Chapter 2

CLONING AND PURIFICATION OF ARTIFICIAL EXTRACELLULAR MATRIX PROTEINS

2.1 Abstract

This work describes the cloning, expression, and purification of the current generation of artificial extracellular matrix (aECM) proteins. These proteins were developed for application in small-diameter vascular grafts and contain domains derived from fibronectin and elastin. aECM-RGD and aECM-PHSRN contain sequences derived from the tenth and ninth type III domains of fibronectin, respectively. Sequence-scrambled aECM-RDG and aECM-NHRPS serve as negative controls. aECM-CS1' was designed to serve as a negative control for a previously developed CS1 protein.[1] The proteins also include a T7 tag, heptahistidine tag, enterokinase cleavage site, and lysine residues that serve as crosslinking sites.

2.2 Introduction

A family of artificial extracellular matrix (aECM) proteins has been developed by this laboratory for application in small-diameter grafts. The first generation was developed by Alyssa Panitch and consisted of proteins that varied the ratio of CS5 cell-binding domains to elastin-like repeats (VPGIG).[2,3] Eric Welsh modified these proteins by introducing a T7 tag for antibody identification and terminal lysine residues to serve as crosslinking sites.[4] Ralf Weberskirch designed the pET28aRW vector by incorporating a polylinker with a T7 tag, heptahistidine tag, and an enterokinase cleavage site. Kathleen Di Zio used the pET28aRW vector to create proteins that contained the CS5 and CS1 cell-binding domains and a scrambled version of the CS5 sequence (SC5).[1,5] She also introduced lysine residues within the elastin-like repeat $(VPGIG)_2VPGKG(VPGIG)_2$ to generate more crosslinking sites.

The current generation of aECM proteins is similar to proteins developed by Di Zio but contains different cell-binding domains (Figure 2.1). The RGD sequence is found in many different extracellular matrix proteins and has been shown in multiple studies to increase attachment and spreading of cells.[6] The $\alpha_v\beta_3$ integrin, the vitronectin receptor, can bind to many different RGD proteins and short peptides, whereas the $\alpha_5\beta_1$ integrin, the fibronectin receptor, binds to RGD in conjunction with a synergy site, PHSRN.[7,8] In competitive peptide inhibition studies with cells cultured on fibronectin, the peptide VPHSRN added to a solution of GRGDS did not exhibit a synergistic effect.[9] When the PHSRN peptide was covalently linked to a mutant protein at a position corresponding to where it is normally found in relation to the RGD sequence, a synergistic effect was observed. These results suggest that the spatial orientation of the

PHSRN sequence in relation to the RGD sequence is critical in $\alpha_5\beta_1$-mediated attachment. Experiments that studied the effect of adding short flexible linkers between the tenth and ninth type III fibronectin domains (where the RGD and PHSRN sequences are found, respectively) provide further evidence that the spatial orientation of the two sites is important. When linkers of two or six amino acids were introduced between the tenth and ninth type III domains, the percentage of well-spread cells and phosphorylation of focal adhesion kinase (FAK) both were reduced.[10] The crystal structure of fibronectin reveals that the RGD and PHSRN sequences are on the same face of the protein.[11] It thus is highly probable that the $\alpha_5\beta_1$ integrin binds both sequences at once.

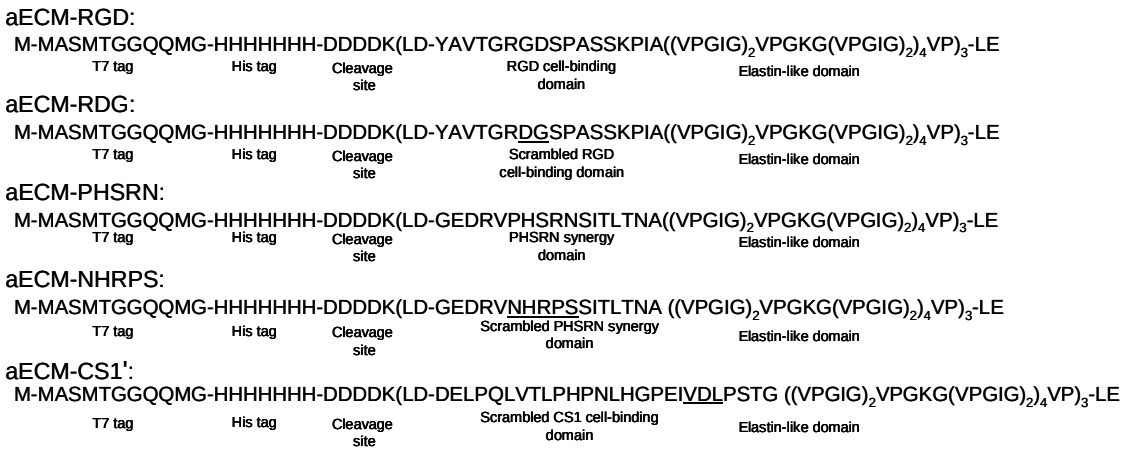

Figure 2.1 Artificial extracellular matrix protein sequences. The bioactive domains include an RGD cell-binding domain, a scrambled RDG sequence, a PHSRN synergy site, a scrambled NHRPS sequence, and a scrambled CS1' sequence. The proteins also include a T7 tag, heptahistidine tag, enterokinase cleavage site, and elastin-like repeats containing lysine residues that serve as crosslinking sites.

One of the objectives of this work was to develop artificial proteins containing the RGD cell-binding domain (aECM-RGD) and the PHSRN synergy site (aECM-PHSRN) and appropriate negative control proteins containing scrambled sequences (aECM-RDG and aECM-NHRPS) (Figure 2.1). The bioactive domains were designed to have the same number of amino acids so that each protein would be approximately the same molecular weight. Although the RGD and PHSRN sequences are both found within one fibronectin molecule, these aECM proteins were constructed so that proteins with different bioactive domains could be mixed together easily in different ratios. Endothelial response to aECM-RGD and aECM-RDG has been characterized extensively (Chapters 3 & 5). Unpublished data demonstrated that mixing aECM-PHSRN with aECM-RGD did not result in any synergistic effects on cell adhesion or spreading. aECM-PHSRN and aECM-NHRPS were therefore not as thoroughly characterized as aECM-RGD and aECM-RDG.

One additional protein was developed (aECM-CS1')—a negative control for the CS1 protein developed by Di Zio. Although this protein was cloned and purified on a small-scale, extensive characterization was not performed because preliminary, unpublished data indicated that cells did not bind to the CS1 protein in a sequence-specific manner. The DNA cloning scheme for the five proteins described in this study is shown in Figure 2.2.

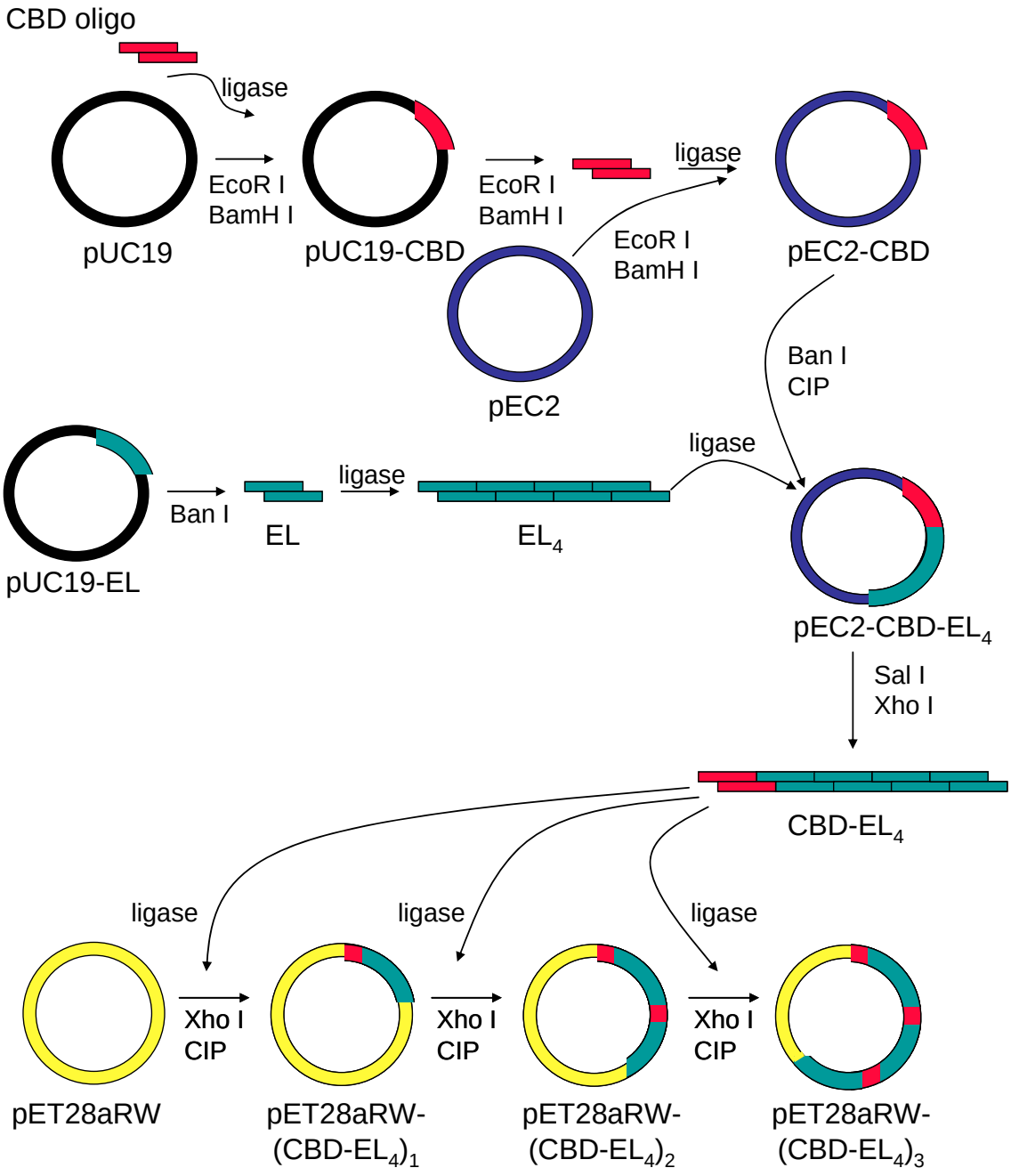

Figure 2.2 DNA cloning scheme showing the development of artificial extracellular matrix proteins. Figure was provided by Marissa Mock and modified.

2.3 Materials and Methods

2.3.1 Genetic Manipulation

Standard methods for cloning, bacterial growth, protein expression, sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), and Western blotting were used.[12,13]

2.3.2 Cloning of pEC2-CBD

Cell-binding domain (CBD) oligonucleotides were obtained from the Caltech DNA Synthesis Facility (Figure 2.3). To anneal the single-stranded DNA, 10 μL of each DNA strand (5 μg/μL in 10 mM Tris buffer at pH 8.1) was added to 80 μL of annealing buffer (100 mM NaCl, 20 mM $MgCl_2$). The solution was boiled for 3 min and allowed to cool in a Styrofoam box overnight. The double-stranded (ds) DNA was run on a 2% agarose gel and the bands (90 bp for RGD, RDG, PHSRN, NHRPS and 117 bp for CS1') were purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). The ds DNA was phosphorylated by T4 polynucleotide kinase (New England Biolabs [NEB], Ipswich, MA) with 1 mM adenosine 5' triphosphate (ATP) for 1 h at 37 °C. It was then isolated and purified on a 2% agarose gel.

In order to amplify the cell-binding domain DNA, it was ligated (T4 ligase [Boehringer Mannheim, Mannheim, Germany], overnight, 16 °C) into a purified pUC19 vector that had been digested with *Eco*RI and *Bam*HI (NEB, 2–3 h, 37 °C) and dephosphorylated with calf intestinal alkaline phosphatase (CIP, NEB, 10 min, 37 °C). The resulting mixtures were transformed directly into the *E. coli* strain DH5αF'.

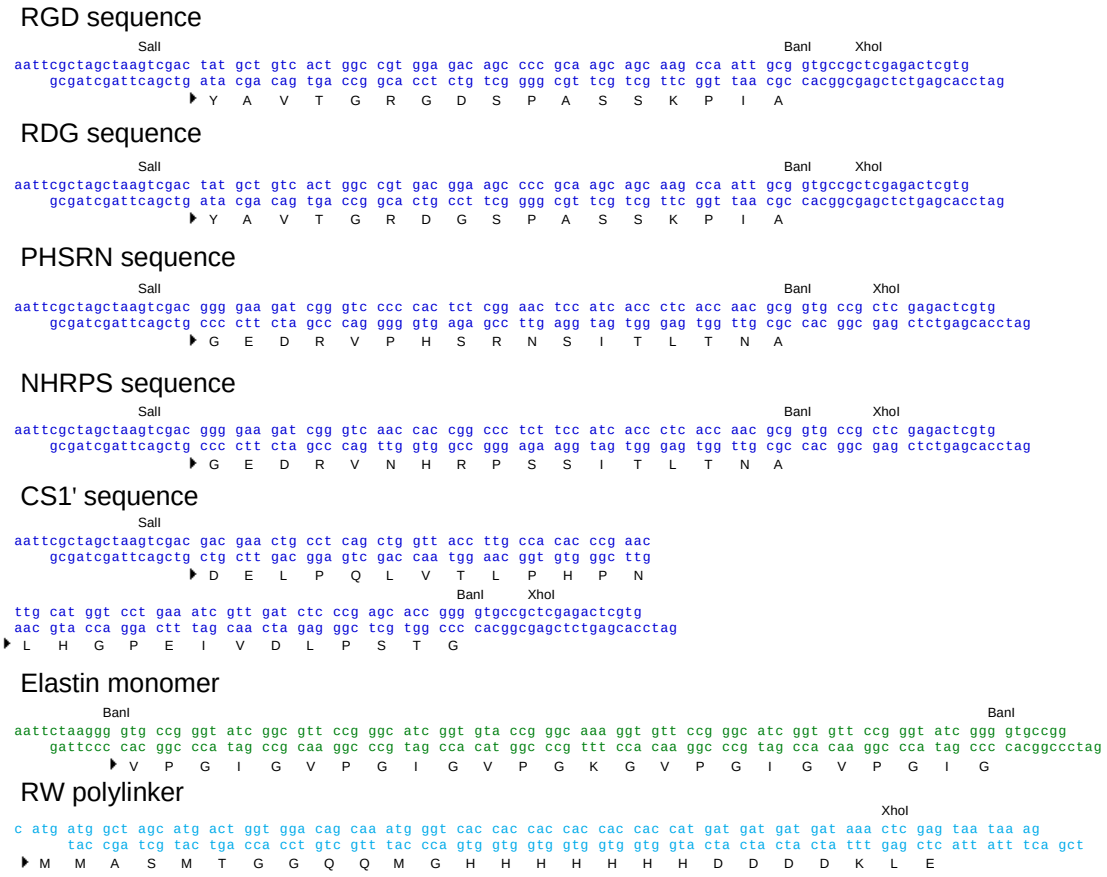

Figure 2.3 DNA oligonucleotide sequences are shown of the following bioactive domains: the RGD cell-binding domain, the scrambled RDG sequence, the PHSRN synergy sequence, the scrambled NHRPS sequence, and the scrambled CS1' sequence. The sequences for the elastin monomer and the RW polylinker are also included.

Individual transformants were grown in 5 mL overnight 2×YT cultures containing 200 mg/mL ampicillin. DNA was isolated with the QIAprep Spin Miniprep Kit (Qiagen), digested with *Eco*RI and *Bam*HI, and run on a 2% agarose gel to verify successful insertion. Sequencing analysis (Caltech DNA Sequencing Facility) with an M13 reverse

primer confirmed that the clones had the correct oligonucleotides inserted in the pUC19 vector. Similar techniques were used to ligate the CBDs into the pEC2 vector[14] between the *Eco*RI and *Bam*HI sites. Overnight cultures were supplemented with 25 mg/L kanamycin.

2.3.3 Cloning of pEC2-CBD-EL$_4$

Cells containing the DNA sequence for the elastin monomer (EL) with internal lysine residues (Figure 2.3) in a pUC19 vector (pUC19-EL) were obtained from Di Zio.[1] The internal lysine was included to facilitate crosslinking throughout the protein. The EL monomer (75 bp) was isolated (*Ban*I [NEB], 37 °C, overnight) and ligated at 16 °C for 35 min to form 300 bp tetramers (EL$_4$).

The purified elastin tetramers were ligated into *Ban*I-digested (37 °C, 2–3 h), dephosphorylated pEC2-CBD vectors. DNA was isolated and digested with *Xho*I and *Sal*I (NEB, 37 °C, 3–4 h) to check for successful insertions (363 bp for RGD, RDG, PHSRN, NHRPS and 390 bp for CS1').

2.3.4 Cloning of pET28aRW-(CBD-EL$_4$)$_3$

The pET28a vector (Novagen, San Diego, CA) with a unique polylinker (Figure 2.3) inserted between the *Nco*I and *Xho*I sites (pET28aRW) was obtained from Ralf Weberskirch. The polylinker contains a heptahistidine tag, a T7 tag for antibody identification, and an enterokinase cleavage site. DNA fragments isolated from digesting pEC2-CBD-EL$_4$ with *Xho*I and *Sal*I were sequentially inserted into the pET28aRW vector at the *Xho*I site. Sequencing analysis with a T7 forward primer was

done after the first insertion to verify the DNA sequences. The final products were digested with *Xho*I and *Xba*I to confirm that the DNA constructs were the expected lengths (1202 bp for RGD, RDG, PHSRN, NHRPS and 1283 bp for CS1'). These DNA sequences were also verified by sequencing analysis with T7 forward and reverse primers. Due to the lengths and repetitive structure of our constructs, there is a ~300–400 bp region in the middle where the sequencing signal is weak and the sequencing is inconclusive. All five constructs have a silent mutation in the last histidine of the heptahistidine tag where the DNA is CAT instead of CAC. The construct for aECM-PHSRN also has a silent mutation in the last glycine codon of the elastin tetramer in which the sequence is GGA instead of GGG.

2.3.5 Protein Expression and Purification

The pET28aRW-$(CBD\text{-}EL_4)_3$ vectors were transformed into a BL21(DE3) pLysS expression host (Stratagene, La Jolla, CA). Expression was performed in a 10 L Bioflow 3000 fermentor (New Brunswick Scientific, Edison, NJ). Cells were grown in Terrific Broth supplemented with 25 mg/L kanamycin and 35 mg/L chloramphenicol to an optical density at 600 nm (OD_{600}) of 5–6. Cells were harvested 1.5–2 h after induction with 2.5 mM isopropyl-1-β-D-thiogalactosidase (IPTG). The cells were resuspended in TEN buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 100 mM NaCl) at a concentration of 0.5 g/mL. The cells were frozen at –20 °C and defrosted at 4 °C with 10 μg/mL of deoxyribonuclease I (Sigma, St. Louis, MO), 10 μg/mL of ribonuclease A (Sigma), and 50 μg/mL of phenylmethylsulfonyl fluoride (Sigma). Water was added to bring the total volume of the solution to 1.3 L. Proteins were readily purified via a series of three

temperature cycles. The pH of the solution was adjusted to 9 and the solution was centrifuged (39,750*g*, 2 h, 4 °C). The resulting supernatant was adjusted to 1 M NaCl at 4 °C, warmed to 37 °C, and centrifuged (39,750*g*, 2 h, 37 °C). The pellet was then redispersed in water at a concentration of 100 mg/mL. This process was repeated twice. The solution was dialyzed at 4 °C for 3 days and lyophilized.

The expected molecular weights of the proteins are: 34,839 for aECM-RGD and aECM-RDG; 35,409 for aECM-PHSRN and aECM-NHRPS; and 38,174 for aECM-CS1'. The purity and molecular weights of the proteins were verified by SDS–PAGE and Western blots with an anti-T7 horseradish peroxidase-conjugated antibody (Novagen). Amino acid analysis was performed at University of California, Davis. Matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS) was conducted with a 20:1 ratio of matrix (10 mg/mL sinapic acid in 0.07% trifluoroacetic acid, 30% acetonitrile) to protein (30 mg/mL protein in 4 °C water) on an Applied Biosystems Voyager DE-PRO spectrometer (Foster City, CA) at the Caltech Mass Spectrometry Facility. The protein solution was spotted onto the plate at 4 °C to prevent protein aggregation.

2.4 Results and Discussion

2.4.1 Protein Yield

The wet cell mass averaged 212 g per 10 L fermentation for aECM-RGD and aECM-RDG and 256 g for aECM-PHSRN and aECM-NHRPS. Typical yields of a 10 L

fermentation were 720 mg for aECM-RGD and aECM-RDG and 1.2 g for aECM-PHSRN and aECM-NHRPS.

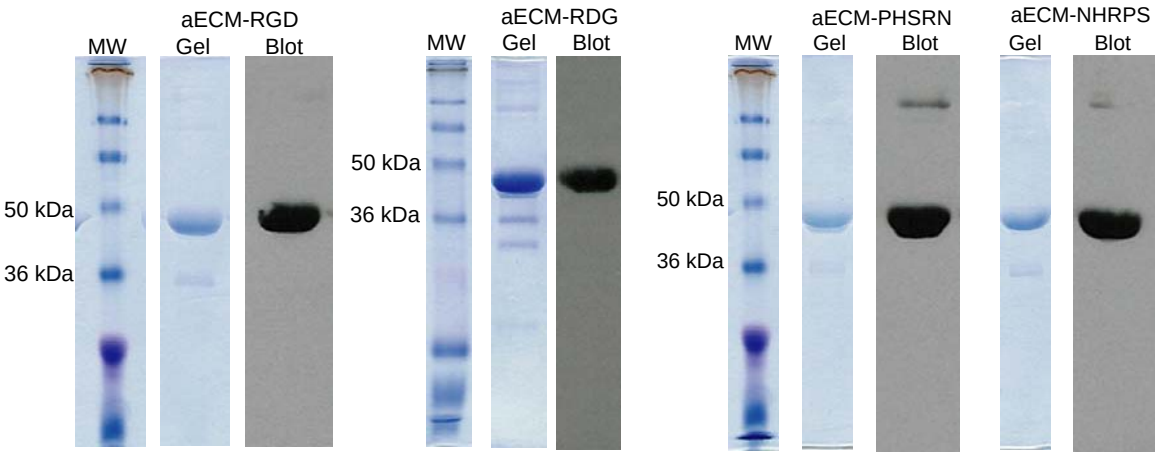

Figure 2.4 SDS–PAGE gels and Western blots of purified aECM-RGD, aECM-RDG, aECM-PHSRN, and aECM-NHRPS.

2.4.2 Protein Purity

Protein gels and Western blots were used to assess the purity and molecular weight of aECM-RGD, aECM-RDG, aECM-PHSRN, and aECM-NHRPS (Figure 2.4). The aECM proteins were the predominant species in the gel, and the protein bands were between the 36 and 50 kDa markers. The apparent molecular weight was therefore higher than the expected 34.8 kDa for aECM-RGD and aECM-RDG and 35.4 kDa for aECM-PHSRN and aECM-NHRPS, but previous work indicates that this family of elastin-like proteins runs at higher molecular weights than predicted.[1]

Amino acid analysis of aECM-RGD and aECM-RDG correlated well with the theoretical amino acid composition (all values are within 1 mol%) and further confirmed the structures of these aECM proteins (Table 2.1). Mass spectrometry indicated that the

molecular weights of aECM-RGD, aECM-RDG, aECM-PHSRN, and aECM-NHRPS were within 0.1% of the predicted value (Figure 2.5). Taken together, the SDS–PAGE, Western blotting, amino acid analysis, and mass spectrometry results demonstrated that the purified aECM proteins were of the correct molecular weight and amino acid composition.

Amino Acid	Theoretical mol%	aECM-RGD mol%	aECM-RDG mol%
Ala (A)	2.6	3.1	3.1
Arg (R)	0.8	0.8	0.8
Asn (N)	0.0	0.0	0.0
Asp (D)	2.6	2.7	2.8
Cys (C)	0.0	0.0	0.0
Gln (Q)	0.5	0.0	0.0
Glu (E)	0.3	0.8	0.8
Gly (G)	33.2	32.4	32.6
His (H)	1.8	1.8	1.8
Ile (I)	13.1	13.1	13.3
Leu (L)	1.0	1.1	1.0
Lys (K)	4.1	4.2	4.2
Met (M)	1.0	0.7	0.6
Phe (F)	0.0	0.0	0.0
Pro (P)	17.7	18.4	18.3
Ser (S)	2.6	2.3	2.3
Thr (T)	1.0	1.0	0.9
Trp (W)	0.0	0.0	0.0
Tyr (Y)	0.8	0.8	0.8
Val (V)	17.0	16.9	16.8
Total	100.1	100.1	100.1

Table 2.1 Theoretical and measured amino acid compositions of aECM-RGD and aECM-RDG. All values are within 1 mol%.

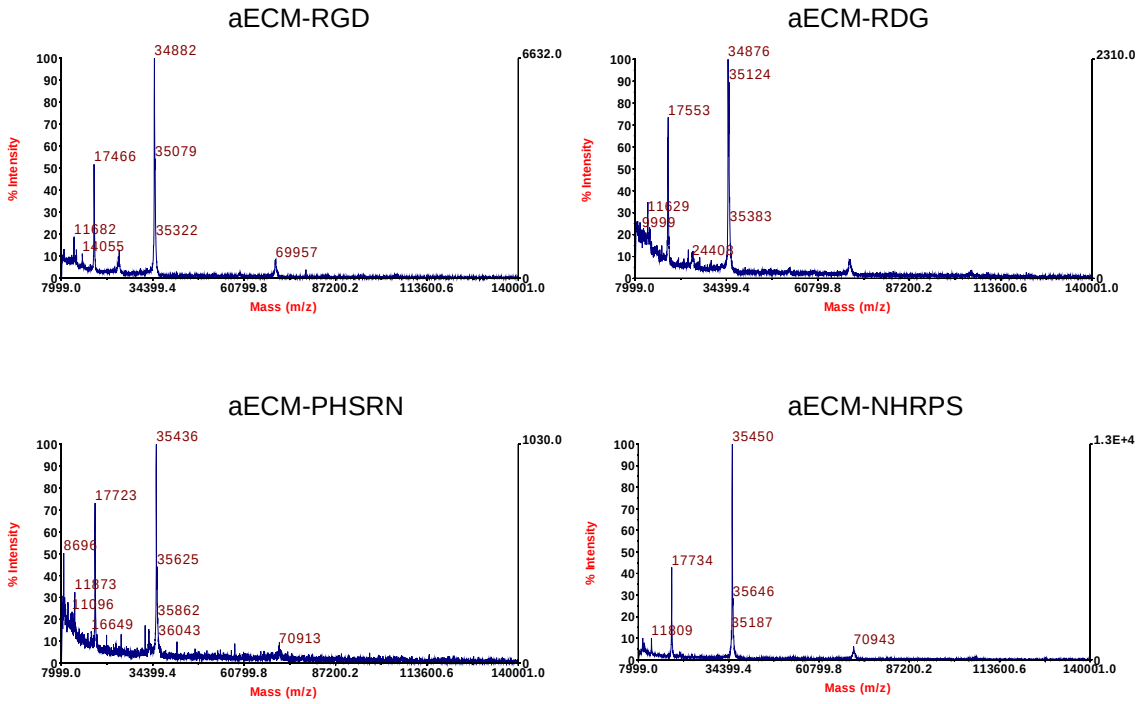

Figure 2.5 Mass spectrometry analysis of aECM-RGD, aECM-RDG, aECM-PHSRN, and aECM-NHRPS. The expected molecular weight is 34,839 for aECM-RGD and aECM-RDG and 35,409 for aECM-PHSRN and aECM-NHRPS. The measured molecular weights are within 0.1% of the expected values.

2.5 Conclusion

The successful cloning of five aECM proteins is described in this work. The average yield of a 10 L fermentation of aECM-RGD, aECM-RDG, aECM-NHRPS, or aECM-NHRPS was 800 mg. SDS–PAGE, Western blots, and mass spectrometry of these four polymers indicated that proteins of the correct molecular weight were readily

purified. Amino acid analysis further confirmed that aECM-RGD and aECM-RDG were of the correct composition.

2.6 References

1. Di Zio K. "Protein matrices engineered for control of cell and tissue behavior," Ph.D. Thesis, University of Massachusetts Amherst, **2002**.

2. Panitch A, Yamaoka T, Fournier MJ, Mason TL, Tirrell DA. Design and biosynthesis of elastin-like artificial extracellular matrix proteins containing periodically spaced fibronectin CS5 domains. *Macromolecules* **1999**, *32*, 1701–1703.

3. Heilshorn SC. "Design and characterization of artificial extracellular matrix proteins for use as small-diameter vascular grafts," Ph.D. Thesis, California Institute of Technology, **2004**.

4. Welsh ER, Tirrell DA. Engineering the extracellular matrix: A novel approach to polymeric biomaterials. I. Control of the physical properties of artificial protein matrices designed to support adhesion of vascular endothelial cells. *Biomacromolecules* **2000**, *1*, 23–30.

5. Di Zio K, Tirrell DA. Mechanical properties of artificial protein matrices engineered for control of cell and tissue behavior. *Macromolecules* **2003**, *36*, 1553–1558.

6. Hersel U, Dahmen C, Kessler H. RGD modified polymers: Biomaterials for stimulated cell adhesion and beyond. *Biomaterials* **2003**, *24*, 4385–4415.

7. Ruoslahti E. RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.* **1996**, *12*, 697–715.

8. Redick SD, Settles DL, Briscoe G, Erickson HP. Defining fibronectin's cell adhesion synergy site by site-directed mutagenesis. *J. Cell Biol.* **2000**, *149*, 521–527.

9. Aota S, Nomizu M, Yamada KM. The short amino acid sequence Pro-His-Ser-Arg-Asn in human fibronectin enhances cell-adhesive function. *J. Biol. Chem.* **1994**, *269*, 24756–24761.

10. Grant RP, Spitzfaden C, Altroff H, Campbell ID, Mardon HJ. Structural requirements for biological activity of the ninth and tenth FIII domains of human fibronectin. *J. Biol. Chem.* **1997**, *272*, 6159–6166.

11. Leahy DJ, Aukhil I, Erickson HP. 2.0 angstrom crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* **1996**, *84*, 155–164.

12. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds.), *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.: New York, NY, **2003**.

13. Sambrook J, Frisch EF, Maniatis T, *Molecular Cloning*. Cold Spring Harbor Laboratory Press: Plainview, NY, **1989**.

14. Cantor EJ. "Effects of amino acid substitution on chain packing in genetically engineered periodic polypeptides," Ph.D. Thesis, University of Massachusetts Amherst, **1997**.