %	26.8
6	78 (±1.6) %	27.5

1 All runs gave a CBM9-GFP purity of > 95% as determined by SDS-PAGE

2 Protein concentrations were quantified by fluorescence as noted in Table 2.3

2.6 Figures



Figure 2.1 Schematic representation of gene fragment coding for the CBM9-S₃N₁₀-IEGR-GFP fusion protein

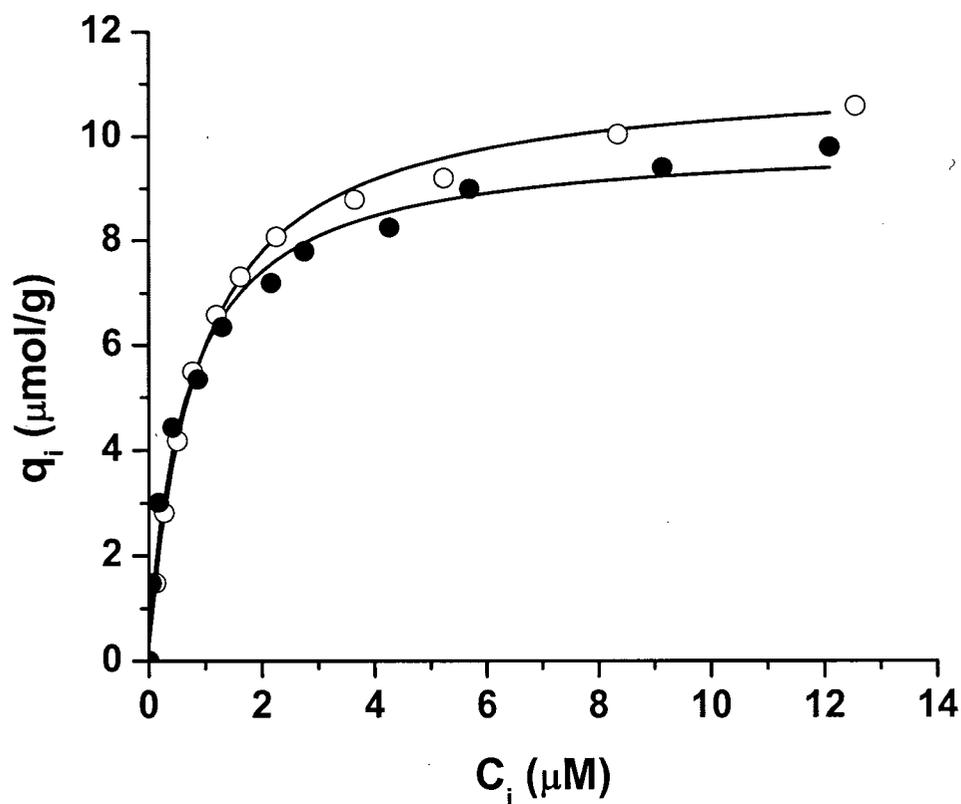


Figure 2.2 Equilibrium adsorption isotherms for binding of CBM9 and CBM9-GFP to Perloza™ MT100 at 4°C. CBM9-GFP binding to Perloza™ MT100 at 4°C in high salt buffer (solid circle). CBM9 binding at 4°C in high-salt buffer (open circle), where q_i is the bound protein concentration and C_i is the equilibrium concentration of protein free in solution.

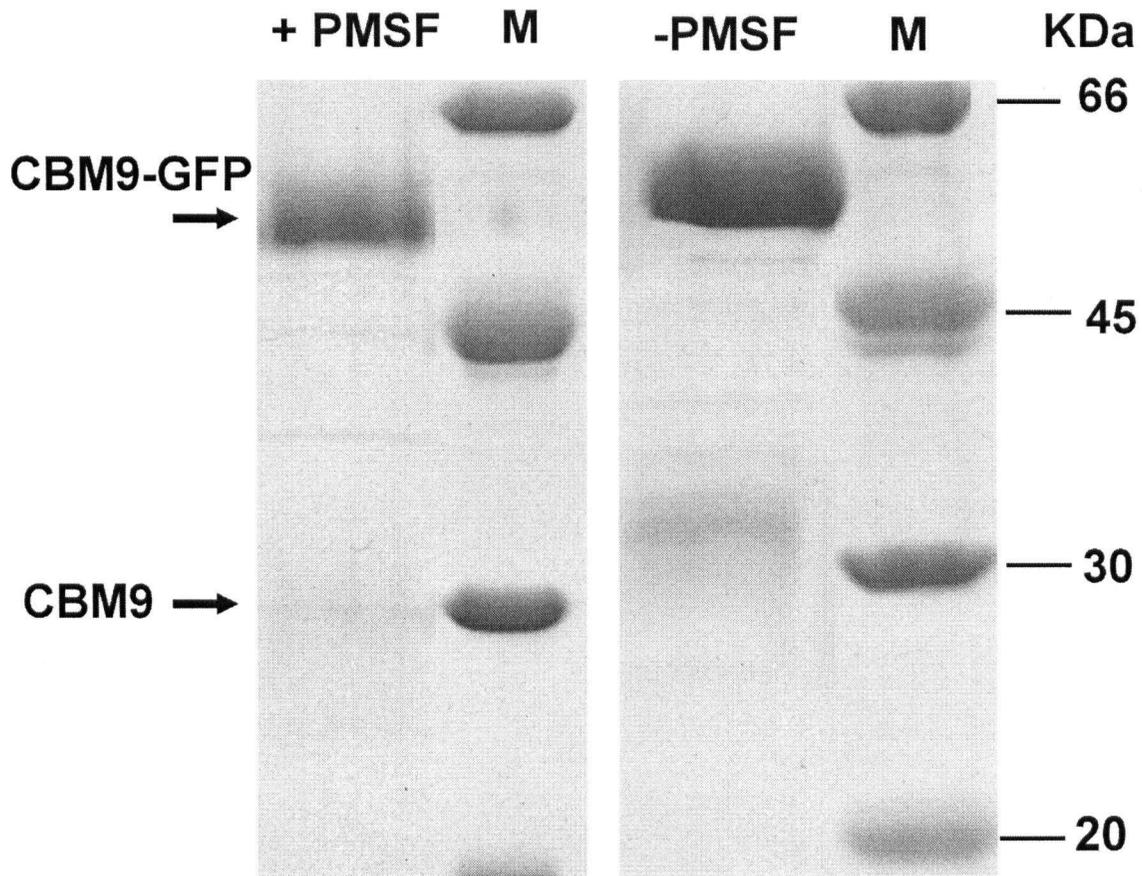


Figure 2.3 Proteolytic stability of the S3N10 linker in CBM9-GFP. 12% SDS-PAGE of CBM9-GFP purified with Perloza™ MT100 in a small batch system. PMSF treated cell extract containing CBM9-GFP was mixed end-over-end, washed with buffer and desorbed with 1-M glucose in TBS8. Despite its molecular weight of 22 kDa, CBM9 runs as a 26-31 kDa protein depending on the length of the linker fragment attached to it (Wassenberg et al. 1997)

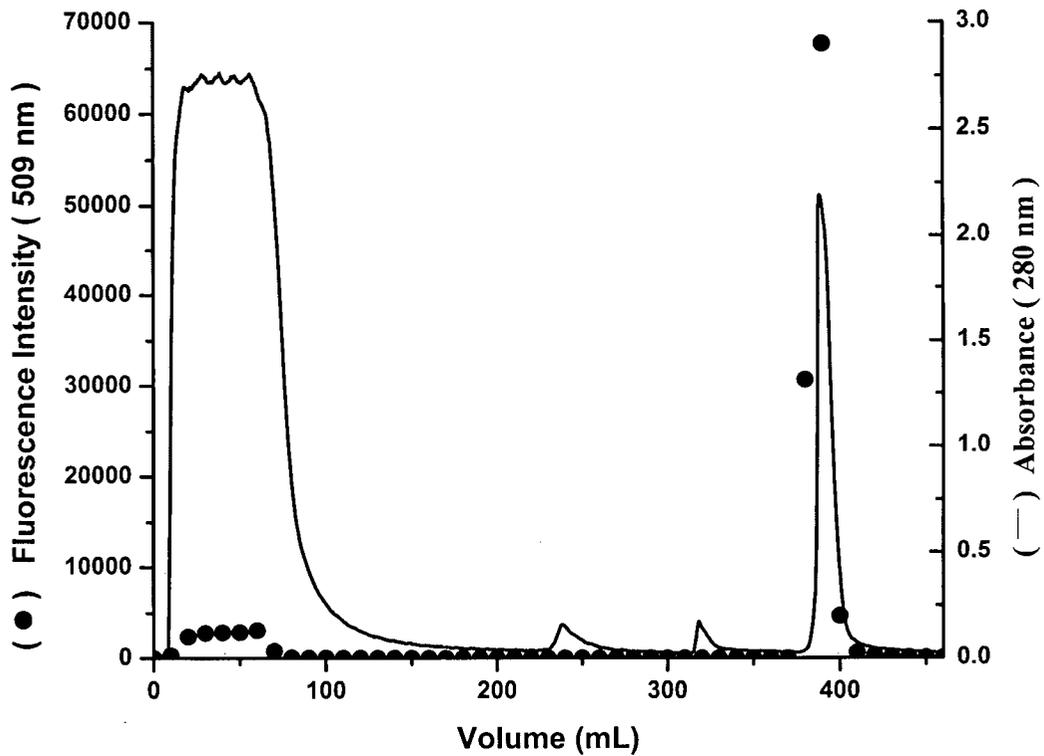


Figure 2.4 Chromatogram of CBM9-GFP purification from an *E. coli* BL21 clarified cell lysate on Perloza™ MT100 at 4°C. 50 mL of clarified cell extract was loaded at 0.2 mL min⁻¹ on a 17 mL column packed with Perloza™ MT100 resin, and then washed with 10 column volumes (CV) high salt buffer and 5 CV low salt buffer. Bound fusion protein was desorbed with 1 M glucose in TBS8. 10 mL fractions were collected and analyze by fluorescence (509 nm) (solid circle) and absorbance at 280 nm (line).

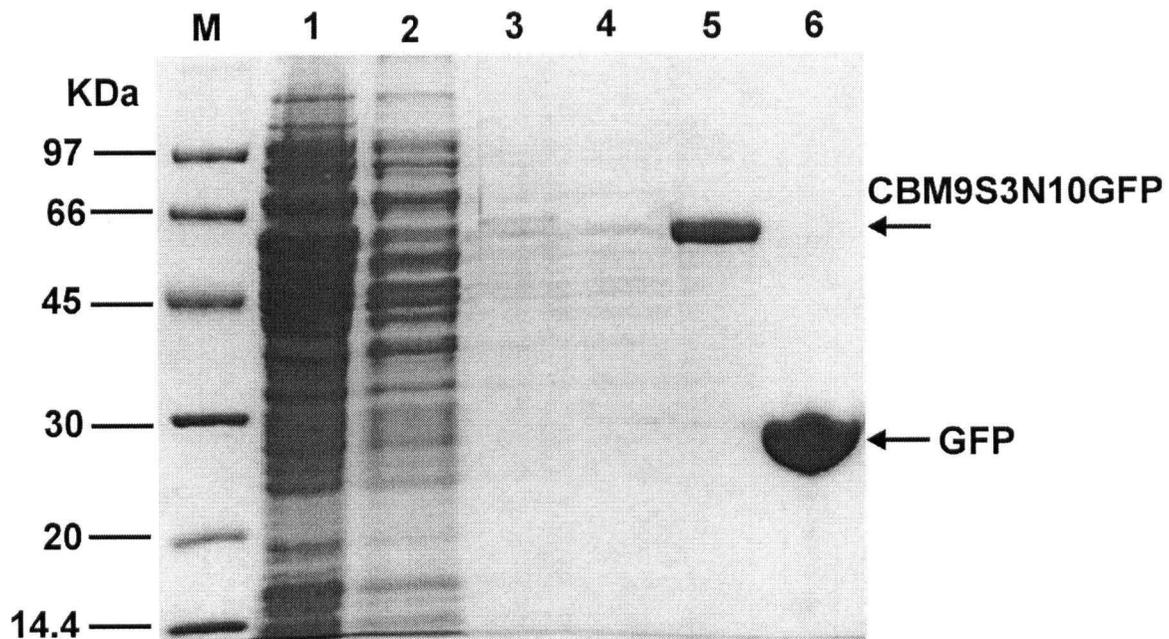


Figure 2.5 SDS-PAGE documentation of the affinity purification of CBM9-GFP. 12% SDS-PAGE of CBM9-GFP purified on a 17 mL Perloza™ MT100 column. All samples dissolved in sample buffer containing 10% SDS. Lane M: molecular mass markers in kg mol^{-1} . Lane 1: clarified cell extract prior to column loading. Lane 2: column flow through. Lane 3: high salt wash. Lane 4: low salt wash. Lane 5: pure CBM9-GFP eluted in TBS8 containing 1-M glucose. Lane 6: purified GFP after affinity-tag removal by immobilized Factor X_a.

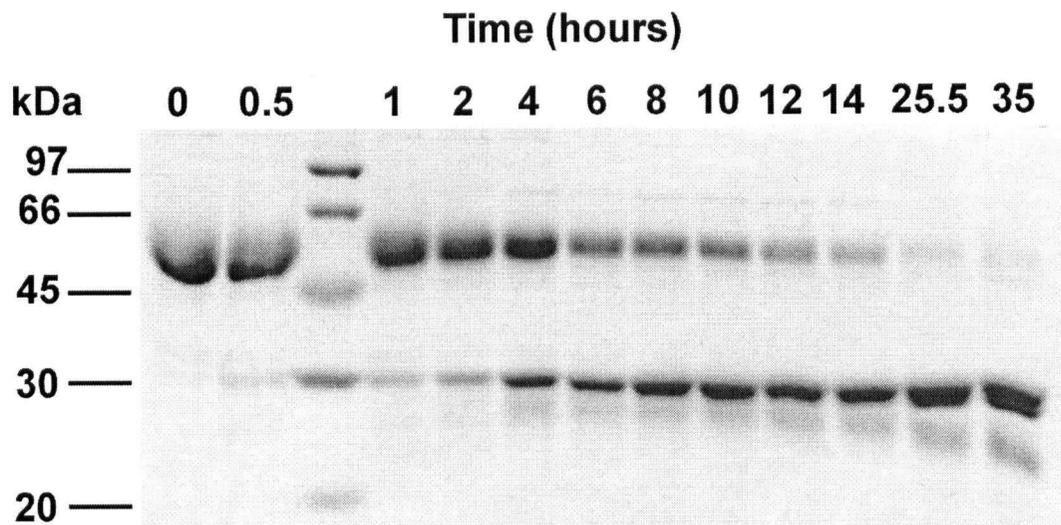


Figure 2.6 Time course of CBM9-GFP cleavage by Factor X_a at 23°C as shown on a 12% SDS-PAGE. A fusion protein to Factor X_a concentration ratio of 1000:1 was used.

2.7 References

- Appa Rao, K. B.; Garg, L. C.; Panda, A. K.; Totey, S. M. (1997). "High-level expression of ovine growth hormone in *Escherichia coli*: single-step purification and characterization". *Protein Expression and Purification* 11(2):201-208.
- Baek, M. C.; Choi, K. H.; Oh, T. G.; Kim, D. H.; Choi, E. C. (1997). "Overexpression of arylsulfate sulfotransferase as fusion protein with glutathione S-transferase". *Protein Expression and Purification* 11(3):257-262.
- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Chalfie, M.; Kain, S. (1998). *Green fluorescent protein : properties, applications, and protocols*. New York: Wiley-Liss. p 385
- Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. (1994). "Green fluorescent protein as a marker for gene expression". *Science* 263(5148):802-805.
- Chang, M.; Bolton, J. L.; Blond, S. Y. (1999). "Expression and purification of hexahistidine-tagged human glutathione S-transferase P1-1 in *Escherichia coli*". *Protein Expression and Purification* 17(3):443-448.
- Chase, H. A. (1984). "Affinity separations utilising immobilised monoclonal antibodies-A new tool for the biochemical engineer". *Chemical Engineering Science* 39(7-8):1099-1125.
- Cramer, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. (1996). "Improved green fluorescent protein by molecular evolution using DNA shuffling". *Nature Biotechnology* 14(3):315-319.
- Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stuber, D.; Henco, K. (1994). "6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification". *Methods in Molecular Biology* 31:371-387.
- Fox, J. D.; Kapust, R. B.; Waugh, D. S. (2001). "Single amino acid substitutions on the surface of *Escherichia coli* maltose-binding protein can have a profound impact on the solubility of fusion proteins". *Protein Science* 10(3):622-630.
- Fox, J. D.; Waugh, D. S. (2003). "Maltose-binding protein as a solubility enhancer". *Methods in Molecular Biology* 205:99-117.

- Guan, K. L.; Dixon, J. E. (1991). "Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Hearn, M. T.; Acosta, D. (2001). "Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins". *Journal of Molecular Recognition* 14(6):323-369.
- Hochuli, E.; Dobeli, H.; Schacher, A. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues". *Journal of Chromatography* 411:177-184.
- Jonasson, P.; Liljeqvist, S.; Nygren, P. A.; Stahl, S. (2002). "Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*". *Biotechnology and Applied Biochemistry* 35(Pt 2):91-105.
- Kwan, E.; Guarna, M. M.; Boraston, A. B.; Gilkes, N. R.; Haynes, C. A.; Kilburn, D. G.; Warren, R. A. (2002). "Self-activating factor X derivative fused to the C-terminus of a cellulose-binding module: Production and properties". *Biotechnology and Bioengineering* 79(7):724-732.
- Ladisich, M. R. (2001). *Bioseparations engineering : principles, practice, and economics*. New York: Wiley.
- Littlejohn, T. K.; Takikawa, O.; Skylas, D.; Jamie, J. F.; Walker, M. J.; Truscott, R. J. (2000). "Expression and purification of recombinant human indoleamine 2, 3-dioxygenase". *Protein Expression and Purification* 19(1):22-29.
- Lowe, C. R.; Lowe, A. R.; Gupta, G. (2001). "New developments in affinity chromatography with potential application in the production of biopharmaceuticals". *Journal of Biochemical and Biophysical Methods* 49(1-3):561-574.
- Mach, H.; Middaugh, C. R.; Lewis, R. V. (1992). "Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins". *Analytical Biochemistry* 200(1):74-80.
- Modesti, A.; Taddei, N.; Bucciantini, M.; Stefani, M.; Colombini, B.; Raugei, G.; Ramponi, G. (1995). "Expression, purification, and characterization of acylphosphatase muscular isoenzyme as fusion protein with glutathione S-transferase". *Protein Expression and Purification* 6(6):799-805.
- Nagai, K.; Thogersen, H. C. (1984). "Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*". *Nature* 309(5971):810-812.

- Nilsson, J.; Stahl, S.; Lundeberg, J.; Uhlen, M.; Nygren, P.-A. (1997). "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins". *Protein Expression and Purification* 11(1):1-16.
- Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation". *Nature* 258(5536):598-599.
- Sambrook, J.; Fritsch, E. F.; Maniatis, T. (1989). *Molecular cloning : a laboratory manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press.
- Shimomura, O.; Johnson, F. H.; Saiga, Y. (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*". *Journal of Cellular and Comparative Physiology* 59:223-239.
- Smith, D. B.; Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase". *Gene* 67(1):31-40.
- Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology* 60(5):523-533.
- Vaillancourt, P.; Simcox, T. G.; Zheng, C. F. (1997). "Recovery of polypeptides cleaved from purified calmodulin-binding peptide fusion proteins". *Biotechniques* 22(3):451-453.
- Wassenberg, D.; Schurig, H.; Liebl, W.; Jaenicke, R. (1997). "Xylanase XynA from the hyperthermophilic bacterium *Thermotoga maritima*: Structure and stability of the recombinant enzyme and its isolated cellulose-binding domain". *Protein Science* 6(8):1718-1726.
- Wilchek, M.; Chaiken, I. (2000). "An overview of affinity chromatography". *Methods in Molecular Biology* 147:1-6.
- Winterhalter, C.; Heinrich, P.; Candussio, A.; Wich, G.; Liebl, W. (1995). "Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*". *Molecular Microbiology* 15(3):431-444.
- Zheng, C. F.; Simcox, T.; Xu, L.; Vaillancourt, P. (1997). "A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins". *Gene* 186(1):55-60.

3 Strategy for Selecting and Characterizing Linker Peptides for CBM9-Tagged Fusion Proteins Expressed in *E. Coli*

3.1 Introduction

Fusion tag technology provides an efficient and generic strategy for high titre production and subsequent purification of recombinant proteins from culture media or clarified cell extracts. The technology is based on genetically linking a short peptide or small protein domain to either the amino- or carboxy-terminus of the target protein and then using the fusion tag to facilitate efficient capture and purification of the chimeric protein on an affinity matrix. Certain fusion partners such as the maltose binding protein (MBP) (di Guan et al. 1988), other carbohydrate binding modules (CBMs) (Ahn et al. 2004; Murashima et al. 2003), thioredoxin (Trx) (LaVallie et al. 1993), and the *E. coli* protein N-utilizing substance A (NusA) (Davis et al. 1999), have been shown to improve product solubility and expression by stabilizing the mRNA (Makrides 1996) or by providing a psuedo-chaperone effect (Samuelsson et al. 1994). Fusion tags have also been used to increase *in vivo* proteolytic stability (Jansson et al. 1990; Martinez et al. 1995; Staahl and Nygren 1997), to serve as effective expression and localization reporters (Gerdes and Kaether 1996; Ozawa 2006), and, when combined with their corresponding leader peptide, to control product localization in or secretion from the expression host (Ford et al. 1991; Johnson et al. 1992; Uhlen and Moks 1990). Skillful selection and positioning of the fusion tag therefore provides a flexible platform for recombinant protein processing.

* A version of this chapter has been published in *Biotechnology and Bioengineering*. [Reference: Mojgan Kavooosi, A. Louise Creagh, Douglas G. Kilburn, Charles A. Haynes, Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *E. coli*. *Biotechnology and Bioengineering*, 98(3); 599-610; (2007)]

The polyhistidine tag (His-tag) (Hochuli et al. 1987; Porath et al. 1975), which permits product capture and purification by immobilized metal affinity chromatography (IMAC), and the *Schistosoma japonicum* glutathione S-transferase (GST) tag (Smith and Johnson 1988), which has specific affinity for immobilized glutathione, are the two most popular affinity tags. However, other affinity tags offer some important advantages (Glatz et al. 1995). For target proteins that express poorly or exhibit low solubility, an MBP or CBM tag is often used to improve expression while at the same time permitting affinity purification on a complimentary carbohydrate resin (Kapust 1999). A number of reviews (Hearn and Acosta 2001; Terpe 2003) discuss several other useful affinity tags, including *Staphylococcus aureus* protein A and its synthetic two-domain variant (Moks 1986; Nilsson 1987), and the increasingly popular FLAG peptide (Einhauer and Jungbauer 2001; Hopp 1988).

The presence of an affinity tag can compromise the function and intended application of the target protein. This problem is generally overcome by incorporating a specific cleavage site for an endopeptidase that may be used to remove the tag either following elution from the column or while the fusion protein is bound to the affinity matrix (Arnau et al. 2006). Two relatively new commercially available endopeptidases, HRV 3C (Libby et al. 1988) and TEV (Shih et al. 2005) are thought, due to their relatively long recognition sites, to minimize non-specific cleavage. However, Factor X_a and enterokinase are still widely used because they allow for recovery of the natural N-terminus of the target.

The specific cleavage site is typically introduced as part of a longer linking peptide connecting the fusion tag and the target protein (Carter 1990; Jenny et al. 2003). Poly-glycine and glycine-rich sequences have been by far the most frequently used linkers because they enhance solubility, are thought to be resistant to proteolysis, and presumably confer conformational flexibility that enhances accessibility of the processing enzyme to its recognition site and allows the adjoining domains to function independently (Brändén and Tooze 1999). For example, the (G₄S)_x motif of Huston *et al.* (Huston et al.

1988) is widely used as the linker between single-chain Fv fragments (scFv) (Huston et al. 1993).

Proline-rich peptides, particularly Pro-Thr repeats, are also frequently employed as linkers due to their apparent resistance to proteolytic degradation and their frequent presence as peptides connecting domains in multidomain proteins (Beck et al. 1997; Wootton and Drummond 1989). Many microbial cellulases and xylanases connect their carbohydrate binding modules to their catalytic module via linkers rich in proline and hydroxylamines (Beguín and Aubert 1994; Gilkes et al. 1991; Gilkes et al. 1988; Knowles et al. 1987; Tomme et al. 1994; Tomme et al. 1988). These natural linkers have been shown to enhance the functional properties of the associated carbohydrate binding module (Ong 1995) and the overall proteolytic stability (Greenwood et al. 1989) of these glycolytic enzymes.

Although comprehensive studies on linker performance are lacking, data drawn from the fusion protein literature suggest that the utility of an affinity tag technology can be compromised by a number of factors sensitive to the chemistry and length of the linker (Brinkmann et al. 1992; Tang et al. 1996), including poor product expression (Furusawa 1976; Sauer 2001; Tang et al. 1996), interference with target protein function (Arvidson et al. 2003; Maeda et al. 1997), susceptibility to proteolysis (Fukuoka et al. 1993; Hampe et al. 2000), low binding efficiency of the affinity tag (Jalaguier et al. 1996; Lucius et al. 1992; Sheffield et al. 1999), and poor enzymatic processing in situations where removal of the affinity tag is required (Davis et al. 1999; Fassina et al. 1994; Guan and Dixon 1991). Here, we use a complimentary set of experiments to evaluate the properties of linkers as part of fusion proteins to the C-terminal family 9 carbohydrate binding module (CBM9) of xylanase 10A from *Thermotoga maritima* (Figure 3.1) (Boraston et al. 2001; Winterhalter et al. 1995). This affinity tag system is used to characterize the performance of the popular poly-glycine and PT-repeat linker chemistries. It is then used to evaluate an alternative strategy for selecting a linker chemistry based on the application of MEROPS™ (Rawlings et al. 2006), a comprehensive bioinformatics database of peptidase (protease) specificities, to predict linker sequences that are resistant to

proteolysis during fusion protein expression and processing in the host organism, which in this case, is the gram negative bacteria *Escherichia coli*.

Previously (Kavoosi et al. 2004), we introduced a novel affinity separation technology based on CBM9 and showed that it has performance characteristics competitive with other commercial affinity tag systems (Vaillancourt et al. 1997). It utilizes a high-capacity (static binding capacities between 90 and 150 mg mL⁻¹) cellulosic resin that is *ca.* 1/50th the cost of an equivalent volume of Ni²⁺-NTA (IMAC) resin used for polyhistidine tag systems, making the CBM9 affinity tag system a cost-competitive and potentially attractive technology for industrial bioprocessing. In this study, the CBM9 affinity tag is fused to the N-terminus of a convenient reporter, the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* (Cramer et al. 1996; Shimomura et al. 1962), through a selected linker sequence and the Factor X_a (FX_a) recognition sequence. Luminescence resonance energy transfer (LRET) with terbium as the donor and GFP as the acceptor is used to determine a characteristic relative distance of separation between the tag and target and its dependence on linker chemistry. Differential scanning calorimetry (DSC) is used to determine the influence of the linker on the thermodynamic stability of the fusion partners. These results are combined with MEROPSTM predictions and tandem mass-spectrometry (MS) data to interpret the effect of linker chemistry and length on production rates, resistance to proteolysis, independent folding and functioning of the two domains, binding isotherms, and the rate of enzymatic cleavage of the CBM9 tag using Factor X_a.

3.2 Materials and Methods

3.2.1 Reagents

Isopropyl-1-thio-β-D-galactoside (IPTG), glucose, kanomycin and all other chemicals were purchased from Sigma (St. Louis, MO, USA). All reagents were analytical grade unless stated otherwise. Restriction enzymes were purchased from New England Biolabs (Beverly, MA). T4-DNA ligase and PWO DNA polymerase were obtained from Roche Molecular Biochemicals (Laval, Quebec). PerlozaTM MT100 chromatography resin having a nominal particle diameter distribution of 50-80 μm was

purchased from Iontosorb Inc. (Czech Republic). Ni²⁺-Sepharose resin and *E. coli* BL21 (DE3) cells were obtained from Novagen (Milwaukee, MI).

3.2.2 Cloning of CBM9-Linker-FX_a-GFP Fusion Proteins

All cloning procedures were performed using standard molecular biology techniques (Sambrook and Russell 2001). The GFP and CBM9 coding regions were amplified from pGFPuv (Clontech, Palo Alto, CA) and pETCBM9 (Boraston et al. 2001), respectively, using the priming oligonucleotides listed in Table 3.1 and following the PCR protocol previously described (Kavoosi et al. 2004). The construction of the pET28-CBM9-P-FX_a-GFP expression vector for production of the CBM9-P-FX_a-GFP fusion protein, where the linker is comprised of a single proline residue P preceding the FX_a processing site comprised of the 4 amino acid sequence IEGR, was achieved as follows: primers 1 and 3, encoding for restriction endonuclease sites Bsp HI and Pvu I, respectively, were used to amplify the CBM9-P coding region. The FX_a-GFP^a coding region was amplified using primers 7 and 8, which encode for restriction sites Pvu I and Not I, respectively. The resulting CBM9-P and FX_a-GFP^a coding regions were digested with Bsp HI/Pvu I and Pvu I/Not I, respectively, and ligated (16°C, 16 h) into pET28 vector (Novagen) previously digested with Nco I and Not I to give pET28-CBM9-P-FX_a-GFP. Other fusion constructs were prepared in a similar manner using the primer sets shown in Table 3.1. DNA sequencing was performed to verify all constructs (NAPS Unit, The Michael Smith Laboratories, The University of British Columbia).

3.2.3 Fusion Protein Production and Purification

Fusion protein was produced in overnight cultures of *E. coli* BL21 (DE3) containing a pET28-CBM9-linker-FX_a-GFP expression vector diluted 100-fold in tryptone-yeast extract-phosphate medium (TYP) supplemented with 50 µg mL⁻¹ of kanamycin. To allow for meaningful comparison of productivities, the cells were grown at 37°C to a cell density (OD_{600 nm}) of 1.0 ± 0.1, induced with IPTG to a final concentration of 0.3 mM, and incubated for a further 16 ± 0.2 h at 30°C. The cells were harvested by centrifugation (8,500 x g) at 4°C for 20 min, resuspended in high salt buffer (1 M NaCl, 50 mM potassium phosphate, pH 7.0), and ruptured by two passages through

a French pressure cell (21000 lb in⁻²). The cell debris was then removed by centrifugation (27,000 x g) at 4°C for 30 min. The final CBM9-linker-FX_a-GFP fusion protein concentration in the cell extract was determined by SDS-PAGE analysis following purification of the fusion by affinity chromatography as described previously (Kavoosi et al. 2004). Cell growth and fusion protein production were also measured in real time by continuously monitoring OD_{620nm} and fluorescence emission at 510 nm (400 nm excitation wavelength) in a SpectraFluorPlus spectrophotometer (Tecan, USA). All experiments were done in replicates of five.

3.2.4 Binding Isotherm Measurement and Analysis

Standard solutions of purified CBM9-linker-FX_a-GFP fusion protein in high salt buffer, ranging in protein concentration from 1 to 30 μM, were mixed with 1 mg (dry weight) of Perloza™ MT100 in high salt buffer to reach a final volume of 1 ml. The samples were incubated at 4°C for 30 hours (time course experiments show equilibrium reached within 3-5 hours) while rotating end over end. The loaded resin was separated by centrifugation at 27,000 x g for 16 min at 4°C. The supernatant was collected and the equilibrium concentration of unbound protein c_i (μM) was determined by UV absorbance at 280 nm using a Cary 100 Spectrophotometer (Varian). The concentration of bound fusion protein q_i (μmol g⁻¹ of Perloza™ MT100) was computed by mass balance. The equilibrium binding constant K_a (M⁻¹) and saturation capacity (q_i^{\max}) were then determined by non-linear regression of the Langmuir adsorption isotherm equation to the experimental isotherm data using GraphPad Prism 3.0 software.

3.2.5 Linker Stability Analysis

To allow clear visualization of differences in rates of linker hydrolysis, clarified cell lysate was incubated for different periods of time at room temperature in the absence of any protease inhibitors, and then analyzed as follows. Following the incubation period, clarified cell lysate (1.5 ml) was mixed with 200 μL of Perloza™ MT100 (94.5 mg dry weight mL⁻¹) for 3.5 h at room temperature while rotating end-over-end. The resin was collected by centrifugation at 8,500 x g for 8 min, washed three times with high salt buffer, 2X with low salt buffer (50 mM potassium phosphate, pH 7.0) and 1X with

TBS8 (15 mM NaCl, 10 mM Tris-HCl, pH 8.0). The fusion protein was then eluted and collected in 400 μ L of 3 M glucose in TBS8 and analyzed by SDS-PAGE to quantify the relative amounts of intact and CBM9 byproducts generated by proteolytic cleavage within the linker region of the fusion.

Intact CBM9-GFP and fusion protein degradation byproducts were sequence analyzed by LC/MS/MS in the Genome BC Proteomics Centre at the University of Victoria (Victoria, British Columbia, Canada) to verify protein sequence and to identify cleavage positions within the linker region. Briefly, 1 to 10 μ g of affinity-purified protein was isolated by SDS-PAGE, excised and extracted into trypsin cleavage buffer, digested overnight with trypsin, dissolved in 0.2% acetic acid, and transferred to autosampler vials for LC/MS/MS analysis. Tandem MS was performed using an LCQDECA IT mass spectrometer (Thermo Finnigan, San Jose, CA) with an in-house fabricated microelectrospray source and an HP1100 solvent delivery system (Agilent, Palo Alto, CA). Peptide sequences were determined using SEQUEST™ (Thermo Finnigan), and Peptide-Prophet™ (Keller et al. 2002) was used to verify correctness of peptide assignments.

3.2.6 Factor X_a Processing Analysis

To simplify our hydrolysis analysis, site-specific proteolytic removal of the CBM9 fusion tag was carried out using Factor X_a immobilized onto Perloza™ MT500, a highly porous (porosity of 0.97) cellulosic media that permits faster mass transfer than Perloza™ MT100 and is therefore better suited as stationary phase substrate for the immobilized enzyme reaction. Immobilization was achieved by producing and purifying Factor X_a as a chimeric fusion to CBM2a, a family 2a carbohydrate-binding module of xylanase 10A of *Cellulomonas fimi* (CBM2a-FX_a_{im}, where “im” indicates immobilized) (Kwan et al. 2002). Unlike CBM9, CBM2a and thus the CBM2a-FX_a fusion bind irreversibly to Perloza™ media, permitting processing of purified CBM9-linker-FX_a-GFP in the solution state by including 1 M glucose in the mobile phase. Purified CBM9-linker-FX_a-GFP (3 mg) was mixed with CBM2a-FX_a (3 μ g) in a TBS8 solution (final [Factor X_a] to [fusion protein] ratio of 1:1000) containing 1 M glucose and maintained at

21°C. Buffer was used in place of CBM2a-FX_{a_{im}} in control experiments. The 300 µL sample was continuously rotated end-over-end with 2 µL samples of the supernatant taken 0, 0.5, 1, 2, 4, 6, 8, 10, 12, 14, 25.5, and 35 h after initial mixing. Each 2 µL sample was diluted 2X in loading buffer and analyzed by SDS-PAGE to determine the percentage of cleaved fusion protein with time, from which was determined the half-life $t_{1/2}$ (h) of each hydrolysis reaction under identical reaction conditions.

3.2.7 Differential Scanning Calorimetry Studies

Melting thermograms for all CBM9-linker-FX_a-GFP fusion proteins were measured by differential scanning calorimetry (DSC) using a VP-DSC Extended Range MicroCalorimeter (MicroCal Incorporated, USA). Each 1.5281 mL sample contained 90 µM of purified protein in 50 mM potassium phosphate buffer (pH 7). Samples were dialyzed against pure buffer with at least five volume exchanges to achieve a stable and reproducible thermogram baseline. Degassed samples were syringe-loaded into the DSC and thermally scanned from 5°C to 110°C at a rate of 1°C min⁻¹. Excess heat capacity data for the protein melting transitions were calculated by subtracting the heat capacity baseline for the solvent obtained by loading buffer in both the sample and reference cells of the calorimeter (Creagh et al. 2005). Melting transitions for the CBM9 and GFP domains of each fusion were analyzed using software provided by MicroCal Inc. to determine the melting temperature (T_m) and the enthalpy of denaturation (ΔH_{cal}) at T_m .

3.2.8 LRET Studies

Hogue *et al.* (Hogue et al. 1992) have demonstrated that calcium-binding sites on proteins can bind Tb(III) and thereby serve as a lanthanide-based luminescent probe for measuring molecular distances. The inherently low absorbance of Tb(III) in solution is overcome in the bound state by the sensitizing effect of the chelating groups which, in the case of the strong calcium binding site of CBM9, bind Tb(III) in a close to perfect octahedral geometry through coordination with Asp, Glu and Val residues (Notenboom et al. 2001). Time-resolved energy transfer from excited Tb(III) in the metal-ion binding site therefore allows determination of the relative distance between the bound terbiums and an acceptor using Förster theory. Since lanthanide emission does not result from a

singlet-to-singlet transition characteristic of fluorescence, the process is termed luminescence resonance energy transfer (LRET).

Each purified CBM9-linker-FX_a-GFP fusion protein was incubated with Chelex 100X resin, mixing end-over-end at 4°C for 48 h, to strip all bound metal from the calcium binding sites of CBM9. The apo-protein was then passed over a Sephadex G10 size-exclusion column to remove any trace contaminants. Chelex-treated fusion protein was diluted to a concentration of 3 μM in nano-pure water containing excess terbium (15 μM) and the solution was mixed for 15-30 min at 4°C, centrifuged (27,000 x g, 10 min) to remove any precipitates, and used directly in the fluorescence and LRET measurements.

Steady-state fluorescence measurements and time-resolved luminescence decay measurements were made on a Cary Eclipse fluorescence spectrophotometer (Varian) equipped with a pulsed excitation source and gating system that takes advantage of the millisecond lifetime of the lanthanide to eliminate any background fluorescence from direct excitation of the acceptor fluorophore (GFP). Tb(III) binding to CBM9 in each CBM9-linker-FX_a-GFP fusion was confirmed by monitoring sensitized luminescence intensity at 512 nm during titration of the apo-protein with Tb(III), which indicated stoichiometric binding to the calcium binding sites of CBM9 (Notenboom et al. 2001). Fluorescence lifetime experiments were conducted at an excitation wavelength of 222 nm with collection of the emission intensity $I(t)$ at 512 nm, a wavelength near the peak maximum in the GFP fluorescence emission spectra (Chalfie and Kain 1998; Kain et al. 1995) that does not overlap with any of the emission peaks for Tb(III) (Jiao et al. 2003). Emission intensities were recorded following a 200 μs delay after the excitation flash to ensure that any signal detected at 512 nm is entirely due to resonance energy transfer and not from direct excitation. All readings were an average of 1000 cycles. Analysis of $I(t)$ data using Förster theory followed the procedure of Selvin *et al.* (Selvin 2002; Selvin and Hearst 1994). Measured $I(t)$ curves were fit to an exponential of the form $I(t) = I(0) \exp(-t/\tau_{da})$, where τ_{da} is the lifetime of the donor in the presence of the acceptor. The fits were always characterized by an $r^2 > 0.999$, confirming that the data follow a single

exponential decay and therefore, that the donor alone makes no significant contribution to the intensity signal recorded at 512 nm.

Förster theory relates the distance between donor and acceptor R to the resonance energy transferred E , given by $(1 - \tau_{da}/\tau_d)$,

$$R = R_o \left[\left(\frac{1}{E} \right) - 1 \right]^{1/6} \quad (3.1)$$

where τ_d is the lifetime of the donor in the absence of the acceptor, and R_o is the Förster distance, where there is 50% energy transfer ($E = 0.5$), given by

$$R_o = 0.211 \frac{\kappa^2 Q_D J}{\eta^4} \quad (3.2)$$

with J , the spectral overlap, given by

$$J = \frac{\int f_D(\lambda) \varepsilon_a(\lambda) \lambda^4 d\lambda}{\int f_D(\lambda) d\lambda} \quad (3.3)$$

In Equation 3.2, κ^2 is an orientation factor having a value of 2/3 in these experiments, Q_D is the quantum yield of the donor (Tb(III) bound CBM9) (Root et al. 1999). The refractive index η is 1.33 for biological samples in water and, in equation 3.3, $f_D(\lambda)$ is the corrected fluorescence intensity of the donor, and ε_a is the extinction coefficient of the acceptor. The value of Q_D taken from (Root et al. 1999) may not fully incorporate the contribution of the sensitizing chromophore, CBM9. As a result, the value of R_o is likely smaller than its true value. Moreover, as noted previously, CBM9 is capable of binding three Tb(III) ions at the reaction conditions used. As a result, energy transfer is the sum of E for each bound Tb(III). Nevertheless, since reaction conditions and R_o were held

constant in these studies, we do not report absolute R values, but rather $\frac{R}{R_p}$ values, where R_p is the measured separation distance between donor and acceptor in the CBM9-P-IEGR-GFP fusion.

3.3 Results and Discussion

Since we first introduced the CBM9 affinity tag (Kavoosi et al. 2004), the technology has been applied to the production and purification of over a dozen fusion proteins harboring the tag at the N-terminus and utilizing the different linkers characterized in this study. Here, we report results obtained for CBM9-linker-FX_a-GFP fusion proteins (Figure 3.1), taking advantage of the unique fluorescence of the GFP reporter to quantify the results. These results are consistent with those obtained for other fusion constructs. Table 3.2 lists the linker sequences studied. The lengths of the poly(G) and poly(PT) sequences were varied to assess the impact of linker length on performance. The poly(PT) linker used in this study is derived from the natural 23-amino acid linker found in endoglucanase A (Cen A) of the bacterium *Cellulomonas fimi* (Shen 1991).

3.3.1 Linker Selection and Screening Using MEROPS™

Although proteolysis is observed within protein domains, it is known to occur far more frequently within flexible or less compact regions of the folded polypeptide chain (*i.e.*, those with large crystallographic B -factors and poorly defined electron density) (Fontana et al. 1997). It is thought that these flexible regions can more easily adopt conformations compatible with the active site of the attacking endopeptidase (Huber and Bennett 1983; Huyton et al. 2003). Inter-domain linkers are therefore particularly susceptible to proteolysis, a feature frequently exploited by crystallographers to identify and isolate domains of interest (Orengo et al. 1997; Wheelan et al. 2000). A number of bioinformatics tools have recently been developed to identify potential cleavage sites using genome-based assignment of peptidases and their associated putative specificities (Boyd et al. 2005). Among the most comprehensive is MEROPS™, a protease specificity database developed by Rawlings *et al.* (Rawlings et al. 2006). The use of

MEROPS™, which includes the complete inventory of known and putative *E. coli* peptidases (Barrett et al. 1998), therefore establishes a potential strategy for *in silico* design of an effective linker for fusion tags to be expressed within a particular host organism. For example, the S₃N₁₀ region of the linker sequence shown in Table 3.2 is predicted by MEROPS™ to be completely resistant to *E. coli* endopeptidases.

This MEROPS™-based screening strategy may be applied to any putative or existing linker to assess its proteolytic stability. For each investigated linker, Table 3.2 reports putative cleavage sites that are predicted by MEROPS™ for all known and assigned endopeptidases of *E. coli*, as well as the cleavage sites observed experimentally by tandem mass spectrometry after production and purification of the fusion protein in *E. coli*. MEROPS™ identifies a number of potential cleavage sites within each poly(G)-IEGR sequence, predicting the potential for cleavage at both Gly/Gly (through the action of the *E. coli* peptidases colicin V processing peptidase, cytotoxin SubA, HtpX peptidase, as well as a number of unassigned peptidases from the M23, M48, and S8 families of *E. coli* peptidases) and Gly/Ile (Lit peptidase, an *E. coli* subfamily U49 peptidase) junctions. Tandem MS data for processed CBM9-poly(G)-FX_a-GFP fusion proteins reveals that cleavage does in fact occur at each of these junction types (Table 3.2). Thus, although widely used for fusion-protein bioprocessing, glycine-rich linker sequences are susceptible to hydrolysis by a number of *E. coli* peptidases. In contrast, poly(PT) sequences are predicted by MEROPS™ and confirmed by tandem MS results to be proteolytically stable in *E. coli*, supporting the dominant rationale for utilizing these “natural” sequences as linking agents for chimeric protein expression in bacteria (Gustavsson et al. 2001).

The FX_a recognition sequence is predicted by MEROPS™ to be stable against *E. coli* peptidases. Tandem MS results, however, revealed cleavage at the Gly/Arg junction within the FX_a recognition sequence, possibly due to endopeptidase mediated hydrolysis upstream of the FX_a cleavage site and subsequent activity by carboxypeptidases to the Gly/Arg junction. As a result, for all linkers studied, a small amount of cleavage within the FX_a recognition sequence was observed in the absence of the FX_a processing enzyme. The use of HRV 3C or TEV in place of the FX_a would not alleviate this problem and in

fact might exacerbate it since both HRV 3C and TEV have potential sites of *E. coli* endopeptidase-catalyzed hydrolysis predicted by MEROPS™.

Table 3.2 also reports sites of hydrolysis within the linker sequence observed by tandem MS but not predicted by MEROPS™. All sites of cleavage within the poly(G) and the poly(PT)_x-P linkers correspond to ones predicted by MEROPS™. For the S₃N₁₀ linker, tandem MS data indicate cleavage within the poly(S) region which is not captured by MEROPS™ predictions. However, the amount of cleavage is extremely small, as the MS Data reveals that the S₃N₁₀ linker remains intact for at least 99.2% of the fusion protein population. Therefore, MEROPS™ appears to be a useful tool in designing and selecting linkers resistant against endogeneous peptidases for expression in a given host.

3.3.2 Impact of the Linker on Fusion-Tag Performance

Our ongoing studies on CBM9-tagged fusion proteins indicate that the choice of linker chemistry and length can have a significant impact on fusion-tag performance. While similar specific growth rates [average doubling time of 0.5614 h⁻¹ (± 0.0256 h⁻¹) as monitored by optical density at 620 nm] were observed for all constructs, fusion protein production rates and yields showed a dependence on linker chemistry (Figure 3.2). A significant drop in the expression rate is observed when tryptophans are introduced into the linker, with the drop in productivity proportional to the number of aromatic amino acids in the sequence, which might explain why tryptophan and phenylalanine are seldom found in natural linkers (Argos 1990). A weaker dependence on linker composition is observed for sequences devoid of aromatic amino acids or significant hydrophobic amino acid content.

In addition to tandem MS analysis, the stability of the linker against long-term exposure to endogeneous *E. coli* peptidases was examined by SDS-PAGE analysis (Figure 3.3). The results confirm that proteolytic stability is strongly correlated to linker composition and length. As both the CBM9 (melting temperature @ pH 7 of 97.9 °C) and GFP (T_m @ pH 7 of 83.7 °C) domains are highly stable, hydrolysis of the fusion protein is largely restricted to the linker region, resulting in a band of hydrolyzed products covering a relatively narrow molecular-weight range lying slightly above the

molecular weight at which CBM9 alone runs on the gel (*ca.* 28 kDa). The results observed by SDS-PAGE analysis (Figure 3.3) are consistent with MEROPSTTM predictions and tandem MS results reported in Table 3.2. The poly(PT)_x-P linkers and the synthetic S₃N₁₀ linker are significantly more stable than the two poly(G) linkers against hydrolysis by *E. coli* peptidases. Densitometry data on gels from independent samples indicate that about 36±5% of the expressed CBM9-(G)₃-IEGR-GFP fusion protein is lost due to hydrolysis within or near the linker region during the fermentation and purification process; these losses are significantly higher (45±5%) for the CBM9-(G)₁₅-IEGR-GFP fusion protein.

For the (PT)_x-P-IEGR linker series, the size of the CBM9-containing fragment released by proteolysis increases with x, consistent with MS data that show no significant cleavage within the (PT)_x-P sequence while detecting cleavage within the FX_a processing site. In contrast, the dominant CBM9-containing fragment released by hydrolysis is the same size for the (G)₃-IEGR and (G)₁₅-IEGR constructs, indicating a common and preferred set of hydrolysis sites within both poly(G) sequences.

The S₃N₁₀ linker was the most proteolytically stable linker studied, with combined losses due to hydrolysis within the linker-region and Factor X_a processing site during long-term incubation studies with *E. coli* proteases constituting less than 7±4% of the purified CBM9-S₃N₁₀-IEGR-GFP product. Virtually all of that loss was due to hydrolysis at the Gly/Arg junction within the FX_a processing site as determined by tandem MS results. Therefore, despite its relatively long length, the S₃N₁₀ construct maintained good proteolytic resistance.

Interdomain linkers can also influence the structure and activity of the connected domains (Arvidson et al. 2003; Deyrup et al. 1999). Equilibrium binding isotherms were therefore performed to determine the influence of the linker on the performance of CBM9 in purifying the fusion protein. Determined by nonlinear regression of the Langmuir equation ($q_i = q_i^{\max} K_a c_i / (1 + K_a c_i)$) to adsorption isotherm data (Figure 3.4), association constants (K_a) near 10⁶ M⁻¹ (Table 3.3) were measured for binding of all CBM9-linker-FX_a-GFP fusion proteins to PerlozaTM MT100. The K_a values are essentially equivalent

to that of free CBM9 ($K_a = 1.2 \times 10^6 \text{ M}^{-1}$), indicating that the linker and the GFP fusion partner do not interfere with the function of the affinity tag. Similarly, saturation capacities (q_i^{max}) of *ca.* 10 μmol protein/g Perloza™ MT100 resin were measured for all fusion proteins (Table 3.3), representing a saturation binding capacity of *ca.* 530 mg protein/g resin (75 mg protein/mL column), which is significantly higher than almost any commercially available affinity matrix.

Half lifes ($t_{1/2}$, representing the time where 50% of the original fusion protein has been cleaved) for FX_a processing of CBM9-Linker-FX_a-GFP fusion proteins were determined by SDS-PAGE analysis (Figure 3.5) and were found to depend strongly on linker chemistry but not on linker length (Table 3.4). Relatively rapid cleavage is observed for fusion proteins containing either a poly(G) linker or the S₃N₁₀ linker. The introduction of a proline residue or, more dramatically, a PT repeat sequence upstream of the FX_a processing site was found to dramatically reduce the rate of cleavage, suggesting that FX_a function is inhibited by these upstream elements.

The proteolytically susceptible poly-glycine linkers had the fastest FX_a processing rates, with a $t_{1/2}$ of $1.4(\pm 4.0) \times 10^4 \text{ s}$ observed for both the G₃ and G₁₅ linkers. These results are not surprising since glycine is often placed directly upstream of the N-terminus of a protease recognition sequence to enhance its rate of cleavage (Guan and Dixon 1991). However, for fusion proteins expressed and localized in the cytoplasm of *E. coli*, the approximately 2-fold enhancement in the rate of FX_a processing relative to the S₃N₁₀ linker is offset by the significantly higher loss of poly(G)-linked fusion proteins due to linker proteolysis by endogenous *E. coli* peptidases.

3.3.3 Determination of Characteristic Distances Using Luminescence Resonance Energy Transfer

Increasing the length of the (PT)_x-P linker had relatively little impact on susceptibility to proteolysis. In contrast, increasing the length of the poly(G) linker from 3 to 15 amino acids resulted in a dramatic increase in linker proteolysis, reflecting both an increase in the number of potential cleavage sites as predicted by MEROPST™ (Table 3.2), and the possibility of an increase in the accessibility of those sites to *E. coli*

peptidases. To explore this latter issue, we used LRET to measure relative distances of separation between the donor chromophore, Tb(III) bound to the calcium-binding sites of CBM9, and the acceptor chromophore, GFP.

Fluorescence lifetime $I(t)$ data, such as that shown in Figure 3.6 for the CBM9-(PT)₂P-IEGR-GFP fusion protein, were collected at 512 nm (222 nm excitation wavelength) and fit to the exponential decay equation to determine τ_{da} , the lifetime of the donor in the presence of the acceptor. Förster theory (equations 3.1 to 3.3) was then used to compute the characteristic distance R between donor and acceptor (Table 3.5). As one would expect, the shortest linker sequence P-IEGR is characterized by the smallest R ($32.6 \pm 1.0 \text{ \AA}$) value, hereafter called R_p . $\frac{R}{R_p}$ is therefore greater than unity for all other fusion proteins studied. For a given linker chemistry, $\frac{R}{R_p}$ increases with increasing linker length. The increase in $\frac{R}{R_p}$ with linker length is particularly strong for the poly(G)-IEGR linker, suggesting that this linker adopts a hydrated, extended configuration that makes it accessible to both *E. coli* peptidases, leading to an enhancement in the proteolytic degradation rate, and FX_a, leading to an improvement in site specific tag-removal kinetics. Thus, the chemistry of the linker is seen to strongly influence overall linker performance and stability in a manner that is not fully captured by MEROPS™. As a result, we find that MEROPS™ can be used as an initial but not an exhaustive linker selection strategy.

3.3.4 Influence of the Linker on the Thermodynamic Stability of CBM9-Linker-FX_a-GFP Fusion Proteins

The thermodynamic stability of the native protein fold is known to be a key determinant in the expression of recombinant proteins in *E. coli*, with a decrease in the thermodynamic stability generally resulting in a decrease in the half-life of the protein (McLendon and Radany 1978). For example, expression yields progressively increase and proteolysis rates decrease for increasingly higher T_m variants of either T4 lysozyme (Inoue and Rechsteiner 1994) or the N-terminal domain of X-repressor protein (Parsell

and Sauer 1989). Studies by Kwon *et al.* (Kwon et al. 1996) on barnase variants indicate that these correlations also hold for recombinant proteins exported to the periplasm of *E. coli*, where a 4.3 °C increase in the T_m of barnase resulted in a 50% increase in protein yield. As their length affects spatial positioning of the fused domains (Table 3.5), linkers therefore may also influence intracellular rates of proteolysis by modulating the thermodynamic stabilities of the fusion partners.

We therefore used the proline-containing set of linkers to explore the connection between linker length and the thermodynamic stabilities of the fused domains. Melting thermograms for CBM9 on its own, for CBM9-P-IEGR-GFP and for CBM9-(PT)₇P-IEGR-GFP (Figure 3.7) show that formation of the fusion does not alter the T_m of GFP, but results in a decrease in the T_m and in ΔH_{cal} , the calorimetric enthalpy of denaturation, of CBM9 (Table 3.6). Both results indicate that fusing CBM9 to GFP through the linker-IEGR sequence thermodynamically destabilizes CBM9, with the degree of destabilization increasing with decreasing length of the (PT)_xP linker such that $\Delta T_m = - 8.4$ °C when the (PT)₀P linker (*i.e.*, the P linker) is used. This result correlates well with intracellular protein yield data reported in Figure 3.2, which show a decrease in fusion protein yield with decreasing length of the (PT)_xP linker.

3.4 Conclusions

Linker design and its importance to the performance of fusion tag technology has received relatively little attention (Carlsson et al. 1996; Gustavsson et al. 2001; Sauer 2001). Here, we have shown that the level of expression, proteolytic stability, thermodynamic stability, and Factor X_a processibility of fusion proteins containing an N-terminal CBM9 tag and expressed in recombinant *E. coli* can depend on the chemistry and length of the linker. This points to the need for methods for designing or selecting an effective linker that take into account the specific cellular environment in which the fusion protein is produced and localized, and from which it must be purified. For example, glycine-rich linker sequences, which are currently by far the most commonly reported linker chemistry, were found in this study to be quite susceptible to hydrolysis by *E. coli* peptidases, emphasizing that linkers should be specifically designed to resist

hydrolysis catalyzed by the peptidases of the host organism. Our strategy for meeting this important design criterion was to exploit bioinformatics tools (MEROPSTTM) for predicting protease specificities to screen putative linker sequences for hydrolytic resistance to endogenous peptidases. This approach builds on earlier efforts to improve linker design by selecting linker sequences derived from naturally occurring linkers (Gustavsson et al. 2001). Both approaches are shown in this work to identify linkers with a considerably higher resistance to endogenous *E. coli* proteolysis than observed for poly(G)-based linkers. The S₃N₁₀ linker selected by MEROPSTTM-based screening was slightly more resistant to proteolysis than those designed from naturally occurring linkers (the (PT)_x-P linker series). However, in both cases, the majority of degradation occurred within the Factor X_a processing site and not within the linker itself.

While MEROPSTTM-based screening was effective in selecting a stable linker, tandem MS based sequencing of the C-terminal peptide released by tryptic digestion of the CBM9-containing hydrolysis product indicated that certain sites of potential hydrolysis predicted by MEROPSTTM were not cleaved, while one site not identified by MEROPSTTM was cleaved, albeit to a very small degree. Thus, the screening method is not exact, possibly due to a dependence of proteolytic susceptibility on local structure (Parsell and Sauer 1989; Wriggers et al. 2005). Nevertheless, it is shown in this work to provide a novel and potentially useful strategy for linker design.

Our results further show that design of an effective linker should not be based on proteolytic stability considerations alone. The thermodynamic stability and spatial positioning of the fused proteins/domains may also influence the overall performance of the fusion tag by altering expression levels and the rate at which the tag can be removed from the target protein following affinity capture and purification. The additional experimental methods (DSC, LRET, SDS-PAGE, etc.) described in this work therefore constitute a more comprehensive approach to linker selection and performance analysis.

3.5 Tables

Table 3.1 Oligonucleotides used in the construction of CBM9-Linker-IEGR-GFP fusion proteins. Restriction sites are underlined and the Factor X_a recognition sequence is in bold.

Primer	Oligonucleotide Sequence (all sequences are 5'-3')
1	TTGCTAGCT <u>TCATGACTAGCGGAATAATGGTAGC</u>
2	AGCGGCCGCAGGCCTACCCT <u>CGATCGGAGTCGGAGTCGGCGTCGG-AGTCGGAAGCTTGATGAGCCTGAGGTTACC</u>
3	AGCGGCCGCCT <u>CGATCGGAAGCTTGATGAGCCTGAGGTTACC</u>
4	CT <u>CGATCGGACCACCACCAAGCTTGATGAGCCTGAGGTT</u>
5	TCCCT <u>CGATCGCGCCGCCACCGCCGCCACCACCACCACCGCCGCCA-CCGCCACCGCCAAGCTTGATGAGCCTGAGGTTACC</u>
6	TCCCT <u>CGATCGCGAGGTTGTTGTTATTGTTATTGTTGTTGTTGTTTCG-AGCTCGAAAGCTTGATGAGCCTGAGGTTACC</u>
7	CCGATCGAGGGTCGTATCATGAGTAAAGGAGA
8	<u>TGCGGCCGCTTTGTAGAGCTCATCCATGCCATGTGTAATCCC</u>
9	AAGAATTCAAGCTTCCGACGCC GATCGAGGGTCGTATGAGTAAAG-GAGAAGA ACTTTTCAC
10	TTCAAGCTTCCGACTCCGACTCCGACGCCGACTCCGACCCCGACTC-CAACTCCG ATCGACGGTCGTATCATGAGTAAAGGAGA

Table 3.2 Linker lengths, sequences, and susceptibility to proteolytic cleavage as predicted by MEROPS™ and confirmed experimentally by LC/MS/MS

Linker	Length (aa)	Cleavage sites predicted by MEROPS™ (↓) and observed by tandem MS (↑)	Cleavage Sites Predicted by MEROPS™
G ₃ -IEGR	3	G _↑ [↓] G _↑ [↓] G _↑ [↓] -IEG _↑ R	G/G (C39; M23; M48; S8); G/I (U49)
G ₁₅ -IEGR	15	G _↑ [↓] -IEG _↑ -R	G/G (C39; M23; M48; S8); G/I (U49)
P-IEGR	1	P-IEG _↑ R	
(PT) ₂ P-IEGR	5	PTPTP - IEG _↑ R	
(PT) ₄ P-IEGR	9	PTPTPTPTP - IEG _↑ R	
(PT) ₇ P-IEGR	15	PTPTPTPTPTPTPTP - IEG _↑ R	
S ₃ N ₁₀ -IEGR	15	S _↑ SSNNNNNNNNNNL [↓] A-IEG _↑ R	L/A (Family S8)

Symbol (↓) denotes putative cleavage site predicted by MEROPS™

Symbol (↑) denotes cleavage site observed by tandem mass spectrometry

Table 3.3 Regressed Langmuir isotherm parameters for binding of CBM9-Linker-IEGR-GFP on Perloza™ MT100 at pH 7 (4°C). Standard deviations (σ) computed from triplicate measurements.

Linker	K_a (M^{-1})	q^{\max} ($\mu\text{mol/g resin}$)
P-IEGR	$2.0 (\pm 0.26) \times 10^6$	10.4 (± 0.18)
(PT) ₂ P-IEGR	$2.4 (\pm 0.36) \times 10^6$	10.4 (± 0.25)
(PT) ₄ P-IEGR	$6.0 (\pm 0.99) \times 10^6$	9.9 (± 0.30)
(PT) ₇ P-IEGR	$3.1 (\pm 0.24) \times 10^6$	10.3 (± 0.14)
G ₃ -IEGR	$4.6 (\pm 0.71) \times 10^6$	9.9 (± 0.27)
G ₁₅ -IEGR	$5.2 (\pm 0.82) \times 10^6$	9.5 (± 0.26)
S ₃ N ₁₀ -IEGR	$1.5 (\pm 0.21) \times 10^6$	9.9 (± 0.31)
CBM9	$1.2 (\pm 0.06) \times 10^6$	11.2 (± 0.15)

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence intervals)

Table 3.4 Half-lives for Factor X_a cleavage of CBM9-Linker-IEGR-GFP fusion proteins. All experiments were conducted in 50 mM potassium phosphate buffer (pH 7, 21 °C) at a fusion protein to Factor X_a concentration ratio of [1000] to [1].

Linker	<i>t</i> _{1/2} (h)
P-IEGR	8 (±1)
(PT) ₂ P-IEGR	25
(PT) ₄ P-IEGR	35
(PT) ₇ P-IEGR	> 35
G ₃ -IEGR	4
G ₁₅ -IEGR	4
S ₃ N ₁₀ -IEGR	6

Table 3.5

Average relative distance of separation R/R_p between the bound terbiums on CBM9 and the fluorescent chromophore of GFP determined by LRET at 21°C.

Linker	R/R_p
P-IEGR	1.00 (± 0.03)
(PT) ₂ P-IEGR	1.08 (± 0.02)
(PT) ₄ P-IEGR	1.11 (± 0.04)
(PT) ₇ P-IEGR	1.14 (± 0.02)
G ₃ -IEGR	1.09 (± 0.02)
G ₁₅ -IEGR	1.51 (± 0.10)
S ₃ N ₁₀ -IEGR	1.04 (± 0.05)

* reported errors represent $\pm\sigma$ (i.e. 60% confidence interval)

Table 3.6 Melting temperatures (T_m) and denaturation enthalpies (ΔH_{cal}) for CBM9 on its own and as part of various CBM9-linker-IEGR-GFP fusion proteins.

Linker	T_m (°C)	ΔH_{cal} (J mol ⁻¹)
None (isolated CBM9)	97.9	1.34 (± 0.06) $\times 10^5$
P-IEGR	89.5	0.71 (± 0.05) $\times 10^5$
(PT) ₂ P-IEGR	89.8	0.83 (± 0.05) $\times 10^5$
(PT) ₄ P-IEGR	91.6	0.96 (± 0.04) $\times 10^5$
(PT) ₇ P-IEGR	93.4	1.12 (± 0.05) $\times 10^5$

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence intervals)

3.6 Figures



Figure 3.1 Schematic representation of the CBM9-Linker-FX_a-GFP fusion protein. FX_a indicates the recognition site (IEGR) for Factor X_a processing.

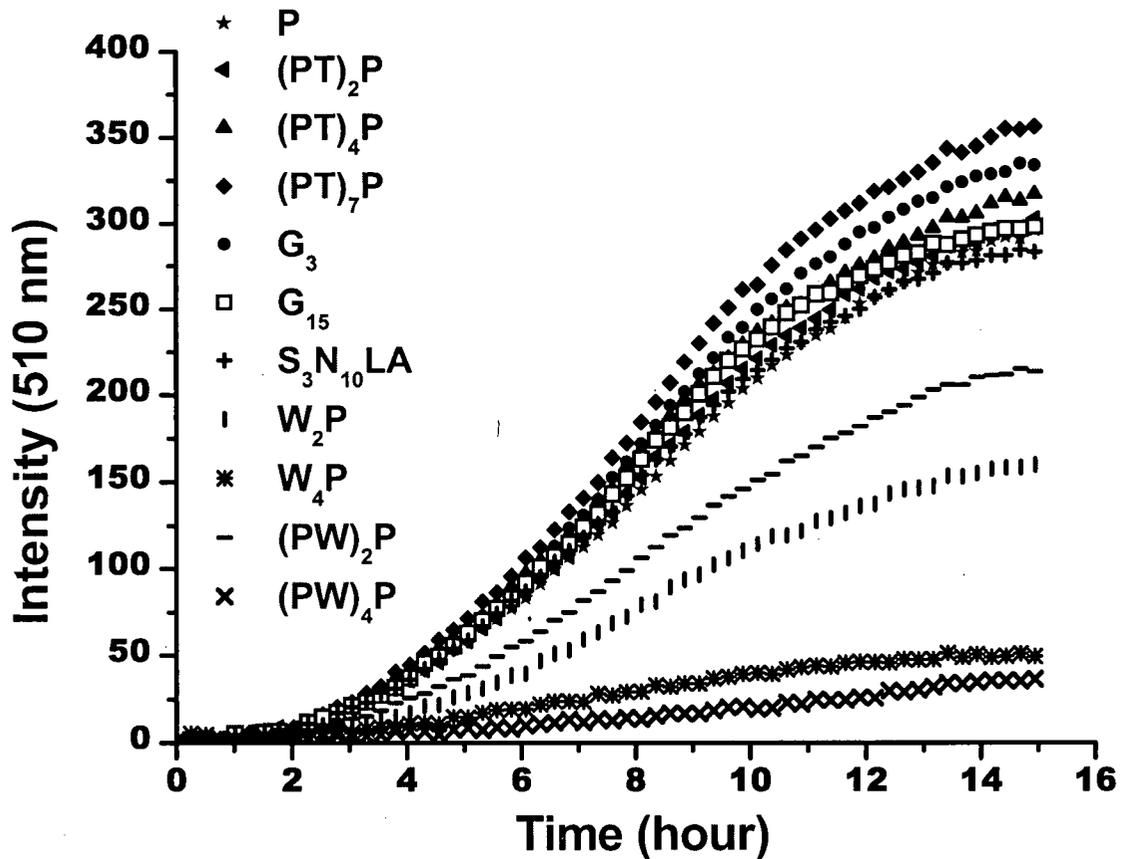


Figure 3.2 Intracellular Expression of CBM9-Linker-FX_a-GFP fusion proteins in *E. coli*. Cells expressing a CBM9-Linker-FX_a-GFP fusion protein were grown at 30°C to an OD₆₂₀ of ~0.9 and protein expression was induced with 0.1 mM IPTG. CBM9-Linker-FX_a-GFP expression was continuously monitored at 510 nm with excitation at 400 nm.

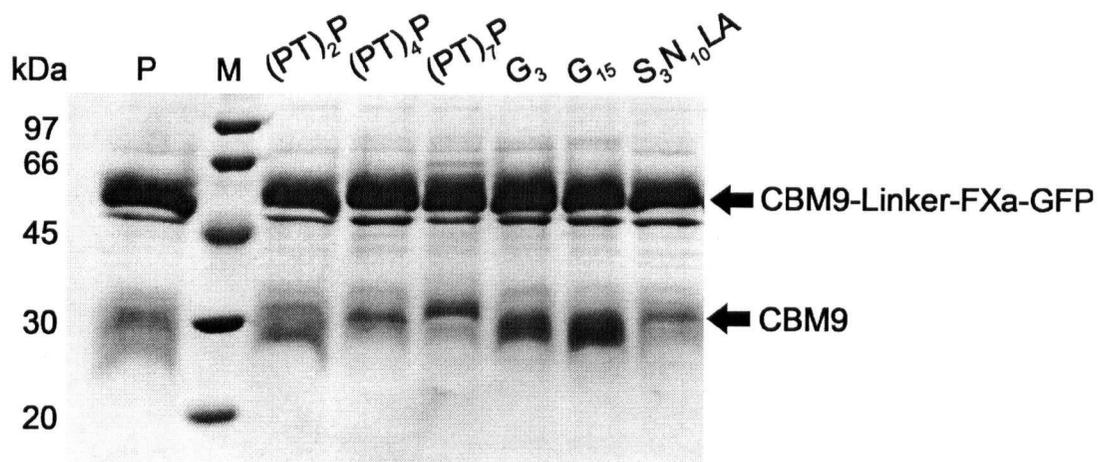


Figure 3.3 Long-term proteolytic stability of the linkers in CBM9-Linker-FX_a-GFP fusion proteins. Clarified cell lysates were incubated at 23°C for 4 hours and then the intact and degraded CBM9-Linker-FX_a-GFP fusion proteins were purified on Perloza™ MT100 in the absence of protease inhibitors. Each lane represents proteins eluted off the Perloza™ MT100 resin with elution buffer containing 3 M glucose. Lane M contains the molecular weight markers.

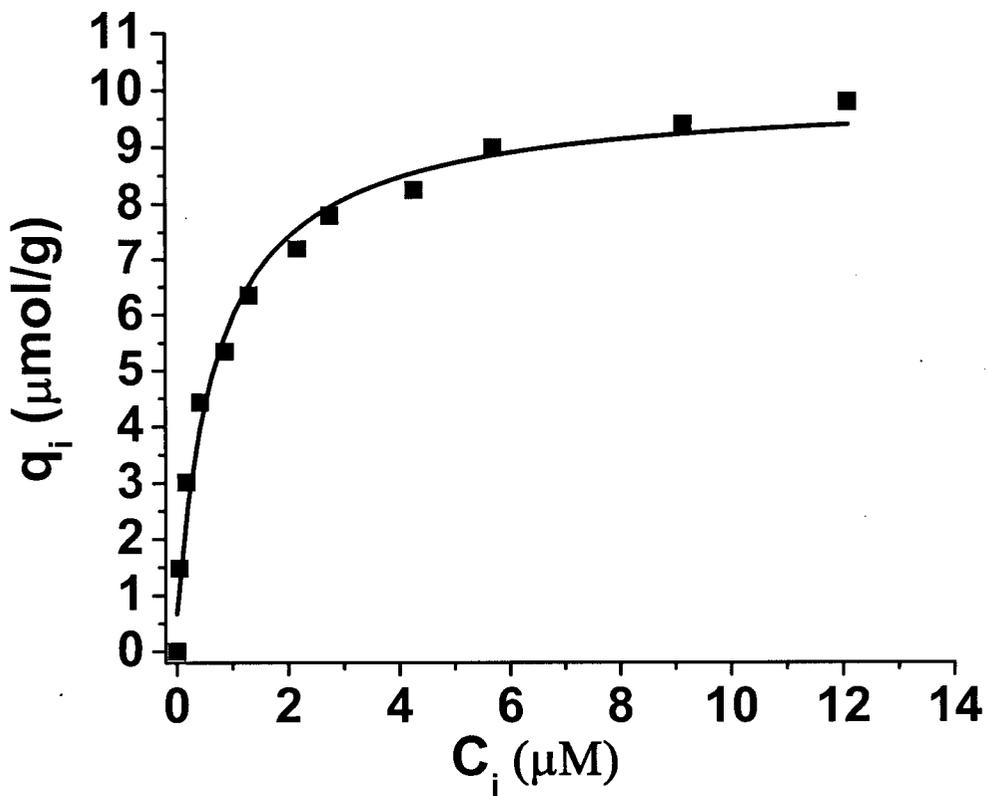


Figure 3.4 Equilibrium adsorption isotherm for binding of CBM9-G₃-IEGR-GFP to Perloza™ MT100. Isotherm measured at 4°C in high salt buffer (pH7; 50 mM potassium phosphate, 1 M NaCl).

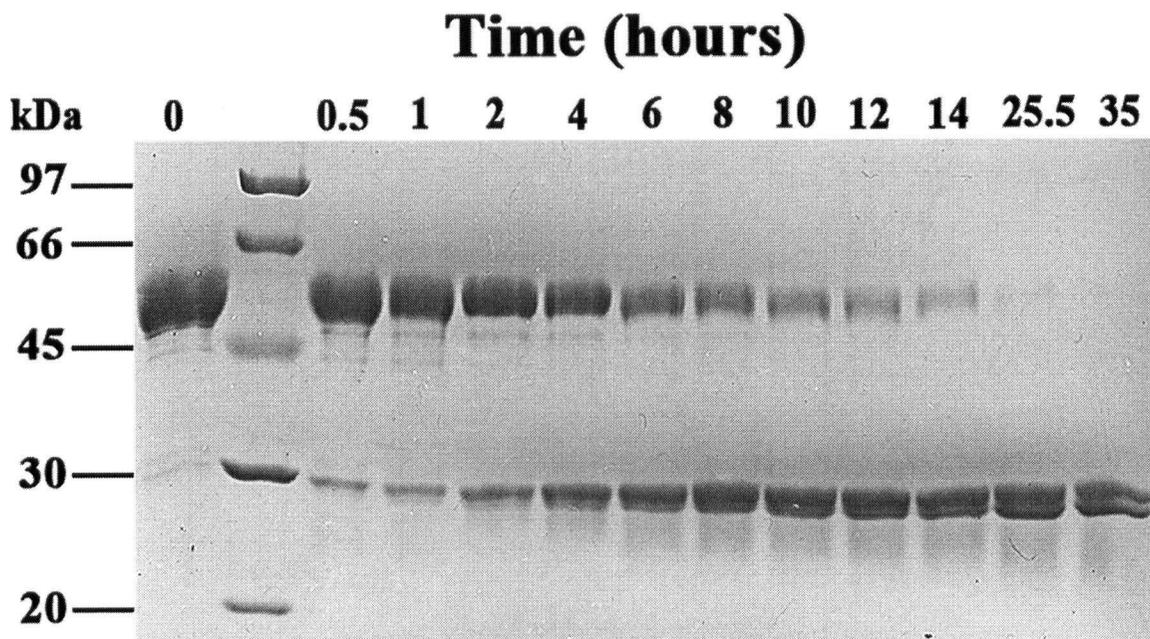


Figure 3.5 FX_a cleavage kinetics. CBM9-P-IEGR-GFP fusion protein was incubated at 23°C with Factor X_a at a fusion protein/ FX_a concentration ratio of 1000:1. Samples were taken at each time point indicated and analyzed by SDS-PAGE.

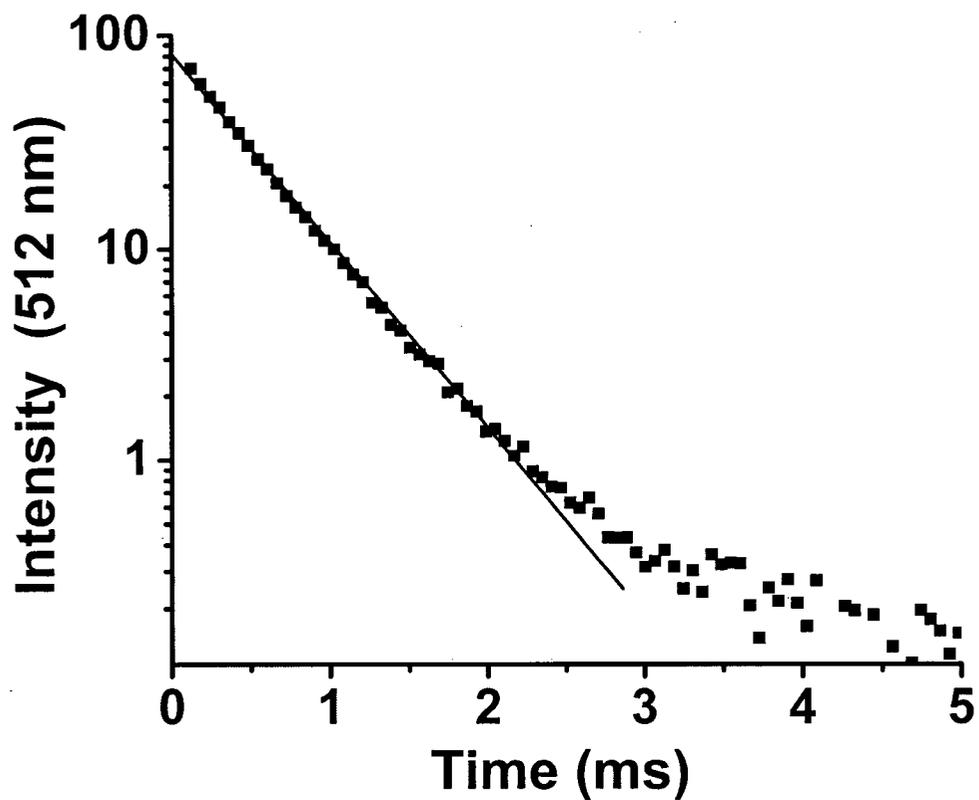


Figure 3.6 Sensitized fluorescence emission lifetime data for CBM9-(PT)₂P-IEGR-GFP at 23°C. Solid line indicates fit to the first-order exponential decay equation ($r^2 = 0.999$).

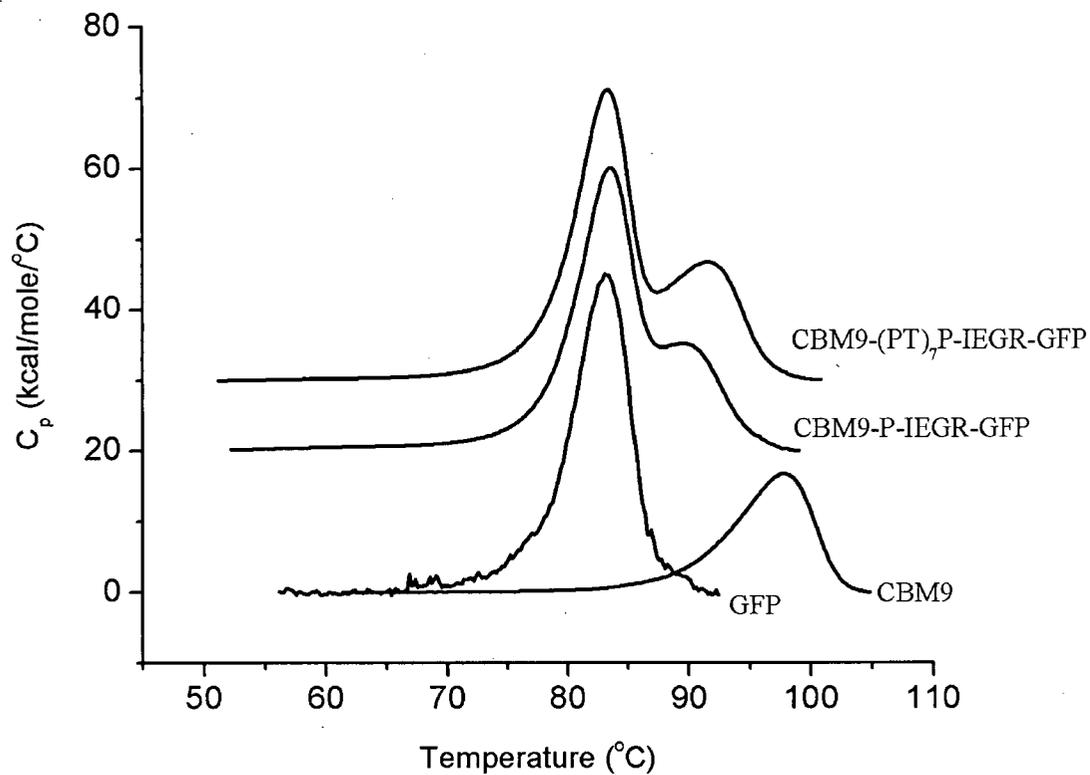


Figure 3.7 Differential scanning calorimetry thermograms for CBM9, GFP, CBM9-P-IEGR-GFP and CBM9-(PT)₇P-IEGR-GFP in 50 mM potassium phosphate buffer (pH 7). Area under the denaturation peak(s) provides the denaturation enthalpy through the relationship $\Delta H_{cal} = \int C_p(T)dT$, where C_p is the heat capacity.

3.7 References

- Ahn, J. O.; Choi, E. S.; Lee, H. W.; Hwang, S. H.; Kim, C. S.; Jang, H. W.; Haam, S. J.; Jung, J. K. (2004). "Enhanced secretion of *Bacillus stearothermophilus* L1 lipase in *Saccharomyces cerevisiae* by translational fusion to cellulose-binding domain". *Applied Microbiology and Biotechnology* 64(6):833-839.
- Argos, P. (1990). "An Investigation of Oligopeptides Linking Domains in Protein Tertiary Structures and Possible Candidates for General Gene Fusion". *Journal of Molecular Biology* 211(4):943-958.
- Arnau, J.; Lauritzen, C.; Petersen, G. E.; Pedersen, J. (2006). "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins". *Protein Expression and Purification* 48(1):1-13.
- Arvidson, B.; Seeds, J.; Webb, M.; Finlay, L.; Barklis, E. (2003). "Analysis of the retrovirus capsid interdomain linker region". *Virology* 308(1):166-177.
- Barrett, A. J.; Rawlings, N. D.; Woessner, J. F. (1998). *Handbook of Proteolytic Enzymes*. London: Elsevier.
- Beck, C.; Marty, R.; Klausli, S.; Hennecke, H.; Gottfert, M. (1997). "Dissection of the transcription machinery for housekeeping genes of *Bradyrhizobium japonicum*". *Journal of Bacteriology* 179(2):364-369.
- Beguin, P.; Aubert, J. P. (1994). "The Biological Degradation of Cellulose". *FEMS Microbiology Reviews* 13(1):25-58.
- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Boyd, S. E.; Pike, R. N.; Rudy, G. B.; Whisstock, J. C.; De La Banda, M. G. (2005). "PoPS: A computational tool for modeling and predicting protease specificity". *Journal of Bioinformatics and Computational Biology* 3(3):551-585.
- Brändén, C.-I.; Tooze, J. (1999). *Introduction to protein structure*. New York: Garland Publishing.
- Brinkmann, U.; Buchner, J.; Pastan, I. (1992). "Independent Domain Folding of *Pseudomonas* Exotoxin and Single-Chain Immunotoxins - Influence of Interdomain Connections". *Proceedings of the National Academy of Sciences of the United States of America* 89(7):3075-3079.

- Carlsson, H.; Ljung, S.; Bulow, L. (1996). "Physical and kinetic effects on induction of various linker regions in beta-galactosidase/galactose dehydrogenase fusion enzymes". *Biochimica et Biophysica Acta* 1293(1):154-160.
- Carter, P. (1990). "Site-Specific Proteolysis of Fusion Proteins". ACS Symposium Series 427:181-193.
- Chalfie, M.; Kain, S. (1998). *Green fluorescent protein : properties, applications, and protocols*. New York: Wiley-Liss.
- Cramer, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. (1996). "Improved green fluorescent protein by molecular evolution using DNA shuffling". *Nature Biotechnology* 14(3):315-319.
- Creagh, A. L.; Tiong, J. W. C.; Tian, M. M.; Haynes, C. A.; Jefferies, W. A. (2005). "Calorimetric studies of melanotransferrin (p97) and its interaction with iron". *Journal of Biological Chemistry* 280(16):15735-15741.
- Davis, G. D.; Elisee, C.; Newham, D. M.; Harrison, R. G. (1999). "New fusion protein systems designed to give soluble expression in *Escherichia coli*". *Biotechnology and Bioengineering* 65(4):382-388.
- Deyrup, A. T.; Krishnan, S.; Singh, B.; Schwartz, N. B. (1999). "Activity and stability of recombinant bifunctional rearranged and monofunctional domains of ATP-sulfurylase and adenosine 5'-phosphosulfate kinase". *Journal of Biological Chemistry* 274(16):10751-10757.
- di Guan, C.; Li, P.; Riggs, P. D.; Inouye, H. (1988). "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein". *Gene* 67(1):21-30.
- Einhauer, A.; Jungbauer, A. (2001). "The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins". *Journal of Biochemical and Biophysical Methods* 49(1-3):455-465.
- Fassina, G.; Merli, S.; Germani, S.; Ciliberto, G.; Cassani, G. (1994). "High-Yield Expression and Purification of Human Endothelin-1". *Protein Expression and Purification* 5(6):559-568.
- Fontana, A.; Zambonin, M.; deLaureto, P. P.; DeFilippis, V.; Clementi, A.; Scaramella, E. (1997). "Probing the conformational state of apomyoglobin by limited proteolysis". *Journal of Molecular Biology* 266(2):223-230.
- Ford, C. F.; Suominen, I.; Glatz, C. E. (1991). "Fusion tails for the recovery and purification of recombinant proteins". *Protein Expression and Purification* 2(2-3):95-107.

- Fukuoka, Y.; Tachibana, T.; Yasui, A. (1993). "Expression of Biologically-Active C3a as Fusion Proteins". *Immunology Letters* 38(2):153-158.
- Furusawa, T., Suzuki, M, Smith, J. M.. (1976). "Rate parameters in heterogeneous catalysis by pulse techniques." *Catalysis Reviews - Science and Engineering* 13(1):43-76.
- Gerdes, H.-H.;Kaether, C. (1996). "Green fluorescent protein: applications in cell biology". *FEBS Letters* 389(1):44-47.
- Gilkes, N. R.; Henrissat, B.; Kilburn, D. G.; Miller, R. C.;Warren, R. A. J. (1991). "Domains in Microbial Beta-1,4-Glycanases - Sequence Conservation, Function, and Enzyme Families". *Microbiological Reviews* 55(2):303-315.
- Gilkes, N. R.; Warren, R. A. J.; Miller, R. C.;Kilburn, D. G. (1988). "Precise Excision of the Cellulose Binding Domains from 2 Cellulomonas-Fimi Cellulases by a Homologous Protease and the Effect on Catalysis". *Journal of Biological Chemistry* 263(21):10401-10407.
- Greenwood, J. M.; Gilkes, N. R.; Kilburn, D. G.; Miller, R. C., Jr.;Warren, R. A. J. (1989). "Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose". *FEBS Letters* 244(1):127-131.
- Guan, K. L.;Dixon, J. E. (1991). "Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Gustavsson, M.; Lehtio, J.; Denman, S.; Teeri, T. T.; Hult, K.;Martinelle, M. (2001). "Stable linker peptides for a cellulose-binding domain-lipase fusion protein expressed in Pichia pastoris". *Protein Engineering* 14(9):711-715.
- Hampe, W.; Voss, R. H.; Haase, W.; Boege, F.; Michel, H.;Reilander, H. (2000). "Engineering of a proteolytically stable human beta(2)-adrenergic receptor/maltose-binding protein fusion and production of the chimeric protein in Escherichia coli and baculovirus-infected insect cells". *Journal of Biotechnology* 77(2-3):219-234.
- Hearn, M. T.;Acosta, D. (2001). "Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins". *Journal of Molecular Recognition* 14(6):323-369.
- Hochuli, E.; Dobeli, H.;Schacher, A. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues". *Journal of Chromatography* 411:177-184.

- Hogue, C. W. V.; Macmanus, J. P.; Banville, D.; Szabo, A. G. (1992). "Comparison of Terbium(III) Luminescence Enhancement in Mutants of Ef Hand Calcium-Binding Proteins". *Journal of Biological Chemistry* 267(19):13340-13347.
- Hopp, T., Prickett, K.S., Price, V.L., Libby, R.T., March, C.J., Ceretti, D.P., Urdal, D.L., Conlon, P.J. (1988). "A short polypeptide marker sequence useful for recombinant protein identification and purification". *Bio/Technology* 6:1204-1210.
- Huber, R.; Bennett, W. S., Jr. (1983). "Functional significance of flexibility in proteins". *Biopolymers* 22(1):261-279.
- Huston, J. S.; Levinson, D.; Mudgett-Hunter, M.; Tai, M. S.; Novotny, J.; Margolies, M. N.; Ridge, R. J.; Brucoleri, R. E.; Haber, E.; Crea, R.; Oppermann, H. (1988). "Protein Engineering of Antibody-Binding Sites - Recovery of Specific Activity in an Anti-Digoxin Single-Chain Fv Analog Produced in *Escherichia-Coli*". *Proceedings of the National Academy of Sciences of the United States of America* 85(16):5879-5883.
- Huston, J. S.; McCartney, J.; Tai, M. S.; Mottola-Hartshorn, C.; Jin, D.; Warren, F.; Keck, P.; Oppermann, H. (1993). "Medical applications of single-chain antibodies". *International Reviews of Immunology* 10(2-3):195-217.
- Huyton, T.; Pye, V. E.; Briggs, L. C.; Flynn, T. C.; Beuron, F.; Kondo, H.; Ma, J.; Zhang, X.; Freemont, P. S. (2003). "The crystal structure of murine p97/VCP at 3.6 Å". *Journal of Structural Biology* 144(3):337-348.
- Inoue, I.; Rechsteiner, M. (1994). "On the Relationship between the Metabolic and Thermodynamic Stabilities of T4 Lysozymes - Measurements in *Escherichia-Coli*". *Journal of Biological Chemistry* 269(46):29241-29246.
- Jalaguier, S.; Mesnier, D.; Leger, J. J.; Auzou, G. (1996). "Putative steroid binding domain of the human mineralocorticoid receptor, expressed in *E-coli* in the presence of heat shock proteins shows typical native receptor characteristics". *Journal of Steroid Biochemistry and Molecular Biology* 57(1-2):43-50.
- Jansson, B.; Palmcrantz, C.; Uhlen, M.; Nilsson, B. (1990). "A dual-affinity gene fusion system to express small recombinant proteins in a soluble form: expression and characterization of protein A deletion mutants". *Protein Engineering* 3(6):555-561.
- Jenny, R. J.; Mann, K. G.; Lundblad, R. L. (2003). "A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa". *Protein Expression and Purification* 31(1):1-11.

- Jiao, H.; Liao, F. H.; Tian, S. J.; Jing, X. P. (2003). "Luminescent properties of Eu³⁺ and Tb³⁺ activated Zn₃Ta₂O₈". *Journal of the Electrochemical Society* 150(9):H220-H224.
- Johnson, K.; Murphy, C. K.; Beckwith, J. (1992). "Protein export in *Escherichia coli*". *Current Opinion in Biotechnology* 3(5):481-485.
- Kain, S. R.; Adams, M.; Kondepudi, A.; Yang, T. T.; Ward, W. W.; Kitts, P. (1995). "Green Fluorescent Protein as a Reporter of Gene-Expression and Protein Localization". *Biotechniques* 19(4):650-655.
- Kapust, R. B., Waugh, D. S. (1999). "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused". *Protein Science* 8(8):1668-1674.
- Kavoosi, M.; Meijer, J.; Kwan, E.; Creagh, A. L.; Kilburn, D. G.; Haynes, C. A. (2004). "Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T-maritima*". *Journal of Chromatography B* 807(1):87-94.
- Knowles, J.; Lehtovaara, P.; Penttila, M.; Teeri, T.; Harkki, A.; Salovuori, I. (1987). "The Cellulase Genes of *Trichoderma*". *Antonie Van Leeuwenhoek Journal of Microbiology* 53(5):335-341.
- Kwan, E.; Guarna, M. M.; Boraston, A. B.; Gilkes, N. R.; Haynes, C. A.; Kilburn, D. G.; Warren, R. A. (2002). "Self-activating factor X derivative fused to the C-terminus of a cellulose-binding module: Production and properties". *Biotechnology and Bioengineering* 79(7):724-732.
- Kwon, W. S.; Da Silva, N. A.; Kellis, J. T., Jr. (1996). "Relationship between thermal stability, degradation rate and expression yield of barnase variants in the periplasm of *Escherichia coli*". *Protein Engineering* 9(12):1197-1202.
- LaVallie, E. R.; DiBlasio, E. A.; Kovacic, S.; Grant, K. L.; Schendel, P. F.; McCoy, J. M. (1993). "A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm". *Bio/Technology* 11(2):187-193.
- Libby, R. T.; Cosman, D.; Cooney, M. K.; Merriam, J. E.; March, C. J.; Hopp, T. P. (1988). "Human Rhinovirus 3C Protease - Cloning and Expression of an Active Form in *Escherichia-Coli*". *Biochemistry* 27(17):6262-6268.
- Lucius, R.; Kern, A.; Seeber, F.; Pogonka, T.; Willenbacher, J.; Taylor, H. R.; Pinder, M.; Ghalib, H. W.; Schulzkey, H.; Soboslay, P. (1992). "Specific and Sensitive Igg4 Immunodiagnosis of Onchocerciasis with a Recombinant 33kd *Onchocerca-Volvulus* Protein (Ov33)". *Tropical Medicine and Parasitology* 43(3):139-145.

- Maeda, Y.; Ueda, H.; Kazami, J.; Kawano, G.; Suzuki, E.; Nagamune, T. (1997). "Engineering of functional chimeric protein G - *Vargula luciferase*". *Analytical Biochemistry* 249(2):147-152.
- Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in *Escherichia coli*". *Microbiological Reviews* 60(3):512-538.
- Martinez, A.; Knappskog, P. M.; Olafsdottir, S.; Doskeland, A. P.; Eiken, H. G.; Svebak, R. M.; Bozzini, M.; Apold, J.; Flatmark, T. (1995). "Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases: Isolation and characterization of the wild-type enzyme". *Biochemical Journal* 306(2):589-597.
- Mclendon, G.; Radany, E. (1978). "Protein Turnover Thermodynamically Controlled". *Journal of Biological Chemistry* 253(18):6335-6337.
- Moks, T., Abrahmsen, L., Nilsson, B., Hellman, U., Sjoquist, J., Uhlen, M. (1986). "Staphylococcal protein A consists of five IgG-binding domains". *European Journal of Biochemistry* 156(3):637-643.
- Murashima, K.; Kosugi, A.; Doi, R. H. (2003). "Solubilization of cellulosomal cellulases by fusion with cellulose-binding domain of noncellulosomal cellulase EngD from *Clostridium cellulovorans*". *Proteins-Structure Function and Genetics* 50(4):620-628.
- Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A., Uhlen, M. (1987). "A synthetic IgG-binding domain based on staphylococcal protein A". *Protein Engineering* 1(2):107-113.
- Notenboom, V.; Boraston, A. B.; Kilburn, D. G.; Rose, D. R. (2001). "Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms". *Biochemistry* 40(21):6248-6256.
- Ong, E., Alimonti, J. B., Greenwood, J. M., Miller, R. C., Jr., Warren, R. A., Kilburn, D. G. (1995). "Purification of human interleukin-2 using the cellulose-binding domain of a prokaryotic cellulase". *Bioseparation* 5(2):95-104.
- Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M. (1997). "CATH - a hierarchic classification of protein domain structures". *Structure* 5(8):1093-1108.
- Ozawa, T. (2006). "Designing split reporter proteins for analytical tools". *Analytica Chimica Acta* 556(1):58-68.

- Parsell, D. A.; Sauer, R. T. (1989). "The Structural Stability of a Protein Is an Important Determinant of Its Proteolytic Susceptibility in *Escherichia-Coli*". *Journal of Biological Chemistry* 264(13):7590-7595.
- Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation". *Nature* 258(5536):598-599.
- Rawlings, N. D.; Morton, F. R.; Barrett, A. J. (2006). "MEROPS: the peptidase database". *Nucleic Acids Research* 34:D270-D272.
- Root, D. D.; Shangguan, X.; Xu, J.; McAllister, M. A. (1999). "Determination of Fluorescent Probe Orientations on Biomolecules by Conformational Searching: Algorithm Testing and Applications to the Atomic Model of Myosin". *Journal of Structural Biology* 127(1):22-34.
- Sambrook, J.; Russell, D. W. (2001). *Molecular cloning : a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.
- Samuelsson, E.; Moks, T.; Nilsson, B.; Uhlen, M. (1994). "Enhanced in-Vitro Refolding of Insulin-Like Growth-Factor-I Using a Solubilizing Fusion Partner". *Biochemistry* 33(14):4207-4211.
- Sauer, J., Christensen, T., Frandsen, T. P., Mirgorodskaya, E., McGuire, K. A., Driguez, H., Roepstorff, P., Sigurskjold, B. W., Svensson, B. (2001). "Stability and function of interdomain linker variants of glucoamylase 1 from *Aspergillus niger*". *Biochemistry* 40(31):9336-9346.
- Selvin, P. R. (2002). "Principles and biophysical applications of lanthanide-based probes". *Annual Review of Biophysics and Biomolecular Structure* 31:275-302.
- Selvin, P. R.; Hearst, J. E. (1994). "Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer". *Proceedings of the National Academy of Sciences of the United States of America* 91(21):10024-10028.
- Sheffield, P.; Garrard, S.; Derewenda, Z. (1999). "Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors". *Protein Expression and Purification* 15(1):34-39.
- Shen, H., Schmuck, M., Pilz, I., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A. (1991). "Deletion of the linker connecting the catalytic and cellulose-binding domains of endoglucanase A (CenA) of *Cellulomonas fimi* alters its conformation and catalytic activity". *Journal of Biological Chemistry* 266(17):11335-11340.

- Shih, Y. P.; Wu, H. C.; Hu, S. M.; Wang, T. F.; Wang, A. H. J. (2005). "Self-cleavage of fusion protein in vivo using TEV protease to yield native protein". *Protein Science* 14(4):936-941.
- Shimomura, O.; Johnson, F. H.; Saiga, Y. (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*". *Journal of Cellular and Comparative Physiology* 59:223-239.
- Smith, D. B.; Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase". *Gene* 67(1):31-40.
- Staahl, S.; Nygren, P. A. (1997). "The use of gene fusions to protein A and protein G in immunology and biotechnology". *Pathologie Biologie* 45(1):66-76.
- Tang, Y.; Jiang, N.; Parakh, C.; Hilvert, D. (1996). "Selection of linkers for a catalytic single-chain antibody using phage display technology". *Journal of Biological Chemistry* 271(26):15682-15686.
- Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology* 60(5):523-533.
- Tomme, P.; Gilkes, N. R.; Miller, R. C., Jr.; Warren, R. A. J.; Kilburn, D. G. (1994). "An internal cellulose-binding domain mediates adsorption of an engineered bifunctional xylanase/cellulase". *Protein Engineering* 7(1):117-123.
- Tomme, P.; Vantilbeurgh, H.; Pettersson, G.; Vandamme, J.; Vandekerckhove, J.; Knowles, J.; Teeri, T.; Claeysens, M. (1988). "Studies of the Cellulolytic System of *Trichoderma-Reesei* Qm-9414 - Analysis of Domain Function in 2 Cellobiohydrolases by Limited Proteolysis". *European Journal of Biochemistry* 170(3):575-581.
- Uhlen, M.; Moks, T. (1990). "Gene fusions for purpose of expression: an introduction". *Methods in Enzymology* 185(Gene Expression Technology):129-143.
- Vaillancourt, P.; Simcox, T. G.; Zheng, C. F. (1997). "Recovery of polypeptides cleaved from purified calmodulin-binding peptide fusion proteins". *Biotechniques* 22(3):451-453.
- Wheelan, S. J.; Marchler-Bauer, A.; Bryant, S. H. (2000). "Domain size distributions can predict domain boundaries". *Bioinformatics* 16(7):613-618.
- Winterhalter, C.; Heinrich, P.; Candussio, A.; Wich, G.; Liebl, W. (1995). "Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*". *Molecular Microbiology* 15(3):431-444.

Wootton, J. C.; Drummond, M. H. (1989). "The Q-Linker - a Class of Interdomain Sequences Found in Bacterial Multidomain Regulatory Proteins". *Protein Engineering* 2(7):535-543.

Wriggers, W.; Chakravarty, S.; Jennings, P. A. (2005). "Control of protein functional dynamics by peptide linkers". *Biopolymers* 80(6):736-746.

4 Direct Measurement of the Kinetics of CBM9 Fusion-Tag Bioprocessing Using Luminescence Resonance Energy Transfer (LRET)

4.1 Introduction

Affinity-tag technology, where a target protein or peptide is produced as a recombinant fusion to an N- or C-terminal affinity tag possessing a highly specific binding partner that can serve as a stable ligand for affinity chromatography, is a general and potentially powerful approach to streamline production of recombinant therapeutics and other biologics (Lowe et al. 2001). In addition to facilitating a highly selective separation, affinity tags, particularly those comprised of carbohydrate binding modules, can improve product solubility, increase *in vivo* proteolytic stability, and control product localization in or secretion from the host organism (Hearn and Acosta 2001).

Recently we introduced the CBM9 affinity tag (Kavoosi et al. 2004), which is based on the C-terminal family 9 carbohydrate binding module of xylanase 10A from *Thermotoga maritima* (Winterhalter et al. 1995). CBM9 exhibits strong and specific affinity for both insoluble cellulose and soluble sugars such as glucose (Boraston et al. 2001). Affinity purification of CBM9-tagged recombinant proteins on a commercially available cellulosic media, Perloza™ MT100, results in highly pure, concentrated products recovered at high yields (Kavoosi et al. 2004). Due to the low cost and high binding capacity of Perloza™ MT100 media, the CBM9 affinity system provides a considerable economic advantage over other commercially available affinity tag technologies.

* A version of this chapter has been prepared for submission to *Biotechnology and Bioengineering*. [Reference: Mojgan Kavoosi, A. Louise Creagh, Douglas G. Kilburn, Charles A. Haynes, Direct measurement of the kinetics of CBM9 fusion-tag bioprocessing using Luminescence Resonance Energy Transfer (LRET).

However, the cost competitiveness of a given affinity-tag technology, particularly at preparative scales, is not solely determined by the cost and maximum cycle number of the capture column. It also depends on the cost and efficiency of the post-purification step required to remove the affinity tag and obtain the desired final product with its natural N- and C-termini. This can be an expensive step, with its economics and yield known to depend on the choice of processing enzyme and the local structure of the cleavage site, most notably the peptide sequence (often denoted the linker sequence) flanking the recognition site on the tag side (Girard et al. 2006; Jenny et al. 2003; Lien et al. 2001). For example, thrombin-catalyzed processing of glutathione-S-transferase (GST) tagged tyrosine phosphatase is greatly enhanced by the incorporation of a glycine linker immediately upstream of the recognition site (Guan and Dixon 1991). The lack of studies on and related methods for selecting the optimal processing enzyme and linker sequence for a particular affinity-tagged protein product is therefore surprising, and points to the need for a simple and effective spectroscopic method to screen a candidate library of processing-enzyme/linker-sequence combinations or, in cases where regulatory or IP constraints predetermine the processing enzyme, a library of linker sequences. A pre-validation screening method of this type could guide bioprocess design and speed translation to the manufacturing scale in a manner that increases downstream process throughput while decreasing cost of goods.

Several recent studies have reported on the use of fluorescence (or Förster) resonance energy transfer (FRET) to measure enzyme activity, and on the potential of FRET in high-throughput kinetic screens (Boskovic et al. 2004; Chen et al. 2005; Cotrin et al. 2004; Medintz et al. 2006; Nishikata et al. 2006; Thomas et al. 2006). In FRET, distances between 15 and 80 Å can be measured through the distance-dependent energy transfer between a donor and an acceptor fluorophore. This energy transfer leads to measurable changes in donor intensity or excited-state lifetime, as well as to spectral changes in the acceptor dye. Placement of the donor and acceptor fluorophores on opposite sides of an enzymatic-cleavage site on a substrate thereby provides a method for real-time monitoring of hydrolysis kinetics. However, while FRET has proven useful, it suffers from well-documented limitations. The maximum distance that can be measured,

which is a direct result of the small and relatively inefficient energy transfer between donor and acceptor, is less than optimal for many biological applications (Selvin 1996). Measurement of the sensitized emission of the acceptor dye is difficult due to a low signal-to-background ratio resulting from interfering fluorescence emission from the donor and direct excitation of the acceptor fluorophore (Selvin 1995). Finally, donor fluorophores used in FRET generally exhibit short (nanosecond) multiexponential lifetimes, necessitating the use and careful synchronization of fast electronics and detectors, which can be difficult and expensive (Selvin 1995).

Luminescence resonance energy transfer (LRET), which overcomes many of the limitations of FRET (Selvin 2002), is a related spectroscopic method that has not been applied to real-time monitoring of hydrolytic enzyme activity. In LRET, a luminescent lanthanide chelate is used in place of a traditional fluorophore as the donor, while the acceptor remains a standard organic dye. Lanthanide emission arises from high-spin to high-spin transitions, and is therefore not related to the singlet-to-singlet transitions characteristic of fluorescence. Excitation of lanthanide ions produces very narrow, non-polarized and strongly Stokes-shifted emission bands with extremely long luminescence lifetimes (milliseconds) that can serve to excite the acceptor fluorophore over a sustained period. As in FRET, direct excitation of the acceptor fluorophore can occur. However, the lifetime of this fluorescence emission is on the order of nanoseconds, and any acceptor emission after this very short-lived event will only be due to energy transfer received by the acceptor from the long-lived lanthanide donor. The resulting high signal-to-background ratio of the LRET measurement provides very significant advantages over the more common FRET experiment. In particular, sub-femtomolar detection limits are possible (Sammes and Yahioğlu 1996), allowing highly accurate kinetic data to be obtained.

Although it has not been used to screen for peptidase specificity and activity, LRET has been used in a number of other applications, including the measurement of conformational changes on the nanometer scale (Callaci et al. 1999; Kapanidis et al. 2001; Xiao et al. 1998) during formation of protein-protein (Kolb et al. 1997; Mathis 1995; Root 1997) and protein-DNA complexes (Churchich 1997; Heyduk et al. 1997). It

has also been used to measure inter- and intramolecular distances between protein domains (Burmeister-Getz et al. 1998), and to characterize structures of duplex DNA (Kapanidis et al. 2001; Selvin and Hearst 1994). Finally, the lanthanides terbium and europium have been extensively used to explore the mechanism of metal-ion binding both to DNA and to metalloproteins (Chaudhuri et al. 1997; Veenstra et al. 1995).

Here, we introduce a LRET-based assay designed to rapidly and accurately measure CBM9 fusion-tag bioprocessing kinetics and their dependence on the choice of linker sequence and processing enzyme. The assay is based on the nonradiative energy transfer between lanthanide-based donors specifically bound to the CBM9 tag and an acceptor fluorophore presented on the target protein or peptide, and thus, on the opposite side of the processing site. Enzyme-catalyzed cleavage of the fusion tag then terminates the resonance energy transfer, resulting in a change in the fluorescence intensity of the sensitized acceptor that can be continuously monitored to quantify substrate concentration over time. CBM9 is particularly well suited for this assay, as it is a calcium (II) binding metalloprotein (Notenboom et al. 2001) capable of binding the terbium (III) lanthanide ion with high affinity in the absence of Ca^{2+} . This Tb^{3+} -binding property can therefore be combined with a number of elegant and efficient methods that have recently been developed to specifically label the C-terminus of a protein with a fluorophore (Becker et al. 2006; Chao et al. 1998; Dursina et al. 2005; Tripet et al. 1996) to provide a general LRET-based assay of the kinetics of CBM9 fusion-tag bioprocessing. Our aim in this work is to prove that such a LRET-based assay can provide accurate kinetic data sets for fusion-tag bioprocessing, allowing one to identify both major and subtle differences in rates of hydrolysis. To do this, we chose as the fusion partner the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* (Shimomura et al. 1962), thereby exploiting the fact that the target protein, GFP, endogenously provides the required fluorescent acceptor group. This permitted the rapid preparation of a library of chimeric proteins (Figure 4.1) comprised of an N-terminal Tb^{3+} -loaded CBM9 (hereafter referred to as Tb^{3+} -CBM9) fused to GFP via a cleavage-site terminated linker sequence specific to the popular processing enzyme enterokinase. Our LRET-based screen was applied to the library to quantify differences in enterokinase

cleavage activity due to changes in the linker sequence immediately upstream of the enterokinase recognition element, thereby providing proof-of-concept of the assay.

4.2 Methods and Materials

4.2.1 Materials

All chemicals and reagents, including $TbCl_3$, were purchased from either Sigma-Aldrich (St. Louis, MO) or Fisher (Pittsburgh, PA) Chemicals and of analytical grade unless otherwise stated. Recombinant enterokinase (EKmaxTM) was purchased from Invitrogen Corp. (Carlsbad, CA). The concentration of EKmaxTM was determined using the extinction coefficient provided by the vendor and verified using the commercial Bradford Protein Assay of BioRad Laboratories (Hercules, CA) using BSA as the standard.

4.2.2 Cloning and purification of fusion protein substrates

All cloning was performed using standard molecular biology techniques (Sambrook and Russell 2001). Previously prepared pET28-CBM9-GFP expression vectors (Kavoosi et al. 2007) were digested with Pvu I and Not I to excise the GFP coding sequence, which was then amplified using the vector pGFPuv (Clontech, Palo Alto, CA). The enterokinase recognition sequence (D₄K) (**bold**) and a Pvu I restriction endonuclease site (*italic*) were introduced at the 5' end of GFP using the oligonucleotide 5'-AAAGCGATCGATGACGACGACAAGGCCATGAGTAA-AGGAGAAGAACT-3' as primer. A Not I site (*italic*) was introduced at the 3' end using the oligonucleotide 5'-TGCGGCCGCTTTGTAGAGCTCATCCATGCCATGTGTAATCCC-3' as primer using a PCR protocol previously described (Kavoosi et al. 2004). The compatible pET28-CBM9-Linker expression vectors comprising the library of CBM9-Linker coding sequences were prepared in a similar primer-assisted manner to allow the D₄K-GFP coding region to be digested with Pvu I/Not I and ligated (16°C, 16 h) into the digested expression vectors to give the appropriate pET28-CBM9-linker-D₄K-GFP expression vector. DNA sequencing was performed to verify all constructs (NAPS Unit, Biotechnology Laboratory, The University of British Columbia). The vectors were

transformed into *E. coli* BL21 (DE3) cells and fusion protein expression and purification were performed as previously described (Kavoosi et al. 2007).

4.2.3 Preparation of apo-fusion protein

Each purified fusion protein was incubated with Chelex 100X resin (Sigma-Aldrich; St. Louis, MO), mixing end-over-end at 4°C for 48 h to strip all bound metal from the calcium binding sites of CBM9. The apo-protein was then passed over a Sephadex G10 size-exclusion column using nano-pure water as solvent and mobile phase to remove any trace contaminants. Each purified apo-fusion protein was stored at 4°C in nano-pure water prior to use in the LRET-based assay.

4.2.4 Isothermal titration calorimetry studies

Isothermal titration calorimetry (ITC) was performed using a VP ITC (MicroCal, Inc., Northampton MA). All samples were pH 6.9 in 20-mM Hepes or Tris buffer, 50-mM NaCl. Titrations were performed by injecting consecutive 5- μ L aliquots of TbCl₃ solution (0.60 mM) into an ITC cell (volume = 1.3528 mL) containing apo-CBM9-GFP (20 μ M). The ITC data were corrected for the heat of dilution of the titrant by subtracting mixing enthalpies for 5- μ L injections of TbCl₃ solution into protein-free buffer. Six and three independent titration experiments were performed for linkers PID₄K and (PT)₂PID₄K, respectively; single titrations were also performed for linkers (PT)₄PID₄K, (PT)₇PID₄K, G₃ID₄K, G₁₅ID₄K, and S₃N₁₀LAID₄K. All titrations were at 25 °C. Binding stoichiometry, enthalpy and equilibrium association constants were determined by fitting the corrected data to a model for two independent binding sites on CBM9-GFP (software provided with instrument).

4.2.5 LRET-based assay of fusion-tag cleavage kinetics

Absorbance spectra for solvated and CBM9-bound lanthanide ion (Tb(III)) were measured in a quartz cuvette over the range of wavelengths 300 to 700 nm on a Cary Eclipse fluorescence spectrophotometer equipped with a xenon pulsed lamp and a Czerny-Turner monochromator. Wavelength scans were performed at a rate of 1 nm s⁻¹.

Enterokinase-catalyzed cleavage of the CBM9 fusion tag can be described by the Michaelis-Menten equation (Boulware and Daugherty 2006; Gasparian et al. 2003), where the reaction rate (v) is related to the substrate (Tb^{3+} -CBM9-linker-ID₄K-GFP) concentration $[S]$ by

$$v = -\frac{d[S]}{dt} = \frac{v_{\max} [S]}{K_M + [S]} \quad (4.1)$$

where v_{\max} is the maximum reaction rate, and K_M is the Michaelis constant. At low substrate concentrations, where $[S] \ll K_M$, equation 4.1 reduces to an apparent first order reaction

$$v = \frac{v_{\max} [S]}{K_M} = \left(\frac{k_{\text{cat}}}{K_M} \right) [E_T] [S] \quad (4.2)$$

where $[E_T]$ is the total enzyme concentration, k_{cat} is the catalytic constant, and the ratio k_{cat}/K_M is a measure of substrate specificity. Equation 4.2 can be solved by integration to give the change in $[S]$ with time

$$\ln(\phi) = \ln\left(\frac{[S]}{[S]_0}\right) = -k_{\text{obs}} t = -\frac{k_{\text{cat}} [E_T]}{K_M} t \quad (4.3)$$

where k_{obs} is the observed rate constant, and ϕ is the fraction of substrate remaining, which is directly measured by our LRET-based assay and which can be used to determine the initial reaction rate v_0 at each initial substrate concentration $[S]_0$.

Our LRET-based assay was used to acquire initial reaction-rate (v_0 and $\phi(t)$) data for $[S]_0$ values ranging from 1 μM to 48 μM . For each $[S]_0$, k_{cat}/K_M was estimated by linear regression of equation 4.3 to individual $\ln(\phi)$ versus t data sets, and the results (duplicates) were averaged to provide the reported k_{cat}/K_M and associated standard deviation. A secondary check of the k_{cat}/K_M value was obtained from the slope of a Michaelis-Menten plot as $[S]$ approaches zero. If the two estimates of k_{cat}/K_M were consistent, a Lineweaver-Burke plot was constructed to obtain a rough estimate of K_M to

confirm that $[S] \ll K_M$. The basic LRET-based assay proceeded as follows. Chelex-prepared apo-fusion protein was incubated with 5x excess terbium in reaction buffer (20 mM HEPES, 50 mM NaCl, pH 6.9) for 15-30 min at room temperature, then centrifuged at 27,000 x g for 10 min to remove any precipitates. Tb(III)-loaded fusion protein (40 μ L) at a concentration $2 \times [S]_0$ was then mixed with an equivalent volume of EKmaxTM (0.74 pmole) solution to reach a final reaction volume of 80 μ L. All reactions were carried out at 30°C in a 384-well microplate (Corning Life Sciences; Big Flats, NY) and monitored on a POLARstar Optima microplate reader (BMG Lab Technologies, Inc.; Germany), a high-performance microplate reader designed for luminescence/fluorescence-based applications (note that the instrument used in this chapter differs from that of chapter 3). The system integrates a high-energy xenon flash lamp with an excitation and emission filter system to generate a clean reproducible signal and maximal sensitivity (filter sets for this instrument required excitation at 235 nm and emission at 520 nm). Black, low binding, 384-well plates (Corning Life Sciences; Big Flats, NY) were used. Samples were excited at 235 nm and sensitized emissions were measured at 520 nm for 1500 μ s following a 200 μ s post-excitation delay to eliminate background noise introduced by direct excitation of the acceptor fluorophore, GFP. Integration of the resulting fluorescence emission signal was then used to obtain the average fluorescence intensity at the given reaction time point. For each well in the microtitre plate, a reading was taken every 15 seconds.

Raw fluorescence data were baseline corrected for losses in the fluorescence intensity of the unreacted Tb³⁺-CBM9-linker-ID₄K-GFP control due to bleaching and other factors. The fluorescence intensity corresponding to $[S]_0$ was determined from EKmaxTM-free control wells to permit construction of the required $\ln(\phi)$ versus t plots. The kinetic parameters reported are the means of two independent experiments with the error computed as the propagation of this average error and the error in the linear fit of the data. The half-life $t_{1/2}$ (h) of each reaction was determined from $t_{1/2} = \ln 2 / k_{obs}$.

4.2.6 Gel-based analysis of EKmax™ cleavage reaction

Tb³⁺-CBM9-PID₄K-GFP fusion protein (32 μM) was prepared as described above, mixed with EKmax™ (0.74 pmole) in reaction buffer to a final volume of 80 μL and incubated at 30°C in an Eppendorf Thermomixer R (Hamburg, Germany) with shaking (5 s every hour). Reaction buffer was used in place of EKmax™ in the control experiment. At each time point (0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 and 24 h after initial mixing), 2 μL of the supernatant was removed and diluted with 2x loading buffer for analysis by SDS-PAGE to determine the percentage of cleaved fusion protein with time, from which the half-life $t_{1/2}$ (h) of each cleavage reaction was estimated.

4.3 Results and Discussion

Terbium (III), like all lanthanide ions, is characterized by an emission spectrum comprised of sharp, fully-resolved peaks with exceptionally long decay times (Sammes and Yahioğlu 1996). Unfortunately, the molar absorptivity of Tb³⁺ is extremely low. Excitation of water-solvated Tb³⁺ can therefore only be accomplished with an intense laser light source. This problem can be overcome by binding the lanthanide ion to a sensitizing agent, such as CBM9, possessing a high molar absorptivity and an excited triplet state comparable to that of the emissive state of the lanthanide metal ion (Selvin 1996). Energy absorbed by the sensitizing agent may then be transferred to the terbium ion. The sensitizer must also function to shield the ion from the quenching effects of water (Parker and Williams 2003). Through its Ca⁺²-ion binding sites (Notenboom et al. 2001), the CBM9 affinity tag provides coordination sites for binding of up to three terbium (III) ions. Tb³⁺ ion binding to each of these sites on the apo-CBM9-PID₄K-GFP fusion protein was characterized by isothermal titration calorimetry (ITC) experiments, which reveal the presence of a single strong Tb³⁺ binding site ($K_a = 2 (\pm 3) \times 10^8 \text{ M}^{-1}$) and two somewhat weaker Tb³⁺ sites ($K_a = 2 (\pm 2) \times 10^5 \text{ M}^{-1}$). Essentially identical ITC results were obtained for the binding of Tb³⁺ to apo-CBM9-linker-ID₄K-GFP fusion proteins containing each of the other linker sequences. Based on these K_a values, all three binding sites are saturated with Tb³⁺ under the loading conditions used in this study.

Figure 4.1 shows the schematic diagram of the fusion protein analyzed in this LRET-based assay, with the linker sequences that make up the fusion-protein library listed in Table 4.1. The bound terbium and GFP correspond to the LRET donor and acceptor, respectively. For resonance energy transfer to occur between the donor and the acceptor, two conditions must be met: (i) the emission spectrum of the donor must overlap with the absorbance spectrum of the acceptor, and (ii) the donor/acceptor pair must be separated by a distance no greater than the wavelength of light λ emitted by the sensitized donor that is frequency matched to the corresponding absorbance peak of the acceptor (in this case meaning a distance less than 490 nm). Under these conditions, the excited terbium transfers a fraction of its energy to GFP via a long-range dipole-dipole coupling mechanism (Heyduk 2001), allowing real time monitoring of the decrease in fluorescence intensity of the acceptor resulting from cleavage of the CBM9-linker-ID₄K affinity tag.

4.3.1 Characterization of energy transfer from bound Tb(III) to GFP

The spectral characteristics of both the donor (Tb³⁺-CBM9 on its own) and acceptor (GFP on its own) are shown in Figure 4.2. Excitation of CBM9-bound Tb³⁺ at 222 nm results in four sharp emission peaks with maxima at 490 nm, 546 nm, 585 nm and 622 nm, respectively. Comparing this luminescence emission spectrum with the absorbance spectra of GFP (acceptor fluorophore) reveals an overlap between the Tb³⁺ emission peak at 490 nm and the minor GFP absorption peak at 475 nm. The spectral overlap between CBM9-bound Tb³⁺ and GFP indicates that the energy levels of the two match in frequency, allowing internal transfer of energy from excited terbium to GFP.

Figure 4.3 shows the fluorescence emission spectrum for the acceptor resulting from energy transfer from the sensitized donor. Excitation of CBM9-bound terbium at 235 nm produces an oscillating electric dipole field. In the absence of a suitable acceptor, the field energy decays in the form of radiative energy, shown by the four characteristic luminescent emission peaks of Tb(III)-CBM9 in Figure 4.3. However, in the intact fusion protein (donor and acceptor), a fraction of the electric field energy of the sensitized donor is nonradiatively transferred to the acceptor (Selvin 1995). Excitation at

235 nm then produces five emission peaks (Figure 4.3): four corresponding to the characteristic emission signal of CBM9-bound Tb^{3+} while the fifth, with a broad maximum at 510 nm, corresponding to sensitized emission from GFP. The average characteristic distance between bound terbiums and GFP in the various fusion proteins characterized in this study was determined (by a method previously described (Kavoosi et al. 2007)) to range from $35.9 \pm 0.3 \text{ \AA}$ to $37.2 \pm 0.4 \text{ \AA}$. These values are much less than λ ($\sim 500 \text{ nm}$), indicating that within the fusion protein, a significant fraction of the donor-derived energy is transferred to fused GFP in a nonradiative manner.

4.3.2 Fusion-tag bioprocessing kinetics

Hydrolysis of the fusion protein by EKmaxTM separates the donor/acceptor pair such that $R \gg \lambda$, leading to a decrease in energy transfer and a corresponding decrease in the sensitized fluorescence emission of the acceptor. Figure 4.4 reports the time-resolved baseline-corrected decrease in GFP fluorescence emission for EKmaxTM-catalyzed hydrolysis of Tb^{3+} -CBM9-PID₄K-GFP at an initial substrate concentration of $32 \mu\text{M}$. The fluorescence emission intensity was monitored at 520 nm, where interfering signals from donor emission are minimized (241:1 maximum signal intensity ratio), as opposed to at 510 nm, where some signal overlap is observed. An exponential decay in fluorescence intensity is observed for the hydrolysis reaction, reaching a minimum intensity of ca. 10000 after 5 hours of reaction. No reduction in this minimum intensity was observed when monitoring the same reaction for up to 48 hours, indicating that the background fluorescence is not related to unreacted substrate. This fact was further verified through SDS-PAGE documentation of the reaction. Instead, the background fluorescence is likely due to stochastic interactions between the cleavage products, Tb(III)-CBM9 and GFP. The resulting $\ln(\phi)$ versus t data (Figure 4.5) are linear ($R^2 = 0.991$) at these reaction conditions for which, as will be shown later, $[S]_0 \ll K_M$. Regression of equation 4.3 to the 520 nm data therefore provides an accurate measure of k_{obs} , from which k_{cat}/K_M can be directly determined (Table 4.1). Similarly, regression of equation 4.2 to the corresponding initial reaction rate v_0 versus $[S]_0$ data (Figure 4.6) also

provides an estimate of k_{cat}/K_M , which, for all fusion proteins studied, was statistically indistinguishable from the average value regressed from the $\ln(\phi)$ versus t data.

K_M can be estimated from the x-intercept of a traditional Lineweaver-Burk plot. However, as is often noted, the error in K_M values obtained with this approach can be large since significant extrapolation to the x-intercept is generally required. This extrapolation bias can be reduced by fixing the slope of the Lineweaver-Burk plot at the previously determined value of k_{obs} and then positioning this slope to best fit the experimental data (Figure 4.7). The resulting K_M estimates (Table 4.1) are then sufficiently accurate to confirm that K_M is indeed much greater than $[S]_0$ for each data set used to obtain estimates of k_{cat}/K_M through regression of equation 4.3.

Finally, confirmation that the decrease in the LRET signal is due to EKmaxTM mediated hydrolysis was obtained by reacting a fusion protein lacking the enterokinase recognition sequence with EKmaxTM. No significant decrease in LRET signal was detected compared to the control reaction containing no enzyme (Figure 4.4).

The accuracy of the kinetic parameters determined using our LRET-based assay was verified by comparing reaction half-lives ($t_{1/2}$) measured using the LRET-based assay with those computed from an independent SDS-PAGE experiment. Figure 4.8 shows an SDS-PAGE documentation of the hydrolysis of Tb³⁺-CBM9-PID₄K-GFP at a fusion protein concentration of 32 μM . The two methods provide very similar reaction half lives (SDS-PAGE $t_{1/2} = 0.75 (\pm 0.25)$ h; LRET $t_{1/2} = 0.61 (\pm 0.02)$ h), validating the use of the LRET-based assay as an alternative and much more convenient method for monitoring fusion-tag processing, particularly in cases where the reaction kinetics are too rapid to permit visualization by SDS-PAGE.

The LRET assay confirms that changes in linker composition and length can impact CBM9 fusion-tag bioprocessing kinetics (Table 4.1). A *ca.* 2-fold difference in $t_{1/2}$ values, and thus bioprocessing times, is observed for relatively small library of Tb³⁺-CBM9-linker-ID₄K-GFP fusion proteins included in this study. The assay is simple, fast and accurate, providing k_{cat}/K_M values that typically contain standard errors of less than 3%. As a result, both substantial and more subtle differences in bioprocessing kinetics

can be measured. For example, a 9.2% increase in k_{cat}/K_M was observed when the length of the poly-glycine linker was increased from 3 to 15 amino acids. The 384-well microtitre plate format of the assay should therefore allow for a number of useful high-throughput studies, including selection of an optimal linker chemistry as well as the preferred processing enzyme.

4.4 Tables

Table 4.1 Michaelis Menten kinetic constants and reaction half-lives determined by the LRET-based assay

EKmaxTM recognition site and linker sequence	k_{cat}/K_M ($M^{-1}s^{-1} \times 10^{-4}$)	$t_{1/2}$ (hr)	K_M (μM)
PIDDDDK	3.43 ± 0.12	0.61 ± 0.02	4301 ± 712
PTPTPIDDDDK	6.62 ± 0.10	0.31 ± 0.01	1439 ± 1203
PTPTPTPIDDDDK	6.54 ± 0.12	0.32 ± 0.01	1328 ± 850
PTPTPTPTPTPTPIDDDDK	5.22 ± 0.13	0.40 ± 0.01	626 ± 299
GGGIDDDDK	3.82 ± 0.11	0.55 ± 0.02	2119 ± 1071
GGGGGGGGGGGGGGGGGGIDDDDK	4.17 ± 0.12	0.50 ± 0.01	541 ± 653
SSSNNNNNNNNNNLAIIDDDDK	4.07 ± 0.11	0.51 ± 0.01	595 ± 334

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence intervals)

4.5 Figures

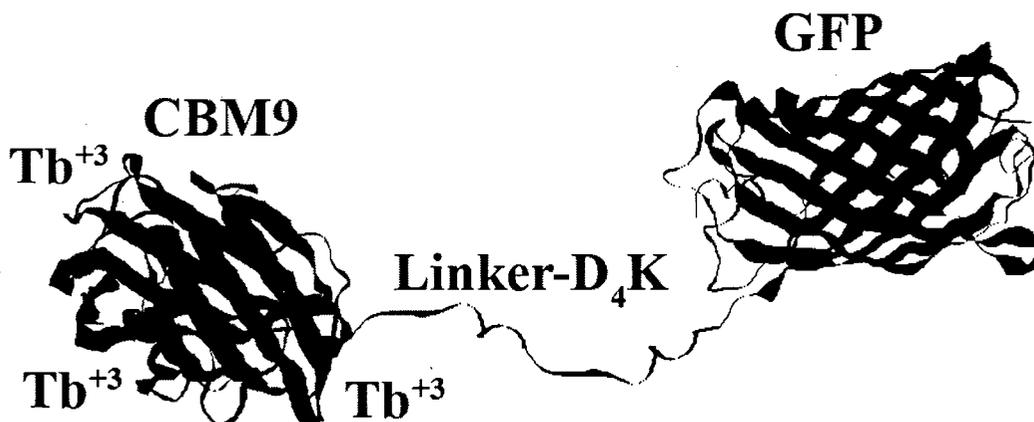


Figure 4.1 Schematic representation of the terbium bound fusion protein used for the LRET-based peptidase assay (not to scale). Excitation of bound Tb(III) at 235 nm results in the transfer of energy at 490 nm to GFP, leading to the emission of a signal at 510 nm. D₄K indicates the recognition site for the serine peptidase, enterokinase.

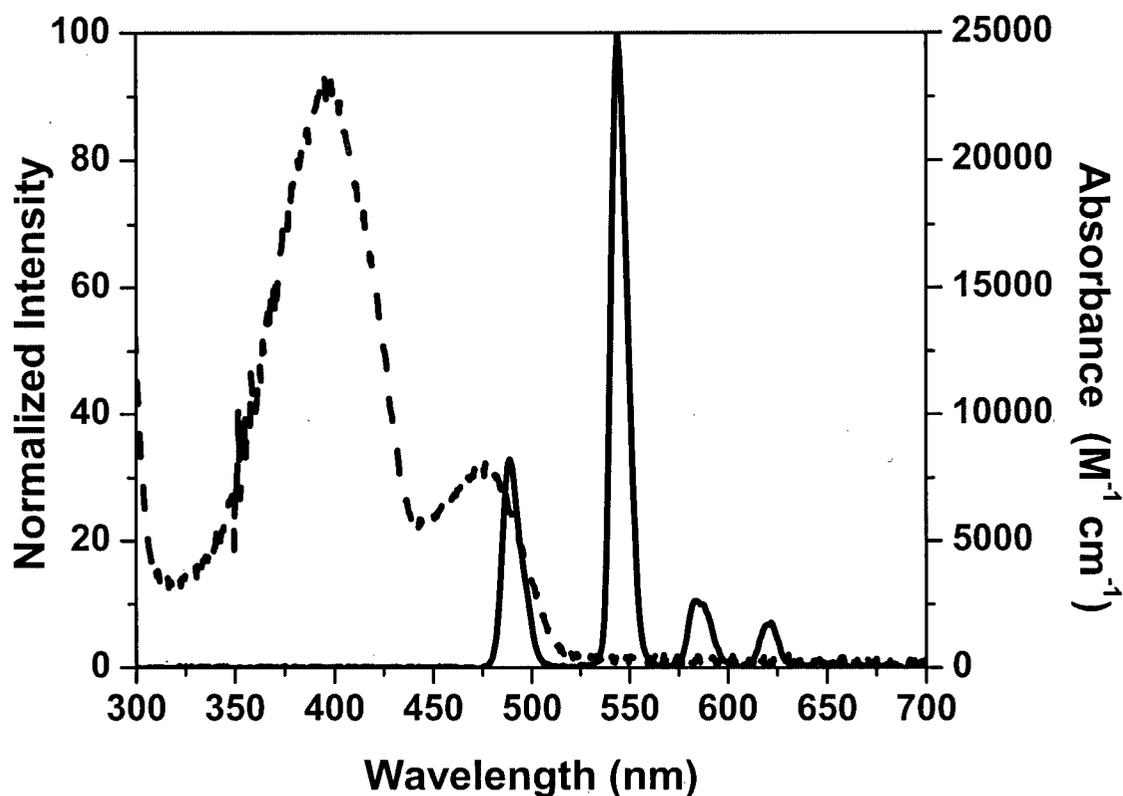


Figure 4.2 Individual spectrum of donor and acceptor. Solid line is the emission spectrum of CBM9-bound terbium (donor) obtained after excitation at 235 nm and a post-excitation delay of 200 μ s. Dashed line is the absorbance spectrum of GFP (acceptor). The spectral overlap (480 nm to 505 nm) between CBM9-bound terbium and GFP allows for energy transfer to take place.

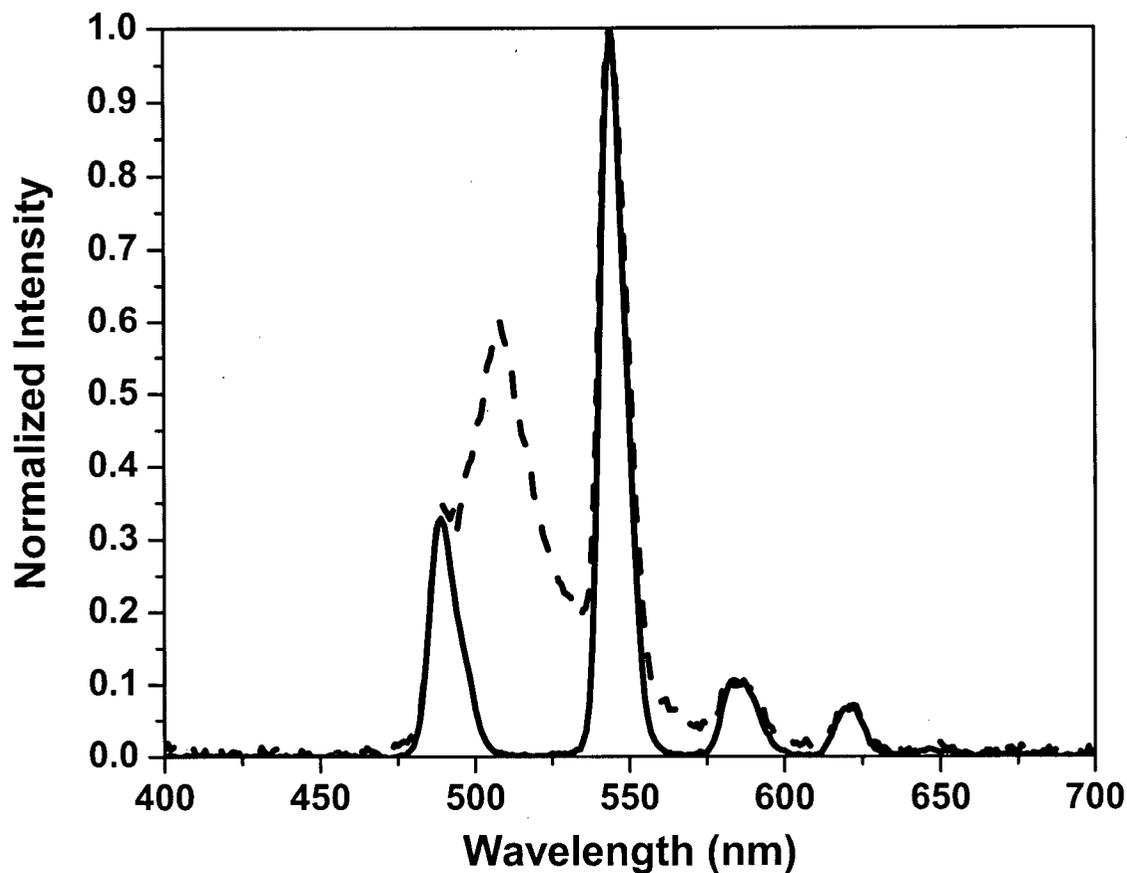


Figure 4.3 Emission spectra of CBM9-bound terbium (donor domain only) (solid line) and Tb³⁺-CBM9-(PT)₂PID4K-GFP fusion protein (dashed line). Samples were excited at 235 nm followed by a 200 μ s delay prior to measurement of emission signal at each wavelength. Scans were performed at a rate of 1 nm/s.

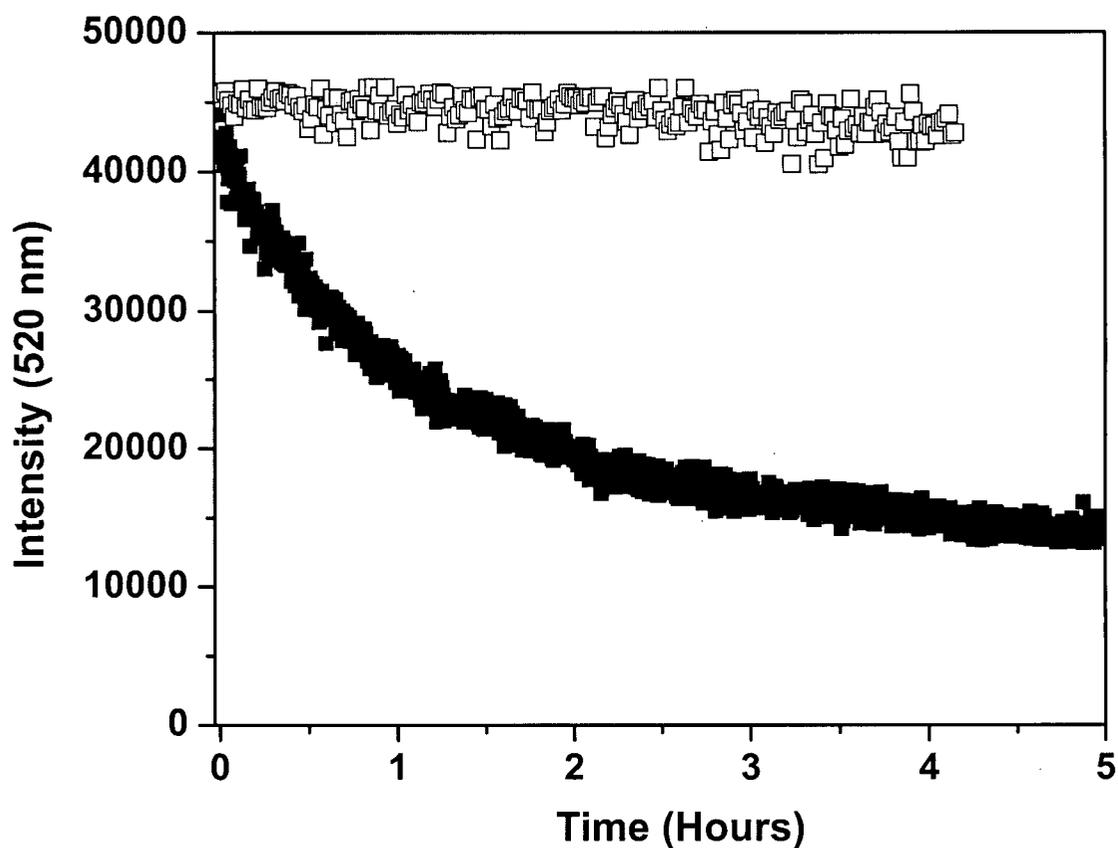


Figure 4.4 Real-time LRET-based signal from EKmaxTM mediated hydrolysis reaction. 32 μM of Tb^{3+} -CBM9-PID₄K-GFP fusion protein was incubated at 30°C with 9.3 nM EKmaxTM (solid squares). The control reaction has buffer in place of enzyme (open squares). Sample was excited at 235 nm and emission was measured at 520 nm following a 200 μs delay. Each measurement is an average of 10 excitations.

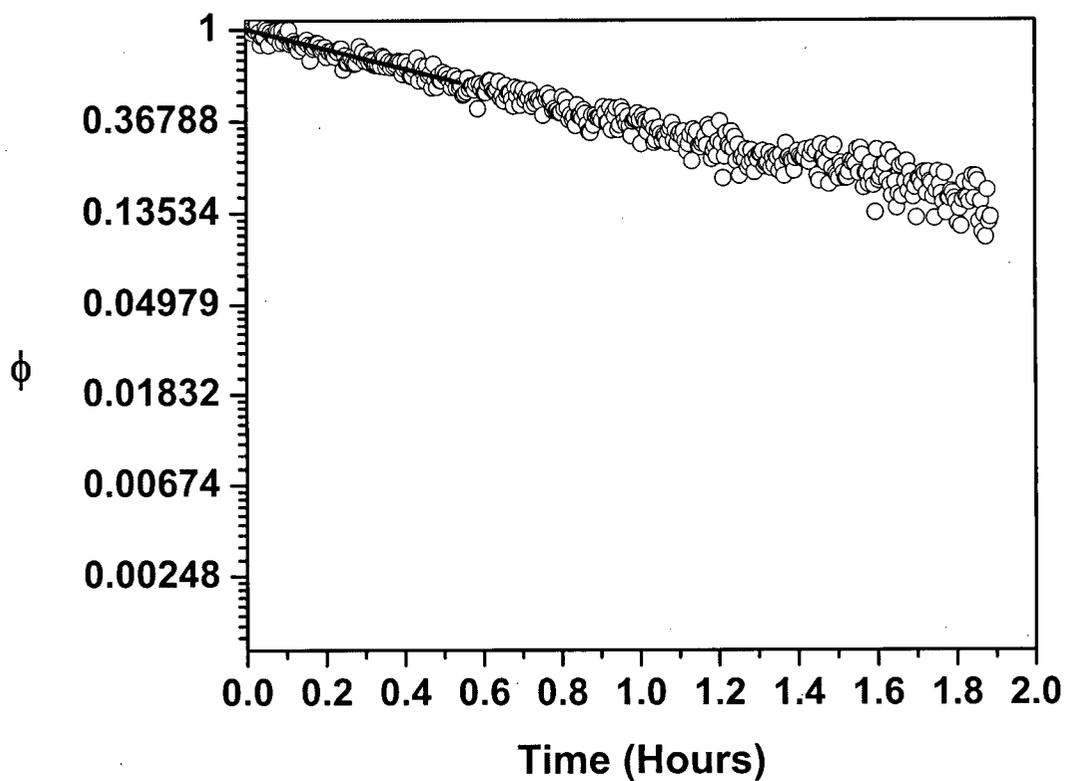


Figure 4.5 Baseline corrected logarithmic decay in $\ln(\phi)$ due to EKmaxTM mediated hydrolysis. Reaction conditions are the same as in Figure 4.4. The solid line represents the linear least-squares fit to the initial data.

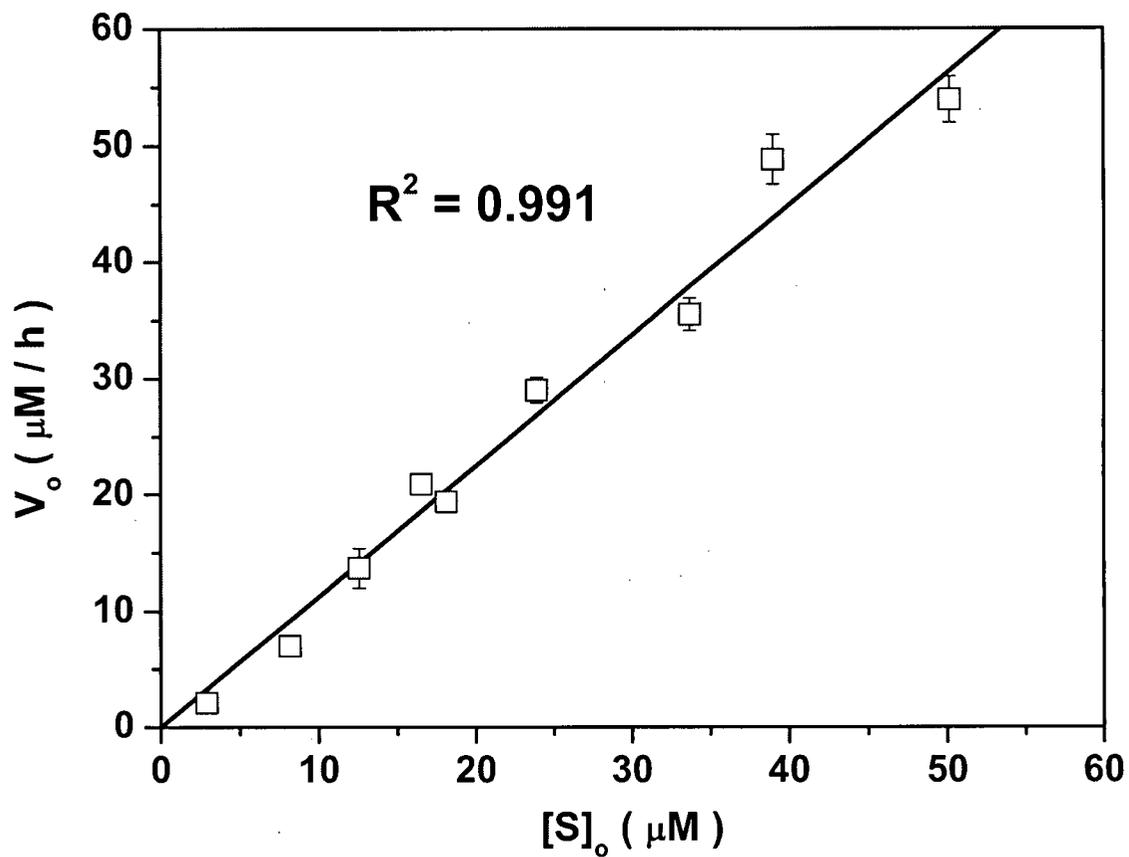


Figure 4.6 Michaelis-Menten analysis. Initial rates of EKmaxTM-catalyzed hydrolysis measured as a function of initial substrate concentration. Tb³⁺-CBM9-PID₄K-GFP was incubated with EKmaxTM (9.3 nM) at 30°C.

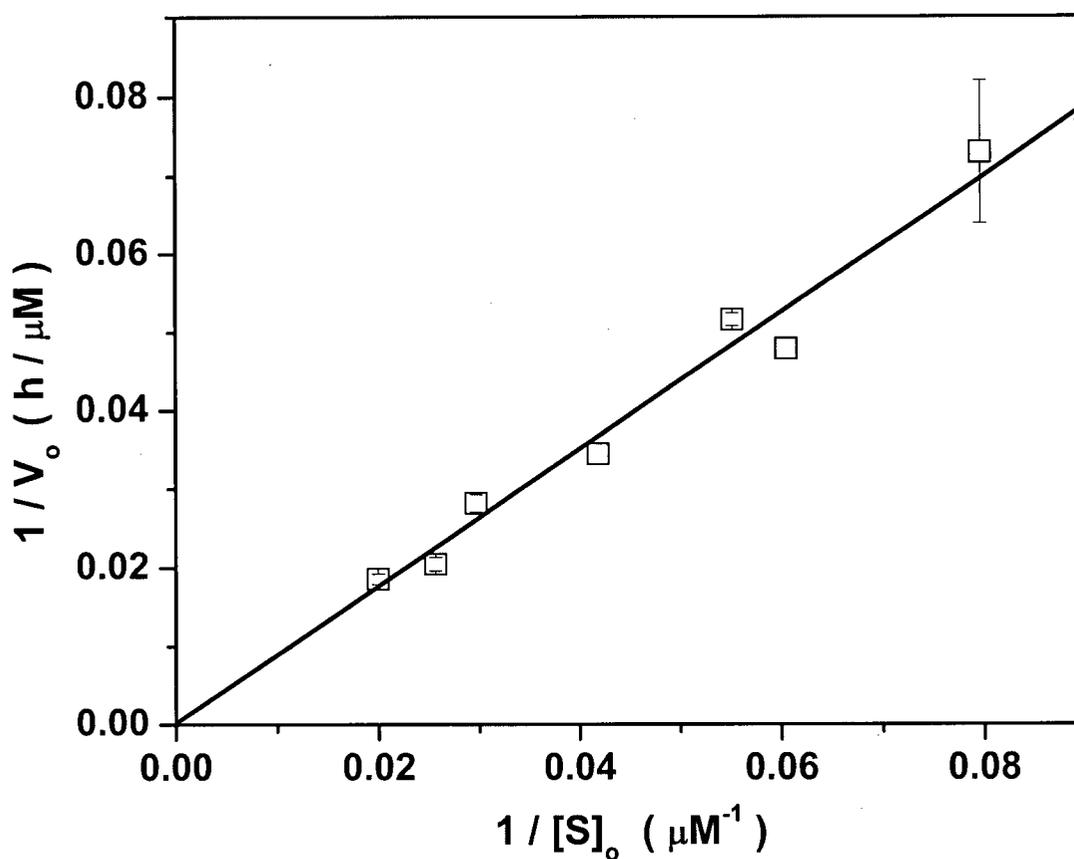


Figure 4.7 Lineweaver-Burk representation of initial rates of EKmaxTM-catalyzed hydrolysis of Tb³⁺-CBM9-PID₄K-GFP measured as a function of initial substrate concentration. Reaction conditions are the same as reported in Figure 4.4.

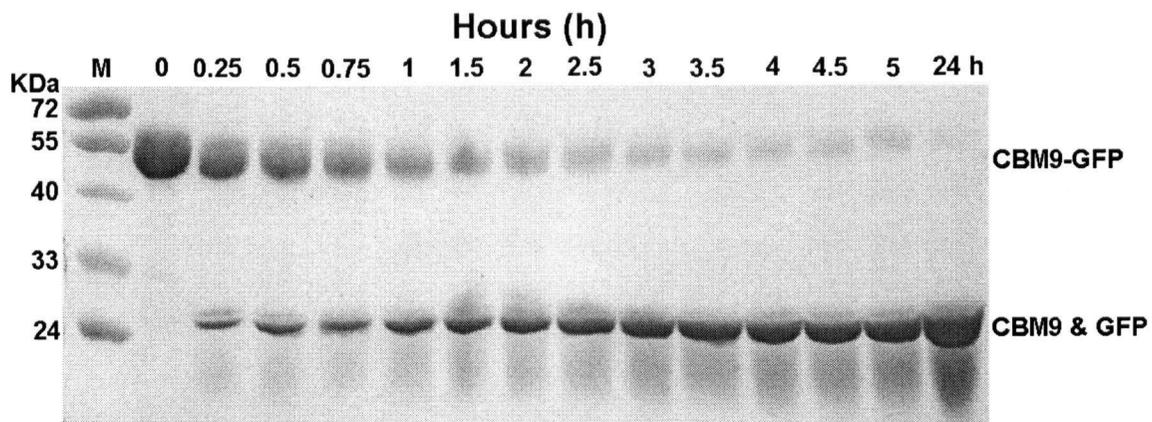


Figure 4.8 SDS-PAGE documentation of EKmaxTM-catalyzed hydrolysis of Tb³⁺-CBM9-PID₄K-GFP. Reaction conditions are the same as reported in Figure 4.4.

4.6 References

- Becker, C. F. W.; Seidel, R.; Jahnz, M.; Bacia, K.; Niederhausen, T.; Alexandrov, K.; Schwille, P.; Goody, R. S.; Engelhard, M. (2006). "C-terminal fluorescence labeling of proteins for interaction studies on the single-molecule level". *ChemBioChem* 7(6):891-895.
- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Boskovic, D. S.; Troxler, T.; Krishnaswamy, S. (2004). "Active site-independent recognition of substrates and product by bovine prothrombinase. A fluorescence resonance energy transfer study". *Journal of Biological Chemistry* 279(20):20786-20793.
- Boulware, K. T.; Daugherty, P. S. (2006). "Protease specificity determination by using cellular libraries of peptide substrates (CLiPS)". *Proceedings of the National Academy of Sciences of the United States of America* 103(20):7583-7588.
- Burmeister-Getz, E.; Cooke, R.; Selvin, P. R. (1998). "Luminescence resonance energy transfer measurements in myosin". *Biophysical Journal* 74(5):2451-2458.
- Callaci, S.; Heyduk, E.; Heyduk, T. (1999). "Core RNA polymerase from E-coli induces a major change in the domain arrangement of the sigma(70) subunit". *Molecular Cell* 3(2):229-238.
- Chao, H. M.; Bautista, D. L.; Litowski, J.; Irvin, R. T.; Hodges, R. S. (1998). "Use of a heterodimeric coiled-coil system for biosensor application and affinity purification". *Journal of Chromatography B* 715(1):307-329.
- Chaudhuri, D.; Horrocks, W. D., Jr.; Amburgey, J. C.; Weber, D. J. (1997). "Characterization of Lanthanide Ion Binding to the EF-Hand Protein S100b by Luminescence Spectroscopy". *Biochemistry* 36(32):9674-9680.
- Chen, S.; Chen, L.-l.; Luo, H.-b.; Sun, T.; Chen, J.; Ye, F.; Cai, J.-h.; Shen, J.-k.; Shen, X.; Jiang, H.-l. (2005). "Enzymatic activity characterization of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer technique". *Acta Pharmacologica Sinica* 26(1):99-106.
- Churchich, J. E. (1997). "Conformational changes at the nucleotide binding of GroEL induced by binding of protein substrates - Luminescence studies". *Journal of Biological Chemistry* 272(32):19645-19648.

- Cotrin, S. S.; Puzer, L.; De Souza Judice, W. A.; Juliano, L.; Carmona, A. K.; Juliano, M. A. (2004). "Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: assays with human cathepsin B". *Analytical Biochemistry* 335(2):244-252.
- Dursina, B. E.; Reents, R.; Niculae, A.; Veligodsky, A.; Breitling, R.; Pyatkov, K.; Waldmann, H.; Goody, R. S.; Alexandrov, K. (2005). "A genetically encodable microtag for chemo-enzymatic derivatization and purification of recombinant proteins". *Protein Expression and Purification* 39(1):71-81.
- Gasparian, M. E.; Ostapchenko, V. G.; Schulga, A. A.; Dolgikh, D. A.; Kirpichnikov, M. P. (2003). "Expression, purification, and characterization of human enteropeptidase catalytic subunit in *Escherichia coli*". *Protein Expression and Purification* 31(1):133-139.
- Girard, J.-M.; Le, K. H. D.; Lederer, F. (2006). "Molecular characterization of laforin, a dual-specificity protein phosphatase implicated in Lafora disease". *Biochimie* 88(12):1961-1971.
- Guan, K. L.; Dixon, J. E. (1991). "Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Hearn, M. T.; Acosta, D. (2001). "Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins". *Journal of Molecular Recognition* 14(6):323-369.
- Heyduk, E.; Heyduk, T.; Claus, P.; Wisniewski, J. R. (1997). "Conformational changes of DNA induced by binding of Chironomus high mobility group protein 1a (cHMG1a) - Regions flanking an HMG1 box domain do not influence the bend angle of the DNA". *Journal of Biological Chemistry* 272(32):19763-19770.
- Heyduk, T. (2001). "Luminescence resonance energy transfer analysis of RNA polymerase complexes". *Methods* 25(1):44-53.
- Jenny, R. J.; Mann, K. G.; Lundblad, R. L. (2003). "A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa". *Protein Expression and Purification* 31(1):1-11.
- Kapanidis, A. N.; Ebright, Y. W.; Ludescher, R. D.; Chan, S.; Ebright, R. H. (2001). "Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution". *Journal of Molecular Biology* 312(3):453-468.
- Kavoosi, M.; Creagh, A. L.; Kilburn, D. G.; Haynes, C. A. (2007). "Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *E. coli*". *Biotechnology and Bioengineering* In Press.

- Kavoosi, M.; Meijer, J.; Kwan, E.; Creagh, A. L.; Kilburn, D. G.; Haynes, C. A. (2004). "Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T-maritima*". *Journal of Chromatography B* 807(1):87-94.
- Kolb, A. J.; Burke, J. W.; Mathis, G. 1997. Homogeneous, time-resolved fluorescence method for drug discovery. In: J. P. Devlin, editor. *High Throughput Screening: The Discovery of Bioactive Substances*. New York: Marcel Dekker. p 345-360.
- Lien, S.; Milner, S. J.; Graham, L. D.; Wallace, J. C.; Francis, G. L. (2001). "Linkers for improved cleavage of fusion proteins with an engineered alpha-lytic protease". *Biotechnology and Bioengineering* 74(4):335-343.
- Lowe, C. R.; Lowe, A. R.; Gupta, G. (2001). "New developments in affinity chromatography with potential application in the production of biopharmaceuticals". *Journal of Biochemical and Biophysical Methods* 49(1-3):561-574.
- Mathis, G. (1995). "Probing Molecular-Interactions with Homogeneous Techniques Based on Rare-Earth Cryptates and Fluorescence Energy-Transfer". *Clinical Chemistry* 41(9):1391-1397.
- Medintz, I. L.; Clapp, A. R.; Brunel, F. M.; Tiefenbrunn, T.; Uyeda, H. T.; Chang, E. L.; Deschamps, J. R.; Dawson, P. E.; Mattoussi, H. (2006). "Proteolytic activity monitored by fluorescence resonance energy transfer through quantum-dot-peptide conjugates". *Nature Materials* 5(7):581-589.
- Nishikata, M.; Yoshimura, Y.; Deyama, Y.; Suzuki, K. (2006). "Continuous assay of protein tyrosine phosphatases based on fluorescence resonance energy transfer". *Biochimie* 88(7):879-886.
- Notenboom, V.; Boraston, A. B.; Kilburn, D. G.; Rose, D. R. (2001). "Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms". *Biochemistry* 40(21):6248-6256.
- Parker, D.; Williams, J. A. G. (2003). Responsive luminescent lanthanide complexes. In: A. Sigel, H. Sigel, editors. *Metal Ions in Biological Systems*. New York: Marcel Dekker p233-280.
- Root, D. D. (1997). "In situ molecular association of dystrophin with actin revealed by sensitized emission immuno-resonance energy transfer". *Proceedings of the National Academy of Sciences of the United States of America* 94(11):5685-5690.
- Sambrook, J.; Russell, D. W. (2001). *Molecular cloning : a laboratory manual*. 3rd ed. New York: Cold Spring Harbor Laboratory Press.

- Sammes, P. G.;Yahioğlu, G. (1996). "Modern bioassays using metal chelates as luminescent probes". *Natural Product Reports* 13(1):1-28.
- Selvin, P. R. (1995). "Fluorescence resonance energy transfer". *Methods in Enzymology* 246:300-334. .
- Selvin, P. R. (1996). "Lanthanide-based resonance energy transfer". *IEEE Journal of Selected Topics in Quantum Electronics* 2(4):1077-1087.
- Selvin, P. R. (2002). "Principles and biophysical applications of lanthanide-based probes". *Annual Review of Biophysics and Biomolecular Structure* 31:275-302.
- Selvin, P. R.;Hearst, J. E. (1994). "Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer". *Proceedings of the National Academy of Sciences of the United States of America* 91(21):10024-10028.
- Shimomura, O.; Johnson, F. H.;Saiga, Y. (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*". *Journal of Cellular and Comparative Physiology* 59:223-239.
- Thomas, D. A.; Francis, P.; Smith, C.; Ratcliffe, S.; Ede, N. J.; Kay, C.; Wayne, G.; Martin, S. L.; Moore, K.; Amour, A.;Hooper, N. M. (2006). "A broad-spectrum fluorescence-based peptide library for the rapid identification of protease substrates". *Proteomics* 6(7):2112-2120.
- Tripet, B.; Yu, L.; Bautista, D. L.; Wong, W. Y.; Irvin, R., T.;Hodges, R. S. (1996). "Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins". *Protein Engineering* 9(11):1029-1042.
- Veenstra, T. D.; Gross, M. D.; Hunziker, W.;Kumar, R. (1995). "Identification of metal-binding sites in rat brain calcium-binding protein". *Journal of Biological Chemistry* 270(51):30353-30358.
- Winterhalter, C.; Heinrich, P.; Candussio, A.; Wich, G.;Liebl, W. (1995). "Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*". *Molecular Microbiology* 15(3):431-444.
- Xiao, M.; Li, H.; Snyder, G. E.; Cooke, R.; Yount, R. G.;Selvin, P. R. (1998). "Conformational changes between the active-site and regulatory light chain of myosin as determined by luminescence resonance energy transfer: the effect of nucleotides and actin". *Proceedings of the National Academy of Sciences of the United States of America* 95(26):15309-15314.

5 A Novel Two-Zone Protein Uptake Model for Affinity Chromatography and Its Application to the Description of Elution Band Profiles of Proteins Fused to a Family 9 Cellulose Binding Module Affinity Tag

5.1 Introduction

Due to its exquisite binding selectivity, affinity chromatography is finding increasingly widespread use in the purification of natural and recombinant protein products at the manufacturing scale (Lowe et al. 2001). The large-scale capture and affinity purification of monoclonal antibodies on immobilized protein A columns is the most widely used and thoroughly studied application of industrial affinity chromatography (Jungbauer and Hahn 2004), but many other important applications exist, including the purification of human tissue plasminogen from blood plasma using immobilized lysine (Deutsch and Mertz 1970) and the purification of ATP-dependent kinases and NAD⁺-dependent dehydrogenases using immobilized 5'-AMP (Mulcahy et al. 2002).

The power of affinity separations can be extended to proteins with no known binding partner through recombinant DNA technology, which enables production of a target protein as a recombinant fusion to an N- or C-terminal affinity tag possessing a highly specific binding partner that can be immobilized to form a stable affinity

* A version of this chapter is published in the *Journal of Chromatography A*. [Reference: Mojgan Kavooosi, Nooshafarin Sanaie, Florian Dismer, Jürgen Hubbuch, Douglas G. Kilburn, Charles A. Haynes, A novel two-zone protein uptake model for affinity chromatography and its application to the description of elution band profiles of proteins fused to a family 9 cellulose binding module affinity tag. *J. Chromatography A*. 1160(1-2), 137-149 (2007)]

chromatography media. A number of affinity tag technologies are commercially available, including the glutathione S-transferase (GST) tag (Guan and Dixon 1991; Smith and Johnson 1988), the calmodulin binding peptide tag (Stofko-Hahn et al. 1992; Vaillancourt et al. 1997; Zheng et al. 1997), the streptavidin tag (Keefe et al. 2001; Wilson et al. 2001), the FLAG peptide tag (Einhauer and Jungbauer 2001; Hopp et al. 1988) and the polyhistidine tag, which permits selective capture and purification of the fusion protein on an immobilized metal affinity chromatography column (Crowe et al. 1994; Hopp et al. 1988; Porath et al. 1975).

Intraparticle mass transport, most notably the rate of diffusion within the pore liquid of the stationary phase, typically limits protein uptake and controls band broadening in adsorptive modes of protein chromatography, including affinity chromatography, where the binding interaction with the stationary phase is typically strong (Arnold et al. 1985a; Farnan et al. 1997). A number of models have therefore been developed to describe intraparticle mass transport inside sorbent particles, with the pore diffusion model (PDM) finding the most widespread use [*e.g.*, (Farnan et al. 2002)]. The PDM assumes that the overall rate of protein uptake is proportional to the concentration gradient in the pore liquid, permitting simplification of the general rate model of chromatography by eliminating concentration gradients in the hydrodynamic film surrounding each sorbent particle and by establishing local equilibrium of the sorbate at each radial position within the sorbent particle. Assuming for the moment that surface diffusion effects are negligible, the rate of protein uptake within the porous sorbent particle is given by the PDM through the relation

$$\varepsilon_p \frac{dc_i}{dt} = \frac{1}{r^2} \frac{d}{dr} \left(\varepsilon_p D_p r^2 \frac{dc_i}{dr} \right) - \frac{dq_i}{dt} \quad (5.1)$$

where, sorbate equilibrium is often described by the Langmuir isotherm,

$$\frac{dq_i}{dt} = \frac{dq_i}{dc_i} \frac{dc_i}{dt} = \frac{q_i^{sat} K_{a_i}}{(1 + K_{a_i} c_i)^2} \frac{dc_i}{dt} \quad (5.2)$$

Thus, knowledge of the stationary phase porosity ε_p , the saturation capacity of the sorbent q_i^{sat} (kg m^{-3}), and the Langmuir equilibrium binding constant K_a (M^{-1}), permits estimation of solute c_i and sorbate q_i concentration profiles within the stationary phase as a function of time. Mass transfer within an interstitial volume element of the column is given by the column continuity equation

$$\frac{dC_i}{dt} = D_L \frac{d^2 C_i}{dz^2} - u_o \frac{dC_i}{dz} - \frac{(1-\varepsilon)}{\varepsilon} \frac{d\bar{s}_i}{dt} \quad (5.3)$$

where C_i is the concentration of protein i in the interstitial mobile-phase liquid, ε is the interstitial void fraction of the column, D_L is the axial dispersion coefficient ($\text{m}^2 \text{s}^{-1}$), u_o is the interstitial velocity of the mobile phase (m s^{-1}), z is the axial positional vector, and \bar{s}_i is the average protein concentration within the stationary phase particles of uniform radius r_p , given by

$$\bar{s}_i = \frac{3}{r_p^3} \int_0^{r_p} (\varepsilon_p c_i + q_i) r^2 dr = \frac{3}{r_p^3} \int_0^{r_p} s_i(r,t) r^2 dr \quad (5.4)$$

The boundary conditions for solving equation 5.1 of the PDM are given by

$$c_i(r=r_p, t) = C_i \quad (5.5)$$

$$\frac{dc_i(r=0, t)}{dr} = 0 \quad (5.6)$$

where C_i is given by solution of equation 5.3 at time t . The boundary condition at $r = 0$ given by equation 5.6 is generally applied in all continuous models of chromatography. However, the full implications of its use are not always appreciated. In particular, since c_i and q_i are both specified by equation 5.1 to be continuous functions of r and t , the application of equation 5.6 necessarily leads to a physically improbable model prediction that both $c_i(r=0, t)$ and $q_i(r=0, t)$ become nonzero immediately upon contact of the stationary phase particle with the mobile phase liquid, where $C_i(r_p, t=0) = C_i^0$. Equation 5.6 has nevertheless been extensively applied to the modeling of many different forms of adsorptive chromatography (Guiochon 2002; Johnston and Hearn 1991; Skidmore et al.

1990), including various forms of affinity chromatography (Arnold et al. 1985a; Arnold et al. 1985b; Arve and Liapis 1987; Katoh et al. 1978; Patwardhan et al. 1995; Sirotti and Emery 1983).

A more general and physically realistic model for protein uptake within a porous stationary phase would predict for sufficiently short contact times a region within the interior of the sorbent particle that contains no protein, while both c_i and q_i would be nonzero and increase with r in the outer shell of the particle. The protein-containing zone would then be predicted to increase with time at the expense of the protein-free zone. This two-zone behavior has been observed in confocal laser scanning microscopy (CLSM) studies of protein uptake in porous chromatography particles, particularly when there is strong interaction between the sorbate and the sorbent, as is typically observed in affinity chromatography systems (Hubbuch et al. 2003; Li et al. 2003; Linden et al. 2002; Ljunglöf and Hjorth 1996). It is reminiscent of the classic shrinking-core model of diffusion-controlled chemical reaction engineering first proposed by Weisz and Goodwin (Weisz and Goodwin 1963). However, that model assumes that local sorbate equilibrium is defined by the rectangular isotherm and thereby predicts an infinitely steep concentration gradient at the core radius r_c separating the protein-free inner core from the sorbent-saturated outer shell of the porous particle (Pinto and Graham 1987).

Here we describe a generalized two-zone model for protein uptake in a porous sorbent particle that relaxes the rectangular-isotherm approximation of the traditional shrinking-core model to allow for simultaneous intraparticle mass transport and sorbent loading within the outer zone of the porous particle and thus, the presence of concentration gradients within the shell region. The model is applied to the description of elution band profiles for fusion proteins tagged at their N-terminus with TmXyn10ACBM9-2 (henceforth referred to as CBM9), the C-terminal family 9 carbohydrate-binding module of xylanase 10A from *Thermotoga maritima* (Winterhalter et al. 1995). In a previous paper (Kavoosi et al. 2004), we introduced the CBM9 affinity tag and demonstrated its application in the affinity purification of recombinant proteins from *E. coli* using an inexpensive, commercially available cellulosic resin, Perloza™ MT100. CBM9 binds specifically and tightly to the reducing ends of both insoluble

cellulose and simple soluble sugars, including glucose (Boraston et al. 2001). These unique binding properties allow for selective binding of CBM9-tagged fusion proteins to a porous cellulose sorbent particle and quantitative elution using 1 M glucose. Perloza™ MT100, a highly porous, beaded cellulosic resin sells for *ca.* \$35 per liter of bulk resin. The extraordinary low cost of this matrix, combined with its high static binding capacity for CBM9-tagged fusion proteins (10 $\mu\text{mol g}^{-1}$ dry resin), offer considerable economic advantages over other commercially available affinity tag technologies. In this work, we fuse CBM9 to the N-terminus of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* (Cramer et al. 1996; Shimomura et al. 1962), and use the natural fluorescence of GFP as a direct and convenient means to track our fusion protein and validate our model. CLSM is used to measure temporally and radially resolved CBM9-GFP fluorescence profiles inside the Perloza™ MT100 sorbent particle, permitting tracking of r_c and intraparticle mass transport of protein in the outer zone of the particle. Because GFP of CBM9-GFP fluoresces naturally, uptake artifacts associated with competition between unlabelled and chemically labelled protein are eliminated, greatly simplifying data analysis (Carta et al. 2005; Martin et al. 2005).

5.2 A Proposed Generalized Two-Zone Model of Affinity Chromatography

As with the PDM, the derivation of our generalized two-zone model (TZM) of adsorptive chromatography is based on the condition that both protein adsorption kinetics and protein transport through the hydrodynamic fluid film surrounding the porous particle are rapid compared to solute diffusion processes within the spherical particles of uniform radius r_p and porosity ε_p . We may therefore apply the well-known parallel pore diffusion model for spherical sorbent particles (Ma et al. 1996), given by

$$\frac{ds_i}{dt} = \frac{1}{r^2} \frac{d}{dr} \left[r^2 \left(\varepsilon_p D_p \frac{dc_i}{dr} + D_s \frac{dq_i}{dr} \right) \right] \quad (5.7)$$

where D_s is the surface diffusivity of the sorbate and the driving force for diffusion in the adsorbed phase is assumed to be given by the sorbate concentration gradient. Surface

diffusion is ignored in most chromatography model developments as D_s is generally thought to be at least two orders of magnitude smaller than D_p . It will not be explicitly accounted for in this study either as independent measurement of D_s was not possible. However, we note that in high capacity chromatography media loaded in the nonlinear region, the surface concentration gradient may be higher than the solute concentration gradient within the pore liquid, thereby resulting in a sorbate flux contribution to intraparticle protein transport despite the significantly lower value of D_s . Extension of the model described here to that situation is straightforward, provided the value of D_s is known.

The parallel flux term in equation 5.7 has therefore been simplified to

$$\frac{ds_i}{dt} = \frac{1}{r^2} \frac{d}{dr} \left[r^2 \varepsilon_p D_p \frac{dc_i}{dr} \right] \quad (5.8)$$

The TZM divides the porous sorbent particle into two zones that meet at r_c , the core radius, which decreases as a function of time due to intraparticle mass transport of the protein. Protein uptake in the outer shell extending from r_c to r_p is defined by equation 5.8 and the adsorption isotherm model selected. The inner core from $r = 0$ to r_c contains no protein. Thus, an inner boundary condition of $c_i = 0$ at $r = r_c$ may be used to solve equation 5.8 to determine protein $c_i(r,t)$ and sorbate $q_i(r,t)$ concentration profiles in the outer shell provided the value of $r_c(t)$ is known. An estimate of $r_c(t)$ can be obtained by numerical iteration. To illustrate the strategy used, which is based on a modification of an iterative solution scheme developed by Pritzker (Pritzker 2003), we consider the simple case where C_i remains equal to C_i^0 at all times.

The rate of sorbate uptake into the porous particle equals the flux of sorbate across the external surface of the particle

$$\frac{4}{3} \pi r_p^3 \frac{d\bar{s}_i}{dt} = 4 \pi r_p^2 \left(\varepsilon_p D_p \frac{dc_i}{dr} \Big|_{r=r_p} \right) \quad (5.9)$$

The right-hand side of this equation may be evaluated by applying the steady-state approximation $\left(i.e., \frac{d\bar{s}_i}{dt} = 0 \right)$ to equation 5.8 to permit its analytical integration within the outer zone. Since $r = r_c(t)$ marks the position of the advancing front of the adsorbate, it requires that $c_i(r_c, t) = 0$, which automatically sets $q_i(r, t) = 0$. The boundary conditions at $r = r_p$ remains the same as in the homogenous model, giving

$$c_i(r, t) = C_i^o \left(\frac{\left(\frac{1}{r_c} - \frac{1}{r} \right)}{\left(\frac{1}{r_c} - \frac{1}{r_p} \right)} \right) \quad (5.10)$$

Differentiating with respect to r and evaluating the result at $r = r_p$, then gives

$$\frac{d\bar{s}_i}{dt} = \frac{3}{r_p^3} C_i^o \frac{1}{\left(\frac{1}{r_c(t)} - \frac{1}{r_p} \right)} \quad (5.11)$$

For a given time t , an assumed value of r_c therefore allows estimation of $\bar{s}_i(t)$ based on the initial condition $\bar{s}_i(t=0) = 0$. The value of $\bar{s}_i(t)$ for an assumed value of $r_c(t)$ may also be determined from equation 5.4, which upon insertion of equation 5.10 may be written as

$$\bar{s}_i = \frac{3}{r_p^3} \int_0^{r_p} \left[\varepsilon_p C_i^o \frac{\left(\frac{1}{r_c} - \frac{1}{r} \right)}{\left(\frac{1}{r_c} - \frac{1}{r_p} \right)} + q_i(r, t) \right] r^2 dr \quad (5.12)$$

Solution of equation 5.12 requires knowledge of $q_i(r, t)$, which may be determined for the assumed value of $r_c(t)$ through insertion of equation 5.10 into the chosen adsorption isotherm relation. If the adsorption process follows the one-component Langmuir adsorption isotherm, we obtain

$$q_i(r,t) = \frac{q_i^{sat} K_{a_i} C_i^o \left(\frac{1}{r_c} - \frac{1}{r} \right) / \left(\frac{1}{r_c} - \frac{1}{r_p} \right)}{\left[1 + K_{a_i} C_i^o \left(\frac{1}{r_c} - \frac{1}{r} \right) / \left(\frac{1}{r_c} - \frac{1}{r_p} \right) \right]} \quad (5.13)$$

A self-consistent estimate of $r_c(t)$ can therefore be obtained by using a Newton-Raphson algorithm to minimize the difference in the value of $\bar{y}_i(t)$ calculated from equations 5.11 and 5.12. It is important to note that this estimate of $r_c(t)$ is not exact since we have invoked the steady-state approximation to derive equation 5.9. However, as we will show, the estimated values of $r_c(t)$ are quantitatively consistent with CLSM data for CBM9-GFP uptake in the stationary-phase media, indicating that the steady-state approximation, though clearly inexact, is sufficiently reliable to permit accurate model predictions.

5.3 Two-Zone Model Solution Algorithm

The set of coupled transport equations (5.3, 5.4 and 5.8) were solved numerically by a finite-difference iteration scheme written in FORTRAN 90. Initial and boundary conditions for the column continuity equation are:

$$\begin{aligned} C_i &= 0 & t = 0, \quad 0 \leq z \leq L \\ C_i &= C_i^{feed} + \frac{D_L}{u_o} \frac{dC_i}{dz} & z = 0, \quad \text{all } t > 0 \\ \frac{\partial C_i}{\partial z} &= 0 & z = L, \quad \text{all } t > 0 \end{aligned} \quad (5.14)$$

Time and space domains were discretized by a Crank-Nicolson scheme (Crank and Nicolson 1996) to approximate differentials by a central difference in time and an average central difference in space. The column was meshed in the z dimension into N (at least 400) volume elements to match (or slightly exceed) the number of theoretical units (NTUs) within the column, and the number of radial volume elements within the stationary phase particle was set equal to 30. Finer meshing within the stationary phase increased computational time without a noticeable improvement in model accuracy. This

discretization of the model equations yields a set of tridiagonal linear algebraic equations that were solved by the Thomas algorithm (Holland and Liapis 1983) and the application of a first-order upwind-corrected power-law scheme (Patankar 1980) to ensure diagonal dominant matrices. Time increments for solution of equations 5.3 and 5.8 were set at $0.01L/Nu$ and $0.002L/Nu$, respectively, where L is the column length and u is the superficial velocity.

5.4 Materials and Methods

5.4.1 Chromatographic Media and Reagents

Perloza™ MT100 was purchased from Iontosorb Inc. (Czech Republic). Perloza™ MT100 is a porous, spherical media derived from regenerated cellulose. The particle diameter (d_p) distribution of Perloza™ MT100 as well as the average d_p was determined by light scattering using a Malvern Mastersizer 2000 (Malvern Instruments, UK). The average particle volume was calculated assuming a spherical geometry from which the mean particle diameter of the sphere was obtained using MIE theory (Barber and Hill 1990). Perloza™ MT100/G15 has a particle diameter (d_p) distribution of 56 μm to 159 μm with a mean value of $84 \pm 0.6 \mu\text{m}$.

Sephadex G15 and Ni^{+2} -Sepharose IMAC media were obtained from Sigma-Aldrich (Mississauga, ON, Canada) and Novagen (Milwaukee, MI), respectively.

5.4.2 Scanning Electron Micrographs

Scanning electron micrographs of Perloza™ MT100 media and its associated pore structure were obtained using a Hitachi S-4700 Field Emission scanning electron microscope operating at an accelerating voltage of 5 kV with a working distance of 5 to 15 mm. Samples were prepared by loading a high-pressure freezing hat with approximately 3 μL of concentrated Perloza™ MT100 in nano-pure water. The hat was immersed for 5-7 seconds in subcooled liquid nitrogen ($-210 \text{ }^\circ\text{C}$), then fractured open by microtome cleavage. The exposed resin surface was mounted and then sputter coated with gold for 30 s to generate the appropriate phase contrast for imaging.

5.4.3 Protein Production

The cloning of CBM9 and CBM9-GFP is reported elsewhere (Kavoosi et al. 2004). CBM9 or CBM9-GFP was produced in a 60 L fermentation as follows. BL21 (DE3) cells containing the pET28-CBM9-GFP expression vector were grown at 37°C in Luria broth (LB) to a cell density (OD_{600}) between 0.8 and 1.0. Protein expression was induced with isopropyl-1-thio- β -D-galactoside (IPTG) to a final concentration of 0.1 mM and the cells allowed to incubate for a further 10-12 h at 30°C. The cell pellet was resuspended in high salt buffer (1 M NaCl, 50 mM potassium phosphate, pH 7.0), ruptured by two passages through a French pressure cell (21000 lb in⁻²), and the cell debris was removed by centrifugation (27,000 x g) for 30 min at 4°C. Highly pure CBM9 or CBM9-GFP was obtained by first passing the clarified cell extract over a Pharmacia XK-16 column (10 cm x 1.6 cm I.D.) packed with the Perloza™ MT100-based composite media. Contaminating proteins were removed by washing the column with 10 column volumes (CV) of high salt buffer, followed by 5 CV of low salt buffer (150 mM NaCl, 50 mM potassium phosphate, pH 7.0). CBM9 or CBM9-GFP was then eluted from the column with 2 CV of 1 M glucose in TBS8 (15 mM NaCl, 10 mM Tris-HCl, pH 8.0) and the eluent peak was injected into a column packed with Ni⁺²-Sepharose IMAC media and purified according to the manufacturer's instructions. The protein eluted from the IMAC column was buffer exchanged into low salt buffer, concentrated in a stirred ultrafiltration (UF) unit (Amicon, Beverly MA) and stored at 4°C until use. The concentration of the purified protein was determined by UV absorbance (280 nm) using a calculated molar extinction coefficient of 43100 M⁻¹ cm⁻¹ (CBM9) or 62870 M⁻¹ cm⁻¹ (CBM9-GFP) (Mach et al. 1992).

5.4.4 Equilibrium binding isotherms

Equilibrium isotherms for binding of CBM9 and CBM9-GFP to Perloza™ MT100 and to the Perloza™ MT100/G15 composite media were measured at pH 7.0 and 4°C. Purified protein at concentrations ranging from 1 to 30 μ M was mixed with stationary phase media (1 mg dry weight) in low salt buffer to a final volume of 1 ml. Samples were then incubated overnight under continuous end-over-end rotation. The

media was removed by centrifugation (27,000 x g) for 16 min. The supernatant was collected and the concentration of free protein was determined by UV absorbance (280 nm) using a Cary 100 Spectrophotometer (Varian).

The resulting isotherm was generated by plotting the concentration of bound protein ($\mu\text{mol g}^{-1}$ dry resin), determined by total mass balance, against the concentration of free protein (mol L^{-1}). Binding parameters were then determined by non-linear regression of the Langmuir adsorption isotherm equation to the experimental data using GraphPad Prism 3.0 software.

5.4.5 Confocal Laser Scanning Microscopy (CLSM)

Time-course fluorescence intensity profiles for uptake of CBM9-GFP into Perloza™ MT100 media packed into an optically transparent two-dimensional chromatography column were measured according to the procedures described in (Hubbich et al. 2002) using an inverted Zeiss LSM 510 confocal laser scanning microscope equipped with a water immersion 63x/NA1.2 C-Apochromat (Zeiss) objective and an argon laser. The GFP chromophore has an absorbance maximum at 475 nm. Excitation was conducted at 488 nm, on the red-shifted shoulder of the absorbance peak, to attenuate the resulting fluorescence emission at 505 nm to avoid saturation of the signal. The laser intensity was kept constant for all experiments. Slight adjustments (± 20 V) of the photomultiplier (PMT) detector gain were necessary to account for signal attenuation effects from neighboring particles in the packed bed. An 80/20 filter and no-band or long-pass filter before the PMT were used to improve signal-to-noise. The fluorescence intensity profile within the central particle was measured every second with 300-500 frames in total. All profiles were stored as eight-bit single scans with a resolution of 512 x 512 pixels representing an area of 146.2 x 146.2 μm^2 . The chosen time interval allowed monitoring of the diffusion both into and out of the particle in one run. Bleaching of GFP was not observed. Before each individual run, the reflection mode of the microscope was used to verify that the focal plane went through the center of the particle. The image frame consisted of the particle of interest, other particles (focal plane

not necessarily through particle center), and interstitial areas between the particles, from which the average bulk protein concentration was determined.

5.4.6 Characterization and Application of Perloza™ MT100/G15 Composite Media Column

A Pharmacia Inc. (GE Healthcare) FPLC system with two P-500 reciprocating pumps, an 8-port mixing and injection valve, a UV-MII flow spectrophotometer (monitoring absorbance at 280 nm), and a Frac-200 fraction collector was used to measure all chromatograms. Columns 7 to 10 mL in volume (Pharmacia Inc. HR-10 column (1.0 cm I.D.) were packed with either degassed Perloza™ MT100 or a 50% Perloza™ MT100 (by mass)/50% Sephadex G15 degassed slurry under a superficial velocity u of $4.25 \times 10^{-4} \text{ m s}^{-1}$ and then equilibrated with degassed loading buffer (50 mM potassium phosphate, 150 mM NaCl, 0.02% NaN₃, pH 7.0) prior to use.

In all columns used, the column length L to diameter d_c ratio (L/d_c) was maintained well above 2 to minimize end effects (Jungbauer 1993). In addition, the volume of solute pulses used for moment analysis was kept well below 0.5% of the column void to minimize any precolumn solute dispersion effects (Ladisich et al. 1984). Wall effects could be ignored since the column diameter to particle diameter ratio was much higher than 30 (Knox et al. 1976).

5.5 Results and Discussion

5.5.1 Geometric and Sorption Properties of Perloza™ MT100/Sephadex G15 Composite Media

Perloza™ MT100 is a highly porous, hydrophilic media derived from regenerated cellulose with its structural elements comprised of partially microcrystalline regions stabilized by interchain hydrogen bonds (Iontosorb, CZ). Scanning electron micrographs (SEM) of the spherical MT100 particles (Figure 5.1A) show that this media offers a network of primarily submicron pores with a small percentage of larger pores up to 2 μm in nominal diameter (Figures 5.1B and 5.1C). Simple hydrodynamic calculations

proposed by Liapis (Heeter and Liapis 1997) predict for the 80- μm particles and range of linear solvent velocities used in this work that intraparticle convective flow requires pores greater than *ca.* 4 to 5 μm in nominal diameter. Thus, diffusion is the primary solute transport process within this stationary phase media.

Adsorption of CBM9 and CBM9-tagged fusion proteins to Perloza™ MT100 has previously been shown to follow Langmuir-type adsorption behavior (Kavoosi et al. 2004). Efficient purification from clarified *E. coli* cell lysates of CBM9-tagged fusion proteins on analytical-scale MT100 columns (*i.e.* less than 5 ml in volume) has also been demonstrated (Kavoosi et al. 2004). Due to the compressible gel-like nature of the MT100 matrix, scale-up of a packed MT100 column to a volume above *ca.* 10 to 15 mL is compromised by stationary-phase compression effects that degrade column performance. Mechanical stabilization of the stationary phase is therefore required and can be achieved through addition of an inert support media, in this case Sephadex G15, that provides a rigid mechanical scaffold stabilizing the active stationary phase to allow stable columns to be prepared on the semi-preparative to preparative scale. Here, however, our focus is on validation of our proposed generalized two-zone model of adsorptive chromatography.

Equilibrium isotherms for pure CBM9-GFP binding to a 50/50 (by dry mass) composite media of MT100/G15 (Figure 5.2), and regression of the Langmuir isotherm parameters (Table 5.1) to those data show that the addition of the G15 mechanical support does not affect the equilibrium association constant (K_a) for CBM9-GFP binding to MT100 but reduces the static capacity of the composite media (q_i^{sat}) to half its value when the stationary phase consists of pure Perloza™ MT100. This is expected since q_i^{sat} is reported in terms of the total volume of the stationary phase, only half of which is MT100 in the composite media suitable for scale-up. Nevertheless, due to the extraordinarily high binding capacity of MT100 for CBM9 tagged fusion proteins (*ca.* 11-12 $\mu\text{mol g}^{-1}$ dry resin), the MT100/G15 composite column is capable of binding 45 to 60 mg of CBM9-GFP per column mL, making it highly competitive with popular commercial affinity chromatography media. Finally, control experiments confirm that the binding interaction is between CBM9 and cellulose (Table 5.1). Interactions between

CBM9-GFP and Sephadex G15 and between untagged GFP and Perloza™ MT100 are insignificant.

5.5.2 Purification of CBM9-GFP on Perloza™ MT100/G15 Composite Media Column

The chromatogram and gel documentation for purification of CBM9-GFP from a clarified cell lysate are shown in Figure 5.3 and Figure 5.4, respectively. Eluent from the column was continuously monitored both for total protein using absorbance at 280 nm (A_{280}) and for CBM9-GFP using the intrinsic fluorescence of GFP (excitation at 395 nm and emission at 509 nm). Initial breakthrough of contaminating proteins was observed just after 1 column void volume. The small fluorescence peak observed within the contaminant breakthrough peak is due to the presence in the feed mixture of free GFP, released as a result of a small amount of degradation (no protease inhibitors were used during the purification) within the linker region connecting CBM9 to GFP (Kavoosi et al. 2007). Two wash steps were used to effectively remove most of the contaminating proteins present in the column. As evident by the overlapping A_{280} and fluorescent signals (Figure 5.3), addition of 1 M glucose to the mobile phase elutes CBM9-GFP as a single sharp peak.

Table 5.2 compares the yield, purity and concentration factor for the affinity purification of CBM9-GFP on the MT100/G15 composite column relative to that achieved on the pure MT100 column (Kavoosi et al. 2004). The performance of the two columns is very similar at the relatively small column scales used in this study (10 mL composite column, 7 mL pure MT100 column).

5.5.3 Characterization of Solute Mass Transfer Within Perloza™ MT100/G15 Composite Columns

Measured solute (CBM9-GFP) mass-transfer and column-geometry parameters for our proposed generalized two-zone model for mass transport in a Perloza™ MT100/G15 composite media column are listed in Table 5.3. Good column packing uniformity, defined according to the guidelines proposed by Bristow and Knox (Bristow

and Knox 1977), is indicated both by a peak asymmetry factor (A_s) close to unity and a reduced plate height value (h) close to 3 for pulse injection of 100 μL at a superficial velocity of $2.1 \times 10^{-3} \text{ cm s}^{-1}$.

The column void fraction ε was determined by pulse injection of blue dextran (MW = 2,000,000 Da) as a function of the superficial velocity, u (m s^{-1}). Elution peaks for this large, non-binding solute were Gaussian or very nearly Gaussian in shape. Direct computation of the first moment (μ_1) of each elution peak and application of the theory of Haynes and Sarma (Haynes and Sarma 1973)

$$\mu_1 = \frac{L}{u} [\varepsilon + (1 - \varepsilon)\varepsilon_p] \quad (5.15)$$

yielded ε from the slope of a plot of μ_1 versus u^{-1} (Figure 5.5) under the approximation that $\varepsilon_p = 0$. The estimated value of ε is slightly higher than the value expected for a column packed uniformly with spherical beads (Ladisich et al. 1984), but agrees well with an independent measure of ε obtained from application of the Blake, Kozeny and Carmen (Allen 1997) equation

$$\Delta P = 36k \frac{u L \eta (1 - \varepsilon)^2}{\varepsilon d_p^2 \varepsilon^3} \quad (5.16)$$

for pressure drop data across the column measured as a function of u , where η is the fluid viscosity ($\text{g m}^{-1} \text{ s}^{-1}$). For spherical packing, the aspect factor k is assumed equal to 5 (Janson and Rydén 1998).

Protein-based probes were used to determine the effective porosity ε_p of the MT100/G15 composite media as a function of protein molecular mass (Figure 5.6). Each data point in the figure was determined by measuring μ_1 as a function of u for the respective marker and application of equation 5.15 using the measured void fraction of 0.425. Interpolation between measured ε_p data was facilitated by fitting to an exponential decay type equation of the form:

$$\varepsilon_p = \alpha e^{-M/\gamma} - \delta \quad (5.17)$$

where α , δ and γ are fitted parameters and M is the protein molecular weight (kg mol^{-1}). The solid curve in Figure 5.6 represents the best fit, for which $\alpha = 0.815$, $\delta = 0.033$ and $\gamma = 207.8$.

Numerical determination of the second moment (σ^2) of the elution peak as a function of the interstitial velocity, u_o , was combined with the Laplace transform results of Haynes and Sarma (Haynes and Sarma 1973) for operation within the linear region of the adsorption isotherm

$$\frac{u_o \sigma^2 L}{2\mu_1^2} = D_L + u_o^2 \frac{1}{\kappa_M} \left(\frac{\varepsilon}{1-\varepsilon} \right) \left[1 + \frac{\varepsilon}{(1-\varepsilon)\varepsilon_p} \right]^{-2} \quad (5.18)$$

to obtain values for the parameters characterizing mass transfer of the CBM9-GFP fusion protein within the Perloza™ MT100/G15 composite media column. In equation 5.18, which has been used in many similar parameter estimation studies (*e.g.* (Giddings 1965; Guiochon et al. 1994; Ladisch 2001)), κ_M is the overall solute mass transfer coefficient, given by

$$\frac{1}{\kappa_M} = R_M = \frac{r_p}{3k_f} + \frac{r_p^2}{15\varepsilon_p D_p} + \frac{1}{k_{ads}} \left(1 + \frac{\varepsilon_p}{(1-\varepsilon_p)K_a} \right)^{-2} \quad (5.19)$$

where R_M is the overall resistance to solute mass transfer, k_f is the film mass-transfer coefficient (m s^{-1}) and k_{ads} is the sorption rate constant ($\text{M}^{-1} \text{s}^{-1}$). Equations 5.18 and 5.19 can be applied under both binding and nonbinding conditions to estimate the axial dispersion coefficient (D_L) ($\text{m}^2 \text{s}^{-1}$) and the overall mass transfer coefficient (κ_M) from the y-intercept and slope, respectively, of a plot of $u_o \sigma^2 L / (2\mu_1^2)$ versus u_o^2 . Results for CBM9-GFP under nonbinding conditions (*i.e.*, in the presence of 2-M glucose) are shown in Figure 5.7, from which the parameters in Table 5.3 were determined following estimation of the film mass transfer coefficient using the correlation of Wilson and Geankoplis (Wilson and Geankoplis 1966)

$$Sh = \frac{2r_p k_f}{D_M} = \left(\frac{1.09}{\varepsilon} \right) Re^{1/3} Sc^{1/3} \quad (5.20)$$

for solute mass transfer in a packed bed of porous, spherical particles at conditions where $0.0016 < Re < 55$ and $165 < Sc < 70600$. The correlation of Young and Carrood (Young et al. 1980) was used to estimate the bulk molecular diffusivity D_M ($m^2 s^{-1}$) of CBM9-GFP

$$D_M = 8.34 \times 10^{-8} T / \eta M^{1/3} \quad (5.21)$$

where T is the temperature (K) and M is the molecular mass of the solute in units of $g \text{ mole}^{-1}$. The estimate of k_f using this approach is in close agreement ($\pm 10\%$) with that estimated using either the correlation of Wakao et al. (Wakao et al. 1958) or of Goto et al. (Goto et al. 1983).

First and second moments analysis of the first derivative of breakthrough curve data for frontal loading of pure CBM9-GFP under binding conditions indicates that the last term on the right-hand side of equation 5.19 makes less than a 1% contribution to R_M . This indicates that $k_{ads} > 0.33 \text{ m}^3 \text{ mole}^{-1} \text{ s}^{-1}$, which is consistent with the value of k_{ads} determined by Jervis et al (Jervis et al. 1997) for the binding of the family 2a carbohydrate binding module (CBM2a) to the surface of crystalline cellulose. The measured mass-transfer and adsorption-kinetics parameters allow calculation of the column Peclet ($Pe = 2r_p u_o / D_p$), Biot ($Bi = k_f r_p / 3D_p$) and Damköhler ($Da = C_i^o k_{ads} r_p^2 / D_p$) numbers (Table 5.3), which together indicate that pore diffusion limits CBM9-GFP mass transport within the composite media column. As a result, the local equilibria approximation can be invoked for the column loading process, with elution band profiles and the rate of solute uptake within the stationary phase best described by an appropriately formulated pore diffusion model of chromatography. As noted previously, relatively slow rates of intraparticle diffusion often limit affinity chromatography processes as well as other forms of adsorptive chromatography (Jungbauer 1996; Ladisch 2001), particularly under conditions where the concentration of solute in the feed is significantly greater than $1/K_a$ (Hall et al. 1966), as was the case in

our experiments. The CBM9-GFP affinity capture process therefore provides a suitable system for investigating the advantages of treating intraparticle solute uptake in pore-diffusion-limited adsorptive chromatography processes using the proposed generalized two-zone model of affinity chromatography relative to using the traditional PDM for spherical sorbent particles encoded in equations 5.1 through 5.6.

5.5.4. CLSM-Derived Rates of CBM9-GFP Uptake

Time-course fluorescence-intensity profiles (Figure 5.8) for batch uptake of CBM9-GFP into a Perloza™ MT100 particle routinely show a two-zone behavior characterized by a region within the interior of the sorbent particle containing no protein, and a second region in which s_i is nonzero and increases with r . In this experiment, the interstitial volume of the 100 μL viewing chamber was rapidly flooded with CBM9-GFP feed solution to an initial concentration of 5.4 μM to permit fluorescence intensities both within the central particle and within the surrounding interstitial volume to be monitored as a function of time. Fluorescence intensities in the outer protein-containing shell of the stationary phase particle do not exhibit the square-wave characteristics predicted by the traditional shrinking-core model, where adsorption equilibrium is described by the rectangular isotherm. Instead, finite concentration gradients within the shell region are observed, whose description is better described through use of a more realistic equilibrium relationship such as the Langmuir-type isotherm used in the traditional PDM and the two-zone model proposed here.

Time-dependent radial profiles of CBM9-GFP uptake within a Perloza™ MT100 particle predicted by the PDM (Figure 5.9A) and by the new TZM (Figure 5.9B) are compared with normalized fluorescence intensities measured by CLSM for an interstitial feed concentration (C_i^0) of 5.4 μM . This feed concentration represents conditions within the linear portion of the adsorption isotherm, and both models accurately capture initial rates and profiles of CBM9-GFP uptake under this moderate loading condition. Divergence of the traditional PDM from the measured solute uptake profiles is observed after *ca.* 10 to 18 minutes of exposure, with the model predicting a faster rate of protein uptake due to an unrealistically rapid accumulation of solute near the center to the

stationary phase particle - an error created by the continuous nature of the model that requires both $c_i(r=0,t)$ and $q_i(r=0,t)$ to become nonzero immediately upon contact of the stationary phase particle with the mobile phase liquid. By eliminating this approximation and incorporating a realistic isotherm model (relative to the traditional shrinking core model), the TZM accurately predicts the experimentally observed rates and profiles of solute uptake within the Perloza™ MT100 particle throughout the loading process. This includes model predictions of $r_c(t)$, which agree with CLSM estimated values to within experimental error (Figure 5.10).

Errors in rates of CBM9-GFP uptake predicted by the PDM decrease with increasing protein load in the feed. When C_i^0 is increased to 49 μM , adsorption equilibrium lies in the nonlinear region of the isotherm, resulting in a more rapid penetration of solute into the sorbent particle and improved agreement of PDM predictions with both experiment and TZM predictions (Figure 5.11). Indeed, under nonlinear or overload conditions, model predictions are more sensitive to the choice of isotherm model than to improvements provided by the TZM.

5.5.5 Simulation of Breakthrough Curves

Figures 5.12A and 5.12B compare model predictions to normalized breakthrough curve data for the cases where binding equilibrium is within the linear and non-linear regions of the isotherm, respectively. Under linear binding conditions, the TZM agrees with experiment while the traditional PDM over-predicts the dynamic capacity of the column, resulting in a significant delay in the predicted onset of breakthrough. Both models provide a reliable prediction of breakthrough behavior when the concentration of CBM9-GFP in the feed is increased to nonlinear loading conditions due to the more rapid rate of protein uptake in the stationary phase. Model agreement, however, is not exact. The experimental breakthrough curve is slightly asymmetric, such that the leading edge of the breakthrough transition is a bit sharper than the approach to saturation. This slight asymmetry, present in the other breakthrough curves measured at low linear velocities, could be due to a number of factors not fully captured in either of the two models,

including non-uniform mixing in some regions of the bed, non-specific adsorption, or conformational changes in the adsorbed protein (Johnston and Hearn 1990).

TZM predictions are compared in Figure 5.13 to raw breakthrough data to show that the model captures changes in elution band profiles and dynamic capacity over a wide range of feed concentrations. The TZM also predicts the dependence of the elution band profile on u (Figure 5.14) and thus the expected changes in solute dispersion and dynamic capacity (Mao et al. 1991; Mao et al. 1995).

5.5.6 CBM9-GFP Breakthrough from a Clarified Cell Extract Feed

Figures 5.12 to 5.14 present breakthrough data for a binary buffer solution containing pure CBM9-GFP. We have previously shown that our custom expression vectors for expression of CBM9-tagged fusion proteins direct high-level production of the chimeric protein, typically yielding concentrations of soluble product between 0.4 to 5 grams per liter (Kavoosi et al. 2004). This high level expression is exemplified in Figure 5.4, where soluble CBM9-GFP comprises approximately half of the total protein within the clarified cell extract. Nonbinding contaminant effects on mass transport and affinity binding of the target protein are therefore likely to be relatively minor, suggesting that TZM predictions using the one-component Langmuir isotherm and pure-component transport parameters (Table 5.3) may be sufficient to predict product elution bands for a clarified cell extract feed containing CBM9-GFP.

Figure 5.15 compares predictions of this simplified TZM to breakthrough data for frontal loading of a clarified cell extract containing $25 \pm 0.7 \mu\text{M}$ CBM9-GFP onto a 7.5 ml Perloza™ MT100/G15 composite column at a superficial velocity of $8.5 \times 10^{-3} \text{ cm s}^{-1}$. The breakthrough of CBM9-GFP was continuously monitored using the intrinsic fluorescence of GFP (excitation at 395 nm, emission at 510 nm). The results confirm that our pseudo-binary solution assumption is accurate provided the fusion-protein makes up a high percentage of the total protein in the feed and the total protein load in the feed is not greater than *ca.* 3 g L^{-1} .

5.6 Conclusions

We have introduced a novel generalized two-zone model for adsorptive chromatography and compared it to the traditional pore-diffusion model for protein uptake within a porous stationary phase. The TZM divides each stationary phase particle into two zones, an inner protein-free core and an outer zone into which a finite mass of protein has penetrated. The TZM replaces the rectangular isotherm of the traditional shrinking-core model with Langmuir theory, thereby allowing for simultaneous intraparticle mass transport and sorbent loading within the outer region of the porous particle by accounting for the presence of a protein concentration gradient within this outer region. The model therefore improves upon the PDM by eliminating a boundary condition that forces the unrealistic prediction of a nonzero protein concentration at the center of the particle immediately upon contact of the empty particle with a protein-containing mobile phase. Confocal laser scanning microscopy results indicate that under linear loading conditions, both models accurately predict the initial rate of solute uptake. As loading time increases, the PDM deviates from experimental results due to errors associated with the continuous nature of the model, while the TZM continues to accurately predict experimental data. Under nonlinear loading conditions, both models performed well, with the choice of isotherm model becoming the more critical factor in determining the accuracy of the prediction. Finally, the TZM was able to predict product breakthrough over a range of feed concentrations and superficial velocities, including accurate prediction of product breakthrough during frontal loading of a clarified cell extract.

5.7 Tables

Table 5.1 Measured Langmuir isotherm parameters for binding of CBM9-GFP and each of its fusion partners to Perloza™ MT100, Sephadex G15, and the composite MT100/G15 media at 4 °C. The solvent consisted of 50 mM potassium phosphate, 150 mM NaCl, pH 7. NB indicates that no binding was observed.

Protein	Perloza™ MT100		MT100/G15 Composite		Sephadex G15	
	$K_a \times 10^6$ (M ⁻¹)	q_i^{max} (μmol/g)	$K_a \times 10^6$ (M ⁻¹)	q_i^{max} (μmol/g)	$K_a \times 10^6$ (M ⁻¹)	q_i^{max} (μmol/g)
CBM9-GFP	1.5 (±0.21)	9.93 (±0.31)	1.2 (±0.15)	5.27 (±0.18)	NB	NB
CBM9	1.2 (±0.06)	11.2 (±0.15)	0.85 (±0.15)	5.67 (±0.32)		
GFP	NB	NB				

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence interval)

Table 5.2 Yield, purity and concentration factor for the affinity purification of CBM9-GFP on a MT100/G15 composite column and on a pure MT100 column. Clarified *E. coli* cell lysate containing CBM9-GFP was loaded onto each column at a superficial velocity of $4.25 \times 10^{-3} \text{ cm s}^{-1}$. Fractions were collected and analyzed both by absorbance at 280 nm and by fluorescence intensity to obtain reported data. The total lysate volume loaded was 65 mL and contained a CBM9-GFP concentration of $5.7 \mu\text{M}$.

	MT100/G15 Column	MT100 Column
Purity*	> 95%	> 95%
Yield	$80 \pm 3\%$	$82 \pm 3\%$
Concentration Factor (C_i/C_i^0)	29 ± 2	31 ± 2

*Determined from SDS-PAGE analysis

Table 5.3 Measured mass-transfer and column-geometry parameters for CBM9-GFP transport in a Perloza™ MT100/G15 composite media column.

Column Property	Value	Units
L	7.9 - 11.4	cm
d_c	1.0	cm
$\langle d_p \rangle$	83.7 ± 0.6	μm
ρ_p	1.33 ± 0.73	kg m^{-3}
ε	0.425 ± 0.015	
ε_p	0.65 ± 0.03	
h (reduced plate height)	2.9 ± 0.3	
A_s (peak asymmetry factor)	1.20 ± 0.03	
D_L	3.0×10^{-5}	$\text{m}^2 \text{s}^{-1}$
k_f	8.45×10^{-6}	m s^{-1}
D_M	1.54×10^{-11}	$\text{m}^2 \text{s}^{-1}$
D_p	7.17×10^{-12}	$\text{m}^2 \text{s}^{-1}$
Pe	500-2000	
Bi	16.4	
Da	≥ 82	

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence interval)

5.8 Figures

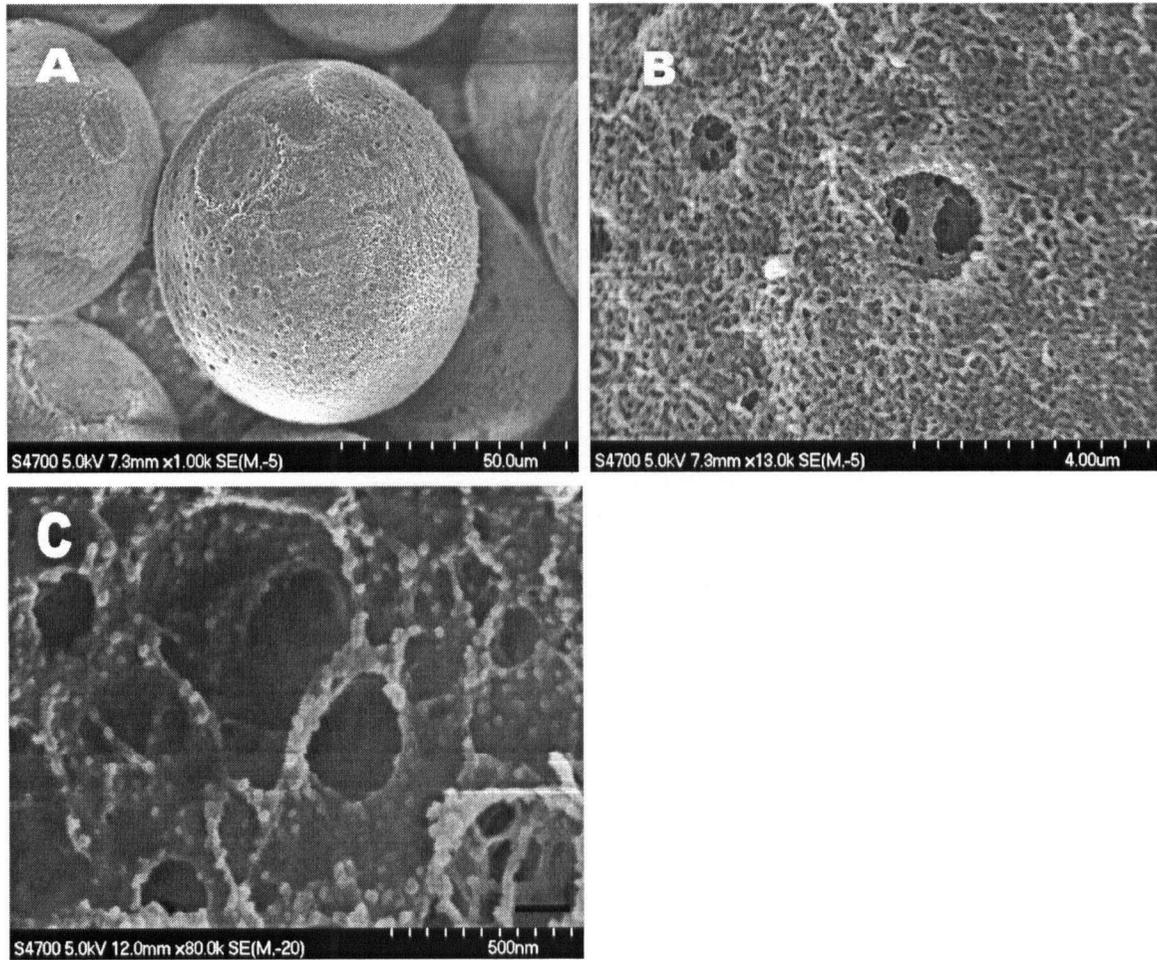


Figure 5.1 Scanning electron micrographs of Perloza™ MT100 beaded media. Figures A, B and C show an MT100 particle at 1.0k, 13.0k and 80.0k magnification, respectively. The pore structure shown at the center of Figure B is further magnified and shown in Figure C.

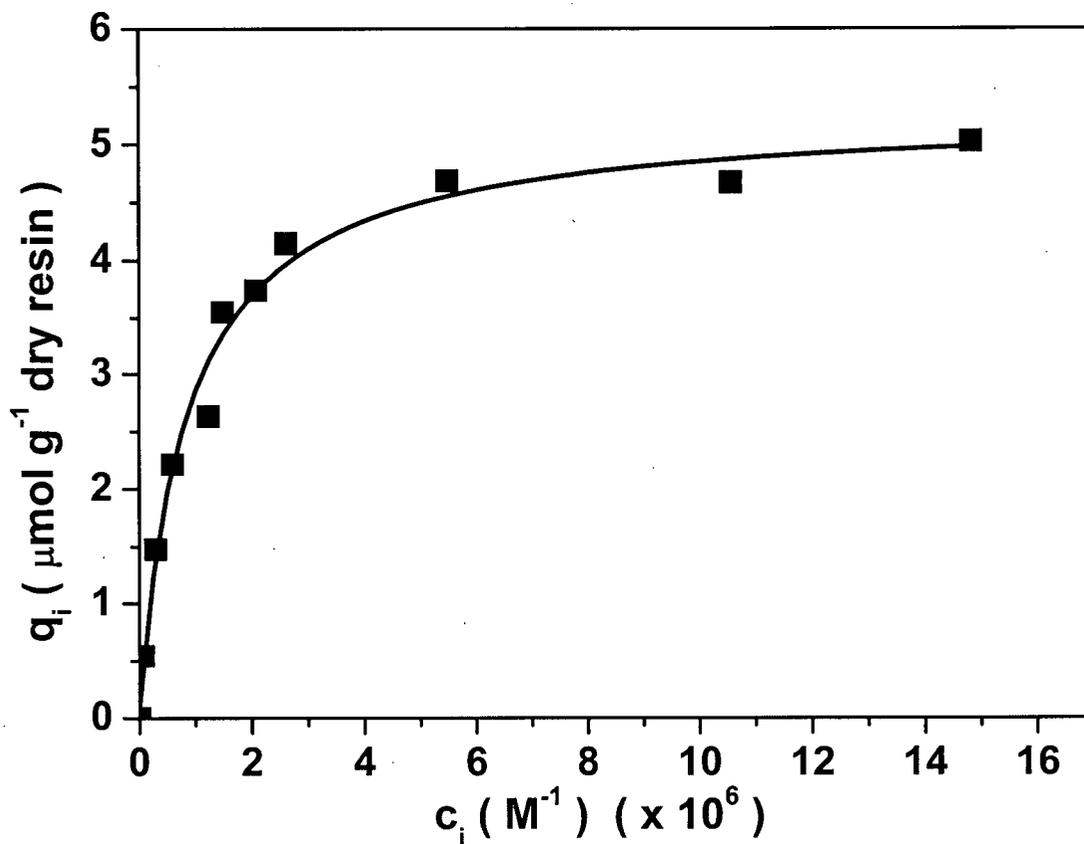


Figure 5.2 Equilibrium adsorption isotherm for batch binding of CBM9-GFP to Perloza™ MT100/G15 composite media at 4°C. The solvent consisted of 50 mM potassium phosphate, 150 mM NaCl, pH 7. The solid curve represents the best fit of the experimental data to the Langmuir isotherm equation.

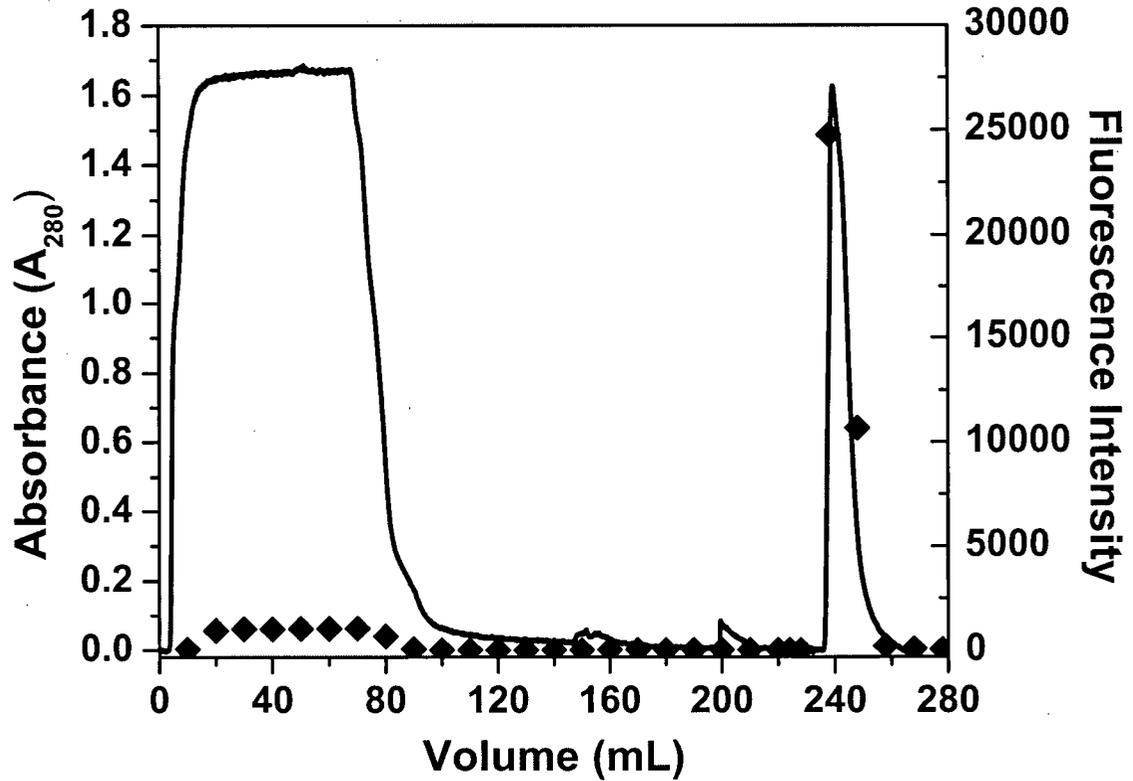


Figure 5.3 Chromatogram for CBM9-GFP purification on Perloza™ MT100/G15 composite media. Clarified *E. coli* cell lysate containing CBM9-GFP was loaded at a superficial velocity of $4.25 \times 10^{-3} \text{ cm s}^{-1}$ onto a 10 mL column. Fractions were collected and analyzed both by absorbance at 280 nm (line) and by fluorescence intensity (solid diamond) as shown.

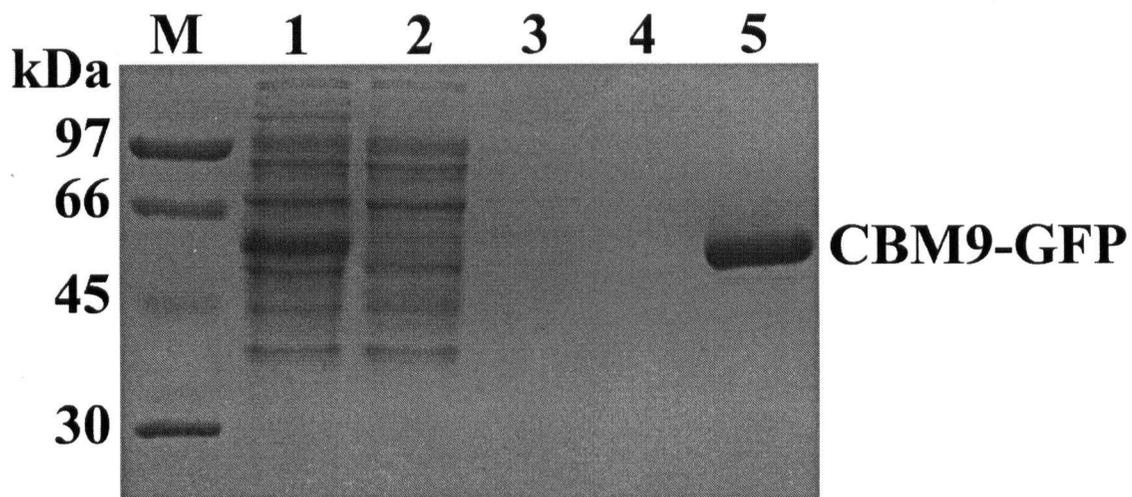


Figure 5.4 SDS-PAGE documentation of CBM9-GFP purification on a 10 mL Perloza™ MT100/G15 column. All samples were dissolved in sample buffer containing 10% SDS. Lane M: molecular mass markers; Lane 1: clarified cell lysate prior to column loading; Lane 2: column flow through; Lane 3: high salt wash; Lane 4: low salt wash; Lane 5: CBM9-GFP eluted in low salt buffer containing 1 M glucose.

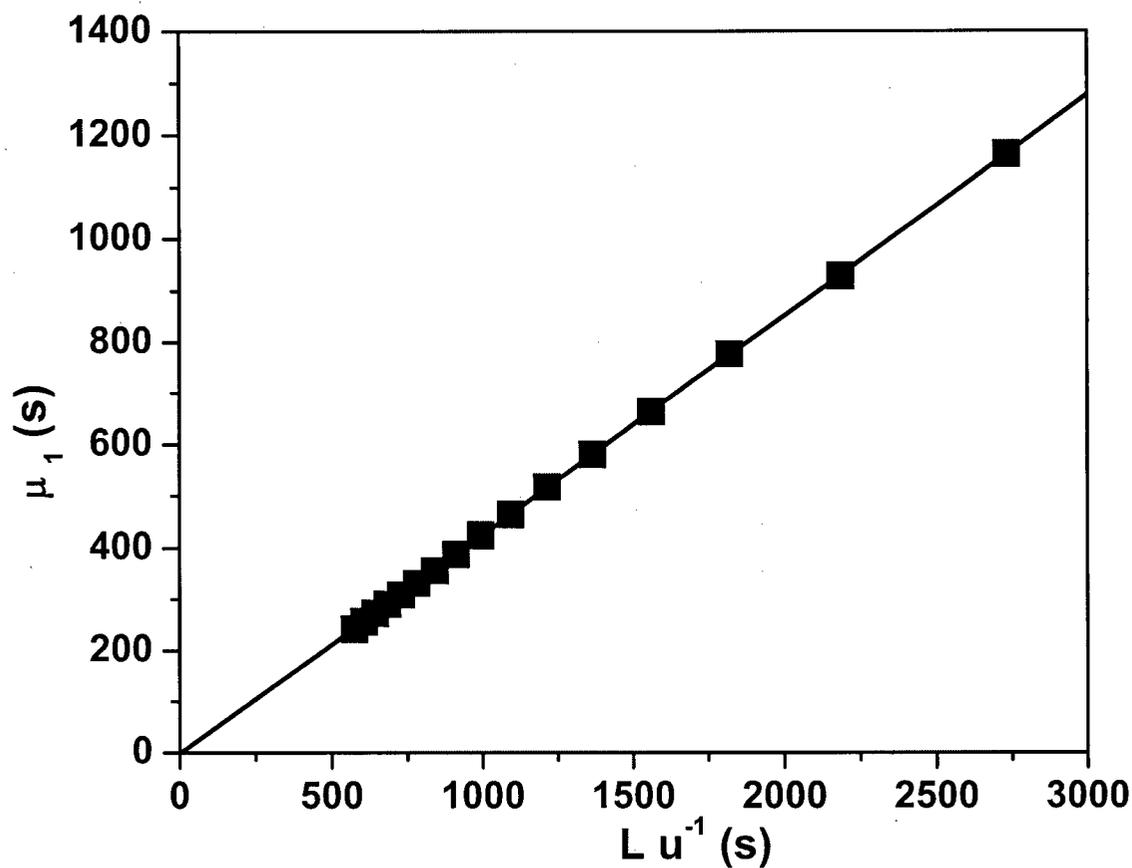


Figure 5.5 First moment (μ_1) analysis for a 10 mL column packed with Perloza™ MT100/G15. Integrated μ_1 values are reported for pulse injections of a 50 μL solution of blue dextran (MW 2000 kDa) over the interstitial velocity range $4.25 \times 10^{-3} \text{ cm s}^{-1}$ to $2.02 \times 10^{-2} \text{ cm s}^{-1}$. The mobile phase consisted of 50 mM phosphate buffer, 150 mM NaCl (pH 7, 4°C).

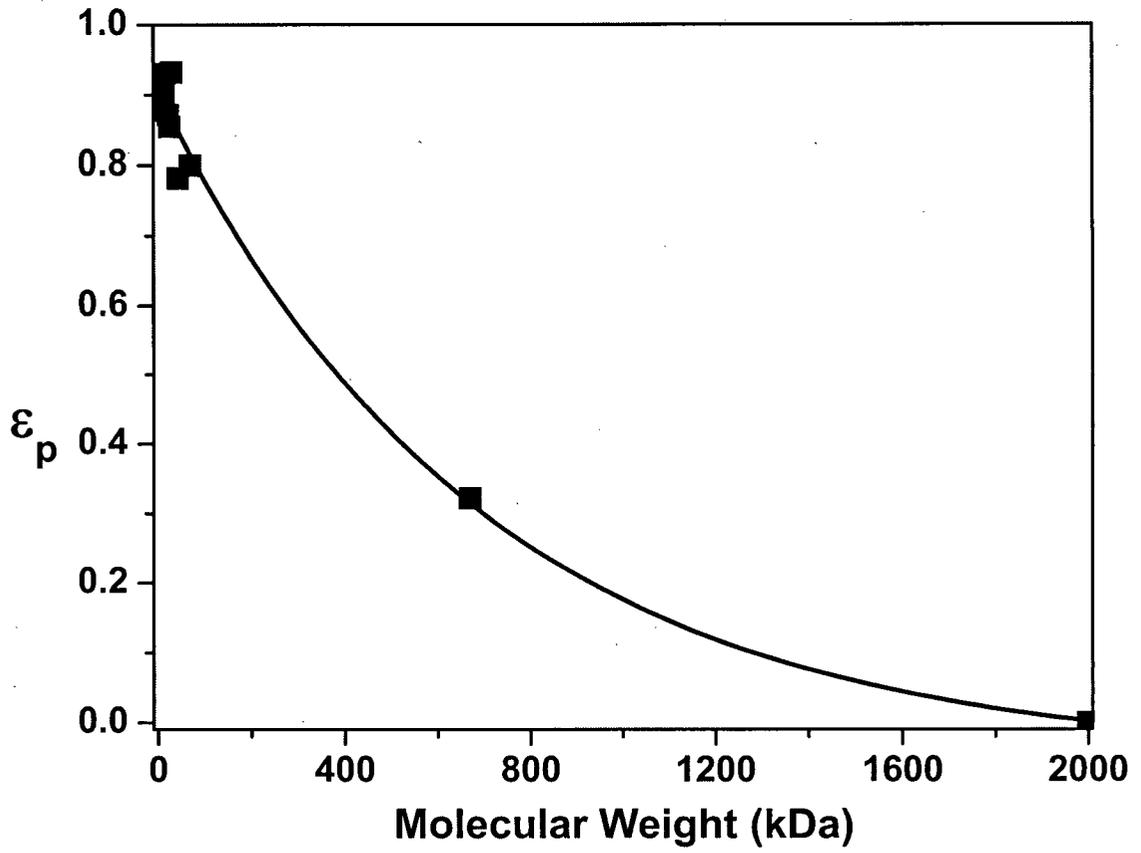


Figure 5.6 Measured average porosity ϵ_p of composite media as a function of molecular weight of standard proteins. Pulse injections of standard molecular weight protein markers were used over the interstitial velocity range $2.1 \times 10^{-3} \text{ cm s}^{-1}$ to $8.5 \times 10^{-3} \text{ cm s}^{-1}$. The porosity was determined from equation 5.15 using the measured void fraction of 0.425.

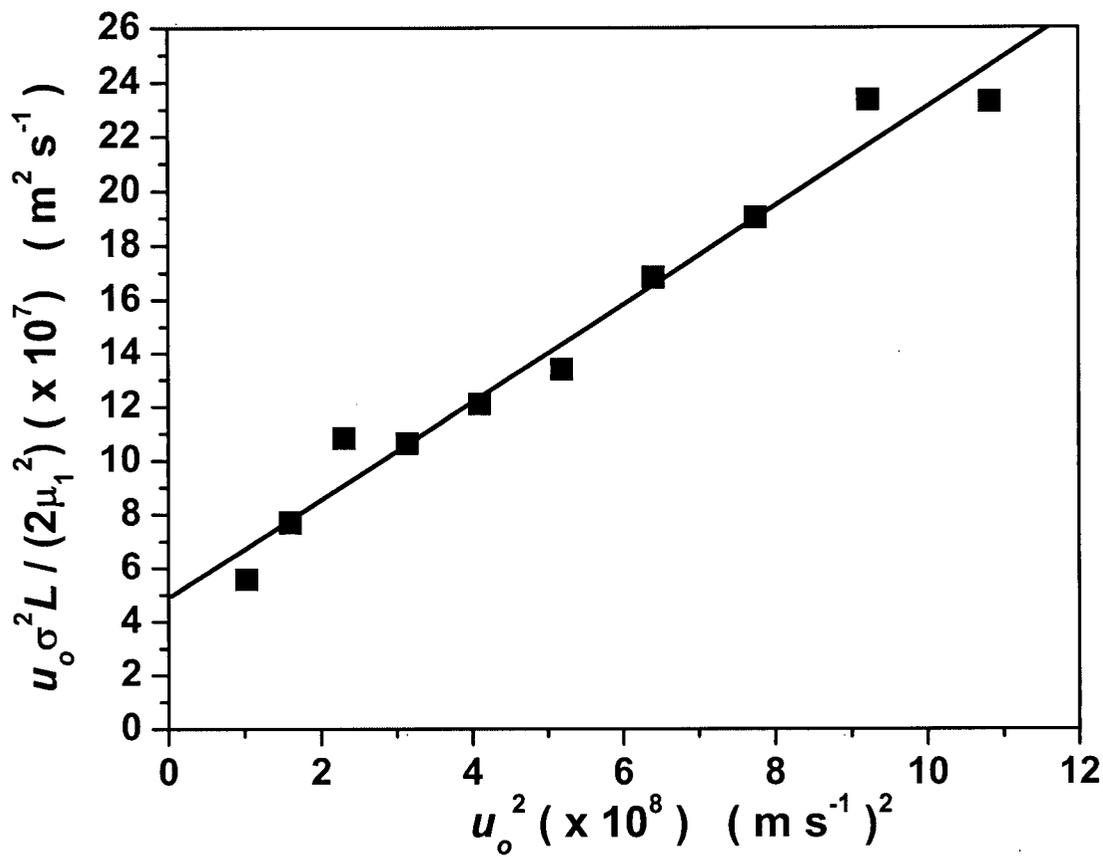


Figure 5.7 Determination of axial dispersion coefficient D_L and overall solute mass-transfer coefficient k_M for injection of CBM-GFP under nonbinding conditions onto a 10 mL column packed with Perloza™ MT100/G15. Pulse injections of a 50 μL solution of CBM9-GFP was used over the interstitial velocity range $4.25 \times 10^{-3} \text{ cm s}^{-1}$ to $1.38 \times 10^{-2} \text{ cm s}^{-1}$. The mobile phase consisted of 50 mM potassium phosphate, 150 mM NaCl (pH 7, 4°C) with 2 M glucose added to achieve nonbinding conditions.

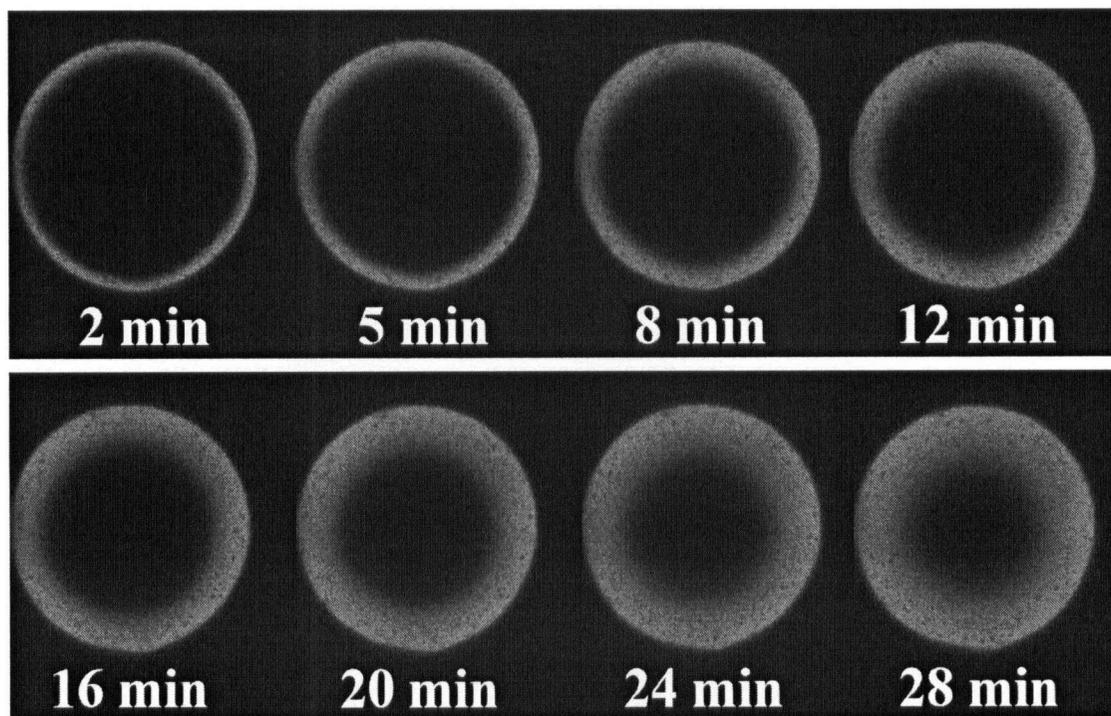


Figure 5.8 Time course fluorescent intensity profile of CBM9-GFP uptake into Perloza™ MT100 particle. Protein uptake monitored by an inverted Zeiss LSM 510 confocal laser scanning microscope with the center of the particle used as the focal plane. Excitation and emission wavelengths were 488 nm and 505 nm, respectively. The initial CBM9-GFP concentration C_i^0 outside the particle was 5.4 μM .

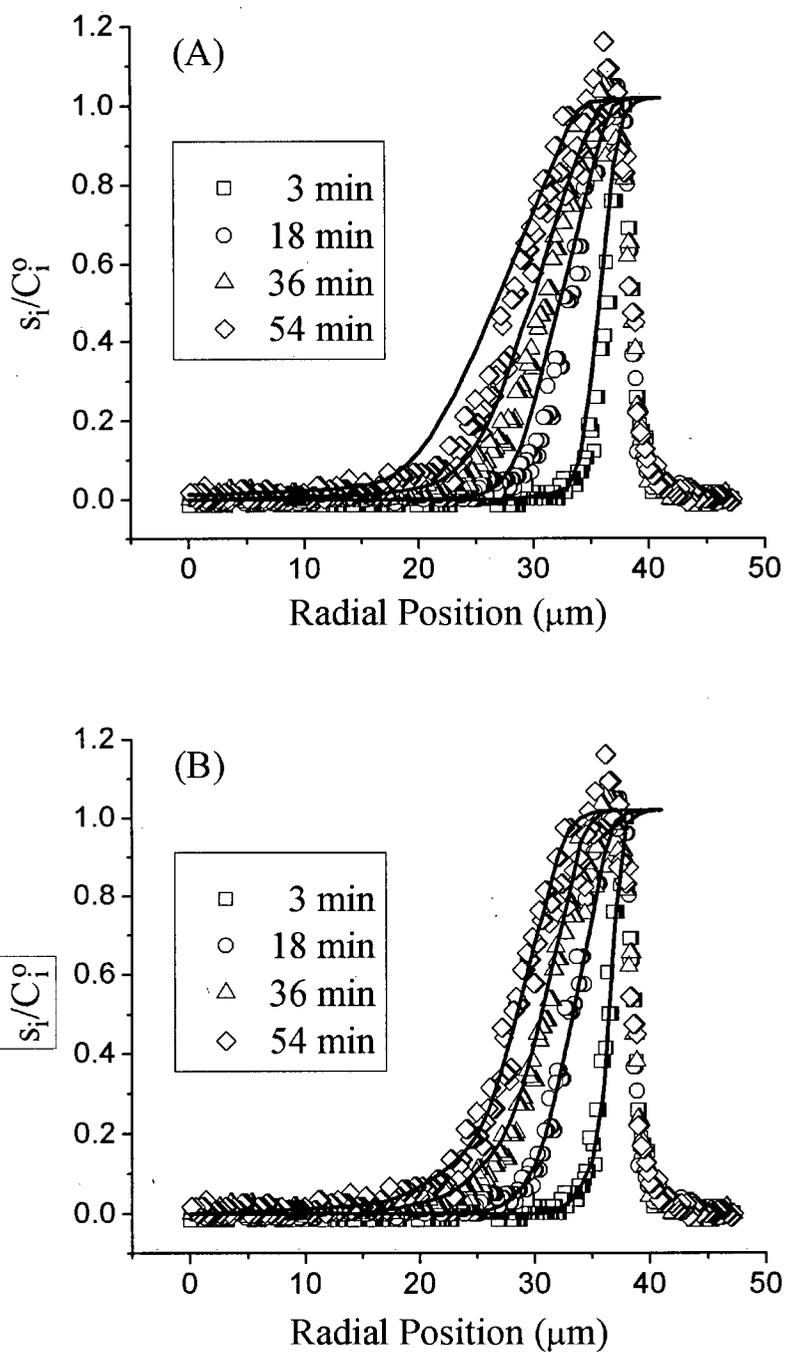


Figure 5.9 Time dependent radial profiles of CBM9-GFP uptake into a Perloza™ MT100 particle for a feed concentration of 5.4 μM . Predicted uptake rates using (Figure 9A) pore-diffusion model and (Figure 9B) two-zone model are compared with experiment. The rapid drop in fluorescence intensity at radial positions above *ca.* 39 μm indicates the position of the outer radius of the bead.

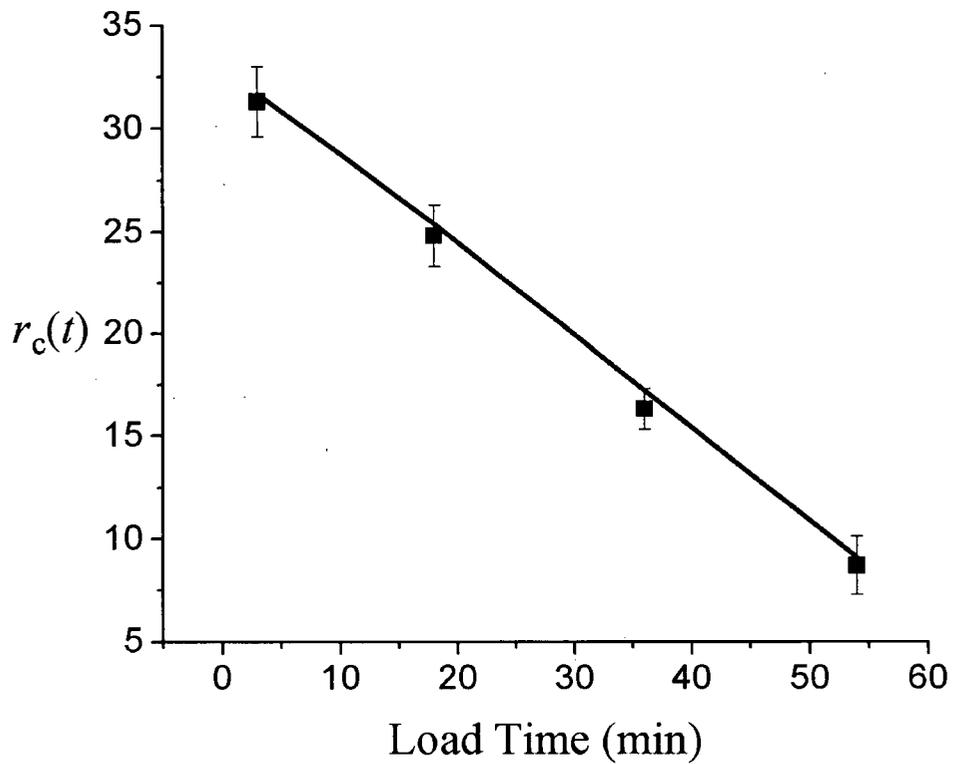


Figure 5.10 Comparison of TZM model predictions of $r_c(t)$ with values computed from CLSM data. CLSM determined core radius reported as the radius at which the measured fluorescence intensity falls below 3X the standard deviation of the background fluorescence. Load conditions same as stated in Figure 5.9.

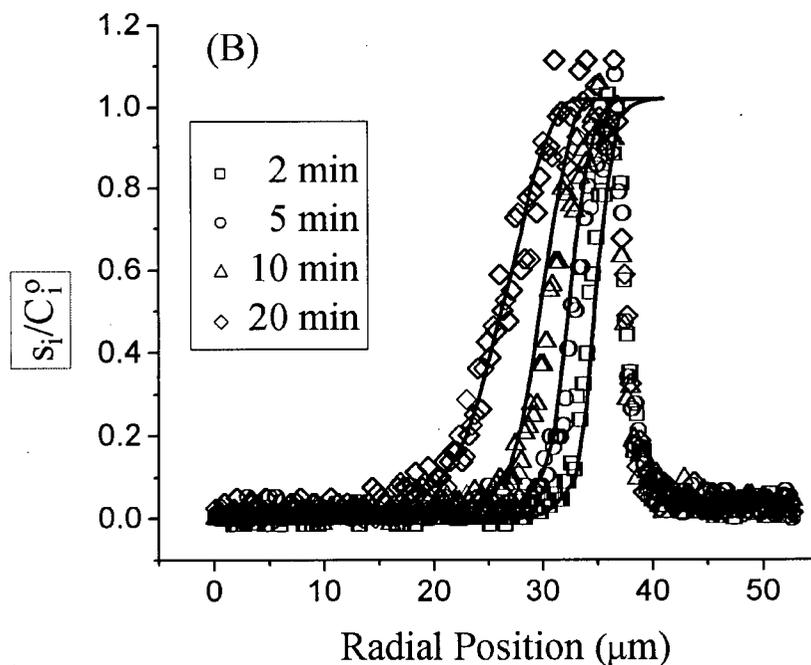
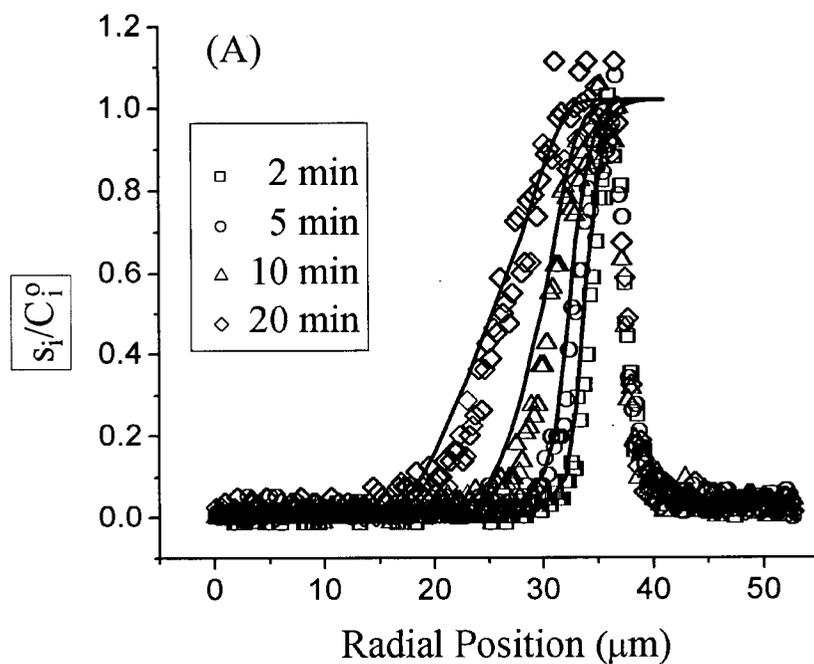


Figure 5.11 Time dependent radial profiles of CBM9-GFP uptake into a Perloza™ MT100 particle for a feed concentration of 49 μM . Predicted uptake rates using (Figure 11A) pore-diffusion model and (Figure 11B) two-zone model are compared with experiment.

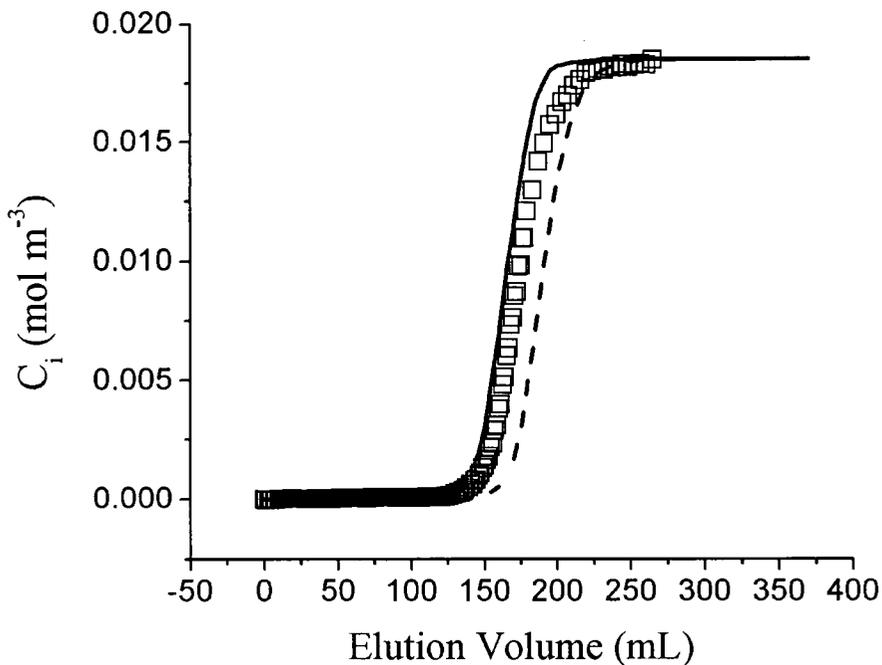
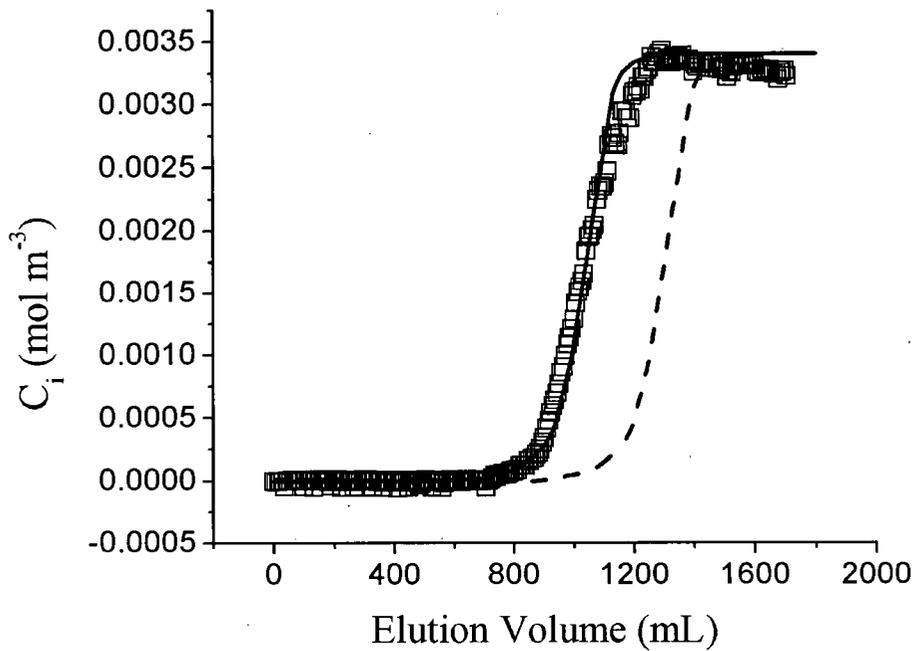


Figure 5.12 Comparison of TZM (solid curve) and PDM (dashed curve) predictions with experimental (points) breakthrough curves. Pure CBM9-GFP loaded at a superficial velocity of $8.5 \times 10^{-3} \text{ cm s}^{-1}$ onto a Perloza™ MT100/G15 composite media column: (A) frontal load of $3.4 \times 10^{-3} \text{ mol m}^{-3}$ CBM9-GFP, (B) frontal load of $1.85 \times 10^{-2} \text{ mol m}^{-3}$ CBM9-GFP.

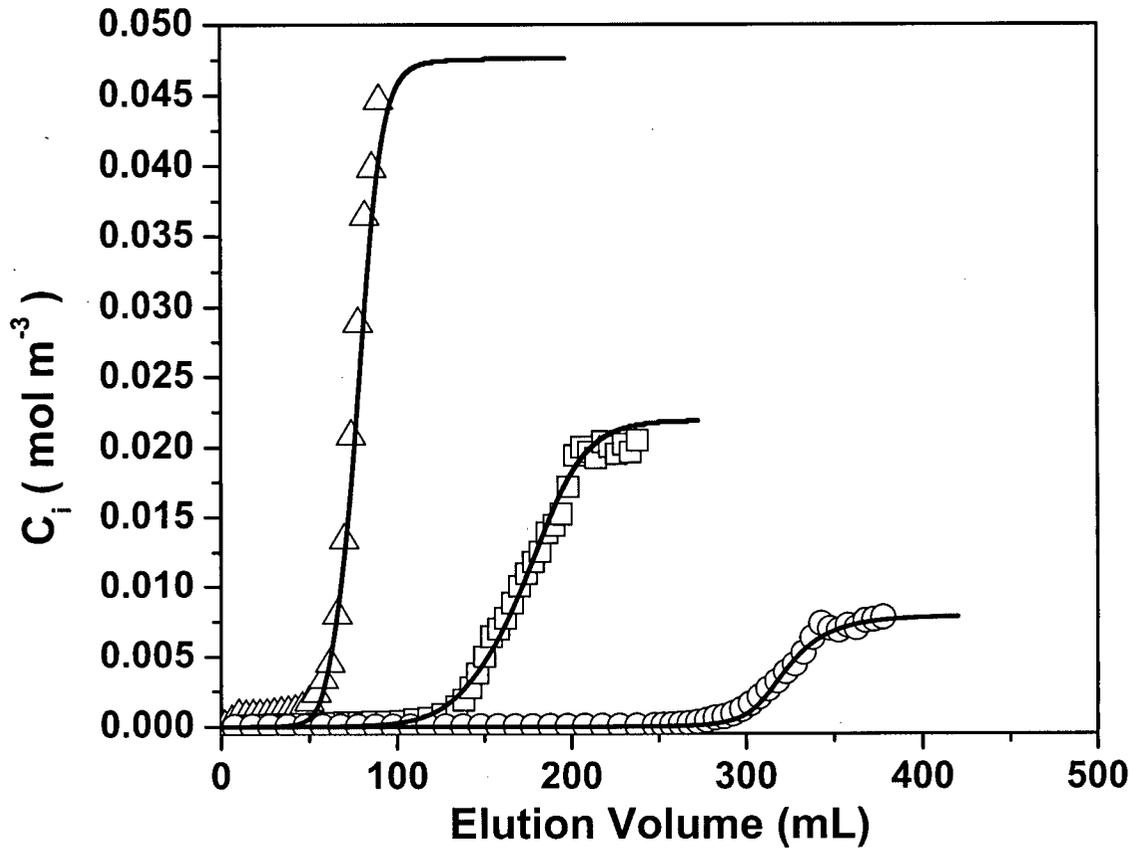


Figure 5.13 TZM predicted (line) and experimental (points) breakthrough curves for pure CBM9-GFP loaded onto a Perloza™ MT100/G15 composite media column at three different feed concentrations: $C_i^o = 4.77 \times 10^{-2} \text{ mol m}^{-3}$ (triangles), $2.2 \times 10^{-2} \text{ mol m}^{-3}$ (squares), and $8.0 \times 10^{-3} \text{ mol m}^{-3}$ (circles). Mobile phase loaded at a flow rate of 0.4 mL min^{-1} .

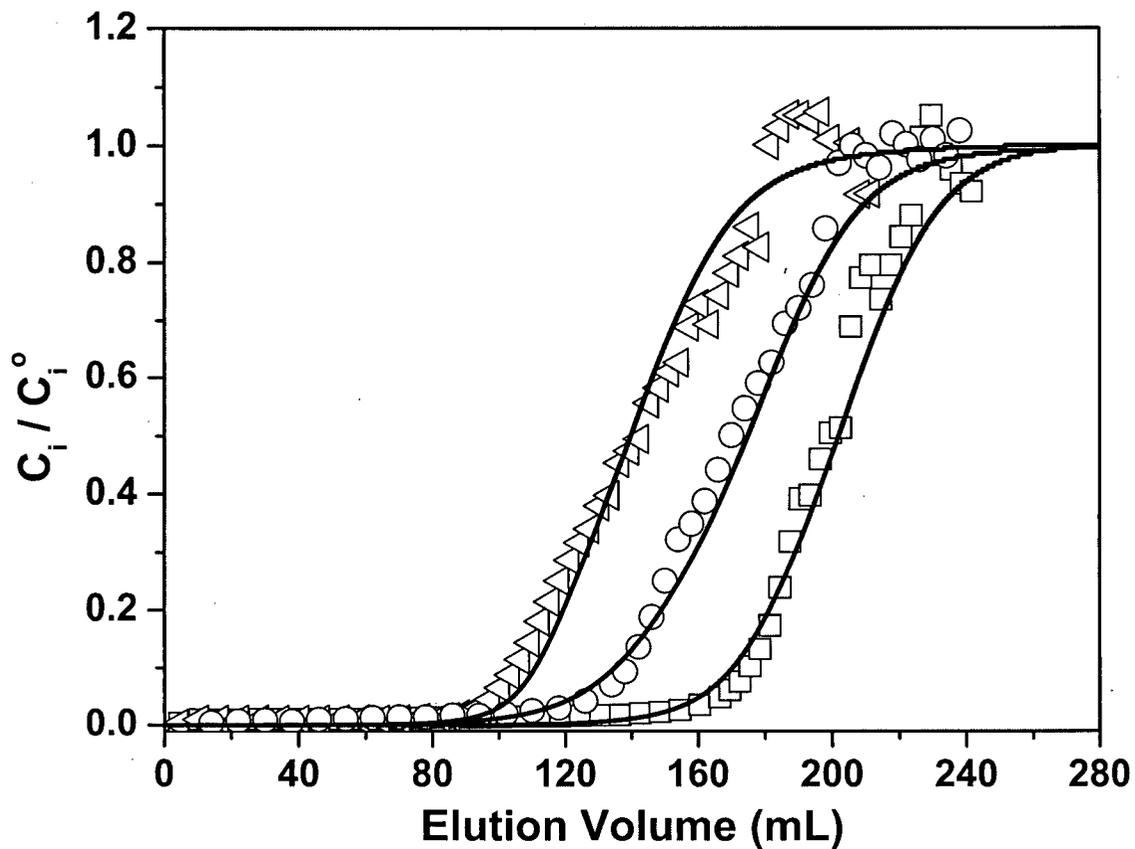


Figure 5.14 TZM predicted (line) and experimental (points) breakthrough curves as a function of interstitial velocity. Pure CBM9-GFP loaded onto a Perloza™ MT100/G15 composite media column: $u = 1.7 \times 10^{-2} \text{ cm s}^{-1}$ (triangles), $8.5 \times 10^{-3} \text{ cm s}^{-1}$ (circles), and $4.2 \times 10^{-3} \text{ cm s}^{-1}$ (squares).

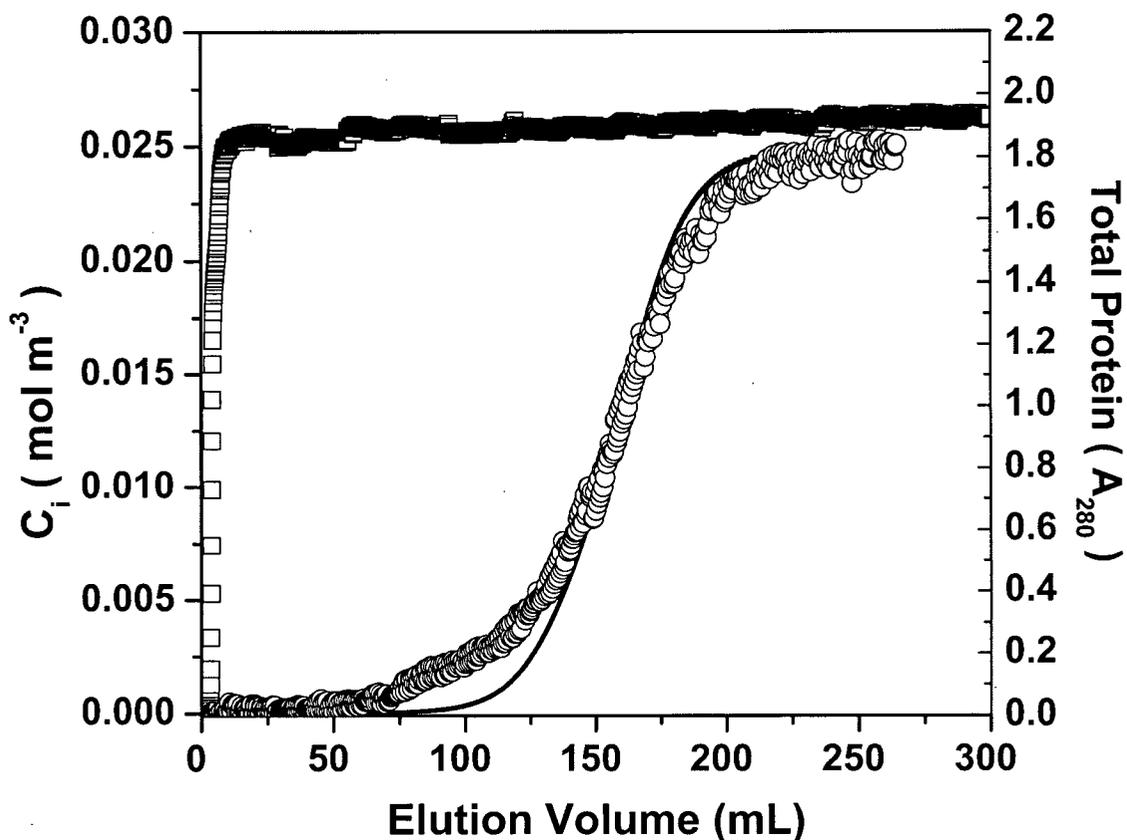


Figure 5.15 Predicted (line) and experimental (points) breakthrough curves for frontal loading of a clarified cell extract onto a Perloza™ MT100/G15 composite media column. The clarified cell extract contained 25 μM CBM9-GFP and was loaded at a superficial velocity of $8.5 \times 10^{-3} \text{ cm s}^{-1}$. Eluent absorbance data at 280 nm (open squares) are also shown to indicate total protein as a function of time.

5.9 References

- Allen, T. (1997). Powder Sampling and Particle Size Measurement. Particle Size Measurement. New York: Chapman & Hall. p 528
- Arnold, F. H.; Blanch, H. W.; Wilke, C. R. (1985a). "Analysis of affinity separations. I: Predicting the performance of affinity adsorbers." *Chemical Engineering Journal* 30(2):B9-23.
- Arnold, F. H.; Blanch, H. W.; Wilke, C. R. (1985b). "Analysis of affinity separations. II: The characterization of affinity columns by pulse techniques." *Chemical Engineering Journal* 30(2):B25-36.
- Arve, B. H.; Liapis, A. I. (1987). "The modeling and analysis of the elution stage of biospecific adsorption in fixed beds." *Biotechnology and Bioengineering* 30(5):638-649.
- Barber, P. W.; Hill, S. C. (1990). Light scattering by particles: computational methods. Singapore World Scientific. p 261.
- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Bristow, P. A.; Knox, J. H. (1977). "Standardization of Test Conditions for High-Performance Liquid-Chromatography Columns". *Chromatographia* 10(6):279-289.
- Carta, G.; Ubiera, A. R.; Pabst, T. M. (2005). "Protein mass transfer kinetics in ion exchange media: Measurements and interpretations". *Chemical Engineering & Technology* 28(11):1252-1264.
- Cramer, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. (1996). "Improved green fluorescent protein by molecular evolution using DNA shuffling". *Nature Biotechnology* 14(3):315-319.
- Crank, J.; Nicolson, P. (1996). "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type". *Advances in Computational Mathematics* 6(3-4):207-226.
- Crowe, J.; Dobeli, H.; Gentz, R.; Hochuli, E.; Stuber, D.; Henco, K. (1994). "6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification". *Methods in Molecular Biology* 31:371-387.

- Deutsch, D. G.;Mertz, E. T. (1970). "Plasminogen: purification from human plasma by affinity chromatography". *Science* 170(962):1095-1096.
- Einhauer, A.;Jungbauer, A. (2001). "The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins". *Journal of Biochemical and Biophysical Methods* 49(1-3):455-465.
- Farnan, D.; Frey, D. D.;Horvath, C. (1997). "Intraparticle mass transfer in high-speed chromatography of proteins". *Biotechnology Progress* 13(4):429-439.
- Farnan, D.; Frey, D. D.;Horvath, C. (2002). "Surface and pore diffusion in macroporous and gel-filled gigaporous stationary phases for protein chromatography". *Journal of Chromatography A* 959(1-2):65-73.
- Giddings, J. C. (1965). *Dynamics of Chromatography*. New York: Dekker.
- Goto, M.; Hayashi, N.;Goto, S. (1983). "Separation of Electrolyte and Non-Electrolyte by an Ion Retardation Resin". *Separation Science and Technology* 18(5):475-484.
- Guan, K. L.;Dixon, J. E. (1991). "Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Guiochon, G. (2002). "Preparative liquid chromatography". *Journal of Chromatography A* 965(1-2):129-161.
- Guiochon, G.; Shirazi, S. G.;Katti, A. M. (1994). *Fundamentals of preparative and nonlinear chromatography*. Boston: Academic Press.
- Hall, K. R.; Eagleton, L. C.; Acrivos, A.;Vermeulen, T. (1966). "Pore- and solid-diffusion kinetics in fixed-bed adsorption under constant-pattern conditions". *Industrial and Engineering Chemistry Fundamentals* 5(2):212-223.
- Haynes, H. W.;Sarma, P. N. (1973). "A model for the application of gas chromatography to measurements of diffusion in bidisperse catalysis". *AIChE Journal* 19:1043-1046.
- Heeter, G. A.;Liapis, A. I. (1997). "Estimation of pore diameter for intraparticle fluid flow in bidisperse porous chromatographic particles". *Journal of Chromatography A* 761(1-2):35-40.
- Holland, C. D.;Liapis, A. I. (1983). *Computer methods for solving dynamic separation problems*. New York: McGraw-Hill. p xvi, 475.
- Hopp, T. P.; Prickett, K. S.; Price, V. L.; Libby, R. T.; March, C. J.; Ceretti, D. P.; Urdal, D. L.;Conlon, P. J. (1988). "A short polypeptide marker sequence useful for

- recombinant protein identification and purification". *Bio/Technology* 6:1204-1210.
- Hubbich, J.; Linden, T.; Knieps, E.; Thoemmes, J.;Kula, M.-R. (2003). "Mechanism and kinetics of protein transport in chromatographic media studied by confocal laser scanning microscopy Part II. Impact on chromatographic separations". *Journal of Chromatography, A* 1021(1-2):105-115.
- Hubbich, J.; Linden, T.; Knieps, E.; Thommes, J.;Kula, M. R. (2002). "Dynamics of protein uptake within the adsorbent particle during packed bed chromatography". *Biotechnology and Bioengineering* 80(4):359-368.
- Janson, J. C.;Rydén, L. (1998). *Protein purification : principles, high-resolution methods, and applications*. 2nd ed. New York: Wiley. p x, 695.
- Jervis, E. J.; Haynes, C. A.;Kilburn, D. G. (1997). "Surface diffusion of cellulases and their isolated binding domains on cellulose". *Journal of Biological Chemistry* 272(38):24016-24023.
- Johnston, A.;Hearn, M. T. W. (1990). "High-performance liquid chromatography of amino acids, peptides and proteins. CIII. Mass transfer resistances in ion-exchange and dye-affinity chromatography of proteins". *Journal of Chromatography* 512:101-114.
- Johnston, A.;Hearn, M. T. W. (1991). "High-performance liquid chromatography of amino acids, peptides and proteins. CXIV. Protein interactions with porous coulombic sorbents: comparison of experimental findings with predictions of several adsorption models." *Journal of Chromatography* 557(1-2):335-358.
- Jungbauer, A. (1993). "Preparative Chromatography of Biomolecules". *Journal of Chromatography* 639(1):3-16.
- Jungbauer, A. (1996). "Insights into the chromatography of proteins provided by mathematical modeling". *Current Opinion in Biotechnology* 7(2):210-218.
- Jungbauer, A.;Hahn, R. (2004). "Engineering protein A affinity chromatography". *Current Opinion in Drug Discovery & Development* 7(2):248-256.
- Katoh, S.; Kambayashi, T.; Deguchi, R.;Yoshida, F. (1978). "Performance of affinity chromatography columns." *Biotechnology and Bioengineering* 20(2):267-280.
- Kavoosi, M.; Creagh, A. L.; Kilburn, D. G.;Haynes, C. A. (2007). "Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *E. coli*". *Biotechnology and Bioengineering* In Press.

- Kavoosi, M.; Meijer, J.; Kwan, E.; Creagh, A. L.; Kilburn, D. G.; Haynes, C. A. (2004). "Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T-maritima*". *Journal of Chromatography B* 807(1):87-94.
- Keefe, A. D.; Wilson, D. S.; Seelig, B.; Szostak, J. W. (2001). "One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag". *Protein Expression and Purification* 23(3):440-446.
- Knox, J. H.; Laird, G. R.; Raven, P. A. (1976). "Interaction of Radial and Axial-Dispersion in Liquid-Chromatography in Relation to Infinite Diameter Effect". *Journal of Chromatography* 122(Jul7):129-145.
- Ladisch, M. R. (2001). *Bioseparations engineering : principles, practice, and economics*. New York: Wiley.
- Ladisch, M. R.; Voloch, M.; Jacobson, B. (1984). "Bioseparations: column design factors in liquid chromatography". *Biotechnology and Bioengineering Symposium 14(6th Symposium on Biotechnology for Fuels and Chemicals)*:525-541.
- Li, P.; Xiu, G.; Rodrigues, A. E. (2003). "Modeling separation of proteins by inert core adsorbent in a batch adsorber". *Chemical Engineering Science* 58(15):3361-3371.
- Linden, T.; Ljunglof, A.; Hagel, L.; Kula, M. R.; Thommes, J. (2002). "Visualizing patterns of protein uptake to porous media using confocal scanning laser microscopy." *Separation Science and Technology* 37(1):1-32.
- Ljunglof, A.; Hjorth, R. (1996). "Confocal microscopy as a tool for studying protein adsorption to chromatographic matrices". *Journal of Chromatography A* 743(1):75-83.
- Lowe, C. R.; Lowe, A. R.; Gupta, G. (2001). "New developments in affinity chromatography with potential application in the production of biopharmaceuticals". *Journal of Biochemical and Biophysical Methods* 49(1-3):561-574.
- Ma, Z.; Whitley, R. D.; Wang, N. H. L. (1996). "Pore and surface diffusion in multicomponent adsorption and liquid chromatography systems". *AIChE Journal* 42(5):1244-1262.
- Mach, H.; Middaugh, C. R.; Lewis, R. V. (1992). "Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins". *Analytical Biochemistry* 200(1):74-80.
- Mao, Q. M.; Johnston, A.; Prince, I. G.; Hearn, M. T. W. (1991). "High-performance liquid chromatography of amino acids, peptides and proteins. CXIII. Predicting

- the performance of non-porous particles in affinity chromatography of proteins." *Journal of Chromatography* 548(1-2):147-163.
- Mao, Q. M.; Prince, I. G.; Hearn, M. T. W. (1995). "High-Performance Liquid-Chromatography of Amino-Acids, Peptides and Proteins .139. Impact of Operating Parameters in Large-Scale Chromatography of Proteins". *Journal of Chromatography A* 691(1-2):273-283.
- Martin, C.; Iberer, G.; Ubiera, A.; Carta, G. (2005). "Two-component protein adsorption kinetics in porous ion exchange media". *Journal of Chromatography A* 1079(1-2):105-115.
- Mulcahy, P.; O'Flaherty, M.; Jennings, L.; Griffin, T. (2002). "Application of kinetic-based biospecific affinity chromatographic systems to ATP-dependent enzymes: studies with yeast hexokinase". *Analytical Biochemistry* 309(2):279-292.
- Patankar, S. V. (1980). *Numerical heat transfer and fluid flow*. New York: McGraw-Hill. p xiii, 197.
- Patwardhan, A. V.; Atai, M. M.; Zenouzi, M. (1995). "A mathematical model of the immobilized metal affinity chromatography for protein separations: simulation of loading and elution conditions." *American Society of Mechanical Engineers, Heat Transfer Division* 322:61-68.
- Pinto, N. G.; Graham, E. E. (1987). "Application of the shrinking-core model for predicting protein adsorption". *Reactive Polymers, Ion Exchangers, Sorbents* 5(1):49-53.
- Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation". *Nature* 258(5536):598-599.
- Pritzker, M. D. (2003). "Model for parallel surface and pore diffusion of an adsorbate in a spherical adsorbent particle". *Chemical Engineering Science* 58(2):473-478.
- Shimomura, O.; Johnson, F. H.; Saiga, Y. (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*". *Journal of Cellular and Comparative Physiology* 59:223-239.
- Sirotti, D. A.; Emery, A. (1983). "Mass transfer parameters in an immobilized glucoamylase column by pulse response analysis." *Biotechnology and Bioengineering* 25(7):1773-1779.
- Skidmore, G. L.; Horstmann, B. J.; Chase, H. A. (1990). "Modeling Single-Component Protein Adsorption to the Cation Exchanger S Sepharose Ff". *Journal of Chromatography* 498(1):113-128.

- Smith, D. B.;Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase". *Gene* 67(1):31-40.
- Stofko-Hahn, R. E.; Carr, D. W.;Scott, J. D. (1992). "A single step purification for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase". *FEBS Letters* 302(3):274-278.
- Vaillancourt, P.; Simcox, T. G.;Zheng, C. F. (1997). "Recovery of polypeptides cleaved from purified calmodulin-binding peptide fusion proteins". *Biotechniques* 22(3):451-453.
- Wakao, N.; Oshima, T.;Yagi, S. (1958). "Mass transfer from packed beds of particles to a fluid". *Journal of Chemical Engineering of Japan* 22:780-785.
- Weisz, P. B.;Goodwin, R. D. (1963). "Combustion of carbonaceous deposits within porous catalyst particles. I. Diffusion-controlled kinetics". *Journal of Catalysis* 2(5):397-404.
- Wilson, D. S.; Keefe, A. D.;Szostak, J. W. (2001). "The use of mRNA display to select high-affinity protein-binding peptides". *Proceedings of the National Academy of Sciences of the United States of America* 98(7):3750-3755.
- Wilson, E. J.;Geankoplis, C. J. (1966). " Liquid mass transfer at very low Reynolds numbers in packed beds." *Industrial and Engineering Chemistry Fundamentals* 5(1):9-14.
- Winterhalter, C.; Heinrich, P.; Candussio, A.; Wich, G.;Liebl, W. (1995). "Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*". *Molecular Microbiology* 15(3):431-444.
- Young, M. E.; Carroad, P. A.;Bell, R. L. (1980). "Estimation of Diffusion-Coefficients of Proteins". *Biotechnology and Bioengineering* 22(5):947-955.
- Zheng, C. F.; Simcox, T.; Xu, L.;Vaillancourt, P. (1997). "A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins". *Gene* 186(1):55-60.

6 A Mechanically Stable Porous Cellulose Media for Affinity Purification of CBM9-Tagged Fusion Proteins

6.1 Introduction

Recombinant DNA technology and affinity chromatography have advanced the field of protein purification such that many proteins and peptides can now be purified without extensive *a priori* knowledge of the physicochemical properties of the desired product. By designing a chimeric protein incorporating a known affinity tag, the protein of interest can be purified from a culture supernatant in a single, highly selective capture step (Ford et al. 1991; Lowe et al. 2001). Affinity tags have been used extensively at the laboratory scale (Hearn and Acosta 2001; Nilsson et al. 1997), yet their application at the preparative scale has been limited, due in large part to the cost and stability (both chemical and mechanical) of the associated affinity matrices (Terpe 2003). For example, IMAC technology, based on the (His)₆ affinity tag, relies on a stationary phase bearing iminodiacetate or imidazole groups to which a transition metal ion may complex. The complexed metal ions act as the coordination sites for selective binding of (His)₆-tagged proteins (Porath et al. 1975; Rulisek and Havlas 2000). Although a popular and effective technology for laboratory-scale applications, the cost associated with the synthesis of the stationary phase along with concerns related to leaching of the bound metal ion into the eluent make this technology less attractive at preparative scales. Other popular affinity-tag systems such as the glutathione S-transferase tag (Guan and Dixon 1991; Smith and Johnson 1988) and the FLAG tag (Einhauer and Jungbauer 2001; Hopp 1988) likewise

* A version of this chapter has been accepted for publication in the *Journal of Chromatography A*. [Reference: Mojgan Kavooosi, Dexter Lam, Jenny Bryan, Douglas G. Kilburn, Charles A. Haynes, A mechanically stable porous cellulose media for affinity purification of CBM9-tagged fusion proteins. *J. Chromatography A*. In Press]

exhibit properties that limit their scalability to preparative applications (Linhult et al. 2005). The successful scale-up of tag-based affinity chromatography would therefore benefit from an affinity tag that provides efficient capture and elution on an inexpensive, yet chemically and mechanically robust stationary phase.

Kavoosi et al. (Kavoosi et al. 2004) recently introduced novel vectors for high-level production in *E. coli* of soluble chimeric proteins comprised of a target protein or peptide fused to CBM9, the family 9 carbohydrate binding module (CBM9) of xylanase 10A of *T. maritima*. CBM9 strongly binds both insoluble cellulose and soluble polysaccharides (Boraston et al. 2001), allowing for the specific capture of CBM9-tagged fusion proteins on a cellulose-based capture column and quantitative elution of the purified protein using an inexpensive soluble sugar such as glucose. The CBM9-cellulose affinity cassette is potentially attractive for preparative-scale applications because cellulose is inexpensive, chemically inert, and can be machined into well defined geometries offering high specific surface area (Rodriguez et al. 2004). Cellulose has also been approved for many pharmaceutical and human uses and is unlikely to introduce any harmful agents into the purified product (Hoenich et al. 1997; Varela et al. 2001). Cellulose beads have therefore been widely utilized as packing materials in chromatography, including serving as the base chemistry for a wide range of size-exclusion and ion-exchange media (Motozato and Hirayama 1984). Kuga (Kuga 1984) has reviewed preparation methods of cellulose beads and their separation properties.

CBM9 binds directly to the reducing ends of cellulose chains (Boraston et al. 2001), permitting most commercially available cellulose-based stationary phases to serve as the affinity media without the need for additional chemical processing. As an example, the porous cellulosic media Perloza™ MT100 (henceforth referred to as MT100) binds CBM9 strongly ($K_a \sim 10^6 \text{ M}^{-1}$), with very high capacity ($q_i^{\text{max}} \sim 500 \text{ mg protein/g dry media}$), and without significant fouling, so that repeated purification cycles from a clarified bacterial cell extract consistently result in high product purities and yields (Kavoosi et al. 2004). However, a limitation of highly porous cellulose-based matrixes such as MT100 is the tendency for bed compression at elevated flow rates or at larger bed dimensions (Colby et al. 1996b; Ostergren et al. 1998). Conventional cellulose

beads, including MT100, are characterized by relatively large particle-size distributions, resulting in an additional flow resistance in columns packed with these media that is associated with smaller particles occupying void spaces among larger ones. Pure cellulose media is also susceptible to compressive deformation, which can not only increase flow instabilities within the packed bed, reducing dynamic capacity and column performance, but can also significantly increase band broadening in extreme cases (Colby et al. 1996a; Mohammad et al. 1992). Scale-up of columns utilizing these media therefore requires a strategy to mechanically stabilize the packed bed.

Sephadex and other popular hydrophilic gel-like chromatographic media typically achieve mechanical strength through interchain cross-linking (Hjerten et al. 1987; Laas 1976; Leonard 1997). In contrast, the mechanical stability of Perloza™ MT100, a highly porous, hydrophilic media derived from regenerated cellulose, is only from interchain hydrogen bonds and dispersion forces between disordered and partially crystalline regions of its regenerated cellulose architecture (Iontosorb, CZ). This suggests that the mechanical stability of MT100 may be improved through cross-linking using an appropriate epoxide or halohydrin (Narayanan and Crane 1990). A particularly promising cross-linking agent for this purpose is epichlorohydrin, which yields hydrophilic linkages, is industrially proven, and is readily available and inexpensive. The resulting epoxide bond is stable at high pH, permitting stringent clean-in-place protocols using caustic. Epichlorohydrin is therefore the preferred crosslinking agent in the preparation of Sephadex (Holmberg et al. 1995) and its reaction chemistry is compatible with cellulose. Ruckenstein and Guo (Guo and Ruckenstein 2001), for example, have shown that cross-linking with epichlorohydrin can be used to increase the mechanical stability of cellulose-based filter paper.

This paper focuses on the use of response surface methodology (RSM) to identify optimal reaction conditions for synthesis of a mechanically stable, porous cellulose affinity chromatography media for use with CBM9 fusion-tag technology. Intraparticle cross-links are introduced into MT100 using epichlorohydrin-based epoxide chemistry with the concentrations of epichlorohydrin and dimethyl sulfoxide (DMSO) representing the two input variables. Treated and untreated media are compared both by monitoring

pressure drop across the column and by identifying the critical superficial velocity u_{crit} for column compression, with the latter being used as the response variable. A mechanically enhanced, cross-linked media is identified and moment analysis, equilibrium binding isotherms and HETP analysis are used to compare the binding and transport properties of this mechanically stabilized media against untreated MT100. The application of this cross-linked media to the affinity purification of a recombinant protein from *E. coli* is also demonstrated.

6.2 Materials and Methods

6.2.1 Reagents

Epichlorohydrin, glucose and all other reagents were analytical grade and purchased from Sigma-Aldrich (Mississauga, ON, Canada) unless otherwise stated. Perloza™ MT100 chromatographic media with a nominal particle diameter distribution of 50 to 80 μm was purchased from Iontosorb Inc. (Czech Republic). *E. coli* BL21 (DE3) was obtained from Novagen (Madison, WI, USA).

6.2.2 Cross-Linking of MT100

Cross-linking reactions were carried out in a 200 mL round-bottom flask immersed in a water-filled jacket. A PolyScience 910 constant-temperature water pump (Preston Ind, USA) circulated heated water through the jacket, maintaining the reaction temperature at 50°C. The reaction flask was charged with a magnetic stir bar and appropriate volumes (mL per gram MT100 cellulose) of 5.0 M sodium hydroxide, epichlorohydrin and DMSO were added. The reagents were allowed to mix and thermally equilibrate before the addition of an aliquot of MT100 slurry equivalent to 2 g of dry media. The resulting mixture was well stirred throughout the reaction, which was allowed to proceed for 2 hours before being quenched with acid addition to neutralize pH and immediate immersion in an ice bath. The cross-linked media was then filtered and repeatedly washed with 200 mL of water before being degassed as a dilute slurry.

The cross-linking reactions were continuously mixed to prevent formation of interparticle linkages and to ensure that all cross-linking occurred inside the resin

particle. Lack of interparticle cross-linking was verified by optical microscope imaging of the reaction products.

6.2.3 Hydrodynamic Characterization

MT100 or cross-linked MT100 media were packed into a Pharmacia HR10 column (I.D. = 1.0 cm) at an elevated flow rate using standard inclined pouring techniques. A silica filter pad was positioned on top of the bed and the flow distributor lowered and secured to achieve finger-pressure compression of the bed. The final bed heights ranged between 8.5 and 10.6 cm.

The flow curve for each column was measured by pumping buffer (50 mM potassium phosphate, 100 mM NaCl, pH 7.0) through the column at specified superficial velocities and measuring the pressure drop at steady state. Compression of the stationary phase, measured as the separation of the bed from the flow distributor was also monitored at each flow rate by recording time-lapse images of the top surface of the bed using a Logitech Quickcam Pro 4000 web camera.

6.2.4 Measurement of Binding Isotherms

Equilibrium adsorption isotherms for binding of CBM9-GFP to MT100 and each cross-linked MT100 preparation were performed at 22°C in low salt buffer (50 mM potassium phosphate, 100 mM NaCl, pH 7.0). Purified protein at concentrations ranging from 1 to 30 μM was mixed with media (1 mg dry weight) in low salt buffer to a final volume of 1 ml. Samples were then incubated overnight at 22°C while mixing end-over-end. The supernatant was collected after centrifugation at 27,000 $\times g$ for 20 min at 22°C and the concentration of unbound protein was determined by UV absorbance (280 nm) using a Cary 100 Spectrophotometer (Varian, Palo Alto, CA) and a calculated molar extinction coefficient of 62870 $\text{M}^{-1} \text{cm}^{-1}$ (Mach et al. 1992). The above protocol was repeated for binding of ovalbumin to identify any non-specific binding that may arise due to the cross-linking reaction. Langmuir-type binding parameters were determined by non-linear regression of the Langmuir isotherm equation to the experimental data using GraphPad Prism 3.0 software.

6.2.5 Affinity Purification of CBM9-GFP

An overnight 5 mL culture of *E. coli* BL21 (DE3) cells carrying the pET28-CBM9-GFP vector (Kavoosi et al. 2004) was used to inoculate 500 mL of Lustral Broth (LB). Cells were then grown at 37°C to an OD_{600 nm} of about 0.8 and protein production was induced by the addition of 0.1 mM isopropyl-1-thio-β-D-galactoside (IPTG). The culture was further incubated at 30°C for another 10-12 hours after which the cells were harvested by centrifugation for 20 min at 8,500 x g and 4°C. The cells were resuspended in high salt buffer (50 mM potassium phosphate, 1 M NaCl, pH 7.0) and ruptured by two passes through a French pressure cell (21,000 lb/in²). The cell debris was removed by centrifugation at 27,000 x g for 30 min and 4°C. The clarified cell extract was frontal loaded onto an XK26 column (GE Healthcare) packed with MT100 or a cross-linked MT100 preparation. Unbound proteins were washed through the column using 210 mL of high salt buffer followed by 150 mL of low salt buffer. Bound CBM9-GFP was isocratically eluted with 100 mL of 1 M glucose in low salt buffer; 10 mL fractions were collected throughout the process.

The chromatogram for the purification process was detected using in-line UV absorbance at 280 nm and off-line measurement of the fluorescence intensity (395 nm excitation of GFP, 509 nm emission) of each collected fraction using a Cary Eclipse fluorescence spectrophotometer (Varian). Process yields were determined by comparing the fluorescence intensity of appropriately diluted elution fractions to that of the original cell lysate. The purity of the combined product peak was analysed by 12% SDS-PAGE.

6.3 Results and Discussion

6.3.1 Response Surface Methodology to Improve Mechanical Properties

Both flow curve and bed-compression visualization data were used to evaluate the mechanical stability of unmodified MT100 and of each chemically modified cellulose stationary phase. As illustrated in Figure 6.1, significant compression of a packed bed of unmodified Perloza™ MT100 is observed at all superficial velocities above 0.64×10^{-4} m

s^{-1} , indicating the need for mechanical stabilization of this media prior to column scale-up to permit acceptable throughput without column compression.

The epichlorohydrin-based cross-linking chemistry proposed to stabilize the mechanical properties of MT100 against pneumatic compression is described in Figure 6.2. Cross-linking reactions involving epoxides such as epichlorohydrin are typically carried out under strong caustic conditions to deprotonate all solvent-exposed hydroxyl groups and increase their reactivity. Previous studies have shown that the reaction product is relatively insensitive to NaOH provided the concentration of NaOH is sufficient to bring the reaction pH above 12 (Guo and Ruckenstein 2001). A polar organic co-solvent such as methanol, or in this case DMSO, is generally used to swell the matrix and improve accessibility of reactive hydroxides to the cross-linking reagent, and the mechanical stability of the reaction product is therefore likely sensitive to the concentration of the co-solvent ([DMSO]) used. The concentration of epichlorohydrin ([Epi]) is also expected to be a determinant of the characteristics of the reaction product.

We investigated the dependence of the mechanical stability (response variable u_{crit} , detailed below) of epichlorohydrin cross-linked media on the MT100-normalized concentration of epichlorohydrin (input variable ξ_1 in mL Epi/g MT100) and DMSO (input variable ξ_2 in mL DMSO/g MT100) used in the formulation reaction. For each column prepared at a given ξ_1, ξ_2 condition, pressure drop was measured as a function of superficial velocity u to determine u_{crit} , the value of u at which the flow curve deviates from linearity, which was used as our response variable quantifying mechanical stability. In all cases, u_{crit} differed by less than 15% from the linear velocity at which incipient bed collapse was observed, suggesting that the onset of flow curve non-ideality is directly related to bed compression.

A general response model for this system can be written as:

$$u_{crit} \equiv y = f(\xi_1, \xi_2) + \Psi \quad (6.1)$$

where Ψ represents random variation (*e.g.*, measurement error or background noise) and is assumed to have a normal distribution with mean zero and variance σ^2 . To identify a

range of ξ_1 , ξ_2 values for which model (6.1) is applicable, we first determined a set of ξ_1 , ξ_2 pairs that yielded a well-defined reaction product and for which a value of u_{crit} could be measured reproducibly. The resulting values suggested an input variable space of $0, 0 \leq \xi_1, \xi_2 \leq 12, 12$ mL/g, which is consistent with the range of epichlorohydrin and DMSO concentrations previously used by Ruckenstein and Guo (Guo and Ruckenstein 2001) to optimize cross-linking of cellulose filters. As the true response function f was unknown, we first performed a nonparametric exploration of the observed response values. To obtain a flexible and smooth representation of the response surface, we used locally weighted polynomial regression, in which the fitted response surface for a target point results from a weighted, least squares fit of a polynomial to the data points in a neighborhood of the target (Cleveland and Devlin 1988). The resulting estimate of the response surface f , depicted in Figure 6.3, showed strong evidence for substantial curvature; specifically, a mound-shaped response surface was suggested by the data. This fit implies that the mechanical stability u_{crit} is maximized when [epichlorohydrin] and [DMSO] are 4.08 and 8.2 mL/g respectively, which is consistent with maximum observed in our experiment at input concentrations of $\xi_1 = 2.78$ mL/g and $\xi_2 = 8.33$ mL/g. The plots of the response surface also suggest that mechanical stability at or near this maximum is only achieved in a relatively small set of DMSO and epichlorohydrin concentration pairings.

For the purposes of conducting inference and interpreting model parameters, it was also desirable to pose a more structured, global model. Interestingly, the standard second-order response surface model can be seen as simply a global implementation of the polynomial fits employed above (Box et al. 2005):

$$u_{crit} = f(\xi_1, \xi_2) = \beta_0 + \beta_1 \xi_1 + \beta_2 \xi_2 + \beta_{11} \xi_1^2 + \beta_{22} \xi_2^2 + \beta_{12} \xi_1 \xi_2 + \Psi \quad (6.2)$$

where β_1 and β_{11} are the linear and quadratic regression coefficients for the main effect of [Epi] on u_{crit} , and β_{12} is the regression coefficient for linear interaction effect between ξ_1 and ξ_2 . The results of fitting model (6.2) are presented in Table 6.2 and Figure 6.4, and the interaction effect β_{12} is seen to be highly insignificant ($p = 0.8664 \gg 0.05$), whereas all other non-intercept terms are highly significant (maximum $p = 0.0134 < 0.05$). The

results of fitting a version of model (6.2) without the interaction term are also presented in Table 6.2 and we find that both the estimates and the strong statistical significance of the remaining non-intercept terms are essentially unchanged. The simplified model therefore permits a direct and unambiguous analysis of the dependence of u_{crit} on the input variables ξ_1 and ξ_2 , each of which was found to influence the mechanical stability. At low DMSO concentrations, the observed improvement in u_{crit} with increasing ξ_2 is consistent with the known solvating effects of DMSO. The parabolic dependence of u_{crit} on ξ_2 is then observed because DMSO concentrations greater than 8.33 mL/g result in significant swelling of the particles, weakening interchain hydrogen bonds between disordered and partially crystalline regions of the regenerated cellulose architecture of MT100 even in the presence of a large concentration of crosslinking agent. Similarly, excessive crosslinking of the media at high epichlorohydrin loads (*i.e.*, > 3 mL/g) results in significant swelling and mechanical breakdown of the media, presumably through the high osmotic pressures and new interchain forces created through the crosslinking reaction. Thus, the experimental and RSM results suggest that chemical crosslinking can improve the mechanical stability of MT100 only under conditions where the reaction does not disrupt the interchain hydrogen-bonding forces that naturally provide mechanical stability to the unmodified MT100 media.

Finally, a two-fold change in the volume of 5.0 M NaOH added was also applied at two randomly chosen reaction conditions ($\xi_1 = 0.56$ mL/g, $\xi_2 = 8.33$ mL/g; and $\xi_1 = 11.11$ mL/g, $\xi_2 = 11.11$ mL/g) to verify, as reported by Ruckenstein and Guo (Guo and Ruckenstein 2001), that the NaOH concentration does not influence the reaction product provided the concentration of NaOH is sufficient to bring the reaction pH above 12.

As quantified by u_{crit} , the mechanical stability of CRL100-7 ($u_{\text{crit}} = 6.0(\pm 0.15) \times 10^{-4}$ m s⁻¹) is *ca.* an order of magnitude greater than that of untreated Perloza™ MT100 ($u_{\text{crit}} = 0.64 \times 10^{-4}$ m s⁻¹). This is verified in Figure 6.5, which compares the measured flow curve for a packed bed of each media to show the significant extension of the desired linear response in pressure drop to increases in flow. As a result, mechanically permissible linear velocities within the CRL100-7 column extend deep onto the linear region of the van Deemter plot (as will be shown later in this manuscript), allowing

column operating conditions to be selected based on the traditional trade-off between throughput and mass-transfer related reductions in column efficiency. Thus, cross-linked media CRL100-7 was selected and further characterized as a potentially scalable matrix for affinity capture and purification of CBM9-tagged fusion proteins.

6.3.2 Comparison of Sorption Properties

Equilibrium adsorption studies were performed to determine the influence of cross-linking on the binding properties of the media. Adsorption isotherms for CBM9-GFP binding to untreated MT100 and to the cross-linked media are shown in Figure 6.6. Both data sets are well described by the Langmuir equation, which shows that the epichlorohydrin cross-linking reaction does not alter binding affinity, but reduces static capacity by *ca.* 25% (Table 6.3). The epichlorohydrin cross-linking chemistry is expected to target free hydroxyls (Figure 6.2), including the reducing ends of solvent-exposed cellulose chains within the media, leading to the observed decrease in the ligand density for CBM9 binding. Nevertheless, the static capacity, which equates to 370 mg CBM9-GFP bound per g cross-linked media, remains significantly higher than that reported for most commercially available media used for fusion-tag based affinity chromatography (Lichty et al. 2005). Finally, negligible binding of ovalbumin was observed (Figure 6.6), indicating that cross-linking of the media produced no chemical alterations that enhance non-specific binding of proteins.

6.3.3 Comparison of Column Parameters and Properties

Using methods described previously (Rodriguez et al. 2004), moments analysis (Kubin 1965; Kucera 1965) was combined with van Deemter theory (Van Deemter et al. 1956) to determine parameters characterizing the geometric properties of columns packed with CRL100-7 media and the mass-transfer behavior of CBM9-GFP within those columns. Experiments were conducted under nonbinding conditions with the column length to diameter (L/d_c) ratio maintained above 5 to minimize end effects (Jungbauer 1993). All wall effects were ignored since the ratio of d_c to particle diameter (d_p) was well above that needed to safely ignore column wall contributions to band broadening (Knox et al. 1976).

Measured parameters and properties are listed in Table 6.4 for both an unmodified MT100 column and our CRL100-7 cross-linked media column. A peak asymmetry factor (A_s) close to unity and a reduced plate height (h) close to 2 are both good indicators (Bristow and Knox 1977) that the cross-linked media packs uniformly. The first moment (μ_1) for CBM9-GFP elution (nonbinding conditions) from the CRL100-7 column shows a linear dependence on solute residence time (Figure 6.7) that, together with a regressed ε value lower than measured for the untreated MT100 column, indicates the more mechanically stable cross-linked media packs somewhat more efficiently than the compressible media.

A decrease in stationary-phase porosity ε_p with protein molecular weight is expected for media with pore diameters within the size range of the partitioning proteins. Stationary-phase porosities were therefore determined with respect to partitioning of CBM9-GFP, a 53 kDa fusion protein. Cross-linking decreases ε_p for this protein probe from its original value of 0.81 for untreated MT100 to 0.67, indicating a significant change in intraparticle pore architecture that is consistent with the 25% drop in static binding capacity reported in Table 6.3. Such changes in stationary-phase architecture might also alter solute mass transport such that gains in mechanical stability achieved through the cross-linking reaction occur at the expense of overall column performance. Second moment (σ^2) data for CBM9-GFP elution peaks were therefore determined as a function of interstitial velocity u_o and combined with the moments theory of Haynes and Sarma (Haynes and Sarma 1973) for operation under nonbinding conditions

$$\frac{u_o \sigma^2 L}{2\mu_1^2} = D_L + u_o^2 \left(\frac{\varepsilon}{1-\varepsilon} \right) \left[1 + \frac{\varepsilon}{(1-\varepsilon)\varepsilon_p} \right]^2 \left(\frac{r_p}{3k_f} + \frac{r_p^2}{15\varepsilon_p D_p} \right) \quad (6.3)$$

to obtain values for the parameters characterizing mass transfer of the CBM9-GFP fusion protein within a column packed with cross-linked MT100 media. The second moment data are plotted in Figure 6.8 according to equation 6.3, where D_L is the axial dispersion coefficient, and k_f and D_p are the film mass-transfer coefficient and pore diffusivity, respectively. Standard correlations for the bulk molecular diffusivity D_M (Young et al. 1980) and k_f (Wilson and Geankoplis 1966) therefore permit estimation of D_L and D_p

through regression of equation 6.3. The subsequent calculation of the column Peclet (Pe) and Biot (Bi) numbers then provides a sound basis for determining if the cross-linking reaction has negatively impacted solute mass transport within the column. In a previous study (Kavoosi et al. 2007b), we determined that uptake of CBM9-tagged proteins within a MT100 column is limited by the rate of solute diffusion within the pores of the stationary phase, with axial dispersion providing no significant contribution to elution band broadening. As shown in Table 6.4, cross-linking of the MT100 media changes this result very little. Axial dispersion remains insignificant, mass transfer across the fluid film is enhanced by the small reduction in column voidage, and the pore diffusivity decreases only slightly compared to the column utilizing untreated MT100. As a result, both Pe and Bi are largely unchanged by the cross-linking reaction, indicating that rates of solute mass transfer and overall mass transfer behavior are quite similar in the two columns.

Finally, as noted previously, the van Deemter plot (Figure 6.9) characterizing CBM9-GFP transport within the CRL100-7 column confirms that this cross-linked media can be reliably operated far into the flow regime where reduced plate height is controlled by solute mass transfer kinetics within the stationary phase.

6.3.4 Preparative Scale Affinity Purification of CBM9-GFP on CRL100-7.

Figure 6.10 shows the chromatogram for the purification of CBM9-GFP from clarified *E. coli* BL21 cell lysate on a 60 mL CRL100-7 column, which represents an 8-fold column scale-up by volume. No protease inhibitors were added to the cell extract and the purification was performed at room temperature (21°C). Although u in this separation is *ca.* three times larger than u_{crit} for untreated Perloza™ MT100, the purification on the CRL100-7 column proceeded in a similar manner to that on the MT100 column (Kavoosi et al. 2004) when both chromatograms are reported in terms of elution volume. Initial breakthrough of contaminating proteins was observed after *ca.* one column void volume. The small fluorescence peak observed within the breakthrough profile for contaminating proteins indicates the presence in the clarified cell lysate of free GFP, released as a result of a small amount of degradation within the linker region

connecting CBM9 to GFP (Kavoosi et al. 2007a). Two wash steps were used to effectively remove most of the contaminating proteins present in the column. As evident by the overlapping A_{280} and fluorescent signals, addition of 1 M glucose to the mobile phase elutes CBM9-GFP as a single sharp peak.

The product yield obtained in the pooled elution peak for the first cycle was $75 \pm 2\%$ (Table 6.5). Most of the lost product can be attributed to CBM9-GFP degradation products in the feed. Nevertheless, the yield from the initial run on the cross-linked column is lower than the 86% yield previously recorded for the same separation performed on an untreated MT100 column at a 3.6-fold lower linear velocity of $0.41 \times 10^{-4} \text{ m s}^{-1}$.

SDS-PAGE gel documentation of the scaled-up purification (Figure 6.11) shows quantitative flowthrough (Lane 2) of contaminating proteins with minimal presence of CBM9-GFP, resulting in elution of the 53 kDa CBM9-GFP product (Lane 5) characterized by a purification factor greater than 95%. The weak 26 kDa elution band in Lane 5 corresponds to CBM9 affinity tag released by proteolytic degradation of the fusion protein within the linker region. As the purification was performed at 21°C with no protease inhibitors, activity from endogeneous proteases was likely responsible for the observed product degradation. Higher product yields may therefore be realized through application of a suitable protease inhibitor.

6.3.5 Column Reusability

The ability of CRL100-7 to provide acceptable and predictable purification performance with repeated column use was also investigated. Three consecutive purifications were performed at a superficial velocity of $1.5 \times 10^{-4} \text{ m s}^{-1}$ to identify any changes in column performance with increasing number of purification cycles. In all three purification cycles, very high product purity (>95%) was observed. Interestingly, product yield increased after the first column use (Table 6.5) to a value comparable to that observed for the untreated MT100 column. This suggests that the cross-linking reaction introduces a small number of sites that exhibit irreversible binding to CBM9-

GFP. These sites are blocked in the first cycle, leading to the higher observed yields in the second and subsequent cycles.

6.4 Conclusions

The mechanical stability of porous Perloza™ MT100 cellulosic matrix was enhanced through an RSM optimized cross-linking reaction using epichlorohydrin to create a chromatographic media for affinity purification of recombinant CBM9-tagged fusion proteins. The chemistry used to cross-link MT100 is simple to conduct and inexpensive, yielding a stable affinity media that is *ca.* 1/50th – 1/100th the cost of popular affinity media currently used for industrial processing. The mechanically enhanced CRL100-7 media can withstand mobile-phase superficial velocities up to 6×10^{-4} m/s, representing an order of magnitude improvement in throughput over unmodified MT100. The cross-linking reaction alters pore architecture, reducing the porosity of the stationary phase by 20% and thereby lowering binding capacity by *ca.* 25%. Despite this loss in capacity, CRL100-7 is capable of binding 7.1 μmol CBM9-GFP/g media (376 mg/g), a value superior to most commercial affinity-tag systems. The mechanical stability of the cross-linked media is maintained during (at least) modest column scale-up. Consecutive purifications on a scaled-up column showed yields equal to or greater than 82%, demonstrating reliable column performance from cycle to cycle.

6.5 Tables

Table 6.1 Cross-linked resins and reaction conditions

Resin	Epichlorohydrin (ml/g cellulose)	DMSO (ml/g cellulose)	u_{crit} (m/s)
MT100	0	0	$0.64 (\pm 0.10) \times 10^{-4}$
CRL100-1	0.28	0	$1.3 (\pm 0.20) \times 10^{-4}$
CRL100-2	2.78	0	$1.1 (\pm 0.17) \times 10^{-4}$
CRL100-3	11.11	0	$1.1 (\pm 0.17) \times 10^{-4}$
CRL100-4	0	8.33	$0.64 (\pm 0.10) \times 10^{-4}$
CRL100-5	0.28	8.33	$1.2 (\pm 0.18) \times 10^{-4}$
CRL100-6	0.56	8.33	$1.5 (\pm 0.23) \times 10^{-4}$
CRL100-7	2.78	8.33	$6.0 (\pm 0.90) \times 10^{-4}$
CRL100-8	0	11.11	$0.74 (\pm 0.11) \times 10^{-4}$
CRL100-9	0.28	11.11	$1.1 (\pm 0.17) \times 10^{-4}$
CRL100-10*	11.11	11.11	$1.2 (\pm 0.18) \times 10^{-4}$

*The reaction for CRL100-10 had 2.8 ml/g of cellulose of 5 M NaOH. All other reactions had 1.4 ml/g of cellulose of 5 M NaOH.

Table 6.2 Results of fitting quadratic response surface models to the data set reported in Table 6.1

Regression Equation

$$u_{crit} = \beta_0 + \beta_1\xi_1 + \beta_2\xi_2 + \beta_{11}\xi_1^2 + \beta_{22}\xi_2^2 + \beta_{12}\xi_1\xi_2 + \Psi$$

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	0.548	0.425	1.289	0.205
β_1	0.710	0.163	4.353	9.79×10^{-5}
β_2	0.432	0.140	3.094	0.0037
β_{11}	-0.068	0.013	-5.085	1.02×10^{-5}
β_{22}	-0.033	0.013	-2.595	0.0134
β_{12}	0.002	0.010	0.169	0.8664

Residual standard error: 1.079 on 38 degrees of freedom

Multiple R-Squared: 0.5281

Adjusted R-squared: 0.466

F-statistic: 8.506 on 5 and 38 DF

p-value: 1.784×10^{-5}

Regression Equation

$$u_{crit} = \beta_0 + \beta_1\xi_1 + \beta_2\xi_2 + \beta_{11}\xi_1^2 + \beta_{22}\xi_2^2 + \Psi$$

Coefficient	Estimate	Standard Error	t-statistic	p-value
β_0	0.512	0.363	1.410	0.167
β_1	0.720	0.148	4.851	2.00×10^{-5}
β_2	0.436	0.135	3.223	0.0026
β_{11}	-0.068	0.013	-5.152	7.75×10^{-6}
β_{22}	-0.033	0.013	-2.622	0.0124

Residual standard error: 1.066 on 39 degrees of freedom

Multiple R-Squared: 0.5278

Adjusted R-squared: 0.4793

F-statistic: 10.9 on 4 and 39 DF

p-value: 4.997×10^{-6}

Table 6.3 Equilibrium adsorption parameters for MT100 and CRL100-7 affinity media determined from two independent experiments

Resin	Affinity (M ⁻¹)	Capacity (μmol/g resin)
MT100	0.71 (±0.05) x 10 ⁶	9.53 (±0.23)
CRL100-7	1.04 (±0.21) x 10 ⁶	7.10 (±0.33)

* reported errors represent ±2σ (i.e. 95% confidence interval)

Table 6.4 Column properties and parameters

Parameter	MT100	CRL100-7	Units
L	9.3	11.4	cm
h	2.3 (\pm 0.02)	2.6 (\pm 0.03)	
A_s	1.1 (\pm 0.03)	1.01 (\pm 0.09)	
ε	0.53 (\pm 0.07)	0.49 (\pm 0.01)	
ε_p	0.81 (\pm 0.02)	0.67 (\pm 0.01)	
D_L	2.9 (\pm 1.8) $\times 10^{-9}$	5.0 (\pm 1.4) $\times 10^{-8}$	m ² /s
k_f^*	7.2 (\pm 0.13) $\times 10^{-6}$	7.8 (\pm 0.23) $\times 10^{-6}$	m/s
D_M	6.54 $\times 10^{-11}$	6.54 $\times 10^{-11}$	m ² /s
D_p^*	4.5 (\pm 0.21) $\times 10^{-12}$	1.7 (\pm 0.07) $\times 10^{-12}$	m ² /s
Pe^*	613 (\pm 29)	1668 (\pm 69)	
Bi^*	17 (\pm 0.8)	51 (\pm 2.3)	

* Properties determined for superficial velocity of 4.25×10^{-5} m/s at 294 K; two independent columns were packed to determine the error in the packing procedure.

* reported errors represent $\pm 2\sigma$ (i.e. 95% confidence interval)

Table 6.5 Yield and purity for consecutive purification cycles of CBM9-GFP on a CRL100-7 column

Column cycle	Yield	Purity*
1	74.8 (± 2)%	>95%
2	88.7 (± 3)%	>95%
3	83.5 (± 3)%	>95%

*Determined by SDS-PAGE analysis

6.6 Figures

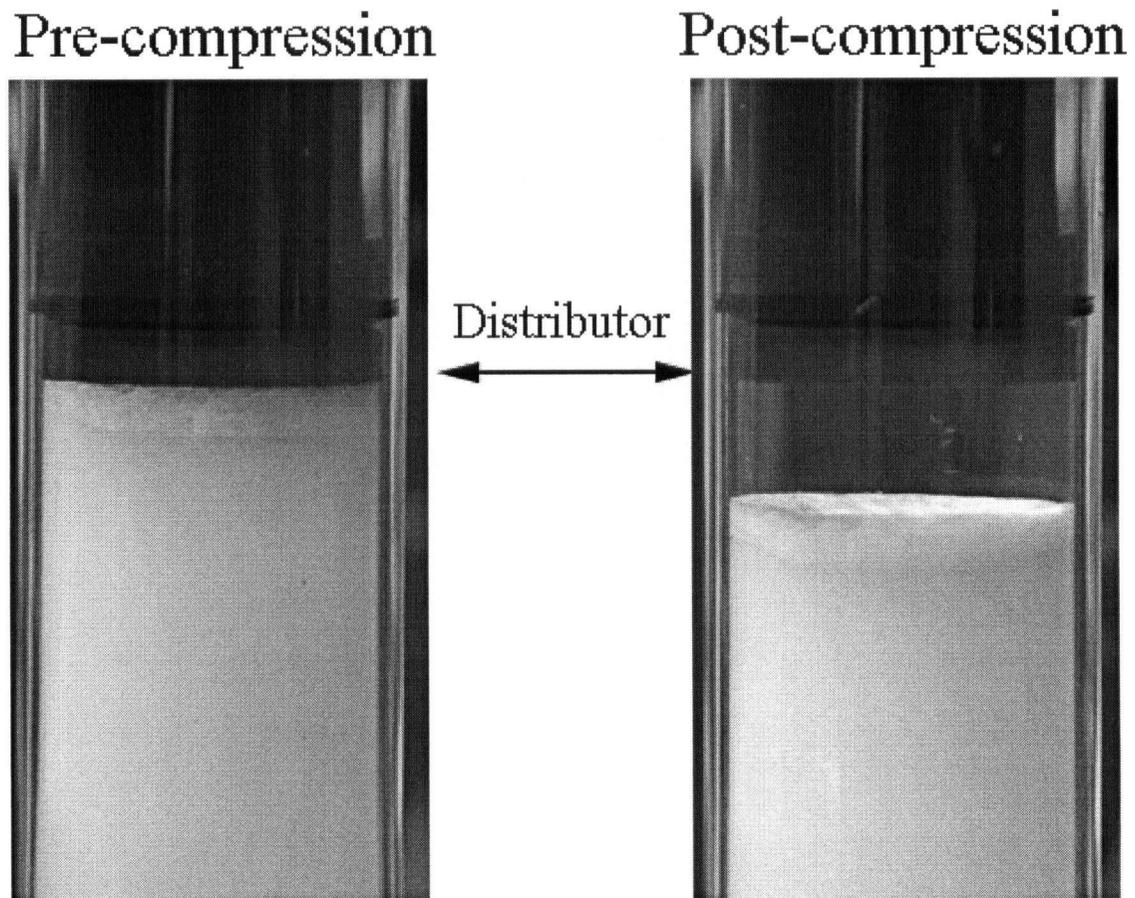


Figure 6.1 Picture of an MT100 column pre- and post-compression. Picture shows bed compression at a superficial velocity of 1.06×10^{-4} m/s.

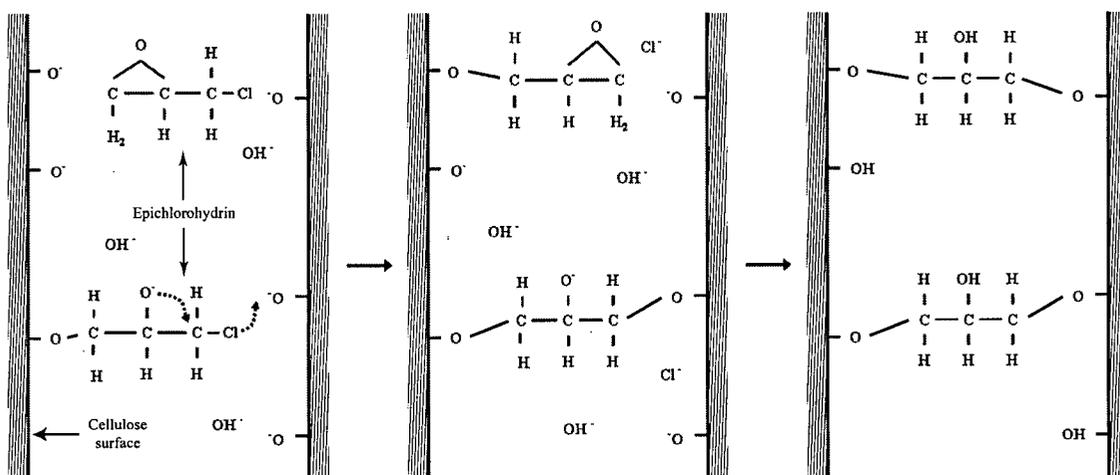


Figure 6.2 Schematic representation of the proposed epoxide-based reaction for cross-linking cellulose with epichlorohydrin.

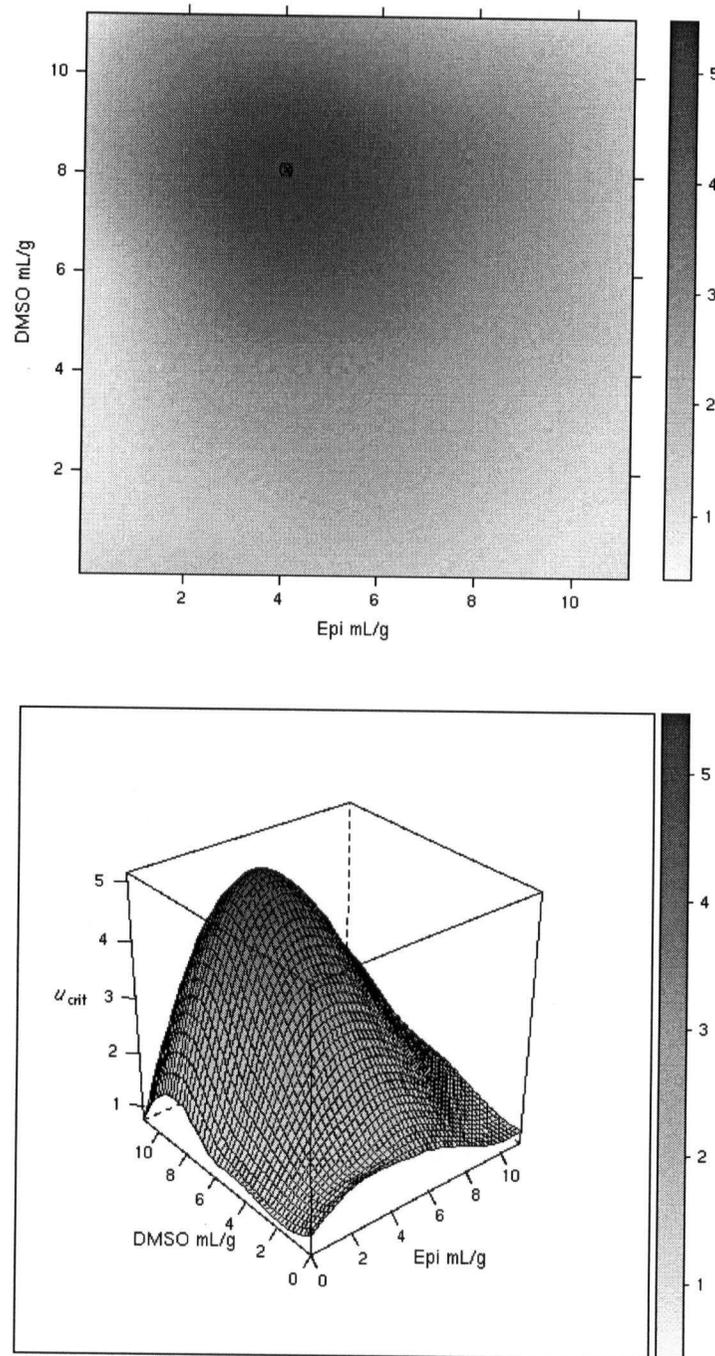


Figure 6.3 Imageplot (A) and wireframe surface (B) for a locally weighted polynomial regression fit to the measured u_{crit} (in $\text{m s}^{-1} \times 10^4$) data set reported in Table 1. Created with the loess, levelplot, and wireframe functions in R (Young et al. 1980) and the add-on package Lattice (Lattice Graphics R package, version 0.14-16), with a span of 0.65 and a degree of 2 (i.e. local fits are quadratic). The location of the maximum is indicated by (\otimes).

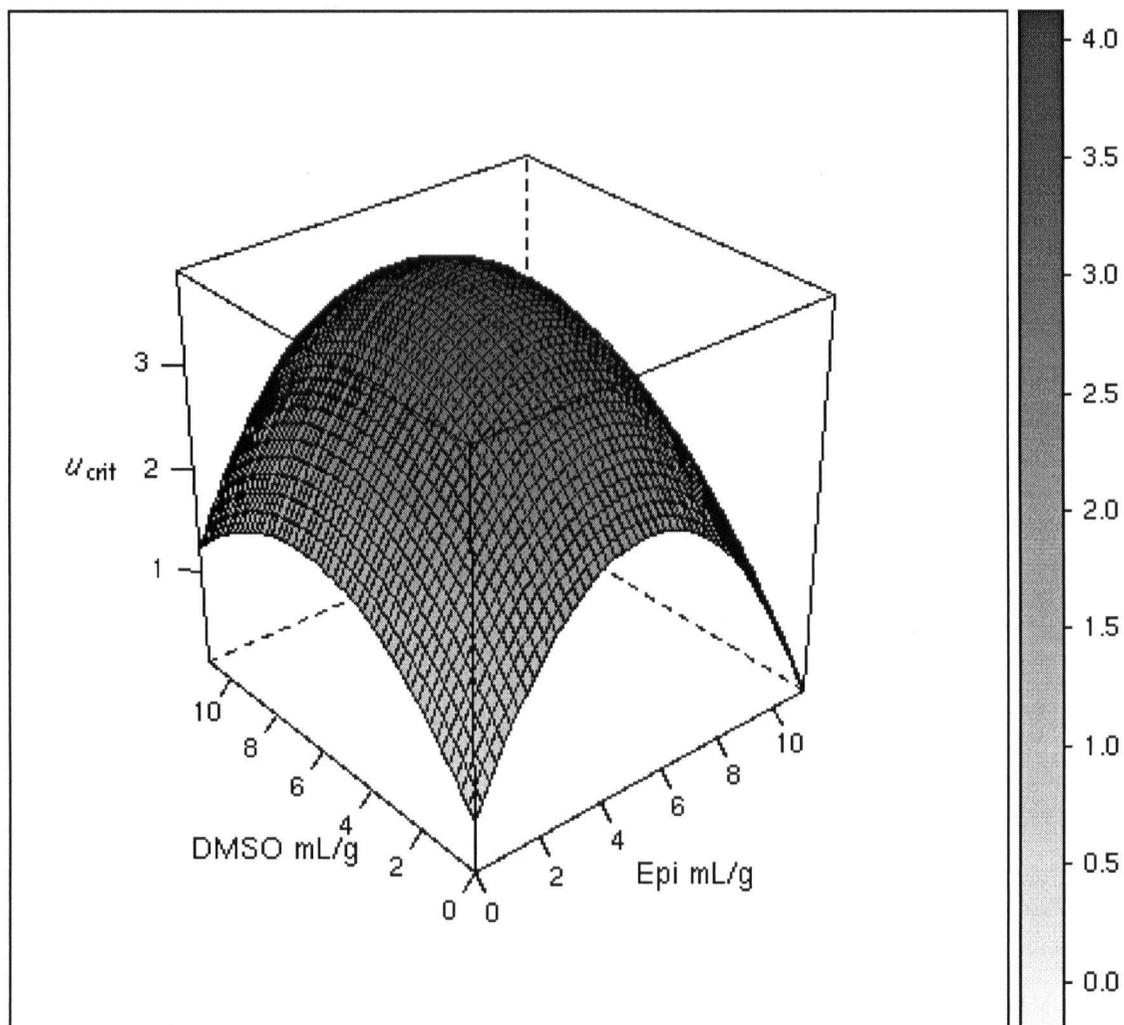


Figure 6.4 Wireframe surface of the results obtained from the standard quadratic response surface model (6.2) when applied to the measured u_{crit} (in $\text{m s}^{-1} \times 10^4$) data set reported in Table 1.

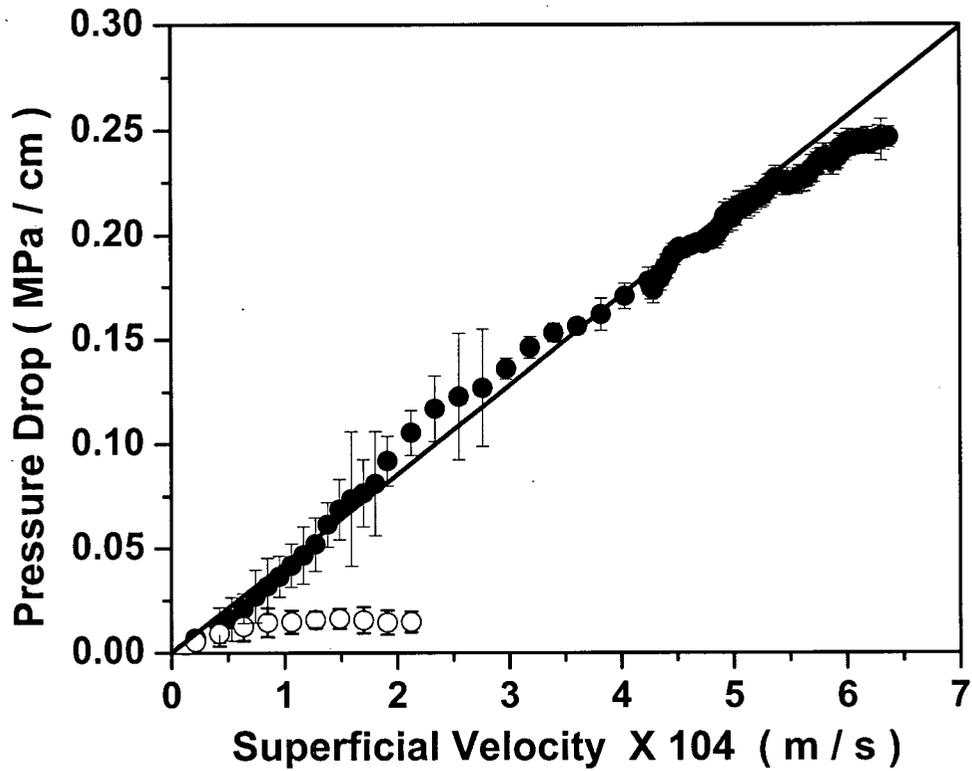


Figure 6.5 Pressure curve for solvent flow through an MT100 (open circles) or CRL100-7 (filled circles) column monitored over a range of superficial velocities. The pressure drop curve for CRL100-7 was compiled from two independent experiments.

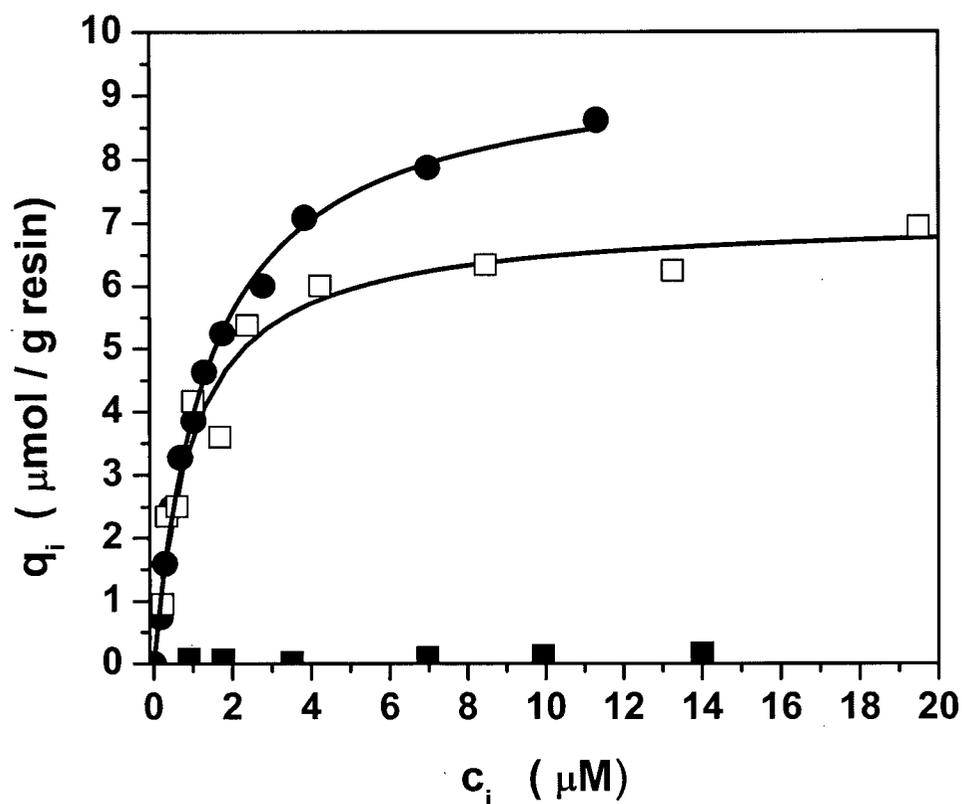


Figure 6.6 Equilibrium adsorption isotherms for binding of CBM9-GFP to MT100 (filled circles) and CRL100-7 (open squares) at 21°C. Control experiment shows ovalbumin binding to CRL100-7 (filled squares) at 21°C. The mobile phase buffer consisted of 50 mM phosphate buffer, 100 mM NaCl, pH 7. The curve represents the best fit of the experimental data to the Langmuir adsorption isotherm equation where q_i is the protein concentration bound to the media surface and c_i is the equilibrium concentration of protein free in solution.

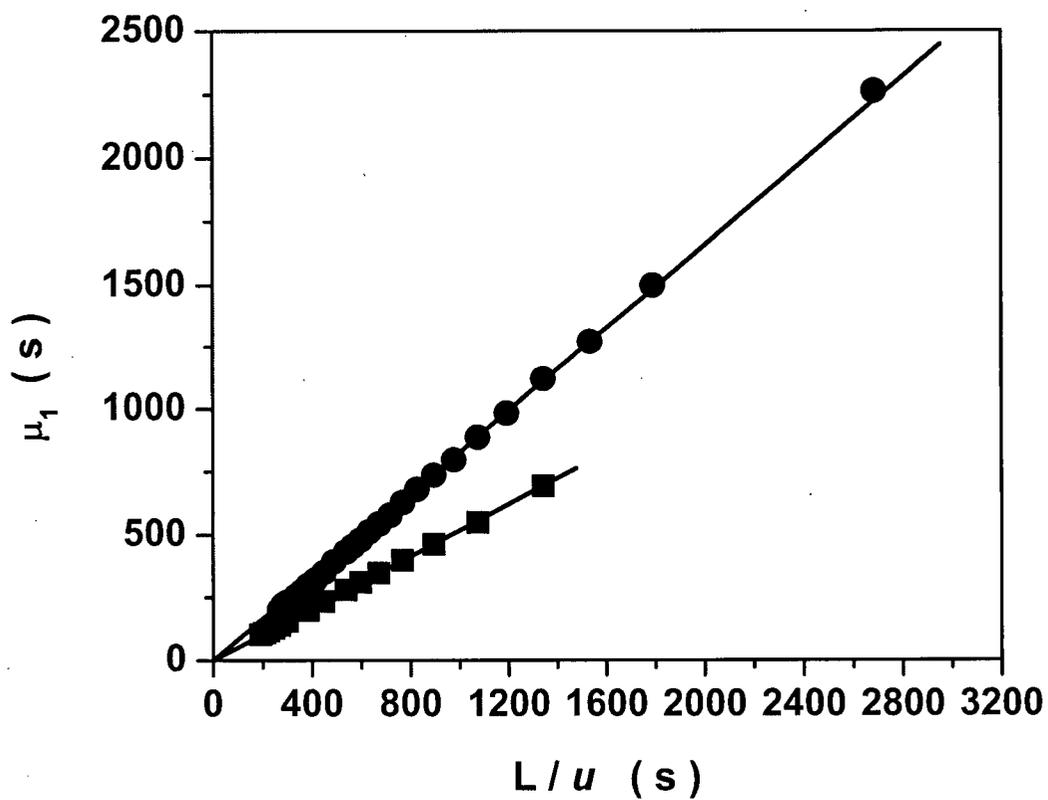


Figure 6.7 Measured first central moment (μ_1) for blue dextran (filled squares) and CBM9-GFP (filled circles) on a column (I.D. 1.0 cm) packed with CRL100-7.

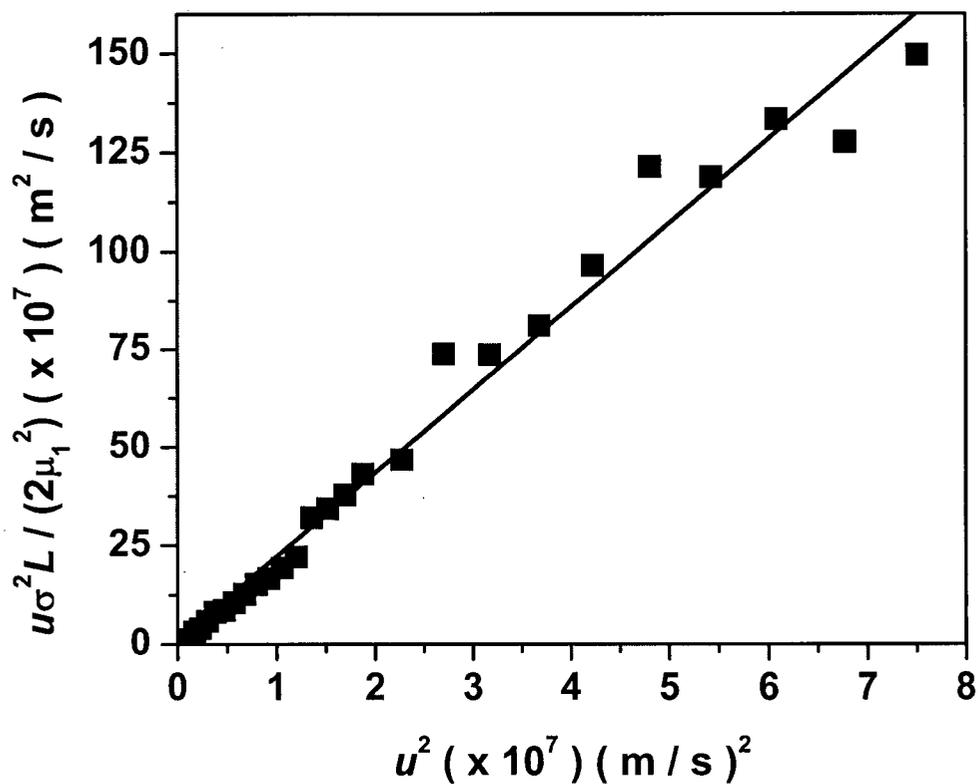


Figure 6.8 Second moment analysis under non-binding conditions for a column (I.D. 1.0 cm) packed with CRL100-7 media. Pulse injections of CBM9-GFP were used over the superficial velocity range 4.25×10^{-5} m/s to 4.25×10^{-4} m/s. The mobile phase consisted of 2 M glucose in 50 mM potassium phosphate, 100 mM NaCl (pH 7, 21°C).

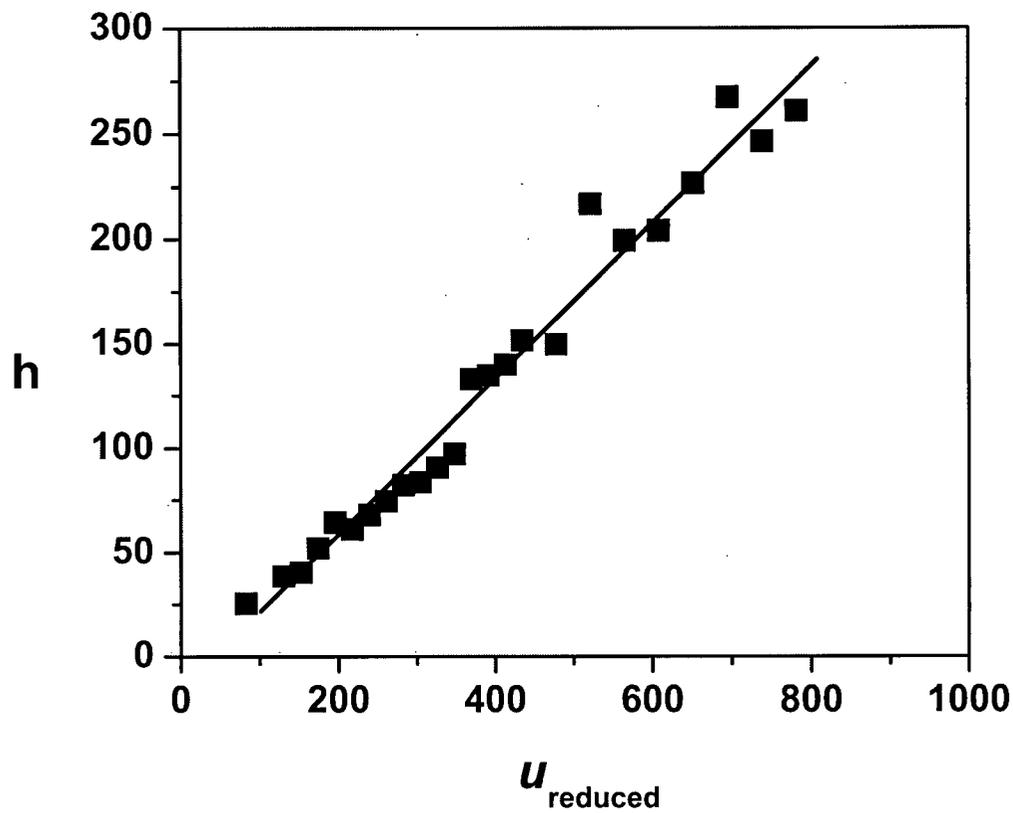


Figure 6.9 Van Deemter plot describing CBM9-GFP transport within the CRL100-7 column (I.D. 1.0 cm X 11.4 cm); the reduced plate height $h = (\sigma^2 L) / (\mu_1^2 d_p)$ is plotted as a function of the reduced linear velocity, given by $(u d_p) / D_M$.

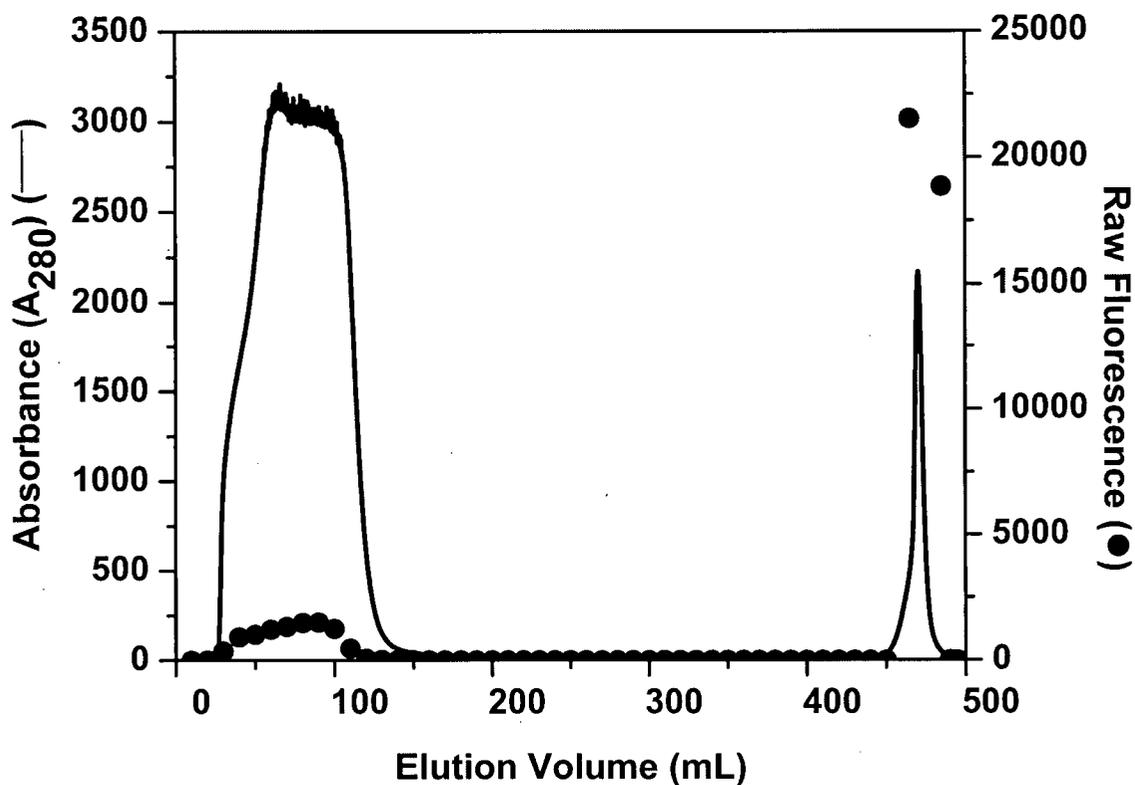


Figure 6.10 Chromatogram for purification of CBM9-GFP on a 60-mL CRL100-7 column at 21°C. Clarified cell extract from *E. coli* BL21 were loaded onto a CRL100-7 column (2.6 cm I.D. X 11.8 cm) at a superficial velocity of 1.5×10^{-2} m/s, washed with *ca* 3 column volumes (CV) of high salt buffer, *ca* 2.5 CV of low salt buffer and bound CBM9-GFP was eluted with 1M glucose in low salt buffer. 10 ml fractions were collected and analyzed by both absorbance at 280 nm (solid line) and fluorescence emission at 509 nm (filled circles).

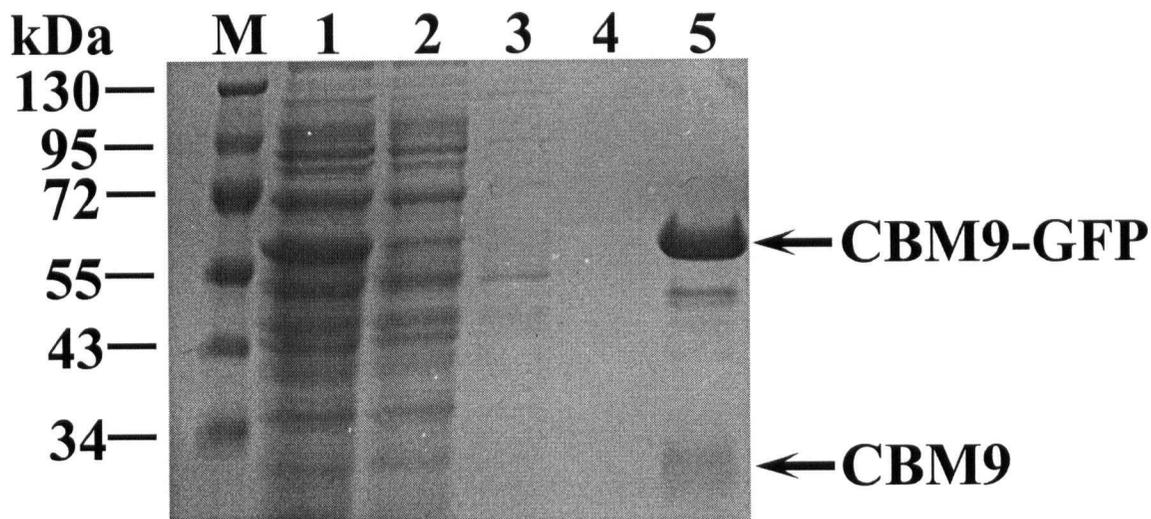


Figure 6.11 12% SDS-PAGE documentation for preparative-scale affinity purification of CBM9-GFP on a CRL100-7 column (I.D. 2.6 cm). All samples were dissolved in sample buffer containing 10% SDS. Lane M: molecular weight markers in kg/mole; Lane 1: clarified cell extract prior to column loading; Lane 2: column flow through; Lane 3: high salt wash; Lane 4: low salt wash; Lane 5: pure CBM9-GFP eluted in low salt buffer containing 1 M glucose.

6.7 References

- Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A.; Kilburn, D. G. (2001). "Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A". *Biochemistry* 40(21):6240-6247.
- Box, G. E. P.; Hunter, J. S.; Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd Edition. New Jersey: John Wiley & Sons, Inc.
- Bristow, P. A.; Knox, J. H. (1977). "Standardization of Test Conditions for High-Performance Liquid-Chromatography Columns". *Chromatographia* 10(6):279-289.
- Cleveland, W. S.; Devlin, S. J. (1988). "Locally Weighted Regression - An Approach to Regression-Analysis by Local Fitting". *Journal of the American Statistical Association* 83(403):596-610.
- Colby, C. B.; O'Neill, B. K.; Middelberg, A. P. J. (1996a). "A modified version of the volume-averaged continuum theory to predict pressure drop across compressible packed beds of sepharose big-beads sp (vol 12, pg 92, 1996)". *Biotechnology Progress* 12(5):728-728.
- Colby, C. B.; O'Neill, B. K.; Vaughan, F.; Middelberg, A. P. J. (1996b). "Simulation of compression effects during scaleup of a commercial ion-exchange process". *Biotechnology Progress* 12(5):662-681.
- Einhauer, A.; Jungbauer, A. (2001). "The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins". *Journal of Biochemical and Biophysical Methods* 49(1-3):455-465.
- Ford, C. F.; Suominen, I.; Glatz, C. E. (1991). "Fusion tails for the recovery and purification of recombinant proteins". *Protein Expression and Purification* 2(2-3):95-107.
- Guan, K. L.; Dixon, J. E. (1991). "Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase". *Analytical Biochemistry* 192(2):262-267.
- Guo, W.; Ruckenstein, E. (2001). "A new matrix for membrane affinity chromatography and its application to the purification of concanavalin A". *Journal of Membrane Science* 182(1-2):227-234.

- Haynes, H. W.;Sarma, P. N. (1973). "A model for the application of gas chromatography to measurements of diffusion in bidisperse catalysis". *AIChE Journal* 19:1043-1046.
- Hearn, M. T.;Acosta, D. (2001). "Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins". *Journal of Molecular Recognition* 14(6):323-369.
- Hjerten, S.; Wu, B.;Liao, J. (1987). "An high-performance liquid chromatographic matrix based on agarose cross-linked with divinyl sulfone". *Journal of Chromatography* 396:101-113.
- Hoenich, N. A.; Woffindin, C.; Cox, P. J.; Goldfinch, M.;Roberts, S. J. (1997). "Clinical characterization of Dicea a new cellulose membrane for haemodialysis". *Clinical Nephrology* 48(4):253-259.
- Holmberg, L.; Lindberg, B.;Lindqvist, B. (1995). "The Reaction between Epichlorohydrin and Polysaccharides - Structural Elements in a Cross-Linked Dextran, Sephadex G-25". *Carbohydrate Research* 272(2):203-211.
- Hopp, T., Prickett, KS, Price, VL, Libby, RT, March, CJ, Ceretti, DP, Urdal, DL, Conlon, PJ. (1988). "A short polypeptide marker sequence useful for recombinant protein identification and purification". *Bio/Technology* 6:1204-1210.
- Jungbauer, A. (1993). "Preparative Chromatography of Biomolecules". *Journal of Chromatography* 639(1):3-16.
- Kavoosi, M.; Creagh, A. L.; Kilburn, D. G.;Haynes, C. A. (2007a). "Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *E. coli*". *Biotechnology and Bioengineering* In Press.
- Kavoosi, M.; Meijer, J.; Kwan, E.; Creagh, A. L.; Kilburn, D. G.;Haynes, C. A. (2004). "Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T-maritima*". *Journal of Chromatography B* 807(1):87-94.
- Kavoosi, M.; Sanaiea, N.; Dismerb, F.; Hubbuch, J.; Kilburn, D. G.;Haynes, C. A. (2007b). "A Novel Two-Zone Protein Uptake Model for Affinity Chromatography and Its Application to the Description of Elution Band Profiles of Proteins Fused to the CBM9 Affinity Tag". *Journal of Chromatography A* Accepted.
- Knox, J. H.; Laird, G. R.;Raven, P. A. (1976). "Interaction of Radial and Axial-Dispersion in Liquid-Chromatography in Relation to Infinite Diameter Effect". *Journal of Chromatography* 122(Jul7):129-145.

- Kubin, M. (1965). "Theory of the chromatography. II. Effect of the external diffusion and of the adsorption in the sorbent particle." *Collection of Czechoslovak Chemical Communications* 30(9):2900-2907.
- Kucera, E. (1965). "Contribution to the theory of chromatography. Linear non-equilibrium elution chromatography". *Journal of Chromatography* 19:237-248.
- Kuga, S. (1984). "Porous cellulose materials for liquid chromatography". *Kami Pa Gikyoshi* 38(2):166-173.
- Laas, T. (1976). "Improved agarose matrixes for biospecific affinity chromatography". *Protides of the Biological Fluids* 23:495-503.
- Leonard, M. (1997). "New packing materials for protein chromatography". *Journal of Chromatography B* 699(1-2):3-27.
- Lichty, J. J.; Malecki, J. L.; Agnew, H. D.; Michelson-Horowitz, D. J.; Tan, S. (2005). "Comparison of affinity tags for protein purification". *Protein Expression and Purification* 41(1):98-105.
- Linhult, M.; Guelich, S.; Hober, S. (2005). "Affinity ligands for industrial protein purification". *Protein and Peptide Letters* 12(4):305-310.
- Lowe, C. R.; Lowe, A. R.; Gupta, G. (2001). "New developments in affinity chromatography with potential application in the production of biopharmaceuticals". *Journal of Biochemical and Biophysical Methods* 49(1-3):561-574.
- Mach, H.; Middaugh, C. R.; Lewis, R. V. (1992). "Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins". *Analytical Biochemistry* 200(1):74-80.
- Mohammad, A. W.; Stevenson, D. G.; Wankat, P. C. (1992). "Pressure-Drop Correlations and Scale-up of Size Exclusion Chromatography with Compressible Packings". *Industrial and Engineering Chemistry Research* 31(2):549-561.
- Motozato, Y.; Hirayama, C. (1984). "Preparation and Properties of Cellulose Spherical-Particles and Their Ion-Exchangers". *Journal of Chromatography* 298(3):499-507.
- Narayanan, S. R.; Crane, L. J. (1990). "Affinity-Chromatography Supports - a Look at Performance Requirements". *Trends in Biotechnology* 8(1):12-16.
- Nilsson, J.; Stahl, S.; Lundeberg, J.; Uhlen, M.; Nygren, P.-A. (1997). "Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins". *Protein Expression and Purification* 11(1):1-16.

- Ostergren, K. C. E.; Tragardh, A. C.; Enstad, G. G.; Mosby, J. (1998). "Deformation of a chromatographic bed during steady-state liquid flow". *AICHE Journal* 44(1):2-12.
- Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation". *Nature* 258(5536):598-599.
- Rodriguez, B.; Kavooosi, M.; Koska, J.; Creagh, A. L.; Kilburn, D. G.; Haynes, C. A. (2004). "Inexpensive and Generic Affinity Purification of Recombinant Proteins Using a Family 2a CBM Fusion Tag". *Biotechnology Progress* 20(5):1479-1489.
- Rulisek, L.; Havlas, Z. (2000). "Theoretical studies of metal ion selectivity. 1. DFT calculations of interaction energies of amino acid side chains with selected transition metal ions (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+} , Cd^{2+} , and Hg^{2+})". *Journal of the American Chemical Society* 122(42):10428-10439.
- Smith, D. B.; Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase". *Gene* 67(1):31-40.
- Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology* 60(5):523-533.
- Van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. (1956). "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography". *Chemical Engineering Science* 5:271-289.
- Varela, M. P.; Kimmel, P. L.; Phillips, T. M.; Mishkin, G. J.; Lew, S. Q.; Bosch, J. P. (2001). "Biocompatibility of hemodialysis membranes: Interrelations between plasma complement and cytokine levels". *Blood Purification* 19(4):370-379.
- Wilson, E. J.; Geankoplis, C. J. (1966). "Liquid mass transfer at very low Reynolds numbers in packed beds." *Industrial and Engineering Chemistry Fundamentals* 5(1):9-14.
- Young, M. E.; Carroad, P. A.; Bell, R. L. (1980). "Estimation of Diffusion-Coefficients of Proteins". *Biotechnology and Bioengineering* 22(5):947-955.

7 Conclusions and Recommendations

Recombinant proteins are the fastest growing sector of the pharmaceuticals industry, providing effective therapy for a number of life-threatening diseases. They are, however, difficult and expensive to produce. Growing this important industry in a manner that meets the public's increasing demand for lower healthcare costs therefore requires new more efficient technologies for producing recombinant protein therapeutics. In this thesis, I have addressed this need by introducing an inexpensive affinity-tag technology for both research and preparative-scale purification of recombinant proteins. The technology centers on fusing a novel affinity tag, the family 9 carbohydrate-binding module (CBM9) of xylanase 10A from *T. maritima*, to either the N- or C- terminus of a target therapeutic. CBM9 has a strong and specific affinity for both soluble sugars and insoluble cellulose, allowing for effective capture of CBM9-tagged fusion proteins on an inexpensive cellulose-based chromatographic media and subsequent selective elution using 1-M glucose. The technology was successfully used to purify CBM9-tagged green fluorescent protein (CBM9-GFP) from a clarified bacterial cell extract with yields of 83 (± 3)% following fusion-tag removal. GFP was selected as the first target protein to exploit its endogenous fluorescence, which provided a reliable means to quantitatively monitor each step of the purification. Subsequently, the CBM9 fusion tag technology has been successfully used to produce and purify a wide range of recombinant proteins in a cost-competitive manner.

Along with verifying the utility of this cost-effective purification technology, the use of GFP as target permitted identification of a number of areas where improvement of the prototype technology could be realized. In particular, losses in fusion protein due to nonspecific cleavage within the short peptide connecting the tag and target were observed, indicating the need to identify linker sequences that provide improved proteolytic resistance. I therefore established two bioinformatic-based strategies for designing CBM9-tagged fusion proteins with enhanced stability within the solvent-exposed linker region connecting the two protein domains. Both strategies, one based on identifying naturally occurring linkers within the proteome of the host organism, and the second based on screening peptidases and their known specificities using the

bioinformatics software MEROPSTTM, were successfully used to design a linker with improved resistant to endogeneous proteases of the host when compared against the traditional poly-glycine linker. Although widely used, the glycine-rich linkers were found by tandem MS data to be susceptible to hydrolysis by *E. coli* peptidases. The natural (PT)_xP and MEROPSTTM-designed S₃N₁₀ linkers were significantly more stable, indicating both strategies provide a useful approach to linker design. As both bioinformatics-based approaches are easily implemented as an *in silico* high-throughput screen, they can potentially be used to optimize linker design for any given fusion protein, an area that deserves further study given the paucity of available strategies for linker design.

Given the low cost of the bacterial production system and the cellulose-based affinity chromatography column, the overall process economics of fusion protein production and purification using the CBM9 affinity tag is largely dictated by the cost of the enzyme-catalyzed cleavage step required to remove the tag and recover the target protein. Use of GFP as the target protein provided a convenient way to show that the choice of the processing enzyme and the local structure of the cleavage site (including the adjacent linker sequence) affect the rate of tag bioprocessing, which in turn, influences the yield and economics of the bioprocess. I therefore introduced a simple spectroscopic method to screen a candidate library of processing-enzyme/linker-sequence combinations to identify a pair that enhances the rate of processing of our target CBM9-tagged fusion protein while providing resistance to nonspecific cleavage by endogenous peptidases. The assay, based on the Luminescence Resonance Energy Transfer (LRET), monitors nonradiative energy transfer from a lanthanide-based donor specifically bound to CBM9 and an acceptor fluorophore presented on the target protein. Enzyme catalyzed hydrolysis of the fusion protein terminates the resonance energy transfer, leading to a decrease in the sensitized fluorescence intensity of the acceptor that can be continuously monitored to quantify substrate concentration over time. The LRET-based assay is simple, fast and provides accurate k_{cat}/K_M values with standard errors of less than 3%, allowing measurement of both substantial and subtle differences in bioprocessing kinetics. Application of this technology to design the CBM9-tagged fusion proteins may

therefore speed product transfer to the manufacturing scale and decrease production costs.

The high cost of affinity resins is another factor currently limiting the industrial application of affinity chromatography. Although not manufactured for this purpose, the extraordinarily inexpensive cellulose-based chromatography media Perloza™ MT100 selectively binds CBM9-tagged fusion proteins with remarkably high capacity (*e.g.* 583 mg/g). However, this economic advantage over other commercially available affinity matrixes cannot be realized at the preparative scale because Perloza™ MT100 is prone to bed compression at elevated flow rates. Epoxide-based cross-linking chemistry was therefore used to mechanically stabilize Perloza™ MT100 at higher linear velocities to improve throughput and permit column scale up. A fixed-effect two-way response surface methodology was used to identify reactant concentrations resulting in an order-of-magnitude improvement in the mechanical stabilization of the media without significantly diminishing column capacity. The purification performance and stability of the cross-linked cellulose column were shown to not be diminished by modest scale-up, suggesting that this modified resin may provide a practical and economical purification system for preparative-scale purification of CBM9-tagged recombinant therapeutics. However, further studies are certainly required to establish a comprehensive understanding of process scale-up.

The use of CBM9 fusion tag technology at the industrial-scale would also benefit from an accurate mathematical model to predict fusion protein mass transfer and elution under a variety of loading conditions and column geometries. The traditional pore-diffusion model (PDM) forms the basis of the most popular models describing protein uptake within porous affinity chromatography columns. The predictive power of this model can be compromised by a boundary condition that can result in unrealistic concentration profiles within the stationary phase. I have relaxed this unnecessary approximation by introducing a generalized two-zone model (TZM) for adsorptive chromatography capable of predicting column performance under both linear and non-linear loading conditions and over a range of superficial velocities. The TZM divides the porous sorbent particle into two zones: an inner protein-free core and an outer zone where

the protein concentration gradient, described by the Langmuir theory, dictates solute loading and adsorption. The novel TZM is able to accurately predict fusion protein loading and breakthrough in the presence of clarified bacterial cell extract, thereby providing an important tool for technology scale-up and simulation.

In this thesis, I have therefore developed a new cost-effective affinity-tag technology for production and purification of recombinant therapeutics, identifying and effectively addressing a number of issues that could impact the use of the technology at larger scales. Although my work has largely focused on issues related to recombinant protein manufacturing, the fusion-tag technology I describe is highly flexible, allowing its potential use in a number of important areas of biotechnology, including the growing field of proteomics. The CBM9 affinity system can be used in a small-scale, high-throughput format to analyze protein expression and purification protocols under different environmental conditions. The technology can therefore be applied to parallel production of proteins within a proteome of interest or to rapid identification of promising strategies for producing a protein of interest at high yield and purity. The complementary technologies developed to assist with the technology transfer to preparative scales can also be applied in a small-scale, high throughput format to other areas of research. For example, the LRET-based technology can be applied to any system for parallel high-throughput monitoring of the rate of hydrolysis. Incorporating a lanthanide binding tag (Sculimbrene and Imperiali 2006) and the K coil of the *de novo*-designed coiled-coil heterodimerization tag technology (Chao et al. 1998; Tripet et al. 1996) at opposing ends of the fusion protein of interest, allows for the selective labeling of the fusion protein with both a lanthanide donor and an appropriate acceptor fluorophore; in the latter case, through binding of a synthesized E coil to which an appropriate fluorophore has been covalently attached (E and K coils will spontaneously associate into a full helical, stable ($K_a \sim 1 \times 10^9 \text{ M}^{-1}$) coiled-coil structure, Chao et al. 1998). These promising applications were not addressed in this thesis, but could form the basis of an important and interesting follow-up research project.

7.1 References

- Chao, H. M.; Bautista, D. L.; Litowski, J.; Irvin, R. T.;Hodges, R. S. (1998). "Use of a heterodimeric coiled-coil system for biosensor application and affinity purification". *Journal of Chromatography B* 715(1):307-329.
- Sculimbrene, B. R.;Imperiali, B. (2006). "Lanthanide-binding tags as luminescent probes for studying protein interactions". *Journal of the American Chemical Society* 128(22):7346-7352.
- Tripet, B.; Yu, L.; Bautista, D. L.; Wong, W. Y.; Irvin, R., T.;Hodges, R. S. (1996). "Engineering a de novo-designed coiled-coil heterodimerization domain off the rapid detection, purification and characterization of recombinantly expressed peptides and proteins". *Protein Engineering* 9(11):1029-1042.