TOWARD RATIONAL DESIGN OF FUNCTIONAL MATERIALS FOR BIOLOGICAL APPLICATIONS

by

Charng-Yu Lin

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Davidson School of Chemical Engineering

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Julie C. Liu, Chair
Davidson School of Chemical Engineering
Dr. Alyssa Panitch
Department of Biomedical Engineering
University of California, Davis
Dr. Chongli Yuan
Davidson School of Chemical Engineering
Dr. You-Yeon Won
Davidson School of Chemical Engineering

Approved by:
Dr. John Morgan
Head of the School Graduate Program

TABLE OF CONTENTS

LIST OF TABLES 7
LIST OF FIGURES 8
ABSTRACT 21
1. INTRODUCTION 24
1.1 Tissue Engineering 24
1.1.1 Three Components in Tissue Engineering 25
1.2 Drug Delivery 27
1.3 Tissue Adhesives 29
1.4 Functional Biomaterials 30
1.4.1 Synthetic Polymers 31
1.4.2 Natural Polymers 33
1.4.3 Recombinant Proteins 36
1.4.3.1 Structural Domains 37
1.4.3.2 Bioactive Domains 38
1.4.3.3 Domains for Higher-Order Control over Material Properties 41
1.5 Thesis Outline and Contributions 43
1.6 References 46
2. REDOX-RESPONSIVE HYDROGELS VIA ONE-POT THIOL-ENE ADDITION AND THIOL OXIDATION WITH DECOUPLED DEGRADATION BEHAVIORS AND MECHANICAL PROPERTIES 59
2.1 Abstract 59
2.2 Introduction 59
2.3 Experimental Methods 62
2.3.1 Materials 62
2.3.2 Hydrogel Formation 62
2.3.3 Rheological Characterization 62
2.3.4 Swelling and Degradation Analysis 63
2.3.5 Quantification of *in vitro* Release of Dextran 63
2.3.6 Cell Encapsulation and Cytocompatibility 64

2.4 Results and Discussion 64
2.4.1 Crosslinker Formulation Independent Hydrogel Formation and Shear Modulus 64
2.4.2 Tunable Swelling and Degradation of the Hydrogels with Similar Mechanical Properties 69
2.4.3 Increased FITC-Dextran Release in the Presence of Glutathione 71
2.4.4 DVS and Ferric EDTA Crosslinking Was Cytocompatible 74
2.5 Conclusions 75
2.6 Acknowledgments 76
2.7 References 76
3. PROTEIN ADHESIVES BASED ON ELASTIN-LIKE POLYPEPTIDES: A COMPARISON BETWEEN CATECHOL- AND THIOL-BASED ADHESION 81
3.1 Abstract 81
3.2 Introduction 81
3.3 Experimental Methods 84
3.3.1 Materials 84
3.3.2 Protein Design and Cloning 84
3.3.3 Protein Expression and Purification 85
3.3.4 Tyrosinase Modification 86
3.3.5 Reduced-Thiol Quantification 86
3.3.6 Lap Shear Adhesion Test 87
3.3.7 Cytocompatibility Test 88
3.3.8 Statistical Analysis 89
3.4 Results and Discussion 89
3.4.1 Protein Design and Production 89
3.4.2 Lap Shear Adhesion Strength Depends on Time and Chemistry 90
3.4.3 Iron Ions Enhanced the Performance of m[YKV-72] 93
3.4.4 Combination of Thiol and DOPA Chemistries Increased the Adhesion Strength. 96
3.4.5 Similar Curing Process between m[YKV-72]-[CKVYKV-72] Mixture and [CKVYKV-72] Along 98

3.4.6 Adhesion Recovery after a Disturbance Depends on the Protein and the Curing Conditions. 100
3.4.7 [CKVYKV-72] with Iron Ions is More Cytocompatible than m[YKV-72] with Iron Ions. 107
3.5 Conclusions 109
3.6 Acknowledgements 110
3.7 Supporting Information 110
3.8 References 114
4. TAILORING THE TEMPERATURE RESPONSIVENESS OF POSITIVELY-CHARGED ELASTIN-LIKE POLYPEPTIDES WITH EXTERNAL SEQUENCES 120
4.1 Abstract 120
4.2 Introduction 120
4.3 Experimental Methods 123
4.3.1 Materials 123
4.3.2 Protein Design and Cloning 123
4.3.3 Protein Expression and Purification 123
4.3.4 LCST Measurement 124
4.4 Results and Discussion 125
4.4.1 Design and Production of the S-tag and I-tag[YKV] Proteins 125
4.4.2 Short Non-ELP Sequences Introduced pH Sensitivity to the LCST Behavior 127
4.4.3 Protein Concentration Affected LCST Behavior and pH Sensitivity 129
4.4.4 pH Sensitivity Resulted from the Protonation State of Histidine 132
4.4.5 The Extent of pH sensitivity Depended on ELP Domain Length 132
4.4.6 Increasing Ionic Strength Shielded pH-Sensitive LCST Behavior Introduced by the Tag Sequence 133
4.4.7 Hydrophobicity of the Tag Affected LCST Behavior 134
4.4.8 Tag Identity Affected the Interaction between ELP Proteins and Hofmeister Ions 137
4.5 Conclusions 139
4.6 Acknowledgements 140

4.7 Supporting Information 141
4.8 References 144
5. CONCLUSIONS 149
5.1 Summary 149
5.2 Future Directions 150
5.3 References 152
APPENDIX A. TUNABLE LIGAND PRESENTATION TO GUIDE STEM CELL DIFFERENTIATION 153
APPENDIX B. NUCLEOPHILE-CATECHOL PROTEIN CONJUGATION AND CORSSLINKING 173
APPENDIX C. RECOMBINANT PROTEIN-BASED SCAFFOLD FOR RAT CALVARIAL CRITICAL DEFECT HEALING 185
APPENDIX D. RECOMBINANT PROTEIN-BASED EPIGENETIC PROBES FOR DNA METHYLATION 214
APPENDIX E. PROTOCOLS 243
APPENDIX F. CLONING SCHEMES AND SEQUENCES 255
VITA 330

LIST OF TABLES

Table 4.1. Reference LCSTs at 1 μM protein and concentration dependence b_{pH} of I-tag[YKV-48] and S-tag[YKV-48].[a] 130

Table B. 1.Amino acid sequences of common bioactive peptides. 174

Table C. 1 Gel clot LAL assay results of Ni-IPA-purified YC12-T.....196

Table D. 1 List of constructs based on the BiFC pair VN173/VC155 222

Table D. 2. List of constructs based on the BiFC pair VN210/VC210 223

LIST OF FIGURES

Figure 2.1. One-pot crosslinking through thiol-ene addition and thiol oxidation. (A) Chemical structures of four-arm poly(ethylene glycol) thiol (PEG-SH), divinyl sulfone (DVS), and sodium ferric ethylenediaminetetraacetic acid (ferric EDTA). (B) Formation of thioether bonds between PEG-SH and DVS through thiol-ene addition and formation of disulfide bonds between two PEG-SH through thiol oxidation promoted by ferric EDTA. (C) Illustration of the hydrogel network with both thioether and disulfide bonds. In the presence of reducing agents, disulfide bonds are reduced and broken. On the other hand, thioether bonds are not reducible and remain intact in the network. (D) Hydrogels crosslinked with 50 mM ferric EDTA with different DVS concentrations showed similar gelation times as determined by tube inversion test. All formulations formed hydrogels within 10 min at 37 °C. Color change was observed during the gelation process. The color was initially deep purple and turned into dark brown then light brown. The rate of this color change depended on the crosslinking formulation..... 66

Figure 2.2. Similar storage moduli of hydrogels crosslinked with 0, 2, 4, and 8 mM DVS were observed in (A) frequency and (B) strain sweeps. All hydrogels displayed linear viscoelastic regions from 0.1 to 10 Hz at 1% strain and from 0.1 to 10% strain at 1 Hz. Data (n = 3) are presented as the average ± the standard deviation..... 67

Figure 2.3. Degradation in 10 μM GSH increased with hydrogels crosslinked with lower DVS concentrations. All hydrogels showed similar swelling ratios when incubated in PBS. When incubated in PBS with 10 μM GSH, hydrogels crosslinked with 6 and 8 mM DVS showed no degradation, but swelling ratios were slightly higher than when incubated in PBS. On the other hand, hydrogels crosslinked with 2 and 4 mM DVS degraded and showed significantly higher swelling ratios compared those when incubated in PBS. The 2-mM DVS hydrogel disintegrated completely after 120h of incubation with 10 μM GSH. Data (n = 3) are presented as the average ± the standard deviation..... 70

Figure 2.4. Faster and higher release of encapsulated FTIC-dextran from the 2- and 4-mM DVS hydrogels in PBS with 10 μM GSH. On the other hand, release profiles from hydrogels crosslinked with 6 and 8 mM DVS were similar when incubated in either PBS only or PBS with 10 μM GSH. Hydrogels crosslinked with 2 and 4 mM DVS released >99% of the encapsulated FITC-dextran after 120 h. Meanwhile, only ~87% was released from the hydrogels when incubated in PBS. Data (n = 3) are presented as the average ± the standard deviation. 72

Figure 2.5. NIH/3T3 mouse fibroblasts encapsulated in the dual-crosslinked hydrogels showed good viability up to 5 days. 50 mM ferric EDTA and 8 mM DVS were used during cell encapsulation. Cells stained green are viable, and cells stained red are dead. Most encapsulated cells were viable on days 1 and 5. The XY projection images are maximum intensity projections from confocal scanned stacks and were created using ImageJ. The unit for the grid in the 3D reconstruction images is μm..... 74

Figure 3.1. Curing kinetics depend on the adhesive chemistry used. The graph shows the adhesion strength of m[YKV-72] and [CKVYKV-72] cured for 1, 6, and 24 h. m[YKV-72] showed an increased strength with longer cure time. All samples of m[YKV-72] cured for 1 h failed during the test setup. [CKVYKV-72] showed similar adhesion strengths among all three cure times. Within each protein group, identical letters indicate statistically similar data as determined by Tukey's HSD test. * indicates a statistical difference ($p < 0.05$) between proteins at the same time point with an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests. 91

Figure 3.2. Addition of iron ions accelerates the curing kinetics of m[YKV-72]. The graph shows the adhesion strengths of m[YKV-72] and [CKVYKV-72] with iron ions cured for 1, 6, and 24 h. Lower failure rates and measurable adhesion strengths were obtained at 1 h cure time with m[YKV-72] with iron ions compared to that the results without Iron ions. [CKVYKV-72] showed similar adhesion strengths among three cure times. Within each protein group, identical letters indicate statistically similar data as determined by Tukey's HSD test. * indicates a statistical difference with $p < 0.05$ as assessed with an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests.. 94

Figure 3.3. Mixing m[YKV-72] and [CKVYKV-72] at a mass ratio of 1-1.5 resulted in a significant increase in adhesion strength. m[YKV-72] and [CKVYKV-72] were mixed at three different mass ratios and were cured for 24 h in the presence of iron ions. The lap shear adhesion test results of each group were shown in (A) using m[YKV-72]-[YKV-72] as the negative control groups (DOPA only) and in (B) using [YKV-72]-[CKVYKV-72] as the negative control groups (thiol only). In both panels, the graphs show the normalized adhesion strengths, and the table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests. In panel (A), because two-way ANOVA indicated significant interaction between the two factors (formula and ratio), one-way ANOVA was performed on each formula separately with Dunnett's test. & indicates a significant difference to the control group of m[YKV-72], iron ions, 24 h from Figure. 2. with $p < 0.05$. # indicates a significant difference to the control group of [CKVYKV-72], iron ions, 24 h with $p < 0.05$. * indicates a statistical difference with $p < 0.05$ with an unpaired *t*-test. In panel (B), two-way ANOVA was performed using formula and ratio as the two factors with Tukey's HSD *post hoc* test. @ indicates the two group does not share the same Tukey group with a $p < 0.05$. # indicates a significant difference to the control group of [CKVYKV-72], iron ions, 24 h with $p < 0.05$ with Dunnett's test. 97

Figure 3.4. m[YKV-72]-[CKVYKV-72] 1-1.5 showed fast curing kinetics that were similar to that of [CKVYKV-72]. The graph shows the adhesion strengths with or without iron ions at three curing times. Within each formula group, identical letters indicate statistically similar data as determined by Tukey's HSD test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). 99

Figure 3.5. (A) An illustration on the curing and test process. (B) In the absence of iron ions, the adhesion strength of m[YKV-72] recovered from disturbances. The graph shows the adhesion strengths of m[YKV-72] from both the first and second applications. When no iron ions were used, the adhesion strengths after the second application were statistically similar to that of m[YKV-72] cured for 24 h undisturbed (Figure 3.1). With iron ions, the adhesion strengths were lower after the second applications compared to that of m[YKV-72] with iron ions cured for 24 h undisturbed (Figure 3.2). & indicates a significant difference to the control group of m[YKV-72], Iron ions, 24 h in Figure 3.2 with $p < 0.05$ in Dunnett's test. * indicates a statistical difference with $p < 0.05$ as determined by an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.1 and 2. (C) The adhesion strength of [CKVYKV-72] did not recover from disturbance after long curing times. The graph shows the adhesion strengths of [CKVYKV-72] after both the first and the second applications. Regardless of the presence of Iron ions, the adhesion strengths of the second applications at 1 h r were statistically similar to the strengths of the respective first applications and to their respective undisturbed groups that were cured for 24 h. The second application without Iron ions at 6 h showed statistically lower strength compared to the strengths after the first application and of the undisturbed 24-h group. The second application with Iron ions showed a markedly lower strength compared to the strength after the first application and had a high failure rate; however, statistical analysis was not performed due to insufficient sample size. # indicates a significant difference as assessed by Dunnett's test to the control group of [CKVYKV-72], 24 h in Figure 3.1 with $p < 0.05$. * indicates a statistical difference as determined by an unpaired *t*-test with $p < 0.05$. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.1 and Figure 3.2. 102

Figure 3.6. In the presence of iron ions, the adhesion strength of m[YKV-72]-[CKVYKV-72] 1-1.5 recovered from a disturbance. The graph shows the adhesion strengths of m[YKV-72]-[CKVYKV-72] 1:1.5 after both the first and second applications. Without iron ions, the adhesion strengths were statistically lower than the corresponding strengths after the first application and for the control group of m[YKV-72]-[CKVYKV-72] 1-1.5 cured for 24 h undisturbed. With iron ions, the adhesion strengths after the first and second applications at 1 h were similar. With iron ions, the second application at 6 h showed a statistically lower strength compared to that of the first application but was similar to m[YKV-72]-[CKVYKV-72] 1-1.5 cured for 24 h undisturbed (Figure 3.4). # indicates a significant difference as assessed by Dunnett's test to the control group of m[YKV-72]-[CKVYKV-72], 24 h with $p < 0.05$. * indicates a statistical difference as determined by an unpaired *t*-test with $p < 0.05$. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.4. 105

Figure 3.7. [CKVYKV-72] showed higher cytocompatibility than m[YKV-72]. Live/dead images of NIH/3T3 fibroblasts cultured on hydrogels of [CKVYKV-72], m[YKV-72], and m[YKV-72]:[CKVYKV-72] 1:1.5 all with iron ions show good viability on all three surfaces up to day 5 of culture. Cells stained green are viable, and cells stained red are dead. No cells were observed on m[YKV-72] at day 3, and these results suggest that the cell attachments were not strong enough to withstand the washing steps during the medium change and live/dead staining steps. Scale bar represents 100 μm. 108

Figure 3.8. Amino acid sequences of [YKV-72] and [CKVYKV-72]. (A) Each repeat of (PGYGVPGKGVPGVGV) contains three ELP pentapeptide sequences with guest residues being Y, K, and V. There are 24 repeats, and thus there is a total of 72 pentapeptide sequences. (B) Each repeat of (PGYGVPGKGVPGVGVPGCGVPGKGVPGVGV)) contains six ELP pentapeptide sequences with guest residues being Y, K, V, C, K, and V. There are 12 repeats that result in a total of 72 pentapeptide sequences 110

Figure 3.9. SDS-PAGE and Western blot results of (A) [YKV-72] and (B) [CKVYKV-72]. Densitometry of SDS-PAGE images showed >95% purity for both proteins. Western blot confirmed the presence of the T7-tag on purified [YKV-72] and the His-tag on purified [CKVYKV-72]. The expected molecular weights of the proteins are 32 kDa for [YKV-72] and 34 kDa for [CKVYKV-72]. 111

Figure 3.10. Electrospray ionization mass spectrometry of (A) [YKV-72] (expected MW: 32160 Da) and (B) [CKVYKV-72] (expected MW: 34006 Da). 112

Figure 3.11. Adhesion strengths in the presence or absence of ferric ions for (A) m[YKV-72] and (B) [CKVYKV -72]. With m[YKV-72], there were statistical difference between groups with and without ferric ions at 6 and 24 h. Statistical analysis was not performed at 1 h due to the 100% failure rate of the group without ferric ions. In the case of [CKVYKV -72], adhesion strengths without ferric ions were statistically higher than those with ferric ions at 1 and 6 h. However, there was no statistical difference at 24 h. * indicates statistical difference as determined by a *t*-test with $p < 0.05$. Data are replotted from Figure 1 and 2 in the main text..... 113

Figure 4.1. Design and production of the I-tag and S-tag[YKV] proteins. (A) Illustration of the differences between the I-tag and S-tag[YKV] proteins. All proteins shared a common ELP YKV domain. I-tag included the sequence of SKGPG. The S-tag was composed of a T7 tag, a 7x His tag, and an enterokinase cleavage site. (B) SDS-PAGE and Western blot images of purified I-tag and S-tag[YKV] proteins. All proteins on the SDS-PAGE gel were close to their expected molecular weights (I-tag[YKV-48]: 21 kDa, S-tag[YKV-48]: 25 kDa, S-tag[YKV-72]: 28.5 kDa, and S-tag[YKV-96]: 32 kDa). Western blot confirmed the presence of the T7 tag on the S-tag[YKV] proteins. Western blot was not applicable for I-tag[YKV-48] due to a lack of the T7 tag epitope used for detection..... 127

Figure 4.2. The LCSTs of I-tag[YKV-48] are insensitive to pH, whereas the LCSTs of S-tag[YKV-48] decrease significantly with increasing pH. LCST measurements of (A) I-tag[YKV-48] and (B) S-tag[YKV-48] at protein concentrations ranging from 50 to 400 μM and pH values ranging from 5.5 to 8.0. Both proteins had lower LCSTs with increasing protein concentration. At a given concentration, the LCSTs of I-tag[YKV-48] were similar at all pH values. On the other hand, S-tag[YKV-48] showed pH-sensitive LCSTs. NaCl was added to the solution for all groups at a final concentration of 0.2 M. The LCST values were presented as the average with standard deviation of three independent samples. . 128

Figure 4.3. LCSTs of I-tag[YKV-48] and S-tag[YKV-48] were well described by a two-parameter equation. LCST measurements with best-fit lines from equation 1 of (A) I-tag[YKV-48] and (B) S-tag[YKV-48]. The LCST values were presented as the average with standard deviation of three independent samples. 130

Figure 4.4. Different pH sensitivity in concentration dependence (b_{pH}) between I-tag[YKV-48] and S-tag[YKV-48]. Fitted b_{pH} of I-tag and S-tag[YKV-48] with standard error were plotted as a function of pH. Linear regression lines were included to show the trend in b_{pH}. The b_{pH} values for I-tag[YKV-48] yielded a nearly horizontal line with an R^2 close to zero and indicated that b_{pH} was not sensitive to pH. On the other hand, the b_{pH} values for S-tag[YKV-48] were sensitive to pH and exhibited two different trends below and above pH 7..... 131

Figure 4.5. Longer YKV domain lengths decreased the extent of pH sensitivity derived from the S-tag sequence. At 200 μM protein and 0.2 M NaCl, pH sensitivity was more prominent at a lower pH region (pH 5.5 to 6.5). The data set of S-tag[YKV-48] is replotted from Figure 4.2B. LCST values were presented as the average with standard deviation of three independent samples. 133

Figure 4.6. High ionic strength shielded the electrostatic interactions between S-tag and the YKV-48 domain. LCST measurements of (A) I-tag and (B) S-tag[YKV-48] at different NaCl concentrations with 200 μM protein. The LCST behavior of I-tag[YKV-48] remained insensitive to pH over all NaCl concentrations. On the other hand, the LCST behavior of S-tag[YKV-48] showed decreasing pH sensitivity with increasing NaCl concentration. In addition, the two proteins responded differently to the increasing NaCl concentration. The LCST of I-tag[YKV-48] initially increased until 0.3 M NaCl then decreased. In contrast, the LCSTs of S-tag[YKV-48] decreased monotonically with increasing NaCl concentration. LCST values were presented as the average with standard deviation of three independent samples. 134

Figure 4.7. High protein concentration and high ionic strength decreased the LCSTs of S-tag[YKV-48] more significantly then those of I-tag[YKV-48]. LCST differences between S-tag and I-tag[YKV-48] with different (A) protein concentrations and (B) NaCl concentrations. S-tag[YKV-48] had higher LCSTs than I-tag[YKV-48] at low pH values and low protein concentrations. As the protein concentration increased, S-tag[YKV-48] only had a higher LCST at pH 5.5. With varying NaCl concentration, S-tag[YKV-48] only had higher LCSTs at NaCl concentrations <0.2 M. The plotted differences were based on the averaged LCSTs reported in Figure 4.2 and 6. 135

Figure 4.8. External tag identity affected the interactions between ELPs and Hofmeister ions. LCST measurements of I-tag[YKV-48] and S-tag[YKV-48] with (A) Na_2SO_4 and (B) NaI. Both I-tag[YKV-48] and S-tag[YKV-48] had decreasing LCSTs at >0.2 M Na_2SO_4. I-tag[YKV-48] showed an increasing trend in LCST from 0.1 to 0.2 M Na_2SO_4 even with a strong kosmotropic ion such as SO_4^{2-}. With NaI, both proteins exhibited a similar trend that LCSTs decreased before 0.4 M NaI and increased after that. LCSTs with Na_2SO_4 were determined at a protein concentration of 50 μM at pH 5.5. LCSTs with NaI were determined at a protein concentration of 200 μM at pH 7.5. The presented LCST values were averages of three independent samples with standard deviation. 138

Figure 4.9. Electrospray ionization mass spectroscopy spectra of (A) S-tag[YKV-72] (expected MW: 32160 Da), (B) S-tag[YKV-96] (expected MW: 46692 Da), and (C) I-tag[YKV-48] (expected MW: 23964 Da). Peaks were labeled with the m/z ratios and the corresponding number of charges on the polymer chains. 141

Figure 4.12. Amino acid analysis results of I-tag[YKV-48]. The result showed a good match between the measured molar fraction and the expected molar fraction of each amino acid. 142

Figure 4.13. (A) Amino acid sequences of I-tag[YKV-48] and S-tag[YKV-48]. Each repeat of (PGYGVPGKGVPGVGV) contains three ELP pentapeptide sequences with guest residues being Y, K, and V. The 16, 24, or 32 repeats contain a total of 48, 72, or 96 pentapeptide sequences, respectively. (B) Charge estimation of I-tag, S-tag, and YKV sequences as indicated in (A). The charge estimation of the amino acid sequence was calculated using the Henderson-Hasselbalch equation by assuming that the pKa value of each residue was the same as the isolated amino acid and was not affected by the protein structure..... 143

Figure 4.14. Relative hydrophobicity of the I-tag and the S-tag sequences. The sequences of each domain used for the estimation were indicated in Supporting Information Figure 4.13A. The relative hydrophobicity was calculated using a method proposed by Trabbic-Carlson et al.[19] This estimation was based on the hydrophobicity scale established by Urry et al.[12] that was originally used to predict the LCST of an ELP with a known guest residue composition. More hydrophobic guest residue compositions resulted in lower LCST values. It should be noted that this estimation had no direct physical meaning and was only used to demonstrate the difference in the relative hydrophobicity between the tags in an ELP-fusion protein context. 144

Figure A. 1. Solubility of HZE-QEQV-RGDGC (top) and HZE-RGDGC (bottom) changes with temperature. Proteins were dissolved in PBS at 30 mg/mL and heated to temperature indicated in the column headings..... 155

Figure A. 2. Oligomerization of RZY10-HZR and HZE-RGDGC. (-) and (+) indicate sample with and without crosslinking, respectively. MW of proteins: RZY10-HZR: 20.2 kDa, HZE_RGDGC: 10.2 kDa, homodimer of RZY10-HZR: 40.4 kDa, homodimer of HZE-RGDGC: 20.4 kDa, heterodimer of RZY10-HZR and HZE-RGDGC: 30.4 kDa. 156

Figure A. 3. Oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC) verified by luminescent anti-T7 tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa..... 158

Figure A. 4. Expression of RZY10-HZR (MW 20.2 kDa). Top: SDS-PAGE gel showing RZY10-HZR being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-T7 tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel. 159

Figure A. 5. Expression of HZE-RGDGC (MW 10.2 kDa). Top: SDS-PAGE gel showing HZE-RGDGC being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by luminescent anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel. 160

Figure A. 6. SDS-PAGE gel of purification of RZY10-HZR in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~30 mg/L culture. FT: flow though; G1 and G2: wash with 6 M guanidine hydrochloride; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9... 161

Figure A. 7. SDS-PAGE gel of purification of HZE-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. CL: cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9..... 162

Figure A. 8. Expression of RZY10-COMPm (MW 20.2 kDa). Top: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-T7 tag Western blot. Bottom: SDS-PAGE gel of purification of RZY10-COMPm in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~30 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9..... 163

Figure A. 9. Expression of COMPm-RGDGC (MW 10.2 kDa). Top: SDS-PAGE gel showing COMPm-RGDGC being expressed for 3 and 5 hours after induction of IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel..... 164

Figure A. 10. SDS-PAGE gel of purification of COMPm-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9..... 165

Figure A. 11. Oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC). Top: Coomassie Blue stained SS-PAGE gel. Bottom: Luminescent anti-His tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa. 166

Figure A. 12. Visualization of RZY10-COMPm and COMPm-RGDGC with different antibodies in the Western blot to show oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC). Luminescent anti-His tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa. 167

Figure A. 13. Expression of LZE-BMP2GC (MW 10.2 kDa). Top: SDS-PAGE gel showing LZ-BMP2GC being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts by colorimetric anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel. 168

Figure A. 14. SDS-PAGE gel of purification of LZE-BMP2GC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. FT: flow though; C1 and C2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9. 169

Figure A. 15. Expression of HZE-QEQV-RGDGC (MW 11.6 kDa). Top: SDS-PAGE gel showing HZE-QEQV-RGDGC being expressed for 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts by anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel. 170

Figure A. 16. SDS-PAGE gel of purification of HZE-QEQV-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was about 50 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9. 171

Figure B. 1. RZY10 and tyrosinase-converted RZY10 (mRZY10) on SDS-PAGE gel. mRZY10 showed a lower electrophoretic mobility that could be due to the increase in molecular weight (MW) and/or the decrease in hydrophobicity. 176

Figure B. 2. RZY10 or mRZY10 conjugation at pH 7.4. RZY10 or mRZY10 in 100 mM acetate buffer with 100 mM ascorbic acid were pH adjusted to 7.4 and were incubated at room temperature for 24 hours. Left: RZY10 appeared as a single band at the expected position on the SDS-PAGE gel. On the other hand, mRZY10 appeared as an upward smear. Right: Anti-T7 tag Western blot better visualized the smear in mRZY10. 176

Figure B. 3. No tag-mRZY10 and RZY10 conjugation at pH 7.4. Proteins were in a cobuffer of 10 mM acetate buffer and 10 mM Tris buffer with 150 mM NaCl, 10 mM $CaCl_2$, and 10 mM ascorbic acid and were incubated at room temperature for 24 hours. Top: SDS-PAGE gel showing high MW products from reaction of no tag-mRZY10 (12.1 kDa) and RZY10 (14.1 kDa). Bottom: Anti-T7 tag Western blot image revealed that the high MW products contained T7 tags from RZY10. Arrows indicate numbers of RZY10 being conjugated to no tag-mRZY10. The band below 25 kDa ladder was single RZY10. 178

Figure B. 4. Conjugation between acRZY10 and mSKGPG-Y2 through DOPA-Histidine bonds was not observed. No changes in the smearing pattern or shifts in MW of the band near 25 kDa ladder were observed at pH 7.4 where the DOPA-Histidine conjugation was expected to occur. At pH 9.5, DOPA-Lysine conjugation was expected to occur and result in intrachain crosslinking of mSKGPG-Y2. The expected result of such conjugation would be a decrease in the intensity of the free mSKGPG-Y2 band around 25 kDa. However, such a decrease was not observed. 181

Figure B. 5. Ascorbic acid prevents the oxidation of DOPA. Less smearing was observed with mSKGPG-Y2 when no ascorbic acid was used in the buffer. This result is consistent with the fact that ascorbic acid decreases the reactivity of quinone towards nucleophiles by preventing the oxidation of DOPA to quinone. In addition, no difference was observed in the absence of ascorbic acid between the group with and without the His-containing peptide. The groups of pH 7.4 10x indicates that the peptide concentration used was 10 times higher then the other groups in the figure. 182

Figure C. 1.The expression of YC12-T in Rosetta2 pLysS was confirmed by anti-His tag Western blot. The expected MW of YC12-T was 43.7 kDa..... 185

Figure C. 2. The purity of the lyophilized protein was >95% for YC12, YC12-M, YC12-D, and YC12-T as determined by densitometry on a Coomassie blue-stained SDS-PAGE gel. An anti-His tag Western blot confirmed the presence of the His tag on the purified proteins..... 186

Figure C. 3. YC12, YC12-M, and YC12-T had similar coated amounts at pH 6.3. To spin coat, 15 μL of protein solution in MQ at 5 mg/mL was mixed with 3.75 μL of 15 wt% H_2O_2 (all pH adjusted to the desired pH) and then dispensed onto the center of an acid-washed coverslip. The coverslip was spun for 60 sec at 4000 rpm and incubated at 37 °C for 30 min. After incubation, the coverslip was sterilized in 70% ethanol for 30 min at room temperature. The amount of coated protein was quantified using a BCA kit. Different letters indicate Tukey groups that are statistically different ($p < 0.05$). 187

Figure C. 4. Coated YC12 proteins were stable in PBS for 7 days. Protein-coated coverslips were incubated in PBS at 37 °C without shaking. At designated time points, the protein remaining on the coverslip was quantified by using a BCA assay. 188

Figure C. 5. Similar number of cells attached on YC12, YC12-M, and YC12-T surfaces. Fischer 344 rat bone MSCs (rBMSCs) and NIH/3T3 mouse fibroblasts were seeded at 3000 cells/cm^2 and counted by nuclei staining 5 h after seeding (n =3)..... 189

Figure C. 6. Similar cell morphology and area of rBMSCs on surfaces with different BMP-2 peptide presentations. Top: Actin (green) and nucleus (red) staining of rBMSCs on protein surfaces 15 h after seeding. Bottom: Cell area quantification based on actin staining (n > 100 cells). Different letters indicate Tukey groups that are statistically different ($p < 0.05$). 190

Figure C. 7. Alkaline phosphatase (ALP) activity increased from day 4 to day 7 on all surfaces. Cells on the tissue culture poly(styrene) (TCPS) surface had the highest ALP activity. Within protein surfaces, YC12 resulted in the highest ALP activity followed by YC12-M then YC12-T. 192

Figure C. 8. More calcium deposition stained by Alizarin Red S (ARS) was observed on the YC12 surfaces than the YC12-M and YC12-T surfaces. Top: Staining images of 24 well plates on day 9 and day 13. Bottom: 4x images at the center of the wells. The trend of ARS staining on day 13 is consistent with the ALP activity results. Together, these results suggested that YC12 promoted a higher extent of osteogenesis..... 193

Figure C. 9. ARS quantification by acid extraction. The result showed a similar trend to what was observed with the staining images in Figure C 8. 194

Figure C. 10. Purification of YC12-T by Ni-IPA column. CL: cleared lysate; Pel: pellet of cell debris after sonciation; FT: flow through; C: wash of buffer C (8 M urea, 10 mM Tris, 100 mM NaH_2PO_4-H_2O, 300 mM NaCl, 10 mM BME, pH 6.3); D: wash of buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME, pH 8.0); IPA/D: wash of IPA-buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME in 60 (v/v)% isopropyl alcohol, pH 8.0); E: elute of buffer E (10 mM Tris, 500 mM imidazole, 300 mM NaCl, pH 8.0). 195

Figure C. 11. Representative images of gel clot formation in LAL assay. 196

Figure C. 12. Images of (A) cell-encapsulated YC12-T/PEG-MAL hydrogel and (B) cell-laden collagen sponge. 198

Figure C. 13. Images of *in vivo* CT at 6 weeks after implantation of the scaffolds. Empty: the group of empty defects; collagen: the group of cell-laden collagen sponges; protein: the group of cell-encapsulated YC12-T/PEG-MAL hydrogels. 199

Figure C. 14. Images of *ex vivo* μCT at 12 weeks after implantation of the scaffolds. Empty: the group of empty defects; collagen: the group of cell-laden collagen sponges; protein: the group of cell-encapsulated YC12-T/PEG-MAL hydrogels. Images are in the same order as in Figure C 13. Images from rats that did not survive to the 12^{th} week are not shown. . 199

Figure C. 15. Haemotoxylin and Eosin staining images of a defect implanted with a protein hydrogel (rat 30). Bone is stained as dark pink. Black arrowheads mark the edge of the defect. 200

Figure C. 16. Haemotoxylin and Eosin staining images of a defect implanted with a collagen sponge (rat 23). Bone is stained as dark pink. Black arrowheads mark the edge of the defect. 201

Figure C. 17. Low ALP activity as assessed by staining rat bone MSCs after 3 weeks of culture in osteogenic medium (ODM). Top: whole well image of cells cultured in growth medium (GM) and in ODM. Bottom: 4x microscope images of the center part of the well. 203

Figure C. 18. Significant calcium deposition was not observed with rat bone MSCs. Two different sources of ARS dyes were used. The ODM group showed slightly more calcium deposition than the GM group, but the overall staining intensity of the ODM group did not suggest strong osteogenesis of rat bone MSCs after culture in ODM for 3 weeks. 204

Figure C. 19. Protein purification using ammonium sulfate salting out with YC12 proteins. Addition of ammonium sulfate at 10 wt% followed by 20 wt% worked well for YC12, YC12-M, and YC12-D. YC12-T partially salted out at 10 wt% (red box). CL: cleared lysate; SF: soluble fraction; Pel: cell pellet; 10S and 10P: 10 wt% supernatant and pellet; 20S and 20P: 20 wt% supernatant and pellet. 205

Figure C. 20. Protein purification using inverse temperature cycling (ITC) with YC12. ITC was carried out in PBS with 200 mM BME. Cold cycles were performed at a concentration of 50 mg pellet/mL at 4 °C. Hot cycles were performed at a concentration of 50 mg pellet/mL, coacervation was induced by 2 M NaCl at RT, and protein solutions were centrifuged at 30 °C. CL: cleared lysate; Pel: cell pellet; 10S and 10P: 10 wt% supernatant and pellet; 20S and 20P: 20 wt% supernatant and pellet; CP and CS: cold cycle pellet and supernatant; HP and HS: hot cycle pellet and supernatant. 206

Figure C. 21. Protein purification using inverse temperature cycling (ITC) with YC12-M, -D, and -T. ITC was carried out in PBS with 2 M urea and 200 mM BME. A final concentration of 2 M urea in PBS aided in resolubilizing protein pellets. Vigorous mixing was necessary to solubilize YC12-T. Cold cycles were performed at a concentration of 50 mg pellet/mL at 4 °C. Hot cycles were performed at a concentration of 50 mg pellet/mL, coacervation was induced by 2 M NaCl at RT, and protein solutions were centrifuged at 30 °C. CP and CS: cold cycle pellet and supernatant; HP and HS: hot cycle pellet and supernatant. 207

Figure C. 22. Endotoxin purification by Ni column with Triton X114 wash (Ni-114 column). Compared to Ni-IPA method, the residual endotoxin level was higher with Ni-114 column. The protein-loaded column was washed with 50 times of the bed volume with Triton X-114 wash buffer (100 mM KCl, 200 NaCl, 10 mM imidazole, 20 mM Tris-HCl, and 0.1 (v/v)% Triton X-114, pH 8.0). Next, the column was washed with 20 times of the bed volume with wash buffer to remove residual Triton X-114. The column was washed with 5 times of the bed volume with wash buffer containing 25-50 mM imidazole. Finally, the column was eluted with wash buffer containing 500 mM imidazole and 1 mM TCEP. 208

Figure C. 23. Gel clot LAL assay of Ni-114 column purified YC12-T. The endotoxin level was between 12.5-125 EU/mg protein which was at least 100-1000 times higher than the same protein purified by the Ni-IPA method. A solid gel was not observed at 1:100 dilution possibly because of the interference from the residual Triton-X114 in the purified protein. 209

Figure C. 24. Gel clot LAL assay control groups. Top: Testing whether TCEP negatively interacted with the assay. TCEP between 0.05-0.5 mM did not significantly interfer with the assay. Bottom: Test result of YC12-T purified by ITC method. The endotoxin content was >1250 EU/mg protein. 210

Figure C. 25. Representative scanning images of *in vivo* CT at 6 weeks after implantation. Empty defects showed the most healing followed by defects with cell-laden collagen sponges. Defects with cell-encapsulated protein gels showed the least extent of healing. 211

Figure C. 26. Scanning resolution comparison between *in vivo* CT and *ex vivo* μCT. Top: defect with cell-laden collagen sponge at 6^{th} week; bottom: defect with cell-encapsulated protein gel at 6^{th} week. *Ex vivo* μCT had a much higher resolution on resolving small volume of new bone formation. Both Rat 14 and 28 died after *in vivo* CT at 6^{th} week. Defect sites were harvested immediately and scanned by *ex vivo* MicroCT..... 212

Figure D. 1 PAGE gel of PPM-17 with 57, m57, and negative control A1 DNA oligonucleotides. Numbers above the image indicate the molar ratio of PPM-17:DNA.220

Figure D. 2. HEK293T cells transfected with constructs based on BiFC pair VN173/VC155. Cells were imaged 24 h after transfection. The green signal indicates complementation from the BiFC domains on the constructs. R20-mCherry (Addgene #49637) was used as a positive control for telomere binding (red signal). 223

Figure D. 3. HEK293T cells transfected with constructs based on BiFC pair VN210/VC210 with the short linker. Cells were imaged 24 h after transfection. 225

Figure D. 4. HEK293T cells transfected with constructs based on BiFC pair VN210/VC210 with the long linker. Cells were imaged 24 h after transfection. R20-mCherry was used as a positive control for telomere binding (red signal)..... 226

Figure D. 5. Expression of PPM-13 (MW 78.5 kDa). Top: SDS-PAGE gel showing PPM-13 being expressed 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts except for BL21(DE3)pLysS by anti-His tag Western blot. 227

Figure D. 6. Expression of PPM-13 (MW 78.5 kDa) and PPM-17 (MW 95.6 kDa) in BL21(DE3). Top: SDS-PAGE gel showing PPM-17 being expressed 5 hours after induction with IPTG. Bottom: The expression was confirmed in expression *E. coli* host BL21(DE3) by anti-His tag Western blot. 228

Figure D. 7. PPM-17 purification. Top: Fractionation of cell lysate based on solubility. Red arrow indicates PPM-17 band in soluble fraction (SF). For detailed purification steps, refer to Appendix E. 6. Bottom: SF of PPM-17 purified by a native Ni column. The purity of E300 was found to be 91% by densitometry. SF: soluble fraction; SP: supernatant of Triton wash; FT: flow through; W1 and W2: wash with 500 mM NaCl, 5 mM imidazole, 20 mM Tris-Cl, pH 7.9; HW: wash with 2M NaCl, 5 mM imidazole, 20 mM Tris-Cl, pH 7.9. E50 to E500-2: Elutes with 500 mM NaCl, 50-500 mM imidazole, 20 mM Tris-Cl, pH 7.9. 229

Figure D. 8. HpaII digestion of 57 and m57. The methylation level of m57 was determined to be 93% by densitometry. 230

Figure D. 9. HpaII digestion of pJB-91 to check methylation level. Methylation of pJB-91 was accomplished by co-transformation with pAIT2, which encodes methyl transferase M.SssI. Similar methylation was achieved among three clones with pJB-91 and pAIT2. 231

Figure D. 10. Illustration of methylation detection of the one-piece probe (PPM-17)... 232

Figure D. 11. Digestion of HRV3C control protein with HRV3C in HRV3C buffer and binding buffer..... 233

Figure D. 12. Binding assay of PPM-17 in HRV3C buffer with 50 ng/μL dIdC polymer. 234

Figure D. 13. HRV3C digestion of PPM-17 with different HRV3C:PPM-17 ratios. PPM-17 arrow indicates undigested PPM-17; TALE arrow indicates TALE domain from digested PPM-17; HRV3C arrow indicates HRV3C protease. 235

Figure D. 14. PPM-17 digestion efficiency assessed by densitometry. Corresponding SDS-PAGE gels can be found in Figure D. 13. 236

Figure D. 15. HRV3C digestion of PPM-17 at different temperatures. PPM-17 arrow indicates undigested PPM-17; TALE arrow indicates TALE domain from digested PPM-17; HRV3C arrow indicates HRV3C protease. 237

Figure D. 16. PPM-17 digestion efficiency assessed by densitometry. Corresponding SDS-PAGE gels can be found in Figure D. 15. 238

Figure D. 17. PPM-17 digestion by HRV3C at room temperature and 16 °C (top) and PPM-17 digestion efficiency assessed by densitometry (bottom). 239

Figure F. 1. Cloning scheme of pET21Q-LZE-BMP2GC 255

Figure F. 2. Cloning scheme of pET21Q-COMP-RGDGC 255

Figure F. 3. Cloning scheme of pUC57-HZE-BMP2GC 256

Figure F. 4. Cloning scheme of pET21Q-HZE-BMP2GC 256

Figure F. 5. Cloning scheme of pUC57-COMP-BMP2GC 257

Figure F. 6. Cloning scheme of pET21Q-COMP-BMP2GC 257

Figure F. 7. Cloning scheme of pJB-RZY10 258

Figure F. 8. Cloning scheme of pJB-RZY10-LZR 258

Figure F. 9. Cloning of pET28aCYL-LZR 259

Figure F. 10. Cloning scheme of pJB-RZY10-COMP 259

Figure F. 11. Cloning scheme of pET28aRW-RZY10-COMP 260

Figure F. 12. Cloning scheme of pJB-HZR 260

Figure F. 13. Cloning scheme of pET21b-RZY10-HZR 261

Figure F. 14. Cloning scheme of pJB-YC12 with recursive directional ligation strategy. This scheme can be applied to other pJB constructs as long as the sequence of interest (e.g., YC8) is flanked by compatible ends that are not compatible with ScaI. 261

Figure F. 15. Cloning scheme of pJB-YC12-M. 262

Figure F. 16. Cloning scheme of pET21Q-SKGPG-YC12-M. 262

Figure F. 17. Cloning scheme of pET21Q-SKGPG-YC12-RGD. 263

ABSTRACT

Author: Lin, Charng-Yu. PhD
Institution: Purdue University
Degree Received: May 2019
Title: Toward Rational Design of Functional Materials for Biological Applications
Committee Chair: Julie C. Liu

Cellular activities are composite responses to stimuli from the surroundings. Materials for biological applications, therefore, must be designed with care such that undesired interactions between cells and the materials will not be elicited while cellular responses that are beneficial to the dedicated applications are promoted. Efforts have been made to construct such materials based on both synthetic polymers and natural polymers including poly(ethylene glycol) (PEG) and proteins. In particular, recombinant proteins have drawn great interest for their similar biocompatibility to natural proteins and the uniformity of material properties that is found in manufacturing of synthetic polymers. Recombinant proteins are designed at the DNA level, which allows precise control over the translated protein sequence. By assembling encoded DNA sequences of amino acids with desired functional groups or protein domains conferring desired functionalities, a recombinant protein-based material can be tailored. In this dissertation, works toward developing functional biomaterials based on both synthetic polymers and recombinant proteins are presented.

The first part of this thesis encompasses the development of a new thiol-based crosslinking approach to achieve independent control over degradability and mechanical properties of a hydrogel system. Thiol chemistry was chosen as the focus here because it can easily be incorporated into recombinant protein designs by inserting cysteine residues. In addition, the low frequency of cysteine residues in natural proteins can reduce unwanted reactions between the hydrogel material and encapsulated biomolecules or cells. We utilized divinyl sulfone (DVS) to form thioether crosslinking through thiol-ene addition and ferric ethylenediaminetetraacetic acid (ferric EDTA) to make disulfide crosslinking via thiol oxidation. By controlling the ratio between the non-reducible thioether bonds to reducible disulfide bonds, hydrogels with similar mechanical properties can be made with

different degradability in reducing conditions. Accelerated degradation and increased release of encapsulated dextran was observed in response to an extracellular reducing condition. Good viability of encapsulated fibroblasts also suggested high cytocompatibility of the crosslinking approach. This work demonstrated the potential of thiol crosslinking by DVS and ferric EDTA for making redox-responsive drug delivery vehicles and tissue engineering scaffolds.

In the second part, we developed protein adhesives using thiol- or catechol-based adhesion. Every year more than 310 million surgeries are performed around the world, and more than 50% of these surgeries used sutures or staples for wound closure.[1-2] Surgical sealants or adhesive can be applied together with sutures and staples to mitigate the risk of infection. Protein-based adhesives could have better biocompatibility than synthetic polymer-based adhesives and have the potential of providing biochemical cues for cellular responses. Many adhesive proteins have been found in nature. Among them, mussel adhesive proteins have been actively studied for their outstanding underwater adhesion. The capability of being able to cure in a wet environment is critical for an ideal surgical sealant and adhesive. Mussels uses both thiols and a catechol, 3, 4-dihydroxyphenylalanine (DOPA), to achieve underwater adhesion. Inspired by mussel adhesive proteins and modular recombinant design, we developed two proteins harboring thiol or DOPA groups with highly similar amino acid sequences. The adhesion performance, including curing kinetics, adhesion strength, and cytocompatibility, were compared between the two proteins. The similarity in the protein sequences allows us to focus on the performance difference between thiol- and DOPA-based adhesion. We also showed that a synergistic increase in the adhesion strength can be achieved when the two proteins are combined. This increase indicates a cross-reaction between thiol and DOPA groups. Our results provide insights into selecting the chemistry for designing adhesives based on the needs of the applications.

In the last part, we studied the lower critical solution temperature (LCST) behavior of elastin-like polypeptides (ELPs) with a series of ELPs with rationally designed charge distributions and chain lengths. The LCST behavior of ELPs are controlled by multiple factors including the amino acid composition, ELP chain length, protein concentration, salt identity, salt concentration, and pH of the solution. Fusion of other non-ELP recombinant

protein domains to ELPs have also been shown to influence the LCST behavior of the fusion ELP protein. Inspired by this effect, we explored the use of short non-ELP sequences as a new way to tailor the LCST behavior of ELP-based proteins. We designed the non-ELP and the ELP sequences with different pH-dependent charge states and showed that pH sensitivity was introduced to the LCST behavior by electrostatic and hydrophobic interactions between the non-ELP and ELP sequences. The electrostatic interactions can be shielded by the ionic strength in the protein solution. The pH sensitivity was introduced by the non-ELP sequences, and this sensitivity decreased when the relative length of the ELP domain increased. We also found that the hydrophobicity of the non-ELP sequences changes the interactions between the proteins and Hofmeister ions in solution. Our results demonstrated the potential of using non-ELP sequences as a new factor in controlling the LCST behavior of ELP proteins.

References:

(1) Bré, L. P.; Zheng, Y.; Pêgo, A. P.; Wang, W. Taking tissue adhesives to the future: from traditional synthetic to new biomimetic approaches. *Biomater. Sci.* **2013,** *1* (3), 239-253.

(2) Weiser, T. G.; Haynes, A. B.; Molina, G.; Lipsitz, S. R.; Esquivel, M. M.; Uribe-Leitz, T.; Fu, R.; Azad, T.; Chao, T. E.; Berry, W. R. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet* **2015,** *385*, S11.

1. INTRODUCTION

Functional materials can be defined as materials possessing capabilities other than bearing loads because of their chemical compositions or physical structures.[1] Functional materials that show good biocompatibility and interact with surrounding cells can be regarded as functional biomaterials.[2] The interactions can be as simple as being recognized by cells as a attaching substrate or a degradable matrix, or the interactions can be a composite cellular response to the material such as cell migration, cell proliferation, or cell differentiation. The origin of a material, however, does not define the material being biomaterials. Both synthetic polymers and natural materials have been designed toward functional biomaterials and used in biological applications including tissue engineering, drug delivery, and tissue adhesives.

1.1 Tissue Engineering

Damaged tissues due to diseases or injuries needs to be replaced to regain the proper functions of the tissues. Ideally, such replacement is completed by transplantation of a healthy and functional tissue from a compatible new source. However, autologous tissue transplantation introduces additional damages to the harvesting site and can be impractical due to the limited availability of the healthy tissue from the patient. Replacement from a different individual of the same species (i.e., allogeneic transplantation) or from a different species (i.e., xenogeneic transplantation) has the risk of transmitting diseases from the donor to the recipient or across species and requires control over immune responses after transplantation. In the past decade, the number of donors per year had grown from 14,207 by 16% to 16,473; however, the number of patients per year had also grown from 100,775 by 15% to 115,759 in the U.S. alone, and such unmet demand of donors is also observed worldwide.[3-4]

To circumvent the shortage of tissue donors, tissue engineering has drawn increasing interests as an approach to regenerate tissues. The ultimate goal of tissue engineering in tissue regeneration is to recreate a healthy and functional tissue with minimal autologous cells with the aid of engineered biomaterials that mimic the

extracellular matrix (ECM) of the target tissue by presenting appropriate bioactive signals. Tissue engineering is truly an interdisciplinary field where material science, biology, medicine, and an engineering mindset meet and work together toward the goal.

1.1.1 Three Components in Tissue Engineering

There are three components that are essential in all kinds of tissue engineering: cells, bioactive signals, and scaffolds. Cells are used to produce proper ECM in the engineered scaffold and to confer biological function to the engineered tissue. Ideally, autologous differentiated cells will be used as the cell source. However, harvesting enough healthy cells from a damaged-tissue can be impractical, and these cells often cannot be expand in large quantities *in vitro*.[5] Pluripotent or multipotent stem cells can be good alternatives autologous differentiated cells because of there ability to differentiate into desired cell types. Mesenchymal stem cells (MSCs) harvested from bone marrow have been extensively studied for their multipotency. With the appropriate stimuli, bone marrow MSCs have been shown to undergo osteogenic, chondrogenic, adipogenic, neurogenic, and myogenic differentiations.[6-7] Other than the multipotency, bone marrow MSCs also can be easily harvested and expanded to a certain extent *in vitro*.[8] In 2006, Takahashi and Yamanaka reported for the first time the induced pluripotent stem cells (iPSCs) from mouse embryonic and adult fibroblast cells by reactivating four embryonic state-related transcription factors, Oct4, Sox2, c-Myc, and Klf4, through viral transfection.[9] They later showed that adult human fibroblasts can be reprogrammed into iPSCs using the same approach.[10] Compared to MSCs, iPSCs have pluripotency similar to that of embryonic stem cells. In addition, the source cells for iPSCs (e.g., fibroblasts) can be harvested much easier than MSCs and embryonic stem cells. iPSCs also do not raise as many ethical concerns as embryonic stem cells. Progresses have been made to produce iPSCs in viral-free and more efficient ways through small molecules, recombinant proteins, or microRNAs.[11-13] Despite its great potential as the new cell source in tissue engineering, a major concern about iPSCs is the genomic instability.[14] Efforts are being made to identify and correct such instabilities, and new methods of efficiently producing safer, more defined iPSCs are being developed. It could be expected that iPSCs will be the new alternative cell source to autologous cells and widen the clinical applications of tissue engineering.

In order to direct differentiation of MSCs or iPSCs into desired cell types or to maintain proper cell functionals of autologous differentiated cells, it is necessary to present appropriate biological signals in the engineered scaffold in tissue engineering. These biological signals can be categorized into four types: biochemical stimuli, biophysical stimuli, ECM cues, and cell-cell interactions.[15] Both biochemical and biophysical stimuli have been widely utilized in tissue engineering to guide cell behaviors. Biochemical stimuli are often growth factors or small molecules that promote certain cellular activities. These molecules can be presented either in the culture media or on the engineered scaffold. For example, the peptide Arg-Gly-Asp (RGD) is well known for its function as a cell binding site. The growth factor bone morphogenetic proteins (BMPs) has been shown to promote bone differentiation of MSCs.[16-17] Transforming growth factor beta (TGF-β) has been used in cartilage tissue engineering to promote chondrogenesis of MSCs.[18] Several growth factors have been identified participating in angiogenesis including vascular endothelial growth factor (VEGF), basic fibroblast growth factor, platelet-derived growth factor, and angiopoietins.[19] Short peptides derived from whole growth factors have also been reported to have similar efficacy in guiding stem cell differentiations. Examples include the BMP-2 peptide derived from the knuckle epitope of whole BMP-2 and the QK peptide derived from VEGF.[20-21] Some small molecules have also been shown to promote stem cell differentiation. Dexamethasone, for instance, is widely used in the osteogenic media for MSCs.

Biophysical signals include the stiffness of the engineered scaffold and the force being applied during the culture. Scaffold stiffness is the key factor when designing the scaffold. MSC differentiation has been shown to be guided by the stiffness of the scaffold both in 2D and 3D.[7, 22] Although most *in vitro* cell cultures are performed in a static way, cells do sense fluid shear stress *in vivo* such as the blood flow in the blood vessels. With proper culture setup, it is possible to induce fluid shear stress during the culture period. Fluid shear stress in culture have been shown to promote osteogenic and cardiomyogenic differentiation.[23-24] Biophysical signals can also act synergistically with biochemical signals. Zouani *et al.* reported that the BMP-2 peptide had no effect on MSCs when being presented on a scaffold with stiffness ranging from 0.5-3.5 kPa.[25] However, increased

efficacy from the BMP-2 peptide was observed when the scaffold stiffness was between 45-49 kPa.

Scaffolds provide the microenvironment mimicking the native ECM for cells. Design factors of an engineered scaffold include mechanical properties that matches the targeting or surrounding tissues, a degradation behavior that matches the production of the new ECM by cells, and the capability to provide proper biological stimuli for cell growth and differentiation.[26-28] Both synthetic and natural materials have been extensively explored for tissue engineering scaffolds and other biological applications. Synthetic materials have the potential to be functionalized though various chemical approaches that are biorthogonal. However, the lack of biocompatibility or interactions with cells also hampers the direct application of synthetic materials. Natural materials, on the other hand, are usually biocompatible and have intrinsic motifs that can interact with cells. Still, natural materials have the drawbacks of limited approaches for further functionalization and difficulties in large-scale production and processing.

1.2 Drug Delivery

In the field of drug delivery, another design criterion of the material besides biocompatibility is the capability for a spatially and/or temporally controlled release of the drug.[29] A controlled drug release can increase the treatment efficiency and efficacy and can reduce the risk of side effects. The controlled release is often achieved by tailoring the material properties. For example, the degradation rate of poly(lactic-co-glycolic acid) (PLGA) can be controlled by altering the ratio between the two repeating segments, poly(lactic acid) (PLA) and poly(glycolic acid) (PGA). This difference in the degradation of the material then reflects on the release of the drugs encapsulated in PLGA vehicles.[30] For controlled release at targeted sites, a common strategy used is to functionalize the vehicle material with ligands and/or receptors that are specific to the targeting site.[31] These are, however, passive approaches that cannot be activated or tuned on demand after the vehicles are in the body.

In order to achieve more flexible control on the drug release, smart biomaterials have drawn greater interests in the field of drug delivery. Smart biomaterials have the properties that respond to changes in environmental stimuli.[32] Smart materials that are

responsive to changes in temperature, pH, or redox potential have been investigated for targeted drug delivery. Poly(N-isopropylacrylamide) (PNIPAM) is one example of temperature-sensitive synthetic polymers. PNIPAM has a lower critical solution temperature around 37 °C at which it undergoes a sol-gel transition.[33] A drug-loaded PNIPAN hydrogel will shrink drastically in volume above its LCST and release the encapsulated drugs. Variants of PNIPAM have also been developed with different LCSTs that can utilized for targeted drug release. Recombinant proteins elastin-like polypeptides (ELPs) are another example of temperature-responsive materials that have been explored for controlled drug delivery. The LCST of an ELP can be easily tuned to be in the range of mild clinical hyperthermia (40-44 °C) by changing the amino acids at the guest residue positions.[34-35] Above the LCST, ELPs undergo hydrophobic collapses and aggregate into coacervates. This coacervation allows drug-loaded ELP complexes accumulate at the targeted site. The increase in the hydrophobicity also promotes endocytosis which increase the treatment efficiency. Taking the advantage of modular recombinant protein design, ELPs can also be fused with cell penetrating peptides to further increase the drug delivery efficiency into cells.[36]

It has been known that there is a pH difference between healthy tissues and cancerous tissues with cancerous tissue being slightly acidic.[37] Organelles such as endosomes and lysomes also have lower pHs than other cellular compartments. Chen *et al.* developed a poly(ethylene glycol) (PEG) variant functionalized with acid-labile carbonate group for target delivery of anti-cancer drugs.[38] The accelerated degradation of the PEG micelle resulted in a faster release of the drug in acid environments. Callahan and coworkers designed a pH-sensitive ELP such that the formation of coacervation is disrupted with decreased pH.[39] They showed that the pH-sensitive ELP can penetrate deeper into the tumor tissue due to the decreasing pH gradient from the surface to the center of the tumor site.

Other than the pH difference, the redox condition between a normal tissue and a tumor tissue is also different. For example, the concentration of the biological-relevant reducing agent glutathione (GSH) is generally four times higher in a cancerous tissue than in a healthy tissue.[40] With the progress in developing new synthetic or natural materials,

smart materials that utilize the redox difference to achieve target drug delivery and controlled release have a growing interest.[41-42]

With the progress in functional biomaterials, the development of smart materials responding to multiple stimuli can be expected.[43] New stimuli such as magnetic field are also being explored to achieve targeted and controlled drug release.[44]

1.3 Tissue Adhesives

More than one million tissue transplants and about 310 million other surgeries were performed in 2017.[45-46] Sutures and staples are usually used for wound closure in many of these surgeries. Unfortunately, the use of sutures and staples inevitably damage the surrounding healthy tissue near the wound and increase the risk of infection.[47] Surgical sealants and tissue adhesives can be applied together with sutures and staples to lower the risk of infection but also have the potential to be used alone for wound closure.[48-49]

An ideal material for surgical adhesive has similar criteria as for tissue engineering scaffold. The material needs to have good biocompatibility and have similar mechanical properties to the surrounding tissue. In addition, the adhesive and cohesive strength from the material also needs to match the surrounding tissue.[50] Most importantly, the material has to have adhesion in a wet environment. There are FDA-approved surgical adhesives, but none of them meets all criteria. FDA-approved cyanoacrylate-based adhesive such as SurgiSeal[51] and Dermabond[52] can set under wet conditions and provide closure strength similar to that of sutures.[53-54] Their brittle nature also reframe the use in areas of tissue that experience tensile stress. In addition, both cyanoacrylate and its degradation products are toxic not only to the patient but also to the surgeons.[55] As the result, the current approve of cyanoacrylate-based adhesives is limited to topical applications only.[56]

Fibrin-based surgical sealants (e.g., Tisseel and Crosseal) have good biocompatibility and degradability and therefore have been approved as hemostatic agents for internal applications.[57] Their relatively low adhesion strength, however, prohibit the applications tissue adhesives for wound closure.[58] The risk of transmitting bloodborne pathogens is also a concern in the production of fibrin-based sealants. The risk of transmitting hepatitis and HIV is about 1 in 10^{15} packages and is about 1 in 10^{7} packages

for parvovirus.[59] The manufactures are still improving the process to lower the risk of such transmissions.

PEG-based sealants can be biocompatible with tunable adhesion strength and fast curing.[60] Different curing mechanisms can be employed due to the synthetic nature of PEG. Coseal utilizes two different PEGs functionalized with glutaryl-succinimidyl esters or thiols. Upon mixing of the two PEGs, the two groups react and form the sealant network. Coseal is approved for sealing vascular grafts. Another product, DuraSeal, uses PEG ester and trilysine amine and is approved for prevent cerebrospinal fluid leakage. PEG-based sealants require, however, a relative dry surface for application in order to get optimal performance. Due to the hydrophilic nature of PEG, PEG-based sealants can swell up to four times by volume.[61] As the result, cautions must be exerted when applying PEG-based sealants in a confined space to avoid pressing pressure to the surrounding tissue.

A recently approved tissue adhesive, TissuGlu, utilizes urethane chemistry to achieve adhesion in a wet environment. TissuGlu is currently approved in abdominoplasties to reduce fluid accumulation between tissue layers in the abdominal region. The prepolymer with end groups of isocyanate reacts with amines groups in the formulation or in the surrounding tissue and with hydroxyl groups in water. The degradation products of TissuGlu does not show acute and chronical toxicities.[62] However, the cure time is on the order of tens of minutes, and the patient is often advised to stay still to allow optimal recovery.[63]

Despite the recent advances in developing materials for surgical sealants and tissue adhesives, ideal materials matching the requirements of being biocompatible, being capable of wet adhesion, and have comparable mechanical properties to tissue have not emerged yet. There are, however, new materials being designed mimicking the adhesion mechanisms found in nature and could bridge the gap between the state-of-the-art and the ideal materials for surgical adhesives.[64]

1.4 Functional Biomaterials

Both synthetic and natural polymers have been used in designing functional biomaterials. The choice between the two is usually made depending on the intended application. In general, synthetic polymers have the advantage of accessing various chemistries for

functionalization and being ready for mass production. However, synthetic polymers with functionalization are usually also lack of material-cell interactions. Natural polymers, on the other hand, often have good interactions with cells, but the functionalization processes are usually limited within a certain range of conditions (e.g., pH, temperature). Natural polymers are also more difficult to be produced in large quantities and have lower consistency in material quality from batch-to-batch. Recombinant protein-based materials are a special category of natural polymers. Due to their protein nature, material-cell interactions are usually expected and can be further promoted by incorporating specific cell-recognition sequences into the material design. In addition, the protein translation machinery ensures minimal batch-to-batch variation in material qualities (e.g., chain length and chain composition). In the following sections, we discuss the current progress of developing functional biomaterials using synthetic polymers, natural polymers, and recombinant protein-based materials.

1.4.1 Synthetic Polymers

PEG is perhaps the most widely used synthetic polymer for developing functional biomaterials. PEG possesses many biocompatible properties including high hydrophilicity and low toxicity.[65] Due to its very low polymer-water interfacial energy, PEG also shows very low protein binding affinity and is often used as an antifouling material.[66] This antifouling property also prohibits most of the material-cell interactions. However, PEG can be easily functionalized to promote desired PEG-cell interactions. For example, PEG can be decorated with peptides harboring a cell binding sequence RGD through various of conjugation chemistries.[65] Peptides with different bioactivities have also been used for PEG functionalization. Peptide sequences that are subjective to matrix metalloproteinase (MMP) cleavage have been incorporated into PEG hydrogel to confer degradability to the hydrogel network.[67] Growth factors or peptides that promote cell differentiation have also been utilized in PEG functionalization. Saik et al. immobilized platelet-derived growth factor-BB (PDGF-BB) and fibroblast growth factor-2 (FGF-2) into PEG hydrogel to promote angiogenesis both *in vitro* and *in vivo*.[68] In the study of done by Leslie-Barbick and coworkers, VEGF-decorated PEG increased tubulogenesis of human umbilical vein endothelial cells (HUVECs) in both 2D and 3D culture forms.[69] Note that a collagenase-

sensitive peptide was also used to confer degradability to the PEG hydrogel in this work. Bioactive peptides such as BMP-2 peptide derived from the knuckle epitope of BMP-2 and VEGF-mimicking QK peptide have also shown success in PEG-functionalization.[70-71] Great reviews on PEG functionalization are available.[65, 72]

PLGA has drawn great interest as a functional biomaterial for its good biocompatibility and intrinsic tunable degradation.[73] PLGA is a copolymer of PLA and PGA. Both PLA and PGA can be degraded via hydrolysis of ester bonds as well as enzymatic digestions. The degraded products, lactic acid and glycolic acid respectively, can be cleaned through the tricarboxylic acid cycle in aerobic organisms. Due to the methyl side chain in the polymer backbone, PLA shows a slower degradation rate than PGA. As the results, the degradation rate of the copolymer PLGA can be tuned by controlling the ratio between the two components. Because of this intrinsic tunable degradation property, PLGA has been widely used in drug delivery applications and is approved for clinical use by the U.S. Food and Drug Administration. Due to its hydrophobicity, hydrophobic drugs including paclitaxel,[74] estradiol,[75] and haloperidol[76] have been loaded into PLGA-based particles with high encapsulation efficiency. PLGA particles have also been used for nucleic acid delivery. Both DNA and RNA can be encapsulated in PLGA particle, however, with a relatively low encapsulation efficiency due to the negative charges on PLGA.[77] Direct-functionalization of PLGA with positively-charged peptides has also been reported to improve DNA delivery efficiency from the functionalized PLGA substrate.[78] Peptide-functionalized PLGA has also been used as tissue engineering scaffolds. Yoon *et al.* showed that bone marrow cells cultured on RGD-conjugated PLGA scaffolds exhibited better attachment and higher alkaline phosphatase (ALP) activity than on RGD-conjugated PLGA negative control scaffolds.[79] Pan and coworker utilized polydopamine-assisted coating method to functionalized PLGA scaffolds with both RGD and BMP-2 derived peptides.[80] Rabbit-derived bone marrow stromal cells showed greater mineralization on the functionalized scaffold surface *in vitro*. In a follow up *in vivo* study, the PLGA scaffold functionalized with BMP-2 derived peptide showed more ectopic bone formation than the plain PLGA scaffold and PLGA scaffold loaded with the peptide.[81]

There are other synthetic polymers show great potentials as functional biomaterials. One example is poly(ϵ-caprolactone) (PCL). PCL is a linear polyester with intrinsic

degradability due to ester hydrolysis.[82] PCL has similar physical and chemical properties to PLGA but with simpler synthesis. Like PLGA, the hydrophobic nature of PCL prevents good cell attachment and interactions to the polymer. As the results, efforts have been made to functionalize PCL to improve the cytocompatibility including coating with extracellular matrix proteins,[83] plasma treatment,[84] or immobilization of biomolecules.[85] Modal et al. wrote a comprehensive review on recent progress and challenges on PCL-based biomaterials for tissue engineering and drug delivery applications.[86] Polyurethanes have also found their applications as biomaterials. Unlike the aforementioned polymers which are defined by their monomers, polyurethanes (PURs) are polymers that polymerized through carbamate links, and different monomers can be used to make PURs. As the result, the physical and chemical properties of PURs have a wide range of distribution. Biostable PURs can be used for biomedical devices that require material stability over a long time period.[87] On the other hand, biodegradable PURs are suitable for tissue engineering or regenerative medicine scaffolds.[88] Functionalization is also possible to enhance material-cell interactions with PURs. Comprehensive reviews are available focusing on different types of tissues including bone,[89] cartilage,[90] and blood vessel.[91]

1.4.2 Natural Polymers

Natural polymers are materials found in organisms in nature. Most of the natural polymers have good biocompatibility and can interact with cells. As the result, efforts have been mostly made on increasing the processability of the natural polymers. Two major types of natural polymers are widely used as functional biomaterials: polysaccharides and proteins. Polysaccharides are carbohydrates composed of mono- or disaccharides joined through glycosidic bonds. Cellulose, chitin/chitosan, and hyaluronic acid have all be used as functional biomaterials. Cellulose is the major component of plant and bacteria cell walls and is the most abundant polysaccharide in nature.[92] Due to the large amount of inter- and intra-molecular hydrogen bonds between cellulose molecules, native cellulose has very low solubility in water and most organic solvents and usually requires modifications to increase its solubility and processability.[92] The extensive hydrogen bonding networks, however, also make cellulose a very stiff material that finds its potential as tissue engineering scaffolds for bone regeneration. Kumbar et al. reported cellulose acetate-based

scaffolds with compressive moduli ranging from 200-300 MPa that match with native bone and show potential as bone grafts.[93]

Chitin is the second most abundant polysaccharide on earth and is mostly found in cell walls of fungi and exoskeletons of arthropods.[94] Similar to cellulose, chitin has low solubilities in water and most organic solvents due to the extensive hydrogen bonding network. Chitin can be deacetylated under alkaline conditions and becomes chitosan. Chitosan has superior solvent solubility and modification potential than chitin because of the amine groups generated after deacetylation.[94] Besides being biocompatible and biodegradable, chitosan also shows antimicrobial activity which makes chitosan a potential material for biomedical implants. Tardajos et al. showed that conjugating chitosan to the surface of PCL-based scaffolds can significantly reduce bacterial growth without compromising cytocompatibility.[95] In a work by Ressler and coworkers, an injectable chitosan/hydroxyapatite hydrogel system was developed.[96] Encapsulated porcine bone marrow MSCs showed higher ALP activity and more mineralization after 14 days of culture, suggesting osteogenic differentiation. In another work done by Lu et al., an anticancer drug for bone tumors, zoledronic acid, was incorporated into chitosan/hydroxyapatite composite scaffolds.[97] The addition of zoledronic acid induced apoptosis in bone tumor cells but not in human MSCs and did not affect the osteogenic effect from hydroxyapatite. The scaffold also significantly reduced colony formation with both Gram-positive and -negative bacteria.

Hyaluronic acid (HA) is a polysaccharide mostly found in connective, epithelial, and neural tissues. Because of its high hydrophilicity, HA regulates water content and functions as a lubricant in the extracellular matrixes.[98] The hydroxyl and carboxyl groups on HA allow crosslinking of HA directly through these functional groups. Glutaraldehyde and divinyl sulfone have been used to crosslinking HA via hydroxyl groups,[99-100] and carbodiimide has been utilized for crosslinking through carboxyl groups.[101] HA can also be functionalized through the hydroxyl and carboxyl groups without altering its bioactivity. Gramlich et al. modified the hydroxyl groups on HA to norbornene groups that allow photocrosslinking with thiol-based crosslinkers.[102] They demonstrated that norbornene photocrosslinked HA hydrogels can be further modified with thiol-containing peptides post-gelation without significantly changing the mechanical properties of the hydrogel. In

another work done by Khetan and coworkers, HA was modified with maleimide and methacrylate groups on the carboxyl groups.[103] This approach allows sequential orthogonal crosslinkings with the primary maleimide-thiol crosslinking providing the hydrogel integrity and the on-demand secondary photocrosslinking between methacrylate groups. They showed that the mechanical properties of the hydrogel can be tuned post gelation and post cell-encapsulation and utilized the system to study the relation between stem cell differentiation and cellular traction.

Proteins are the major macromolecules in the mammalian ECM.[104] Collagen, elastin, fibronectin, and laminin are the four most abundant ECM proteins. Being the major protein providing structural integrity, collagen has been widely explored as functional biomaterials. Collagens can be divided into two main categories based on their ability to undergo fibrillogenesis. Among all types of collagens, type I and type II fibrillar collagen are the most explored two in the literature. Type I collagen is often found in skin and bone tissues, and type II collagen is mostly found in cartilage tissue. Both type I and II collagens can form hydrogels spontaneously through fibrillogenesis. The resulting mechanical properties of the hydrogels are, however, usually on the soft end. Type II collagen hydrogels have lower stiffness than type I collagen hydrogels due to the higher degree of glycosylation on type II collagen which impedes the formation of higher order fibril structures.[105] Vázquez-portalatín and Kilmer and coworkers showed that the mechanical properties of type II collagen hydrogel can be increased by blending with type I collagen.[106] The extent of the increase depended on the ratio between type I and type II collagen. The blending of type I and type II collagens did not alter the fibril dimeter distribution but had an effect on the pore structure in the fibril network.

Chemical crosslinking with glutaraldehyde or 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and N-hydroxysuccinimide (NHS) have been utilized to increase the mechanical properties of collagen hydrogels.[107-108] Collagen hydrogels can also be reinforced through photocrosslinking. Zhang and coworkers showed that concentrated collagen solution can be crosslinked directly by γ-radiation.[109] The compressive moduli of the crosslinked hydrogels can be controlled by the radiation dosage. The crosslinked collagen hydrogels showed good cytocompatibility both *in vivo* and *in vitro*. Implantation of these crosslinked collagen hydrogels into rats did not inflict

observable inflammation, and the extent of cell infiltration and degradation was related to the γ-radiation dosage used for crosslinking. UV-riboflavin photocrosslinking is another approach that have been used clinically for cornea repair.[110] Tirella et al. showed that this approach can be applied to collagen hydrogels, and the stiffness of the hydrogels was tunable by varying the collagen concentration.[111] In another work by Heo and coworkers, UV-riboflavin crosslinked collagen hydrogels showed slower *in vivo* and *in vitro* degradation than non-crosslinked collagen hydrogels.[112] Fibrochondrocytes encapsulated in the crosslinked hydrogels also showed high viability.

Being the major component in the ECM, collagens have multiple intrinsic bioactivities. Type I collagen has intrinsic RGD cell binding motifs in the sequence.[113] Taubenberger et al. showed that These RGD motifs can be better exposed to cells by partially heat-denaturing type I collagens.[114] The increased accessibility to RGD motifs not only enhanced initial cell spreading and migration but also promoted mineralization of pre-osteoblasts. Type II collagen is the major protein component in cartilage and has been explored as the materials for cartilage tissue engineering. Bosnakovski and coworkers showed that bovine MSCs showed higher chondrogenic gene expressions in the type II collagen/alginate hydrogel than in the type I collagen/alginate or alginate hydrogels both in chondrogenic and plain media.[115] Collagens can also be functionalized to have additional bioactivities. One such example is the work done by Wissink *et al.* where heparin is immobilized onto crosslinked collagen hydrogels.[116] Type I collagen has been used as a coating on vascular grafts to promote growth of endothelial cells. However, the thrombogenic activity of collagen can cause premature graft failure before the graft is covered by endothelial cells. This work demonstrated that thrombin activation can be decreased by conjugating heparin onto collagens at an optimal heparin density.

1.4.3 Recombinant Proteins

This section is adapted from a manuscript by Lin C-Y and Liu JC published in *Current Opinion in Biotechnology*, Volume. 40, Pages 56-63, 2016. Reproduced with the permission of Elsevier.

Progress in molecular biology has enabled the facile production of recombinant proteins. Furthermore, maturation in large-scale production techniques makes protein-based materials more economically feasible. New techniques such as incorporation of non-canonical amino acids further expand the possibilities of protein-based biomaterials.

A recombinant protein is designed at the DNA level, and DNA sequences encoding different protein domains can be assembled in the desired order with a specified number of repeats. This modularity in recombinant protein design enables the production of recombinant protein-based materials with diverse properties. The translation machinery also has high fidelity so that the desired recombinant protein will have the specified amino acid sequence. On the other hand, synthetic polymers or proteins harvested from nature often have dispersity in chain length or composition. Thus, the high fidelity in recombinant protein production promises precise control over material properties.

In general, a recombinant protein designed for tissue engineering applications is composed of modular domains that confer specific functions. For example, structural domains provide mechanical properties, and biological domains facilitate the interactions between cells and the materials. In addition, domains can be used that respond to environmental stimuli or enable spatiotemporal control over material properties. In this section, we focus on domains that are being widely used in protein-based materials or are being incorporated into composite materials for added functionality.

1.4.3.1 Structural Domains

Elastin-like polypeptides (ELPs) are based on sequences derived from native elastin and are being actively studied. Elastin is the major component that provides elasticity to the ECM. ELPs are able to mimic the mechanical properties of native elastin and are mostly composed of a repeating amino acid sequence $(VPGXG)_n$, where X is a guest residue consisting of any amino acid except proline.[117] The flexibility in choosing the guest residue expands the functionality of ELPs as a structural domain. One example is the use of lysine as a guest residue. This choice allows crosslinking through the primary amine side chain, and a range of mechanical moduli can be achieved with different degrees of crosslinking.[118] Another example is the use of cysteine as a guest residue to facilitate surface immobilization of ELPs or crosslinking of free ELPs.[119] ELPs are also thermo-responsive,

and this feature can be modulated using the guest residues. ELPs exhibit LCST behavior; they are soluble below the LCST and form a coacervate, or a dense polymer-rich liquid phase, above the LCST. The Rodríguez-Cabello group has utilized the LCST behavior of ELPs to facilitate hydrogel formation of ELP-fusion proteins.[120-121] In addition, spherical structures of ELPs have been triggered by the thermo-responsive behavior.[122] Structures formed by ELPs include hollow spheres or spheres with a dense core, and their size distributions can be controlled by salt concentration[123] or by the number of repeating units.[124] ELP spheres have been used as templates for nanoparticle synthesis[125] and have been applied to other applications, such as drug delivery vehicles.[34]

Resilin, which is found in insect cuticle, has drawn great interest with its high elasticity, high resilience, and heat stability.[126] The Elvin group first developed resilin-like polypeptides (RLPs) based on sequences from *Drosophila melanogaster* (Dros16, $(GGRPSDSYGAPGGGN)_n$) and *Anopheles gambiae* (An16, $(AQTPSSQYGAP)_n$), and both RLPs possess mechanical properties and heat stability that are similar to those of native resilin.[126] Unlike ELPs, there are no guest residues in RLP sequences; however, RLP amino acid sequences in which lysine residues have been inserted to serve as a crosslinking site still retain resilin-like characteristics.[127-128] Photochemical crosslinking through tyrosine residues in RLPs can be mediated through Ru(II), and crosslinked RLPs have been explored for various applications. For example, Lv *et al.* reported a recombinant protein based on Dros16 and an RGD-containing domain from tenascin-C.[129] Crosslinked hydrogels of this material had tunable mechanical properties controlled by protein concentration and facilitated spreading of human lung fibroblasts. Another example is the use of photocrosslinked RLPs to modify tissue culture polystyrene surfaces.[130] High coating concentrations prevented fibroblast attachment and spreading, but cell attachment could be restored by incorporating an RGD peptide in the coating.

1.4.3.2 Bioactive Domains

One advantage of modular recombinant proteins is that bioactive cues can be directly incorporated among the structural domains at the desired location and valency. In tissue engineering, the following issues have been addressed by the strategic fusion of bioactive

domains: cell-material interactions, cell fate determination, and material response to cellular activities.

Cell attachment to the material is often the first consideration when designing biomaterials. Many cell-adhesive domains are derived from ECM proteins such as fibronectin, collagen, and laminin. The RGD sequence is a classic cell-binding domain that is derived from the 10th domain of fibronectin type III (FNIII). The RGD sequence is often presented with the PHSRN synergy site, which is derived from the 9th domain of FNIII, to increase cell adhesion and target the $\alpha_5\beta_1$ integrin. The CS5 sequence, with the minimum sequence of REDV, is another cell-binding domain derived from FNIII. Hsueh and coworkers reported that ELP proteins containing the CS5 domain supported murine Schwann cell proliferation.[131]

Another set of cell-binding domains is derived from laminins, which are heterotrimeric proteins found in the basal lamina. The PPFLMLLKGSTR sequence is derived from the laminin-5 α3 chain globular domain 3 (LG3). Including this domain within an ELP facilitated human keratinocyte attachment.[132] This work also examined keratinocyte attachment with ELPs containing two other binding domains: fibronectin domains containing the RGD and PHSRN sequences and the GEFYFYDLRLKGDK sequence derived from the α1 chain of collagen type IV. Keratinocytes attached more quickly to proteins with the laminin or fibronectin domains compared to the collagen domain. The authors found that keratinocyte attachment to the laminin and fibronectin domains could be reduced with antibodies against the α_3 and α_5 integrin subunits, respectively, and thus concluded that the keratinocytes were likely utilizing the $\alpha_3\beta_I$ and $\alpha_5\beta_I$ integrins, respectively, to adhere to those proteins. Many other cell-binding domains have been reported (e.g., DGEA from collagen type I and IKVAV and YIGSR from laminin) and are described in a review.[133] However, new cell-binding domains are still being identified. Lee and coworkers recently showed that the C-terminal RKRK sequence cannot completely account for cell attachment to human tropoelastin and suggested that there is a new binding domain in domains 17 and 18 of tropoelastin.[134]

Promoting stem cell differentiation into the desired cell lineage and preventing committed cells from de-differentiating are critical to the success of tissue engineering. Because growth factors play an important role in morphogenesis, it is logical that many of

the bioactive domains that are capable of determining cell commitment are derived from the corresponding growth factors. Bone morphogenetic proteins (BMPs) are a family of growth factors that are involved in the development of different tissues, but they are best known for their roles in bone morphogenesis. Recombinant human BMP-2 and -7 are approved by the FDA for clinical applications. The KIPKASSVPTELSAISTLYL peptide is derived from the knuckle epitope of human BMP-2 and has been widely used as a bioactive domain to promote osteogenesis. Kim *et al*. recently showed that the BMP-2 peptide retained its bioactivity when fused in an RLP backbone and that the fusion protein enhanced osteogenic differentiation.[135] The peptide has also been reported to promote chondrogenesis of human mesenchymal stem cells (hMSCs) in pellet culture.[136] Three peptides have been derived from human BMP-7: SNVILKKYRN, KPSSAPTQLN, and KAISVLYFDDS. In a recent work by Tao and coworkers, these three peptides were incorporated into self-assembling peptides and promoted ECM secretion by human degenerated nucleus pulposus cells.[137] These results demonstrate the potential that these peptides have for intervertebral disc regeneration, which is a major clinical application of recombinant human BMP-7. It is expected that these BMP-7 derived peptides could be easily incorporated into recombinant protein-based materials through the modular design approach.

Growth factors such as VEGF and PDGF play an important role in blood vessel formation in regenerating tissue. The KLTWQELYQLKYKGI sequence (named QK) is based on VEGF. The QK peptide retained its bioactivity and promoted endothelial cell behavior when crosslinked to a protein backbone[138] or presented as a fusion protein.[139] Besides its well-known effects on angiogenesis, VEGF also has protective effects on neuronal cells. Verheyen and coworkers showed that the QK peptide also displayed this function and protected neurons from paclitaxel toxicity and hyperglycemic stress.[140]

Biomaterial degradation is an important factor in tissue engineering. Ideally, degradation should synchronize with cellular regeneration so that there will be room for newly formed tissue. Peptide sequences that are sensitive to proteases can be used as degradation domains in recombinant protein-based materials. For example, ELPs are usually used as structural domains; however, they can also serve as degradation domains due to their sensitivity to elastase.[141] Degradation domains can also be explicitly

incorporated into modular designs. A popular choice is MMP-sensitive sequences. MMPs are a family of endopeptidases that can degrade ECM proteins including collagen, elastin, fibronectin, and laminin. MMP-sensitive sequences have been widely used in applications requiring material degradation. For example, Price and coworkers incorporated an MMP-sensitive sequence into a silk-elastin-like protein hydrogel for viral-mediated gene delivery for cancer treatment.[142] When implanted into mice, MMP-sensitive hydrogels had higher cell invasion compared to MMP-insensitive hydrogels. Tumor-bearing mice treated with MMP-sensitive hydrogels had the highest survival rate.

MMPs also play important roles in many physiological events such as morphogenesis and tissue remodeling. For example, Fonseca and coworkers profiled the *in vitro* gene expression of hMSCs grown in basal or osteogenic medium and found that in osteogenic medium there was an increase in MMP-14 gene expression levels and ALP activity at one week.[143] Thus, it is anticipated that by selecting specific MMP-sensitive sequences it is possible to design materials that are selectively degraded by desired cells at specific stages of differentiation. For example, Sridhar *et al.* used an MMP-sensitive sequence, KCGPQGIWGQCK, as a degradable crosslinker for PEG hydrogels. Cells encapsulated in the degradable gels showed higher glycosaminoglycan (GAG) and collagen deposition compared to those in non-degradable hydrogels.[67] In general, MMPs share common features in their cleavage sites. A recent study by Kukreja and coworkers analyzed target sequences of 18 MMPs and identified information for predicting and designing MMP-cleavable sequences.[144] Overall, understanding the specificities of MMPs and identifying their cleavage sequences will expand the possibilities of using degradation domains for targeted applications.

1.4.3.3 Domains for Higher-Order Control over Material Properties

In an ideal functional biomaterial, the material properties are precisely controlled, and the material is responsive to dynamic cellular activities and changes in environmental conditions. Therefore, domains that enable material responsiveness and spatial or temporal control over material properties have been explored. The Davies group utilized two MMP sequences with different enzyme specificities (one sequence is recognized by many MMPs whereas the other sequence is recognized by a specific MMP) and successfully modulated

the *in vitro* invasion rates of fibroblasts and vascular smooth muscle cells into PEG hydrogels.[145] They also demonstrated control over *in vivo* cellular invasion by making hydrogels with a mixture of MMP-sensitive sequences.[146]

Leucine zippers have been utilized to control ligand accessibility and density. The stability and oligomerization states of leucine zippers can be easily tuned by changing their amino acid sequences. A heterodimeric leucine zipper pair was used to reversibly enable access to an RGD cell-binding domain. Exposure of gold nanorods to near-infrared (NIR) light resulted in a photothermal effect, which effectively denatured the leucine zippers and enabled access to the RGD sequence.[147] Leucine zippers have also been utilized to increase ligand avidity. For example, a trimeric leucine zipper, cartilage matrix protein (CMP), was recently used to present a ligand for epidermal growth factor receptor (EGFR). The EGFR ligand was recombinantly fused to a CMP domain, and the fusion proteins oligomerized to form trimers through the CMP domain. A monomeric control was constructed by fusing the EGFR ligand to a mutated CMP domain, which was unable to oligomerize to form a trimer. When the ligand was presented as a CMP-facilitated trimer, it demonstrated enhanced binding strength compared to the monomeric ligand.[148]

Conditional-splicing inteins are an intriguing domain in the recombinant protein toolbox. Unlike normal inteins, conditional-splicing inteins only splice after an environmental stimulus. Informative reviews of recent progress on inteins are available.[149-150] Recently, photoactivatable inteins have been achieved by incorporating non-canonical, photoactive amino acids. For example, both cysteine and serine residues in inteins have been modified with a photocage, and these photoactivatable inteins are promising protein-labeling tools with exquisite spatiotemporal control that can be directly used in live mammalian cells.[151-152] Inteins with photoactive tyrosines have also been used as a tool for making cyclic peptides.[153]

In conclusion, biomaterials with sophisticated control over properties are increasingly needed to address questions regarding cellular behavior for tissue engineering applications. These materials need to meet demanding requirements for specific mechanical properties, macromolecular structure, cell-material interactions, and responsiveness towards environmental changes. Protein domains and peptide sequences provide the desired functionality to recombinant proteins or to composite materials. In

particular, recombinant proteins are promising materials because their modularity enables the facile combination of domains that confer the desired structure, bioactivity, and functionality. As we continue to unlock the sequence-structure-function relationship of natural proteins, we can expand the number of domains available as part of our recombinant protein toolbox. Thus, protein domains can be used to design materials that can address a larger variety of questions not only in the field of tissue engineering but also in stem cell and developmental biology, pharmaceutical engineering, and clinical practice.

1.5 Thesis Outline and Contributions

The goal of this dissertation was to explore different rational design approaches for functional materials for various biological applications. We show in Chapter 2 that a new thiol-crosslinking approach for hydrogel formation that allows independently tunable mechanical properties and degradation behavior in response to reducing agents. In Chapter 3, we describe the development of two protein-based adhesives utilizing catechol or thiol chemistry and compare their adhesion performances in a wet setting. In Chapter 4, we investigate the effect of electrostatic interactions in the LCST behavior of ELPs and show that such effect can be harnessed as a new factor tailoring the LCST behavior of ELPs. For this dissertation, I designed and produced all proteins, planned and conducted all experiments, and wrote all the contents unless otherwise stated.

Chapter 2 describes a one-pot thiol-ene addition and thiol oxidation crosslinking for redox-sensitive hydrogels. Thiol-functionalized PEG (PEG-SH) was chose as the model polymer for hydrogel formation. Divinyl sulfone (DVS) was used as the crosslinker in thiol-ene additions and formed non-reducible thioether bonds. Ferric ethylenediaminetetraacetic acid (Ferric EDTA) was used to promote thiol oxidation to form reducible disulfide bonds. By changing the ratio between DVS and Ferric EDTA, the ratio between thioether and disulfide bonds was varied, which determined the degradability of the hydrogel in response to reducing agents. Hydrogels with different degradability were characterized rheologically with similar storage moduli. Swelling, degradation, and release of the encapsulated dextran in response to reducing agents were assessed. Cytocompatibility of DVS and ferric EDTA were evaluated by live/dead assay on encapsulated fibroblast. Our results show that the hydrogels crosslinked with the

combination of DVS and ferric EDTA have a tunable degradation rate in response to GSH at an ECM concentration of 10 μM while maintaining similar stiffnesses. The crosslinking approach is highly cytocompatibility, which suggests its potential in tissue engineering and drug delivery.

Chapter 3 describes the development and the performance comparison of two protein-based adhesives utilizing catechol or thiol chemistry. Taking the advantage of the flexibility in ELP sequence, two adhesive ELPs were designed to have high consensus in sequence. The catechol-featured ELP was rich in tyrosine residue which was later converted to 3,4-dihydroxyphenylalanine (DOPA) enzymatically. The thiol-featured ELP was abundant in cysteine residue which presents thiol groups. The adhesion strength with porcine skins in a wet setting of both the ELP adhesives were tested with different curing times and with or without the addition of ferric EDTA. The curing kinetic and the tolerance to disturbance during curing of the two adhesives were compared. In addition, we showed that a synergistic increase in the adhesion strength can be achieved by mixing the two adhesives because of the cross-reactions between DOPA and thiol groups. Finally, the cytocompatibility of these ELP-based adhesives was evaluated. Our data show the potential of these ELP-based materials as tissue adhesives and provide insights in designing adhesives with different chemistries.

Chapter 4 describes the investigation of the effect of electrostatic interactions on the LCST behavior of ELPs. Taking the advantage of recombinant protein modular design, we constructed two ELPs with the same ELP domain but with external sequences exhibiting different charge states in response to pH changes. By combining the ELP domain to the external sequence with a pH-sensitive charge state, we demonstrated that an additional pH sensitivity can be introduced to the LCST behavior of the ELP protein. We further showed that the effect from the pH-sensitive external sequence can be reduced by decreasing the ratio between the external sequence and the ELP domain or by increasing the ionic strength in the solution. Last but not least, we reported that the identity of the external sequence alters the interactions between the ELP proteins and Hoffmeister ions and thus changes the LCST behavior in response to different Hoffmeister ions. Our results suggest that external sequences can be harness as a new factor for tailoring the LCST

behavior of ELPs and provide insights in designing ELPs for thermo-responsive and/or pH-responsive hydrogels and drug delivery vehicles.

Chapter 5 contains the summary and major conclusions of this research and outlines future directions.

1.6 References

(1) Thomas, A. Functional Materials: From Hard to Soft Porous Frameworks. *Angew. Chem., Int. Ed.* **2010,** *49* (45), 8328-8344.

(2) Sakiyama-Elbert, S.; Hubbell, J. Functional biomaterials: design of novel biomaterials. *Annu. Rev. Mater. Res.* **2001,** *31* (1), 183-201.

(3) Tao, H.; Wu, Y.; Li, H.; Wang, C.; Zhang, Y.; Li, C.; Wen, T.; Wang, X.; He, Q.; Wang, D. BMP7-Based Functionalized Self-Assembling Peptides for Nucleus Pulposus Tissue Engineering. *ACS Appl. Mater. Interfaces* **2015,** *7* (31), 17076-17087.

(4) Stephan, A. Organ Shortage: Can We Decrease the Demand? *Exp. Clin. Transplant.* **2017,** *15* (Suppl 1), 6-9.

(5) Koh, C. J.; Atala, A. Tissue Engineering, Stem Cells, and Cloning: Opportunities for Regenerative Medicine. *J. Am. Soc. Nephrol.* **2004,** *15* (5), 1113.

(6) Pittenger, M. F.; Mackay, A. M.; Beck, S. C.; Jaiswal, R. K.; Douglas, R.; Mosca, J. D.; Moorman, M. A.; Simonetti, D. W.; Craig, S.; Marshak, D. R. Multilineage Potential of Adult Human Mesenchymal Stem Cells. *Science* **1999,** *284* (5411), 143.

(7) Engler, A. J.; Sen, S.; Sweeney, H. L.; Discher, D. E. Matrix Elasticity Directs Stem Cell Lineage Specification. *Cell* **2006,** *126* (4), 677-689.

(8) Mareschi, K.; Ferrero, I.; Rustichelli, D.; Aschero, S.; Gammaitoni, L.; Aglietta, M.; Madon, E.; Fagioli, F. Expansion of mesenchymal stem cells isolated from pediatric and adult donor bone marrow. *J. Cell. Biochem.* **2006,** *97* (4), 744-754.

(9) Takahashi, K.; Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **2006,** *126* (4), 663-676.

(10) Takahashi, K.; Tanabe, K.; Ohnuki, M.; Narita, M.; Ichisaka, T.; Tomoda, K.; Yamanaka, S. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **2007,** *131* (5), 861-872.

(11) Zhou, H.; Wu, S.; Joo, J. Y.; Zhu, S.; Han, D. W.; Lin, T.; Trauger, S.; Bien, G.; Yao, S.; Zhu, Y.; Siuzdak, G.; Schöler, H. R.; Duan, L.; Ding, S. Generation of Induced Pluripotent Stem Cells Using Recombinant Proteins. *Cell Stem Cell* **2009,** *4* (5), 381-384.

(12) Warren, L.; Manos, P. D.; Ahfeldt, T.; Loh, Y.-H.; Li, H.; Lau, F.; Ebina, W.; Mandal, P. K.; Smith, Z. D.; Meissner, A.; Daley, G. Q.; Brack, A. S.; Collins, J. J.; Cowan, C.; Schlaeger, T. M.; Rossi, D. J. Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **2010,** *7* (5), 618-630.

(13) Hou, P.; Li, Y.; Zhang, X.; Liu, C.; Guan, J.; Li, H.; Zhao, T.; Ye, J.; Yang, W.; Liu, K.; Ge, J.; Xu, J.; Zhang, Q.; Zhao, Y.; Deng, H. Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. *Science* **2013**, 1239278.

(14) Yoshihara, M.; Hayashizaki, Y.; Murakawa, Y. Genomic Instability of iPSCs: Challenges Towards Their Clinical Applications. *Stem Cell Rev. Rep.* **2017,** *13* (1), 7-16.

(15) Brafman, D. A. Constructing stem cell microenvironments using bioengineering approaches. *Physiol. Genomics* **2013,** *45* (23), 1123-1135.

(16) Bessa, P. C.; Casal, M.; Reis, R. L. Bone morphogenetic proteins in tissue engineering: the road from the laboratory to the clinic, part I (basic concepts). *J. Tissue Eng. Regen. Med.* **2008,** *2* (1), 1-13.

(17) Bessa, P. C.; Casal, M.; Reis, R. L. Bone morphogenetic proteins in tissue engineering: the road from laboratory to clinic, part II (BMP delivery). *J. Tissue Eng. Regen. Med.* **2008,** *2* (2- 3), 81-96.

(18) Madry, H.; Rey-Rico, A.; Venkatesan, J. K.; Johnstone, B.; Cucchiarini, M. Transforming Growth Factor Beta-Releasing Scaffolds for Cartilage Tissue Engineering. *Tissue Eng., Part B* **2013,** *20* (2), 106-125.

(19) Rouwkema, J.; Khademhosseini, A. Vascularization and Angiogenesis in Tissue Engineering: Beyond Creating Static Networks. *Trends Biotechnol.* **2016,** *34* (9), 733-745.

(20) Saito, A.; Suzuki, Y.; Ogata, S.-i.; Ohtsuki, C.; Tanihara, M. Activation of osteo-progenitor cells by a novel synthetic peptide derived from the bone morphogenetic protein-2 knuckle epitope. *Biochim. Biophys. Acta, Proteins Proteomics* **2003,** *1651* (1), 60-67.

(21) Zhang, G.; Zhou, J.; Fan, Q.; Zheng, Z.; Zhang, F.; Liu, X.; Hu, S. Arterial–venous endothelial cell fate is related to vascular endothelial growth factor and Notch status during human bone mesenchymal stem cell differentiation. *FEBS Lett.* **2008,** *582* (19), 2957-2964.

(22) Huebsch, N.; Arany, P. R.; Mao, A. S.; Shvartsman, D.; Ali, O. A.; Bencherif, S. A.; Rivera-Feliciano, J.; Mooney, D. J. Harnessing traction-mediated manipulation of the cell/matrix interface to control stem-cell fate. *Nat. Mater.* **2010,** *9*, 518.

(23) Datta, N.; P. Pham, Q.; Sharma, U.; Sikavitsas, V. I.; Jansen, J. A.; Mikos, A. G. In vitro generated extracellular matrix and fluid shear stress synergistically enhance 3D osteoblastic differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (8), 2488.

(24) Huang, Y.; Jia, X.; Bai, K.; Gong, X.; Fan, Y. Effect of Fluid Shear Stress on Cardiomyogenic Differentiation of Rat Bone Marrow Mesenchymal Stem Cells. *Arch. Med. Res.* **2010,** *41* (7), 497-505.

(25) Zouani, O. F.; Kalisky, J.; Ibarboure, E.; Durrieu, M.-C. Effect of BMP-2 from matrices of different stiffnesses for the modulation of stem cell fate. *Biomaterials* **2013,** *34* (9), 2157-2166.

(26) Hutmacher, D. W. Scaffolds in tissue engineering bone and cartilage. In *The Biomaterials: Silver Jubilee Compendium*; Williams, D. F., Ed.; Elsevier Science: Oxford, 2000; pp 175-189.

(27) Drury, J. L.; Mooney, D. J. Hydrogels for tissue engineering: scaffold design variables and applications. *Biomaterials* **2003,** *24* (24), 4337-4351.

(28) Cosson, S.; Otte, E. A.; Hezaveh, H.; Cooper-White, J. J. Concise Review: Tailoring Bioengineered Scaffolds for Stem Cell Applications in Tissue Engineering and Regenerative Medicine. *Stem Cells Transl. Med.* **2015,** *4* (2), 156-164.

(29) Hans, M. L.; Lowman, A. M. Biodegradable nanoparticles for drug delivery and targeting. *Curr. Opin. Solid State Mater. Sci.* **2002,** *6* (4), 319-327.

(30) Kapoor, D. N.; Bhatia, A.; Kaur, R.; Sharma, R.; Kaur, G.; Dhawan, S. PLGA: a unique polymer for drug delivery. *Ther. Delivery* **2015,** *6* (1), 41-58.

(31) Masood, F. Polymeric nanoparticles for targeted drug delivery system for cancer therapy. *Mater. Sci. Eng., C* **2016,** *60*, 569-578.

(32) Hollingshead, S.; Lin, C.-Y.; Liu, J. C. Designing Smart Materials with Recombinant Proteins. *Macromol. Biosci.* **2017,** *17* (7), 1600554.

(33) Ashraf, S.; Park, H.-K.; Park, H.; Lee, S.-H. Snapshot of phase transition in thermoresponsive hydrogel PNIPAM: Role in drug delivery and tissue engineering. *Macromol. Res.* **2016,** *24* (4), 297-304.

(34) MacEwan, S. R.; Chilkoti, A. Applications of elastin-like polypeptides in drug delivery. *J. Control. Release* **2014,** *190*, 314-330.

(35) Rodríguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A. Elastin-like polypeptides in drug delivery. *Adv. Drug Delivery Rev.* **2016,** *97*, 85-100.

(36) Massodi, I.; Bidwell, G. L.; Raucher, D. Evaluation of cell penetrating peptides fused to elastin-like polypeptide for drug delivery. *J. Control. Release* **2005,** *108* (2), 396-408.

(37) Gerweck, L. E.; Seetharaman, K. Cellular pH Gradient in Tumor versus Normal Tissue: Potential Exploitation for the Treatment of Cancer. *Cancer Res.* **1996,** *56* (6), 1194.

(38) Chen, W.; Meng, F.; Li, F.; Ji, S.-J.; Zhong, Z. pH-responsive biodegradable micelles based on acid-labile polycarbonate hydrophobe: synthesis and triggered drug release. *Biomacromolecules* **2009,** *10* (7), 1727-1735.

(39) Callahan, D. J.; Liu, W.; Li, X.; Dreher, M. R.; Hassouneh, W.; Kim, M.; Marszalek, P.; Chilkoti, A. Triple stimulus-responsive polypeptide nanoparticles that enhance intratumoral spatial distribution. *Nano Lett.* **2012,** *12* (4), 2165-2170.

(40) Kuppusamy, P.; Li, H.; Ilangovan, G.; Cardounel, A. J.; Zweier, J. L.; Yamada, K.; Krishna, M. C.; Mitchell, J. B. Noninvasive Imaging of Tumor Redox Status and Its Modification by Tissue Glutathione Levels. *Cancer Res.* **2002,** *62* (1), 307.

(41) Cheng, R.; Feng, F.; Meng, F.; Deng, C.; Feijen, J.; Zhong, Z. Glutathione-responsive nano-vehicles as a promising platform for targeted intracellular drug and gene delivery. *J. Control. Release* **2011,** *152* (1), 2-12.

(42) Zhang, P.; Wu, J.; Xiao, F.; Zhao, D.; Luan, Y. Disulfide bond based polymeric drug carriers for cancer chemotherapy and relevant redox environments in mammals. *Med. Res. Rev.* **2018,** *38* (5), 1485-1510.

(43) Mahdavinia, G. R.; Etemadi, H.; Soleymani, F. Magnetic/pH-responsive beads based on caboxymethyl chitosan and κ-carrageenan and controlled drug release. *Carbohydr. Polym.* **2015,** *128*, 112-121.

(44) Ulbrich, K.; Hola, K.; Subr, V.; Bakandritsos, A.; Tucek, J.; Zboril, R. Targeted drug delivery with polymers and magnetic nanoparticles: covalent and noncovalent approaches, release control, and clinical studies. *Chem. Rev.* **2016,** *116* (9), 5338-5431.

(45) Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

(46) Weiser, T. G.; Haynes, A. B.; Molina, G.; Lipsitz, S. R.; Esquivel, M. M.; Uribe-Leitz, T.; Fu, R.; Azad, T.; Chao, T. E.; Berry, W. R.; Gawande, A. A. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet* **2015,** *385*, S11.

(47) Bhagat, V.; Becker, M. L. Degradable Adhesives for Surgery and Tissue Engineering. *Biomacromolecules* **2017,** *18* (10), 3009-3039.

(48) Duarte, A. P.; Coelho, J. F.; Bordado, J. C.; Cidade, M. T.; Gil, M. H. Surgical adhesives: Systematic review of the main types and development forecast. *Prog. Polym. Sci.* **2012,** *37* (8), 1031-1050.

(49) Scognamiglio, F.; Travan, A.; Rustighi, I.; Tarchi, P.; Palmisano, S.; Marsich, E.; Borgogna, M.; Donati, I.; de Manzini, N.; Paoletti, S. Adhesive and sealant interfaces for general surgery applications. *J. Biomed. Mater. Res., Part B* **2016,** *104* (3), 626-639.

(50) Strausberg, R. L.; Link, R. P. Protein-based medical adhesives. *Trends Biotechnol.* **1990,** *8*, 53-57.

(51) Zhang, S.; Ruiz Sr, R. Sterilized liquid compositions of cyanoacrylate monomer mixtures. U.S. Patent 8,652,510, 2014.

(52) Quinn, J. V. *Tissue adhesives in clinical medicine*, PMPH-USA: 2005; Vol. 1.

(53) Shapiro, A. J.; Dinsmore, R. C.; North Jr, J. H. Tensile strength of would closure with cyanoacrylate glue. *The American Surgeon* **2001,** *67* (11), 1113.

(54) Singer, A. J.; Perry, L. A Comparative Study of the Surgically Relevant Mechanical Characteristics of the Topical Skin Adhesives. *Acad. Emerg. Med.* **2012,** *19* (11), 1281-1286.

(55) Leggat, P. A.; Smith, D. R.; Kedjarune, U. Surgical applications of cyanoacrylate adhesives: A review of toxicity. *Aust. N. Z. J. Surg.* **2007,** *77* (4), 209-213.

(56) Singer, A. J.; Quinn, J. V.; Hollander, J. E. The cyanoacrylate topical skin adhesives. *Am. J. Emerg. Med.* **2008,** *26* (4), 490-496.

(57) Spotnitz, W. D.; Prabhu, R. Fibrin Sealant Tissue Adhesive−Review and Update. **2005,** *15* (3), 245-270.

(58) Spotnitz, W. D. Commercial fibrin sealants in surgical care. *Am. J. Surg.* **2001,** *182* (2, Supplement 1), S8-S14.

(59) Horowitz, B.; Busch, M. Estimating the pathogen safety of manufactured human plasma products: application to fibrin sealants and to thrombin. *Transfusion (Paris)* **2008,** *48* (8), 1739-1753.

(60) Mehdizadeh, M.; Yang, J. Design Strategies and Applications of Tissue Bioadhesives. *Macromol. Biosci.* **2013,** *13* (3), 271-288.

(61) Spotnitz, W. D.; Burks, S. Hemostats, sealants, and adhesives: components of the surgical toolbox. *Transfusion (Paris)* **2008,** *48* (7), 1502-1516.

(62) Hunstad, J. P.; Michaels, J.; Burns, A. J.; Slezak, S.; Stevens, W. G.; Clower, D. M.; Rubin, J. P. A Prospective, Randomized, Multicenter Trial Assessing a Novel Lysine-Derived Urethane Adhesive in a Large Flap Surgical Procedure without Drains. *Aesthetic Plast. Surg.* **2015,** *39* (4), 616-624.

(63) TissuGlu® Surgical Adhesive Patient Information Brochure. FDA, Ed. 2014.

(64) Park, K. H.; Seong, K.-Y.; Yang, S. Y.; Seo, S. Advances in medical adhesives inspired by aquatic organisms' adhesion. *Biomater. Res.* **2017,** *21* (1), 16.

(65) Zhu, J. Bioactive modification of poly(ethylene glycol) hydrogels for tissue engineering. *Biomaterials* **2010,** *31* (17), 4639-4656.

(66) Krishnan, S.; Weinman, C. J.; Ober, C. K. Advances in polymers for anti-biofouling surfaces. *J. Mater. Chem.* **2008,** *18* (29), 3405-3413.

(67) Sridhar, B. V.; Brock, J. L.; Silver, J. S.; Leight, J. L.; Randolph, M. A.; Anseth, K. S. Development of a Cellularly Degradable PEG Hydrogel to Promote Articular Cartilage Extracellular Matrix Deposition. *Adv. Healthcare Mater.* **2015,** *4* (5), 702-713.

(68) Saik, J. E.; Gould, D. J.; Watkins, E. M.; Dickinson, M. E.; West, J. L. Covalently immobilized platelet-derived growth factor-BB promotes angiogenesis in biomimetic poly(ethylene glycol) hydrogels. *Acta Biomater.* **2011,** *7* (1), 133-143.

(69) Leslie-Barbick, J. E.; Moon, J. J.; West, J. L. Covalently-immobilized vascular endothelial growth factor promotes endothelial cell tubulogenesis in poly (ethylene glycol) diacrylate hydrogels. *J. Biomater. Sci. Polym. Ed.* **2009,** *20* (12), 1763-1779.

(70) He, X.; Ma, J.; Jabbari, E. Effect of Grafting RGD and BMP-2 Protein-Derived Peptides to a Hydrogel Substrate on Osteogenic Differentiation of Marrow Stromal Cells. *Langmuir* **2008,** *24* (21), 12508-12516.

(71) Leslie-Barbick, J. E.; Saik, J. E.; Gould, D. J.; Dickinson, M. E.; West, J. L. The promotion of microvasculature formation in poly(ethylene glycol) diacrylate hydrogels by an immobilized VEGF-mimetic peptide. *Biomaterials* **2011,** *32* (25), 5782-5789.

(72) Bakaic, E.; Smeets, N. M.; Hoare, T. Injectable hydrogels based on poly (ethylene glycol) and derivatives as functional biomaterials. *RSC Adv.* **2015,** *5* (45), 35469-35486.

(73) Gentile, P.; Chiono, V.; Carmagnola, I.; Hatton, P. An overview of poly (lactic-co-glycolic) acid (PLGA)-based biomaterials for bone tissue engineering. *Int. J. Mol. Sci.* **2014,** *15* (3), 3640-3659.

(74) Fonseca, C.; Simões, S.; Gaspar, R. Paclitaxel-loaded PLGA nanoparticles: preparation, physicochemical characterization and in vitro anti-tumoral activity. *J. Control. Release* **2002,** *83* (2), 273-286.

(75) Mittal, G.; Sahana, D. K.; Bhardwaj, V.; Ravi Kumar, M. N. V. Estradiol loaded PLGA nanoparticles for oral administration: Effect of polymer molecular weight and copolymer composition on release behavior in vitro and in vivo. *J. Control. Release* **2007,** *119* (1), 77-85.

(76) Budhian, A.; Siegel, S. J.; Winey, K. I. Production of haloperidol-loaded PLGA nanoparticles for extended controlled drug release of haloperidol. *J. Microencapsul.* **2005,** *22* (7), 773-785.

(77) Danhier, F.; Ansorena, E.; Silva, J. M.; Coco, R.; Le Breton, A.; Préat, V. PLGA-based nanoparticles: An overview of biomedical applications. *J. Control. Release* **2012,** *161* (2), 505-522.

(78) Nie, H.; Khew, S. T.; Lee, L. Y.; Poh, K. L.; Tong, Y. W.; Wang, C.-H. Lysine-based peptide-functionalized PLGA foams for controlled DNA delivery. *J. Control. Release* **2009,** *138* (1), 64-70.

(79) Jin Yoon, J.; Ho Song, S.; Sung Lee, D.; Park, T. G. Immobilization of cell adhesive RGD peptide onto the surface of highly porous biodegradable polymer scaffolds fabricated by a gas foaming/salt leaching method. *Biomaterials* **2004,** *25* (25), 5613-5620.

(80) Pan, H.; Zheng, Q.; Yang, S.; Guo, X. Effects of functionalization of PLGA-[Asp-PEG]n copolymer surfaces with Arg-Gly-Asp peptides, hydroxyapatite nanoparticles, and BMP-2-derived peptides on cell behavior in vitro. *J. Biomed. Mater. Res., Part A* **2014,** *102* (12), 4526-4535.

(81) Pan, H.; Zheng, Q.; Guo, X.; Wu, Y.; Wu, B. Polydopamine-assisted BMP-2-derived peptides immobilization on biomimetic copolymer scaffold for enhanced bone induction in vitro and in vivo. *Colloids and Surfaces B: Biointerfaces* **2016,** *142*, 1-9.

(82) Dash, T. K.; Konkimalla, V. B. Poly-ϵ-caprolactone based formulations for drug delivery and tissue engineering: A review. *J. Control. Release* **2012,** *158* (1), 15-33.

(83) Zander, N. E.; Orlicki, J. A.; Rawlett, A. M.; Beebe, T. P. Surface-modified nanofibrous biomaterial bridge for the enhancement and control of neurite outgrowth. *Biointerphases* **2010,** *5* (4), 149-158.

(84) Venugopal, J.; Low, S.; Choon, A. T.; Kumar, A. B.; Ramakrishna, S. Electrospun-modified nanofibrous scaffolds for the mineralization of osteoblast cells. *J. Biomed. Mater. Res., Part A* **2008,** *85A* (2), 408-417.

(85) Zhang, H.; Lin, C.-Y.; Hollister, S. J. The interaction between bone marrow stromal cells and RGD-modified three-dimensional porous polycaprolactone scaffolds. *Biomaterials* **2009,** *30* (25), 4063-4069.

(86) Mondal, D.; Griffith, M.; Venkatraman, S. S. Polycaprolactone-based biomaterials for tissue engineering and drug delivery: Current scenario and challenges. *Int. J. Polym. Mater. Polym. Biomater.* **2016,** *65* (5), 255-265.

(87) Gunatillake, P. A.; Dandeniyage, L. S.; Adhikari, R.; Bown, M.; Shanks, R.; Adhikari, B. Advancements in the Development of Biostable Polyurethanes. *Polym. Rev. (Philadelphia, PA, U. S.)* **2018**, 1-27.

(88) Guelcher, S. A. Biodegradable Polyurethanes: Synthesis and Applications in Regenerative Medicine. *Tissue Eng., Part B* **2008,** *14* (1), 3-17.

(89) Marzec, M.; Kucińska-Lipka, J.; Kalaszczyńska, I.; Janik, H. Development of polyurethanes for bone repair. *Mater. Sci. Eng., C* **2017,** *80*, 736-747.

(90) Zuber, M.; Zia, F.; Zia, K. M.; Tabasum, S.; Salman, M.; Sultan, N. Collagen based polyurethanes—A review of recent advances and perspective. *Int. J. Biol. Macromol.* **2015,** *80*, 366-374.

(91) Adipurnama, I.; Yang, M.-C.; Ciach, T.; Butruk-Raszeja, B. Surface modification and endothelialization of polyurethane for vascular tissue engineering applications: a review. *Biomater. Sci.* **2017,** *5* (1), 22-37.

(92) Shelke, N. B.; James, R.; Laurencin, C. T.; Kumbar, S. G. Polysaccharide biomaterials for drug delivery and regenerative engineering. *Polym. Adv. Technol.* **2014,** *25* (5), 448-460.

(93) Kumbar, S.; Toti, U.; Deng, M.; James, R.; Laurencin, C.; Aravamudhan, A.; Harmon, M.; Ramos, D. Novel mechanically competent polysaccharide scaffolds for bone tissue engineering. *Biomedical Materials* **2011,** *6* (6), 065005.

(94) Wasupalli, G. K.; Verma, D. 3 - Polysaccharides as biomaterials. In *Fundamental Biomaterials: Polymers*; Thomas, S.; Balakrishnan, P.; Sreekala, M. S., Eds.; Woodhead Publishing: 2018; pp 37-70.

(95) Tardajos, M. G.; Cama, G.; Dash, M.; Misseeuw, L.; Gheysens, T.; Gorzelanny, C.; Coenye, T.; Dubruel, P. Chitosan functionalized poly-ε-caprolactone electrospun fibers and 3D printed scaffolds as antibacterial materials for tissue engineering applications. *Carbohydr. Polym.* **2018,** *191*, 127-135.

(96) Ressler, A.; Ródenas-Rochina, J.; Ivanković, M.; Ivanković, H.; Rogina, A.; Gallego Ferrer, G. Injectable chitosan-hydroxyapatite hydrogels promote the osteogenic differentiation of mesenchymal stem cells. *Carbohydr. Polym.* **2018,** *197*, 469-477.

(97) Lu, Y.; Li, M.; Li, L.; Wei, S.; Hu, X.; Wang, X.; Shan, G.; Zhang, Y.; Xia, H.; Yin, Q. High-activity chitosan/nano hydroxyapatite/zoledronic acid scaffolds for simultaneous tumor inhibition, bone repair and infection eradication. *Mater. Sci. Eng., C* **2018,** *82*, 225-233.

(98) Liao, Y.-H.; Jones, S. A.; Forbes, B.; Martin, G. P.; Brown, M. B. Hyaluronan: Pharmaceutical Characterization and Drug Delivery. *Drug Delivery* **2005,** *12* (6), 327-342.

(99) Larsen, N. E.; Pollak, C. T.; Reiner, K.; Leshchiner, E.; Balazs, E. A. Hylan gel biomaterial: Dermal and immunologic compatibility. *J. Biomed. Mater. Res.* **1993,** *27* (9), 1129-1134.

(100) Tomihata, K.; Ikada, Y. Crosslinking of hyaluronic acid with glutaraldehyde. *J. Polym. Sci., Part A: Polym. Chem.* **1997,** *35* (16), 3553-3559.

(101) Tomihata, K.; Ikada, Y. Crosslinking of hyaluronic acid with water-soluble carbodiimide. *J. Biomed. Mater. Res.* **1997,** *37* (2), 243-251.

(102) Gramlich, W. M.; Kim, I. L.; Burdick, J. A. Synthesis and orthogonal photopatterning of hyaluronic acid hydrogels with thiol-norbornene chemistry. *Biomaterials* **2013,** *34* (38), 9803-9811.

(103) Khetan, S.; Guvendiren, M.; Legant, W. R.; Cohen, D. M.; Chen, C. S.; Burdick, J. A. Degradation-mediated cellular traction directs stem cell fate in covalently crosslinked three-dimensional hydrogels. *Nat. Mater.* **2013,** *12*, 458.

(104) Frantz, C.; Stewart, K. M.; Weaver, V. M. The extracellular matrix at a glance. *J. Cell Sci.* **2010,** *123* (24), 4195.

(105) Notbohm, H.; Nokelainen, M.; Myllyharju, J.; Fietzek, P. P.; Müller, P. K.; Kivirikko, K. I. Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. *J. Biol. Chem.* **1999,** *274* (13), 8988-8992.

(106) Vázquez-Portalatín, N.; Kilmer, C. E.; Panitch, A.; Liu, J. C. Characterization of Collagen Type I and II Blended Hydrogels for Articular Cartilage Tissue Engineering. *Biomacromolecules* **2016,** *17* (10), 3145-3152.

(107) Sheu, M.-T.; Huang, J.-C.; Yeh, G.-C.; Ho, H.-O. Characterization of collagen gel solutions and collagen matrices for cell culture. *Biomaterials* **2001,** *22* (13), 1713-1719.

(108) Nam, K.; Kimura, T.; Kishida, A. Preparation and characterization of cross-linked collagen–phospholipid polymer hybrid gels. *Biomaterials* **2007,** *28* (1), 1-8.

(109) Zhang, X.; Xu, L.; Huang, X.; Wei, S.; Zhai, M. Structural study and preliminary biological evaluation on the collagen hydrogel crosslinked by γ-irradiation. *J. Biomed. Mater. Res., Part A* **2012,** *100A* (11), 2960-2969.

(110) Hovakimyan, M.; Guthoff, R. F.; Stachs, O. Collagen cross-linking: current status and future directions. *J. Ophthalmol.* **2012,** *2012*.

(111) Tirella, A.; Liberto, T.; Ahluwalia, A. Riboflavin and collagen: New crosslinking methods to tailor the stiffness of hydrogels. *Mater. Lett.* **2012,** *74*, 58-61.

(112) Heo, J.; Koh, R. H.; Shim, W.; Kim, H. D.; Yim, H.-G.; Hwang, N. S. Riboflavin-induced photo-crosslinking of collagen hydrogel and its application in meniscus tissue engineering. *Drug Delivery Transl. Res.* **2016,** *6* (2), 148-158.

(113) Bernard, M. P.; Myers, J. C.; Chu, M. L.; Ramirez, F.; Eikenberry, E. F.; Prockop, D. J. Structure of a cDNA for the pro. alpha. 2 chain of human type I procollagen. Comparison with chick cDNA for pro. alpha. 2 (I) identifies structurally conserved features of the protein and the gene. *Biochemistry* **1983,** *22* (5), 1139-1145.

(114) Taubenberger, A. V.; Woodruff, M. A.; Bai, H.; Muller, D. J.; Hutmacher, D. W. The effect of unlocking RGD-motifs in collagen I on pre-osteoblast adhesion and differentiation. *Biomaterials* **2010,** *31* (10), 2827-2835.

(115) Bosnakovski, D.; Mizuno, M.; Kim, G.; Takagi, S.; Okumura, M.; Fujinaga, T. Chondrogenic differentiation of bovine bone marrow mesenchymal stem cells (MSCs) in different hydrogels: Influence of collagen type II extracellular matrix on MSC chondrogenesis. *Biotechnol. Bioeng.* **2006,** *93* (6), 1152-1163.

(116) Wissink, M. J. B.; Beernink, R.; Pieper, J. S.; Poot, A. A.; Engbers, G. H. M.; Beugeling, T.; van Aken, W. G.; Feijen, J. Immobilization of heparin to EDC/NHS-crosslinked collagen. Characterization and in vitro evaluation. *Biomaterials* **2001,** *22* (2), 151-163.

(117) Roberts, S.; Dzuricky, M.; Chilkoti, A. Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Lett.* **2015,** *589* (19), 2477-2486.

(118) Wang, H.; Cai, L.; Paul, A.; Enejder, A.; Heilshorn, S. C. Hybrid elastin-like polypeptide-polyethylene glycol (ELP-PEG) hydrogels with improved transparency and independent control of matrix mechanics and cell ligand density. *Biomacromolecules* **2014,** *15* (9), 3421-8.

(119) Yuan, Y.; Koria, P. Proliferative activity of elastin- like- peptides depends on charge and phase transition. *J. Biomed. Mater. Res., Part A* **2015**.

(120) Fernandez-Colino, A.; Arias, F. J.; Alonso, M.; Rodriguez-Cabello, J. C. Self-organized ECM-mimetic model based on an amphiphilic multiblock silk-elastin-like corecombinamer with a concomitant dual physical gelation process. *Biomacromolecules* **2014,** *15* (10), 3781-93.

(121) Fernandez-Colino, A.; Arias, F. J.; Alonso, M.; Rodriguez-Cabello, J. C. Amphiphilic elastin-like block co-recombinamers containing leucine zippers: Cooperative interplay between both domains results in injectable and stable hydrogels. *Biomacromolecules* **2015,** *16* (10), 3389-98.

(122) Park, W. M.; Champion, J. A. Thermally triggered self-assembly of folded proteins into vesicles. *J. Am. Chem. Soc.* **2014,** *136* (52), 17906-9.

(123) Pinedo-Martín, G.; Santos, M.; Testera, A. M.; Alonso, M.; Rodríguez-Cabello, J. C. The effect of NaCl on the self-assembly of elastin-like block co-recombinamers: Tuning the size of micelles and vesicles. *Polymer* **2014,** *55* (21), 5314-5321.

(124) Hassouneh, W.; Zhulina, E. B.; Chilkoti, A.; Rubinstein, M. Elastin-like polypeptide diblock copolymers self-assemble into weak micelles. *Macromolecules* **2015,** *48* (12), 4183-4195.

(125) Han, W.; MacEwan, S. R.; Chilkoti, A.; Lopez, G. P. Bio-inspired synthesis of hybrid silica nanoparticles templated from elastin-like polypeptide micelles. *Nanoscale* **2015,** *7* (28), 12038-44.

(126) Su, R. S.; Kim, Y.; Liu, J. C. Resilin: Protein-based elastomeric biomaterials. *Acta Biomater.* **2014,** *10* (4), 1601-11.

(127) Renner, J. N.; Cherry, K. M.; Su, R. S.; Liu, J. C. Characterization of resilin-based materials for tissue engineering applications. *Biomacromolecules* **2012,** *13* (11), 3678-85.

(128) Li, L.; Tong, Z.; Jia, X.; Kiick, K. L. Resilin-like polypeptide hydrogels engineered for versatile biological functions. *Soft Matter* **2013,** *9* (3), 665-673.

(129) Lv, S.; Bu, T.; Kayser, J.; Bausch, A.; Li, H. Towards constructing extracellular matrix-mimetic hydrogels: An elastic hydrogel constructed from tandem modular proteins containing tenascin FnIII domains. *Acta Biomater.* **2013,** *9* (5), 6481-91.

(130) Vashi, A. V.; Ramshaw, J. A.; Glattauer, V.; Elvin, C. M.; Lyons, R. E.; Werkmeister, J. A. Controlled surface modification of tissue culture polystyrene for selective cell binding using resilin-inspired polypeptides. *Biofabrication* **2013,** *5* (3), 035005.

(131) Hsueh, Y. S.; Savitha, S.; Sadhasivam, S.; Lin, F. H.; Shieh, M. J. Design and synthesis of elastin-like polypeptides for an ideal nerve conduit in peripheral nerve regeneration. *Mater. Sci. Eng., C* **2014,** *38*, 119-26.

(132) Tjin, M. S.; Chua, A. W.; Ma, D. R.; Lee, S. T.; Fong, E. Human epidermal keratinocyte cell response on integrin-specific artificial extracellular matrix proteins. *Macromol. Biosci.* **2014,** *14* (8), 1125-34.

(133) Rahmany, M. B.; Van Dyke, M. Biomimetic approaches to modulate cellular adhesion in biomaterials: A review. *Acta Biomater.* **2013,** *9* (3), 5431-7.

(134) Lee, P.; Bax, D. V.; Bilek, M. M.; Weiss, A. S. A novel cell adhesion region in tropoelastin mediates attachment to integrin $\alpha_v\beta_5$. *J. Biol. Chem.* **2014,** *289* (3), 1467-77.

(135) Kim, Y.; Renner, J. N.; Liu, J. C. Incorporating the BMP-2 peptide in genetically-engineered biomaterials accelerates osteogenic differentiation. *Biomater. Sci.* **2014,** *2* (8), 1110-1119.

(136) Renner, J. N.; Kim, Y.; Liu, J. C. Bone morphogenetic protein-derived peptide promotes chondrogenic differentiation of human mesenchymal stem cells. *Tissue Eng. Part A* **2012,** *18* (23-24), 2581-9.

(137) Tao, H.; Wu, Y.; Li, H.; Wang, C.; Zhang, Y.; Li, C.; Wen, T.; Wang, X.; He, Q.; Wang, D.; Ruan, D. BMP7-based functionalized self-assembling peptides for nucleus pulposus tissue engineering. *ACS Appl. Mater. Interfaces* **2015,** *7* (31), 17076-87.

(138) Cai, L.; Dinh, C. B.; Heilshorn, S. C. One-pot synthesis of elastin-like polypeptide hydrogels with grafted VEGF-mimetic peptides. *Biomater Sci* **2014,** *2* (5), 757-765.

(139) Mulyasasmita, W.; Cai, L.; Hori, Y.; Heilshorn, S. C. Avidity-controlled delivery of angiogenic peptides from injectable molecular-recognition hydrogels. *Tissue Eng. Part A* **2014,** *20* (15-16), 2102-14.

(140) Verheyen, A.; Peeraer, E.; Lambrechts, D.; Poesen, K.; Carmeliet, P.; Shibuya, M.; Pintelon, I.; Timmermans, J. P.; Nuydens, R.; Meert, T. Therapeutic potential of VEGF and VEGF-derived peptide in peripheral neuropathies. *Neuroscience* **2013,** *244*, 77-89.

(141) Sanchez-Ferrero, A.; Mata, A.; Mateos-Timoneda, M. A.; Rodriguez-Cabello, J. C.; Alonso, M.; Planell, J.; Engel, E. Development of tailored and self-mineralizing citric acid-crosslinked hydrogels for in situ bone regeneration. *Biomaterials* **2015,** *68*, 42-53.

(142) Price, R.; Poursaid, A.; Cappello, J.; Ghandehari, H. In vivo evaluation of matrix metalloproteinase responsive silk-elastinlike protein polymers for cancer gene therapy. *J. Control. Release* **2015,** *213*, 96-102.

(143) Fonseca, K. B.; Maia, F. R.; Cruz, F. A.; Andrade, D.; Juliano, M. A.; Granja, P. L.; Barrias, C. C. Enzymatic, physicochemical and biological properties of MMP-sensitive alginate hydrogels. *Soft Matter* **2013,** *9* (12), 3283-3292.

(144) Kukreja, M.; Shiryaev, S. A.; Cieplak, P.; Muranaka, N.; Routenberg, D. A.; Chernov, A. V.; Kumar, S.; Remacle, A. G.; Smith, J. W.; Kozlov, I. A.; Strongin, A. Y. High-throughput multiplexed peptide-centric profiling illustrates both substrate cleavage redundancy and specificity in the MMP family. *Chem. Biol.* **2015,** *22* (8), 1122-33.

(145) Bracher, M.; Bezuidenhout, D.; Lutolf, M. P.; Franz, T.; Sun, M.; Zilla, P.; Davies, N. H. Cell specific ingrowth hydrogels. *Biomaterials* **2013,** *34* (28), 6797-803.

(146) Goetsch, K. P.; Bracher, M.; Bezuidenhout, D.; Zilla, P.; Davies, N. H. Regulation of tissue ingrowth into proteolytically degradable hydrogels. *Acta Biomater.* **2015,** *24*, 44-52.

(147) Yang, J.; Yao, M.-H.; Du, M.-S.; Jin, R.-M.; Zhao, D.-H.; Ma, J.; Ma, Z.-Y.; Zhao, Y.-D.; Liu, B. A near-infrared light-controlled system for reversible presentation of bioactive ligands using polypeptide-engineered functionalized gold nanorods. *Chem. Commun.* **2015,** *51* (13), 2569-2572.

(148) Kim, D.; Kim, S. K.; Valencia, C. A.; Liu, R. Tribody: Robust self-assembled trimeric targeting ligands with high stability and significantly improved target-binding strength. *Biochemistry* **2013,** *52* (41), 7283-94.

(149) Shah, N. H.; Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem Sci* **2014,** *5* (1), 446-461.

(150) Wood, D. W.; Camarero, J. A. Intein applications: From protein purification and labeling to metabolic control methods. *J. Biol. Chem.* **2014,** *289* (21), 14512-9.

(151) Ren, W.; Ji, A.; Ai, H. W. Light activation of protein splicing with a photocaged fast intein. *J. Am. Chem. Soc.* **2015,** *137* (6), 2155-8.

(152) Jung, D.; Sato, K.; Min, K.; Shigenaga, A.; Jung, J.; Otaka, A.; Kwon, Y. Photo-triggered fluorescent labelling of recombinant proteins in live cells. *Chem. Commun.* **2015,** *51* (47), 9670-9673.

(153) Bocker, J. K.; Friedel, K.; Matern, J. C.; Bachmann, A. L.; Mootz, H. D. Generation of a genetically encoded, photoactivatable intein for the controlled production of cyclic peptides. *Angew. Chem. Int. Ed. Engl.* **2015,** *54* (7), 2116-20.

2. REDOX-RESPONSIVE HYDROGELS VIA ONE-POT THIOL-ENE ADDITION AND THIOL OXIDATION WITH DECOUPLED DEGRADATION BEHAVIORS AND MECHANICAL PROPERTIES

This chapter consists of a manuscript by Lin C-Y and Liu JC, prepared for submission in 2019.

2.1 Abstract

Disulfide-crosslinked hydrogels have been widely used for biological applications because of their ability to degrade in response to redox stimuli. However, the degradability often depends on the polymer concentration, which also influences the hydrogel mechanical properties. Here we describe a one-pot crosslinking approach utilizing both thiol-ene addition through divinyl sulfone (DVS) and thiol oxidation promoted by ferric EDTA. The thiol-ene addition results in non-reducible thioether bonds, whereas the thiol oxidation forms reducible disulfide bonds. By varying the DVS concentration used, the ratio between the thioether and disulfide bonds in the hydrogel is modulated and and therefore the degradability can be controlled. Hydrogels were created that had similar shear moduli but different degradation rates. These gels also had different release rates of encapsulated dextran when exposed to 10 μM glutathione. Fibroblast encapsulation suggested good cytocompatibility of the crosslinking reactions. This work shows the potential of combining DVS and ferric EDTA to create thiol-crosslinked hydrogels as redox-responsive drug delivery vehicles and tissue engineering scaffolds with tunable degradability.

2.2 Introduction

Hydrogels have generated great interest as scaffolds for tissue engineering applications due to their high water content, good biocompatibility, and tunable mechanical properties.[1-2] The properties of a hydrogel largely depend not only on the choice of the material but also on the crosslinking reaction. Synthetic polymers, such as poly(ethylene glycol) (PEG) and poly(hydroxyethyl methacrylate) (PHEMA), or natural polymers, including alginate and

collagen, have been used for hydrogel fabrication.[3] The available crosslinking approaches depend on the material used.[4]

Many different crosslinking mechanisms have been used to create hydrogels. For example, crosslinking based on free radicals is often applied to synthetic polymers and requires the use of initiators. However, depending on the type of initiators, the biocompatibility of the resulting hydrogel can be compromised.[5-7] Physical crosslinking is mostly achieved through electrostatic or hydrophobic interactions but affords less control in tuning the mechanical properties of the crosslinked hydrogels.[8-9] Chemical crosslinking can be used with both synthetic and natural polymers as long as the polymers have the necessary functional groups for the reaction.

Carboxylic acids and amines are commonly-used functional groups for chemical crosslinking because of their abundance in both synthetic and natural polymers. One robust carboxylic acid-reactive crosslinker is 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) that links these two functional groups by forming an amide bond. Despite its wide application in hydrogel crosslinking, the acidic pH (4.0-6.0) required for optimal efficiency prevents the usage of EDC for fabrication of cell-encapsulated hydrogels. Hydrogels crosslinked with high concentrations of EDC have been shown to have moderate cytotoxicity and to elicit non-native cellular interactions.[10-11] N-hydroxysuccinimide esters (NHS esters) can be used to achieve amine-to-amine crosslinking at pH 7-9. However, it is difficult to directly incorporate biomolecules, such as growth factors, into hydrogel formation with NHS ester crosslinking due to the abundance of amine groups on those molecules.[12]

Chemical crosslinking utilizing functional groups that are not commonly present on biomolecules has also been developed. For example, hydrogels crosslinked by strain-promoted azide-alkyne cycloaddition or by the Diels-Alder reaction have been shown to have good biorthogonality and cytocompatibility.[13] However, natural polymers need to be functionalized with the corresponding reactive groups before being used for hydrogel formation.[14-15]

Thiol-based crosslinking is another promising choice for fabricating biocompatible hydrogels. Thiols can be incorporated into synthetic polymers[16] and can also be found in cysteine residues in protein-based materials. Therefore, thiol-based crosslinking can easily

be utilized to form hydrogels of both kinds of polymers. In addition, cysteine is the least abundant amino acid and is often found in buried sites of biomolecules.[17] Therefore, undesired crosslinking between the hydrogel material and the encapsulated biomolecules or cells is less likely to occur. Thiol-ene addition is one of the most commonly-used thiol crosslinking chemistries.[18] Maleimide and divinyl sulfone (DVS) are successful examples of compounds used to create thiol-ene crosslinked hydrogels with good biocompatibility.[13] Thiols can also form disulfide bonds through oxidation, and oxidants, including periodate, hydrogen peroxide, and ferricyanide, can be used to accelerate the oxidation.[19-20]

Disulfide bonds can be reversibly cleaved by reducing agents. This property can be harnessed to make the hydrogels that are responsive to changes in redox condition. Kar *et al.* synthesized two PEG derivatives with or without disulfide bonds in the backbone.[21] The hydrogel degradation rates in response to various reducing agents were modulated by changing the ratio between the two PEG variants while keeping the total PEG concentration constant. Faster degradation was observed with hydrogels having higher ratios of the disulfide-containing PEG variant. In experiments performed by Yang et al., disulfide bond formation was utilized directly as the crosslinking method for PEG hydrogel formation.[22-23] The disulfide-crosslinked hydrogels were responsive to glutathione (GSH), a reducing agent commonly found in the extracellular space, at a physiological concentration of 10 μM. The authors demonstrated that these redox-responsive hydrogels could be used for cell encapsulation or as drug delivery vehicles. When bone morphogenetic protein-2 (BMP-2) was encapsulated during the hydrogel formation, its bioactivity was not compromised and was able to promote bone healing in vivo. However, the degradation behavior of these hydrogels depended on the PEG concentration in the hydrogels and therefore could not be tuned independently from the mechanical properties of the hydrogels.

In this work, we show that the degradation rate of thiol-crosslinked PEG hydrogels can be tuned independently of the mechanical properties by using thiol-ene addition and thiol oxidation simultaneously. DVS was used to crosslink thiol-functionalized PEG by forming non-reducible thioether bonds, and ferric ethylenediaminetetraacetic acid (ferric EDTA) was chosen as the oxidant to facilitate disulfide bond crosslinking (Figure 2.1). Hydrogels crosslinked with different ratios of DVS to ferric EDTA were characterized for their gelation times and rheological properties. Swelling behavior and the release of

encapsulated dextran profile were investigated under physiological reducing conditions, and cytocompatibility of cells encapsulated within the dual-crosslinked hydrogels was evaluated. The results demonstrate the potential of a one-pot DVS and ferric EDTA reactions as a new crosslinking approach to generate thiol-crosslinked hydrogels with independently tunable mechanical properties and degradation behavior.

2.3 Experimental Methods

2.3.1 Materials

All materials and reagents were purchased from Sigma-Aldrich (St. Louis, MO) or AMRESCO, Inc. (Radnor, PA) unless otherwise specified. Four-arm PEG thiol (PEG-SH, Mn 10000 g mol^{-1}) was purchased from JenKem Technology USA (Plano, TX). Milli-Q water was used in all experiments except for protein expression and dialysis where reverse-osmosis purified water was used.

2.3.2 Hydrogel Formation

The PEG-SH solution was prepared at 12 wt% in phosphate buffered saline (PBS) at pH 7.4. The DVS solution was prepared at 80 mM in PBS at pH 7.4. To prepare the solution of ferric EDTA, ferric sulfate was dissolved in a 0.5 M EDTA solution. The mixture was adjusted to pH 7.4, and 10x PBS and MQ were added to make final concentrations of 100 mM ferric ion, 110 mM EDTA, and 1x PBS at pH 7.4. To make a hydrogel, the ferric EDTA solution was first mixed with the DVS solution then with the PEG-SH solution. The final concentration in the hydrogel was 5 wt% PEG-SH with 50 mM ferric ion and 8 mM DVS. To make hydrogels crosslinked with lower DVS concentrations, the DVS solution was first diluted with PBS before mixing with the ferric EDTA solution. After mixing the solutions, the mixture was incubated at 37 °C in a humidified environment until gelation. The gelation time was determined with three independent samples by the tube inversion test, which is widely used for PEG-based hydrogels.[24-25]

2.3.3 Rheological Characterization

To prepare hydrogels for rheological analysis, 170 μL of pre-polymer solution was dispensed on a poly(tetrafluoroethylene)-coated slide (Tekdon, Inc., Myakka City, FL,

Slide ID: 1-200). The hydrogels were incubated on the slides at 37 °C for 24 h in a humidified environment. An AR-G2 rheometer (TA Instruments, New Castle, DE) was used with a 20 mM plate geometry. The bottom plate was maintained at 37 °C during the tests. The hydrogel and slide were allowed to reach temperature equilibrium with the plate for 1 min before the rheological test started. Frequency sweeps were performed from 0.1 to 10 Hz at 1% strain. Strain sweeps were performed from 0.1-10% strain at 1 Hz.

2.3.4 Swelling and Degradation Analysis

To prepare hydrogels for swelling and degradation experiments, 50 uL of pre-polymer solution was dispensed onto an acid-washed 12-mm coverslip. The hydrogel was then incubated at 37 °C for 24 h in a humidified environment. After incubation, the hydrogel was rinsed extensively in MQ and was lyophilized. The dry weight of the lyophilized hydrogel (W_d) was measured, and the hydrogel was incubated at 37 °C in 1 mL of PBS or PBS with 10 μM GSH. At predetermined time points, the swollen weight of the hydrogel (W_s) was measured. The incubation solution was changed every 24 h to maintain a constant GSH concentration. The swelling ratio (S) was calculated using the equation $S = 100 \times (W_s-W_d)/W_d$. Complete degradation of the hydrogel was determined when no remaining hydrogel fragments were observed when measuring W_s.

2.3.5 Quantification of *in vitro* Release of Dextran

Fluorescein isothiocyanate-dextran (FITC-dextran) (average MW 150 kDa or 2 MDa) was used as a model molecule for *in vitro* release experiments. The FITC-dextran stock solution (1xPBS, pH 7.4) was mixed with the PEG-SH solution then mixed with the DVS/ferric EDTA solution. The final FITC-dextran concentration was 1 mg/mL. To form hydrogels, 75 μL of the gel solution was pipetted into a syringe mold with a diameter of 4.5 mm and incubated at 37 °C for 24 h in a humidified environment.

To detect FITC-dextran release, the FITC-dextran-encapsulated hydrogels were transferred to 1.5 mL PBS with or without 10 μM GSH. The gels were incubated at 37 °C with gentle shaking at 100 rpm. At predetermined time points, the incubation solution was collected and replenished with fresh solution. The FITC-dextran concentration in the collected incubation solution was determined by the FITC fluorescence intensity. The

fluorescence reading was performed on a SpectraMax M2^e instrument (Molecular Devices, Downingtown, PA) with an excitation wavelength at 492 nm and an emission wavelength at 518 nm with a cutoff at 515 nm. FITC-dextran standards in PBS or in PBS with 10 μM GSH were prepared to convert fluorescence readings to mass concentrations.

2.3.6 Cell Encapsulation and Cytocompatibility

To investigate the cytocompatibility of the crosslinking chemistry, NIH/3T3 mouse fibroblasts were encapsulated in hydrogels crosslinked with 8 mM DVS. For encapsulation, cells were resuspended in PEG-SH solution first, and the cell resuspension was then mixed with the ferric EDTA and DVS solutions in a μ-Slide Angiogenesis (ibidi GmbH, Martinsried, Germany with a final volume of 10 μL). The final cell concentration was 10^6 cells/mL. The cell-encapsulated hydrogels were incubated at 37 °C in a humidified environment for 2 h before the culture medium was added to the hydrogels. Cell-encapsulated hydrogels were cultured at 37 °C and 5% CO_2 in high-glucose Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum (Lonza, Basel, Switzerland, Cat. #: 14-501F) and 100 U/mL penicillin-streptomycin (Thermo-Fisher Scientific, Waltham, MA). Medium was changed every other day.

Cell viability was assessed 1 and 5 days after encapsulation using a LIVE/DEAD^{TM} Viability/Cytotoxicity Kit (Thermo-Fisher Scientific). On the day of the assay, the hydrogels were rinsed in pre-warmed PBS for 10 min for 3 times and then incubated with 2 μM calcein acetoxymethyl ester and 4 μM ethidium homodimer-1 in PBS for 30 min at 37 °C. The solution was changed to fresh PBS before the hydrogels were imaged on a Nikon Ti-E C1+ confocal (Nikon, Tokyo, Japan). The z-axis scanning was performed over an 800-μm thickness with a step size of 50 μm.

2.4 Results and Discussion

2.4.1 Crosslinker Formulation Independent Hydrogel Formation and Shear Modulus

To create thiol-crosslinked hydrogels with tunable degradability independent of mechanical properties, we utilized thiol-ene addition and thiol oxidation in a one-pot reaction as illustrated in Figure 2.1. Thioether bonds formed between PEG-SH and DVS

are not reducible by reducing agents such as GSH and maintain hydrogel integrity under reducing conditions. On the other hand, disulfide bonds formed between PEG-SH through oxidation promoted by ferric EDTA can be reduced and therefore provide hydrogel degradability in a reducing environment.

The gelation times of 5 wt% PEG-SH crosslinked with 50 mM ferric ion and various concentration of DVS were determined by tube inversion tests. Unless otherwise specified, the concentration of ferric ion was kept constant at 50 mM. As shown in Figure 2.1D, no flow upon tube inversion was observed with 2-8 mM DVS after 10 min of incubation at 37 °C. When the test was performed with shorter intervals of 1 min, hydrogels crosslinked with 8 mM DVS did not flow after 4 min of incubation. Five minutes and ten minutes were required for gels with 4 and 2 mM DVS, respectively, to resist flowing during inversion.

The color of the hydrogel mixture changed during gelation (Figure 2.1D). Immediately after mixing the PEG-SH solution with the DVS/ferric EDTA mixture, the color of the solution turned from light brown to a deep purple. Over time, the color changed to dark brown and then back to light brown. This change in the solution color roughly correlated with the extent of gelation. The gel crosslinked with 8 mM DVS had the shortest gelation time and after 10 min of incubation had a lighter brown color than the other two formulations. When the hydrogels were allowed to further crosslink for another 20 min, all three gels showed a similar light brown color.

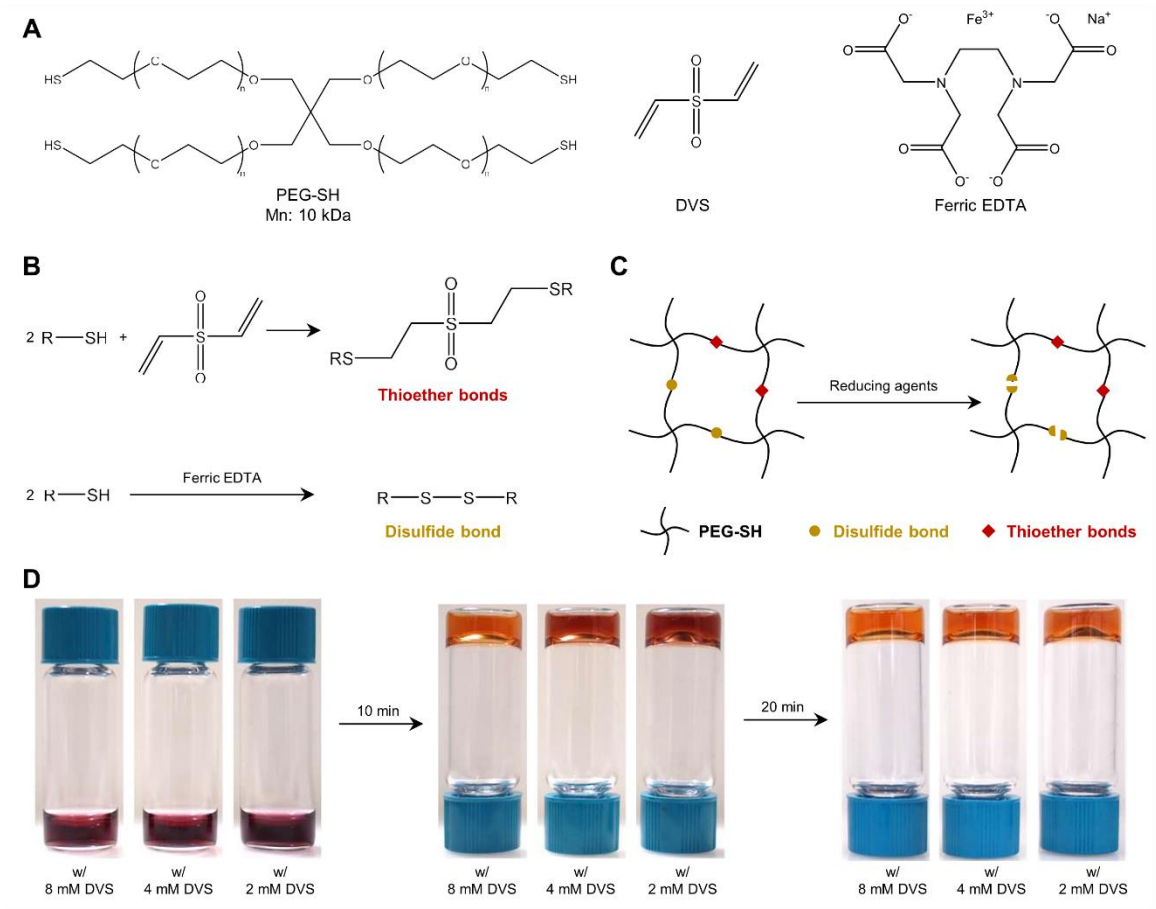


Figure 2.1. One-pot crosslinking through thiol-ene addition and thiol oxidation. (A) Chemical structures of four-arm poly(ethylene glycol) thiol (PEG-SH), divinyl sulfone (DVS), and sodium ferric ethylenediaminetetraacetic acid (ferric EDTA). (B) Formation of thioether bonds between PEG-SH and DVS through thiol-ene addition and formation of disulfide bonds between two PEG-SH through thiol oxidation promoted by ferric EDTA. (C) Illustration of the hydrogel network with both thioether and disulfide bonds. In the presence of reducing agents, disulfide bonds are reduced and broken. On the other hand, thioether bonds are not reducible and remain intact in the network. (D) Hydrogels crosslinked with 50 mM ferric EDTA with different DVS concentrations showed similar gelation times as determined by tube inversion test. All formulations formed hydrogels within 10 min at 37 °C. Color change was observed during the gelation process. The color was initially deep purple and turned into dark brown then light brown. The rate of this color change depended on the crosslinking formulation.

To quantify the mechanical properties of the hydrogels, frequency and strain sweeps were performed using a rheometer. Hydrogels were crosslinked with 0, 2, 4, or 8 mM DVS and incubated on poly(tetrafluoroethylene)-coated slides at 37 ° for 24 h in a

humidified incubator before the rheological tests. The results of the frequency and strain sweeps are shown in Figure 2.2. All hydrogels showed linear trends in storage moduli (G') over the tested frequency range (0.1 to 10 Hz) at 1% strain and over the tested strain range (0.1% to 10%) at 1 Hz. At 1% strain at 1 Hz, the average G' for hydrogels crosslinked with 0, 2, 4, and 8 mM DVS were 357.5 ± 80.8, 364.3 ± 33.1, 494.2 ± 65.8, and 463.5 ± 91.0 Pa, respectively. ANOVA and Tukey's *post hoc* tests were performed, and the average G' of the 0- and 2-mM-DVS hydrogels were statistically lower than that of the 4-mM-DVS hydrogel ($p < 0.05$).

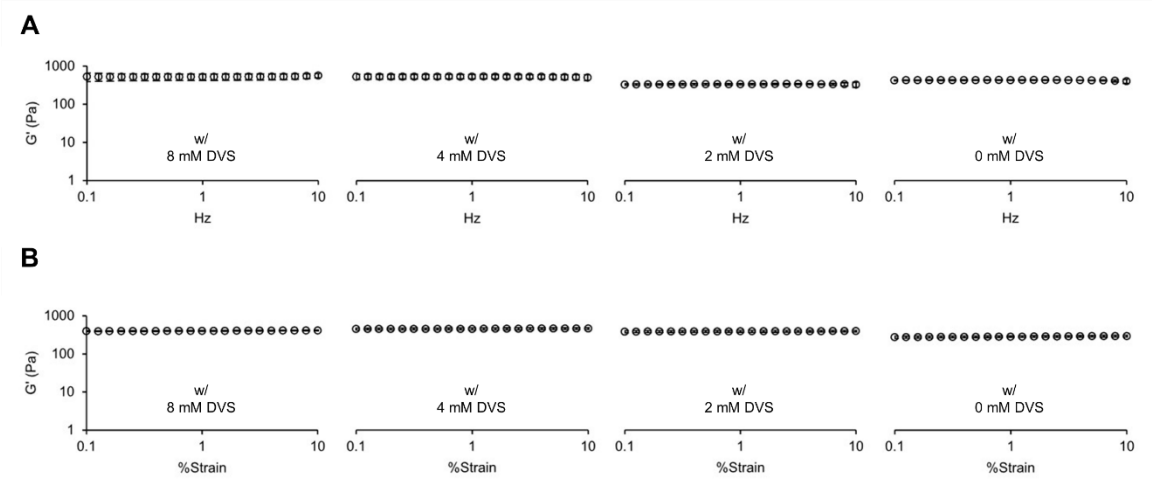


Figure 2.2. Similar storage moduli of hydrogels crosslinked with 0, 2, 4, and 8 mM DVS were observed in (A) frequency and (B) strain sweeps. All hydrogels displayed linear viscoelastic regions from 0.1 to 10 Hz at 1% strain and from 0.1 to 10% strain at 1 Hz. Data (n = 3) are presented as the average ± the standard deviation.

Thiol groups have generated great interest as a reactive group for hydrogel crosslinking. Because14 of its near-neutral pKa (~8) and high nucleophilicity, crosslinking via thiols can be performed under mild, physiological conditions.[15] Moreover, thiols are the least abundant functional group found in proteins and are usually not exposed on protein surfaces.[17] Therefore, thiol-based crosslinking is less likely to negatively affect biomolecules or cells encapsulated within the hydrogels due to unwanted side reactions.[26-27]

Two common mechanisms for thiol-based crosslinking are thiol-ene addition and disulfide formation through oxidation. Thiol-ene addition is often used because of its fast

reaction kinetics.[28] The reaction rate is largely dependent on the choice of the alkene-containing compound. When maleimides are used, crosslinking can be complete within seconds.[29-30] This fast crosslinking, however, results in inhomogeneous hydrogel network structures. DVS, on the other hand, has a slower reaction rate with thiols. The gelation time of thiol-DVS crosslinking depends on the polymer type and concentration but is generally on the order of minutes.[31-33] The gelation time from thiol oxidation crosslinking depends on whether any oxidants are used and the type of oxidant used.[19] The gelation times of our hydrogels were in the range of minutes, which is similar to the thiol-DVS crosslinking time, despite the fact that both crosslinking through thiol-DVS and thiol oxidation occurred in one pot. In our system, the gelation time was 25 min when no DVS and only 50 mM ferric EDTA was used for crosslinking, and no gelation was observed at 24 h when neither ferric EDTA nor DVS was used. Note that the gelation time was 8 min when no ferric EDTA and only 8 mM DVS was used. Because of the slower reaction rate and gelation time of ferric EDTA, we expect that thiol groups predominantly react with DVS in the early stage of gelation. Since DVS was the limiting reagent in all crosslinking conditions in this work, crosslinking through thiol oxidation could take place after most of the DVS had reacted.

A color change was observed during the gelation process in our system. Whereas the PEG-SH solution was clear, the ferric EDTA solution was light brown as expected because of the presence of ferric ions. Upon mixing the ferric EDTA solution with the DVS solution (which was also clear), only color change from dilution was observed. Interestingly, after mixing the PEG-SH and the DVS/ferric EDTA solutions, a dark purple color was observed. We further investigated this color change by mixing the PEG-SH solution with only ferric EDTA and still observed the purple color, and this result suggests that formation of the purple color does not require the presence of DVS. Transition metals have been known to show different colors depending on their oxidation or coordination states.[34] It has also been shown that thiolates can form coordination complexes with iron.[35] Thus, we hypothesize that the color change into purple was a combined result of the reduction of ferric to ferrous ions and the coordination between thiolates on PEG-SH and the iron ion coordination center. The color of the hydrogel eventually changed back to the light brown color of the ferric EDTA solution as gelation progressed. This result supports our hypothesis that thiolates were involved in the formation of the purple color since, at

the end of gelation, most thiolates had reacted with DVS or formed disulfides. Thus, the observed color change has the potential to be a colorimetric method to determine the extent of gelation in our system.

The rheological analysis showed that hydrogels crosslinked with different concentrations of DVS had similar storage moduli. Despite the statistical difference in storage moduli between the 2-mM- and 4 -mM-DVS hydrogels, the difference was within 27% of each other. The mechanical properties of a hydrogel depend on several factors including the polymer concentration and the crosslinking mechanism.[36] Because the polymer concentration was kept constant at 5 wt% for all of our hydrogels, the gels differed only in crosslinking mechanism. Specifically, the ratio between crosslinking through thiol-ene addition and through disulfide formation was varied based on the concentration of DVS (the limiting reagent) used in hydrogel formation. To minimize the change in mechanical properties from crosslinking through thiol-ene addition, DVS was chosen as the alkene-containing compound due to its relatively small size. As a result, the crosslinking junctions formed with DVS will be similar in length as the disulfide-crosslinked junctions in the hydrogel network. A similar strategy was used to make hydrogels that were crosslinked with different mechanisms but had similar mechanical properties.[37] In addition, hydrogel stiffness could still be changed independently by varying the polymer concentration even when multiple crosslinkers were involved.[38] Therefore, we expect that the stiffness of our hydrogels can be tuned by changing the polymer concentration and proportionally scaling the amount of DVS and ferric EDTA.

2.4.2 Tunable Swelling and Degradation of the Hydrogels with Similar Mechanical Properties

To test if hydrogel responsiveness to a reducing environment depends on the DVS concentration used for crosslinking, hydrogels crosslinked with 2 to 8 mM DVS were incubated at 37 °C in PBS or in PBS with 10 μM GSH. Under non-reducing conditions (i.e., PBS), all hydrogels showed similar swelling behavior regardless of the DVS concentration used for crosslinking (Figure 2.3). An equilibrium swelling ratio of ~22 was reached within 24 h and maintained up to 248 h. When hydrogels were incubated in reducing conditions with 10 μM GSH, different swelling behavior was observed. Hydrogels crosslinked with 6 or 8 mM DVS showed stable swelling ratios that were

slightly higher than those in PBS, and the gels remained intact up to 248 h. On the other hand, hydrogels crosslinked with 2 or 4 mM DVS swelled more significantly in the presence of 10 μM GSH than in PBS alone. The swelling ratio of 2-mM DVS gel reached a maximum of 60 at 96 h and decreased afterwards. After 120 h of incubation, the 2-mM DVS hydrogel disintegrated completely, and no further measurements were made. The 4-mM DVS hydrogel reached its highest swelling ratio of 49 after 175 h and showed a slow decreasing trend afterwards. However, complete degradation of the hydrogel was not observed at 248 h.

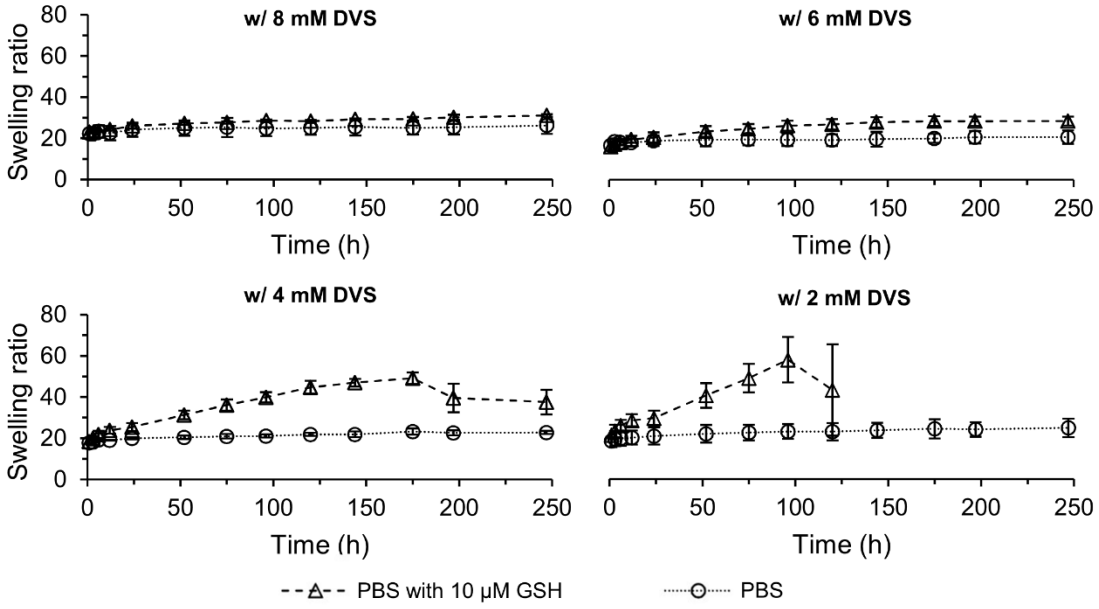


Figure 2.3. Degradation in 10 μM GSH increased with hydrogels crosslinked with lower DVS concentrations. All hydrogels showed similar swelling ratios when incubated in PBS. When incubated in PBS with 10 μM GSH, hydrogels crosslinked with 6 and 8 mM DVS showed no degradation, but swelling ratios were slightly higher than when incubated in PBS. On the other hand, hydrogels crosslinked with 2 and 4 mM DVS degraded and showed significantly higher swelling ratios compared those when incubated in PBS. The 2-mM DVS hydrogel disintegrated completely after 120h of incubation with 10 μM GSH. Data (n = 3) are presented as the average ± the standard deviation.

Degradability is an important hydrogel property in biomedical applications. Complete degradation with a tunable rate is often desired in drug delivery vehicles or tissue engineered scaffolds.[39-40] Hydrogels with disulfide moieties in the network show good degradability through the reduction of the disulfide bond by various reducing agents

including cysteine, dithiothreitol, and tris(2-carboxyethyl)phosphine.[21] Glutathione, a more biologically relevant reducing agent, can trigger hydrogel degradation at physiological concentrations by reduction of disulfides.[21-23] The glutathione concentration in cells ranges from 0.5 to 10 mM, and the extracellular concentration of glutathione ranges from 2 to 20 μM.[41]

Kar *et al.* synthesized a PEG derivative with a disulfide moiety to create hydrogels that were degradable by 0.5 mM GSH within 24 h of incubation.[21] Yang and coworkers used hydrogen peroxide as an oxidant to make PEG hydrogels crosslinked through disulfide bonds and showed that the hydrogels degraded in 10 μM GSH.[22-23] By changing the PEG concentration in the hydrogel, they tuned the degradation time from ~3 days for the 3 wt% hydrogel and ~2 weeks for the 10 wt% hydrogel. However, the storage moduli between their 3 wt% and 10 wt% gels differed by about two orders of magnitude. On the other hand, our hydrogels crosslinked with DVS and ferric EDTA showed similar storage moduli (Figure 2.2) but different degradation behavior in 10 μM GSH (Figure 2.3). The consistency in storage moduli was achieved by keeping the polymer concentration and the number of thiol groups the same, and the degradation was tuned by the ratio of ferric EDTA to DVS. Since both ferric EDTA and DVS crosslinked through thiols and had the same crosslinking stoichiometry, the number of crosslinking junctions in the hydrogel network remained the same. As a result, the increased number of crosslinks with DVS did not significantly change the stiffness of the hydrogel but drastically increased the number of non-degradable crosslinks and hence increased the time needed to completely degrade the hydrogel.

2.4.3 Increased FITC-Dextran Release in the Presence of Glutathione

FITC-dextran was used a model molecule for the release of encapsulated cargo from the hydrogels. The cumulative release profiles of both 150 kDa and 2 MDa FITC-dextran from hydrogels crosslinked with 2 to 8 mM DVS in non-reducing (PBS) or reducing (PBS with 10 μM GSH) conditions are shown in Figure 2.4. For both 150 kDa and 2 MDa FITC-dextran, more dextran molecules were released from all hydrogels, regardless of the DVS concentration used for crosslinking, in reducing conditions versus non-reducing conditions at all time points. However, the difference in the amount of FITC-dextran released between

the two conditions increased as the DVS concentration used for crosslinking decreased. For 150 kDa FTIC-dextran, the differences in the cumulative release between reducing and non-reducing conditions were 2.4%, 6.7%, 11%, and 13% for hydrogels crosslinked with 8, 6, 4, and 2 mM DVS after 120 h of incubation, respectively. For 2 MDa FITC-dextran, the difference in the release amount between the reducing and the non-reducing environments became more significant as the DVS concentration for the crosslinking decreased from 8 to 2 mM. After 120 h of incubation, the differences in the cumulative release between reducing and non-reducing conditions were 2.1%, 13%, 27%, and 45% for hydrogels crosslinked with 8, 6, 4, and 2 mM DVS, respectively.

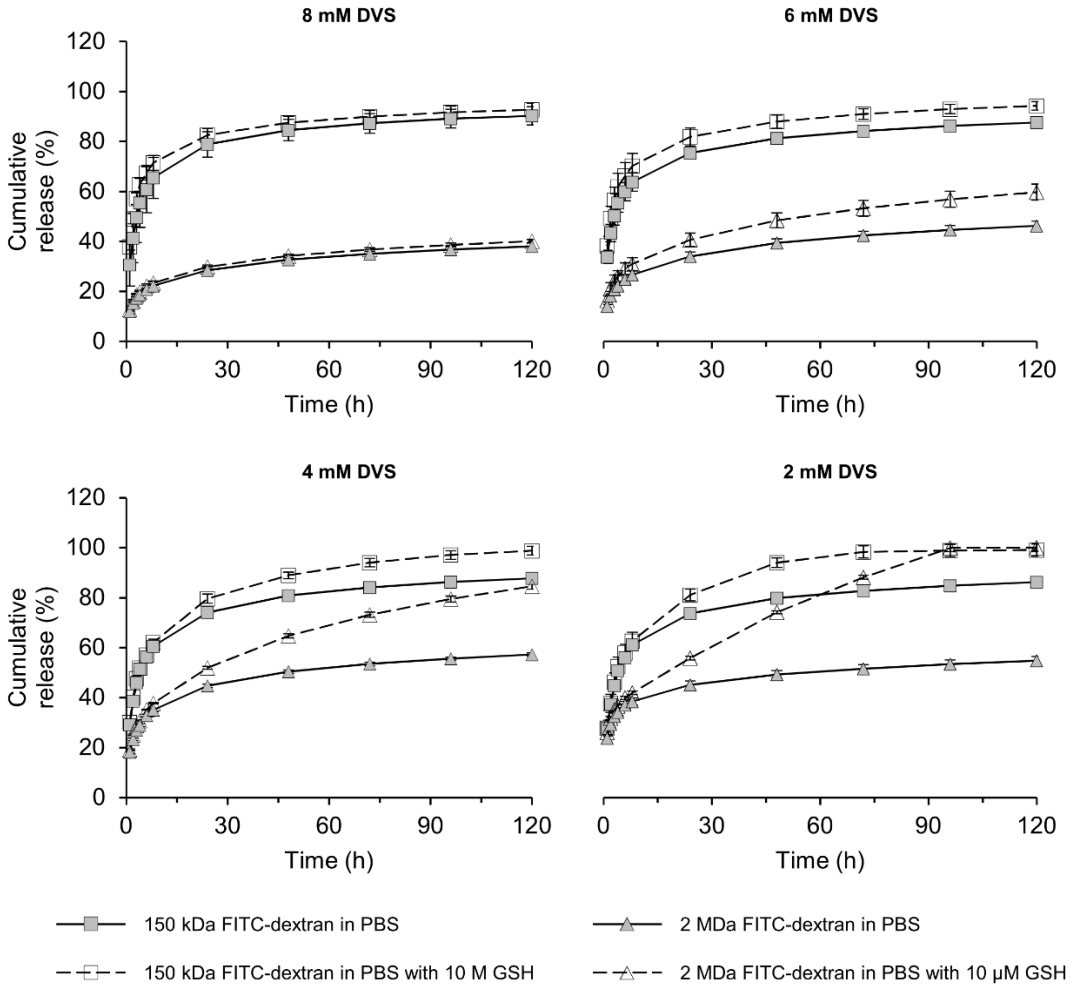


Figure 2.4. Faster and higher release of encapsulated 150 kDa or 2 MDa FTIC-dextrans from the 2- and 4-mM DVS hydrogels in PBS with 10 μM GSH. On the other hand, release profiles from hydrogels crosslinked with 6 and 8 mM DVS were similar when incubated in either PBS only or PBS with 10 μM GSH. Data (n = 3) are presented as the average ± the standard deviation.

Designing redox-responsive vehicles is a major aspect of targeted drug delivery because of the varying GSH concentrations found in different tissue types or between healthy and cancerous tissues.[42] Disulfides have been harnessed to confer such redox-responsiveness to materials, and previous studies have demonstrated GSH-sensitive release kinetics with disulfide-containing hydrogels.[22-23, 43] In general, higher release rates were observed with hydrogels undergoing faster degradation. This trend was also observed with the release profiles of encapsulated FITC-dextran from our dual-crosslinked hydrogels. Hydrogels crosslinked with 2 mM DVS showed both the fastest degradation and the fastest release of FITC-dextran (Figure 2.2 and 2.3). On the other hand, hydrogels crosslinked with 8 mM DVS showed no significant degradation by GSH as indicated by the constant swelling ratio (Figure 2.3) and showed nearly identical release profiles with or without GSH (Figure 2.4). Also note that both the 8- and 6-mM DVS hydrogels showed slightly higher amounts of FITC-dextran released in 10 μM GSH than in PBS. Since both hydrogels also showed slightly higher swelling ratios when incubated in 10 μM GSH, the more swollen and looser network structures could allow faster diffusion of the encapsulated FITC-dextran.

We note that ≥80% of the encapsulated 150 kDa FITC-dextran was released regardless of the extent of DVS crosslinking and the presence of GSH. To understand this finding, we compared the mesh size of our hydrogels to the hydrodynamic radius of FITC-dextran. The mesh size was estimated with the Flory-Rehner theory using volumetric swelling as described previously.[44-45] For hydrogels crosslinked with 8 and 6 mM DVS, the estimated mesh size was ~30 nm in both 10 μM GSH and PBS. For hydrogels crosslinked with 4 mM DVS, the mesh size increased to a maximum of ~52 nm after 175 h of incubation in 10 μM GSH. The maximum mesh size of the 2-mM DVS hydrogel was ~58 nm at 96 h of incubation with 10 μM GSH. On the other hand, the reported hydrodynamic radius of FITC-dextran with a molecular weight of 150 kDa is 8.25 nm.[46] When the mesh size of the hydrogel network is larger than the hydrodynamic radius of the encapsulated molecules, the diffusion of the molecules is less sterically hindered. However, it has been shown that diffusion of molecules within a hydrogel was influenced by changes in mesh size even when the mesh sizes were larger than the hydrodynamic radii of the molecules.[45] In our hydrogels, the release of 150 kDa FITC-dextran was also influenced

by the increasing mesh size over the course of degradation, especially with the hydrogels crosslinked with 2 and 4 mM DVS. On the other hand, the hydrodynamic radius of 2 MDa dextran is around 26.89 nm.[47] This similar radius to the estimated hydrogel mesh size resulted in the low release amount in the hydrogel crosslinked with 8 mM or 6 mM DVS in both the reducing and the non-reducing environment as well as in the early time points from 4- and 2-mM DVS crosslinked hydrogels. As the hydrogels started to be degraded by GSH, the increase in the mesh size accelerated the release of 2 MDa FITC-dextran.

2.4.4 DVS and Ferric EDTA Crosslinking Was Cytocompatible

NIH/3T3 mouse fibroblasts were used to evaluate the cytocompatibility of crosslinking with ferric EDTA and DVS. Cells were encapsulated at 10^6 cells/mL in hydrogels crosslinked with 50 mM ferric ion and 8 mM DVS, which is the highest DVS concentration used in this work. After 1 and 5 days of encapsulation, cell viability was assessed by performing a live/dead assay. As shown in Figure 2.5, most cells were viable on both days 1 and 5. In addition, the majority of the cells were distributed uniformly in the z-direction of the hydrogels. This result suggests that the crosslinking rate of our approach was fast enough to minimize settlement of encapsulated cells or particles.

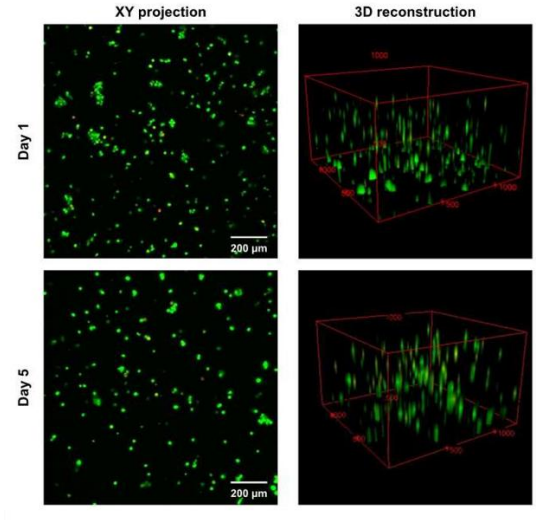


Figure 2.5. NIH/3T3 mouse fibroblasts encapsulated in the dual-crosslinked hydrogels showed good viability up to 5 days. 50 mM ferric EDTA and 8 mM DVS were used during cell encapsulation. Cells stained green are viable, and cells stained red are dead. Most encapsulated cells were viable on days 1 and 5. The XY projection images are maximum intensity projections from confocal scanned stacks and were created using ImageJ. The unit for the grid in the 3D reconstruction images is μm.

Thiol-ene addition through DVS has been widely used to make hydrogels for cell encapsulation. Both free DVS and DVS-functionalized crosslinkers have been shown to have good cytocompatibility.[48-49] Degradation products of DVS-crosslinked hydrogels also have been reported to elicit minimal immune responses.[50] Free ferric ions have been shown to be cytotoxic, but the toxicity can be mitigated by chelation.[51] Our results also suggested that cells can survive with high viability with concentrations as high as 8 mM DVS and 50 mM ferric EDTA used for crosslinking. Furthermore, because of the relatively large mesh size of the hydrogel (~30 nm) compared to the hydrodynamic radius of ferric EDTA (0.42 nm),[52] fast diffusion of ferric EDTA from the hydrogel is expected. The resulting fast decrease in the ferric EDTA concentration in the hydrogel is also expected to lower any cytotoxic effect from ferric EDTA. In fact, ferric EDTA has been widely used as a food additive for iron fortification.[53] Therefore, the release of ferric EDTA from our hydrogels should not negatively impact the potential for its use in *in vivo* applications.

2.5 Conclusions

We developed a one-pot thiol crosslinking based on thiol-ene addition and thiol oxidation. The thiol-ene addition between thiols on PEG-SH and DVS molecules resulted in thioether bonds that were not reducible by reducing agents. In addition, ferric EDTA was used to promote thiol oxidation that forms disulfide bonds. Unlike thioether bonds, disulfide bonds were reversible by reduction. By varying the amount of DVS used during crosslinking, the ratio between thioether and disulfide bonds in the hydrogel network was tuned without significantly changing the mechanical properties of the hydrogel. The change in the ratio between the two bonds, however, allowed us to create hydrogels with different degradation behavior under reducing conditions. At the same polymer concentration, hydrogel degradation was modulated from extremely low degradability to complete degradation by 10 μM GSH within 5 days by lowering the DVS concentration used for crosslinking. The release rate of encapsulated FITC-dextran was also tunable using the same approach. Finally, we showed that the crosslinking with DVS and ferric EDTA showed good cytocompatibility. Our results suggested that the combination of DVS and ferric EDTA provides a crosslinking method for thiol-based hydrogels with tunable degradability independent of mechanical properties. These gels provide a new platform for investigating

redox-responsive drug delivery and cell-material interactions in tissue engineering scaffolds with decoupled mechanical properties and degradation behaviors in response to reduction stimuli.

2.6 Acknowledgments

This work was supported by the American Heart Association (12SDG8980014) and the National Institutes of Health (NIDCR-R03-DE021755).

2.7 References

(1) Caló, E.; Khutoryanskiy, V. V. Biomedical applications of hydrogels: A review of patents and commercial products. *Eur. Polym. J.* **2015,** *65*, 252-267.

(2) Gibas, I.; Janik, H. Synthetic polymer hydrogels for biomedical applications. *Chem. Chem. Technol.* **2010,** *4* (4), 297-304.

(3) Gyles, D. A.; Castro, L. D.; Silva Jr, J. O. C.; Ribeiro-Costa, R. M. A review of the designs and prominent biomedical advances of natural and synthetic hydrogel formulations. *Eur. Polym. J.* **2017,** *88*, 373-392.

(4) Hoffman, A. S. Hydrogels for biomedical applications. *Adv. Drug Delivery Rev.* **2012,** *64*, 18-23.

(5) Mironi-Harpaz, I.; Wang, D. Y.; Venkatraman, S.; Seliktar, D. Photopolymerization of cell-encapsulating hydrogels: Crosslinking efficiency versus cytotoxicity. *Acta Biomater.* **2012,** *8* (5), 1838-1848.

(6) Wilems, T. S.; Lu, X.; Kurosu, Y. E.; Khan, Z.; Lim, H. J.; Smith Callahan, L. A. Effects of free radical initiators on polyethylene glycol dimethacrylate hydrogel properties and biocompatibility. *J. Biomed. Mater. Res., Part A* **2017,** *105* (11), 3059-3068.

(7) Williams, C. G.; Malik, A. N.; Kim, T. K.; Manson, P. N.; Elisseeff, J. H. Variable cytocompatibility of six cell lines with photoinitiators used for polymerizing hydrogels and cell encapsulation. *Biomaterials* **2005,** *26* (11), 1211-1218.

(8) Parhi, R. Cross-Linked Hydrogel for Pharmaceutical Applications: A Review. *Advanced pharmaceutical bulletin* **2017,** *7* (4), 515-530.

(9) Ullah, F.; Othman, M. B. H.; Javed, F.; Ahmad, Z.; Akil, H. M. Classification, processing and application of hydrogels: A review. *Mater. Sci. Eng., C* **2015,** *57*, 414-433.

(10) Bax, D. V.; Davidenko, N.; Gullberg, D.; Hamaia, S. W.; Farndale, R. W.; Best, S. M.; Cameron, R. E. Fundamental insight into the effect of carbodiimide crosslinking on cellular recognition of collagen-based scaffolds. *Acta Biomater.* **2017,** *49*, 218-234.

(11) Powell, H. M.; Boyce, S. T. EDC cross-linking improves skin substitute strength and stability. *Biomaterials* **2006,** *27* (34), 5821-5827.

(12) Koniev, O.; Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **2015,** *44* (15), 5495-5551.

(13) Jiang, Y.; Chen, J.; Deng, C.; Suuronen, E. J.; Zhong, Z. Click hydrogels, microgels and nanogels: emerging platforms for drug delivery and tissue engineering. *Biomaterials* **2014,** *35* (18), 4969-4985.

(14) Madl, C. M.; Heilshorn, S. C. Bioorthogonal Strategies for Engineering Extracellular Matrices. *Adv. Funct. Mater.* **2018,** *28* (11), 1706046.

(15) Patenaude, M.; Smeets, N. M. B.; Hoare, T. Designing Injectable, Covalently Cross-Linked Hydrogels for Biomedical Applications. *Macromol. Rapid Commun.* **2014,** *35* (6), 598-617.

(16) Goethals, F.; Frank, D.; Du Prez, F. Protected thiol strategies in macromolecular design. *Prog. Polym. Sci.* **2017,** *64*, 76-113.

(17) Marino, S. M.; Gladyshev, V. N. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J. Mol. Biol.* **2010,** *404* (5), 902-916.

(18) Nair, D. P.; Podgorski, M.; Chatani, S.; Gong, T.; Xi, W.; Fenoli, C. R.; Bowman, C. N. The thiol-Michael addition click reaction: a powerful and widely used tool in materials chemistry. *Chem. Mater.* **2013,** *26* (1), 724-744.

(19) Asai, D.; Xu, D.; Liu, W.; Quiroz, F. G.; Callahan, D. J.; Zalutsky, M. R.; Craig, S. L.; Chilkoti, A. Protein polymer hydrogels by in situ, rapid and reversible self-gelation. *Biomaterials* **2012,** *33* (21), 5451-5458.

(20) Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.* **2015,** *80*, 148-157.

(21) Kar, M.; Shih, Y.-R. V.; Velez, D. O.; Cabrales, P.; Varghese, S. Poly (ethylene glycol) hydrogels with cell cleavable groups for autonomous cell delivery. *Biomaterials* **2016,** *77*, 186-197.

(22) Yang, F.; Wang, J.; Cao, L.; Chen, R.; Tang, L.; Liu, C. Injectable and redox-responsive hydrogel with adaptive degradation rate for bone regeneration. *J. Mater. Chem. B* **2014,** *2* (3), 295-304.

(23) Yang, F.; Wang, J.; Hou, J.; Guo, H.; Liu, C. Bone regeneration using cell-mediated responsive degradable PEG-based scaffolds incorporating with rhBMP-2. *Biomaterials* **2013,** *34* (5), 1514-1528.

(24) Shi, K.; Wang, Y.-L.; Qu, Y.; Liao, J.-F.; Chu, B.-Y.; Zhang, H.-P.; Luo, F.; Qian, Z.-Y. Synthesis, characterization, and application of reversible PDLLA-PEG-PDLLA copolymer thermogels in vitro and in vivo. *Sci. Rep.* **2016,** *6*, 19077.

(25) Zustiak, S. P.; Leach, J. B. Hydrolytically Degradable Poly(Ethylene Glycol) Hydrogel Scaffolds with Tunable Degradation and Mechanical Properties. *Biomacromolecules* **2010,** *11* (5), 1348-1357.

(26) Choh, S.-Y.; Cross, D.; Wang, C. Facile synthesis and characterization of disulfide-cross-linked hyaluronic acid hydrogels for protein delivery and cell encapsulation. *Biomacromolecules* **2011,** *12* (4), 1126-1136.

(27) McCall, J. D.; Anseth, K. S. Thiol–ene photopolymerizations provide a facile method to encapsulate proteins and maintain their bioactivity. *Biomacromolecules* **2012,** *13* (8), 2410-2417.

(28) Hoyle, C. E.; Bowman, C. N. Thiol–ene click chemistry. *Angew. Chem., Int. Ed.* **2010,** *49* (9), 1540-1573.

(29) Darling, N. J.; Hung, Y.-S.; Sharma, S.; Segura, T. Controlling the kinetics of thiol-maleimide Michael-type addition gelation kinetics for the generation of homogenous poly (ethylene glycol) hydrogels. *Biomaterials* **2016,** *101*, 199-206.

(30) Jansen, L. E.; Negrón-Piñeiro, L. J.; Galarza, S.; Peyton, S. R. Control of thiol-maleimide reaction kinetics in PEG hydrogel networks. *Acta Biomater.* **2018,** *70*, 120-128.

(31) Jin, R.; Teixeira, L. M.; Krouwels, A.; Dijkstra, P. J.; Van Blitterswijk, C.; Karperien, M.; Feijen, J. Synthesis and characterization of hyaluronic acid–poly (ethylene glycol) hydrogels via Michael addition: An injectable biomaterial for cartilage repair. *Acta Biomater.* **2010,** *6* (6), 1968-1977.

(32) Lin, C.; Zhao, P.; Li, F.; Guo, F.; Li, Z.; Wen, X. Thermosensitive in situ-forming dextran–pluronic hydrogels through Michael addition. *Mater. Sci. Eng., C* **2010,** *30* (8), 1236-1244.

(33) Stewart, S.; Coulson, M.; Zhou, C.; Burke, N.; Stöver, H. Synthetic hydrogels formed by thiol–ene crosslinking of vinyl sulfone-functional poly (methyl vinyl ether-alt-maleic acid) with α, ω-dithio-polyethyleneglycol. *Soft Matter* **2018,** *14* (41), 8317-8324.

(34) Bartecki, A. Color of Transition Metal Compounds. *Rev. Inorg. Chem.* **1992,** *12* (1), 35-74.

(35) Solomon, E. I.; Gorelsky, S. I.; Dey, A. Metal–thiolate bonds in bioinorganic chemistry. *J. Comput. Chem.* **2006,** *27* (12), 1415-1428.

(36) Oyen, M. L. Mechanical characterisation of hydrogel materials. *Int. Mater. Rev.* **2014,** *59* (1), 44-59.

(37) Brown, T. E.; Carberry, B. J.; Worrell, B. T.; Dudaryeva, O. Y.; McBride, M. K.; Bowman, C. N.; Anseth, K. S. Photopolymerized dynamic hydrogels with tunable viscoelastic properties through thioester exchange. *Biomaterials* **2018**.

(38) Sawicki, L. A.; Kloxin, A. M. Design of thiol–ene photoclick hydrogels using facile techniques for cell culture applications. *Biomater. Sci.* **2014,** *2* (11), 1612-1626.

(39) Cosson, S.; Otte, E. A.; Hezaveh, H.; Cooper-White, J. J. Concise Review: Tailoring Bioengineered Scaffolds for Stem Cell Applications in Tissue Engineering and Regenerative Medicine. *Stem Cells Transl. Med.* **2015,** *4* (2), 156-164.

(40) Li, J.; Mooney, D. J. Designing hydrogels for controlled drug delivery. *Nat. Rev. Mater.* **2016,** *1*, 16071.

(41) Wu, G.; Fang, Y.-Z.; Yang, S.; Lupton, J. R.; Turner, N. D. Glutathione metabolism and its implications for health. *J. Nutr.* **2004,** *134* (3), 489-492.

(42) Estrela, J. M.; Ortega, A.; Obrador, E. Glutathione in Cancer Biology and Therapy. *Crit. Rev. Clin. Lab. Sci.* **2006,** *43* (2), 143-181.

(43) Pan, Y.-J.; Chen, Y.-Y.; Wang, D.-R.; Wei, C.; Guo, J.; Lu, D.-R.; Chu, C.-C.; Wang, C.-C. Redox/pH dual stimuli-responsive biodegradable nanohydrogels with varying responses to dithiothreitol and glutathione for controlled drug release. *Biomaterials* **2012,** *33* (27), 6570-6579.

(44) Canal, T.; Peppas, N. A. Correlation between mesh size and equilibrium degree of swelling of polymeric networks. *J. Biomed. Mater. Res.* **1989,** *23* (10), 1183-1193.

(45) Lee, S.; Tong, X.; Yang, F. The effects of varying poly (ethylene glycol) hydrogel crosslinking density and the crosslinking mechanism on protein accumulation in three-dimensional hydrogels. *Acta Biomater.* **2014,** *10* (10), 4167-4174.

(46) Jain, R. K. Transport of Macromolecules In Tumor Microcirculation. *Biotechnol. Prog.* **1985,** *1* (2), 81-94.

(47) Armstrong, J. K.; Wenby, R. B.; Meiselman, H. J.; Fisher, T. C. The Hydrodynamic Radii of Macromolecules and Their Effect on Red Blood Cell Aggregation. *Biophys. J.* **2004,** *87* (6), 4259-4270.

(48) Clevenger, T. N.; Hinman, C. R.; Rubin, R. K. A.; Smither, K.; Burke, D. J.; Hawker, C. J.; Messina, D.; Van Epps, D.; Clegg, D. O. Vitronectin-Based, Biomimetic Encapsulating Hydrogel Scaffolds Support Adipogenesis of Adipose Stem Cells. *Tissue Eng. Part A* **2016,** *22* (7-8), 597-609.

(49) Ma, Y.; Neubauer, M. P.; Thiele, J.; Fery, A.; Huck, W. T. S. Artificial microniches for probing mesenchymal stem cell fate in 3D. *Biomater. Sci.* **2014,** *2* (11), 1661-1671.

(50) Sabbieti, M. G.; Dubbini, A.; Laus, F.; Paggi, E.; Marchegiani, A.; Capitani, M.; Marchetti, L.; Dini, F.; Vermonden, T.; Di Martino, P.; Agas, D.; Censi, R. In vivo biocompatibility of p(HPMAm-lac)-PEG hydrogels hybridized with hyaluronan. *J. Tissue Eng. Regen. Med.* **2017,** *11* (11), 3056-3067.

(51) Singh, A. V.; Vyas, V.; Montani, E.; Cartelli, D.; Parazzoli, D.; Oldani, A.; Zeri, G.; Orioli, E.; Gemmati, D.; Zamboni, P. Investigation of in vitro cytotoxicity of the redox state of ionic iron in neuroblastoma cells. *J. Neurosci. Rural Pract.* **2012,** *3* (3), 301-310.

(52) Leaist, D. G.; Hao, L. Tracer diffusion of some metal ions and Metal–EDTA complexes in aqueous sodium chloride solutions. *J. Chem. Soc., Faraday Trans.* **1994,** *90* (1), 133-136.

(53) Blanco-Rojo, R.; Vaquero, M. P. Iron bioavailability from food fortification to precision nutrition. A review. *Innovative Food Sci. Emerging Technol.* **2018**.

3. PROTEIN ADHESIVES BASED ON ELASTIN-LIKE POLYPEPTIDES: A COMPARISON BETWEEN CATECHOL- AND THIOL-BASED ADHESION

This chapter consists of a manuscript by Lin C-Y and Liu JC, prepared for submission in 2019.

3.1 Abstract

Different chemistries have been utilized for protein-based materials to achieve adhesion in a humidified environment. L-3,4-dihydroxyphenylalanine (DOPA) found in marine mussel adhesive proteins has generated great interest because DOPA participates in multiple reaction mechanisms that confer the ability to adhere in wet conditions. However, because DOPA is a non-canonical amino acid, it requires additional steps to be incorporated into recombinant proteins. On the other hand, thiol groups can be easily incorporated into recombinant protein design via cysteine residues. Although thiol-based chemistry has been used in developing adhesive materials, its potential in achieving wet adhesion has not been fully explored. In this work, we took advantage of the sequence flexibility in elastin-like polypeptides (ELPs) and designed two ELP-based adhesives with highly similar sequences that are rich in DOPA or thiol residues. The sequence similarity between the two ELP adhesives allowed us to focus on the differences between DOPA- and thiol--based adhesion. Curing kinetics, adhesion strength in a wet setting, capability to recover from disturbance in the curing process, and cytocompatibility of the two adhesives were compared. Finally, a synergistic increase in the adhesion strength was achieved by utilizing both DOPA- and thiol-based chemistry simultaneously. Our results suggested that the performance of an adhesive material depends on the chemistry used, and our work demonstrated the potential of DOPA-thiol chemistry in protein-based adhesives for wet adhesion.

3.2 Introduction

More than 310 million surgeries are performed each year worldwide,[1] and more than half of these surgeries involves would closure using sutures or staples.[2] However, these

methods inevitably damage the surrounding healthy tissue and have a higher possibility of infection.[3] Surgical sealants and adhesives can be used together with sutures or staples to mitigate such risks.[4-5]

Two important criteria for surgical sealants and adhesives are to be biocompatible and to not cause additional complications.[6] Many of the current FDA-approved surgical sealants or adhesives, however, do not fully meet these criteria. For example, BioGlue® is an FDA-approved tissue adhesive using bovine albumin, which is the most abundant protein in mammalian plasma, and glutaraldehyde to achieve sealing. Glutaraldehyde reacts with amine groups on albumin or surrounding tissue, and the reaction is independent of the blood clotting mechanism in the body. However, glutaraldehyde can be cytotoxic and cause inflammation near the application site.[7-8] The use of bovine albumin can also trigger immune responses.[7-8] Products that circumvent the use of glutaraldehyde have been developed. One example is ProGel™, which replaces glutaraldehyde with poly(ethylene glycol) (PEG) functionalized with N-hydroxylsuccinimide (NHS)esters that react with amine groups. However, PEG is hydrophilic and can swell up to four times by volume.[9] Such swelling can cause pressure to the surrounding tissue. Tisseel and Crosseal™ use fibrinogen and thrombin to trigger the blood clotting mechanism to achieve hemostasis and sealing action. FloSeal and TachoSil® augment fibrin by adding gelatin or collagen, respectively, to increase the sealing efficiency. Although the probability is low, fibrin-based sealants do have a risk of transmitting bloodborne pathogens.[10]

Other than the blood clotting mechanism, nature has more to offer in creating protein-based adhesion. *Drosophila nasuta* uses highly glycosylated proteins to secure puparia onto surfaces.[11] The red-legged salamander *Plethodon shermani* utilizes a sticky protein solution that hardens within seconds as a way to immobilize predators.[12] Upon provocation, the Australian frog *Notaden bennetti* secretes a viscous protein solution as a defense mechanism.[13] A velvet worm, *Euperipatoides rowelli*, captures its prey with a similar protein solution.[14] Proteomic analysis reveals that the amino acid compositions of the last two sticky protein solutions are similar in having high proline/hydroxyproline and charged residues.[15] Protein segments low in structural complexity are hypothesized to form tandem repeats via supramolecular assembly and thus result in the adhesion of these protein solutions. Some plants also use proteins to stick to and climb onto surfaces. English ivy

(*Hedera helix*) produces fluid that is rich in spherical hydroxyproline-rich glycoproteins.[16] In the presence of calcium ions, these proteins can form strong electrostatic interactions with pectic polysaccharides found in the extracellular matrix of English ivy. The spherical structure of the proteins also increase the wetting ability of the fluid and allow stronger adhesion to the surface.

Among all the adhesive materials created by organisms in nature, mussel adhesive proteins have perhaps drawn the most significant interest. Mussels produce proteins that are rich in charged residues (e.g., lysine) and a non-canonical amino acid 3,4-dihydroxyphenylalanine (DOPA) which confers strong adhesion even in a wet environment or underwater.[17] Such adhesion is achieved through a combination of mechanisms, including oxidation of DOPA to quinone, hydrogen bonding, covalent bonding, and metal chelation (e.g., iron).[18] Mussels also secrete proteins that have a relatively high thiol content.[19] In one of these thiol-rich proteins, the majority of the thiols are in the form of disulfide bonds.[20] These disulfide bonds are proposed to provide the structural integrity of the mussel adhesive protein complex. Meanwhile, thiols on another thiol-rich mussel protein are found to be presented mostly in their reduced form.[21] These free thiols can participate in the redox reactions of DOPA and form additional covalent bonds with DOPA.[21-22]

Efforts have been made in developing recombinant protein-based adhesives utilizing the DOPA chemistry found in mussel adhesive proteins.[23-25] However, because DOPA is a non-canonical amino acid, the incorporation of DOPA into recombinant protein production requires post-expression modification to convert tyrosine to DOPA. Direct incorporation of DOPA during recombinant mussel adhesive protein expression in *Escherichia coli* has been reported with good incorporation rate but low yield.[26] Recombinant proteins based on different amino acid sequences that feature DOPA to confer adhesion have also been developed. For example, Brennan *et al.* designed a mussel-inspired adhesive elastin-like polypeptide (ELP) with good adhesion strength in both dry and humid conditions.[27] Tyrosine was incorporated into the ELP design to introduce DOPA by tyrosinase conversion. Burke et al. constructed a silk-based adhesive by chemically modifying serine to DOPA.[25] The DOPA-modified silk showed higher adhesion strength than the non-modified silk and remained cytocompatible to human mesenchymal stem cells.

While DOPA is the key component in many mussel-inspired adhesives, the thiol aspect of the mussel adhesive proteins has not been systematically studied. In this work, we created mussel-inspired ELP adhesives based on DOPA or thiol chemistry. ELPs have garnered great interest as biomaterials because of their elastic mechanical properties and good cytocompatibility.[28-29] ELPs are composed of a repeating pentapeptide sequence of Val-Pro-Gly-Xaa-Gly, where the identity of the guest residue Xaa can be any amino acid except for proline.[30] By inserting tyrosine or cysteine in the guest residue positions, we created a DOPA-containing ELP after tyrosinase conversion of the tyrosine or a thiol-containing ELP. We compared the adhesion strengths and the curing kinetics of these two ELP adhesives in the presence or absence of iron ions. We found that the thiol-based ELP adhesives showed faster curing and higher adhesion strength compared to the DOPA-based ELP adhesives but were less tolerant to disturbances during early curing. In addition, a higher adhesion strength was achieved when the two ELP adhesives were mixed at a 1:1 DOPA:thiol molar ratio. Finally, we assessed the cytocompatibility of our ELP adhesives. We found that more cells attached initially to the thiol-based ELP than DOPA-based ELP; however, cells seeded on both ELP adhesives had high viability after 5 days of culture. Overall, the difference in the performance of these DOPA- and thiol--based adhesives provides insights into the choice of the chemistry used when designing protein-based adhesives.

3.3 Experimental Methods

3.3.1 Materials

All materials and reagents were purchased from Sigma-Aldrich (St Louis, MO) and AMRESCO, Inc. (Radnor, PA) unless otherwise specified. Four-arm polyethylene glycol maleimide (4-arm PEG-Mal, MW 10000 g mol^{-1}) was purchased from JenKem Technology USA (Plano, TX). Milli-Q water (MQ) was used in all experiments except for protein expression and dialysis where reverse-osmosis purified water was used.

3.3.2 Protein Design and Cloning

The tyrosine-containing ELP, [YKV-72], was designed with a total of 72 pentapeptide repeats with a guest residue composition of tyrosine, lysine, and valine evenly distributed

(Supporting Information Figure 3.8). The cloning of [YKV-72] was performed with an encoding fragment from our previous work,[27, 31] recursive directional ligation,[32] and a cloning scheme previously developed by our lab.[33] For the cysteine-containing ELP, [CKVYKV-72], a DNA oligo encoding an ELP fragment of six pentapeptide repeats with a guest residue composition of Cys-Lys-Val-Tyr-Lys-Val was synthesized by Sigma-Aldrich. The cloning of [CKVYKV -72] was achieved by elongating the ELP fragment 12 times as described above.

3.3.3 Protein Expression and Purification

Proteins were expressed in the Rosetta2(DE3)pLysS *E. coli* strain (EMD Chemicals, Gibbstown, NJ) in 2xYT medium with 35 μg/mL chloramphenicol and either 50 μg/mL kanamycin (for [YKV-72]) or 200 μg/mL ampicillin (for [CKVYKV-72]). The expression was induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Corning Inc., Corning, NY) when the optical density at 600 nm was between 0.6 to 0.8. Three hours after induction, the cells were harvested by centrifugation at 5380*g* for 15 min at 4 °C. Cell pellets were resuspended in a denaturing buffer (8 M urea, 100 mM NaH_2PO_4, 100 mM Tris-Cl, pH 8.0) at 4 mL/g wet pellet; the resuspension was stored at -80 °C.

To lyse cells of [YKV-72], sonication (Misonix XL-2000, Qsonica, Newtown, CT) was performed at 25 watts with 15-sec on/off cycles on ice. After 20 cycles, the lysate was centrifuged at 10000*g* for 15 min at 4 °C to remove insoluble debris. The cleared lysate was added with 10 (w/v)% ammonium sulfate. The lysate was incubated on ice for 10 min before centrifugation at 10000*g* for 15 min at 4 °C. The supernatant was collected and incubated with another 10 (w/v)% ammonium sulfate before centrifugation. The centrifuged pellet was resuspended in water at 200 mg wet pellet/mL, and the resuspension was heated to 75 °C for 5 min with constant stirring, vortexed, and heated again as before. The heated resuspension was cooled on ice until ice cold and centrifuged at 10000*g* for 25 min at 4 °C. The supernatant was dialyzed against reverse osmosis water at 4 °C. The water was changed at least 3 h apart from the previous change until a 10^6 dilution was achieved. The dialyzed solution was centrifuged at 10000*g* for 15 min at 4 °C, and the supernatant was frozen in liquid nitrogen before lyophilization.

For [CKVYKV-72], cells were lysed and salted out as described above with the addition of 200 mM 2-mercaptoethanol (2-ME) throughout the process. The salted-out pellet was resuspended in ice-cold phosphate buffered saline (PBS) with 200 mM 2-ME at 50 mg wet pellet/mL and purified using an inverse transition cycling (ITC) method.[34] Briefly, the resuspension was centrifuged at 10000*g* for 15 min at 4 °C. The supernatant was treated with 2 M NaCl and incubated at 25 °C for 30 min before centrifugation at 10000*g* for 15 min at 28 °C. The pellet was then resuspended and centrifuged again in ice-cold PBS with 200 mM 2-ME but without NaCl. The supernatant was then collected and incubated with 50 mM dithiothreitol (DTT) for 10 min before dialysis.

Protein expression and purification were confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot using an anti-His tag antibody (GenScript, Piscataway, NJ). To quantify purity, gels stained with Coomassie Brilliant Blue R-250 were analyzed by densitometry using ImageJ software (National Institutes of Health, NIH, Bethesda, MD). The molecular weights of the proteins were confirmed using electrospray ionization mass spectrometry (ESI-MS) (Advion, Ithaca, NY).

3.3.4 Tyrosinase Modification

DOPA was introduced into [YKV-72] by converting tyrosine residues in the protein using tyrosinase as described previously.[27] Briefly, the protein was dissolved at 2 mg/mL in 0.1 M sodium acetate buffer with 100 mM ascorbic acid at pH 5.5. The solution was mixed with 150 U/mL of mushroom tyrosinase, and the mixture was incubated at 37 ° C for 2 h with shaking at 200 rpm. After incubation, 20 μL of 6 N HCl was added per mL of the mixture to quench the tyrosinase conversion. The solution was then dialyzed in 5% acetic acid and lyophilized. The conversion efficiency was assessed after lyophilization using a difference spectrophotometry method with free DOPA as standards.[27, 35]

3.3.5 Reduced-Thiol Quantification

The amount of reduced thiol content in [CKVYKV-72] was determined by 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) using L-cysteine as standards. Briefly, protein samples were prepared in 0.1 M phosphate buffer (pH 8.0) with 5 mM

ethylenediaminetetraacetic acid (EDTA) such that the total thiol concentration was 0.5 mM. The L-cysteine standard solutions were prepared from 0.015 to 1.5 mM. The DTNB solution was prepared in the same buffer at 0.08 mg mL^{-1}. The protein samples or the standards were mixed with the DTNB solution at a 1:10 ratio in a 96-well plate. The mixture was incubated at room temperature for 15 min and then measured for absorbance at 412 nm using a SpectraMax $M2^e$ plate reader (Molecular Devices, San Jose, CA).

3.3.6 Lap Shear Adhesion Test

Lap shear adhesion tests using porcine skin were set up following ASTM standard F2255-05.[36] Porcine skin sheets (Stellen Medical, St. Paul, MN) were cut into 1.5 by 1.5 mm square pieces and stored at -80 °C. To set up an adherend, porcine skin pieces were thawed in PBS at room temperature, blotted on a paper towel to remove excess PBS, and affixed to the aluminum substrates with cyanoacrylate glue (All Purpose Krazy Glue® Pen (Elmer's Products, High Point, NC) or Gorilla Super Glue® Brush & Nozzle (Gorilla Glue Company, Cincinnati, OH)). The affixed porcine skin pieces were then trimmed to the edge of the aluminum substrates. The porcine skin pieces were kept moist during the process.

[YKV-72] and [CKVYKV-72] stock solutions were prepared at the desired concentration in water, and the pH was adjusted to 7.4. Although the proteins were dissolved in water, the pH remained steady due to the intrinsic buffer capacity of the proteins. A stock solution of ferric EDTA (Fe-EDTA) was prepared by dissolving ferric sulfate in a 0.5 M EDTA solution. MQ and 10x PBS were then added to the solution to make a final concentration of 400 mM ferric ions, 440 mM EDTA, and 1x PBS at pH 7.4. To prepare the adhesive formulations, the protein stock solutions were first mixed at the desired ratio and then mixed with the stock Fe-EDTA solution. The final concentrations were 100 mg/mL protein and 100 mM iron ion. For the adhesive formulations without iron ions, the volume of the stock Fe-EDTA solution was replaced with 1x PBS. All solutions were kept on ice during the preparation.

Immediately after mixing, 10 μL of the adhesive mixture was applied to one porcine skin substrate, and the other porcine skin substrate was then overlaid with an overlap area of 1.44 cm^2. The samples were covered by a moistened paper towel and secured by a 50-

gram weight above the overlap area. The samples were incubated in a humidified chamber at 37 °C for the predetermined curing time. Before the lap shear adhesion test, the samples were removed from the humidified chamber and allowed to equilibrate for 10 min at room temperature while wrapped with moistened paper towels.

Lap shear adhesion tests were performed on a BOSE Electroforce 3200 series III (TA Instruments, New Castle, DE) by pulling the samples at a constant rate of 0.2 cm/min until failure. After the test, the actual overlap area was measured using a caliper, and this area was used to determine the adhesion strength. At least 5 samples were tested for each group. For each set of experiments, a group of [CKVYKV-72] cured for 24 h was always included as an internal control to account for batch-to-batch variation in porcine skin substrates.

3.3.7 Cytocompatibility Test

The NIH/3T3 fibroblasts were maintained in high-glucose Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 100 U/mL penicillin-streptomycin (Thermo-Fisher Scientific, Waltham, MA) and 10% fetal bovine serum (Lonza, Switzerland, catalog #: 14-501F) at 37 °C and 5% CO_2. Cells were subcultured at 60-80% confluency.

Hydrogels with the same compositions of the tested adhesive formulations were made in a silicone mold of 4.5-mm diameter and 200-μm thickness placed on an acid-washed 12-mm coverslip as the bottom of the mold. Hydrogels were allowed to polymerize for 6 h in a humidified chamber and were then extensively rinsed in PBS. Cells were then seeded on top of the gels at a density of 3000 cells/cm^2. Medium was changed every other day. On day 3 and day 5 after seeding, cell viability was assessed using a LIVE/DEAD™ Viability/Cytotoxicity Kit (Thermo-Fisher Scientific). After rinsing in PBS, cells were incubated with 1.5 μM ethidium homodimer-1 and 0.5 μM calcein-acetoxymethyl ester in PBS for 30 min at 37 °C. After incubation, cells were rinsed three times with PBS and imaged with a Nikon Ti-E C-1 Plus microscope (Nikon, Tokyo, Japan). The PBS used in these steps contained 0.01% $CaCl_2$ and $MgCl_2$ to enhance cell attachment.

3.3.8 Statistical Analysis

When more than two groups were compared, Levene's test was used to determine whether variances were equal, and two-way analysis of variance (ANOVA) was performed. If the interaction between the two tested factors was statistically insignificant, Tukey's Honestly Significant Difference (HSD) was used. If the interaction was significant, groups of each factor were analyzed separately by one-way ANOVA with Tukey's Honestly Significant Difference or Dunnett's test if they had equal variances. Games-Howell *post hoc* test was used if the groups had unequal variances. The normality of the ANOVA residuals was tested with the Kolmogorov-Smirnov test. If the residuals were not normal, the data were subjected to a Box-Cox transformation with an optimal λ, and the transformed data were analyzed again as described above. When only two groups were compared, an unpaired *t*-test was used instead. A p-value ≤ 0.05 was considered significant for all tests. The analyses were performed with Minitab 18 (State College, PA). Experimental groups that had high failure rates and yielded less than three valid data points were not included in the above statistical analyses.

3.4 Results and Discussion

3.4.1 Protein Design and Production

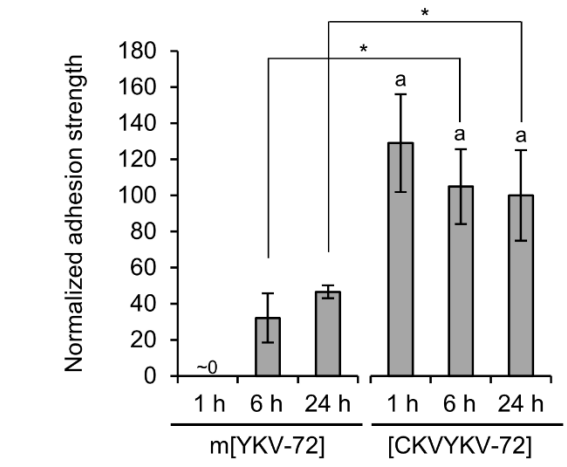
The goal of this study was to compare the performance of protein-based adhesives that use either a DOPA-based chemistry or a thiol-based chemistry. We utilized the sequence flexibility residing in the guest residue positions of ELP pentapeptide repeats to create two ELPs, [YKV-72] and [CKVYKV-72], with tyrosine or cysteine residues as the guest residues. The complete amino acid sequences of the proteins are shown in Supporting Information Figure 3.8. The expression of the proteins was confirmed by Western blot, and the purity of the proteins was assessed to be >95% by densitometry analysis of SDS-PAGE (Supporting Information Figure 3.9). The molecular weight of the proteins was confirmed by ESI-MS, as shown in Supporting Information Figure 3.10. Tyrosine residues in [YKV-72] were converted to DOPA using exogenous mushroom tyrosinase. Difference spectrophotometry of the converted [YKV-72], or m[YKV-72], showed a 67% conversion

efficiency, which was similar to the efficiency reported with a similar but shorter ELP construct.[27]

To retain thiol groups in [CKVYKV-72] in their reduced form, a reducing environment was maintained during the purification process by adding 2-ME and DTT. To avoid high temperatures, which could oxidize thiol groups, during the ITC purification process, 2 M NaCl was used to decrease the lower critical solution temperature of [CKVYKV-72] below room temperature. The purified [CKVYKV-72] had 82% of the thiol groups in the reduced form as determined by the DTNB assay. This percentage of reduced thiol groups was lower but comparable to other percentages achieved with similar thiol-containing ELPs using tris(2-carboxyethyl)phosphine (TCEP) as the reducing agent during purification.[37] However, it has been reported that the negatively-charged TCEP can bind to proteins with positively-charged residues and cannot be effectively removed.[38] To avoid interference from residual TCEP, non-charged reducing agents were used for the purification of [CKVYKV-72].

3.4.2 Lap Shear Adhesion Strength Depends on Time and Chemistry

We first tested the adhesion strength with varying curing times using 100 mg/mL of m[YKV-72] or [CKVYKV-72]. The adhesion strength of m[YKV-72] with a 1 h cure time could not be determined due to a 100% failure rate of the adhesive bond when the samples were transferred from the humidified chamber to the mechanical testing system. Statistically similar adhesion strengths were obtained with m[YKV-72] at curing times of 6 and 24 h (Figure 3.1). On the other hand, the adhesion strength from [CKVYKV-72] was less dependent on curing time. Statistically similar adhesion strengths were obtained from samples cured for only 1 h compared to those cured for 6 and 24 h. When comparing the adhesion strengths between the two proteins at the same curing time, [CKVYKV-72] showed statistically higher strengths than m[YKV-72] at 6 and 24 h. [CKVYKV-72] also seemed to have lower failure rates than m[YKV-72], especially at shorter curing times.



Formula	m[YKV-72]			[CKVYKV-72]		
Cure time (h)	1	6	24	1	6	24
Failure (%)	100	20	17	0	0	0
Replicates	6	5	6	5	5	5
Base (kPa)	5.1 ± 1.0			5.5 ± 1.4		

Figure 3.1. Curing kinetics depend on the adhesive chemistry used. The graph shows the adhesion strength of m[YKV-72] and [CKVYKV-72] cured for 1, 6, and 24 h. m[YKV-72] showed an increased strength with longer cure time. All samples of m[YKV-72] cured for 1 h failed during the test setup. [CKVYKV-72] showed similar adhesion strengths among all three cure times. Within each protein group, identical letters indicate statistically similar data as determined by Tukey's HSD test. * indicates a statistical difference ($p < 0.05$) between proteins at the same time point with an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests.

DOPA-based adhesion to tissue surfaces involves several mechanisms with different reaction kinetics.[39] Two main mechanisms are hydrogen bonding between DOPA and hydrogen bond donors in the system and covalent bond formation between dopaquinones (i.e., oxidized DOPA) or between dopaquinones and amine groups.[39] While adhesion from hydrogen bonding can occur immediately after applying the adhesives to the substrates, adhesion from covalent bond formation requires the oxidation of DOPA to take place first. In the case of thiol-based adhesion, the main mechanism is the oxidation of thiols to form disulfide bonds. Both oxidation reactions have similar redox potentials at pH 7.0 with -0.22 V for DOPA oxidation[40] and -0.24 V for cysteinyl thiol oxidation[41] and therefore should be thermodynamically favorable. However, the difference in the oxidation

kinetics between DOPA and cysteinyl thiol oxidations also plays a role in the curing process and can affect the adhesion strength formed with short curing times.

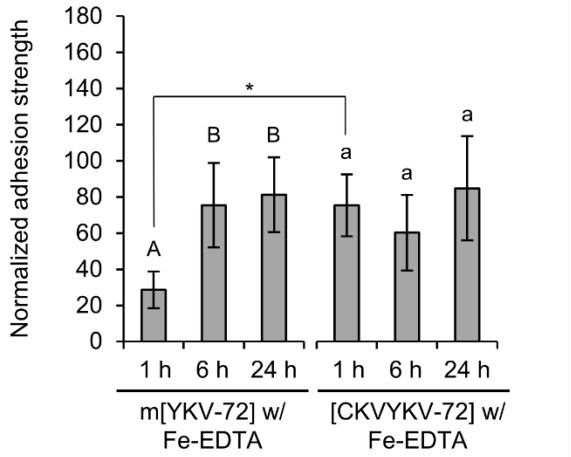
The adhesion strengths of [CKVYKV-72] were higher than those of m[YKV-72] at all curing times tested. Since the mass concentrations used were the same, the performance of [CKVYKV-72] is better than m[YKV-72] on a mass basis. The adhesion strength is expected to be proportional to the number of reactive DOPA or thiol groups present in the adhesives. With a 100 mg/mL mass concentration, the DOPA molar concentration in m[YKV-72] was ~45 mM, and the thiol molar concentration in [CKVYKV-72] was ~29 mM. Therefore, [CKVYKV-72] still outperformed m[YKV-72] when comparing on a molar basis of the reactive groups. Because DOPA can form not only covalent bonds but also hydrogen bonds, we hypothesize that some of the DOPA groups in m[YKV-72] formed hydrogen bonds to the porcine skin substrates. Since the bond energy of a hydrogen bond is generally lower than that of a covalent bond, less energy would be required to break the adhesion created from hydrogen bonds than that from covalent bonds.[42-43] As a result, lower adhesion strengths were obtained from m[YKV-72] even with higher molar concentrations of DOPA compared to the case of [CKVYKV-72].

To the best of our knowledge, our work is the first one to compare the adhesion strength of DOPA- and thiol-containing ELPs with high sequence similarities. Direct comparisons to other systems also utilizing DOPA or thiol chemistries are difficult due to the differences in protein sequences, material properties, formulations, and curing and testing setups. Ninan *et al.* reported that adhesive protein extracts from mussel feet yielded an adhesion strength of 0.93 MPa on porcine skin substrates when cured for 24 h in humid conditions.[44] The authors also reported that a significant increase in the adhesion strength from 12 h to 24 h curing time. Longer cure time (48 h) did not result in a significantly higher adhesion strength (0.95 MPa). The DOPA content in the extract was not specified, but 200 mg of the extract was applied on to an area of ~0.35 cm x1 cm. Wang and coworkers synthesized DOPA-containing polyaspartamides with different overall hydrophobicity.[45] They reported adhesion strengths ranging from 4 to 19 kPa on porcine skins following the trend of increasing hydrophobicity of the polyasparamides. The polyaspartamides had DOPA contents between 47-52 mol%. 50 mg of the polyaspartamides was applied to an area of 2.0 cm x1.5 cm. The curing conditions were

not specified. The adhesion strength from our m[YKV-72] was on the low end compared to the above to works. The DOPA content in m[YKV-72] was ~4 mol%, and 10 mg m[YKV-72] was applied on a 1.2 cm x 1.2 cm area. The lower value of adhesive/area could contribute to the lower adhesion strength observed. Granskog *et al.* developed a PEG-based adhesive utilizing thiol-ene click chemistry and reported adhesion strengths ranging from 3.7-21 kPa with porcine skins.[46] This adhesive was a two-component system with a alkene-functionalized dendritic PEG and a thiol-containing dendritic crosslinker. Another two-component adhesive based on thiol-ene reaction reported adhesion strength from ~16-48 kPa on rat skin substrates.[47] Light-activated curing was used in the above two studies on moistened skin substrates in dry conditions which was different from our curing of [CKVYKV-72] in humidified condition.

3.4.3 Iron Ions Enhanced the Performance of m[YKV-72]

Iron ions were utilized as mild oxidants to accelerate the curing process of both m[YKV-72] and [CKVYKV-72]. To maintain a minimum of 100% molar excess of the oxidant compared to the DOPA or thiol groups, 100 mM of iron ions were used in the adhesive formulations. At 1 h of curing time, m[YKV-72] with iron ions showed an adhesion strength of ~1.5 kPa (Figure 3.2). The adhesion strength increased at longer cure times and were statistically similar at cure times of 6 and 24 h. On the other hand, there were no statistical differences in obtained adhesion strengths for [CKVYKV-72] among the three curing times tested. This result was similar to the case when no iron ions were present (Figure 3.1). When the two proteins were compared at each curing time, m[YKV-72] with iron ions showed a statistically lower strength only at 1 h when compared to that of [CKVYKV-72]. When compared to groups of m[YKV-72] without iron ions with the same curing time, the adhesion strengths of m[YKV-72] cured for 6 and 24 h with iron ions were statistically higher (Supporting Information Figure 3.11A). Interestingly, [CKVYKV-72] with iron ions showed statistically lower adhesion strengths at 1 and 6 h curing time when compared to those without iron ions; however, statistically similar adhesion strengths were observed at 24 h (Supporting Information Figure 3.11B).



Formula	m[YKV-72] w/ Fe-EDTA			[CKVYKV-72] w/ Fe-EDTA		
Cure time (h)	1	6	24	1	6	24
Failure (%)	17	0	0	0	17	0
Replicates	6	6	6	6	5	5
Base (kPa)	5.1 ± 1.0			5.5 ± 1.4		

Figure 3.2. Addition of iron ions accelerates the curing kinetics of m[YKV-72]. The graph shows the adhesion strengths of m[YKV-72] and [CKVYKV-72] with iron ions cured for 1, 6, and 24 h. Lower failure rates and measurable adhesion strengths were obtained at 1 h cure time with m[YKV-72] with iron ions compared to that the results without Iron ions. [CKVYKV-72] showed similar adhesion strengths among three cure times. Within each protein group, identical letters indicate statistically similar data as determined by Tukey's HSD test. * indicates a statistical difference with $p < 0.05$ as assessed with an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests.

Iron ions have been shown to oxidize DOPA to dopaquinones[48] and to oxidize thiols to disulfides.[49] The addition of iron ions thus accelerated the curing process of m[YKV-72] by promoting the formation of dopaquinones and enabling the subsequent covalent bond formation between dopaquinones and other molecules. As a result, a non-zero adhesion strength was achieved with m[YKV-72] in the presence of iron ions. Since [CKVYKV-72] had already reached a stable adhesion strength at a 1h cure time in the absence of iron ions (Figure 3.1), iron ions did not appear to accelerate the curing process. However, iron ions could have an effect at curing times less than one hour. Iron ions also increased the adhesion strengths of m[YKV-72] at 6 and 24 h. These results suggest that due to more DOPA being oxidized into dopaquinone, more covalent bonds were formed and

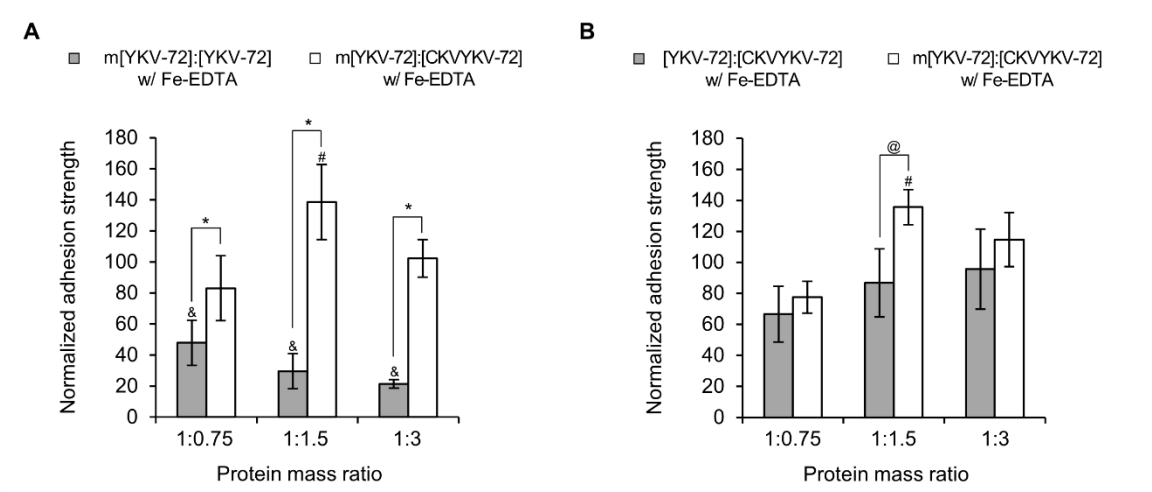
contributed to the increased adhesion strengths. Note that the adhesion strengths of m[YKV-72] at 6 and 24 h were comparable to those of [CKVYKV-72] at 6 and 24 h respectively in the presence of Iron ions even though more DOPA molecules were present in m[YKV-72] than that the number of thiols in [CKVYKV-27]. It is possible that some of the DOPA in the system still formed hydrogen bonds to the porcine skin substrate and did not participate in covalent bond formation. DOPA-iron ions complexes could also form and contribute to the adhesion. Since both hydrogen bonds and chelation bonds have lower bond energies than covalent bonds, the resulting adhesion strength would be lower than when only covalent bonds are formed.

Iron ions have been used to improve adhesion of adhesives based on mussel adhesive proteins extract or other DOPA-containing materials. Doraiswamy *et al*. extracted adhesive proteins from mussel feet and showed an increase in adhesion strength on porcine skin from ~0.25 to ~0.5 MPa when at a 1:1 iron ions:DOPA ratio.[50] The iron ions were from $FeCl_3$. Increasing the iron ions:DOPA ratio to 10:1 did not significantly change the resulted adhesion strength, but a further increase to 100:1 yielded a statistically similar strength to that with no iron ions. In our work, the ratio was about 2:1 iron ions:DOPA with m[YKV-72], which was within the optimal ratio suggested by Doraiswamy *et al*. Shi and coworkers developed a DOPA-containing polyester and showed an increase in the adhesion strength on porcine skin substrates from 6.2 to 13 kPa when iron ions were included in the formulation.[51] The iron ions source was ferric acetylacetonate, but the ratio between iron ions and DOPA was not specified. Meredith *et al*. compared the efficacy of different iron compounds on increasing the adhesion strength of DOPA-containing polystyrene.[52] Potassium ferrate showed an improved adhesion strength over that of the polymer alone, but ferric nitrate did not significantly increase the observed strength. The addition of ferric acetylacetonate resulted in a strength drastically lower than that of the polymer alone. Their results suggested that both the oxidation potential of the iron ions and the availability of the iron ions for DOPA-iron complexation can affect the efficacy of iron ions on increasing the adhesion strength. Based on our results with m[YKV-72], ferric EDTA seems to be a potential candidate for increasing DOPA-based adhesion. Thiols can be oxidized by strong oxidants such as hydrogen peroxide; however, the oxidation can go further beyond disulfide into sulfonic acid.[53] Transition metal ions, including iron ions, can oxidize thiols

to disulfides.[54] Asai and coworkers reported that potassium ferricyanide can crosslink thiol-containing ELPs into hydrogels within hours.[37] Our results with [CKVYKV-72] suggested that ferric EDTA can also be a potential compound for thiol oxidation.

3.4.4 Combination of Thiol and DOPA Chemistries Increased the Adhesion Strength.

Marine mussels utilize both DOPA and thiols to achieve their strong adhesion to rock surfaces.[55] Inspired by mussels, we combined our m[YKV-72] and [CKVYKV-72] proteins to create a protein-based adhesive utilizing both DOPA and thiol groups. Three mass ratios of m[YKV-72]:[CKVYKV-72] were tested: 1:0.75, 1:1.5, and 1:3 (corresponding to molar ratios of DOPA:thiol 1:0.5, 1:1, and 1:2, respectively). All groups were cured for 24 h with iron ions. Negative controls of each ratio were made with [YKV-72] replacing either m[YKV-72] or [CKVYKV-72]. The results are shown in Figure 3.3A (with negative controls of [YKV-72] substituting for [CKVYKV-72]) and Figure 3.3B (with negative controls of [YKV-72] replacing m[YKV-72]). The adhesion strength from a ratio 1:1.5 was statistically higher than both of its negative controls and was also statistically higher than [CKVYKV-72] cured under the same conditions with Iron ions. The other two ratios showed statistically higher adhesion strengths compared to their respective negative control groups consisting of m[YKV-72]:[YKV-72] but were statistically similar to the negative control groups containing [YKV-72]:[CKVYKV-72]. These results suggest that the thiol groups contributed more to the observed adhesion than the DOPA groups in the combination of m[YKV-72] and [CKVYKV-72].



Formula	m[YKV-72]:[YKV-72] w/ Fe-EDTA			m[YKV-72]:[CKVYKV-72] w/ Fe-EDTA		
Ratio	1:0.75	1:1.5	1:3	1:0.75	1:1.5	1:3
Failure (%)	17	0	0	0	17	0
Replicates	6	5	5	6	6	6
Base (kPa)	6.5 ± 1.2					

Formula	[YKV-72]:[YKVKC-72] w/ Fe-EDTA			m[YKV-72]:[CKVYKV-72] w/ Fe-EDTA		
Ratio	1:0.75	1:1.5	1:3	1:0.75	1:1.5	1:3
Failure (%)	0	0	0	0	0	0
Replicates	5	6	6	6	6	6
Base (kPa)	5.7 ± 1.1					

Figure 3.3. Mixing m[YKV-72] and [CKVYKV-72] at a mass ratio of 1-1.5 resulted in a significant increase in adhesion strength. m[YKV-72] and [CKVYKV-72] were mixed at three different mass ratios and were cured for 24 h in the presence of iron ions. The lap shear adhesion test results of each group were shown in (A) using m[YKV-72]-[YKV-72] as the negative control groups (DOPA only) and in (B) using [YKV-72]-[CKVYKV-72] as the negative control groups (thiol only). In both panels, the graphs show the normalized adhesion strengths, and the table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h) for each set of the tests. In panel (A), because two-way ANOVA indicated significant interaction between the two factors (formula and ratio), one-way ANOVA was performed on each formula separately with Dunnett's test. & indicates a significant difference to the control group of m[YKV-72], iron ions, 24 h from Figure. 2. with $p < 0.05$. # indicates a significant difference to the control group of [CKVYKV-72], iron ions, 24 h with $p < 0.05$. * indicates a statistical difference with $p < 0.05$ with an unpaired *t*-test. In panel (B), two-way ANOVA was performed using formula and ratio as the two factors with Tukey's HSD *post hoc* test. @ indicates the two group does not share the same Tukey group with a $p < 0.05$. # indicates a significant difference to the control group of [CKVYKV-72], iron ions, 24 h with $p < 0.05$ with Dunnett's test.

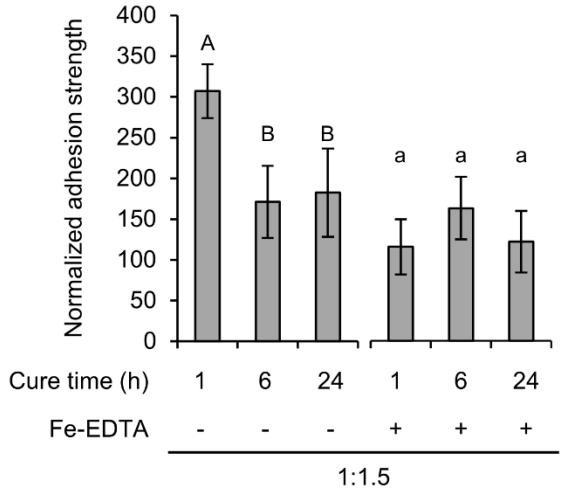
It has been shown that mussels produce adhesion proteins that are rich in free thiols,[19] and the free thiols have been shown to act as sacrificial reductants to reduce oxidized DOPA.[22] Furthermore, as free thiols being consumed by the reductions of oxidized DOPA, the remaining ones have been suggested to form thiol-DOPA adducts.

The formation of thiol-DOPA adducts have been explored to enhance the adhesion performance of DOPA-based adhesives,[56-59] and therefore we hypothesize that the formation of thiol-DOPA adducts is one of the contributor to the synergistic increase in the adhesion strength observed with m[YKV-72]:[CKVYKV-72] at 1:1.5.Ryu and coworkers reported that the storage moduli of their chitosan/pluronic hydrogels increased as the thiol-to-DOPA ratio increased.[56] On the other hand, Sparks et al. reported a peak Young's modulus with an intermediate DOPA content in their hydrogel system.[57] Because DOPA can participate in multiple bonding mechanisms including DOPA-Iron ions complexation, hydrogen bonding, and the redox reactions between thiols and dopaquinones, it is difficult to predict the optimal ratio between DOPA and thiols that reaches the optimal balance between all the bonding mechanisms and result in the optimal adhesion strength. Nonetheless, we speculate that, in our systems, at a 1:0.75 ratio of m[YKV-72]:[CKVYKV-72], there were more DOPA residues than thiol groups, and therefore, it was possible that not every DOPA could form a thiol-DOPA adduct. At the ratio of 1:3, there were more thiol groups than DOPA, and excessive thiols could drive the redox of DOPA to the reduced form. As a result, less dopaquinone was formed and fewer covalent bonds were formed, and the major contribution to the adhesion strength from DOPA was shifted to be from the hydrogen bonds and/or complexes with Iron ions. The 1:1 molar ratio between DOPA and thiols achieved at a protein mass ratio of 1:1.5 allowed the majority of DOPA to be in a thiol-DOPA adduct and hence resulted in the most significant increase in adhesion strength observed among the three ratios tested.

3.4.5 Similar Curing Process between m[YKV-72]-[CKVYKV-72] Mixture and [CKVYKV-72] Along

Based on the above results, we decided that m[YKV-72]-[CKVYKV-72] at 1-1.5 had the most significant synergistic effect from combining DOPA and thiol chemistries. and were interested in assessing the curing kinetics of this formula. As shown in Figure 3.4, in the absence of iron ions, a statistically higher adhesion strength was observed with a 1 h cure time, and then the adhesion strength stabilized at 6 and 24 h. A similar trend towards a higher adhesion strength at 1 h was also observed with [CKVYKV-72] (Figure 3.1), but the effect was not statistically significant. When the 1-1.5 ratio was cured with iron ions,

statistically similar adhesion strengths were observed among three curing times, and this result was again similar to the curing kinetics of [CKVYKV-72] with iron ions (Figure 3.2).



Formula	1:1.5			1:1.5 w/ Fe-EDTA		
Cure time (h)	1	6	24	1	6	24
Failure (%)	0	0	0	0	0	0
Replicates	6	6	6	6	6	6
Base (kPa)	3.5 ± 0.75					

Figure 3.4. m[YKV-72]-[CKVYKV-72] 1-1.5 showed fast curing kinetics that were similar to that of [CKVYKV-72]. The graph shows the adhesion strengths with or without iron ions at three curing times. Within each formula group, identical letters indicate statistically similar data as determined by Tukey's HSD test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h).

The decreasing adhesion strength of m[YKV-72]-[CKVYKV-72] from 1 h to 6 h was not expected from our results in Figure 3.1 and Figure 3.2. It has been suggested that a reversible intermediate forms early in the formation of thiol-DOPA adducts.[60] Such a reversible intermediate could explain the peak adhesion strength observed at a 1 h cure time with m[YKV-72]-[CKVYKV-72] 1-1.5. We hypothesize that, in the absence of iron ions, some of the thiol and DOPA groups initially participated in the formation of thiol-DOPA adduct intermediates could undergo the reverse reaction and participate in other reaction mechanisms contributing to the adhesion. Further studies on how the increased oxidation propensity induced by iron ions and DOPA-metal chelation alter the balance

among all possible reactions would benefit designing adhesives utilizing DOPA and thiol chemistries.

3.4.6 Adhesion Recovery after a Disturbance Depends on the Protein and the Curing Conditions.

The ability of our protein-based adhesives to self-heal were tested. Adhesives were applied to the substrates and allowed to cure for the designated time and then were separated with the mechanical tester. For the sample that failed during setting up for the mechanical tester, adherends were separated manually to ensure complete separation between two porcine skin pieces. The substrates were then carefully overlapped again and allowed to cure further until a total cure time of 24 h was achieved (

Figure 3.5A). The adhesion strengths after this second application and additional cure time were then measured on the mechanical tester as before.

In

Figure 3.5B, the adhesion strengths of m[YKV-72] from the first and second applications were compared. When no iron ions were used during the cure, m[YKV-72] showed a statistically similar adhesion strength after the second application at 6 h. Furthermore, both groups reapplied after 1 and 6 h showed statistically similar strengths to m[YKV-72] cured for 24 h without a disturbance (Figure 3.1 and

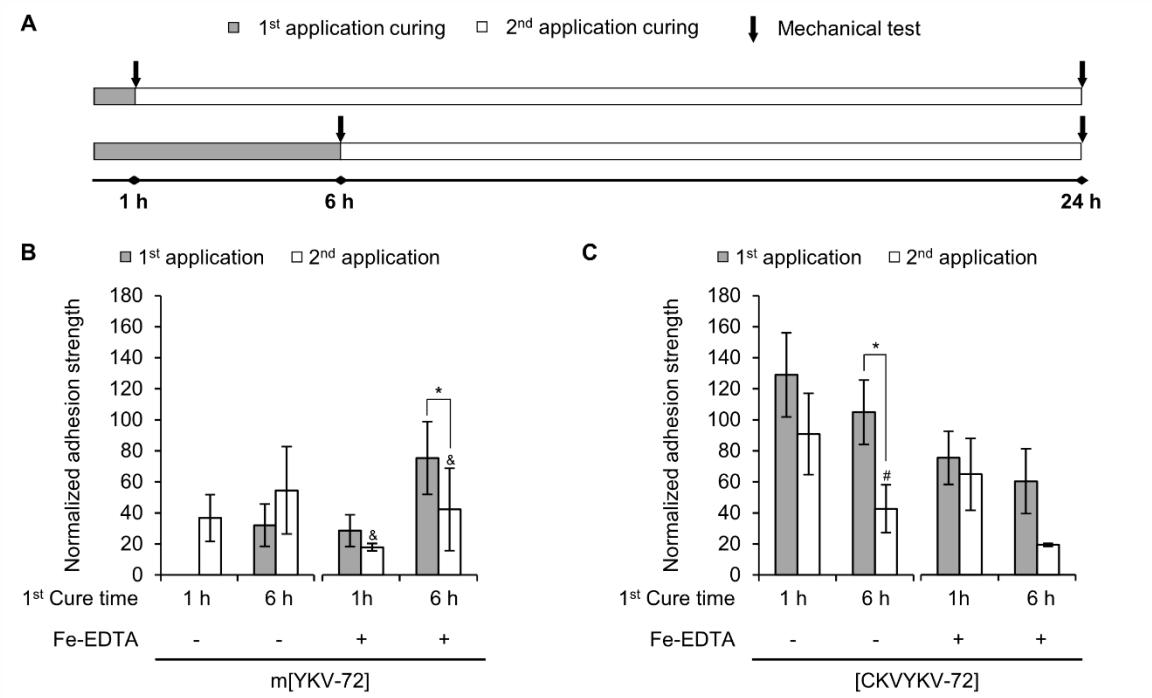
Figure 3.5B). On the other hand, m[YKV-72] with iron ions did not tolerate disturbances well during curing (

Figure 3.5B). Both of the second applications at 1 and 6 h showed statistically lower strengths compared to m[YKV-72] cured for 24 h with Fec ion (Figure 3.2 and

Figure 3.5B). The statistical difference in strengths between the first application and the second application at 6 h also suggests that the adhesion did not recover from the disturbance (

Figure 3.5B).

Figure 3.5. (A) An illustration on the curing and test process. (B) In the absence of iron ions, the adhesion strength of m[YKV-72] recovered from disturbances. The graph shows the adhesion strengths of m[YKV-72] from both the first and second applications. When no iron ions were used, the adhesion strengths after the second application were statistically similar to that of m[YKV-72] cured for 24 h undisturbed (Figure 3.1). With iron ions, the adhesion strengths were lower after the second applications compared to that of m[YKV-72] with iron ions cured for 24 h undisturbed (Figure 3.2). & indicates a significant difference to the control group of m[YKV-72], Iron ions, 24 h in Figure 3.2 with $p < 0.05$ in Dunnett's test. * indicates a statistical difference with $p < 0.05$ as determined by an unpaired *t*-test. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.1 and 2. (C) The adhesion strength of [CKVYKV-72] did not recover from disturbance after long curing times. The graph shows the adhesion strengths of [CKVYKV-72] after both the first and the second applications. Regardless of the presence of Iron ions, the adhesion strengths of the second applications at 1 h r were statistically similar to the strengths of the respective first applications and to their respective undisturbed groups that were cured for 24 h. The second application without Iron ions at 6 h showed statistically lower strength compared to the strengths after the first application and of the undisturbed 24-h group. The second application with Iron ions showed a markedly lower strength compared to the strength after the first application and had a high failure rate; however, statistical analysis was not performed due to insufficient sample size. # indicates a significant difference as assessed by Dunnett's test to the control group of [CKVYKV-72], 24 h in Figure 3.1 with $p < 0.05$. * indicates a statistical difference as determined by an unpaired *t*-test with $p < 0.05$. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.1 and Figure 3.2.



Formula	m[YKV-72]				m[YKV-72] w/ Fe-EDTA			
Group	1 h		6 h		1 h		6 h	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Failure (%)	100	17	20	0	17	33	0	0
Replicates	6	6	5	6	6	6	6	6
Base (kPa)	5.1 ± 1.0							

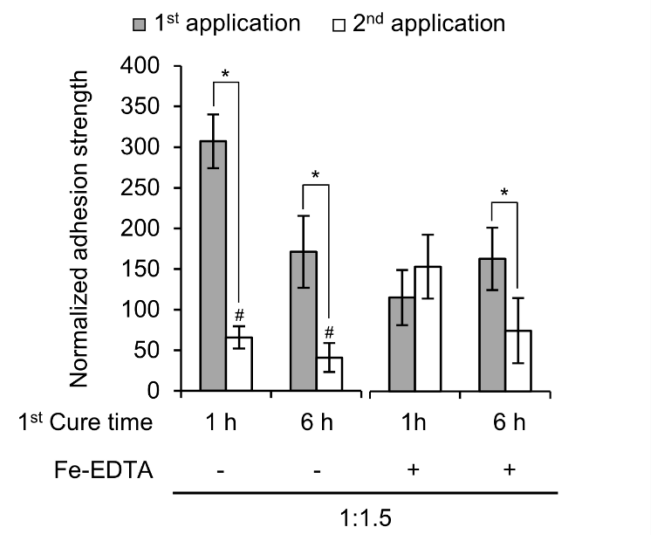
Formula	[CKVYKV-72]				[CKVYKV-72] w/ Fe-EDTA			
Group	1 h		6 h		1 h		6 h	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Failure (%)	0	0	0	17	0	0	17	67
Replicates	5	5	5	6	6	6	5	6
Base (kPa)	5.5 ± 1.4							

The adhesion strengths of [CKVYKV-72] after reapplication are shown in

Figure 3.5C. In the presence or absence of iron ions, the second applications of [CKVYKV-72] showed statistically similar adhesion strengths to their corresponding 24 h group when the adherends were separated 1 h after curing (Figure 3.1, 3.2, and 5C). The second applications at 1h also showed slightly lower yet comparable strength to the corresponding strength after the first application (

Figure 3.5). However, when the second application was performed 6 h after curing without iron ions, the adhesion strength was statistically lower than the strengths of both the 24 h group and the first application that cured for 6 h (Figure 3.1 and 5C). In the presence of iron ions, the adhesion strength of the second application at 6 h showed a higher failure rate than that of the corresponding first application or the 24 h group (Figure 3.2 and 5C). Statistical analysis could not be performed on this group because of the low sample size due to a high failure rate (67%) upon setting up the samples for the mechanical tester.

Figure 3.6 shows the comparison between the first and the second applications of m[YKV-72]-[CKVYKV-72] 1-1.5. In the absence of iron ions, the adhesion strengths from the second applications were statistically lower than the corresponding strengths after the first applications at 1 and 6 h. Both second applications were also statistically weaker than the group cured for 24 h without disturbance (Figure 3.4 and 3.6). These results suggest that the adhesion from m[YKV-72]-[CKVYKV-72] 1-1.5 did not recover well from separation during the curing. Interestingly, curing with iron ions seemed to improve such recovery for m[YKV-72]-[CKVYKV-72] 1-1.5. Even though there was a statistically significant difference between the strengths after the first and second application at 6 h, the adhesion strengths after the second applications at 1 and 6 h were statistically similar to the group cured undisturbed for 24 h with iron ions (Figure 3.4 and 3.6).



Formula	1:1.5				1:1.5, Fe-EDTA			
Group	1 h		6 h		1 h		6 h	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Failure (%)	0	17	0	17	0	0	0	0
Replicates	6	6	6	6	6	6	6	6
Base (kPa)	3.5 ± 0.75							

Figure 3.6. In the presence of iron ions, the adhesion strength of m[YKV-72]-[CKVYKV-72] 1-1.5 recovered from a disturbance. The graph shows the adhesion strengths of m[YKV-72]-[CKVYKV-72] 1:1.5 after both the first and second applications. Without iron ions, the adhesion strengths were statistically lower than the corresponding strengths after the first application and for the control group of m[YKV-72]-[CKVYKV-72] 1-1.5 cured for 24 h undisturbed. With iron ions, the adhesion strengths after the first and second applications at 1 h were similar. With iron ions, the second application at 6 h showed a statistically lower strength compared to that of the first application but was similar to m[YKV-72]-[CKVYKV-72] 1-1.5 cured for 24 h undisturbed (Figure 3.4). # indicates a significant difference as assessed by Dunnett's test to the control group of m[YKV-72]-[CKVYKV-72], 24 h with $p < 0.05$. * indicates a statistical difference as determined by an unpaired *t*-test with $p < 0.05$. The table shows the failure rate of each group and the absolute adhesion strength of the normalization base group ([CKVYKV-72] cured for 24 h). Data for the first applications are from Figure 3.4.

DOPA chemistry, hydrogen bonding and iron complexation in particular, has been explored for self-healing hydrogels.[61] Iron-crosslinked DOPA-based hydrogels seem to have a near-100% recovery in the mechanical properties. For example, Holten-Andersen and coworkers reported a self-healing PEG hydrogel system crosslinked by DOPA-iron

complexation.[62] The fractured hydrogel recovered within minutes to its original storage modulus. Krogsgaard *et al.* also reported a similar recovery in the storage modulus of their polyallylamine-based hydrogel with a longer self-healing time at ~45 min.[63] In our m[YKV-72] adhesive system, DOPA groups on the protein not only can interact with iron ions but also with other reactive groups on the porcine skin substrate. As the result, some of the DOPA groups can participate in reactions other than the DOPA-iron complexation that contributes to the self-healing behavior, which resulted in a lower recovery in adhesion strength with our m[YKV-72]. Thiol-crosslinked hydrogels have been shown to have self-healing capability as well through disulfide exchange.[64] Yu and coworker reported a self-healing thiol-crosslinked hydrogel using cyclic disulfides.[65] The tension in the cyclic disulfide promoted disulfide exchanges under mild condition. Full recovery in storage modulus was observed when the gel was disrupted under shear strain. Pepel *et al.* investigated the self-healing capability of different thiol-crosslinked polymers and proposed that the self-healing process depends on the pKa of the thiol groups and the pH.[66] They also reported that the rate of self-healing depended on the time between the material production and the material destruction. To the best of our knowledge, both DOPA or thiol self-healing capabilities have not been fully investigated for the applications of adhesives. Based on the above results, the adhesion of m[YKV-72] appears to have a higher tolerance to disturbances during the curing process, and this phenomenon may be because of the slower curing kinetics. [CKVYKV-72] adhesion, on the other hand, only recovered from disturbances early in the curing process, and this result may be because most of the reactive thiols are quickly consumed. The poor adhesion recovery of m[YKV-72]:[CKVYKV-72] 1-1.5 without iron ions can also be explained by its fast curing process. However, the good recovery of m[YKV-72]-[CKVYKV-72] 1-1.5 with iron ions was unexpected since the addition of iron ions appeared to accelerate the curing process. Because iron ions can participate in multiple reaction mechanisms with DOPA and thiol groups, it is possible that the capability of the adhesion to recover from disturbances is not a simple correlation to the curing kinetics. For example, the reaction extent on formation of thiol-DOPA adducts can be affected due to the complexation between DOPA and iron ions. Further experiments on elucidating the reaction mechanisms involving DOPA, thiols, and iron ions that occur

during the curing process would provide insights into devising adhesion with these molecules.

3.4.7 [CKVYKV-72] with Iron Ions is More Cytocompatible than m[YKV-72] with Iron Ions.

To investigate the cytocompatibility of our m[YKV-72] and [CKVYKV-72] adhesives, NIH/3T3 fibroblasts were seeded on top of hydrogels made with the same compositions as the adhesives tested in the lap shear adhesion tests. Three adhesive formulations were tested: [CKVYKV-72], m[YKV-72], and m[YKV-72]:[CKVYKV-72] 1:1.5 all with iron ions. The representative images of the live/dead assays of each group on days 3 and 5 are shown in Figure 3.7. Cells showed good attachment and proliferation with high viability on the [CKVYKV-72] gel. On the other hand, very few cells were observed on the m[YKV-72] gel surface on day 3. More cells were on the gel surface on day 5, but the overall density was still lower than that on the [CKVYKV-72] gel. Nonetheless, the cells on the m[YKV-72] gel were still viable. The m[YKV-72]:[CKVYKV-72] 1:1.5 hydrogel on day 3 showed cell counts and viability similar to those on the [CKVYKV-72] surface. However, the proliferation appeared to be slower than that on the [CKVYKV-72] surface as can be seen by the slightly lower cell density on the gel.

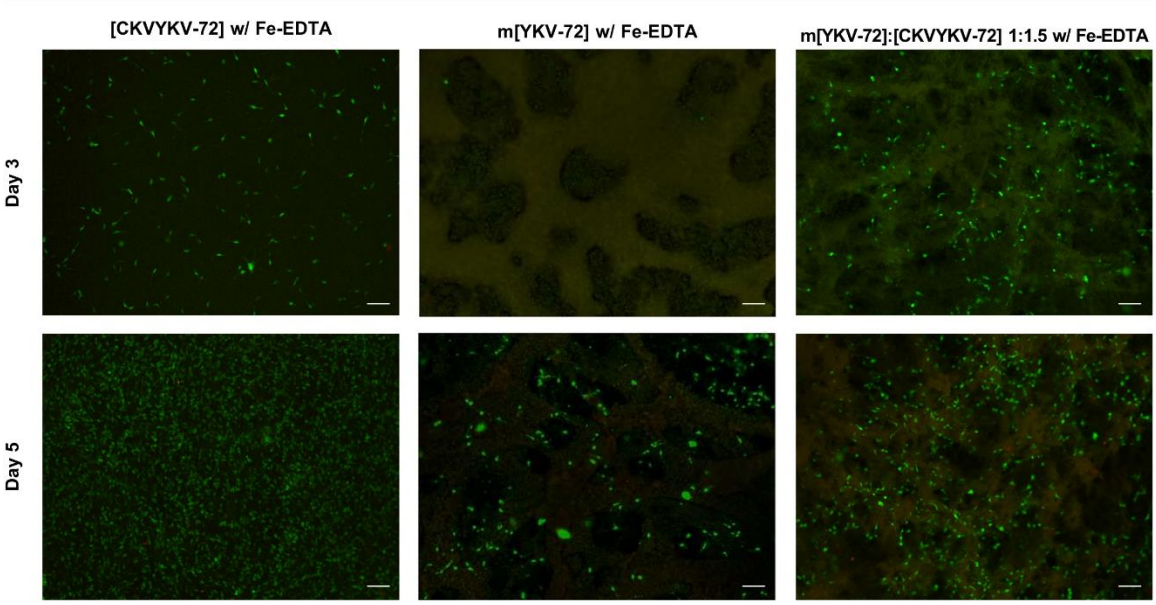


Figure 3.7. [CKVYKV-72] showed higher cytocompatibility than m[YKV-72]. Live/dead images of NIH/3T3 fibroblasts cultured on hydrogels of [CKVYKV-72], m[YKV-72], and m[YKV-72]:[CKVYKV-72] 1:1.5 all with iron ions show good viability on all three surfaces up to day 5 of culture. Cells stained green are viable, and cells stained red are dead. No cells were observed on m[YKV-72] at day 3, and these results suggest that the cell attachments were not strong enough to withstand the washing steps during the medium change and live/dead staining steps. Scale bar represents 100 μm.

The poor initial cell attachment on the m[YKV-72] gel at day 3 was consistent with previous results with a shorter version of m[YKV-72].[27] In the previous study, normal cell morphology with less spreading of fibroblasts was observed when the protein was coated on a cell culture surface without iron ions, but the cells did not attach to the surface firmly. Together with our live/dead result on m[YKV-72] gel on day 5 that cells were able to proliferate, m[YKV-72] with DOPA-iron ion complexes seems to be cytocompatible with suboptimal cell attachment property. The cytocompatibility of the complexes of iron ions with DOPA has not been fully investigated yet, but previous works suggested no negative effect on cell attachment and metabolism.[67-68] Cell attachment depends on both the chemical and the physical properties of the substrate surface.[69] New material designs such as the incorporation cell-binding domains or altering the guest residue composition in ELP designs[70] could also improve the cytocompatibility of these ELP-based adhesive materials. Future studies investigating the other mechanical properties (e.g., porosity, elasticity) of

[CKVYKV-72] and m[YKV-72] would be beneficial in understanding the interactions between these materials and the cells.

3.5 Conclusions

We designed two protein adhesives, [CKVYKV-72] and m[YKV-72], based on thiol or DOPA chemistries, respectively. Their performance results were compared based on curing kinetics, adhesion strengths, recovery after disturbance, and cytocompatibility. [CKVYKV-72] exhibited faster curing kinetics and higher adhesion strengths than m[YKV-72] when cured in a humidified environment. On the other hand, m[YKV-72] showed better recovery from disturbances during the curing process than [CKVYKV-72]. In the presence of iron ions, which acts as a mild oxidant to both DOPA and thiol groups and as a chelation center for only DOPA groups, the curing kinetics of m[YKV-72] were significantly accelerated, and the resulting adhesion strengths were comparable to those of [CKVYKV-72] with iron ions. However, the addition of iron ions decreased the ability of m[YKV-72] to recover from disturbances. By mixing m[YKV-72] and [CKVYKV-72] at a 1:1.5 protein mass ratio without Iron ions, the achieved adhesion strength with 1 h cure time increased by ~300% compared to [CKVYKV-72] cured for 24 h. The adhesion strength of this formula decreased with longer cure times but was still about 170% of the strength of [CKVYKV-72] cured under the same condition. The formula of m[YKV-72] and [CKVYKV-72] at a 1:1.5 protein mass ratio with Iron ions also showed an increased adhesion strength regardless of the curing time by on average ~130% compared to [CKVYKV-72] cured for 24 h under the same conditions. This formula also showed good recovery from disturbances and had a recovered adhesion strength statistically similar to that without any disturbance. Cell culture results showed that m[YKV-72] alone could have suboptimal cell attachment than [CKVYKV-72] or the combination of m[YKV-72] and [CKVYKV-72] in the presence of iron ions. All three formulations showed good cell viability on day 5. Overall, our results suggest that, in addition to the choice of the base material, the chemistry used to create adhesion should be considered when designing adhesive materials.

3.6 Acknowledgements

This work was supported by the American Heart Association (12SDG8980014) and the National Institutes of Health (NIDCR-R03-DE021755). We thank Dr. Sydney E. Hollingshead for help with developing the protocol for lap shear adhesion tests

3.7 Supporting Information

(A) [YKV-72]

M-MASMTGGQQMG-HHHHHHH-DDDDK-LDGTL-*(PGYGVPGKGVPGVGV)*$_{24}$-PVADRGMRLE

(B) [CKVYKV-72]

M-SKGPG-VDGTL-*(PGYGVPGKGVPGVGVPGCGVPGKGVPGVGV)*$_{12}$-PVADRGMRLE-HHHHHH

Figure 3.8. Amino acid sequences of [YKV-72] and [CKVYKV-72]. (A) Each repeat of (PGYGVPGKGVPGVGV) contains three ELP pentapeptide sequences with guest residues being Y, K, and V. There are 24 repeats, and thus there is a total of 72 pentapeptide sequences. (B) Each repeat of (PGYGVPGKGVPGVGVPGCGVPGKGVPGVGV)) contains six ELP pentapeptide sequences with guest residues being Y, K, V, C, K, and V. There are 12 repeats that result in a total of 72 pentapeptide sequences

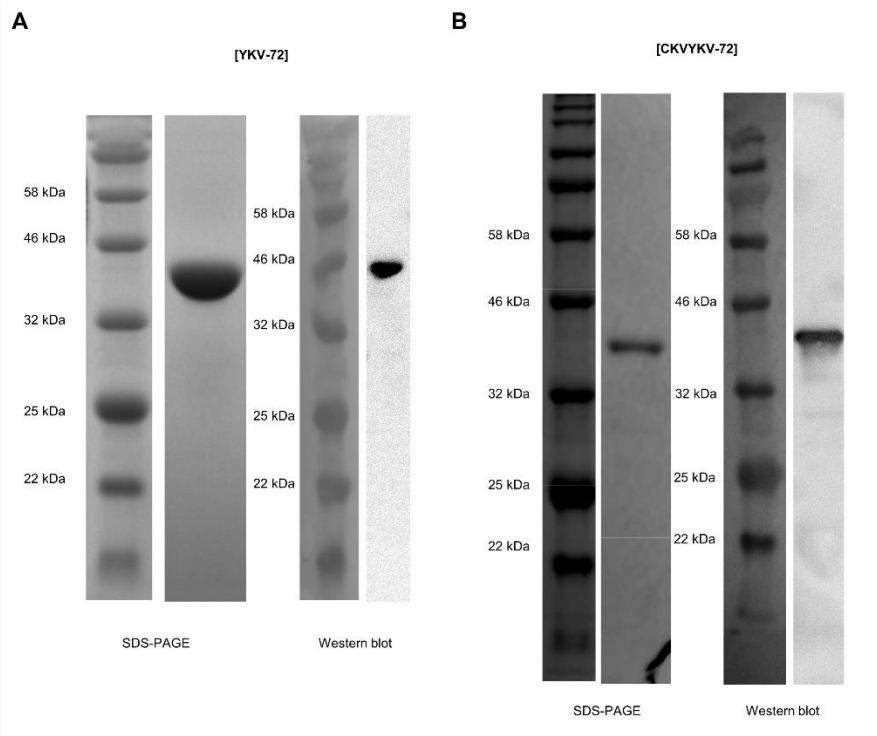


Figure 3.9. SDS-PAGE and Western blot results of (A) [YKV-72] and (B) [CKVYKV-72]. Densitometry of SDS-PAGE images showed >95% purity for both proteins. Western blot confirmed the presence of the T7-tag on purified [YKV-72] and the His-tag on purified [CKVYKV-72]. The expected molecular weights of the proteins are 32 kDa for [YKV-72] and 34 kDa for [CKVYKV-72].

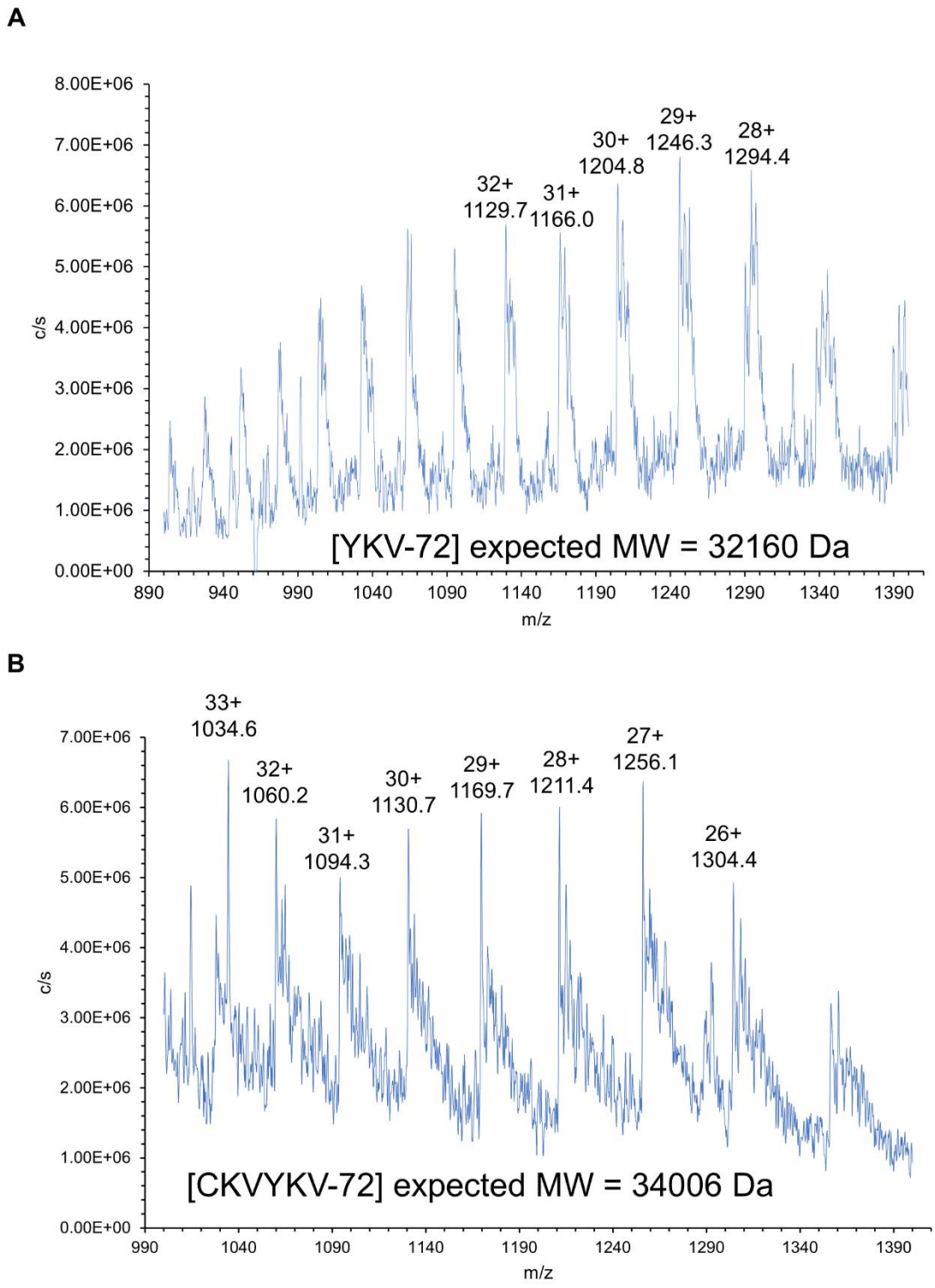


Figure 3.10. Electrospray ionization mass spectrometry of (A) [YKV-72] (expected MW: 32160 Da) and (B) [CKVYKV-72] (expected MW: 34006 Da).

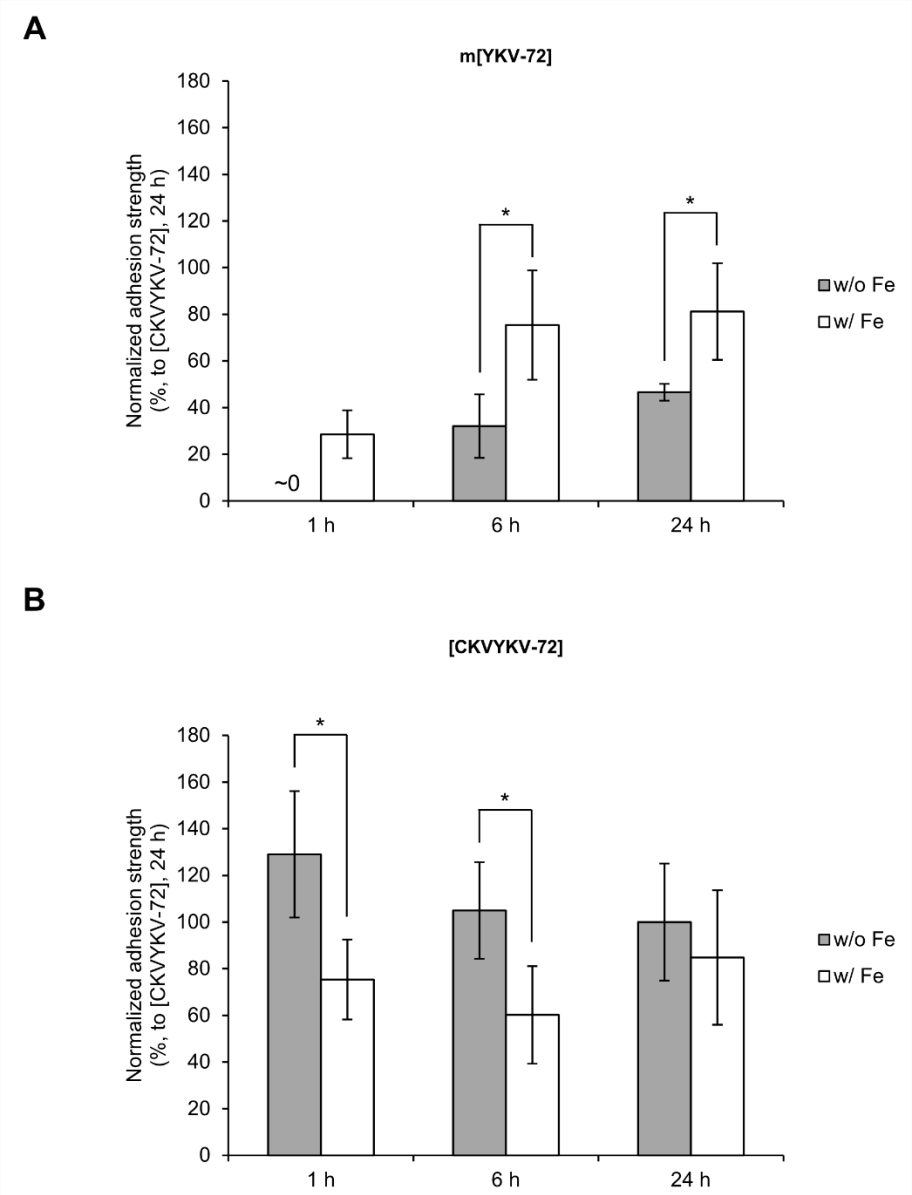


Figure 3.11. Adhesion strengths in the presence or absence of ferric ions for (A) m[YKV-72] and (B) [CKVYKV -72]. With m[YKV-72], there were statistical difference between groups with and without ferric ions at 6 and 24 h. Statistical analysis was not performed at 1 h due to the 100% failure rate of the group without ferric ions. In the case of [CKVYKV -72], adhesion strengths without ferric ions were statistically higher than those with ferric ions at 1 and 6 h. However, there was no statistical difference at 24 h. * indicates statistical difference as determined by a *t*-test with $p < 0.05$. Data are replotted from Figure 1 and 2 in the main text.

3.8 References

(1) Weiser, T. G.; Haynes, A. B.; Molina, G.; Lipsitz, S. R.; Esquivel, M. M.; Uribe-Leitz, T.; Fu, R.; Azad, T.; Chao, T. E.; Berry, W. R. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *The Lancet* **2015,** *385*, S11.

(2) Bré, L. P.; Zheng, Y.; Pêgo, A. P.; Wang, W. Taking tissue adhesives to the future: from traditional synthetic to new biomimetic approaches. *Biomater. Sci.* **2013,** *1* (3), 239-253.

(3) Bhagat, V.; Becker, M. L. Degradable Adhesives for Surgery and Tissue Engineering. *Biomacromolecules* **2017,** *18* (10), 3009-3039.

(4) Duarte, A.; Coelho, J.; Bordado, J.; Cidade, M.; Gil, M. Surgical adhesives: Systematic review of the main types and development forecast. *Prog. Polym. Sci.* **2012,** *37* (8), 1031-1050.

(5) Scognamiglio, F.; Travan, A.; Rustighi, I.; Tarchi, P.; Palmisano, S.; Marsich, E.; Borgogna, M.; Donati, I.; de Manzini, N.; Paoletti, S. Adhesive and sealant interfaces for general surgery applications. *J. Biomed. Mater. Res., Part B* **2016,** *104* (3), 626-639.

(6) Strausberg, R. L.; Link, R. P. Protein-based medical adhesives. *Trends Biotechnol.* **1990,** *8*, 53-57.

(7) Bhatia, S. K. Traumatic Injuries. In *Biomaterials for Clinical Applications*; Springer New York: New York, NY, 2010; Chapter 10, pp 213-258.

(8) Fürst, W.; Banerjee, A. Release of glutaraldehyde from an albumin-glutaraldehyde tissue adhesive causes significant in vitro and in vivo toxicity. *The Annals of thoracic surgery* **2005,** *79* (5), 1522-1528.

(9) Spotnitz, W. D.; Burks, S. Hemostats, sealants, and adhesives: components of the surgical toolbox. *Transfusion (Paris)* **2008,** *48* (7), 1502-1516.

(10) Horowitz, B.; Busch, M. Estimating the pathogen safety of manufactured human plasma products: application to fibrin sealants and to thrombin. *Transfusion (Paris)* **2008,** *48* (8), 1739-1753.

(11) Ramesh, S.; Kalisch, W.-E. Glue proteins inDrosophila nasuta. *Biochem. Genet.* **1988,** *26* (7-8), 527-541.

(12) von Byern, J.; Dicke, U.; Heiss, E.; Grunwald, I.; Gorb, S.; Staedler, Y.; Cyran, N. Morphological characterization of the glandular system in the salamander Plethodon shermani (Caudata, Plethodontidae). *Zoology* **2015,** *118* (5), 334-347.

(13) Graham, L. D.; Glattauer, V.; Huson, M. G.; Maxwell, J. M.; Knott, R. B.; White, J. W.; Vaughan, P. R.; Peng, Y.; Tyler, M. J.; Werkmeister, J. A. Characterization of a protein-based adhesive elastomer secreted by the Australian frog Notaden bennetti. *Biomacromolecules* **2005,** *6* (6), 3300-3312.

(14) Haritos, V. S.; Niranjane, A.; Weisman, S.; Trueman, H. E.; Sriskantha, A.; Sutherland, T. D. Harnessing disorder: onychophorans use highly unstructured proteins, not silks, for prey capture. *Proceedings of the Royal Society of London B: Biological Sciences* **2010,** *277* (1698), 3255-3263.

(15) Graham, L. D.; Glattauer, V.; Peng, Y. Y.; Vaughan, P. R.; Werkmeister, J. A.; Tyler, M. J.; Ramshaw, J. A. M. An Adhesive Secreted by Australian Frogs of the Genus Notaden. In *Biological Adhesives*; Smith, A. M., Ed.; Springer International Publishing: Cham, 2016; Chapter 11, pp 223-243.

(16) Huang, Y.; Wang, Y.; Tan, L.; Sun, L.; Petrosino, J.; Cui, M.-Z.; Hao, F.; Zhang, M. Nanospherical arabinogalactan proteins are a key component of the high-strength adhesive secreted by English ivy. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (23), E3193-E3202.

(17) Wilker, J. J. The iron- fortified adhesive system of marine mussels. *Angew. Chem., Int. Ed.* **2010,** *49* (44), 8076-8078.

(18) Wilker, J. J. Biomaterials: Redox and adhesion on the rocks. *Nat. Chem. Biol.* **2011,** *7* (9), 579.

(19) Kord Forooshani, P.; Lee, B. P. Recent approaches in designing bioadhesive materials inspired by mussel adhesive protein. *J. Polym. Sci., Part A: Polym. Chem.* **2017,** *55* (1), 9-33.

(20) Rzepecki, L. M.; Hansen, K. M.; Waite, J. H. Characterization of a cystine-rich polyphenolic protein family from the blue mussel Mytilus edulis L. *Biol. Bull.* **1992,** *183* (1), 123-137.

(21) Zhao, H.; Waite, J. H. Linking adhesive and structural proteins in the attachment plaque of Mytilus californianus. *J. Biol. Chem.* **2006,** *281* (36), 26150-26158.

(22) Yu, J.; Wei, W.; Danner, E.; Ashley, R. K.; Israelachvili, J. N.; Waite, J. H. Mussel protein adhesion depends on interprotein thiol-mediated redox modulation. *Nat. Chem. Biol.* **2011,** *7* (9), 588.

(23) Castillo, J. J.; Shanbhag, B. K.; He, L. Comparison of Natural Extraction and Recombinant Mussel Adhesive Proteins Approaches. In *Food Bioactives: Extraction and Biotechnology Applications*; Puri, M., Ed.; Springer International Publishing: Cham, 2017; pp 111-135.

(24) Zhong, C.; Gurry, T.; Cheng, A. A.; Downey, J.; Deng, Z.; Stultz, C. M.; Lu, T. K. Strong underwater adhesives made by self-assembling multi-protein nanofibres. *Nat. Nanotechnol.* **2014,** *9*, 858.

(25) Burke, K. A.; Roberts, D. C.; Kaplan, D. L. Silk Fibroin Aqueous-Based Adhesives Inspired by Mussel Adhesive Proteins. *Biomacromolecules* **2016,** *17* (1), 237-245.

(26) Yang, F.; Wang, J.; Cao, L.; Chen, R.; Tang, L.; Liu, C. Injectable and redox-responsive hydrogel with adaptive degradation rate for bone regeneration. *J. Mater. Chem. B* **2014,** *2* (3), 295-304.

(27) Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

(28) Rodríguez Cabello, J. C.; De Torre, I. G.; Cipriani, F.; Poocza, L. 12 - Elastin-like materials for tissue regeneration and repair. In *Peptides and Proteins as Biomaterials for Tissue Regeneration and Repair*; Barbosa, M. A.; Martins, M. C. L., Eds.; Woodhead Publishing: 2018; pp 309-327.

(29) Rodríguez-Cabello, J. C.; González de Torre, I.; Ibañez-Fonseca, A.; Alonso, M. Bioactive scaffolds based on elastin-like materials for wound healing. *Adv. Drug Delivery Rev.* **2018,** *129*, 118-133.

(30) Javier Arias, F.; Santos, M.; Fernández-Colino, A.; Pinedo, G.; Girotti, A. Recent contributions of elastin-like recombinamers to biomedicine and nanotechnology. *Curr. Top. Med. Chem.* **2014,** *14* (6), 819-836.

(31) Brennan, M. J.; Hollingshead, S. E.; Wilker, J. J.; Liu, J. C. Critical factors for the bulk adhesion of engineered elastomeric proteins. *Royal Society open science* **2018,** *5* (5), 171225.

(32) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **2010,** *11* (4), 944-952.

(33) Renner, J. N.; Kim, Y.; Cherry, K. M.; Liu, J. C. Modular cloning and protein expression of long, repetitive resilin-based proteins. *Protein Expr. Purif.* **2012,** *82* (1), 90-96.

(34) Meyer, D. E.; Chilkoti, A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115.

(35) Marumo, K.; Waite, J. Optimization of hydroxylation of tyrosine and tyrosine-containing peptides by mushroom tyrosinase. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1986,** *872* (1-2), 98-103.

(36) ASTM International, W. C., PA, Standard Test Method for Strength Properties of Tissue Adhesives in Lap-Shear by Tension Loading. In *ASTM F2255-05(2015)*, 2015.

(37) Asai, D.; Xu, D.; Liu, W.; Quiroz, F. G.; Callahan, D. J.; Zalutsky, M. R.; Craig, S. L.; Chilkoti, A. Protein polymer hydrogels by in situ, rapid and reversible self-gelation. *Biomaterials* **2012,** *33* (21), 5451-5458.

(38) McGann, C. L.; Akins, R. E.; Kiick, K. L. Resilin-PEG Hybrid Hydrogels Yield Degradable Elastomeric Scaffolds with Heterogeneous Microstructure. *Biomacromolecules* **2016,** *17* (1), 128-140.

(39) Mehdizadeh, M.; Yang, J. Design Strategies and Applications of Tissue Bioadhesives. *Macromol. Biosci.* **2013,** *13* (3), 271-288.

(40) Nicklisch, S. C.; Waite, J. H. Mini-review: The role of redox in Dopa-mediated marine adhesion. *Biofouling* **2012,** *28* (8), 865-877.

(41) Gilbert, H. F. Thiol/disulfide exchange and redox potentials of proteins. In *Bioelectrochemistry of biomacromolecules*; Lenaz, G.; Milazzo, G., Eds.; Birkhäuser: Switzerland, 2012; Chapter 5, pp 256-324.

(42) Lee, H.; Scherer, N. F.; Messersmith, P. B. Single-molecule mechanics of mussel adhesion. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (35), 12999-13003.

(43) Anderson, T. H.; Yu, J.; Estrada, A.; Hammer, M. U.; Waite, J. H.; Israelachvili, J. N. The contribution of DOPA to substrate–peptide adhesion and internal cohesion of mussel-inspired synthetic peptide films. *Adv. Funct. Mater.* **2010,** *20* (23), 4196-4205.

(44) Ninan, L.; Monahan, J.; Stroshine, R. L.; Wilker, J. J.; Shi, R. Adhesive strength of marine mussel extracts on porcine skin. *Biomaterials* **2003,** *24* (22), 4091-4099.

(45) Wang, B.; Jeon, Y.; Bhang, S.; Kim, J. Bioinspired dopamine-conjugated polyaspartamide as a novel and versatile adhesive material. *Express Polymer Letters* **2017,** *11* (8).

(46) Granskog, V.; Andrén, O. C. J.; Cai, Y.; González-Granillo, M.; Felländer-Tsai, L.; von Holst, H.; Haldosen, L.-A.; Malkoch, M. Linear Dendritic Block Copolymers as Promising Biomaterials for the Manufacturing of Soft Tissue Adhesive Patches Using Visible Light Initiated Thiol–Ene Coupling Chemistry. *Adv. Funct. Mater.* **2015,** *25* (42), 6596-6605.

(47) Zeng, Z.; Mo, X.-m.; He, C.; Morsi, Y.; El-Hamshary, H.; El-Newehy, M. An in situ forming tissue adhesive based on poly (ethylene glycol)-dimethacrylate and thiolated chitosan through the Michael reaction. *J. Mater. Chem. B* **2016,** *4* (33), 5585-5592.

(48) El-Ayaan, U.; Herlinger, E.; Jameson, R. F.; Linert, W. Anaerobic oxidation of dopamine by iron (III). *Journal of the Chemical Society, Dalton Transactions* **1997,** (16), 2813-2818.

(49) Taylor, J. E.; Yan, J. F.; Wang, J.-l. The iron (III)-catalyzed oxidation of cysteine by molecular oxygen in the aqueous phase. An example of a two-thirds-order reaction. *J. Am. Chem. Soc.* **1966,** *88* (8), 1663-1667.

(50) Doraiswamy, A.; Dunaway, T. M.; Wilker, J. J.; Narayan, R. J. Inkjet printing of bioadhesives. *J. Biomed. Mater. Res., Part B* **2009,** *89B* (1), 28-35.

(51) Shi, Y.; Zhou, P.; Jérôme, V. r.; Freitag, R.; Agarwal, S. Enzymatically degradable polyester-based adhesives. *ACS Biomater. Sci. Eng.* **2015,** *1* (10), 971-977.

(52) Meredith, H. J.; Jenkins, C. L.; Wilker, J. J. Enhancing the Adhesion of a Biomimetic Polymer Yields Performance Rivaling Commercial Glues. *Adv. Funct. Mater.* **2014,** *24* (21), 3259-3267.

(53) Chauvin, J.-P. R.; Pratt, D. A. On the Reactions of Thiols, Sulfenic Acids, and Sulfinic Acids with Hydrogen Peroxide. *Angew. Chem.* **2017,** *129* (22), 6351-6355.

(54) Witt, D. Recent Developments in Disulfide Bond Formation. *Synthesis* **2008,** *2008* (16), 2491-2509.

(55) Waite, J. H. Mussel adhesion–essential footwork. *J. Exp. Biol.* **2017,** *220* (4), 517-530.

(56) Ryu, J. H.; Lee, Y.; Kong, W. H.; Kim, T. G.; Park, T. G.; Lee, H. Catechol-functionalized chitosan/pluronic hydrogels for tissue adhesives and hemostatic materials. *Biomacromolecules* **2011,** *12* (7), 2653-2659.

(57) Sparks, B. J.; Hoff, E. F.; Hayes, L. P.; Patton, D. L. Mussel-inspired thiol–ene polymer networks: Influencing network properties and adhesion with catechol functionality. *Chem. Mater.* **2012,** *24* (18), 3633-3642.

(58) Sabbieti, M. G.; Dubbini, A.; Laus, F.; Paggi, E.; Marchegiani, A.; Capitani, M.; Marchetti, L.; Dini, F.; Vermonden, T.; Di Martino, P.; Agas, D.; Censi, R. In vivo biocompatibility of p(HPMAm-lac)-PEG hydrogels hybridized with hyaluronan. *J. Tissue Eng. Regen. Med.* **2017,** *11* (11), 3056-3067.

(59) Zeng, Z.; Mo, X. Rapid in situ cross-linking of hydrogel adhesives based on thiol-grafted bio-inspired catechol-conjugated chitosan. *J. Biomater. Appl.* **2017,** *32* (5), 612-621.

(60) Jameson, G. N.; Zhang, J.; Jameson, R. F.; Linert, W. Kinetic evidence that cysteine reacts with dopaminoquinone via reversible adduct formation to yield 5-cysteinyl-dopamine: an important precursor of neuromelanin. *Org. Biomol. Chem.* **2004,** *2* (5), 777-782.

(61) Krogsgaard, M.; Nue, V.; Birkedal, H. Mussel-Inspired Materials: Self-Healing through Coordination Chemistry. *Chemistry – A European Journal* **2016,** *22* (3), 844-857.

(62) Holten-Andersen, N.; Harrington, M. J.; Birkedal, H.; Lee, B. P.; Messersmith, P. B.; Lee, K. Y. C.; Waite, J. H. pH-induced metal-ligand cross-links inspired by mussel yield self-healing polymer networks with near-covalent elastic moduli. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (7), 2651.

(63) Krogsgaard, M.; Behrens, M. A.; Pedersen, J. S.; Birkedal, H. Self-Healing Mussel-Inspired Multi-pH-Responsive Hydrogels. *Biomacromolecules* **2013,** *14* (2), 297-301.

(64) Azcune, I.; Odriozola, I. Aromatic disulfide crosslinks in polymer systems: Self-healing, reprocessability, recyclability and more. *Eur. Polym. J.* **2016,** *84*, 147-160.

(65) Yu, H.; Wang, Y.; Yang, H.; Peng, K.; Zhang, X. Injectable self-healing hydrogels formed via thiol/disulfide exchange of thiol functionalized F127 and dithiolane modified PEG. *J. Mater. Chem. B* **2017,** *5* (22), 4121-4127.

(66) Pepels, M.; Filot, I.; Klumperman, B.; Goossens, H. Self-healing systems based on disulfide–thiol exchange reactions. *Polymer Chemistry* **2013,** *4* (18), 4955-4965.

(67) Choi, Y. C.; Choi, J. S.; Jung, Y. J.; Cho, Y. W. Human gelatin tissue-adhesive hydrogels prepared by enzyme-mediated biosynthesis of DOPA and Fe 3+ ion crosslinking. *J. Mater. Chem. B* **2014,** *2* (2), 201-209.

(68) Kim, B. J.; Kim, S.; Oh, D. X.; Masic, A.; Cha, H. J.; Hwang, D. S. Mussel-inspired adhesive protein-based electrospun nanofibers reinforced by Fe (III)–DOPA complexation. *J. Mater. Chem. B* **2015,** *3* (1), 112-118.

(69) Jafari, M.; Paknejad, Z.; Rad, M. R.; Motamedian, S. R.; Eghbal, M. J.; Nadjmi, N.; Khojasteh, A. Polymeric scaffolds in tissue engineering: a literature review. *J. Biomed. Mater. Res., Part B* **2017,** *105* (2), 431-459.

(70) Yuan, Y.; Koria, P. Proliferative activity of elastin-like-peptides depends on charge and phase transition. *J. Biomed. Mater. Res., Part A* **2016,** *104* (3), 697-706.

4. TAILORING THE TEMPERATURE RESPONSIVENESS OF POSITIVELY-CHARGED ELASTIN-LIKE POLYPEPTIDES WITH EXTERNAL SEQUENCES

This chapter consists of a manuscript by Lin C-Y and Liu JC, prepared for submission in 2019.

4.1 Abstract

Elastin-like polypeptides (ELPs) are recombinant protein domains exhibiting lower critical solution temperature (LCST) behavior. This LCST behavior not only is controlled by intrinsic factors including amino acid composition and polypeptide chain length but also is affected by external fusion domains. Here, we demonstrate that short external sequences can be harnessed as a new variable for designing the LCST behavior of ELP proteins. pH sensitivity near a neutral range is introduced to a positively-charged ELP fusion protein with small external sequences. The expanded pH sensitivity results from both electrostatic and hydrophobic interactions between the external non-ELP sequence and the positively-charged ELP domain. The hydrophobicity of the external non-ELP sequences also alters the ELP interaction with Hofmeister salts by changing the overall hydrophobicity of the fusion protein.

4.2 Introduction

Because of applications in various fields such as tissue engineering,[1] biosensors,[2] and drug delivery,[3] there is great interest in understanding how to modulate the responsiveness of temperature-sensitive polymers. Poly(*N*-isopropylacrylamide), or PNIPAM, is a representative temperature-responsive polymer with a lower critical solution temperature (LCST) of 32 °C.[4] The PNIPAM network is extended and hydrated below its LCST but collapses at temperatures above the LCST due to hydrophobic interactions. Based on this property, PNIPAM-coated surfaces have been used to harvest cultured cell sheets without disturbing the secreted extracellular matrix or the cells themselves.[5] Variants of synthetic

polymers based on PNIPAM have also been developed with different LCSTs to broaden the possible applications.[4]

Certain natural polymers also display temperature responsiveness. For example, elastin, which is found in mammalian extracellular matrix, shows a similar hydrophobic collapse above 37 °C.[6] Elastin-like polypeptides (ELPs) are recombinant protein domains derived from the elastin amino acid sequence. ELPs are composed of the pentapeptide repeats Val-Pro-Gly-Xaa-Gly where Xaa is referred to as a guest residue that can be any amino acid except proline.[7-8] ELPs also exhibit LCST behavior in solution similar to that of PNIPAM.[9-11] Below the LCST, the ELP chain is unstructured and hydrated. When the solution temperature is above the LCST, the ELP chain collapses and expels its waters of hydration. The collapsed ELP chains further aggregate due to hydrophobic interactions, and there is macroscopic phase separation into ELP-poor and ELP-rich phases. Upon cooling to the LCST, the solution goes back to a single phase, and the collapsed ELP chains become extended and hydrated. The LCST of an ELP has a strong dependence on the guest residue composition and the resulting hydrophobicity.[12] Hydrophobicity can be tuned by increasing the mole fraction of hydrophobic guest residues or by incorporating amino acids with higher hydrophobicity. Increasing the hydrophobicity lowers the LCST.

Other factors can influence ELP LCST behavior as well, including ELP chain length, protein concentration, salt identity, salt concentration, and pH of the solution.[13] For a fixed guest residue composition, both longer ELP chains and higher ELP concentrations result in lower LCSTs.[14] Hofmeister salts also affect the LCST behavior of ELPs, and both the type of salt and its concentration determine the effect on the LCST.[15-16] The LCST behavior of ELPs is sensitive to pH when ionizable amino acids are incorporated at guest residue positions. For example, MacKay and coworkers constructed two series of ELPs containing either Glu or His as guest residues and showed that the LCST behaviors were pH sensitive near their respective pKa values.[17] This pH sensitivity is a result of protonation or deprotonation of the ionizable amino acids. When the ionizable amino acids bear charges, the ELPs exhibit higher LCSTs due to a decrease in the overall hydrophobicity. At the pH where the ionizable residues are uncharged, the increased hydrophobicity leads to lower LCSTs.

ELP-based fusion proteins exhibit LCST behavior, which has been utilized for facile purification.[18] However, the LCST observed with the fusion protein can be different from the LCST of the ELP domain alone. The Chilkoti group compared the LCSTs of different ELP fusion proteins and found that the shift in LCST correlated with the surface hydrophobicity of the fusion partners.[19] In a follow-up study, the surface-exposed residues on the fusion partners were categorized into nonpolar, polar, and charged groups to compare their effects on the LCST shift.[20] The results showed that the charged amino acids have the most significant impact on the shift in the fusion protein LCST. In recent studies by Qin *et al.*, the relative position between the ELP domain and the fusion partner was found to have an effect on the LCST behavior and self-assembly of mCherry-ELP fusion proteins.[21-22]

Inspired by the shift in LCST behavior caused by non-ELP domains, we investigated the use of short non-ELP sequences as a new method to tailor the LCST behavior of ELP-based proteins. Previous studies clearly established that non-ELP domains were involved in the LCST behavior of ELP fusion proteins; however, the majority of the non-ELP domains studied were highly structured, and the effect on the LCST shift from non-ELP domains could be a result of both amino acid composition and domain structure. In this study, we extended the utility of using non-ELP sequences to alter LCST by choosing short sequences that are rich in charged amino acids. These external sequences have been shown to have no significant effect on the overall protein structure[23] or have been frequently incorporated for protein crystallization.[24-26] Therefore, the effect on the LCST shift from amino acid composition can be decoupled from structural changes. In addition, small non-ELP sequences can be more easily incorporated into modular recombinant protein designs and thus are a more favorable method for controlling the LCST behavior of ELP fusion proteins. In this paper, small non-ELP sequences introduced pH sensitivity into the LCST behavior of a positively charged ELP. We also report for the first time that external non-ELP sequences can alter the ELP interaction with Hofmeister salts. These results have important implications for rational design of pH- and temperature-responsive ELP-based proteins.

4.3 Experimental Methods

4.3.1 Materials

All materials and reagents were purchased from Sigma-Aldrich unless otherwise stated. Milli-Q water was used in all experiments except for protein expression and dialysis where reverse-osmosis purified water was used.

4.3.2 Protein Design and Cloning

S-tag[YKV-48] was previously described and referred to as ELY_{16} or ELP[$K_2Y_2V_2$-48].[27-28] S-tag[YKV-72] and S-tag[YKV-96] were cloned using the encoding fragment of the YKV domain, recursive directional ligation,[29] and a cloning scheme previously developed by our lab.[30] To construct I-tag[YKV-48], pET21b plasmid (a gift from Dr. Chongli Yuan, Purdue University) was first modified by removing the T7-tag and His-tag followed by inserting a DNA oligo encoding a leading sequence with the amino acid sequence Ser-Lys-Gly-Pro-Gly. The previously cloned YKV-encoding DNA fragment was then inserted into the modified pET21b after the leading sequence.

4.3.3 Protein Expression and Purification

S-tag[YKV] proteins and I-tag[YKV-48] were produced as described previously.[27] Briefly, both proteins were transformed into the Rosetta2(DE3)pLysS *E. coli* expression host (EMD Chemicals, Gibbstown, NJ). 2xYT medium was inoculated with appropriate antibiotics. For the S-tag[YKV] proteins, 50 μg/mL kanamycin and 35 μg/mL chloramphenicol were used, and, for the I-tag[YKV-48] protein, 200 μg/mL ampicillin and 35 μg/mL chloramphenicol were used. Cells were grown for 16-18 h at 275 rpm and 37 °C. The overnight culture was used to inoculate 0.5-1 L of 2xYT medium with appropriate antibiotics in a 4-L flask at 1:100. The culture was grown at 300 rpm and 37 °C and induced with 1.25 mM isopropyl β-$_D$-1-thiogalactopyranoside (IPTG, EMD Chemicals) when the optical density at 600 nm was between 0.6 to 0.8. The cells were harvested 5 h after induction by centrifugation at 5380*g* for 15 min at 4 °C. Cell pellets were resuspended in Buffer B (8 M urea, 100 mM NaH_2PO_4, 100 mM Tris-Cl, pH 8.0) at 2 mL/g wet pellet; the resuspension was stored at -80 °C.

Cells were lysed by 1-2 freeze-thaw cycles and sonication (Misonix XL-2000, Qsonica, Newtown, CT) at 15 watts with 30 sec on/off cycles on ice. After 30 cycles, the lysate was centrifuged at 10000*g* for 25 min at 4 °C to remove insoluble debris. The cleared lysate was collected, 10 (w/v)% ammonium sulfate was added, and the solution was incubated on ice for 10 min before centrifugation at 10000*g* for 25 min at 4 °C. Another 10 (w/v)% ammonium sulfate was added to the cleared supernatant before a second centrifugation. The pellet was collected and resuspended in water at 200 mg wet pellet/mL. The resuspension was heated to 75 °C for 5 min with constant stirring, vortexed, and heated again as before. The heated resuspension was cooled on ice until ice cold and centrifuged at 10000*g* for 25 min at 4 °C. The supernatant was diluted by half with ice-cold water before being dialyzed against reverse osmosis water at 4 °C. The water was changed at least 3 h apart from the previous change until the total reverse osmosis water volume exceeded 10^6 of the supernatant volume. The dialyzed solution was centrifuged at 10000*g* for 25 min at 4 °C, and the supernatant was frozen at -80 °C for overnight before lyophilization.

Protein expression and purification was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot. SDS-PAGE gels were stained with Coomassie Brilliant Blue R-250, and purity was quantified by densitometry using ImageJ software (National Institutes of Health, NIH, Bethesda, MD). The molecular weights of the proteins were confirmed using electrospray ionization mass spectrometry (Research Instrumentation Center, Purdue Univserity) (Supporting Information Figure 4.9). The amino acid compositions of I-tag[YKV-48] was confirmed by amino acid analysis (Supporting Information Figure 4.10) by the Molecular Structure Facility at the University of California, Davis.

4.3.4 LCST Measurement

A tri-buffer system (APT buffer) was prepared by combining 10 mM acetate buffer, 10 mM phosphate buffer (at pH 7), and 10 mM Tris buffer. Stock APT buffers with different salts were prepared similarly with the addition of 1 M NaCl, 2 M NaI, or 2 M Na_2SO_4. APT buffers with different salt concentrations were prepared by mixing the stock solutions.

Proteins were dissolved in ice-cold APT buffer, and the solutions were incubated on ice for at least 10 min to ensure complete solubilization. To prepare solutions for LCST measurements at different protein concentrations, a 400 μM stock solution of the protein was prepared first. Solutions with lower protein concentrations were then prepared by sequential dilution. The solutions were warmed to room temperature before adjusting to the desired pH with concentrated acids and bases. Acids and bases were chosen based on the identity of the salt in the buffer and were HCl and NaOH for solutions with NaCl, acetic acid and NaOH for solutions with NaI, and sulfuric acid and NaOH for solutions with Na_2SO_4.

The LCSTs of the solutions were determined using a Crystal16 (Technobis Group, Alkmaar, the Netherlands) by measuring transmission over a temperature range. Solutions were held at least 15 °C lower than estimated LCSTs for 10 min, and the temperature was increased at 1 °C/min until temperatures at least 20 °C higher than the estimated LCSTs were reached. For solutions with LCSTs higher than 75 °C, the temperature ramp was terminated at 95 °C to avoid the boiling point of water. After the temperature ramp, solutions were cooled back to 4 °C to check the reversibility of the LCST behavior. For solutions that had LCSTs lower than 15 °C, the solutions were held at 4 °C for 10 min to avoid the freezing point of water. The LCST was determined as the temperature at which the minimum of the first derivative of the transmission curve Δ(transmission)/ΔT occurred. Three samples were prepared independently and measured separately to report an average LCST for each condition. All numerical calculation and regression were performed using Matlab (MathWorks, Natick, MA).

4.4 Results and Discussion

4.4.1 Design and Production of the S-tag and I-tag[YKV] Proteins

We designed a set of recombinant proteins based on the same ELP domain with a guest residue composition of Tyr, Lys, and Val (YKV domain). A short sequence comprised of Ser-Lys-Gly-Pro-Gly (SKGPG) was inserted to the N-terminal of a YKV domain with 48 pentapeptide repeats (Figure 4.1A). The charge state of this short sequence was expected to be insensitive to changes in pH near neutral and was named I-tag. The protein was

designated as I-tag[YKV-48]. The SKGPG sequence has been reported to improve the expression yield of other ELP-based recombinant proteins[31-32] and has been incorporated in many ELP-based recombinant proteins used to investigate LCST behavior.[17, 33]

Another recombinant protein, S-tag[YKV-48], was created, and its amino acid sequence differed from that of I-tag[YKV-48] at the N-terminus. Specifically, the non-ELP leading sequence of S-tag[YKV-48] was composed of a T7 tag (Met-Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly), a His tag (7xHis), and an enterokinase cleavage site (Asp-Asp-Asp-Asp-Lys) (Figure 4.1A). This leading sequence was expected to have pH-sensitive charge states because of the abundance in ionizable amino acids and was thus named S-tag. We chose these sequences because of their broad application in recombinant protein design. The T7 tag is a short epitope tag that allows identification of recombinant proteins using immunoblotting. In addition, the T7 tag has the potential to enhance recombinant protein expression in *E. coli*.[24, 34] The His tag has been widely used as a purification tag for immobilized metal affinity chromatography (IMAC). Although ELP-based recombinant proteins can be purified by an inverse transition cycling (ITC) method utilizing the LCST behavior,[35] chromatographic purification could be beneficial for applications where removal of endotoxin is necessary.[36-37] To address this need, removal of endotoxin using His-tag-based IMAC has been developed.[38-39] The enterokinase site is often harnessed to remove epitope and purification tags. The advantage of the enterokinase site is that the cleavage occurs after the Lys position and leaves no trace of the cleavage site in the protein of interest.[40] In addition to S-tag[YKV-48], we constructed two more S-tag proteins with longer YKV domains (i.e., S-tag[YKV-72] and S-tag[YKV-96] composed of 72 and 96 pentapeptide repeats, respectively).

Both the I-tag and S-tag[YKV] proteins were successfully purified using a previously described salting and heating method (Figure 4.1B).[27] Based on densitometry analysis of the SDS-PAGE gels, the purity of the lyophilized product was >95% for all proteins. Western blot using an anti-T7 tag antibody confirmed the presence of the T7 tag on the purified S-tag[YKV] proteins (Figure 4.1B). Western blot was not applicable for detecting I-tag[YKV-48] because there was no epitope tag for antibody recognition. The typical yield was 100-120 mg/L culture for the S-tag[YKV] proteins and 60-80 mg/L

culture for I-tag[YKV-48]. The difference in yields is consistent with previous reports that the T7 tag can improve recombinant protein expression in *E. coli*.[34, 41]

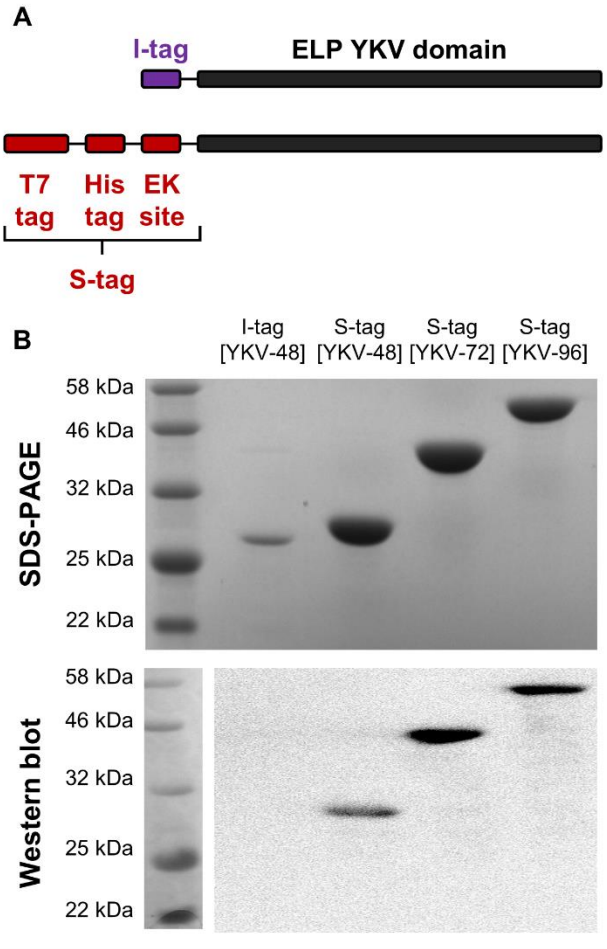


Figure 4.1. Design and production of the I-tag and S-tag[YKV] proteins. (A) Illustration of the differences between the I-tag and S-tag[YKV] proteins. All proteins shared a common ELP YKV domain. I-tag included the sequence of SKGPG. The S-tag was composed of a T7 tag, a 7x His tag, and an enterokinase cleavage site. (B) SDS-PAGE and Western blot images of purified I-tag and S-tag[YKV] proteins. All proteins on the SDS-PAGE gel were close to their expected molecular weights (I-tag[YKV-48]: 21 kDa, S-tag[YKV-48]: 25 kDa, S-tag[YKV-72]: 28.5 kDa, and S-tag[YKV-96]: 32 kDa). Western blot confirmed the presence of the T7 tag on the S-tag[YKV] proteins. Western blot was not applicable for I-tag[YKV-48] due to a lack of the T7 tag epitope used for detection.

4.4.2 Short Non-ELP Sequences Introduced pH Sensitivity to the LCST Behavior

To assess the effect of pH on LCST behavior of I-tag[YKV-48] and S-tag[YKV-48], we first determined the LCST by varying the pH from 5.5 to 8.0 in increments of 0.5 pH units

and focused on one protein concentration (50 μM protein solutions in APT buffer with 0.2 M NaCl). As shown in Figure 4.2A, the LCSTs of I-tag[YKV-48] at a concentration of 50 μM differ only by 2 °C across the whole pH range. Because the LCST behavior of ELP occurs over a temperature span of 2-3 °C,[42] a difference within 2-3 °C between two determined LCSTs is not considered to be significant. Therefore, we consider the LCST behavior of I-tag[YKV-48] to be pH insensitive. On the other hand, the LCSTs of S-tag[YKV-48] under the same conditions resulted in a much wider distribution than that of I-tag[YKV-48] (Figure 4.2B). The LCST of S-tag[YKV-48] was 76.3 °C at pH 5.5 and decreased to 48.1 °C at pH 8.0. Between pH 5.5 to 7.5, there was a >3 °C difference in LCSTs when comparing between two consecutive pH values. The difference in LCST between pH 7.5 and 8.0 was 2.3 °C. Altogether, at a 50 μM concentration, S-tag[YKV-48] showed pH-sensitive LCST behavior, whereas I-tag[YKV-48] did not.

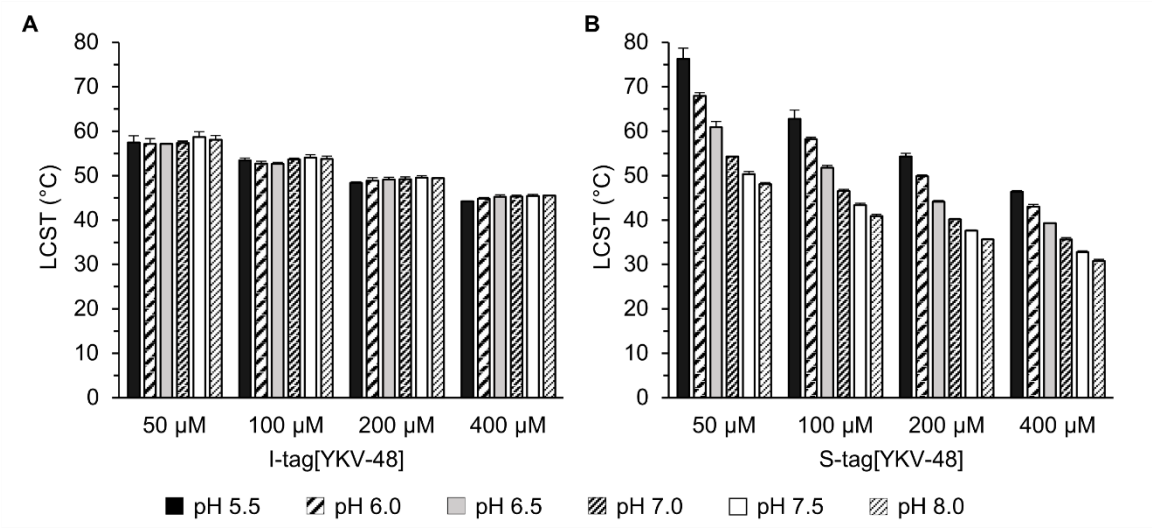


Figure 4.2. The LCSTs of I-tag[YKV-48] are insensitive to pH, whereas the LCSTs of S-tag[YKV-48] decrease significantly with increasing pH. LCST measurements of (A) I-tag[YKV-48] and (B) S-tag[YKV-48] at protein concentrations ranging from 50 to 400 μM and pH values ranging from 5.5 to 8.0. Both proteins had lower LCSTs with increasing protein concentration. At a given concentration, the LCSTs of I-tag[YKV-48] were similar at all pH values. On the other hand, S-tag[YKV-48] showed pH-sensitive LCSTs. NaCl was added to the solution for all groups at a final concentration of 0.2 M. The LCST values were presented as the average with standard deviation of three independent samples.

4.4.3 Protein Concentration Affected LCST Behavior and pH Sensitivity

To investigate the effect of protein concentration on LCST behavior and pH sensitivity, LCSTs were quantified at concentrations of 100, 200, and 400 μM for both I-tag and S-tag[YKV-48]. Similar to the results for 50 μM, pH-insensitive LCST behavior for I-tag[YKV-48] was observed for all concentrations tested (Figure 4.2A). However, the overall LCST decreased as protein concentration increased. This concentration dependence is characteristic of LCST behavior of ELPs.[14]

For S-tag[YKV-48], the LCST behavior was pH sensitive for all concentrations tested (Figure 4.2B). Furthermore, as observed with I-tag[YKV-48], the LCST of S-tag[YKV-48]at a given pH value decreased as the protein concentration increased. Examining the pH sensitivity of the LCST behavior in more detail revealed that there was a relationship to protein concentration. For example, at the lowest protein concentration of 50 μM the difference between LCSTs at pH 5.5 and at pH 8.0 was 28.2 °C whereas at the highest concentration of 400 μM the difference was 15.6 °C. Thus, as the protein concentration increased, the LCST behavior of S-tag[YKV-48] became less sensitive to pH.

The concentration dependence of the LCST behavior of ELPs can be described by an empirical equation[14, 17]

$$LCST = LCST|_{1\ \mu M,pH} - b_{pH} \ln[ELP] \quad (1)$$

where $LCST|_{1\ \mu M,pH}$ is the LCST at a reference protein concentration of 1 μM at a given pH, b_{pH} is the concentration dependence of the LCST behavior at a given pH, and [ELP] is the concentration of ELPs in μM. The LCST results at a given pH were fitted to equation 1 using the regstats function in Matlab and were plotted with the regression line generated using the best-fit parameters (Table 1) as shown in Figure 4.3. Both I-tag[YKV-48] and S-tag[YKV-48] measurements were described well by equation 1. The fitted LCSTs at 1 μM suggested that the LCST behavior of S-tag[YKV-48] was pH sensitive and that this sensitivity could be accurately described by equation 1. It should be noted, however, that some of the fitted LCST values were $\geq$100 °C and would not be observable in aqueous solutions under normal pressure.

Table 4.1. Reference LCSTs at 1 μM protein and concentration dependence b_{pH} of I-tag[YKV-48] and S-tag[YKV-48].[a]

	I-tag[YKV-48]		S-tag[YKV-48]	
pH	$LCST|_{1\ \mu M,pH}$ (°C)	b_{pH} (°C/ln[μM])	$LCST|_{1\ \mu M,pH}$ (°C)	b_{pH} (°C/ln[μM])
5.5	82.7 ± 1.0	6.4 ± 0.2	130.2 ± 6.8	14.2 ± 1.4
6.0	80.0 ± 0.6	5.9 ± 0.1	114.2 ± 3.4	12.0 ± 0.7
6.5	79.1 ± 1.0	5.7 ± 0.2	100.9 ± 4.9	10.5 ± 1.0
7.0	80.5 ± 0.7	5.9 ± 0.1	88.7 ± 3.7	9.0 ± 0.7
7.5	83.6 ± 0.6	6.4 ± 0.1	82.8 ± 2.5	8.4 ± 0.5
8.0	81.8 ± 0.6	6.1 ± 0.1	79.7 ± 2.8	8.2 ± 0.6

[a] Data on Figure 4.2 were fitted to equation 1. Parameters were reported as the fitted value with standard error.

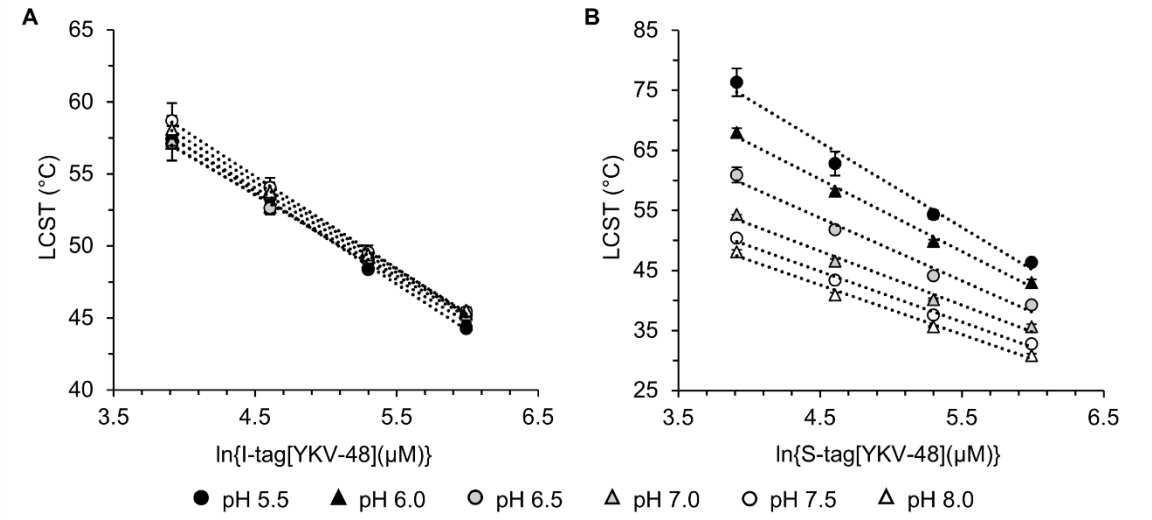


Figure 4.3. LCSTs of I-tag[YKV-48] and S-tag[YKV-48] were well described by a two-parameter equation. LCST measurements with best-fit lines from equation 1 of (A) I-tag[YKV-48] and (B) S-tag[YKV-48]. The LCST values were presented as the average with standard deviation of three independent samples.

Figure 4.4 shows the fitted b_{pH} parameters of S-tag[YKV-48] and I-tag[YKV-48] as a function of pH. I-tag[YKV-48] showed a nearly constant b_{pH} across the pH range with an R^2 value of 0.0002 when a linear regression was performed. The low R^2 value indicated that the response variable (b_{pH}) was not dependent on the predictor variable (pH) and therefore the concentration dependence of the LCST behavior of I-tag[YKV-48] was pH insensitive. On the other hand, S-tag[YKV-48] showed a decreasing trend in b_{pH} as pH increased. This trend indicated that S-tag[YKV-48] exhibited a pH-sensitive concentration dependence in LCST behavior, and the LCST behavior of S-tag[YKV-48] was more sensitive to changes in protein concentration at lower pH values than at higher pH values. As a result, the LCST distribution of S-tag[YKV-48] was narrower at higher protein concentrations (Figure 4.2B). Furthermore, at lower pH values (5.5 to 7.0), the slope was steeper than at higher pH values (7.0 to 8.0). This change in the slope indicated that the concentration dependence of S-tag[YKV-48] became less sensitive to pH at high pH values.

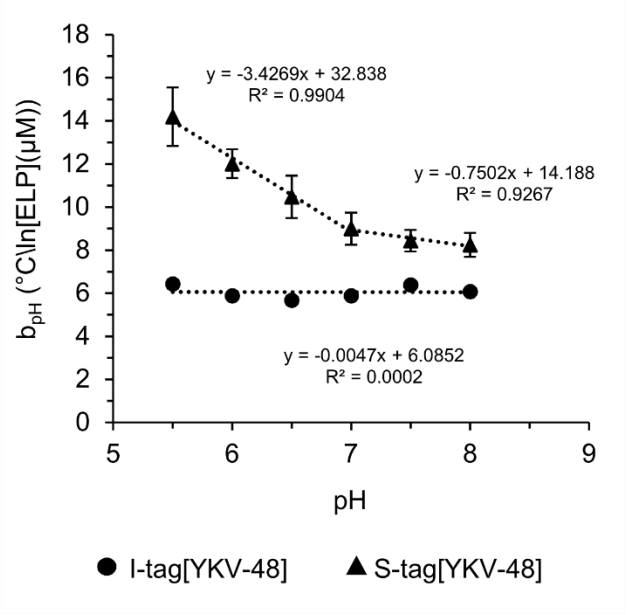


Figure 4.4. Different pH sensitivity in concentration dependence (b_{pH}) between I-tag[YKV-48] and S-tag[YKV-48]. Fitted b_{pH} of I-tag and S-tag[YKV-48] with standard error were plotted as a function of pH. Linear regression lines were included to show the trend in b_{pH}. The b_{pH} values for I-tag[YKV-48] yielded a nearly horizontal line with an R^2 close to zero and indicated that b_{pH} was not sensitive to pH. On the other hand, the b_{pH} values for S-tag[YKV-48] were sensitive to pH and exhibited two different trends below and above pH 7.

4.4.4 pH Sensitivity Resulted from the Protonation State of Histidine

To explain different trends in b_{pH}, we analyzed the protein charge as a function of pH. The charge of the amino acid sequence was calculated using the Henderson-Hasselbalch equation by assuming that the pKa value of each residue was the same as the isolated amino acid and was not affected by the protein structure. The calculations were performed separately on the YKV domain and the N-terminal sequences as indicated in Supporting Information Figure 4.11A. The YKV domain had a nearly constant charge from pH 5.5 to pH 8.0 (Supporting Information Figure 4.11B). This constant charge was expected because the most abundant ionizable residues were Lys with a pKa of 10.7 and Tyr with a pKa of 10.1. The charge on the I-tag was also nearly constant over the same pH range, whereas the S-tag showed a pH-dependent charge because of the ionizable His residues in the His tag. Histidine had a pKa of 6.0 and was positively charged below pH 6.0. On the other hand, Asp residues in the enterokinase cleavage site carried negative charges when the pH was above the Asp pKa of 3.7. As a result, the charge on the S-tag transitioned from positive to negative at approximately pH 6.3 as His residues became deprotonated and uncharged, and this pH value was close to where the b_{pH} trend changed in slope at a pH of approximately 7.0. The deprotonation of histidine also corresponded to the pH-sensitive LCST behavior of S-tag[YKV-48] observed from pH 5.5 to 8.0.

Previous studies have utilized histidine at the guest residue position to construct ELPs with pH-sensitive LCST behavior over the same pH range.[17] Deprotonation of His resulted in increased hydrophobicity and decreased LCST. Placing histidine adjacent to ELP domains was also reported to lower the LCST of an uncharged ELP at pH 7.4.[19] However, it has not been shown whether the change in hydrophobicity and charge state of His outside of an ELP domain can affect the pH-sensitive LCST behavior. Our S-tag[YKV-48] demonstrated that His, when not incorporated as a guest residue, can still influence the ELP LCST behavior and can be harnessed to introduce pH sensitivity to the LCST behavior.

4.4.5 The Extent of pH sensitivity Depended on ELP Domain Length

We next examined to what extent the pH sensitivity introduced by the S-tag persisted when the ratio of the lengths of the external sequences to the ELP was varied. To do so, the YKV domain was elongated to 72 (S-tag[YKV-72]) and 96 repeats (S-tag[YKV-96]). All

proteins were prepared at 200 μM in APT buffer with 0.2 M NaCl, and the LCSTs are shown in Figure 4.5. As the ELP domain length increased, the LCST became less sensitive to pH (Figure 4.5). Furthermore, at a given pH, LCSTs decreased with longer YKV domains. This trend was consistent with previous work in which longer ELPs with the same guest residue composition exhibited lower LCSTs.[14, 41]

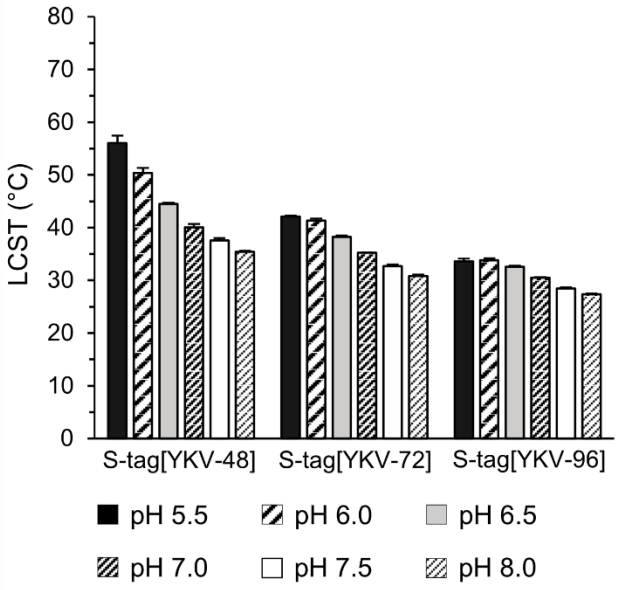


Figure 4.5. Longer YKV domain lengths decreased the extent of pH sensitivity derived from the S-tag sequence. At 200 μM protein and 0.2 M NaCl, pH sensitivity was more prominent at a lower pH region (pH 5.5 to 6.5). The data set of S-tag[YKV-48] is replotted from Figure 4.2B. LCST values were presented as the average with standard deviation of three independent samples.

4.4.6 Increasing Ionic Strength Shielded pH-Sensitive LCST Behavior Introduced by the Tag Sequence

Because both the S-tag and the YKV domain are charged, we hypothesized that interactions between the S-tag and the YKV domain involved electrostatic interactions. Increasing the ionic strength (i.e., salt concentration) was expected to shield the interactions between the S-tag and YKV domain. As a result, the LCST behavior of S-tag[YKV-48] was expected to become less pH sensitive as salt concentration increased.

LCSTs were measured with increasing NaCl concentration from 0.1 to 0.8 M at a fixed protein concentration of 200 μM. As shown in Figure 4.6A, the LCSTs of I-tag[YKV-48] depended on NaCl concentration but remained pH insensitive. This result reflected the

constant charge on both the I-tag and YKV domain over this pH range. On the other hand, the LCST behavior of S-tag[YKV-48] remained pH sensitive over all the NaCl concentrations (Figure 4.6B). It should be noted that as NaCl concentration increased, the LCST became less sensitive to pH but was still pH sensitive even at the highest NaCl concentrations tested. In fact, the differences in LCSTs between two consecutive pH values were <2 °C at 0.6 and 0.8 M NaCl, but the differences between pH 5.5 and pH 8.0 were significant and still larger than 2 °C (5.6 and 4.4 °C, respectively) at 0.6 and 0.8 M NaCl.

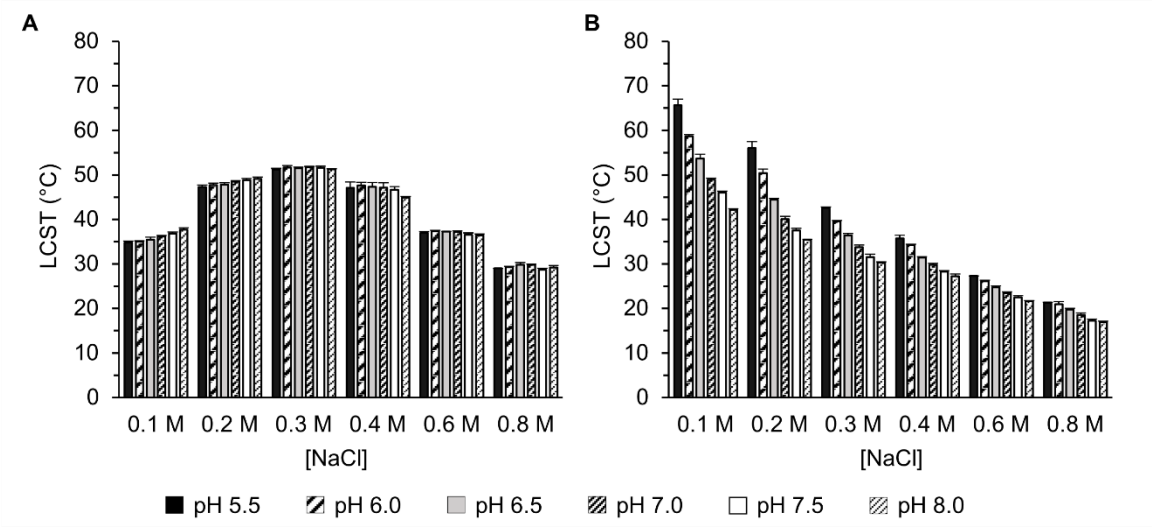


Figure 4.6. High ionic strength shielded the electrostatic interactions between S-tag and the YKV-48 domain. LCST measurements of (A) I-tag and (B) S-tag[YKV-48] at different NaCl concentrations with 200 μM protein. The LCST behavior of I-tag[YKV-48] remained insensitive to pH over all NaCl concentrations. On the other hand, the LCST behavior of S-tag[YKV-48] showed decreasing pH sensitivity with increasing NaCl concentration. In addition, the two proteins responded differently to the increasing NaCl concentration. The LCST of I-tag[YKV-48] initially increased until 0.3 M NaCl then decreased. In contrast, the LCSTs of S-tag[YKV-48] decreased monotonically with increasing NaCl concentration. LCST values were presented as the average with standard deviation of three independent samples.

4.4.7 Hydrophobicity of the Tag Affected LCST Behavior

The hydrophobicity of ELP domains and external sequences also determined the LCST behavior of ELP proteins. To examine the hydrophobicity difference between the two tags as a function of pH, we used an estimation method proposed by Trabbic-Carlson *et al.*[19] This estimation was based on the hydrophobicity scale established by Urry[12] that was originally used to predict the LCST of an ELP with a known guest residue composition. It

should be noted that this estimation had no direct physical meaning and was only used to demonstrate the difference in the relative hydrophobicity between the tags in an ELP-fusion protein context. Similar to its charge state, the I-tag had a nearly constant relative hydrophobicity, and the S-tag had a pH-dependent relative hydrophobicity (Supporting Information Figure 4.12). The pH dependence of the S-tag also resulted from the deprotonation of His and subsequent increase in hydrophobicity at pH values above its pKa. In addition, the S-tag was relatively more hydrophobic than the I-tag from pH 5.5 to 8.0, and thus S-tag[YKV-48] should have lower LCST values than I-tag[YKV-48] under the same conditions. However, this finding based on the hydrophobicity analysis was in conflict with the results shown in Figure 4.2 and 6. By plotting the difference in LCSTs between the S-tag[YKV-48] and I-tag[YKV-48] proteins, it was clearly shown that S-tag[YKV-48] had higher LCSTs than I-tag[YKV-48] from pH 5.5 to 6.5 at a protein concentration of 50 μM, from 5.5 to 6.0 at protein concentrations of 100 and 200 μM, and only at pH 5.5 at a protein concentration of 400 μM (Figure 4.7A). With increasing ionic strength ($\geq$0.3 M NaCl), S-tag[YKV-48] had lower LCSTs than I-tag[YKV-48] at all pH values (Figure 4.7B).

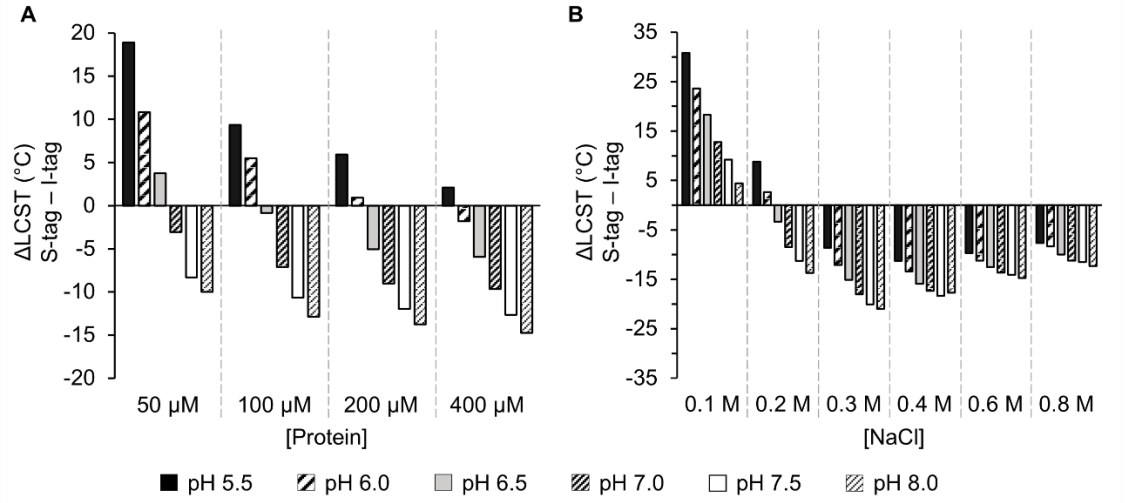


Figure 4.7. High protein concentration and high ionic strength decreased the LCSTs of S-tag[YKV-48] more significantly then those of I-tag[YKV-48]. LCST differences between S-tag and I-tag[YKV-48] with different (A) protein concentrations and (B) NaCl concentrations. S-tag[YKV-48] had higher LCSTs than I-tag[YKV-48] at low pH values and low protein concentrations. As the protein concentration increased, S-tag[YKV-48] only had a higher LCST at pH 5.5. With varying NaCl concentration, S-tag[YKV-48] only had higher LCSTs at NaCl concentrations $\leq$0.2 M. The plotted differences were based on the averaged LCSTs reported in Figure 4.2 and 6.

Based on the charge estimation (Supporting Information Figure 4.11B), both the S-tag and the YKV domain were positively charged from pH 5.5-6.5 and oppositely charged above pH 6.5 where the S-tag became negatively charged. We therefore hypothesized that the interaction between the S-tag and the YKV domain is a combination of electrostatic and hydrophobic interactions, and the dominant type of interaction depended on pH. At low ionic strengths and low pH values, the repulsive charge interactions between the S-tag and the YKV domain hindered the collapse of the protein chain and thus led to higher LCSTs. At low ionic strengths and high pH values, the attractive charge interactions between the S-tag and the YKV domain worked synergistically with the increasing hydrophobicity of the S-tag and resulted in lower LCSTs. The electrostatic interactions were shielded by increasing ionic strength, and the hydrophobic interactions dominated the LCST behavior. As a result, S-tag[YKV-48] exhibited lower LCSTs than I-tag[YKV-48] at higher pH values and ≥0.3 M NaCl, and the LCST differences between the two proteins became smaller as the NaCl concentration increased above 0.3 M (Figure 4.7B). This trend also supported the hypothesis that the interaction between the S-tag and the YKV domain involved electrostatic interactions. The attractive electrostatic interactions between the S-tag and the YKV domain at pH values >6.5 facilitated the collapse of the protein chain, and therefore contributed to much lower LCSTs than those of I-tag[YKV-48] at 0.3 M NaCl. At higher ionic strengths, the attractive electrostatic interactions were shielded, so the LCST behavior was mainly driven by hydrophobic interactions, and the difference in LCSTs between S-tag[YKV-48] and I-tag[YKV-48] became smaller.

It should be noted that the LCST behavior observed at 0.1 M NaCl was not fully explained by our hypothesis. Specifically, S-tag[YKV-48] had higher LCSTs than I-tag[YKV-48] from pH 7.0 to 8.0. Based on our hypothesis, under this low ionic strength, both attractive electrostatic interactions between the S-tag and the YKV domain and the increased hydrophobicity on the S-tag should lead to LCSTs that are lower than those of I-tag[YKV-48]. This inconsistency suggested that there were more factors that needed to be considered when explaining the difference in the LCST behaviors reported here. It has been proposed that the effect of the interfacial tension caused by ions at the water-protein interface can be critical when describing LCST behavior of charged systems, and both cations and anions contributed to the effect.[43] Therefore, we further explored the effects

from the tags on changing the water-protein interfacial tension by determining LCSTs of I-tag[YKV-48] and S-tag[YKV-48] with different cosolutes.

4.4.8 Tag Identity Affected the Interaction between ELP Proteins and Hofmeister Ions

In Figure 4.6, it was clear that the LCST behavior of I-tag[YKV-48] and S-tag[YKV-48] responded differently to an increased NaCl concentration. While S-tag[YKV-48] showed a monotonic decreasing trend, I-tag[YKV-48] exhibited an increasing LCST below 0.3 M NaCl and a decreasing trend above 0.3 M NaCl. The initial increasing LCST trend has been reported as a salting-in effect on the ELP LCST behavior with chaotropic Hofmeister anions.[15] ELP LCST behavior was influenced by the interaction among Hofmeister ions, water molecules, and ELP protein chains.[44] Because both I-tag[YKV-48] and S-tag[YKV-48] were positively charged over the pH range investigated, we expected interactions with anions to dominate. Cl^- is generally considered to be in the middle of Hofmeister series.[15, 45] Therefore, it was surprising to observe a strong salting-in effect with I-tag[YKV-48]. Since Cl^- exerted a typical salting-out effect on S-tag[YKV-48], we hypothesized that the salting-in effect from Cl^- arose from the distinct tag sequences and not from the YKV domain.

To assess the effect of the tags on interactions with Hofmeister anions, LCST measurements were taken in the presence of SO_4^{2-}, a kosmotropic anion, or I^-, which is a chaotrope. Figure 4.8A shows that SO_4^{2-} had a strong salting-out effect on S-tag[YKV-48]. The LCST decreased by 62.9 °C from 0.1 to 0.6 M SO_4^{2-}. In fact, SO_4^{2-} is generally considered to be one of the strongest kosmotropes in the Hofmeister series. Remarkably, as shown in Figure 4.8A, SO_4^{2-} showed a slight salting-in effect on I-tag[YKV-48]. The LCST increased by only 3.4 °C from 0.1 to 0.15 M SO_4^{2-} and by merely 0.9 °C from 0.15 to 0.2 M SO_4^{2-}. Note that, unlike the data with NaCl, the measurements were taken at a protein concentration of 50 μM instead of 200 μM and at a pH of 5.5 because the LCST was <4 °C or the protein was insoluble at 200 μM at higher SO_4^{2-} concentrations.

Figure 4.8B shows the LCST of 200 μM S-tag and I-tag[YKV-48] with 0.1 to 1.2 M I^- at pH 7.5. Both proteins had elevated LCSTs at higher I^- concentrations (>0.2 M with S-tag[YKV-48] and >0.4 M with I-tag[YKV-48]). This salting-in effect was expected because I^- is on the chaotropic end of the Hofmeister series. However, the unusual

decreasing trend in LCST before the salting-in concentration was surprising. To the best of our knowledge, only one similar trend has been reported with another positively-charged ELP with a guest residue composition of one Lys and six Val.[46]

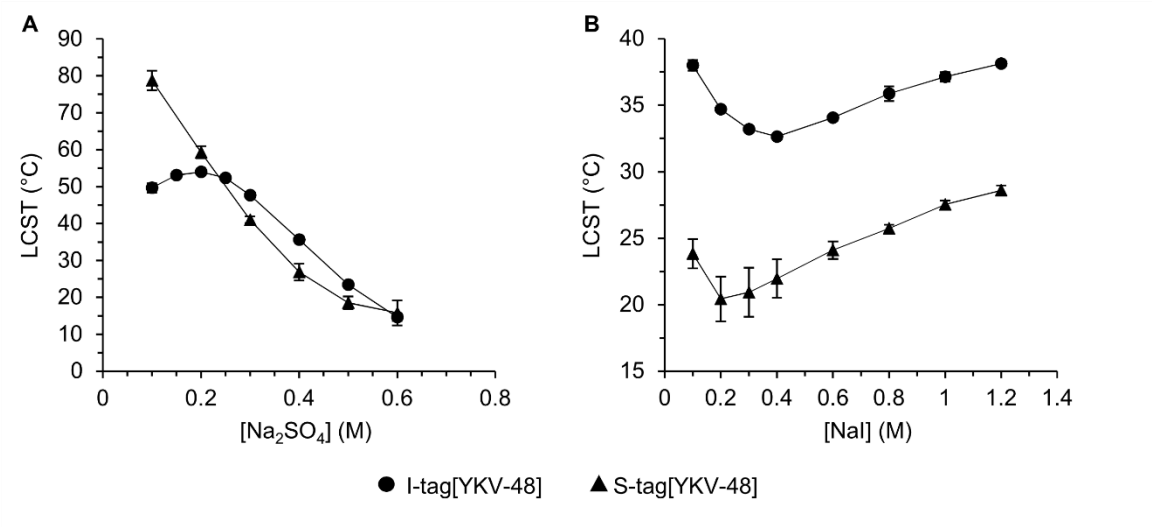


Figure 4.8. External tag identity affected the interactions between ELPs and Hofmeister ions. LCST measurements of I-tag[YKV-48] and S-tag[YKV-48] with (A) Na_2SO_4 and (B) NaI. Both I-tag[YKV-48] and S-tag[YKV-48] had decreasing LCSTs at >0.2 M Na_2SO_4. I-tag[YKV-48] showed an increasing trend in LCST from 0.1 to 0.2 M Na_2SO_4 even with a strong kosmotropic ion such as SO_4^{2-}. With NaI, both proteins exhibited a similar trend that LCSTs decreased before 0.4 M NaI and increased after that. LCSTs with Na_2SO_4 were determined at a protein concentration of 50 μM at pH 5.5. LCSTs with NaI were determined at a protein concentration of 200 μM at pH 7.5. The presented LCST values were averages of three independent samples with standard deviation.

The observed responses from the LCST behavior of I-tag[YKV-48] to three Hofmeister anions suggested that a broader range of anions can exert a salting-in effect when compared to S-tag[YKV-48]. Different mechanisms have been proposed to explain the salting-in effect induced by chaotropic anions. Examples include electrostatic interactions between anions and backbone amides[15] and hydrophobic interactions between chaotropic anions and hydrophobic pockets in the protein.[47] In the electrostatic interaction hypothesis, direct binding of anions to backbone amides is enthalpically favorable toward weakly hydrated anions (i.e., chaotropic anions), and the binding increases the net charge on the protein and thus prevents hydrophobic collapse. Because of the sequence differences between I-tag and S-tag, the enthalpy gains from binding of anions to backbone amides could be different and result in the shift in the Hofmeister series observed between I-

tag[YKV-48] and S-tag[YKV-48]. In the hydrophobic interaction hypothesis, binding of weakly hydrated anions is entropically favorable. This binding competes with the interaction between two hydrophobic surfaces and therefore destabilizes the protein by increasing solubility of hydrophobic pockets. As shown in Supporting Information Figure 4.12, the I-tag was more hydrophilic than the S-tag based on the relative hydrophobicities. Therefore, I-tag[YKV-48] should have had weaker hydrophobic interactions, which resulted in a broader range of hydrated anions in the Hoffmeister series that exerted a salting-in effect on the LCST behavior of I-tag[YKV-48].

4.5 Conclusions

The LCST behavior of ELPs is largely determined by the guest residue composition of the ELP domain, but few studies have explored the interaction between short external non-ELP sequences and ELP domains. In this work, we systematically studied the interactions between non-ELP sequences and ELP domains by constructing a series of ELP proteins bearing the same positively-charged ELP domain. S-tag was designed to have pH-dependent charge and hydrophobicity whereas I-tag had constant charge and hydrophobicity over the tested pH range. Our results showed that pH sensitivity was introduced to the LCST behavior by interactions between the pH-dependent S-tag and the positively-charged YKV domain. We hypothesized that the interaction involved both electrostatic and hydrophobic interactions, and the two interaction types affected the LCST behavior differently. Increasing the ionic strength shielded the electrostatic interactions and resulted in hydrophobic interactions being the main determinant of LCST behavior. The identity of non-ELP sequences also changed the response to different Hofmeister anions. The more hydrophilic non-ELP sequence, I-tag, resulted in a broader range of anions capable of causing a salting-in effect on the LCST behavior. Our results suggested that both electrostatic and hydrophobic interactions between non-ELP and ELP sequences participated in LCST behavior, and rational design of non-ELP external sequences can be harnessed to construct ELP proteins with LCST behavior that can respond more precisely to environmental changes.

4.6 Acknowledgements

This work was supported by the National Institutes of Health (NIAMS R21AR065644 and NIDCR R03DE021755), the National Science Foundation (CBET-1512285) and the American Heart Association Scientist Development Grant (12SDG8980014). We thank the Purdue University College of Pharmacy for access to their Crystal16 for LCST measurements.

4.7 Supporting Information

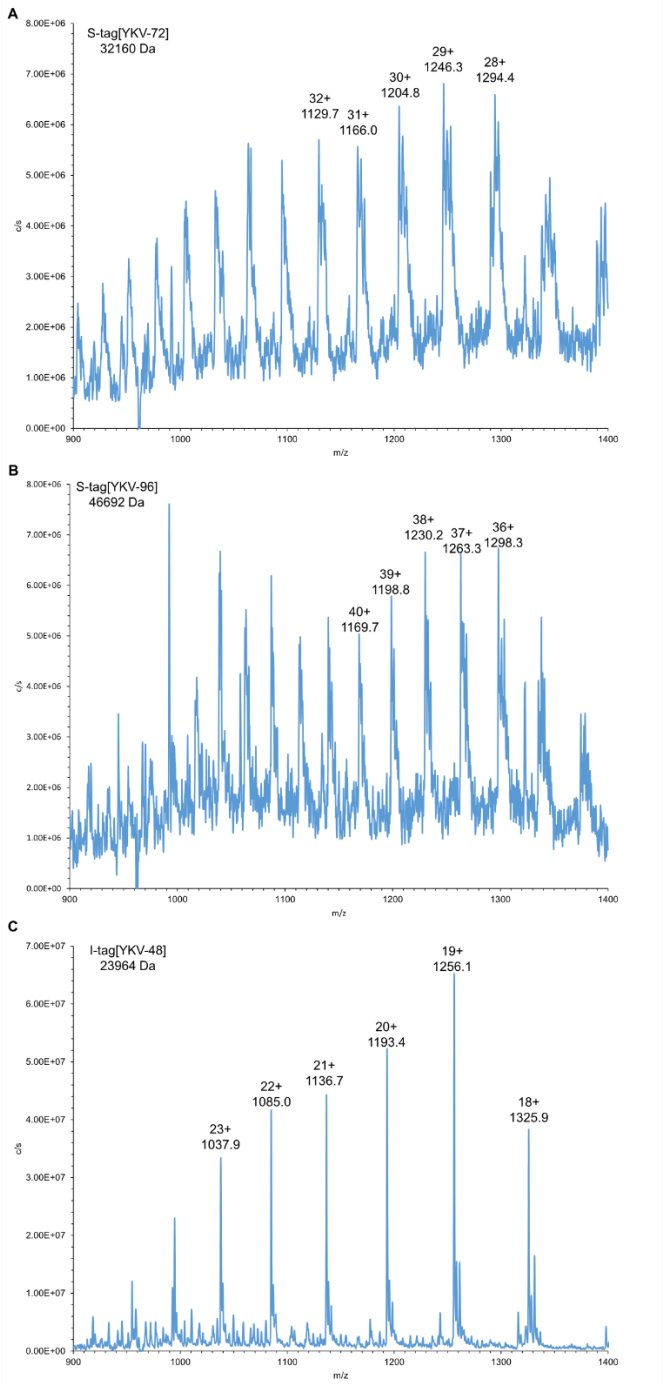


Figure 4.9. Electrospray ionization mass spectroscopy spectra of (A) S-tag[YKV-72] (expected MW: 32160 Da), (B) S-tag[YKV-96] (expected MW: 46692 Da), and (C) I-tag[YKV-48] (expected MW: 23964 Da). Peaks were labeled with the m/z ratios and the corresponding number of charges on the polymer chains.

	I-tag[YKV-48]		
	expected mol%	measured mol%	error
Asp+Asn	1.15	1.58	0.43
Thr	0.38	0.65	0.27
Ser	0.38	0.50	0.12
Glu+Gln	0.38	1.36	0.98
Pro	19.23	18.23	1.00
Gly	38.46	36.99	1.47
Ala	0.38	1.07	0.69
Val	25.38	23.73	1.65
Ile	0.00	0.26	0.26
Leu	0.77	1.37	0.60
Tyr	6.15	5.83	0.32
Phe	0.00	0.49	0.49
His	0.00	0.11	0.11
Lys	6.54	6.96	0.42
Arg	0.77	0.86	0.09

Figure 4.10. Amino acid analysis results of I-tag[YKV-48]. The result showed a good match between the measured molar fraction and the expected molar fraction of each amino acid.

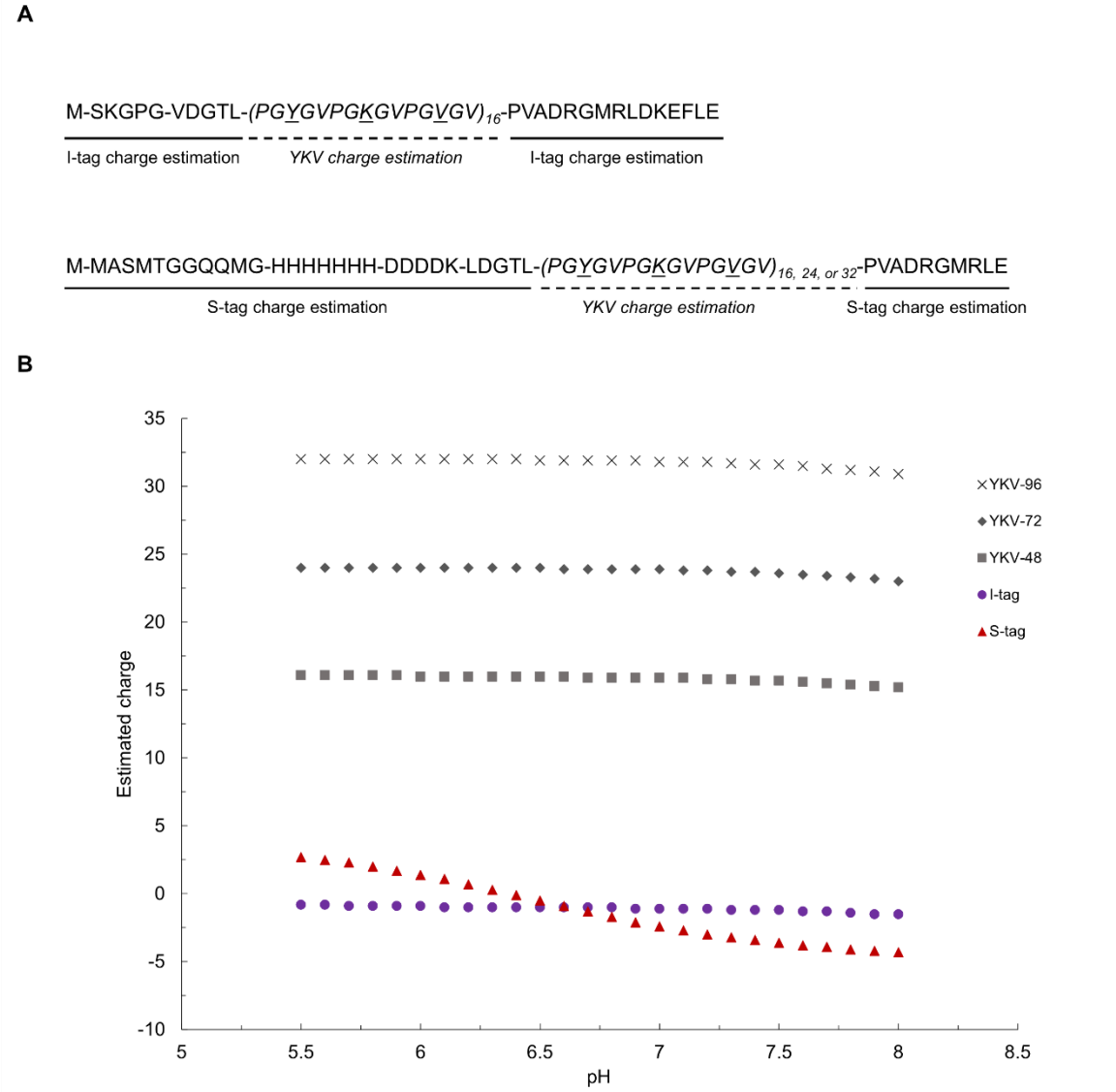


Figure 4.11. (A) Amino acid sequences of I-tag[YKV-48] and S-tag[YKV-48]. Each repeat of (PGYGVPGKGVPGVGV) contains three ELP pentapeptide sequences with guest residues being Y, K, and V. The 16, 24, or 32 repeats contain a total of 48, 72, or 96 pentapeptide sequences, respectively. (B) Charge estimation of I-tag, S-tag, and YKV sequences as indicated in (A). The charge estimation of the amino acid sequence was calculated using the Henderson-Hasselbalch equation by assuming that the pKa value of each residue was the same as the isolated amino acid and was not affected by the protein structure.

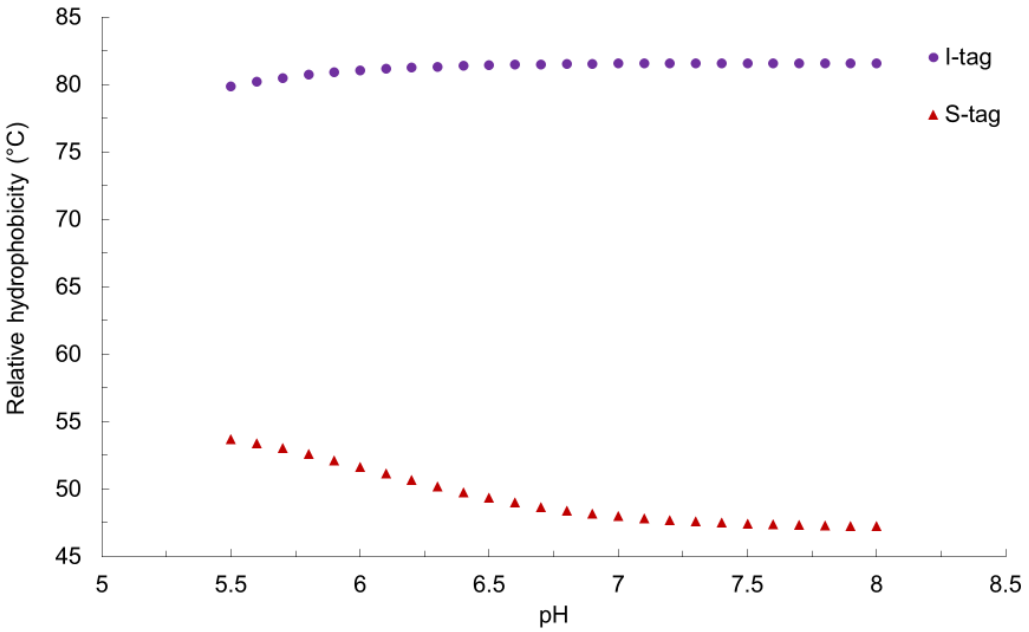


Figure 4.12. Relative hydrophobicity of the I-tag and the S-tag sequences. The sequences of each domain used for the estimation were indicated in Supporting Information Figure 4.11A. The relative hydrophobicity was calculated using a method proposed by Trabbic-Carlson et al.[19] This estimation was based on the hydrophobicity scale established by Urry et al.[12] that was originally used to predict the LCST of an ELP with a known guest residue composition. More hydrophobic guest residue compositions resulted in lower LCST values. It should be noted that this estimation had no direct physical meaning and was only used to demonstrate the difference in the relative hydrophobicity between the tags in an ELP-fusion protein context.

4.8 References

(1) Kim, Y.-J.; Matsunaga, Y. T. Thermo-responsive polymers and their application as smart biomaterials. *J. Mater. Chem. B* **2017,** *5* (23), 4307-4321.

(2) Wei, M.; Gao, Y.; Li, X.; Serpe, M. J. Stimuli-responsive polymers and their applications. *Polym. Chem.* **2017,** *8* (1), 127-143.

(3) Hrubý, M.; Filippov, S. K.; Štěpánek, P. Smart polymers in drug delivery systems on crossroads: Which way deserves following? *Eur. Polym. J.* **2015,** *65*, 82-97.

(4) Ashraf, S.; Park, H.-K.; Park, H.; Lee, S.-H. Snapshot of phase transition in thermoresponsive hydrogel PNIPAM: Role in drug delivery and tissue engineering. *Macromol. Res.* **2016,** *24* (4), 297-304.

(5) Nash, M. E.; Healy, D.; Carroll, W. M.; Elvira, C.; Rochev, Y. A. Cell and cell sheet recovery from pNIPAm coatings; motivation and history to present day approaches. *J. Mater. Chem.* **2012,** *22* (37), 19376-19389.

(6) Cox, B. A.; Starcher, B. C.; Urry, D. W. Coacervation of tropoelastin results in fiber formation. *J. Biol. Chem.* **1974,** *249* (3), 997-998.

(7) Rodríguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A. Elastin-like polypeptides in drug delivery. *Adv. Drug Delivery Rev.* **2016,** *97*, 85-100.

(8) MacEwan, S. R.; Chilkoti, A. Elastin-like polypeptides: Biomedical applications of tunable biopolymers. *Pept. Sci.* **2010,** *94* (1), 60-77.

(9) Kowalczyk, T.; Hnatuszko-Konka, K.; Gerszberg, A.; Kononowicz, A. K. Elastin-like polypeptides as a promising family of genetically-engineered protein based polymers. *World J. Microbiol. Biotechnol.* **2014,** *30* (8), 2141-2152.

(10) Li, B.; Alonso, D. O.; Daggett, V. The molecular basis for the inverse temperature transition of elastin. *J. Mol. Biol.* **2001,** *305* (3), 581-592.

(11) Li, N. K.; Quiroz, F. G.; Hall, C. K.; Chilkoti, A.; Yingling, Y. G. Molecular description of the LCST behavior of an elastin-like polypeptide. *Biomacromolecules* **2014,** *15* (10), 3522-3530.

(12) Urry, D. W.; Gowda, D.; Parker, T. M.; Luan, C. H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992,** *32* (9), 1243-1250.

(13) Smits, F. C.; Buddingh, B. C.; van Eldijk, M. B.; van Hest, J. C. Elastin-like polypeptide based nanoparticles: Design rationale toward nanomedicine. *Macromol. Biosci.* **2015,** *15* (1), 36-51.

(14) Meyer, D. E.; Chilkoti, A. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules* **2004,** *5* (3), 846-851.

(15) Cho, Y.; Zhang, Y.; Christensen, T.; Sagle, L. B.; Chilkoti, A.; Cremer, P. S. Effects of Hofmeister anions on the phase transition temperature of elastin-like polypeptides. *J. Phys. Chem. B* **2008,** *112* (44), 13765-13771.

(16) Kherb, J.; Flores, S. C.; Cremer, P. S. Role of carboxylate side chains in the cation Hofmeister series. *J. Phys. Chem. B* **2012,** *116* (25), 7389-7397.

(17) MacKay, J. A.; Callahan, D. J.; FitzGerald, K. N.; Chilkoti, A. Quantitative model of the phase behavior of recombinant pH-responsive elastin-like polypeptides. *Biomacromolecules* **2010,** *11* (11), 2873-2879.

(18) Meyer, D. E.; Chilkoti, A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **1999,** *17* (11), 1112-1115.

(19) Trabbic- Carlson, K.; Meyer, D.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A. Effect of protein fusion on the transition temperature of an environmentally responsive elastin-like polypeptide: A role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66.

(20) Christensen, T.; Amiram, M.; Dagher, S.; Trabbic- Carlson, K.; Shamji, M. F.; Setton, L. A.; Chilkoti, A. Fusion order controls expression level and activity of elastin-like polypeptide fusion proteins. *Protein Sci.* **2009,** *18* (7), 1377-1387.

(21) Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D. Topological effects on globular protein-ELP fusion block copolymer self-assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

(22) Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D. Effect of ELP sequence and fusion protein design on concentrated solution self-assembly. *Biomacromolecules* **2016,** *17* (3), 928-934.

(23) Carson, M.; Johnson, D. H.; McDonald, H.; Brouillette, C.; DeLucas, L. J. His-tag impact on structure. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007,** *63* (3), 295-301.

(24) Stevens, R. C. Design of high-throughput methods of protein production for structural biology. *Structure* **2000,** *8* (9), R177-R185.

(25) Homma, K.; Moriyama, H. Crystallization and crystal-packing studies of Chlorella virus deoxyuridine triphosphatase. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2009,** *65* (10), 1030-1034.

(26) Lomakin, I. B.; Stolboushkina, E. A.; Vaidya, A. T.; Zhao, C.; Garber, M. B.; Dmitriev, S. E.; Steitz, T. A. Crystal structure of the human ribosome in complex with DENR-MCT-1. *Cell Rep.* **2017,** *20* (3), 521-528.

(27) Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

(28) Brennan, M. J.; Hollingshead, S. E.; Wilker, J. J.; Liu, J. C. Critical factors for the bulk adhesion of engineered elastomeric proteins. *R. Soc. Open Sci.* **2018,** *5* (5), 171225.

(29) McDaniel, J. R.; MacKay, J. A.; Quiroz, F. G.; Chilkoti, A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **2010,** *11* (4), 944-952.

(30) Renner, J. N.; Kim, Y.; Cherry, K. M.; Liu, J. C. Modular cloning and protein expression of long, repetitive resilin-based proteins. *Protein Expr. Purif.* **2012,** *82* (1), 90-96.

(31) McDaniel, J. R.; Weitzhandler, I.; Prevost, S.; Vargo, K. B.; Appavou, M.-S.; Hammer, D. A.; Gradzielski, M.; Chilkoti, A. Noncanonical self-assembly of highly asymmetric genetically encoded polypeptide amphiphiles into cylindrical micelles. *Nano Lett.* **2014,** *14* (11), 6590-6598.

(32) Mukerji, R.; Schaal, J.; Li, X.; Bhattacharyya, J.; Asai, D.; Zalutsky, M. R.; Chilkoti, A.; Liu, W. Spatiotemporally photoradiation-controlled intratumoral depot for combination of brachytherapy and photodynamic therapy for solid tumor. *Biomaterials* **2016,** *79*, 79-87.

(33) Meyer, D. E.; Chilkoti, A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: Examples from the elastin-like polypeptide system. *Biomacromolecules* **2002,** *3* (2), 357-367.

(34) Kimple, M. E.; Brill, A. L.; Pasker, R. L. Overview of affinity tags for protein purification. *Curr. Protoc. Protein Sci.* **2013,** *73* (1), 9.9. 1-9.9. 23.

(35) Fletcher, E. E.; Yan, D.; Kosiba, A. A.; Zhou, Y.; Shi, H. Biotechnological applications of elastin-like polypeptides and the inverse transition cycle in the pharmaceutical industry. *Protein Expr. Purif.* **2019,** *153*, 114-120.

(36) Magalhães, P. O.; Lopes, A. M.; Mazzola, P. G.; Rangel-Yagui, C.; Penna, T.; Pessoa Jr, A. Methods of endotoxin removal from biological preparations: A review. *J. Pharm. Pharm. Sci.* **2007,** *10* (3), 388-404.

(37) Saraswat, M.; Musante, L.; Ravidá, A.; Shortt, B.; Byrne, B.; Holthofer, H. Preparative purification of recombinant proteins: Current status and future trends. *BioMed Res. Int.* **2013,** *2013*.

(38) Franken, K. L.; Hiemstra, H. S.; van Meijgaarden, K. E.; Subronto, Y.; Den Hartigh, J.; Ottenhoff, T. H.; Drijfhout, J. W. Purification of his-tagged proteins by immobilized chelate affinity chromatography: The benefits from the use of organic solvent. *Protein Expr. Purif.* **2000,** *18* (1), 95-99.

(39) Reichelt, P.; Schwarz, C.; Donzeau, M. Single step protocol to purify recombinant proteins with low endotoxin contents. *Protein Expr. Purif.* **2006,** *46* (2), 483-488.

(40) Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **2011,** *80* (2), 283-293.

(41) McDaniel, J. R.; Radford, D. C.; Chilkoti, A. A unified model for de novo design of elastin-like polypeptides with tunable inverse transition temperatures. *Biomacromolecules* **2013,** *14* (8), 2866-2872.

(42) Despanie, J.; Dhandhukia, J. P.; Hamm-Alvarez, S. F.; MacKay, J. A. Elastin-like polypeptides: Therapeutic applications for an emerging class of nanomedicines. *J. Controlled Release* **2016,** *240*, 93-108.

(43) Zhang, Y.; Cremer, P. S. Chemistry of Hofmeister anions and osmolytes. *Annu. Rev. Phys. Chem.* **2010,** *61*, 63-83.

(44) Okur, H. I.; Hladílková, J.; Rembert, K. B.; Cho, Y.; Heyda, J.; Dzubiella, J.; Cremer, P. S.; Jungwirth, P. Beyond the Hofmeister series: Ion-specific effects on proteins and their biological functions. *J. Phys. Chem. B* **2017,** *121* (9), 1997-2014.

(45) Zhang, Y.; Cremer, P. S. Interactions between macromolecules and ions: The Hofmeister series. *Curr. Opin. Chem. Biol.* **2006,** *10* (6), 658-663.

(46) Cho, Y. H. Thermodynamics and applications of elastin-like polypeptides. Doctor of Philosophy Texas A&M University, College Station, Texas, 2009.

(47) Gibb, C. L.; Gibb, B. C. Anion binding to hydrophobic concavity is central to the salting-in effects of Hofmeister chaotropes. *J. Am. Chem. Soc.* **2011,** *133* (19), 7344-7347.

5. CONCLUSIONS

5.1 Summary

This work showcased developments of process approaches and materials toward functional biomaterials. First, a new thiol-crosslinking approach was developed and was demonstrated for its potential in biological applications. The thiol basis was chosen for its easy incorporation into materials designs of both synthetic materials and recombinant protein-based materials, relatively biorthogonal reactivity, and its dual function as a crossing group and redox-responsive junction. Hydrogels crosslinked with one-pot thiol-ene addition and thiol oxidation at different ratios showed tunable degradation in response to the presence of reducing agents. In addition, changing the ratio between the two reactions did not alter the mechanical properties of the hydrogels. Release of the encapsulated dextran demonstrate the potential of our crosslinking approach for making redox-responsive drug delivery vehicles. Good cytocompatibility of the crosslinking approach was shown with encapsulated mouse fibroblasts. This result suggests the potential in redox-degradable hydrogels for tissue engineering scaffolds.

Next, we developed two ELP-based adhesives using thiol or DOPA chemistries and compared their adhesion performances in a wet setting. The adhesion strength of the two ELP-base adhesives were compared at different curing times with or without the presence of ferric EDTA which functions as a mild oxidant for both adhesives. In general, the thiol-based ELP adhesive had higher adhesion strength with shorter curing times than the DOPA-based one. Higher adhesion strength was achieved when the two adhesives were mixed at a 1-1.5 mass ratio with a constant total mass concentration. This synergistic increase is hypothesized to be a result of the formation of thiol-DOPA adducts. The DOPA-based ELP adhesive had higher tolerance to disturbances during the cure process than the thiol-based one. The cytocompatibility of these ELP-based adhesives was evaluated by direct culture of mouse fibroblasts on the adhesive hydrogels. Both thiol- and DOPA-based adhesives showed high cell viability with the DOPA-based one had lower initial cell attachment. This low initial cell attachment was mitigated when the thiol-based adhesive was blended with the DOPA-based one.

In the last part, we investigated the effect of electrostatic interaction on the LCST behavior of ELPs. Taking the advantage of modular design of recombinant proteins, two charged ELPs were constructed with higher sequence similarities but with different charge states in response to change in pH. Both ELPs had the same positively charged ELP domain but with different external sequences. We showed that additional pH sensitivity in the LCST behavior can be introduced via the external sequence with a pH-dependent charge state. When both the ELP domain and the external sequence were positively charged, higher LCSTs of the fusion ELP protein were observed. On the other hand, when the ELP domain and the external sequence were oppositely charged, lower LCSTs were measured. Such effect was not observed when the external sequence had a pH-independent charge state. We also showed that this effect from the external sequence could be reduced by increasing the ionic strength in the solution or by increasing the length of the ELP domain. We also showed that the identity of the external sequence affects the interaction between the ELP fusion proteins and Hoffmeister ions and therefore change the LCST behavior of the proteins in response to the presence of different Hoffmeister ions. The results provided insights into designing ELPs with pH-sensitive LCST behaviors for pH-triggered drug delivery vehicles.

5.2 Future Directions

This dissertation presented the developments of new approaches to design and process functional biomaterials. The cytocompatibility of our one-pot thiol-ene addition and thiol oxidation crosslinking was demonstrated using thiol-functionalized PEG. This crosslinking method can also be used with other thiol-containing materials, including recombinant proteins with cysteine residues. The modular design of recombinant proteins also allows incorporation of bioactive peptide sequences such as the BMP-2 peptide or the QK peptide. Our lab had previously shown that both peptides can promote osteogenesis and angiogenesis of MSCs in 2D culture systems, respectively.[1-2] These two bioactive peptides do not have cysteine residues in the sequences and will not interfere the thiol-based crosslinking. Therefore, the new thiol-crosslinking can be used to make hydrogels presenting these bioactive peptides to investigate their efficacies in 3D. Because cysteine is the least abundant amino acids in human proteins,[3] there are potentials in incorporation

of other growth factors or bioactive peptides derived from growth factors into thiol-crosslinked hydrogels.

Next, we developed two ELP-based adhesives using thiol or DOPA chemistries. We focused on the effects on adhesion strength from extrinsic factor such as cure time and the presence of oxidants, but there are other extrinsic factors that have been shown to also play important roles in the adhesion performance of ELP-based adhesives.[4-5] Future studies on protein concentration, pH, cure temperature, and other oxidants and crosslinkers are expected to be insightful for understanding the differences in the adhesion mechanism and performance between thiol-based and DOPA-based adhesion. Optimizing the synergistic effect from combining thiol and DOPA chemistries to achieve better adhesion performance is another potential direction. In the present work, such combination was achieved by simply mixing the two proteins. With the aid of modular design, it is also possible to design the protein such that both thiol and DOPA groups are presented on the same protein chain. The modular design also allows the incorporation of bioactive peptides such as the RGD peptide or the BMP-2 peptide, which could provide addition functionalities to these protein-based materials other than tissue adhesives. *In vivo* evaluation of these protein-based adhesives on the adhesion performance and biocompatibility could also be the next step of the current work.

Finally, we studied the role of electrostatic interaction in the LCST behavior of positively charged ELPs. We demonstrated that increasing repulsive charge interaction within the protein chain resulted in elevated LCST and vice versa. Future studies with negatively charged ELPs, non-charged ELPs, and zwitterionically charged ELPs can prove the versatility and robustness of using the external sequences to tune the ELP LCST behavior. Previous studies have shown that the LCST behavior of an ELP can be changed when being fused to other proteins.[6-9] The extent of a such change depends on the hydrophobicity and surfaces charges of the fusion protein. However, the exact interactions between the ELP and the fusion protein were not clear due to the complexity from the fusion protein structures. The studies of using charged short external sequences with different ELPs will also help with better understanding the importance of electrostatic interaction in the LCST behavior of ELPs.

5.3 References

(1) Kim, Y.; Renner, J. N.; Liu, J. C. Incorporating the BMP-2 peptide in genetically-engineered biomaterials accelerates osteogenic differentiation. *Biomater. Sci.* **2014,** *2* (8), 1110-1119.

(2) Kim, Y.; Liu, J. C. Protein-engineered microenvironments can promote endothelial differentiation of human mesenchymal stem cells in the absence of exogenous growth factors. *Biomater. Sci.* **2016,** *4* (12), 1761-1772.

(3) Marino, S. M.; Gladyshev, V. N. Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *J. Mol. Biol.* **2010,** *404* (5), 902-916.

(4) Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

(5) Brennan, M. J.; Hollingshead, S. E.; Wilker, J. J.; Liu, J. C. Critical factors for the bulk adhesion of engineered elastomeric proteins. *R. Soc. Open Sci.* **2018,** *5* (5).

(6) Christensen, T.; Amiram, M.; Dagher, S.; Trabbic-Carlson, K.; Shamji, M. F.; Setton, L. A.; Chilkoti, A. Fusion order controls expression level and activity of elastin-like polypeptide fusion proteins. *Protein Sci.* **2009,** *18* (7), 1377-1387.

(7) Trabbic- Carlson, K.; Meyer, D.; Liu, L.; Piervincenzi, R.; Nath, N.; LaBean, T.; Chilkoti, A. Effect of protein fusion on the transition temperature of an environmentally responsive elastin- like polypeptide: a role for surface hydrophobicity? *Protein Eng. Des. Sel.* **2004,** *17* (1), 57-66.

(8) Qin, G.; Glassman, M. J.; Lam, C. N.; Chang, D.; Schaible, E.; Hexemer, A.; Olsen, B. D. Topological Effects on Globular Protein-ELP Fusion Block Copolymer Self-Assembly. *Adv. Funct. Mater.* **2015,** *25* (5), 729-738.

(9) Qin, G.; Perez, P. M.; Mills, C. E.; Olsen, B. D. Effect of ELP Sequence and Fusion Protein Design on Concentrated Solution Self-Assembly. *Biomacromolecules* **2016,** *17* (3), 928-934.

APPENDIX A. TUNABLE LIGAND PRESENTATION TO GUIDE STEM CELL DIFFERENTIATION

PROJECT BACKGROUND

The goals of this project were to: 1) Develop recombinant protein-based materials with tunable ligand presentations, 2) Study the effect of ligand avidity on stem cell differentiation, 3) Investigate the effect of ligand affinity to the backbone materials on stem cell differentiation, and 4) Explore the synergistic effect of ligand presentations and scaffold mechanical properties on guiding stem cell differentiation.

Specifically, we utilized leucine zipper recombinant protein domains to achieve tunable presentations on ligand avidity and affinity. Leucine zippers are coiled coil domains where two or more helical peptides wrap together in a superhelical fashion. The association between leucine zipper peptides can be modulated by changing the length of the peptides, the hydrophobicity of the peptide sequences, and the electrostatic interaction between the peptides.[1] Different oligomerization states of leucine zippers, including dimers, trimers and pentamers, have been reported. In this project, we used one high affinity heterodimeric leucine zipper (HZE/HZR), one low affinity heterodimeric leucine zipper (LZE/LZR), and one high affinity homopentameric leucine zipper (COMPm) to achieve tunable ligand presentation.

We chose the BMP-2 peptide as the model ligand to investigate the effect of ligand presentation on stem cell differentiation. The BMP-2 peptide is derived from the knuckle epitope of human BMP-2.[2] It has been shown that BMP-2 functions mostly as a homodimer.[3] The signaling pathway of the dimeric BMP-2 involves endocytosis of the ligand-receptor complexes. Since the BMP-2 peptide did not have the dimerization domain found in the whole BMP-2 protein, it is expected that dimerization of the peptide will not occur spontaneously and thus decrease its efficacy in guiding stem cell differentiation. We hypothesized that by combining the BMP-2 peptide to oligomeric leucine zippers, we could achieve oligomeric presentation of the BMP-2 peptides and thus mimic the dimerization of native BMP-2. The pentameric COMPm was chosen to explore the potential of higher avidity in increasing the peptide efficacy. The two different dimeric leucine zippers (HZE/HZR and LZE/LZR) were designed to mimic different extents of endocytosis of the

ligand-receptor complexes. The low affinity pair LZE/LZR would allow dissociation of the BMP-2 peptide from the material and allow endocytosis of the BMP-2 peptide. On other hand, the high affinity pair HZE/HZR would hinder the endocytosis process.

To create a hydrogel system that harbors the aforementioned design of leucine zippers with the BMP-2 peptide and allows tunable mechanical properties, an elastomeric recombinant domain, resilin-like polypeptide (RLP), was incorporated into the material design. Our group previously developed a Lys-containing variant (RZY10) from the RLP derived from *Anopheles gambiae* and demonstrated that the Lys-crosslinked RLP hydrogels showed tunable mechanical properties by changing both protein concentration and stoichiometry in the crosslinking chemistry.[4]

RESULTS AND DISCUSSION

Recombinant proteins with all three different leucine zippers were designed and cloned, and seven proteins were expressed and purified using Ni-NTA columns. The proteins were named: RZY10-LZR, RZY10-HZR, RZY10-COMPm, HZE-RGDGC, COMPm-RGDGC, LZE-BMP2GC, and HZE-QEQV-RGDGC. Here the cell binding sequence, Arg-Gly-Asp (RGD), was used instead of the BMP-2 peptide sequence for the pilot cell studies. QEQV was a short sequence of an ELP domain with the guest residues of Gln, Glu, and Val to improve the solubility of leucine zipper-containing proteins. We found that most of our zipper proteins suffered from very low solubility in PBS (~5 mg/mL for RZY10-LZR and RZY10-HZR and <2 mg/mL for HZE-RGDGC, COMPm-RGDGC, and LZE-BMP2GC). Similar issues of poor solubility have been reported with other leucine zipper-containing proteins.[5-6] The construction of HZE-QEQV-RGDGC was an attempt to rescue the low solubility by incorporating a negatively charged hydrophilic ELP domain into the protein. When dissolved in PBS, HZE-QEQV-RGDGC was soluble at 30 mg/mL at 50 °C (**Error! Reference source not found.**). On the other hand, HZE-RGDGC remained completely insoluble at 30 mg/mL at 50 °C. However, both proteins were insoluble at 37 °C in PBS at 30 mg/mL.

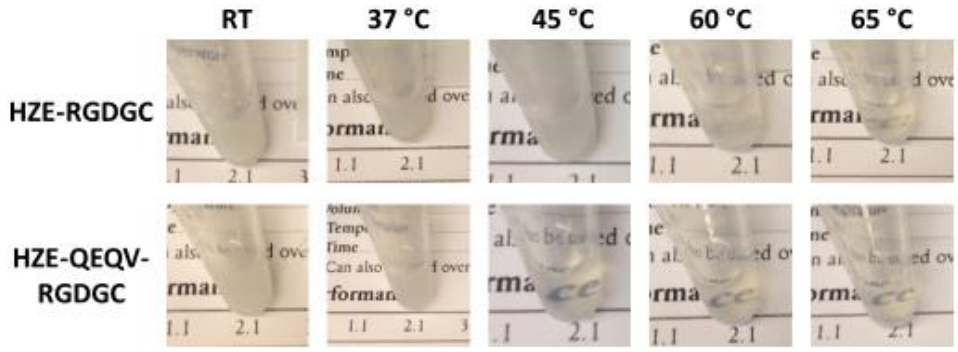


Figure A. 1. Solubility of HZE-QEQV-RGDGC (top) and HZE-RGDGC (bottom) changes with temperature. Proteins were dissolved in PBS at 30 mg/mL and heated to temperature indicated in the column headings.

Despite the low solubility issue, we next verified if our leucine zipper proteins could form desired oligomers in PBS. RZY10-HZR and HZE-RGDGC were prepared at 1 mg/mL in PBS and mixed at a 1:1 volumetric ratio. The mixture was incubated at room temperature overnight to allow association between the two proteins to occur. After the incubation, the protein complexes were crosslinked via a carboxylic-hydrazide reaction (ADH) and analyzed by SDS-PAGE. As shown in Figure A. 2, RZY10-HZR alone without crosslinking appeared as a single band around its expected MW (20.2 kDa); however, when crosslinked, there was a faint high band around 46 kDa. This high band was around the expected MW of a homodimer of RZY10-HZR (40.4 kDa). We also noticed that the band corresponding to monomer of RZY10-HZR ran higher than that in the non-crosslinked sample. It was possible that the ADH crosslinker reacted with the monomer RZY10-HZR without crosslinking to another protein chain and thus changed the electrophoretic mobility of the monomeric protein. Similarly, HZE-RGDGC without crosslinking appeared as a single band around its MW (10.2 kDa). However, in the crosslinked samples, there was a significant decrease in intensity of the HZE-RGDGC monomer band. Furthermore, a smear around 22 kDa was observed. We hypothesized that the smear could be the homodimer of HZE-RGDGC and that the higher K_d (10^{-5} M as reported previously[7]) resulted in a less stable homodimer, which appeared as a smear. Interestingly, homodimer formation of RZY10-HZR seemed to be less than that of HZE-RGDGC as judged by SDS-PAGE even though the K_d of the RZY10-HZR homodimer is expected to be around 10^{-9} M. It is speculated that the long RZY10 chain hindered HZR homodimer formation. Finally, when

a sample of RZY10-HZR/HZE-RGDGC was analyzed without crosslinking, both proteins appeared as two single bands. However, when crosslinked, a distinct band around 32 kDa was observed, and this band was close to the expected MW of an RZY10-HZR/HZE-RGDGC heterodimer (30.4 kDa). Moreover, the smear around 22 kDa was not observed. This observation suggested that the majority of HZE-RGDGC formed heterodimers instead of HZE homodimers. Altogether, we concluded the dimerization of high affinity leucine zipper HZE/HZR was successful. However, both proteins, especially HZE-RGDGC, had low solubility at physiological conditions, and the dimerization efficiency was limited by low solubility.

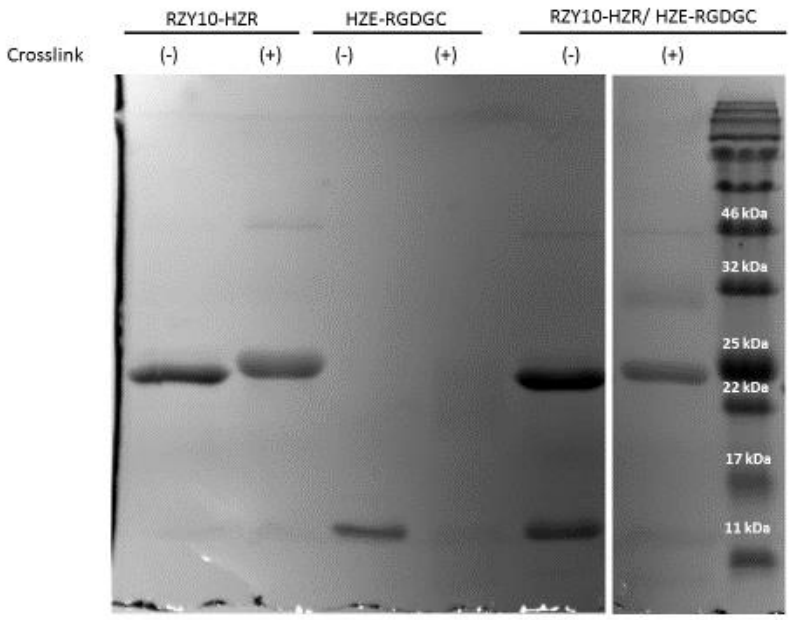


Figure A. 2. Oligomerization of RZY10-HZR and HZE-RGDGC. (-) and (+) indicate sample with and without crosslinking, respectively. MW of proteins: RZY10-HZR: 20.2 kDa, HZE_RGDGC: 10.2 kDa, homodimer of RZY10-HZR: 40.4 kDa, homodimer of HZE-RGDGC: 20.4 kDa, heterodimer of RZY10-HZR and HZE-RGDGC: 30.4 kDa.

The second leucine zipper system tested was the pentameric COMPm system. Two proteins, RZY10-COMPm (20.2 kDa) and COMPm-RGDGC (10.2 kDa) were expressed to test if RZY10-COMPm and COMPm-RGDGC formed heteropentamers. Both proteins were dissolved in 8 M urea to destroy any preformed homopentamers. The urea solutions were then mixed and incubated at room temperature overnight before being subjected to a step-down in urea concentration. Briefly, urea concentration in the dialysis buffer was

lowered gradually toward zero. It was observed that precipitation occurred for both COMPm-RGDGC alone and RZY10-COMPm/COMPm-RGDGC when they were dialyzed against dialysis buffer (phosphate buffer) without urea. In contrast, RZY10-COMPm remained clear throughout the dialysis. This observation was consistent with the solubility of these proteins in PBS since RZY10-COMPm had higher solubility than COMPm-RGDGC. The precipitates from the mixture were analyzed by SDS-PAGE and were found to be composed of primarily COMPm-RGDGC.

Supernatants of each group after dialysis were crosslinked by ADH and were analyzed by SDS-PAGE, anti-T7 tag and anti-His tag Western blots. In the protein design, RZY10-COMPm has N-terminal T7 and 6xHis tags, whereas COMPm-RGDGC only has an N-terminal 6xHis tag. Therefore, we used anti-T7 tag antibodies in a Western blot to identify proteins bands having RZY10-COMPm. As shown in Figure A. 3, non-crosslinked RZY10-COMPm and RZY10-COMPm/COMPm-RGD samples had a 25 kDa band which corresponded to monomeric RZY10-COMPm. Faint bands below 25 kDa could be degradation or truncation products of RZY10-COMPm. COMPm-RGDGC did not appear on anti-T7 tag Western blot, as expected. When examining crosslinked samples, we found that the sample containing only RZY10-COMPm had higher bands with MW corresponding to dimers, trimers, tetramers, and pentamers of RZY10-COMPm; however, signals from these oligomers were much fainter compared to the monomeric RZY10-COMPm band. This result suggested that the oligomerization efficiency was low when RZY10-COMPm complexed with itself. When the mixture RZY10-COMPm and COMPm-RGDGC was crosslinked, signals from the oligomers became stronger, especially the bands at 46 kDa and 135 kDa. Although these bands were at similar positions as the homodimer and the homopentamer in the RZY10-COMPm sample, we could not conclude if they were homo-oligomers of RZY10-COMPm or hetero-oligomers with COMPm-RGDGC. Interestingly, a band at 30 kDa that was not observed in RZY10-COMPm sample was observed in the mixture sample. Therefore, it was speculated that this 30-kDa band could be a heterodimer of RZY10-COMPm (20.2 kDa) and COMPm-RGDGC (10.2 kDa). Surprisingly, anti-His tag Western blot failed to detect any crosslinked proteins (Supporting Information Figure A. 12). We speculated that crosslinking of proteins resulted in a structural hindrance of access to the 6xHis tag.

Altogether, we concluded that oligomerization of the COMPm system was successful; however, the efficiency was low, and the process needs to be optimized to achieve the desired hetero-oligomeric product. Also, the current design of COMPm-RGDGC would benefit by having an exclusive epitope tag for Western blot detection. Last, both RZY10-COMPm and COMPm-RGDGC had low solubility at physiological conditions.

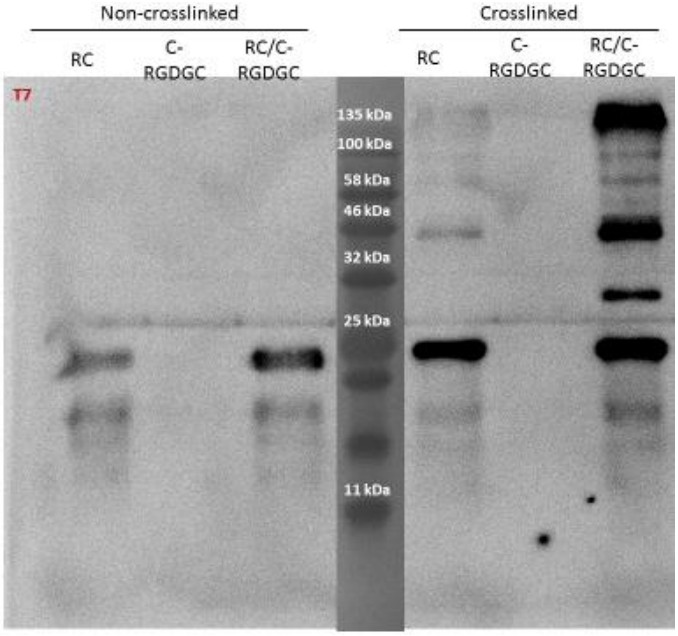


Figure A. 3. Oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC) verified by luminescent anti-T7 tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa.

We observed that recombinant proteins with leucine zippers had low solubility in aqueous solutions. Thus, we decided to verify whether our low affinity leucine zipper system LZE/LZR has low solubility as well. Lyophilized LZE-BMP2GC was dissolved in PBS; however, it did not completely dissolve at 10 mg/mL. Since we expected the working concentration of these leucine zipper-based proteins to be around 30-50 mg/mL, we concluded that our current proteins (RZY10-HZR, HZE-RGDGC, RZY10-COMPm, COMPm-RGDGC, LZE-RGDGC, and HZE-QEQV-RGDGC) would not be feasible for the proposed applications.

Supporting Information

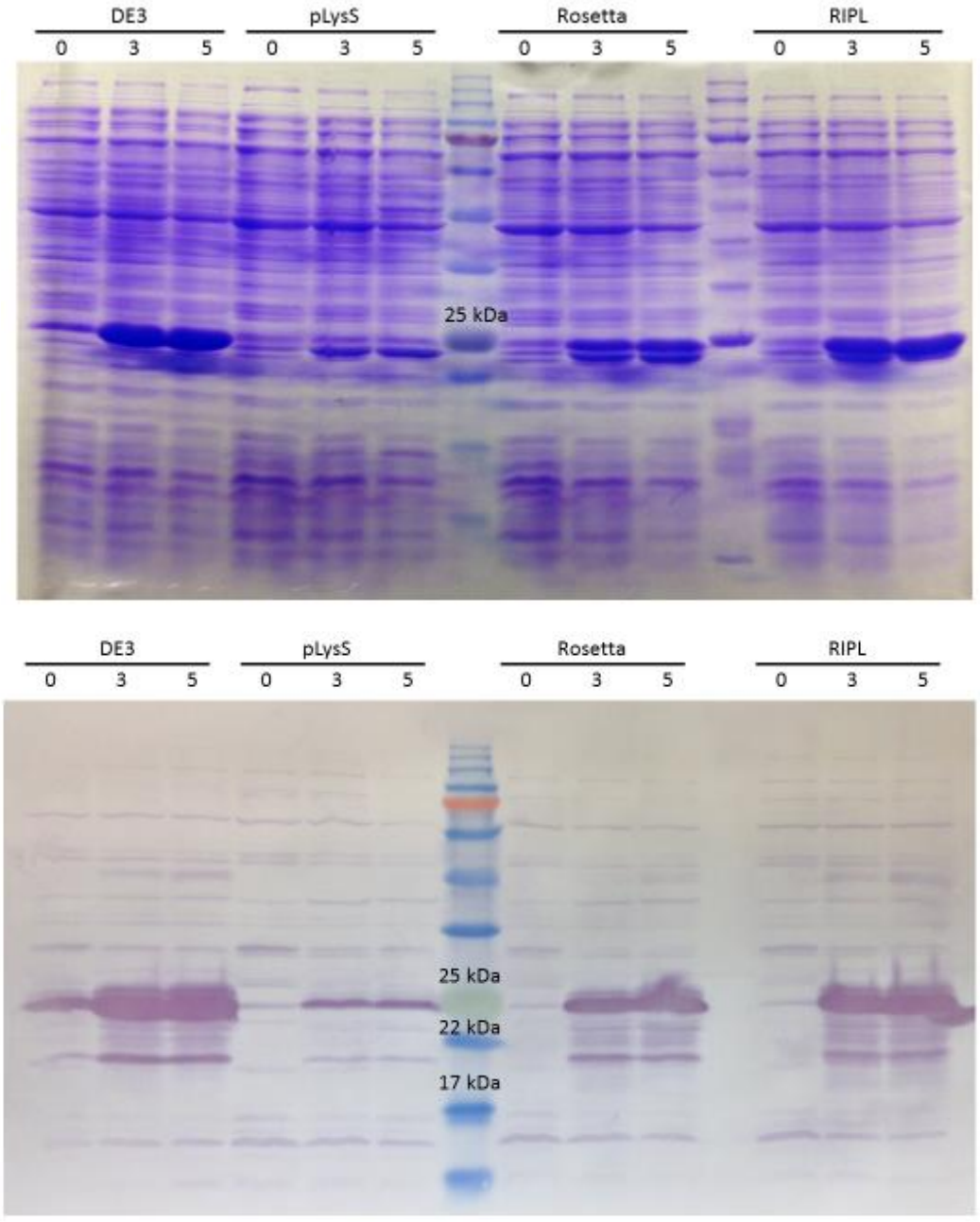


Figure A. 4. Expression of RZY10-HZR (MW 20.2 kDa). Top: SDS-PAGE gel showing RZY10-HZR being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-T7 tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel.

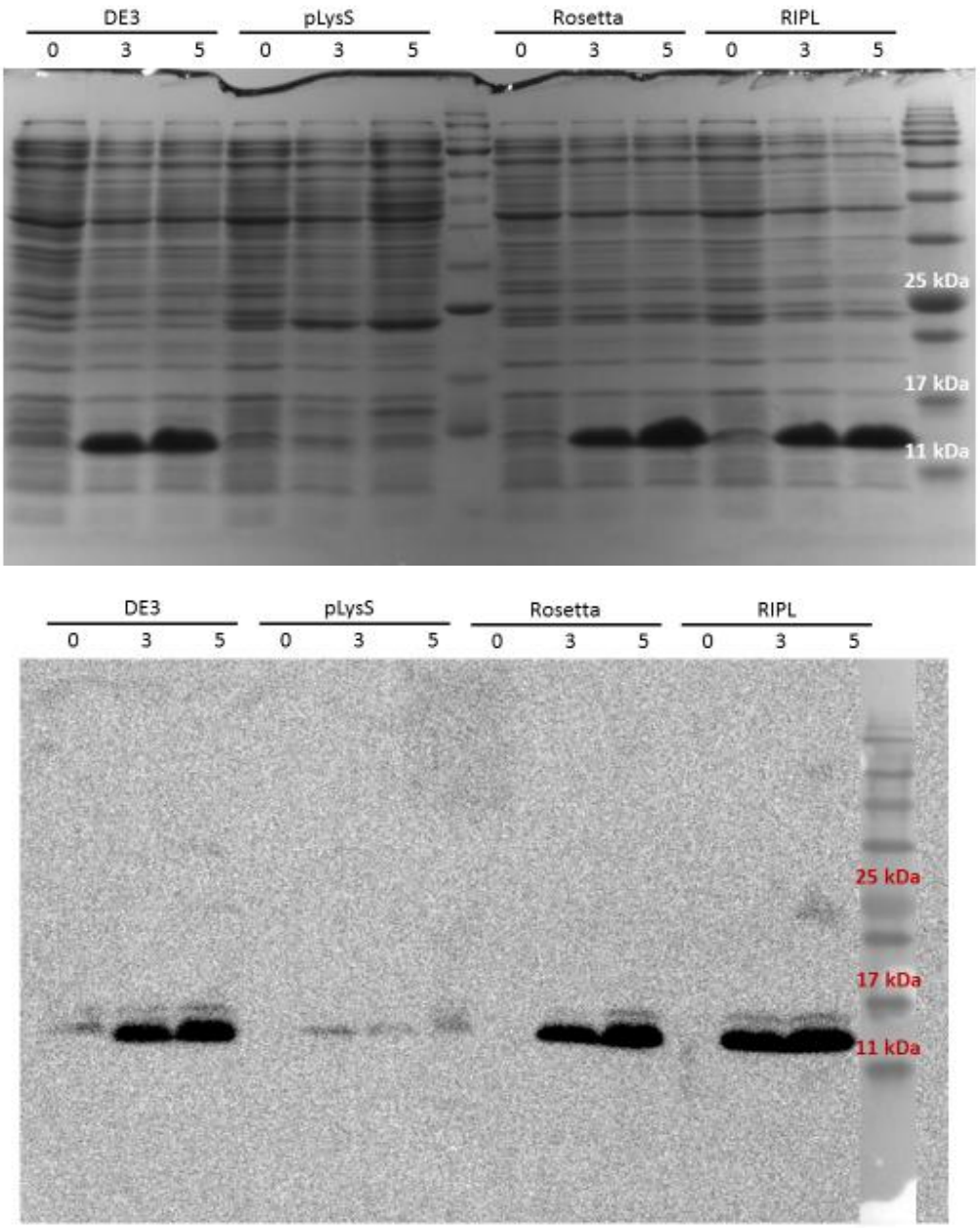


Figure A. 5. Expression of HZE-RGDGC (MW 10.2 kDa). Top: SDS-PAGE gel showing HZE-RGDGC being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by luminescent anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel.

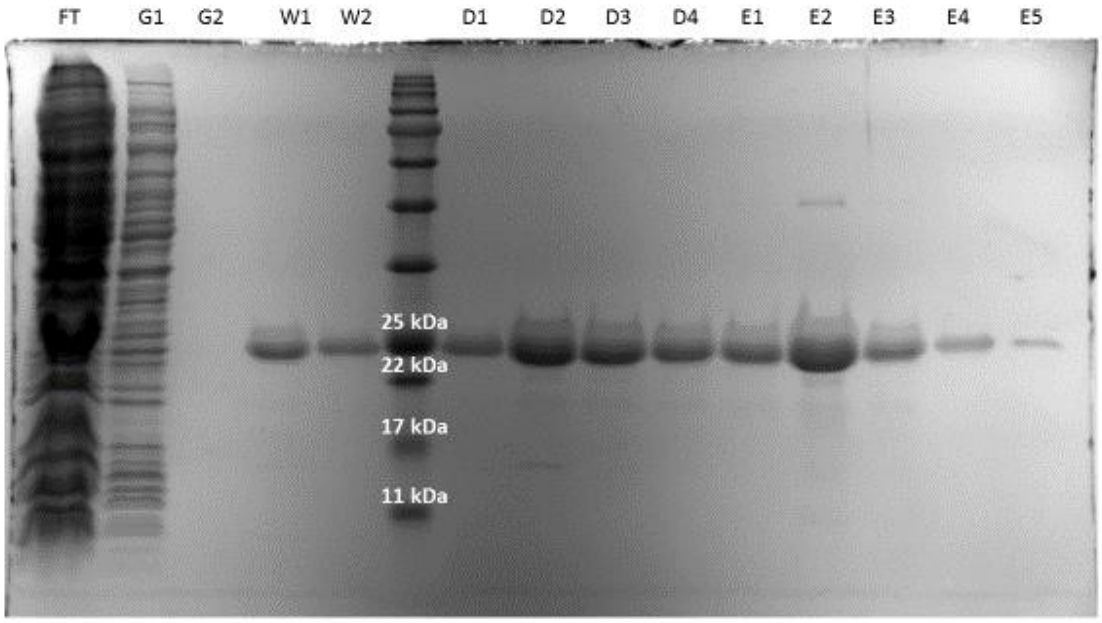


Figure A. 6. SDS-PAGE gel of purification of RZY10-HZR in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~30 mg/L culture. FT: flow though; G1 and G2: wash with 6 M guanidine hydrochloride; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

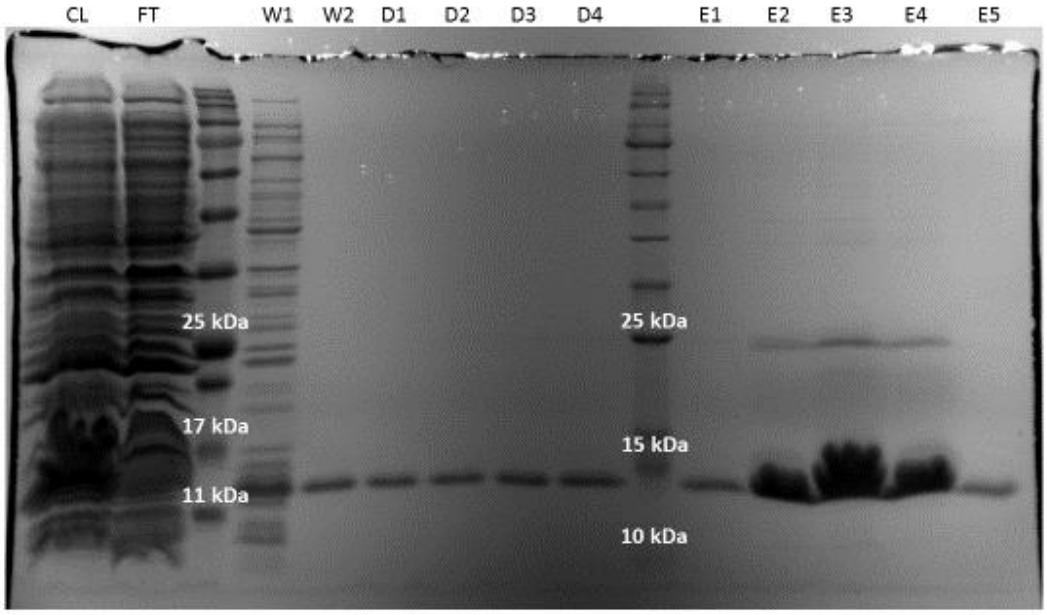


Figure A. 7. SDS-PAGE gel of purification of HZE-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. CL: cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

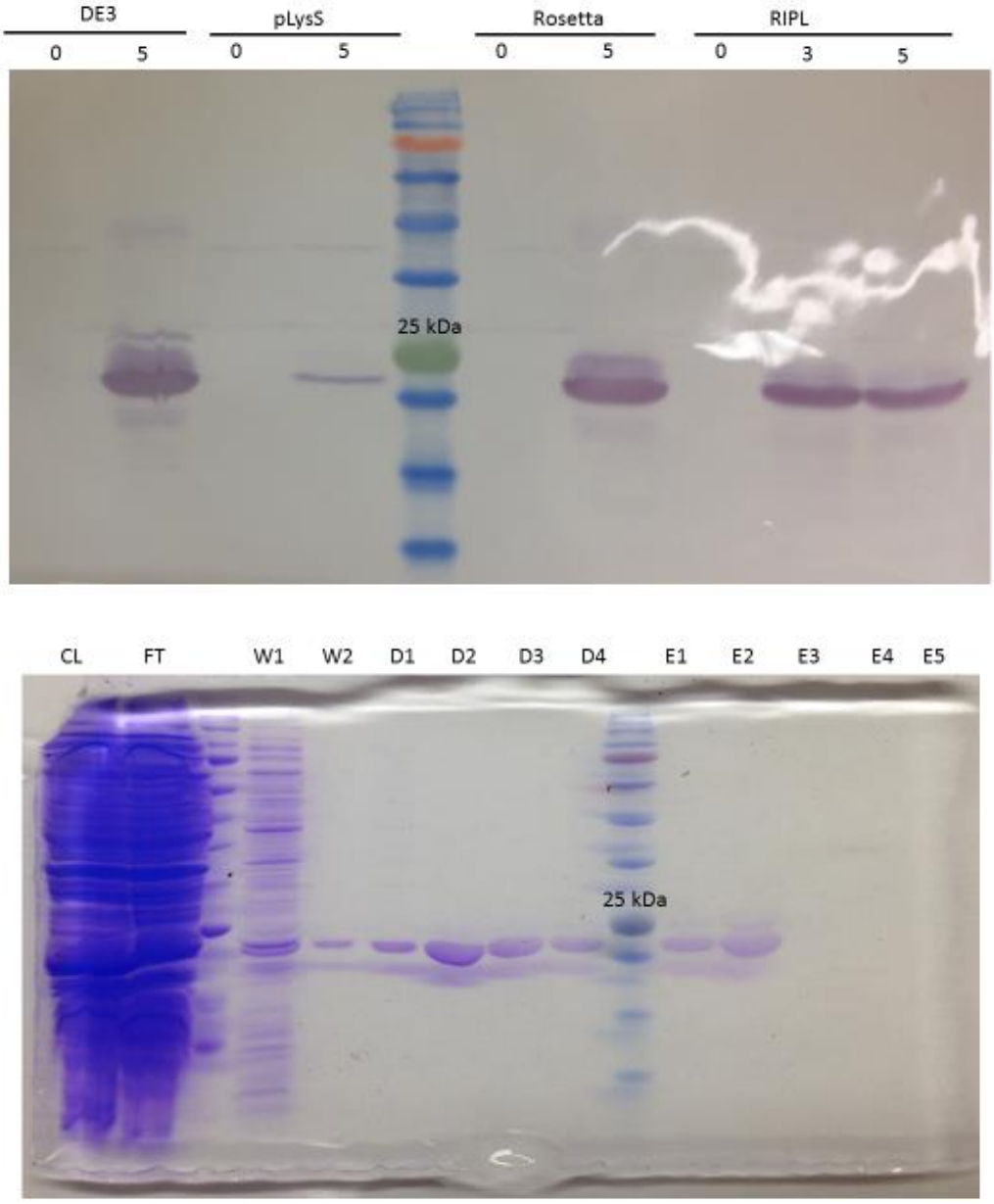


Figure A. 8. Expression of RZY10-COMPm (MW 20.2 kDa). Top: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-T7 tag Western blot. Bottom: SDS-PAGE gel of purification of RZY10-COMPm in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~30 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

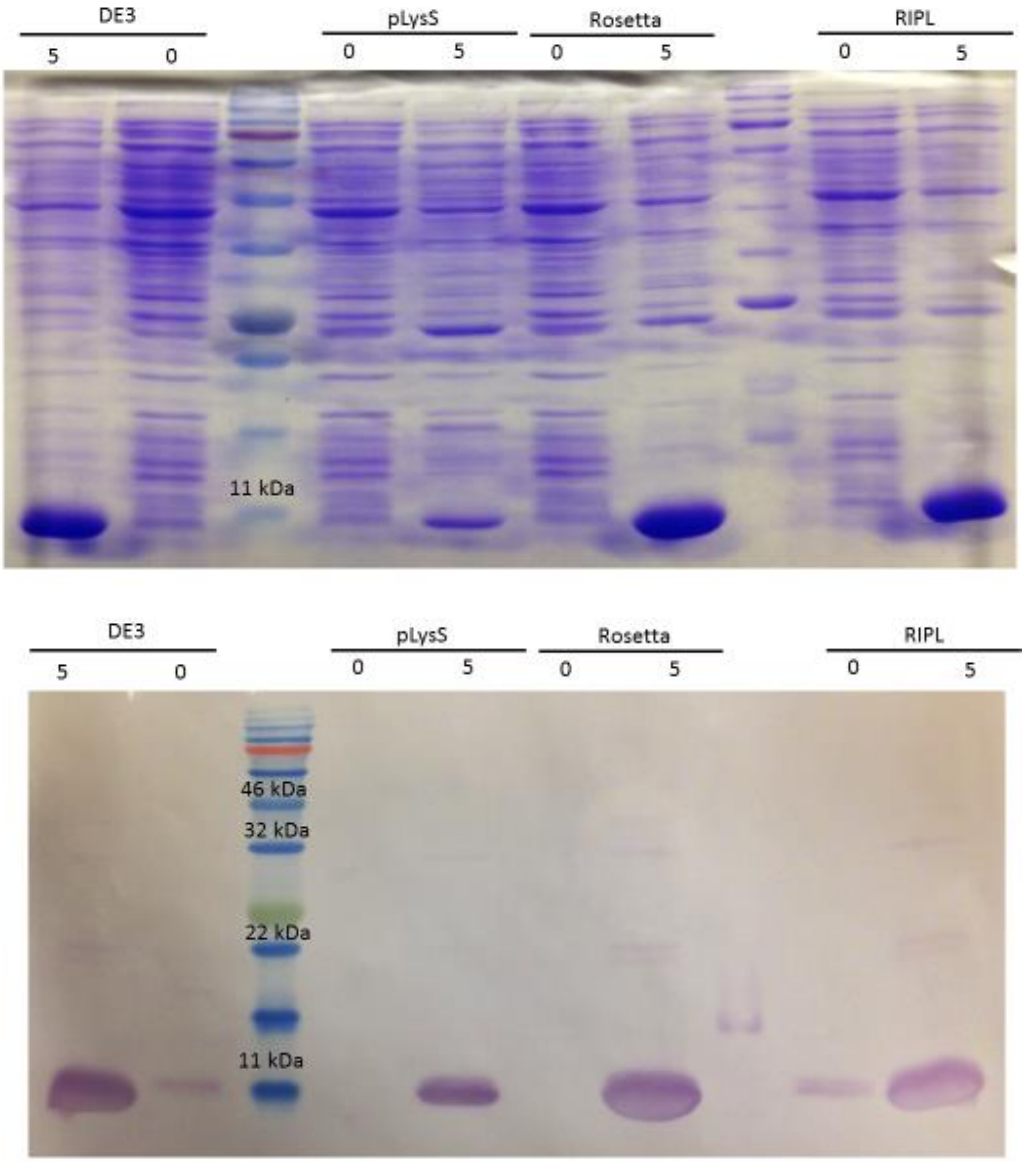


Figure A. 9. Expression of COMPm-RGDGC (MW 10.2 kDa). Top: SDS-PAGE gel showing COMPm-RGDGC being expressed for 3 and 5 hours after induction of IPTG. Bottom: The expression was confirmed in four expression *E. coli* hosts by colorimetric anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel.

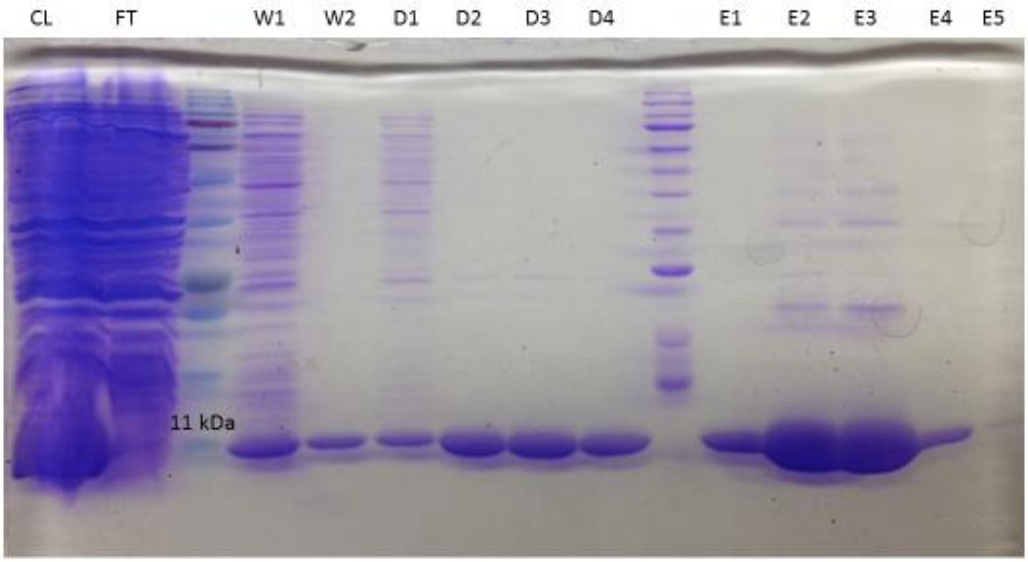


Figure A. 10. SDS-PAGE gel of purification of COMPm-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

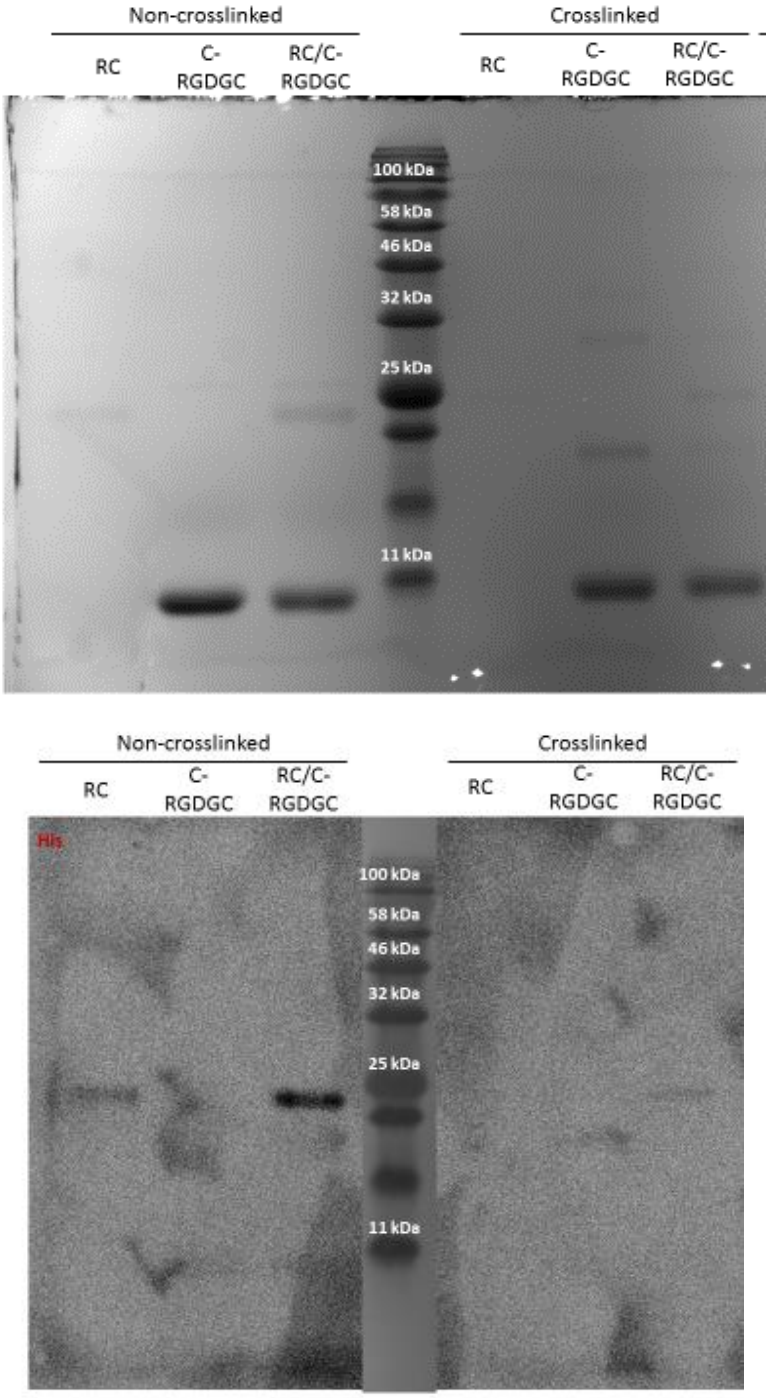


Figure A. 11. Oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC). Top: Coomassie Blue stained SS-PAGE gel. Bottom: Luminescent anti-His tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa.

Antibody	RZY10-COMP	COMP-RGDGC	RZY10-COMP/ COMP-RGDGC
Anti-T7 tag	+	-	+
Anti-His tag	+	+	+

Figure A. 12. Visualization of RZY10-COMPm and COMPm-RGDGC with different antibodies in the Western blot to show oligomerization of RZY10-COMPm (RC) and COMPm-RGDGC (C-RGDGC). Luminescent anti-His tag Western blot. MW of RZY10-COMPm: 20.2 kDa. MW of COMPm-RGDGC: 10.2 kDa.

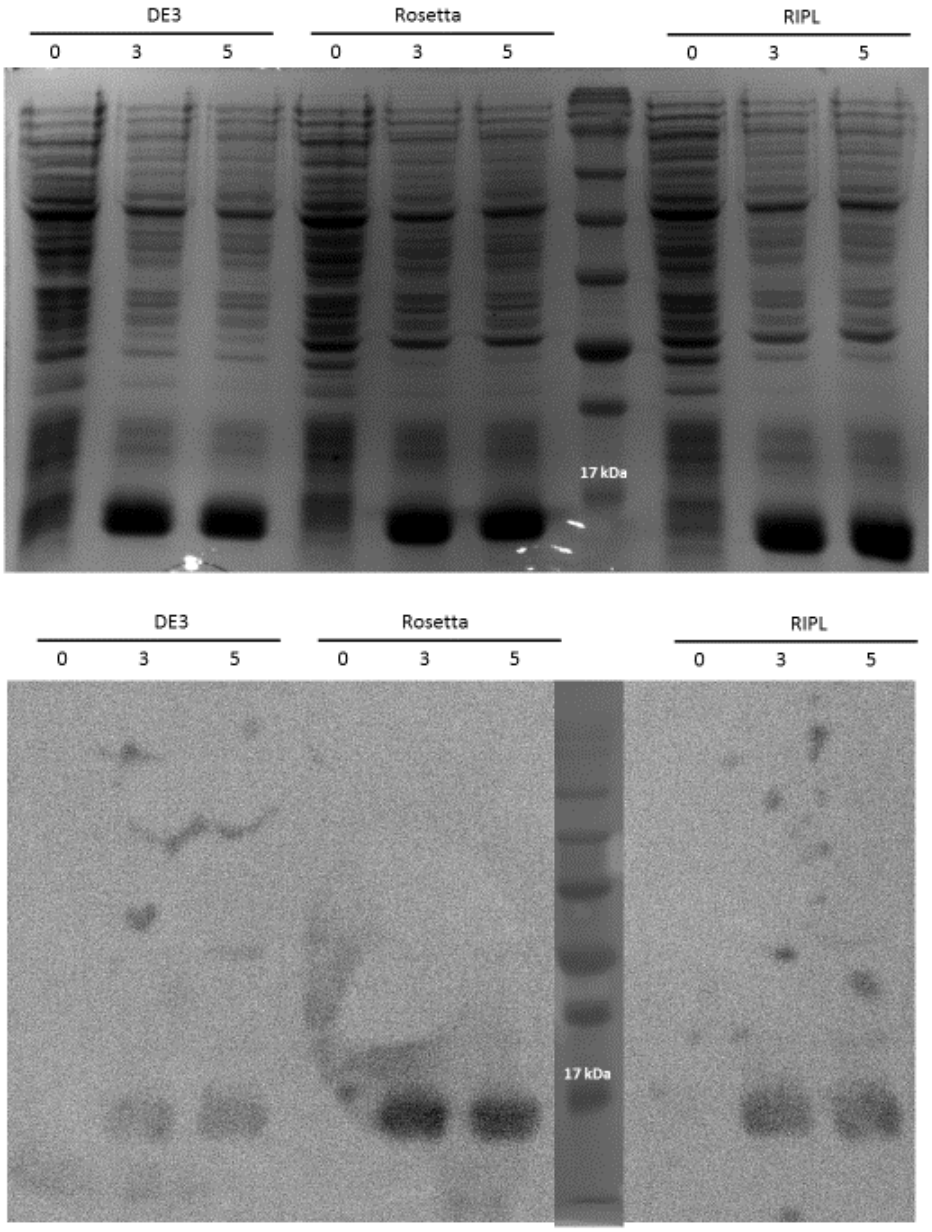


Figure A. 13. Expression of LZE-BMP2GC (MW 10.2 kDa). Top: SDS-PAGE gel showing LZ-BMP2GC being expressed for 3 and 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts by colorimetric anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel.

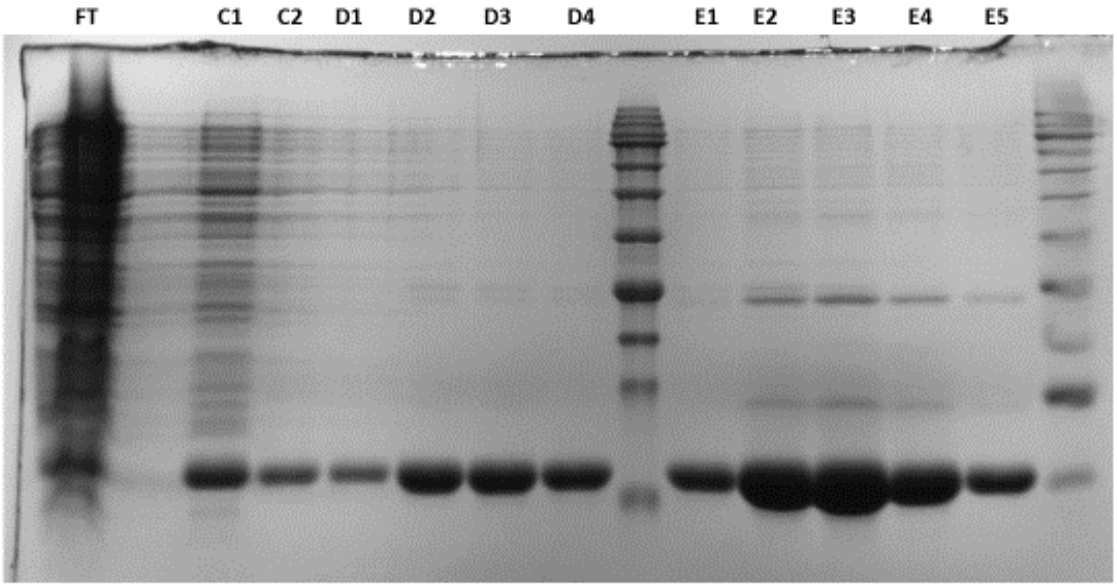


Figure A. 14. SDS-PAGE gel of purification of LZE-BMP2GC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was ~50 mg/L culture. FT: flow though; C1 and C2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

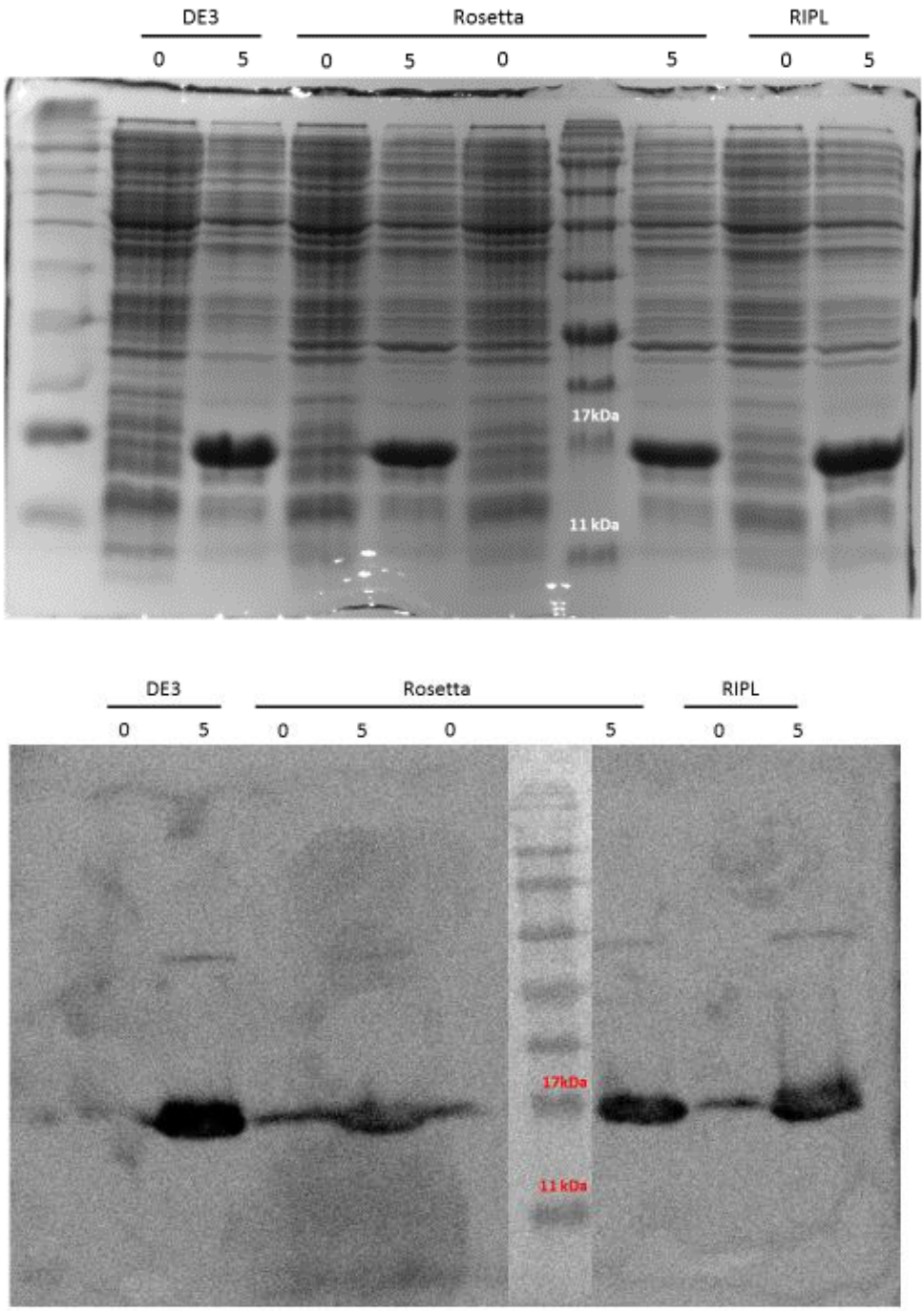


Figure A. 15. Expression of HZE-QEQV-RGDGC (MW 11.6 kDa). Top: SDS-PAGE gel showing HZE-QEQV-RGDGC being expressed for 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts by anti-His tag Western blot. The expression host and the corresponding expression time (h) were specified on top of each gel.

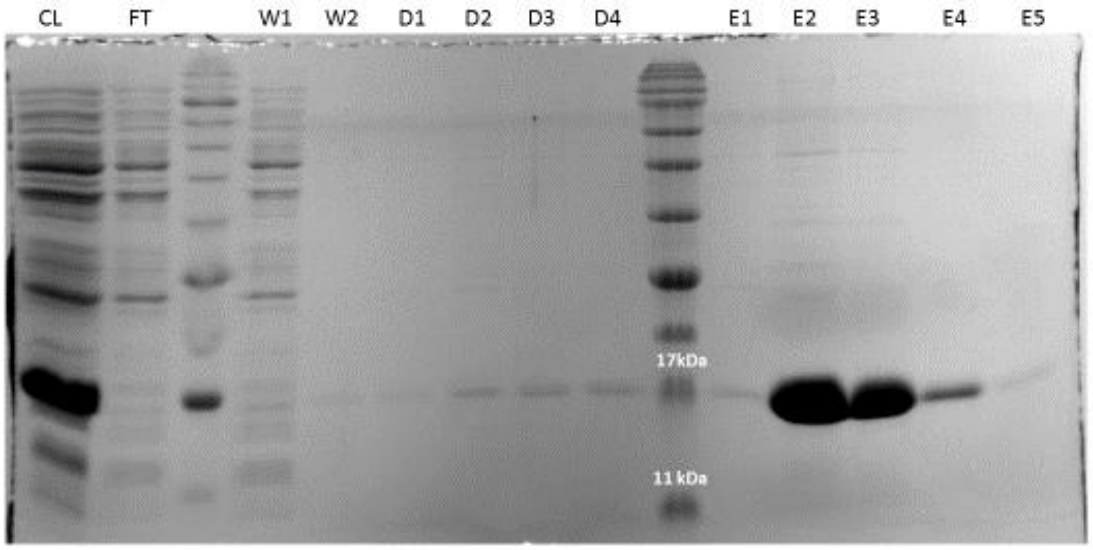


Figure A. 16. SDS-PAGE gel of purification of HZE-QEQV-RGDGC in Rosetta by denatured Ni-NTA affinity chromatography. The final yield was about 50 mg/L culture. CL: Cell lysate; FT: flow though; W1 and W2: wash with 8 M urea, pH 6.3; D1 to D4: elutes with 8 M urea, pH 5.9; E1 to E5: elutes with 8 M urea, pH 4.9.

REFERENCES

(1) Apostolovic, B.; Danial, M.; Klok, H.-A. Coiled coils: attractive protein folding motifs for the fabrication of self-assembled, responsive and bioactive materials. *Chem. Soc. Rev.* **2010,** *39* (9), 3541-3575.

(2) Saito, A.; Suzuki, Y.; Ogata, S.-i.; Ohtsuki, C.; Tanihara, M. Activation of osteo-progenitor cells by a novel synthetic peptide derived from the bone morphogenetic protein-2 knuckle epitope. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2003,** *1651* (1), 60-67.

(3) Bessa, P. C.; Casal, M.; Reis, R. Bone morphogenetic proteins in tissue engineering: the road from the laboratory to the clinic, part I (basic concepts). *J. Tissue Eng. Regen. Med.* **2008,** *2* (1), 1-13.

(4) Su, R. S. C.; Gill, E. E.; Kim, Y.; Liu, J. C. Characterization of resilin-like proteins with tunable mechanical properties. *J. Mech. Behav. Biomed. Mater.* **2019,** *91*, 68-75.

(5) Zhang, K.; Diehl, M. R.; Tirrell, D. A. Artificial polypeptide scaffold for protein immobilization. *J. Am. Chem. Soc.* **2005,** *127* (29), 10136-10137.

(6) Zhang, K.; Sugawara, A.; Tirrell, D. A. Generation of Surface- Bound Multicomponent Protein Gradients. *Chembiochem* **2009,** *10* (16), 2617-2619.

(7) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. Designed heterodimerizing leucine zippers with a ranger of pIs and stabilities up to 10− 15 M. *Protein Sci.* **2001,** *10* (3), 649-655.

APPENDIX B. NUCLEOPHILE-CATECHOL PROTEIN CONJUGATION AND CORSSLINKING

PROJECT BACKGROUND

The goal of this project was to demonstrate protein-protein conjugation through nucleophile-catechol Michael addition and to determine the condition for reactions to occur that favored histidine over lysine.

When choosing the chemistry for protein conjugation or crosslinking, it is desired to use chemistry that does not involve functional groups that are found in the protein of interest in order to minimize potential changes in structure or in bioactivity of the conjugated product. Common protein conjugation or crosslinking chemistries use amines, carboxylic acids, or thiols. However, as shown in Table B. 1, amines and carboxylic acids are commonly found in bioactive proteins and peptides due to the abundance of Lys, Glu, and Asp residues (as well as N-terminal amine and C-terminal carboxylic acid groups). Thiol groups found in Cys have a much lower frequency but are prone to be oxidized and lose reactivity. As a result, thiol conjugation usually requires reducing agents such as 2-mercaptoethanol or dithiothreitol to break existing disulfide bonds immediately prior to conjugation. Moreover, including cysteine residues in small proteins or peptides increases the possibility for proteins or peptides to precipitate because of self-aggregation through disulfide bond formation. Therefore, it is of great interest to develop a new approach for protein conjugation through functional groups other than amines, carboxyls, or thiols.

Catechols, or 1,2-dihydroxybenzene, have been explored for bioorthogonal crosslinking because of their high reactivity towards nucleophiles. Catechol groups can be introduced into protein-based materials through the incorporation of an non-canonical amino acid, 3,4-dihydroxyphenylalanine (DOPA). In nature, DOPA can be produced by converting tyrosine with tyrosinase. Many organisms utilize DOPA-based crosslinking. One example is the wet-setting adhesive proteins produced by murine mussels.[1] Another example is the sclerotization of insect cuticles from catechol-histidine crosslinking.[2]

Table B. 1. Amino acid sequences of common bioactive peptides.

	Sequence	#of Lys	# of Asp, Glu	# of Cys	# of His	Ref
RGD	GRGDSPASSK	1	1	0	0	[3]
BMP-2	KIPKASSVPTELSAISTLYL	2	1	0	0	[4]
BMP-7	SNVILKKYRN	2	0	0	0	[5]
	KPSSAPTQLN	1	0	0	0	
	KAISVLYFDDS	1	2	0	0	
QK	KLTWQELYQLKYKGI	3	1	0	0	[6]
Laminin peptide	IKVAV	1	0	0	0	[7]
IGF-1 peptide	GRVDWLQRNANFYDWFVAELG	0	3	0	0	[8]
TGF-beta 1	ANVAENA	0	1	0	0	[9]
TGF-beta 2	LIAEAK	1	1	0	0	

The Kodadek group showed that in the presence of periodate as an oxidant, DOPA/ortho-quinone conjugated with N-terminal amines, the ε-amine of lysine, thiols in cysteine, and the imidazole group in histidine.[10].. Surface immobilization of proteins on titanium[11] or polymer substrates[12] was achieved by dopamine coating; however, the amino acids being conjugated were not identified in these studies. The Messersmith group demonstrated that choice of pH determined the selectivity for which amino acid would conjugate.[13] A substrate was coated with dopamine and then incubated in the solution of a model peptide containing both histidine and lysine. The immobilization was largely achieved through histidine when pH was between 6.4 and 7.4, a pH range that was close to the pKa of histidine (6.04). On the other hand, lysine (pKa 10.7) appeared to be the major reactive site in the immobilization when the pH ranged from 8.4 to 9.5. The authors suggested that the deprotonation of these nucleophilic side chains was necessary for catechol-nucleophile conjugation to occur, and when both were deprotonated, the primary amine (ε-amino group of lysine) was more reactive than the secondary amine (imidazole of histidine).

We proposed to incorporate catechol-nucleophile Michael addition into the modular design of recombinant proteins to utilize it as a new approach for protein or

peptide conjugation that would add to the traditional methods of amine, carboxyl, or thiol chemistry. Our approach was to insert tyrosine residues into the recombinant protein sequence and then convert these tyrosine residues into DOPA by tyrosinase. Because the recombinant design starts at the DNA level, it is possible to precisely control the location of tyrosine residues in the protein chain and thus to achieve conjugation in a position-specific manner. DOPA residues in one protein would be utilized to conjugate to histidine residues another polypeptide.

RESULTS AND DISCUSSION

As a preliminary test of DOPA-His conjugation, we started with an elastomeric resilin-like polypeptide, RZY10. The amino acid sequence of RZY10 contained five repeats of AQTPSSKQYGAP-AQTPSSQYGAP with N-terminal T7 and 6xHis tags. Tyrosinase-converted RZY10 (mRZY10) was analyzed by SDS-PAGE. The expected MW of mRZY10 was 14.3 kDa, and that of RZY10 was 14.1 kDa. However, RLPs have been reported to have lower electrophoretic mobility in SDS-PAGE gels,[14] and RZY10 was previously shown to run at ~20 kDa in our lab. The mRZY10 band shifted slightly upward as shown **Error! Reference source not found.**, and this finding is consistent with previous observations from our lab that tyrosinase-converted ELP runs higher than the unmodified ELP. This decrease in electrophoretic mobility could be a synergistic effect of increase in molecular weight and decrease in hydrophobicity. The mRZY10 solution was adjusted to pH 7.4 and was incubated at room temperature for 24 hours to allow DOPA to react with histidine residues in the 6xHis tag. The solution was analyzed by SDS-PAGE and anti-T7 tag Western blot immediately after incubation. An RZY10 solution treated with the same conditions was used as a negative control group. As shown in Figure B. 2, RZY10 appeared as one single band while mRZY10 was observed with a smear shifting upward. We hypothesized that the smear was a result of intrachain conjugation on one mRZY10 protein. That is, histidine residues in the 6xHis tag were conjugated to DOPAs on the same protein chain and therefore formed a circular structure. It has been reported that proteins with circular structures could have lower electrophoretic mobility.[15] The smear could reflect different circular structures from conjugation to DOPA at different locations on the protein

chain. However, we failed to see any laddering which is considered as an indication for interchain conjugation products.

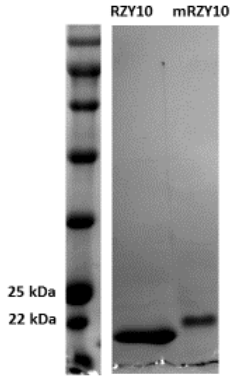


Figure B. 1. RZY10 and tyrosinase-converted RZY10 (mRZY10) on SDS-PAGE gel. mRZY10 showed a lower electrophoretic mobility that could be due to the increase in molecular weight (MW) and/or the decrease in hydrophobicity.

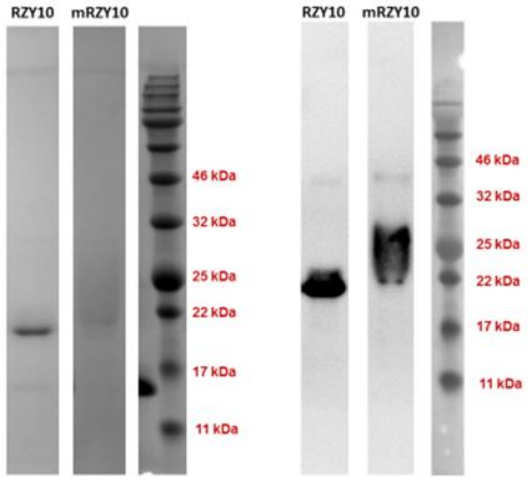


Figure B. 2. RZY10 or mRZY10 conjugation at pH 7.4. RZY10 or mRZY10 in 100 mM acetate buffer with 100 mM ascorbic acid were pH adjusted to 7.4 and were incubated at room temperature for 24 hours. Left: RZY10 appeared as a single band at the expected position on the SDS-PAGE gel. On the other hand, mRZY10 appeared as an upward smear. Right: Anti-T7 tag Western blot better visualized the smear in mRZY10.

In order to prevent intrachain conjugation, we used thrombin to cleave off the N-terminal T7 tag and 6xHis tag. The cleaved RZY10 (hereafter no tag-RZY10) was further converted by tyrosinase in a cobuffer of 10 mM acetate buffer/10 mM Tris buffer with 150 mM NaCl, 10 mM $CaCl_2$, and 10 mM ascorbic acid. Tris buffer was used to provide better buffer capacity near neutral pH. NaCl and $CaCl_2$ were necessary for thrombin cleavage and

were carried over to the following steps. The converted no tag-RZY10 (hereafter no tag-mRZY10) was incubated with RZY10 in the cobuffer at pH 7.4 for 24 hours at room temperature. Control groups were no tag-RZY10 alone, no tag-mRZY10 alone, and RZY10 with no tag-RZY10 prepared as described above. Because there was no histidine in no tag-mRZY10, we expected that the putative intrachain conjugation between histidine and DOPA would not occur and therefore expected to observe interchain conjugation products (that is, conjugation between DOPA on no tag-mRZY10 and histidine on RZY10).

Samples were analyzed by SDS-PAGE and anti-T7 tag Western blot as shown in Figure B. 3. Surprisingly, we failed to clearly visualize the protein band of no tag-RZY10 (lane 2, top image) and no tag-mRZY10 (lane 4, top image) samples with Coomassie blue staining. It was possible that the increased process steps (thrombin cleavage, purification of the cleaved products) resulted in significant protein losses. No clear high MW bands were observed in no tag-RZY10/RZY10 (lane 3, top image). On the other hand, we were able to see high MW products in no tag-mRZY10/RZY10 (lane 5, top image) with Coomassie blue staining. Moreover, these high bands in lane 5 had MWs corresponding to conjugation products between one no tag-mRZY10 (12.1 kDa) and one, two, three, and four RZY10 (14.1 kDa). These high bands also showed up on anti-T7 tag Western blot (lane 5, bottom image), and this result suggested that these bands were proteins with T7 tags. Because the T7 tag in no tag-mRZY10 was removed by thrombin, the only species with a T7 tag in the mixture was RZY10. However, from lane 3 in Figure B. 3 (bottom), it could be concluded that RZY10 by itself did result in any significant high MW bands. These two results together led us to conclude that the high MW bands we observed could only be conjugation products between no tag-mRZY10 and RZY10.

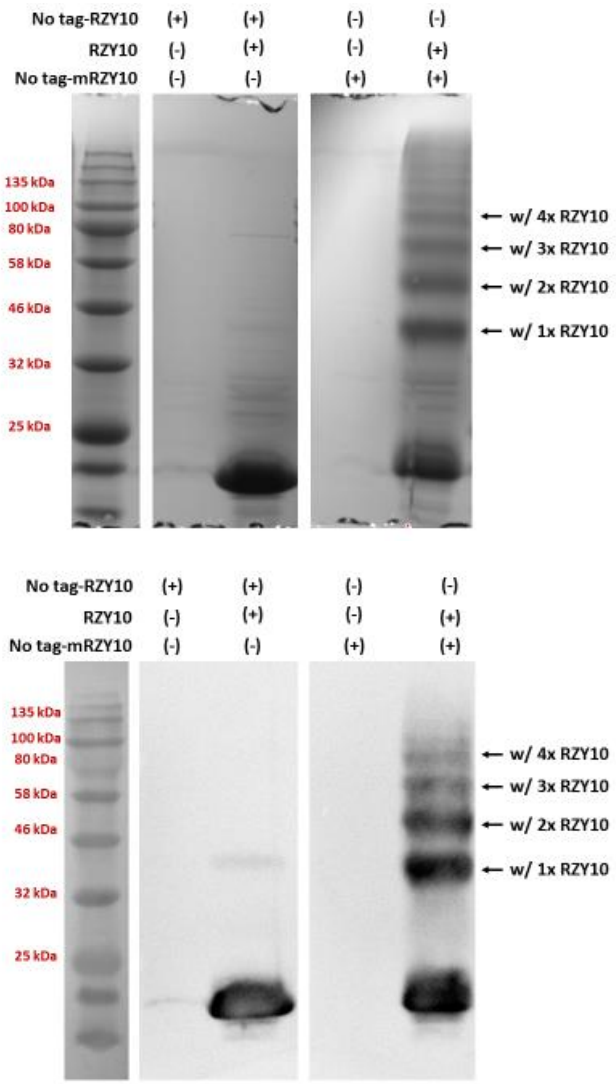


Figure B. 3. No tag-mRZY10 and RZY10 conjugation at pH 7.4. Proteins were in a cobuffer of 10 mM acetate buffer and 10 mM Tris buffer with 150 mM NaCl, 10 mM $CaCl_2$, and 10 mM ascorbic acid and were incubated at room temperature for 24 hours. Top: SDS-PAGE gel showing high MW products from reaction of no tag-mRZY10 (12.1 kDa) and RZY10 (14.1 kDa). Bottom: Anti-T7 tag Western blot image revealed that the high MW products contained T7 tags from RZY10. Arrows indicate numbers of RZY10 being conjugated to no tag-mRZY10. The band below 25 kDa ladder was single RZY10.

The above results supported the idea of protein/peptide conjugation via catechol-nucleophile Michael addition by DOPA residues. However, both histidine and lysine were in the protein sequence and can react with DOPA. To clarify whether lysine or histidine participated in the conjugation, we used acetic anhydride to acetylate the ε-amine of lysine in RZY10. The acetylation level was determined by a 2,4,6-trinitrobenzene sulfonic acid (TNBSA) assay. The acetylation level of acetylated RZY10 (acRZY10) was found to be ~99%. We also constructed another elastomeric ELP, SKGPG-Y2 that had no histidine in the sequence but still had tyrosine residues that could be converted to DOPAs. Tyrosinase-converted SKGPG-Y2 (mSKGPG-Y2) was incubated with acRZY10 at a one-to-one molar ratio at room temperature for 24 hours. Samples with mSKGPG-Y2 alone, acRZY10 alone, and SKGPG-Y2 with acRZY10 were also prepared as control groups. All samples were analyzed by SDS-PAGE. The results of the conjugation between mSKGPG-Y2 and acRZY10 are shown in Figure B. 4. Note that no ascorbic acid was used in the conjugation buffer. Since amines on acRZY10 were acetylated, the only possible conjugation between acRZY10 and mSKGPG-Y2 was between the His residues on acRZY10 and DOPA residues on mSKGPG-Y2. Based on the results in Figure B. 5, we did not expect to see either intrachain or interchain conjugations between mSKGPG-Y2. However, we also failed to see possible conjugations between acRZY10 and mSKGPG-Y2. No changes in the smearing pattern or shifts in MW were observed. Conjugation experiments were also carried out with a short peptide with the sequence of His-Arg-Gly-Asp-His. We verified that the presence of 200 mM ascorbic acid reduced the formation of smearing of mSKGPG-Y2 at molecular weights >25 kDa (Figure B. 5). Since ascorbic acid prevents the oxidation of DOPA, it is possible that the lower extent of DOPA oxidation resulted in less intrachain crosslinking within mSKGPG-Y2. When ascorbic acid was absent, there were no differences in the smearing between the groups with and without the peptide, and we thus excluded the possibility that the smearing was a result of crosslinking between mSKGPG-Y2 and the peptide. However, it was unexpected that interchain crosslinking, which was observed with mRZY10 in Figure B. 3, was not seen with mSKGPG-Y2.

We conclude that the reaction between DOPA and His or Lys was difficult to control possibly due to the high reactivity of DOPA towards nucleophiles and spontaneous oxidation of DOPA. An improved method of tyrosinase conversion used in Chapter 2 or in

our previous publication[16] can be used to prepare DOPA-containing proteins for this project such that a higher percentage of free DOPA can be preserved for the conjugation. It is also suggested to improve the design of materials for experiments to test the reaction specificity towards His over Lys. For example, fluorescently labeled peptides with His or Lys could be used together with a tyrosinase-converted ELP that has no His or Lys in the sequence. Fluorescently-labeled peptides would allow conjugation products to be detected using fluorescence techniques with a higher detection sensitivity than SDS-PAGE. Small proteins could also be used instead of peptides to increase the change in MW of the conjugation products and thus result in easier observation in SDS-PAGE gels.

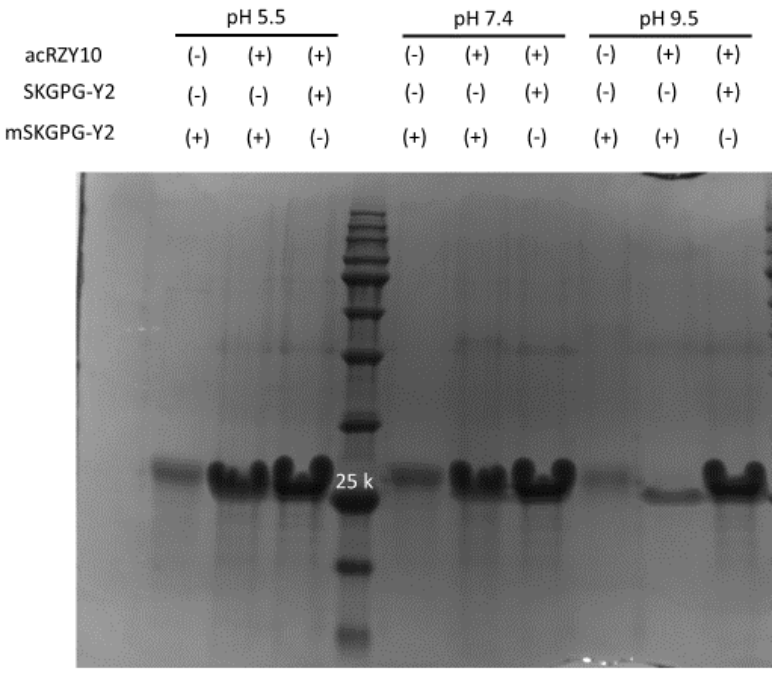


Figure B. 4. Conjugation between acRZY10 and mSKGPG-Y2 through DOPA-Histidine bonds was not observed. No changes in the smearing pattern or shifts in MW of the band near 25 kDa ladder were observed at pH 7.4 where the DOPA-Histidine conjugation was expected to occur. At pH 9.5, DOPA-Lysine conjugation was expected to occur and result in intrachain crosslinking of mSKGPG-Y2. The expected result of such conjugation would be a decrease in the intensity of the free mSKGPG-Y2 band around 25 kDa. However, such a decrease was not observed.

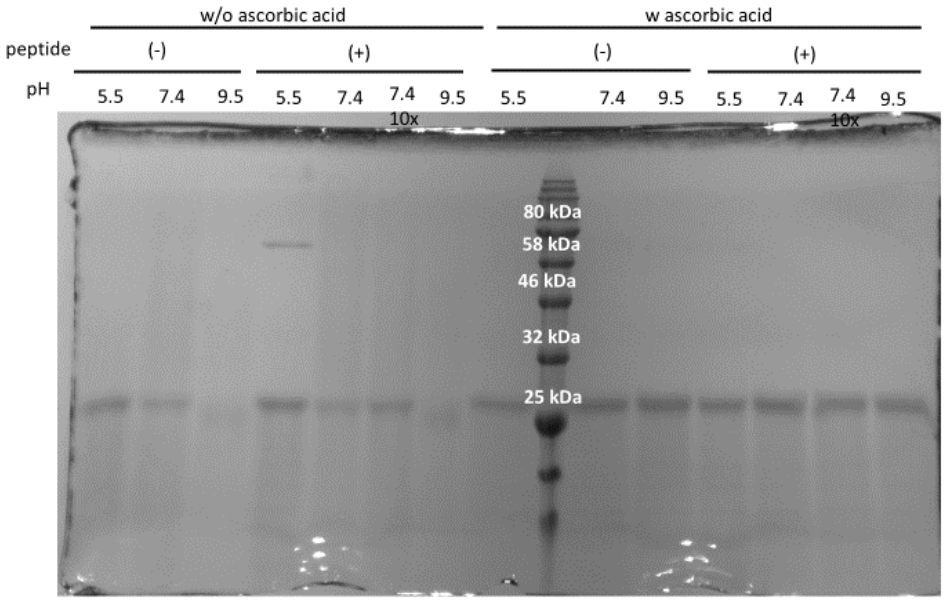


Figure B. 5. Ascorbic acid prevents the oxidation of DOPA. Less smearing was observed with mSKGPG-Y2 when no ascorbic acid was used in the buffer. This result is consistent with the fact that ascorbic acid decreases the reactivity of quinone towards nucleophiles by preventing the oxidation of DOPA to quinone. In addition, no difference was observed in the absence of ascorbic acid between the group with and without the His-containing peptide. The groups of pH 7.4 10x indicates that the peptide concentration used was 10 times higher then the other groups in the figure.

REFERENCES

(1) Seo, S.; Das, S.; Zalicki, P. J.; Mirshafian, R.; Eisenbach, C. D.; Israelachvili, J. N.; Waite, J. H.; Ahn, B. K. Microphase Behavior and Enhanced Wet-Cohesion of Synthetic Copolyampholytes Inspired by a Mussel Foot Protein. *J. Am. Chem. Soc.* **2015,** *137* (29), 9214-9217.

(2) Kerwin, J. L.; Turecek, F.; Xu, R.; Kramer, K. J.; Hopkins, T. L.; Gatlin, C. L.; Yates, J. R. Mass spectrometric analysis of catechol–histidine adducts from insect cuticle. *Anal. Biochem.* **1999,** *268* (2), 229-237.

(3) Kirchhofer, D.; Grzesiak, J.; Pierschbacher, M. Calcium as a potential physiological regulator of integrin-mediated cell adhesion. *J. Biol. Chem.* **1991,** *266* (7), 4471-4477.

(4) Saito, A.; Suzuki, Y.; Ogata, S.-i.; Ohtsuki, C.; Tanihara, M. Activation of osteo-progenitor cells by a novel synthetic peptide derived from the bone morphogenetic protein-2 knuckle epitope. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2003,** *1651* (1), 60-67.

(5) Tao, H.; Wu, Y.; Li, H.; Wang, C.; Zhang, Y.; Li, C.; Wen, T.; Wang, X.; He, Q.; Wang, D. BMP7-Based Functionalized Self-Assembling Peptides for Nucleus Pulposus Tissue Engineering. *ACS Appl. Mater. Interfaces* **2015,** *7* (31), 17076-17087.

(6) D'Andrea, L. D.; Iaccarino, G.; Fattorusso, R.; Sorriento, D.; Carannante, C.; Capasso, D.; Trimarco, B.; Pedone, C. Targeting angiogenesis: structural characterization and biological properties of a de novo engineered VEGF mimicking peptide. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (40), 14215-14220.

(7) Silva, G. A.; Czeisler, C.; Niece, K. L.; Beniash, E.; Harrington, D. A.; Kessler, J. A.; Stupp, S. I. Selective differentiation of neural progenitor cells by high-epitope density nanofibers. *Science* **2004,** *303* (5662), 1352-1355.

(8) Pillutla, R. C.; Hsiao, K.-c.; Beasley, J. R.; Brandt, J.; Østergaard, S.; Hansen, P. H.; Spetzler, J. C.; Danielsen, G. M.; Andersen, A. S.; Brissette, R. E. Peptides identify the critical hotspots involved in the biological activation of the insulin receptor. *J. Biol. Chem.* **2002,** *277* (25), 22590-22594.

(9) Lam, H.; Li, S.; Lou, N.; Chu, J.; Bhatnagar, R. In *Synthetic peptides cytomodulin-1 (CM-1) and cytomodulin-2 (CM-2) promote collagen synthesis and wound healing in vitro*,

Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, IEEE: 2004; pp 5028-5030.

(10) Liu, B.; Burdine, L.; Kodadek, T. Chemistry of periodate-mediated cross-linking of 3, 4-dihydroxylphenylalanine-containing molecules to proteins. *J. Am. Chem. Soc.* **2006,** *128* (47), 15228-15235.

(11) Chien, C.-Y.; Tsai, W.-B. Poly (dopamine)-assisted immobilization of Arg-Gly-Asp peptides, hydroxyapatite, and bone morphogenic protein-2 on titanium to improve the osteogenesis of bone marrow stem cells. *ACS Appl. Mater. Interfaces* **2013,** *5* (15), 6975-6983.

(12) Wouters, O. Y.; Ploeger, D. T.; van Putten, S. M.; Bank, R. A. 3, 4-Dihydroxy-L-Phenylalanine as a Novel Covalent Linker of Extracellular Matrix Proteins to Polyacrylamide Hydrogels with a Tunable Stiffness. *Tissue Engineering Part C: Methods* **2016**.

(13) Lee, H.; Rho, J.; Messersmith, P. B. Facile conjugation of biomolecules onto surfaces via mussel adhesive protein inspired coatings. *Adv. Mater.* **2009,** *21* (4), 431-434.

(14) Brownlee, A. G.; Elvin, C. M.; Wong, D. C. C.; Lesieur, E.; Kim, M.; Pearson, R. D.; Lyons, R. E.; Huson, M. G.; Nairn, K. M. Design and facile production of recombinant resilin-like polypeptides: gene construction and a rapid protein purification method. *Protein Engineering, Design and Selection* **2007,** *20* (1), 25-32.

(15) Evans, T. C.; Benner, J.; Xu, M.-Q. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem.* **1999,** *274* (26), 18359-18363.

(16) Brennan, M. J.; Kilbride, B. F.; Wilker, J. J.; Liu, J. C. A bioinspired elastin-based protein for a cytocompatible underwater adhesive. *Biomaterials* **2017,** *124*, 116-125.

APPENDIX C. RECOMBINANT PROTEIN-BASED SCAFFOLD FOR RAT CALVARIAL CRITICAL DEFECT HEALING

PROJECT BACKGROUND

The goal of this project was to construct a recombinant protein-based hydrogel for stem cell encapsulation that can promote bone healing in a rat calvarial model. After the biocompatibility and the efficacy of the system was verified, the hydrogel system was adapted to study the effect of bioactive ligand presentation on guiding the encapsulated stem cell differentiation.

RESULTS AND DISCUSSION

We chose thiol-maleimide crosslinking chemistry for the hydrogel formation because of its cytocompatibility for cell encapsulation. A series of Cys-containing ELPs were designed with different number of repeats of the BMP-2 peptide followed by a C-terminal His tag. They were YC12 (no BMP-2 peptide), YC12-M (one BMP-2 peptide), YC12-D (two BMP-2 peptides), and YC12-T (four BMP-2 peptides). The expression of YC12-T in Rosetta2 pLysS was confirmed by anti-His tag Western blot (**Error! Reference source not found.**).

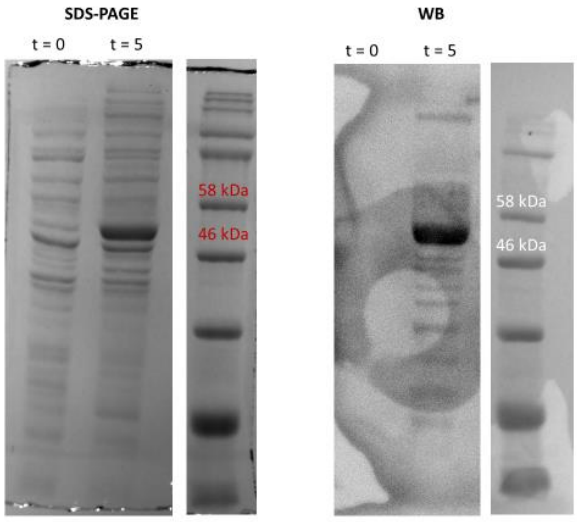


Figure C. 1.The expression of YC12-T in Rosetta2 pLysS was confirmed by anti-His tag Western blot. The expected MW of YC12-T was 43.7 kDa.

YC12 proteins were successfully purified by a modified inverse temperature cycling (ITC) method. Briefly, the cell pellet was resuspended in 8 M urea, sonicated, and centrifuged to get clear lysate. The clear lysate was then salted out with 10-20 wt% (See Appendix E. 10) ammonium sulfate to get the crude protein pellet. The crude protein pellet was then resuspended in PBS for YC12 and YC12-M or in PBS with 2 M urea for YC12-D and YC12-T before the ITC process. The cold spin was done at 4 °C, and the hot spin was triggered by the addition of 3 M NaCl and centrifuged at 30 °C. β-mercaptoethanol (BME) was used throughout the purification, and a final reduction with 50 mM DTT was performed before the purified protein solution was subjected to a 24-h dialysis in water. After dialysis, the supernatant was collected and flash-frozen in liquid nitrogen for lyophilization. The purity of the lyophilized protein was >95% for all four proteins (Figure C. 2).

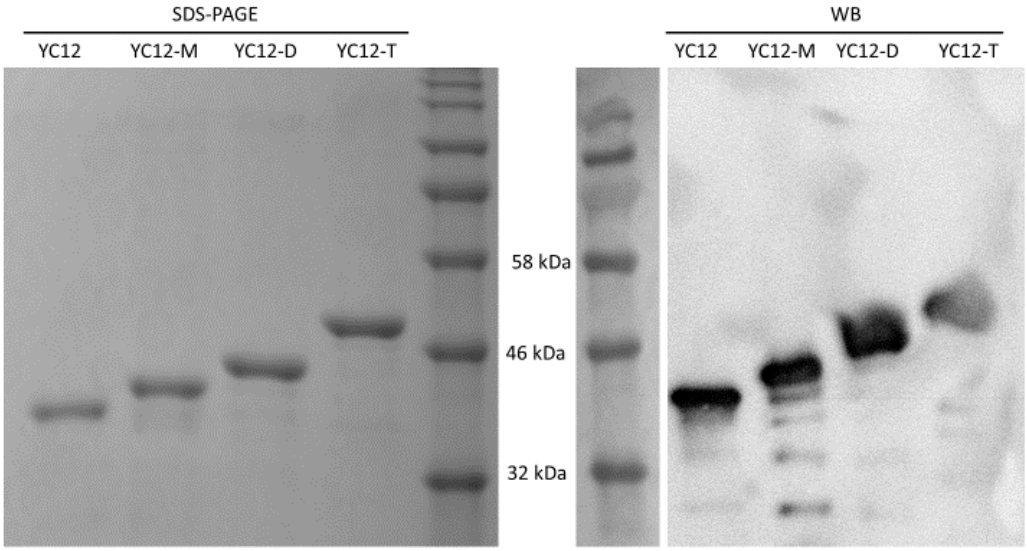


Figure C. 2. The purity of the lyophilized protein was >95% for YC12, YC12-M, YC12-D, and YC12-T as determined by densitometry on a Coomassie blue-stained SDS-PAGE gel. An anti-His tag Western blot confirmed the presence of the His tag on the purified proteins.

We evaluated the osteoconductivity of YC12 proteins by 2D culture of human mesenchymal stem cells (MSCs) on spin-coated coverslips. We first screened the spin coating condition such that YC12, YC12-M, and YC12-T had the same coated amount. We

found that pH was the main factor for the coating efficiency and that all three proteins had similar coating efficiency at pH 6.3 (Figure C. 3).

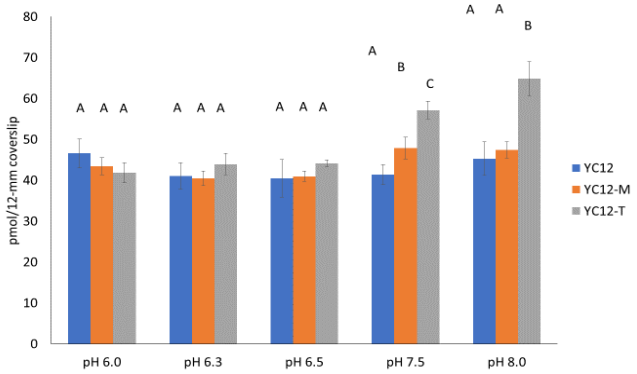


Figure C. 3. YC12, YC12-M, and YC12-T had similar coated amounts at pH 6.3. To spin coat, 15 µL of protein solution in MQ at 5 mg/mL was mixed with 3.75 µL of 15 wt% H_2O_2 (all pH adjusted to the desired pH) and then dispensed onto the center of an acid-washed coverslip. The coverslip was spun for 60 sec at 4000 rpm and incubated at 37 °C for 30 min. After incubation, the coverslip was sterilized in 70% ethanol for 30 min at room temperature. The amount of coated protein was quantified using a BCA kit. Different letters indicate Tukey groups that are statistically different ($p < 0$ .05).

We next tested the stability of the coated protein. The coverslips were prepared as described above, incubated in PBS at 37 °C without shaking, and quantified with BCA assay at predetermined time points. As shown in Figure C. 4, all three proteins showed similar stability over the 7-day incubation period.

After we confirmed that all coated proteins had similar stability on the coverslips, we examined if the presentation of different numbers of the BMP-2 peptide repeats affected cell attachment on the coated surfaces. To account for the difference in the number of the peptide repeats, coverslips of dimeric and tetrameric presentation of the BMP-2 peptides were prepared by mixing YC12-D or YC12-T with YC12 to achieve the same overall peptide surface density as the surface with monomeric presentation (with YC12-M). Two different densities (one BMP-2 peptide per YC12 backbone and ten BMP-2 peptides per YC12 backbone) were tested. Fischer 344 rat bone MSCs or NIH/3T3 mouse fibroblasts were seeded at 3000 cell/cm^2 on the protein-coated coverslips and stained for actin fibers and nuclei 5 and 15 h after seeding. Cell number and cell area were quantified (Figure C. 5 and Figure C. 6).

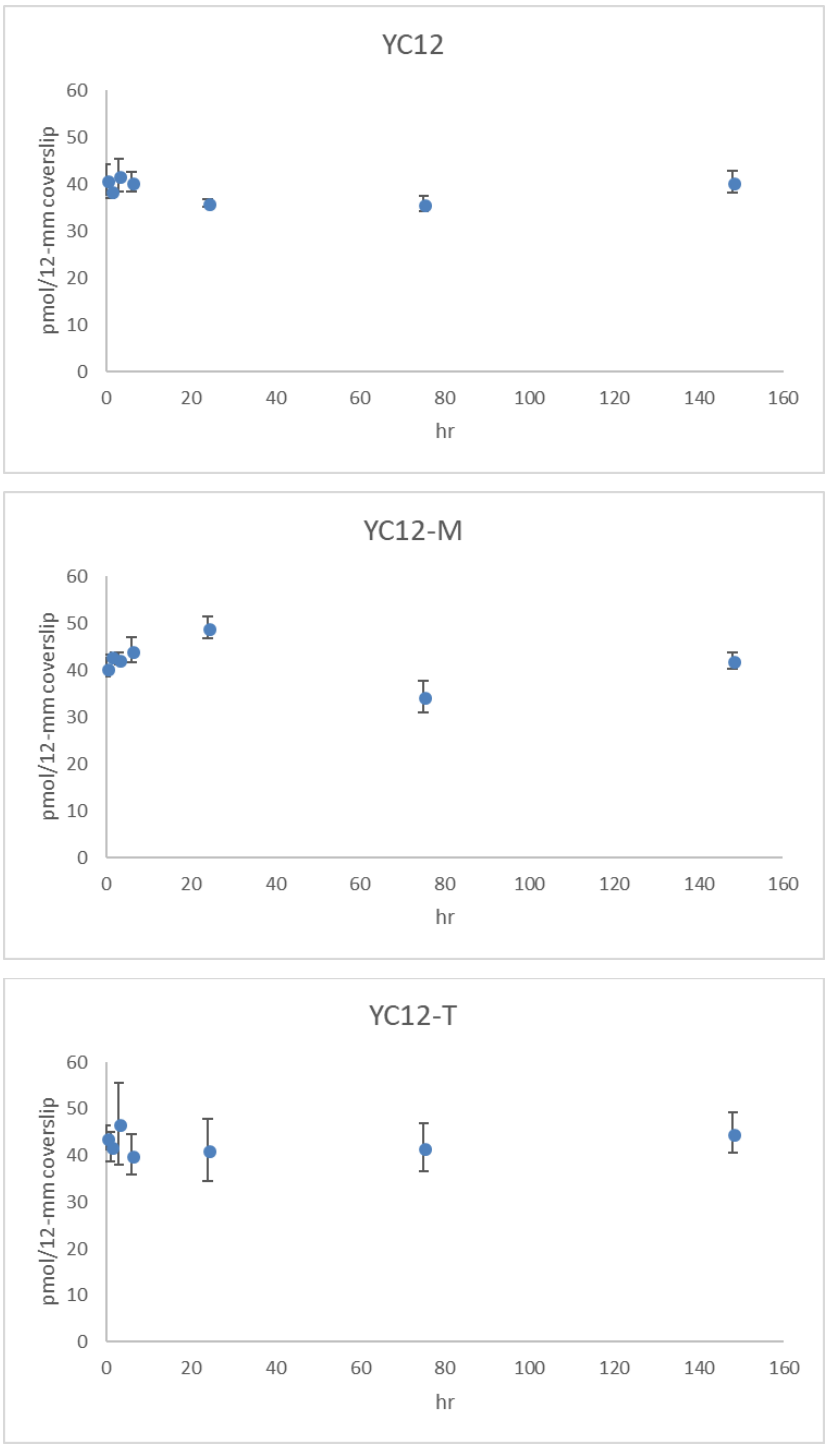


Figure C. 4. Coated YC12 proteins were stable in PBS for 7 days. Protein-coated coverslips were incubated in PBS at 37 °C without shaking. At designated time points, the protein remaining on the coverslip was quantified by using a BCA assay.

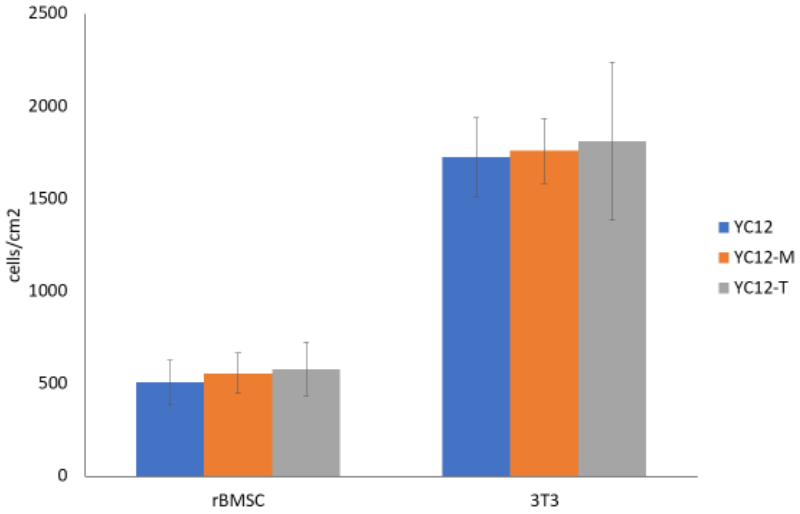


Figure C. 5. Similar number of cells attached on YC12, YC12-M, and YC12-T surfaces. Fischer 344 rat bone MSCs (rBMSCs) and NIH/3T3 mouse fibroblasts were seeded at 3000 cells/cm^2 and counted by nuclei staining 5 h after seeding (n =3).

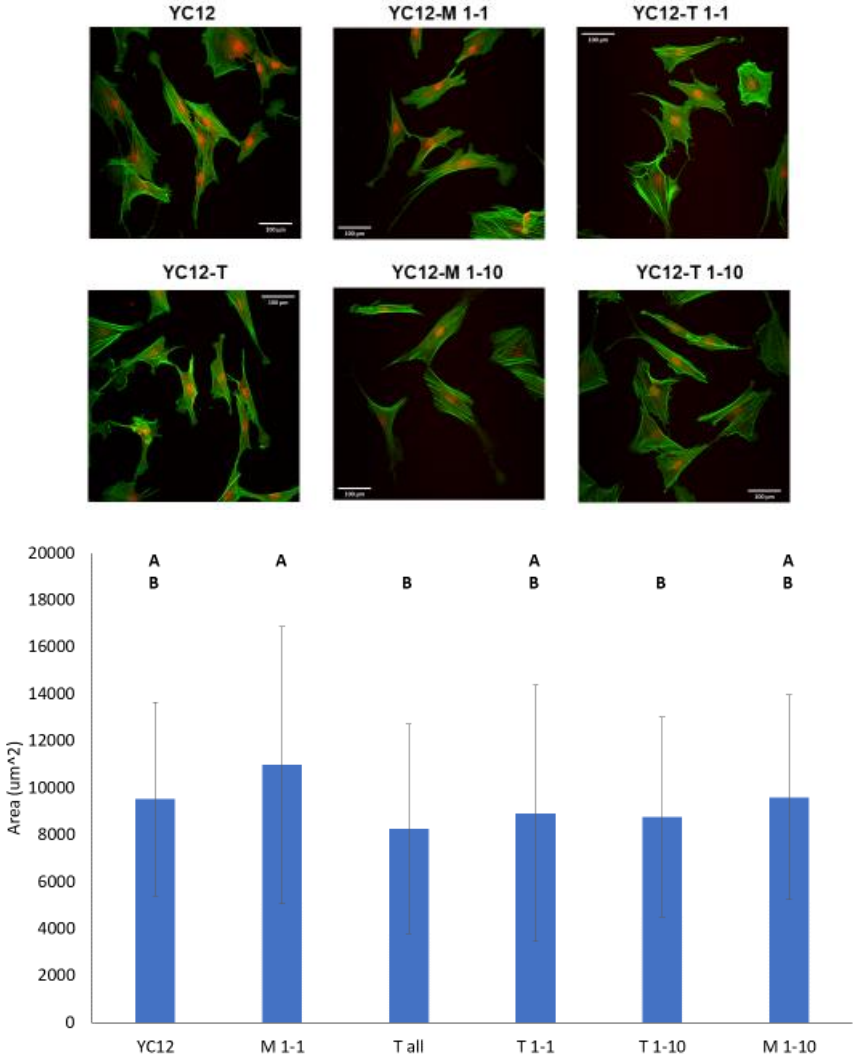


Figure C. 6. Similar cell morphology and area of rBMSCs on surfaces with different BMP-2 peptide presentations. Top: Actin (green) and nucleus (red) staining of rBMSCs on protein surfaces 15 h after seeding. Bottom: Cell area quantification based on actin staining (n > 100 cells). Different letters indicate Tukey groups that are statistically different (p < 0.05).

Knowing that the presentation did not have a significant effect on cell attachment and morphology, we decided to test the osteoconductivity of surfaces with solely YC12, YC12-M, and YC12-T. Note that the total amount of protein coated was kept the same and therefore there were more BMP-2 peptides presented on the YC12-T surface in a tetrameric presentation. Human MSCs (hMSCs) were cultured on the surfaces until ~80% confluency then switched to osteogenic differentiation medium (ODM) following the regular bone differentiation protocol. On days 4 and 7 in ODM, cells were lysed and assayed for alkaline phosphatase activity (ALP) (Figure C. 7). On days 9 and 13 in ODM, cells were stained with Alizarin Red S (ARS) for calcium deposition (Figure C. 8 and Figure C. 9). Interestingly, we found that Alizarin Red staining was most intense on YC12 surfaces. YC12-M and YC12-T surfaces showed similar staining results (Figure C. 8), and there was no statistical difference in extracted ARS amount between YC12-M and YC12-T surfaces (Figure C. 9). These results were the opposite of the expected trend. Since the efficacy of the BMP-2 peptide is concentration dependent, we hypothesized that the overall concentration of the YC12-T surfaces was outside of the optimal range and hence the lower extent of osteogenesis than that of YC12-M. However, it was not clear why the YC12 surface without any BMP-2 peptides had more intense staining than the other two surfaces.

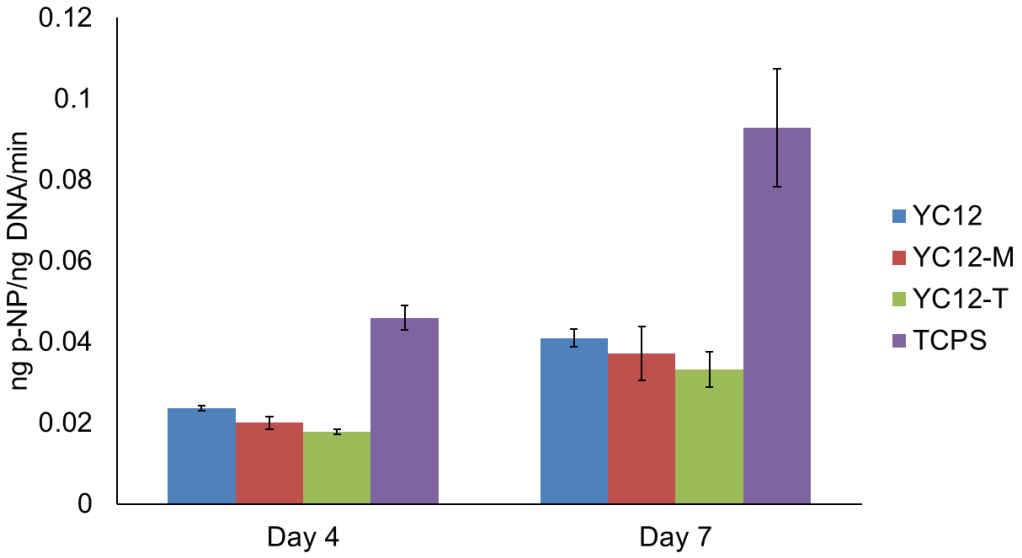


Figure C. 7. Alkaline phosphatase (ALP) activity increased from day 4 to day 7 on all surfaces. Cells on the tissue culture poly(styrene) (TCPS) surface had the highest ALP activity. Within protein surfaces, YC12 resulted in the highest ALP activity followed by YC12-M then YC12-T.

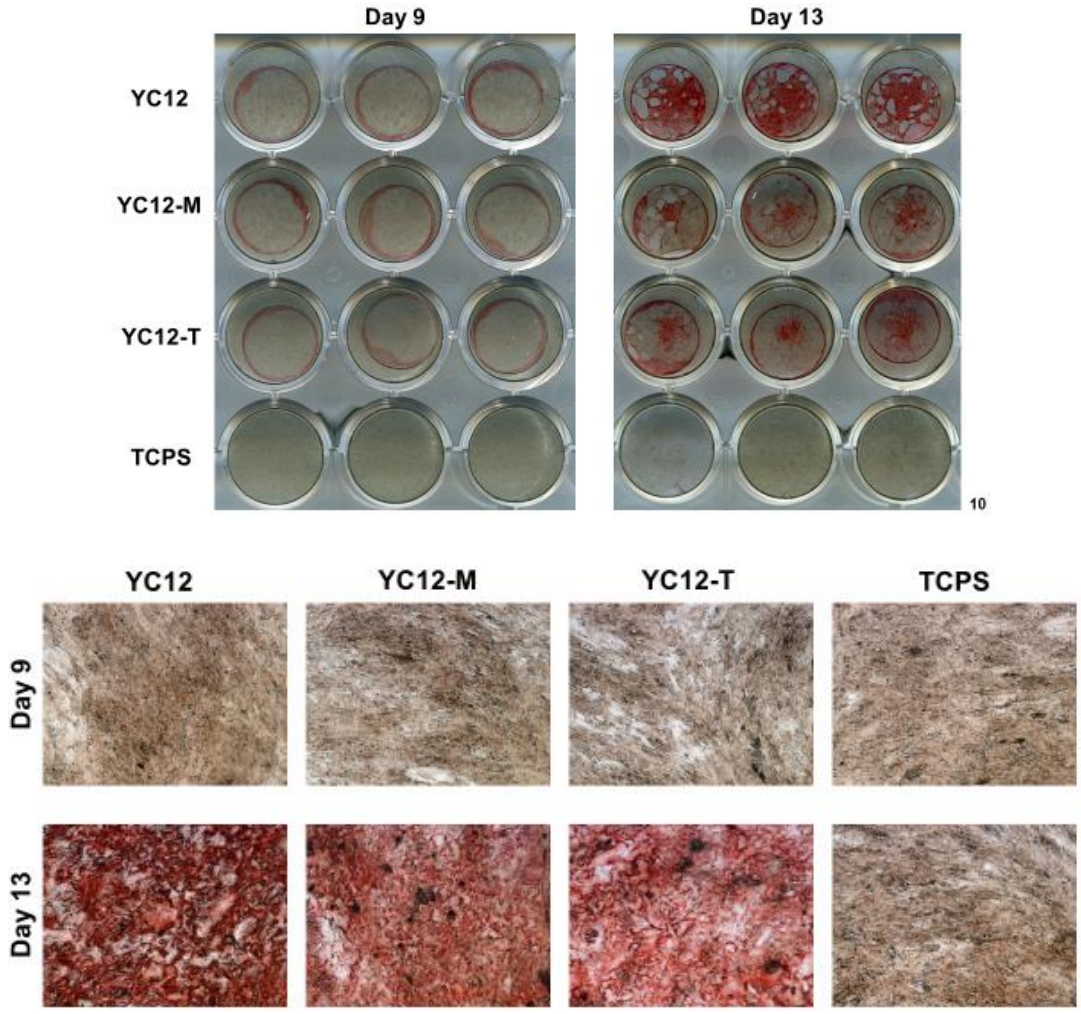


Figure C. 8. More calcium deposition stained by Alizarin Red S (ARS) was observed on the YC12 surfaces than the YC12-M and YC12-T surfaces. Top: Staining images of 24 well plates on day 9 and day 13. Bottom: 4x images at the center of the wells. The trend of ARS staining on day 13 is consistent with the ALP activity results. Together, these results suggested that YC12 promoted a higher extent of osteogenesis.

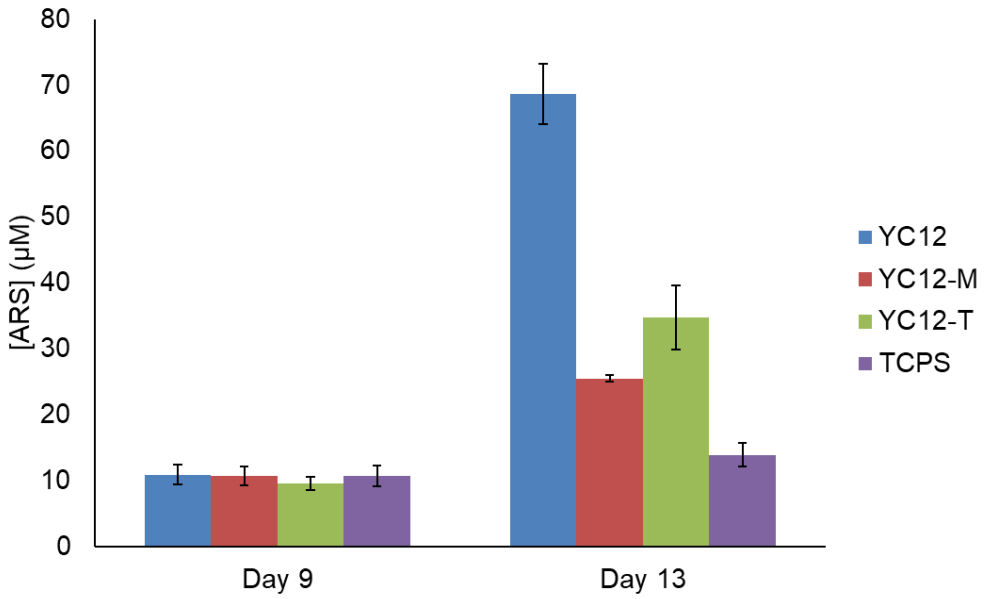


Figure C. 9. ARS quantification by acid extraction. The result showed a similar trend to what was observed with the staining images in Figure C. 8.

Meanwhile, we had decided to use YC12-T as the material for an *in vivo* study in rats. In order to remove endotoxin from the protein, a modified Ni column (Ni-IPA) was used for protein purification. Briefly, the column was loaded with protein and then repeatedly washed by a native wash buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME, pH 8.0) followed by an IPA-buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME in 60 (v/v)% isopropyl alcohol, pH 8.0). A representative purification gel is shown in Figure C. 10. Throughout the endotoxin purification, all plasticware was non-pyrogenic, and all glassware was washed in 0.1 M NaOH for >30 min and/or baked at 250 °C for >30 min for depyrogenation.

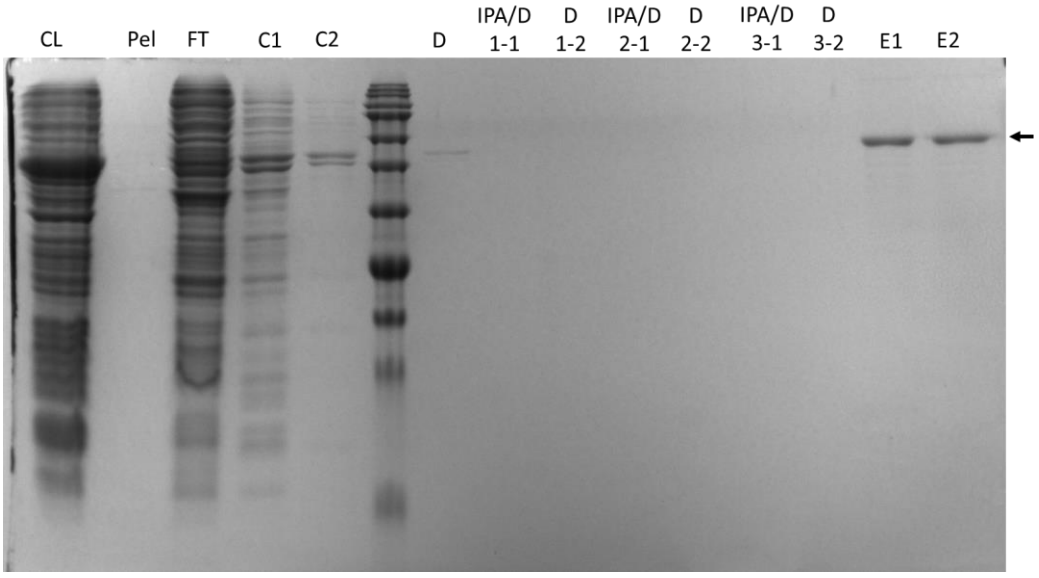


Figure C. 10. Purification of YC12-T by Ni-IPA column. CL: cleared lysate; Pel: pellet of cell debris after sonciation; FT: flow through; C: wash of buffer C (8 M urea, 10 mM Tris, 100 mM NaH_2PO_4-H_2O, 300 mM NaCl, 10 mM BME, pH 6.3); D: wash of buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME, pH 8.0); IPA/D: wash of IPA-buffer D (10 mM Tris, 300 mM NaCl, and 10 mM BME in 60 (v/v)% isopropyl alcohol, pH 8.0); E: elute of buffer E (10 mM Tris, 500 mM imidazole, 300 mM NaCl, pH 8.0).

A semi-quantitative gel clot limulus amebocyte lysate (LAL) assay (ToxinSensor^{TM} Gel Clot Endotoxin Assay Kit, GenScript, Cat. No. L00351) was used to evaluate the level of residual endotoxin in the Ni-IPA purified proteins. Briefly, the presence of endotoxin in the sample of interest will react with LAL assay reagents to form a gel clot (Figure C. 11). This formation depends on the endotoxin concentration. Therefore, by comparing the gelation results from the samples that were serially diluted to those from the endotoxin

standards, a range of the endotoxin content in the sample of interest can be determined. The result of the assay suggested that our Ni-IPA purified proteins had an endotoxin concentration <0.125 EU/mg protein (Table C. 1). This result was considered to be satisfactory for not eliciting severe immune responses in a rat of 250 g implanted with 10 mg protein since the exposure limit of a rat is EU/kg body weight.

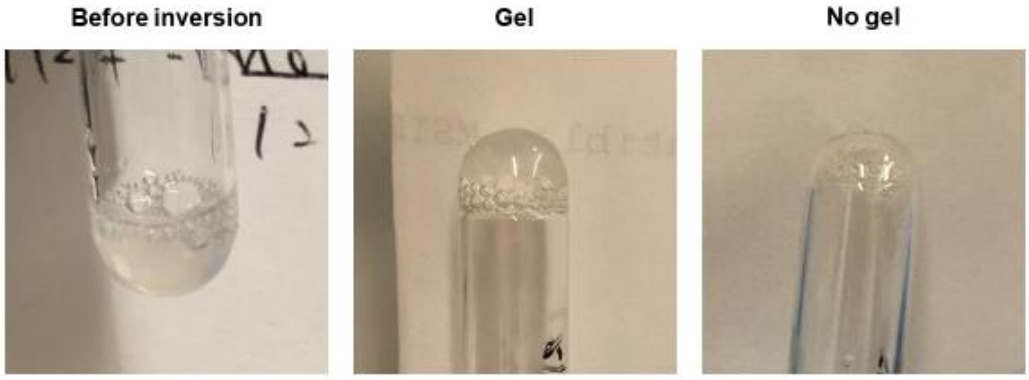


Figure C. 11. Representative images of gel clot formation in LAL assay.

Table C. 1 Gel clot LAL assay results of Ni-IPA-purified YC12-T

Group	Result	Note
IPA – no LAL	-	ELP gelation did not happen
IPA – 1: 100	-	< 0.125 EU/mg protein
IPA – $1:10^3$	-	
IPA – $1:10^4$	-	
IPA – $1:10^5$	-	
IPA – $1:10^6$	-	

A 4-arm PEG maleimide (PEG-MAL) was chosen for the maleimide-containing material (10 kDa, SKU: A7018-1 / 4ARM-MAL-10K, JenKem Technology USA). To prepare the cell-encapsulated gel for implantation, YC12-T was dissolved in MQ at 155 mg/mL and filtered. PEG-MAL was dissolved in PBS at 155 mg/mL and filtered as well. Solutions were kept on ice throughout the process. A cell pellet of 5×10^5 of previously-

harvested rat bone MSCs (passage 1) from 12-week old Fischer 344 rats were resuspended in 6 μL of PBS first and added with 5 μL of PEG-MAL. The resuspension was then pipetted into a sterilized circular silicone mold (8 mm in diameter and 0.5 mm in thickness) and kept on ice. Meanwhile, 4 μL of ice-cold 10x PBS was added into 40 μL of the YC12-T solution to make a final concentration of 1x PBS. The protein solution was then mixed with the cell resuspension in the mold with a positive-displacement pipette. Gel formation was observed immediately after mixing due to the fast reaction between thiols and maleimides. The final gel composition was 16 wt% of polymers with a 1-1 protein-to-PEG ratio. The rat bone MSCs were encapsulated at 5×10^6 cells/mL.

The gel in the silicone mold was transferred into a 6-well plate and moved to a cell culture incubator for an hour for further crosslinking. After the incubation, 3 mL of the osteogenic differentiation medium (ODM, Low glucose-DMEM with 1% FBS, 1% P/S, 10 mM β-glycerol phosphate, 100 nM dexamethasone, and 50 μM ascorbic acid-2-phosphate) was added into each well to cover the gel in the silicone mold (Figure C. 12). The gel was cultured overnight before the implantation the next day. Collagen sponges loaded with recombinant human BMP-2 (rhBMP-2) were prepared as positive controls. Briefly, an 8-mm disk of collagen sponge was punched out from a collagen sheet (Helistat collagen, Integra Lifesciences #3410ZX) using a biopsy punch. A pellet of 5×10^5 rat bone MSCs were resuspended in 56 μL of culture medium with 2 μg of rhBMP-2 and loaded onto both sides of the disk with half of the resuspension at a time. The cell-laden collagen sponge was incubated for an hour in the cell culture incubator without medium to allow cell penetration into the sponge (Figure C. 12).

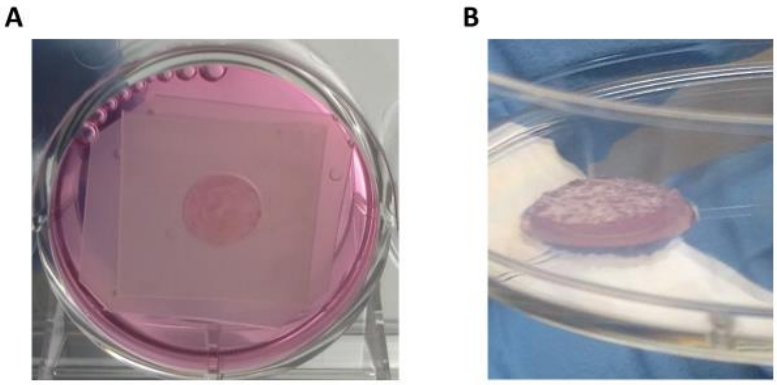


Figure C. 12. Images of (A) cell-encapsulated YC12-T/PEG-MAL hydrogel and (B) cell-laden collagen sponge.

The surgery to create the critical size defect in the calvaria of Fischer 344 rats (12-week-old males) and implant the cell-laden scaffolds was performed by Dr. Gert J. Breur (Department of Veterinary Clinical Sciences, Purdue University) following a previously published protocol.[1] Rats were divided into three groups: cell-encapsulated protein hydrogel, cell-laden collagen sponge, and empty defects as negative controls. Each group had 7-8 rats. Six weeks after implantation, *in vivo* CT was performed using MiLabs VECTor+ (Purdue Bioscience Imaging Facility). Twelve weeks after implantation, rats were euthanized, and the defect sites were harvested and preserved in formalin before *ex vivo* μCT (Scanco μCT u40, Purdue Bone & Body Composition Core).

As shown in Figure C. 13, no new bone formation was observed in the group with protein hydrogels. On the other hand, there was some new bone formation in the collagen group. Unexpectedly, the group with empty defects showed a significant extent of healing at the 6th week. The literature suggested that a defect size of 8 mm in diameter in rat models was a critical size defect and that spontaneous healing was not expected. The *ex vivo* μCT results at the 12th week (Figure C. 14) did not show any significant new bone formation compared to the 6th week in any of the groups. Haemotoxylin and Eosin staining of the defect site filled with the protein hydrogel (Figure C. 15) or the collagen sponge (Figure C. 16) suggested the formation of scar tissue in the defect sites and slow degradation of the protein hydrogel.

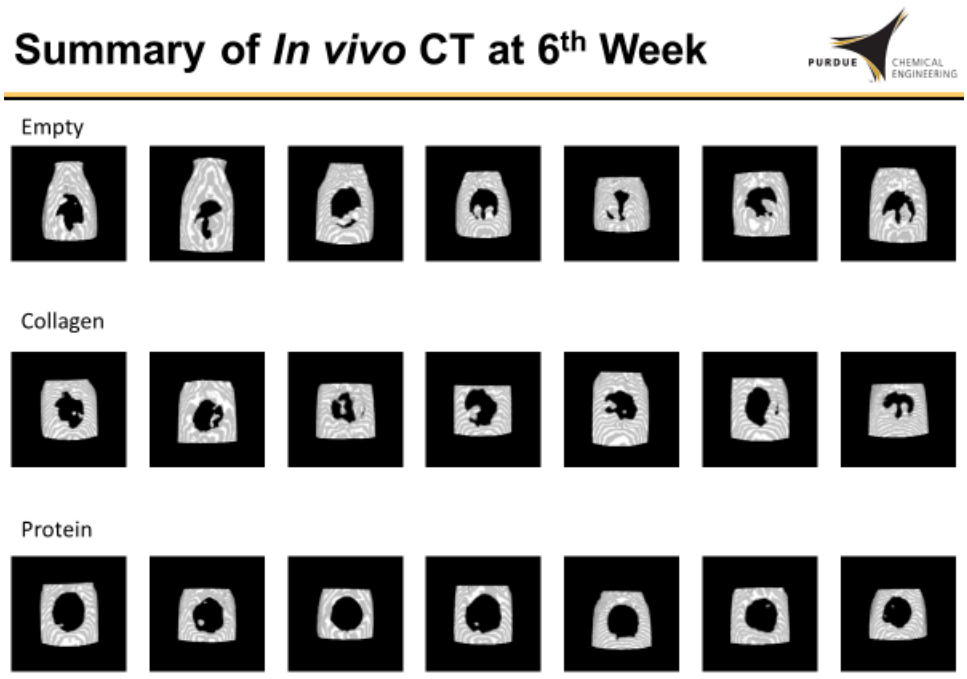


Figure C. 13. Images of *in vivo* CT at 6 weeks after implantation of the scaffolds. Empty: the group of empty defects; collagen: the group of cell-laden collagen sponges; protein: the group of cell-encapsulated YC12-T/PEG-MAL hydrogels.

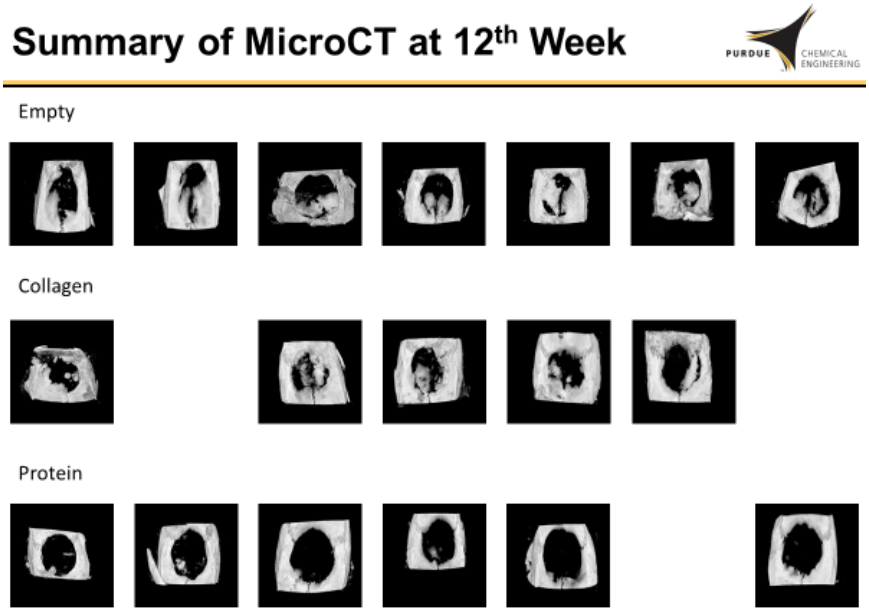


Figure C. 14. Images of *ex vivo* μCT at 12 weeks after implantation of the scaffolds. Empty: the group of empty defects; collagen: the group of cell-laden collagen sponges; protein: the group of cell-encapsulated YC12-T/PEG-MAL hydrogels. Images are in the same order as in Figure C. 13. Images from rats that did not survive to the 12th week are not shown.

.

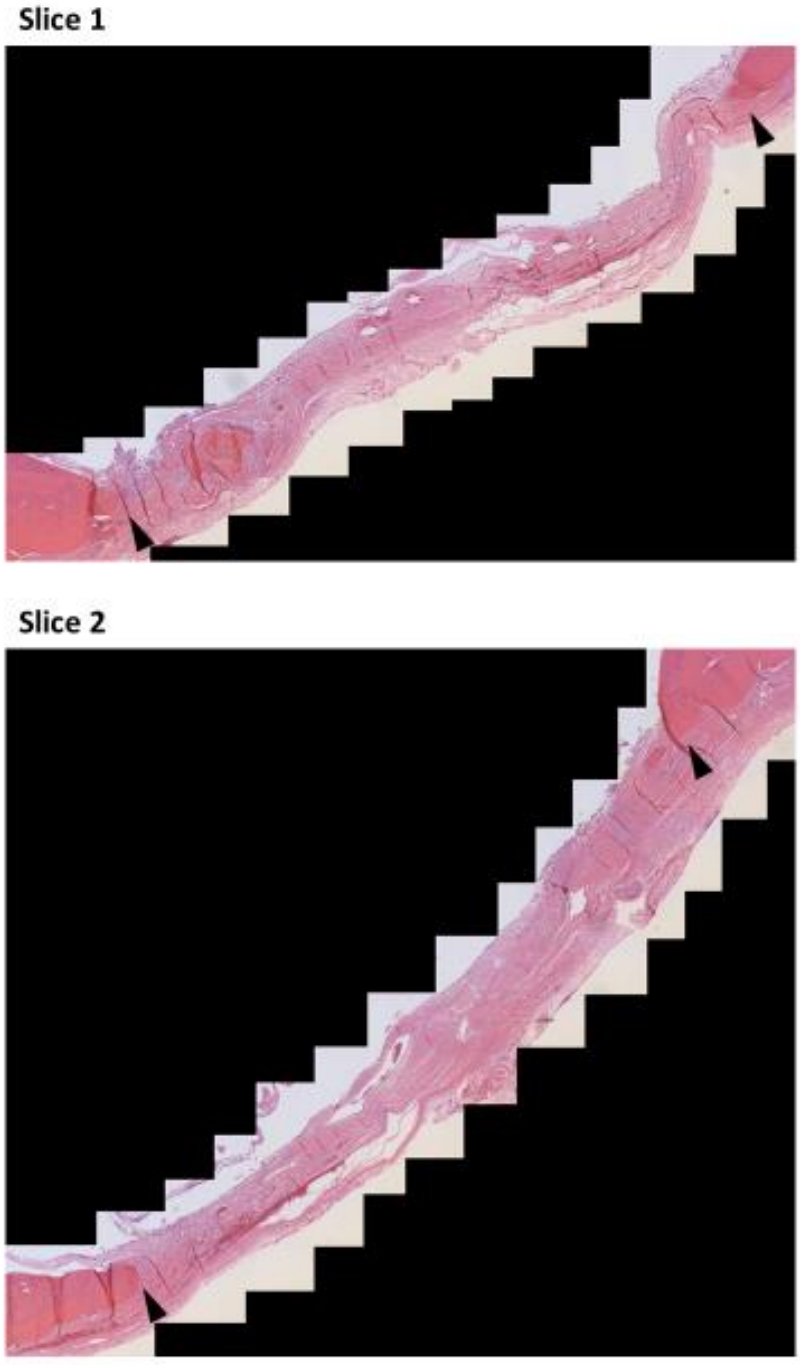


Figure C. 15. Haemotoxylin and Eosin staining images of a defect implanted with a protein hydrogel (rat 30). Bone is stained as dark pink. Black arrowheads mark the edge of the defect.

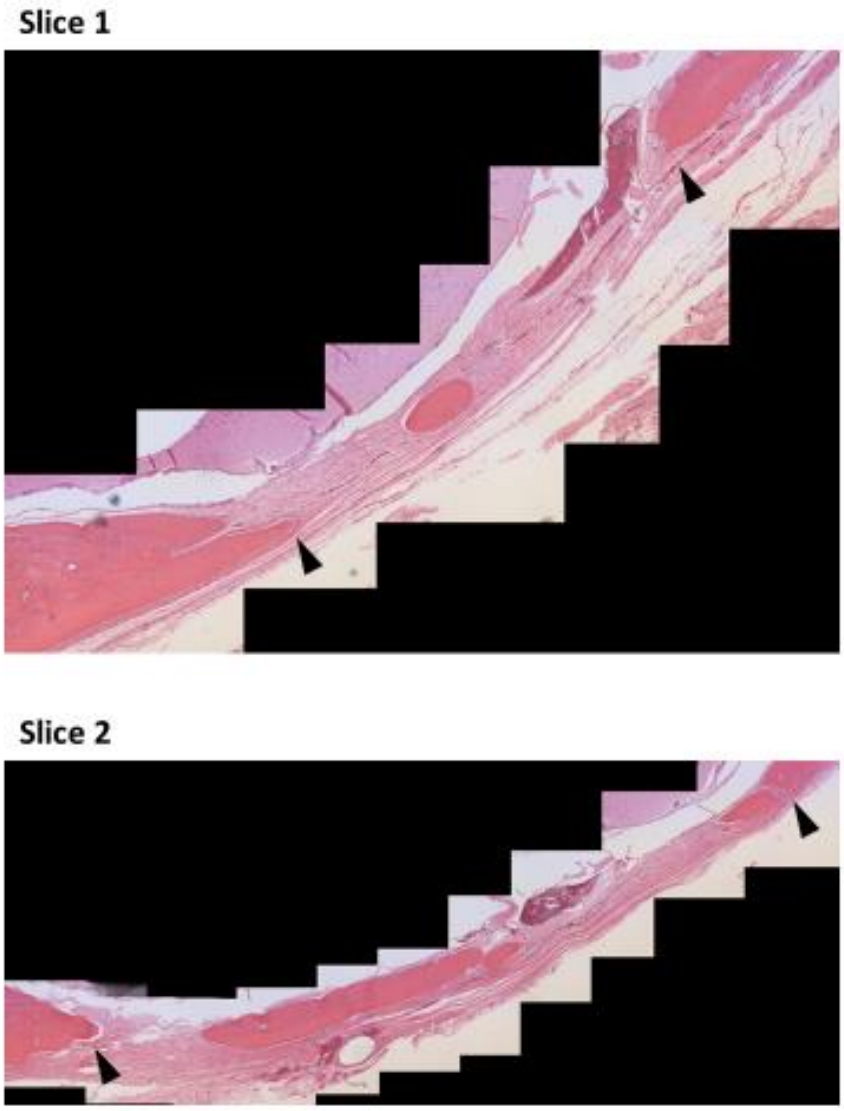


Figure C. 16. Haemotoxylin and Eosin staining images of a defect implanted with a collagen sponge (rat 23). Bone is stained as dark pink. Black arrowheads mark the edge of the defect.

We also evaluated the *in vitro* bone differentiation potential of the rat bone MCSs used for the implantation. Cells were cultured on TCPS and switched to ODM when the confluency reached ~70%. Cells were stained for ALP activity and calcium deposition on week 3 in ODM. The control group was cultured in growth medium (GM) without switching to ODM for the same period of time. As shown in Figure C. 17, slightly higher ALP activity was observed with the GM group. This finding at a late time point was expected due to slower differentiation without the inductive properties of ODM. No calcium deposition was observed with the GM group (Figure C. 18). However, the ODM group did not show significant calcium deposition either. This result suggested that the harvested rat bone MSCs were not able to effectively differentiate into the bone lineage. It is also possible that the cells we isolated from 12-week-old rats were less multipotent than those harvested from younger rats.

In summary, we successfully designed and produced thiol-containing ELPs (YC12 proteins) that were used to form hydrogels through thiol crosslinking. A series of YC12 proteins with different number of repeats of the BMP-2 peptide were constructed. Although the 2D differentiation results did not suggest any advantage of presenting the peptide in tandem repeats, it is possible that a different effect would be observed in 3D culture. We also demonstrated successful removal of endotoxins from proteins produced by *E. coli*, and these proteins were used successfully in rat *in vivo* studies without eliciting any significant immune response. Despite the unexpected results from this pilot study, this system of using YC12 proteins still merits further exploration to construct a 3D gel with different bioactive peptides that can be used both *in vitro* and *in vivo*.

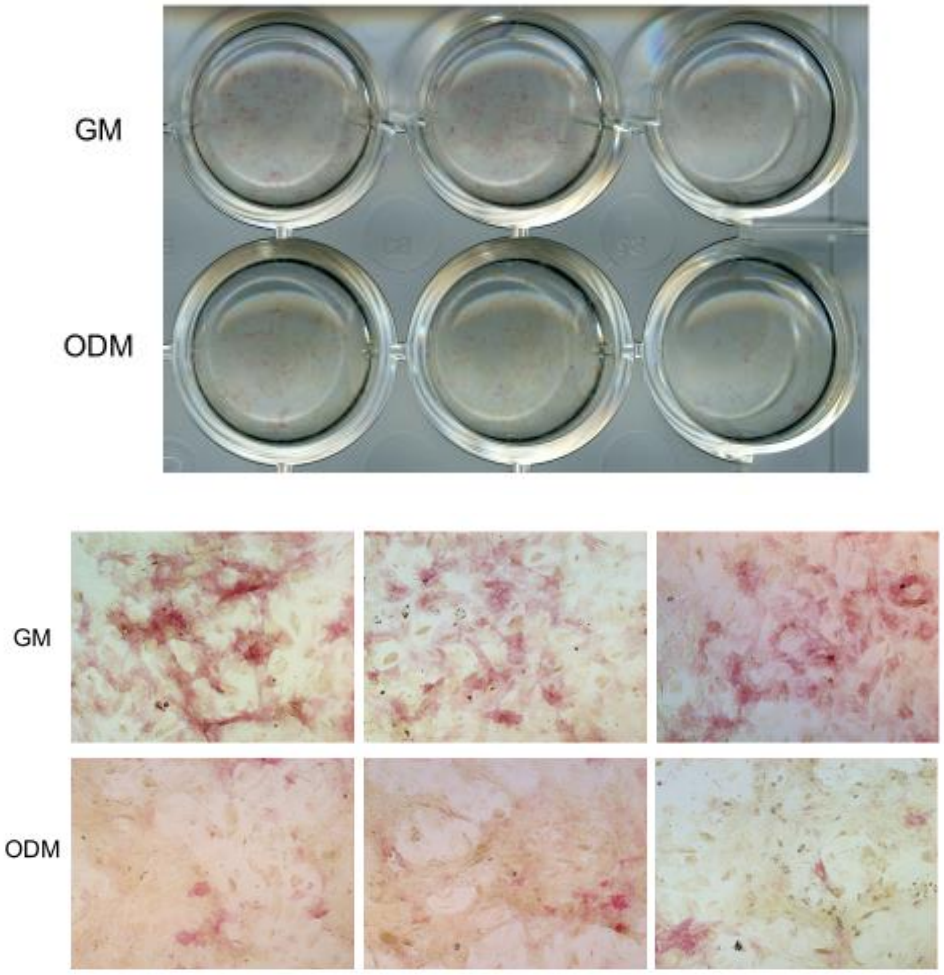


Figure C. 17. Low ALP activity as assessed by staining rat bone MSCs after 3 weeks of culture in osteogenic medium (ODM). Top: whole well image of cells cultured in growth medium (GM) and in ODM. Bottom: 4x microscope images of the center part of the well.

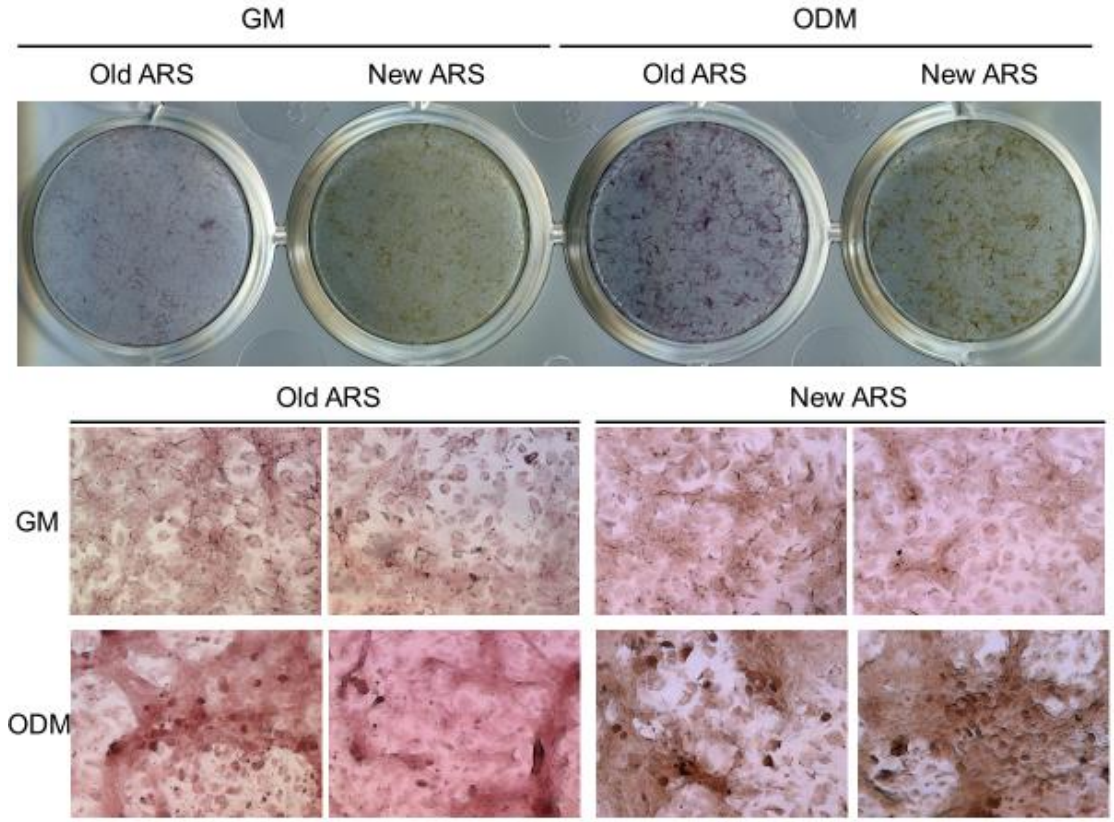


Figure C. 18. Significant calcium deposition was not observed with rat bone MSCs. Two different sources of ARS dyes were used. The ODM group showed slightly more calcium deposition than the GM group, but the overall staining intensity of the ODM group did not suggest strong osteogenesis of rat bone MSCs after culture in ODM for 3 weeks.

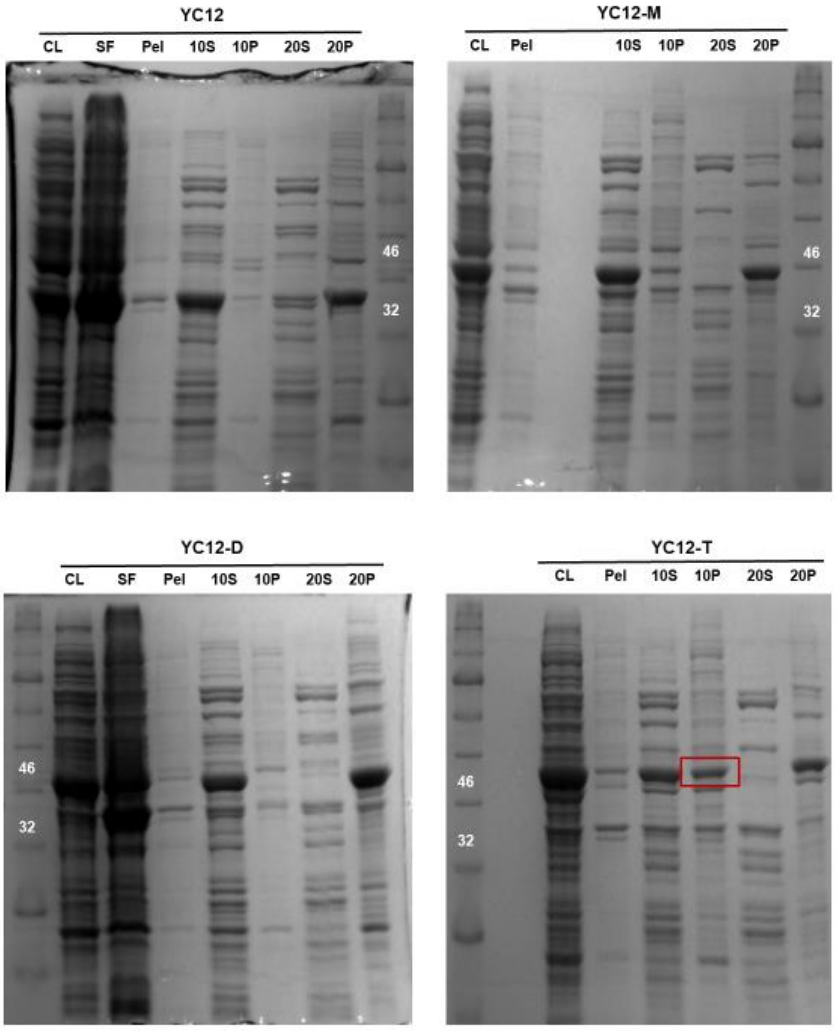


Figure C. 19. Protein purification using ammonium sulfate salting out with YC12 proteins. Addition of ammonium sulfate at 10 wt% followed by 20 wt% worked well for YC12, YC12-M, and YC12-D. YC12-T partially salted out at 10 wt% (red box). CL: cleared lysate; SF: soluble fraction; Pel: cell pellet; 10S and 10P: 10 wt% supernatant and pellet; 20S and 20P: 20 wt% supernatant and pellet.

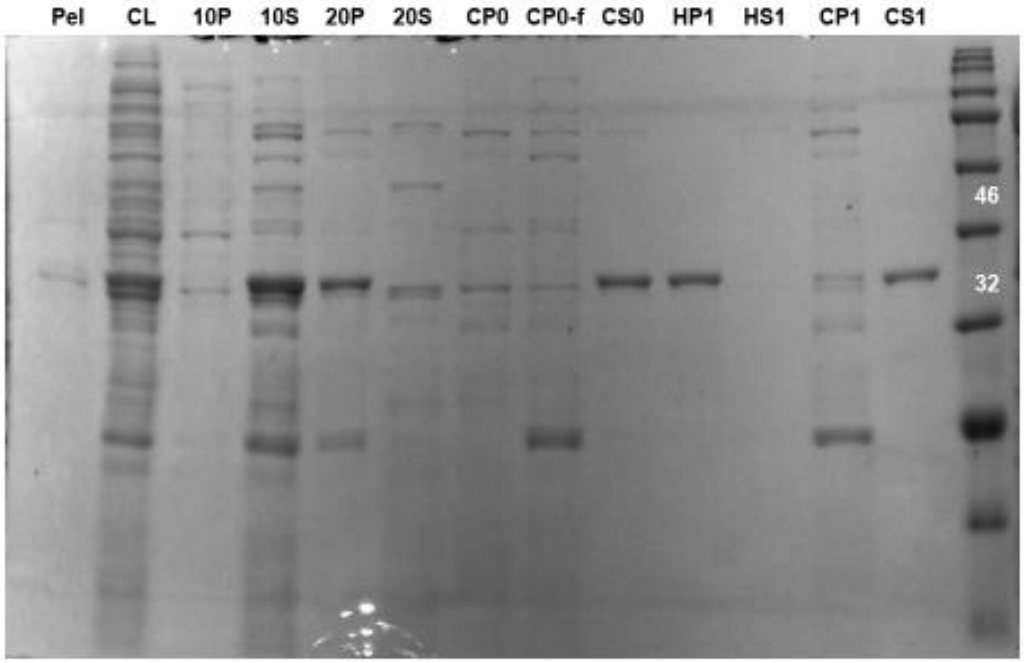


Figure C. 20. Protein purification using inverse temperature cycling (ITC) with YC12. ITC was carried out in PBS with 200 mM BME. Cold cycles were performed at a concentration of 50 mg pellet/mL at 4 °C. Hot cycles were performed at a concentration of 50 mg pellet/mL, coacervation was induced by 2 M NaCl at RT, and protein solutions were centrifuged at 30 °C. CL: cleared lysate; Pel: cell pellet; 10S and 10P: 10 wt% supernatant and pellet; 20S and 20P: 20 wt% supernatant and pellet; CP and CS: cold cycle pellet and supernatant; HP and HS: hot cycle pellet and supernatant.

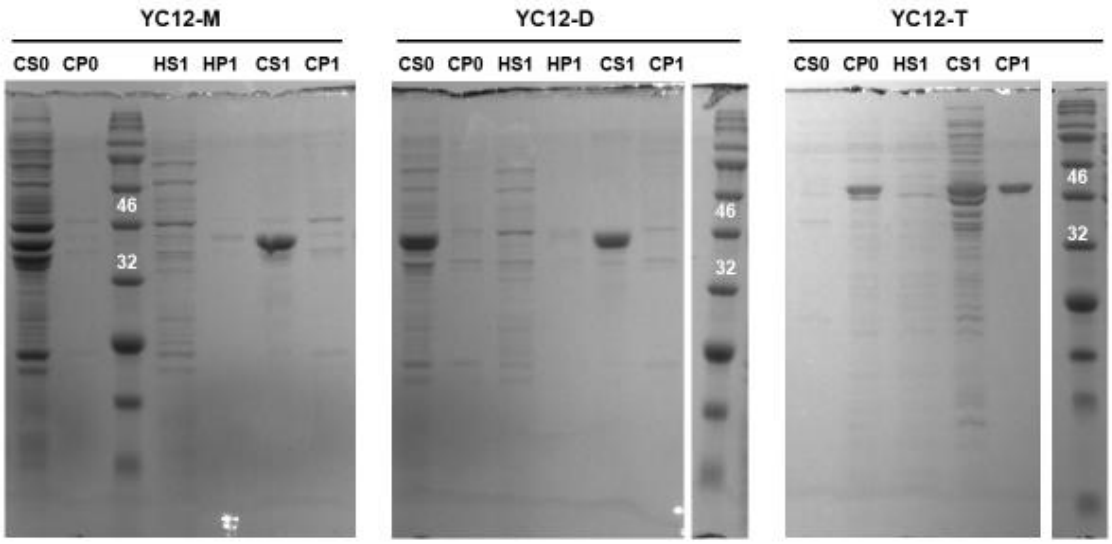


Figure C. 21. Protein purification using inverse temperature cycling (ITC) with YC12-M, -D, and -T. ITC was carried out in PBS with 2 M urea and 200 mM BME. A final concentration of 2 M urea in PBS aided in resolubilizing protein pellets. Vigorous mixing was necessary to solubilize YC12-T. Cold cycles were performed at a concentration of 50 mg pellet/mL at 4 °C. Hot cycles were performed at a concentration of 50 mg pellet/mL, coacervation was induced by 2 M NaCl at RT, and protein solutions were centrifuged at 30 °C. CP and CS: cold cycle pellet and supernatant; HP and HS: hot cycle pellet and supernatant.

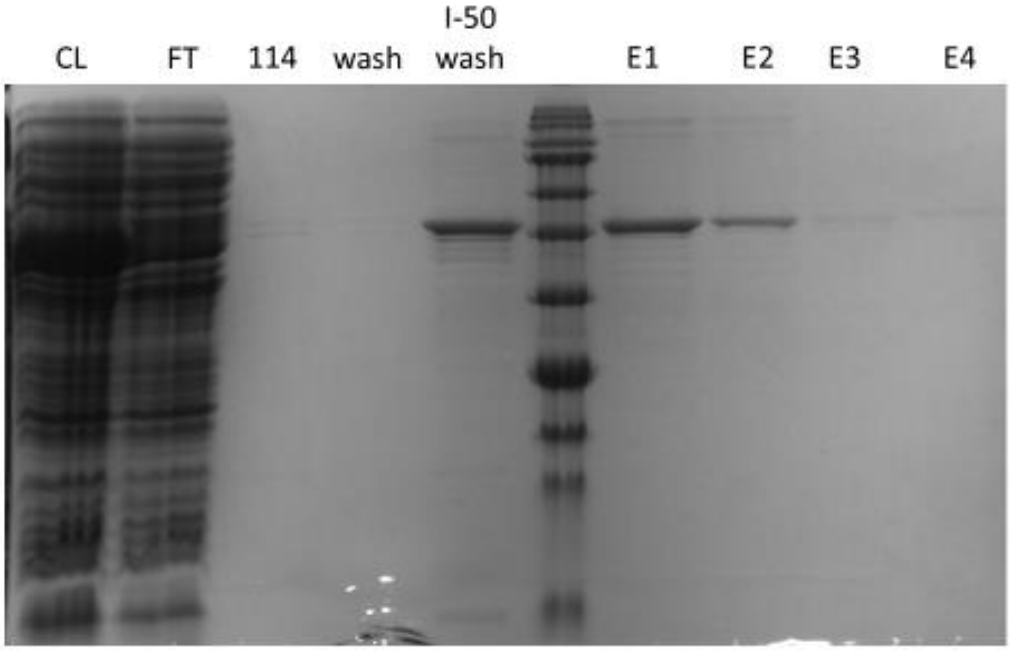


Figure C. 22. Endotoxin purification by Ni column with Triton X114 wash (Ni-114 column). Compared to Ni-IPA method, the residual endotoxin level was higher with Ni-114 column. The protein-loaded column was washed with 50 times of the bed volume with Triton X-114 wash buffer (100 mM KCl, 200 NaCl, 10 mM imidazole, 20 mM Tris-HCl, and 0.1 (v/v)% Triton X-114, pH 8.0). Next, the column was washed with 20 times of the bed volume with wash buffer to remove residual Triton X-114. The column was washed with 5 times of the bed volume with wash buffer containing 25-50 mM imidazole. Finally, the column was eluted with wash buffer containing 500 mM imidazole and 1 mM TCEP.

Endotoxin Assay – Ni-114

- Protein purified with Ni-114 column
- About 100-1000 times higher endotoxin then the target value

Group	Result	Note
114 – no LAL	-	
114 – 1: 100	-	Solution was turbid
114 – $1:10^3$	+	
114 – $1:10^4$	+	> 12.5 EU/mg protein
114 – $1:10^5$	-	< 125 EU/mg protein
114 – $1:10^6$	-	

Figure C. 23. Gel clot LAL assay of Ni-114 column purified YC12-T. The endotoxin level was between 12.5-125 EU/mg protein which was at least 100-1000 times higher than the same protein purified by the Ni-IPA method. A solid gel was not observed at 1:100 dilution possibly because of the interference from the residual Triton-X114 in the purified protein.

Group	Result	Note
Positive	+	Successful setup of the test
Negative	-	No contamination during setup
TCEP – 5 mM	-	Interferes with the test Solution was turbid and viscous, but flowed upon inverting
TCEP – 0.5 mM	+	No interference
TCEP – 0.05 mM	+	
TCEP – 0.005 mM	- (?)	Similar to the TCEP-5 mM group

Group	Result	Note
ITC – no LAL	-	
ITC – 1: 100	+	
ITC – $1:10^3$	+	
ITC – $1:10^4$	+	
ITC – $1:10^5$	+	
ITC – $1:10^6$	+	> 1250 EU/mg protein

Figure C. 24. Gel clot LAL assay control groups. Top: Testing whether TCEP negatively interacted with the assay. TCEP between 0.05-0.5 mM did not significantly interfer with the assay. Bottom: Test result of YC12-T purified by ITC method. The endotoxin content was >1250 EU/mg protein.

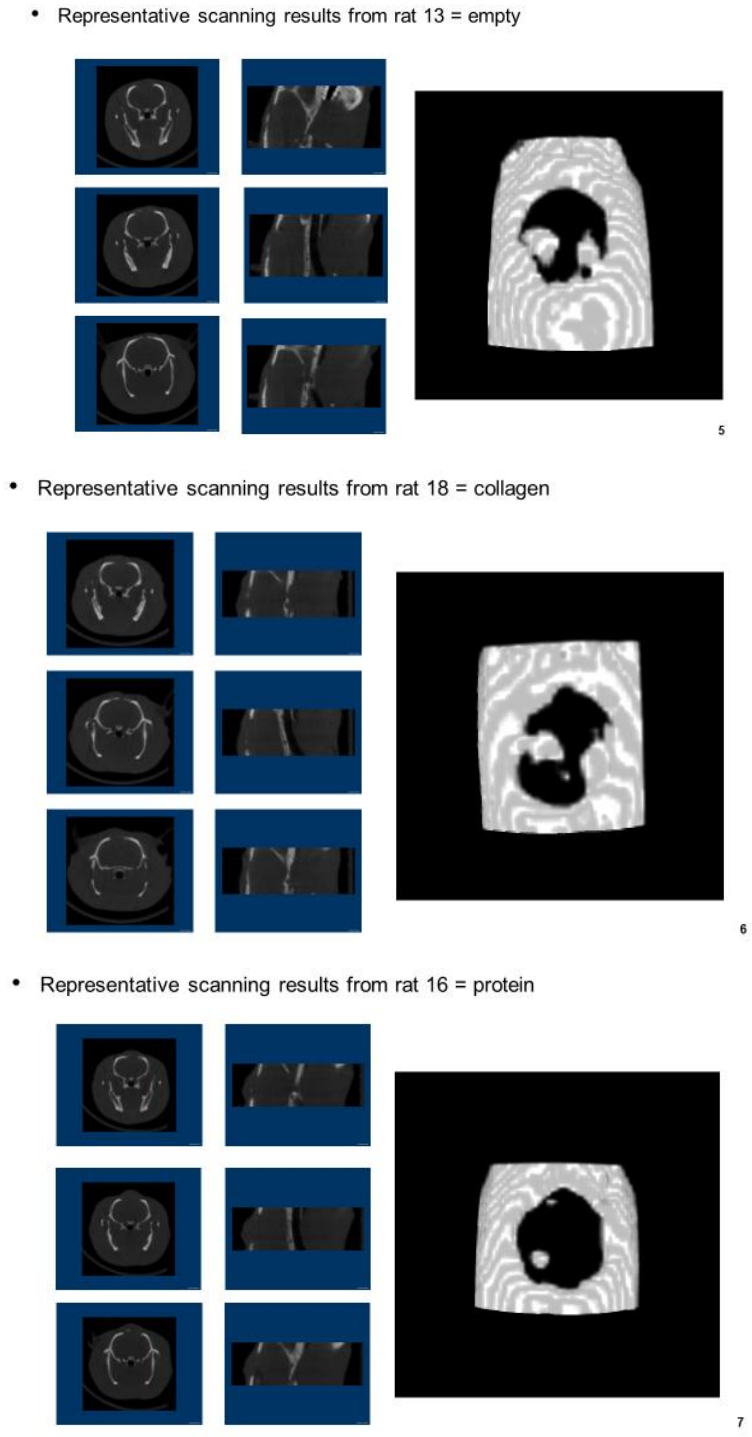


Figure C. 25. Representative scanning images of *in vivo* CT at 6 weeks after implantation. Empty defects showed the most healing followed by defects with cell-laden collagen sponges. Defects with cell-encapsulated protein gels showed the least extent of healing.

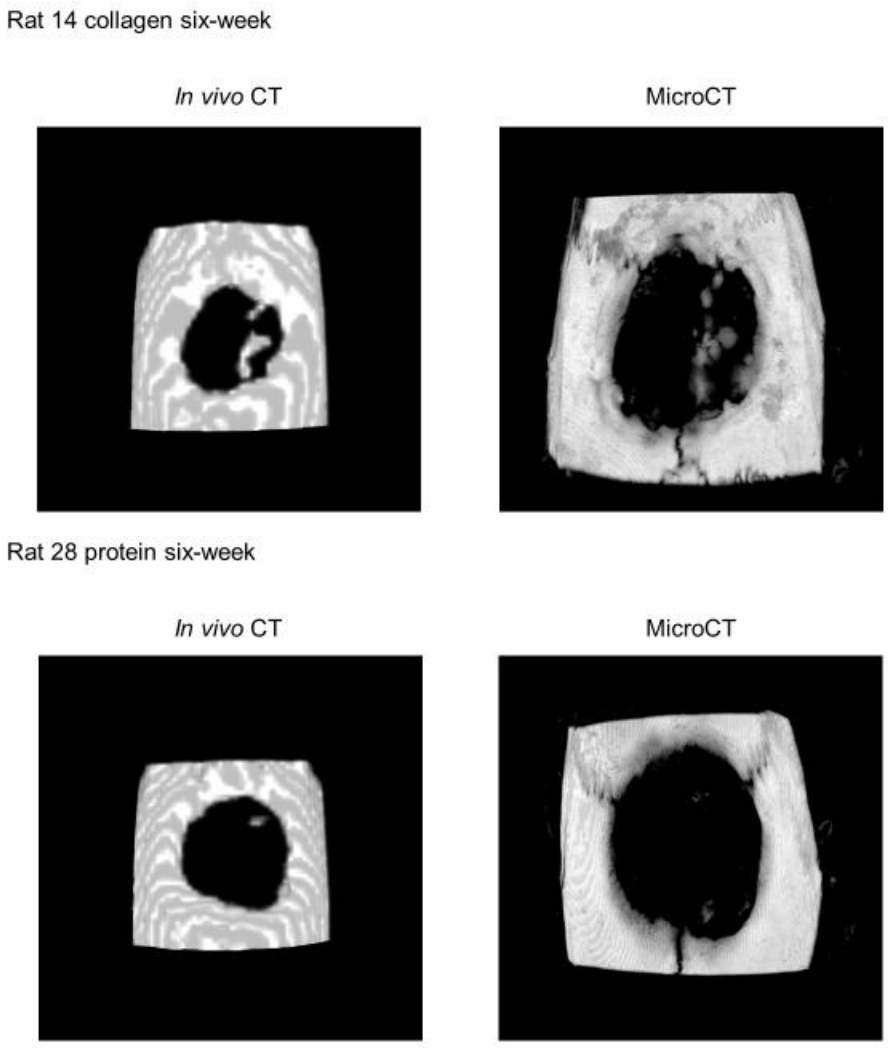


Figure C. 26. Scanning resolution comparison between *in vivo* CT and *ex vivo* μCT. Top: defect with cell-laden collagen sponge at 6th week; bottom: defect with cell-encapsulated protein gel at 6th week. *Ex vivo* μCT had a much higher resolution on resolving small volume of new bone formation. Both Rat 14 and 28 died after *in vivo* CT at 6th week. Defect sites were harvested immediately and scanned by *ex vivo* MicroCT.

REFERENCE

(1) Spicer, P. P.; Kretlow, J. D.; Young, S.; Jansen, J. A.; Kasper, F. K.; Mikos, A. G. Evaluation of bone regeneration using the rat critical size calvarial defect. *Nat. Protoc.* **2012,** *7*, 1918.

APPENDIX D. RECOMBINANT PROTEIN-BASED EPIGENETIC PROBES FOR DNA METHYLATION

PROJECT BACKGROUND

The goals of this project were to 1) develop a recombinant protein probe composed of a transcription activator-like effector (TALE) domain and a methyl-CpG-binding (MBD) domain for *in vitro* sequence-specific methylation detection and 2) develop a system for single-cell sequence-specific methylation detection using TALE, MBD, and bimolecular fluorescence complementation (BiFC).

Despite the fact that all cells in an organism have the same genome, different cell types exist with different gene activities. Recent progress in molecular biology has demonstrated that gene expression is regulated not only by the DNA sequence of a gene, but also by the epigenetic modifications to the gene.[1] While histone modifications alter the chromatin structure and thus change the expression level of a gene, DNA methylation directly changes the DNA nucleotides and hence regulates gene expression. In pluripotent cells found during development, genes that are necessary for the later stage of development can be temporarily deactivated by histone modifications.[2] As cells commit to a certain lineage, or become somatic cells, genes that are related to pluripotency are silenced by DNA CpG methylation. Specifically, a methyl group is transferred to the C5 position of cytosine when it is found in a CpG dinucleotide-site, in which cytosine is followed by guanine. When the promoter of a gene is hypermethylated, its activity is downregulated and vice versa.[3]

The genome methylation state in a cell population is related to development stages. It has been shown that genes related to self-renewal are hypomethylated in embryonic stem cells.[4] On the other hand, genes regulating stimuli responses and signaling transduction pathways are hypermethylated. For example, as embryonic stem (ES) cells differentiate, stem cell marker genes become methylated.[5] When human mesenchymal stem cells (MSCs) are expanded *in vitro*, changes in DNA methylation patterns have been observed as well. A study revealed that genes related to apoptosis and replicative senescence are hypermethylated in early passages.[6] On the other hand, genes regulating cell cycle and DNA replication are hypermethylated in late passages. This result matches the observation that MSC proliferation rate decreases or stops in late passages. Genes related to lineage

differentiation also become hypermethylated in late passages. Bork *et al.* reported that the homeobox gene DLX5 becomes hypermethylated in later passages of human MSCs.[7] DLX5 was previously shown to be able to transactivate gene expression of RUNX2, which is the master regulatory gene of osteogenesis.[8] The authors suggested that the hypermethylation of DLX5 in later passages of human MSCs could result in less transactivation of RUNX2 and further result in lower potential for osteogenesis. These results suggest that the capability of stem cells to proliferate and differentiate are related to methylation levels of specific genes and that information on the methylation states can be valuable when controlling stem cell differentiation in tissue engineering applications.

Current methods to detect DNA methylation patterns are based on sodium bisulfite conversion. The principle behind this method is that bisulfite will convert unmethylated cytosine to uracil but will leave methylated cytosine unchanged. The newly converted uracil can be detected by polymerase chain reaction (PCR). PCR-amplified fragments can be sequenced by conventional sequencing methods. By comparing this sequence to the unconverted sequence, unmethylated cytosine can be detected by observing a change from cytosine to thymine.

Different DNA methylation detection methods have been developed. Methylation-specific PCR (MSP) uses primers that overlap with methylation sites.[9] These primers can be designed such that they will only anneal to a converted methylated sequence or unmethylated sequence. The methylation status can be distinguished by PCR amplification results.[9] Another method, which is called MethyLight, incorporates a fluorescently labeled probe into MSP.[10] Reduced representation bisulfite sequencing (RRBS) uses the CpG methylation-sensitive restriction enzyme MspI to fragment the target sequence.[11] In all the above methods, bisulfite conversion is still an essential step. Another approach to methylation detection involves enrichment of methylated fragments followed by high-throughput sequencing. Enrichment can be achieved by binding of methyl-CpG-binding domain protein 2 (MBD2 protein) to the methylated CpG sites (MBD-Seq).[12] The methylation level can be determined by the frequency of a sequence being captured by MBD2 proteins. However, one thing in common for both bisulfite methods and enrichment methods is that manipulation of DNA samples is necessary, and hence these methods are both unfeasible for real-time methylation detection in living cells. In addition, it is difficult

to determine methylation states in a single cell with both methods. However, methylation information at single-cell level could be valuable for applications with MSCs because the MSC population is heterogeneous and the methylation states may not be the same for every cell in the same population. Therefore, developing a molecular tool that allows us to perform real-time DNA methylation detection in a single living cell is critical and could be helpful for tissue engineering applications with stem cells.

Methyl-CpG-binding domains (MBDs) are protein domains of ~70 amino acids, and they specifically bind to symmetrically methylated CpG sites.[13] The unique α/β-sandwich structure of MBDs allows binding to a single methylated CpG site.[14] There are five human MBD proteins: MeCP2, MBD1, MBD2, MBD3, and MBD4. Among them, MBD1, MBD2, and MBD4 are able to bind to methylated CpG sites *in vitro*.[15] On the other hand, MBD3 did not show any specific binding to methylated CpG sites *in vitro*.

Because of their specific binding to methylated CpG sites, MBDs have been utilized as a tool to detect DNA methylation. One example is the MBD2 protein used in MBD-Seq. Another widely used MBD protein is MBD1 because of its high binding affinity. The Bird group evaluated the binding ability of recombinant proteins containing one or four repeats of an MBD1-derived MBD sequence.[16] An electrophoretic mobility shift assay (EMSA) was used to determine the dissociation constant (K_d) of one MBD to be 30 μM and of four MBD domains to be decreased to 0.5 μM. The binding ability was also increased when more methylated CpG sites were present on target DNA. With three methylated CpG sites, the K_d values were lowered to 2 μM and 0.02 μM for one MBD and four MBD domains, respectively. The Yuan group developed a one-pot methylation detection platform with a fluorescently labeled probe with one repeat of a mutated MBD1 domain coupled with fluorescence correlation spectroscopy (FCS).[17] The limit of detection (LOD) of the platform was shown to be 20 nM of target DNA for methylation levels ranging from 5% to 100%.

While MBD domains have not been reported to have binding preferences for sequences surrounding methylated CpG sites, sequence preference has been introduced into MBD-based methylation detection platforms by combining DNA binding domains with domains that recognize specific sequences (e.g., zinc fingers).[18] Badran and coworkers developed a platform composed of an MBD1 domain, luciferase, and a zinc finger.[19] In

this platform, the N-terminal fragment of luciferase was attached to the MBD1 domain, and the C-terminal fragment was attached to the zinc finger. When methylated CpG sites were in close proximity to the DNA sequence recognized by the zinc finger, binding of both the MBD1 domain and the zinc finger occurred and thus brought the N-terminal fragment and the C-terminal fragments of luciferase together. As a result, luciferase regained its enzymatic function and oxidized its substrate luciferin, and luminescence could be detected. The LOD of this platform was 2 nM; however, the performance was affected by the spacing between the CpG sites. The luminescence signal decreased to 60% when two CpG sites were 21 base pairs (bp) apart compared to 6 bp apart.

In light of these results, we proposed a one-piece recombinant probe containing both a methylation detection domain (MBD1) and a sequence-specific DNA-binding domain (a transcription activator-like effector (TALE) domain) for site-specific DNA methylation detection. In addition, green fluorescent protein (GFP) will be fused to the MBD1 domain to allow for analysis via fluorescence spectroscopy. Unlike platforms with two components, the one-piece design will simplify the preparation of the platform. The choice of TALE also allows us to change our targeting site easily due to the modularity of the TALE assembly. TALEs are sequence-specific DNA binding proteins found in the bacterial genus *Xanthomonas* as a tool to activate certain host plant promoters. TALEs are composed of a number of repeats consisting of 34 amino acids. In each repeat, the sequence is almost identical except the 12th and 13th residues, which are referred to as repeat variable di-residues (RVDs). It is these two residues that confer the nucleotide-specificity to the whole 34-amino acid repeat and thus enables the sequence-specificity of TALEs. Different RVDs have been identified along with their nucleotide specificity, and various TALEs targeting different DNA sequences have been constructed.[20-22] Commercialized TALE assembly kits, such as Golden Gate TALEN and TAL Effector Kit 2.0 (Addgene), are also available to allow customized TALE constructs.[23] As a result, the targeting site for methylation detection can be changed easily within our TALE-based one-piece probe.

We also hypothesized that the performance of site-specific detection will be further improved with the one-piece probe. In platforms where the methylation detection and sequence-specific DNA-binding components are separated, binding of both components is required for positive detection. Moreover, sequence-specific DNA binding domains such

as zinc fingers or TALEs usually have very low K_d values (nM range for zinc fingers[24] and sub μM to nM for TALE[25]). On the other hand, MBD1 has a K_d value in the μM range, and as a consequence, the binding of the methylation detection component is not as effective as that of the DNA-binding component. This difference in K_d could result in a high false negative detection rate for the two-component platforms. Also, even when multiple methylated CpG sites decreased the K_d value of MBD1 and thus increased binding affinity, the detection sensitivity decreased. With our one-piece probe, the methylation detection domain MBD1 will be enriched near methylated CpG sites close to the targeting site because of the anchoring effect from TALE, the sequence-specific domain. This local enrichment will increase the effective concentration of MBD1 in our probe and thus lower the effective K_d and increase sensitivity in methylation detection.

RESULTS AND DISCUSSION

The one-piece probe is composed of an N-terminus 3xFLAG tag followed by a TALE domain, a HRV3C-sensitive sequence, an MBD domain, and a C-terminus His tag. The TALE targeting sequence (TGGAGCCTTCGGCTGAC) is derived from the human CDKN2A gene (26^{th} to 42^{nd} bp). Two TALEs were assembled: TALE13 (targeting CDKN2A 26^{th}-38^{th} bp) and TALE17 (targeting CDKN2A 26^{th}-42^{nd} bp) in the vector PPM (courtesy of Dr. Yuan's group, modified from Addgene Plasmid #40786). The sequences of both constructs, PPM-TALE13 (hereafter PPM-13) and PPM-TALE17 (hereafter PPM-17) were verified by DNA sequencing. Expression of PPM-13 in four different *E. coli* expression hosts (i.e., BL21(DE3), BL21(DE3)pLysS, Rosetta2(DE3)pLysS, and BL212(DE3)-RIPL) were tested, and expression by IPTG induction was confirmed by SDS-PAGE and anti-His Western blots (Supporting Information Figure D. 5). There was no detectable expression in the BL21(DE3)pLysS host, and the highest expression level was achieved in BL21(DE3). Therefore, we expressed PPM-17 in the BL21(DE3) host and found even higher expression levels with PPM-17 (Supporting Information Figure D. 6). Based on its higher expression levels and the expectation that PPM-17 would have a higher TALE binding affinity, we decided to proceed with PPM-17. PPM-17 was purified using a Ni column. The final purity was found to be 91%, as shown in Supporting Information Figure D. 7.

A 57-bp DNA oligo was prepared. It had a TALE binding site and a CpG site as shown in Appendix F. 37. Methylation of the CpG site on the 57-bp oligo (57) was performed by methyltransferase M.SssI (New England Biolabs (NEB), MA) following the recommended protocol. The methylation level of the methylated 57-bp oligo (m57) was assessed by digestion with a CpG methylation-sensitive restriction enzyme, HpaII (NEB). The digestion showed 93% methylation of m57 (Supporting Information Figure D. 8).

The *in vitro* binding ability of PPM-17 was tested with 57, m57, and a negative control DNA sequence (with no TALE binding sites or methylated CpG sites). The EMSA result (Figure D. 1) showed that there was a high band at 400 bp, and this result indicated that the specific binding of PPM-17 to the 57 oligo increased as the PPM-17:DNA molar ratio increased from 1 to 3. There was also a second high band at 800 bp for the PPM-17: DNA ratio of 3; however, this band had a much lower intensity. This second high band could result from non-specific binding of a second PPM-17 protein. When the ratio was further increased to 10, the second high band intensified, and a third high band appeared at 1 kb. PPM-17 binding to the m57 oligo had a similar pattern. However, all bands of bound complexes smeared significantly with the m57 oligo. The smearing with m57 could be a result of MBD interactions with the methylated CpG site. Also, there was a decrease in intensity of the specific binding band at 400 bp at a ratio 10. Because MBD has a much lower K_d than TALE, it is possible that PPM-17 binding to the m57 oligo had a mixture of complexes with bound and unbound MBD, and this mixture showed up as a smeared band on the gel. As for the binding with the negative control DNA A1, there was no high band from bound complexes observed with ratios of 1 and 3; only at a ratio of 10 was there a high band around 760 bp. There were faint high bands around 1.5 kb with ratios of 0, 1, and 3; however, since A1 DNA was prepared from digestion of plasmid DNA, it is possible that those high bands resulted from plasmid DNA contamination. This possibility was also supported by the appearance of the high band at a ratio of 0. In summary, PPM-17 showed specific *in vitro* TALE binding with potential MBD binding of the methylated CpG site. However, further tests are necessary to evaluate MBD binding within PPM-17.

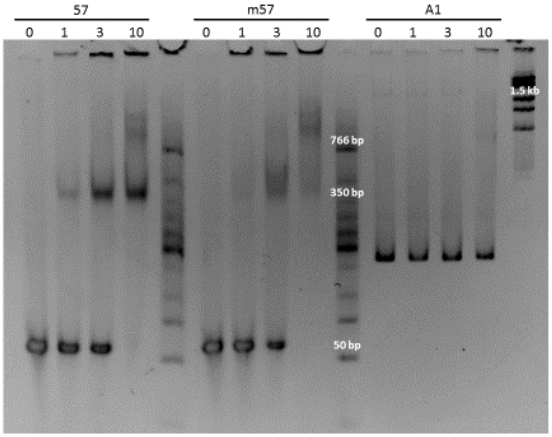


Figure D. 1 PAGE gel of PPM-17 with 57, m57, and negative control A1 DNA oligonucleotides. Numbers above the image indicate the molar ratio of PPM-17:DNA.

We also redesigned the DNA oligo used in the binding assay. There were two issues with the 57-bp oligo. First, the short length of the oligo makes it difficult to stain and visualize by ethidium bromide (EtBr), and as a result, higher DNA concentrations had to be used in EMSA. Second, there was an intrinsic CpG site in the TALE binding sequence derived from the 26th-42nd bp of CDKN2A. To accommodate this methylated cytosine, the RVD at that position was NG (recognizes thymine) instead of HD (recognizes cytosine) based on a previous study showing that NG recognizes methylated cytosine better than HD.[24] However, as a result, the TALE binding specificity is not the same for the 57 and m57 oligos because NG preferentially binds to m57. To address these issues, the redesigned DNA oligo was extended to 91 bp, and the intrinsic CpG site in the TALE binding sequence was removed by changing the cytosine to thymine. We also added three CpG sites in the redesigned DNA oligo (hereafter 91-bp oligo) to increase MBD binding affinity. The two extra CpG sites were centered around the original one. We also used a different methylation approach to prepare the methylated 91 (m91)-bp oligos by co-transforming the plasmid pAIT2 encoding M.SssI (courtesy of Dr. Yuan's group) with the plasmid containing the 91-bp oligo sequence. The methylation level of the m91-bp oligo achieved in this approach was found to be ~90% by densitometry (Supporting Information Figure D. 9). The 91- and m91-bp oligos can be used for future assays.

MBD binding can be visualized by digestion of PPM-17 with HRV3C protease (Thermo Fisher, MA), as illustrated in Supporting Information Figure D. 10. In short,

unbound MBD will not be bound to DNA after HRV3C digestion. On the other hand, MBD bound to methylated CpG sites should remain attached to DNA after digestion. We found that HRV3C activity was lower in the binding buffer used for EMSA with the positive control protein (Supporting Information Figure D. 11). Because we found no binding occurred with PPM-17 in the recommended HRV3C digestion buffer (Supporting Information Figure D. 12), we decided to optimize HRV3C digestion in the binding buffer. The following factors were varied: HRV3C to protein ratio (Supporting Information Figure D. 13 and Figure D. 14), digestion temperature (Supporting Information Figure D. 15 and Figure D. 16), and the competitor polymer (dIdC) concentration (Supporting Information Figure D. 17). A higher HRV3C: PPM-17 ratio increased the digestion efficiency. On the other hand, dIdC concentration and digestion temperature did not have significant effects on digestion efficiency. The optimal condition was determined to be 1U: 0.62 μg HRV3C:PPM-17 with 50 ng/μL dIdC at room temperature for one hour. The digestion efficiency with this condition was ~70%. Because of the moderate digestion efficiency, we decided to redesign the probe by adding a GFP after the MBD domain. The fluorescence from GFP will allow us to track MBD cleavage with higher sensitivity and thus compensate for the lower digestion efficiency with HRV3C.

The above results suggested that TALE and MBD had the potential to be used for DNA methylation detection in live cells. The detection was visualized using bimolecular fluorescence complementation (BiFC). A TALE domain (R20) (Addgene plasmid #49637) targeting a telomere repeating sequence (5’- AACCCTAACCCTAACCCTA-3’) or a dimeric MBD (dMBD) domain were fused with BiFC domains of mVenus fragments (Addgene plasmid #22010 for VN173 fragment and #22011 for VC155 fragment). For all constructs, the DNA binding domains were always on the N-terminus, and the BiFC domains were always on the C-terminus. Negative controls for sequence specificity were created by deleting the sequence-binding region in the TALE domain. The constructs were transfected into human embryonic kidney cells 293 T (HEK293T) using Lipofectamine 3000 (Invitrogen).

Table D. 1 List of constructs based on the BiFC pair VN173/VC155

Construct name	DNA-binding domain	BiFC domain
BG-173	None	VN173
BG-155	None	VC155
dMBD-173	Dimeric MBD	VN173
R20-155	Telomere-binding TALE	VC155

As shown in Figure D. 2, cells transfected with negative control constructs showed no specific punctate methylation binding-pattern but non-specific protein aggregation in the nuclei, and these results suggested that the negative control constructs had no sequence-specific binding. But the protein concentrations of the transfected constructs were high enough to induce complementation in the nuclei. To verify the binding of the DNA-binding domains fused to the BiFC domains, DNA-binding constructs (dMBD-173 and R20-155) were co-transfected with the corresponding negative control constructs. In both co-transfection (dMBD-173/BG-155 and BG-173/R20-155), punctate patterns were observed in the nuclei. These punctate patterns suggested that the dMBD and R20 domains bound to the target DNA sequence properly but non-specific BiFC complementation occurred. We also co-transfected dMBD-173/R20-155 with a positive control construct R20-mCherry to verify the sequence specificity of the BiFC signal from dMBD-173/R20-155. More punctate patterns were observed from dMBD-173/R20-155 (green) with some overlap with the pattern from R20-mCherry (red). The BiFC signals that did not overlap with R20-mCherry fluorescence were likely the results of non-specific BiFC complementation from off-target binding of our constructs (dMBD-173 and R20-155) or excess dMBD-173 and R20-155 proteins in the nucleus space. The results shown here suggested higher specificity of BiFC complementation of our constructs was necessary to ensure the observed signals were from methylation near the target DNA sequence.

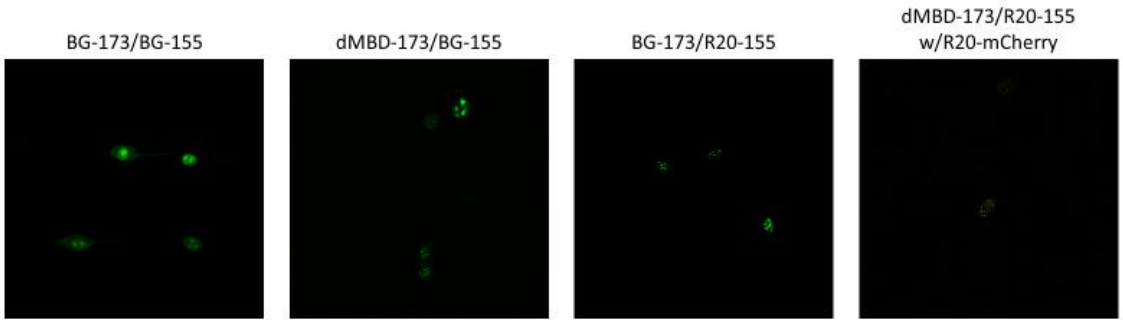


Figure D. 2. HEK293T cells transfected with constructs based on BiFC pair VN173/VC155. Cells were imaged 24 h after transfection. The green signal indicates complementation from the BiFC domains on the constructs. R20-mCherry (Addgene #49637) was used as a positive control for telomere binding (red signal).

To improve the BiFC complementation specificity, we switched to a different pair of mVenus fragments (VN210 and VC210) that had been reported to have lower non-specific complementation. We also investigated the BiFC complementation efficiency by varying the linker length between the DNA-binding domains and the BiFC domains on our constructs.

Table D. 2. List of constructs based on the BiFC pair VN210/VC210

Construct name	DNA-binding domain	Linker length	BiFC domain
BG-VN210	None	2 Gly-Ser	VN210
BG-VC210	None	2 Gly-Ser	VC210
BG-l-VN210	None	12 Gly-Ser	VN210
BG-l-VC210	None	12 Gly-Ser	VC210
dMBD-VN210	Dimeric MBD	2 Gly-Ser	VN210
dMBD-VC210	Dimeric MBD	2 Gly-Ser	VC210
dMBD-l-VN210	Dimeric MBD	12 Gly-Ser	VN210
dMBD-l-VC210	Dimeric MBD	12 Gly-Ser	VC210
R20-VN210	Telomere-binding TALE	2 Gly-Ser	VN210
R20-VC210	Telomere-binding TALE	2 Gly-Ser	VC210
R20-l-VN210	Telomere-binding TALE	12 Gly-Ser	VN210
R20-l-VC210	Telomere-binding TALE	12 Gly-Ser	VC210

As shown in Figure D. 3, cells co-transfected with the negative controls and DNA binding constructs had significantly less non-specific complementation. This result suggested that the BiFC pair VN210/VC210 had a much lower non-specific complementation than the VN173/VC155 pair. Co-transfection of R20-VN210/dMBD-VC210 and dMBD-VN210/R20-VC210 did not result in the punctate pattern that was expected if both R20 and dMBD bound their targets. Since BiFC complementation is sensitive to distance, we varied the linker length between the BiFC domains and the R20 or dMBD domains.

The constructs with a long linker between the DNA-binding domains (dMBD and R20) and the BiFC domains showed low non-specific complementation (Figure D. 4) similar to that observed for the ones with a short linker. When R20-l-VN210 (or R20-l-VC210) and dMBD-l-VN210 (or R20-l-VC210) were co-transfected with R20-mCherry, no overlap between the BiFC signal (green) and the mCherry signal (red) was observed. Our results seemed to suggest that the DNA-binding domains (TALE, MBD) can remain functional when fused with the BiFC domains and the importance of choosing ae BiFC pair with low non-specific complementation. To use our system to visualize sequence-specific DNA methylation, the relative orientation and the distance between the DNA-binding domains and the BiFC domains need to be designed to reach a balance between suppressing the BiFC non-specific complementation and the desired fluorescence resulting from the simultaneous binding of both TALE and MBD DNA-binding domains. Lungu and coworkers reported success in visualizing sequence-specific DNA methylation and histone 3 lysine 9 trimethylation using a very similar BiFC approach with zinc fingers, TALE, and CRISPR/dCas9 recombinant domains.[26]

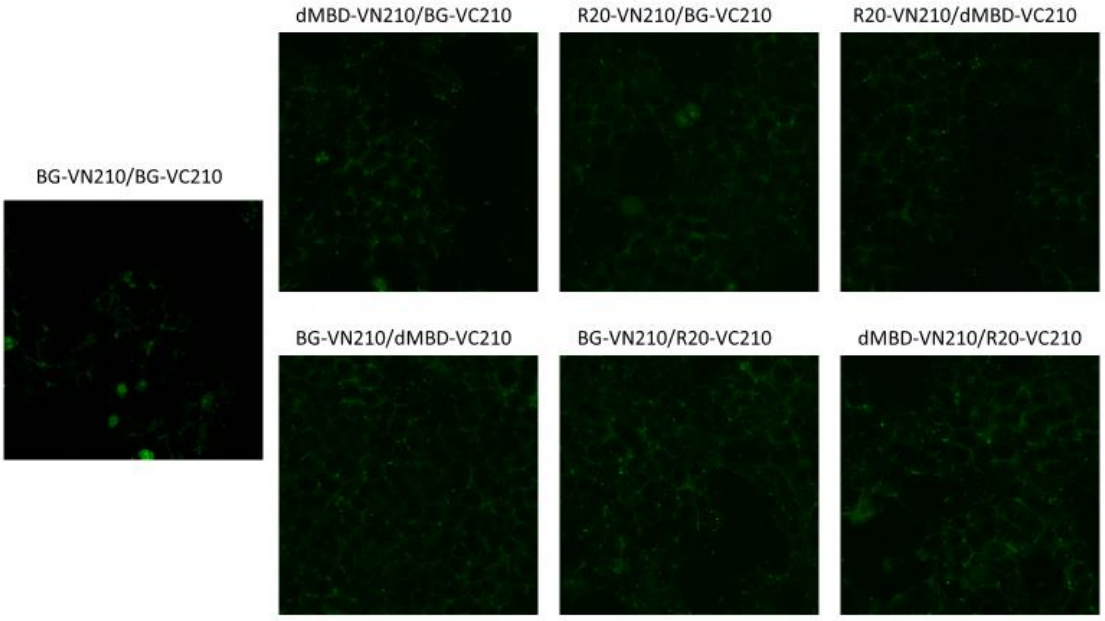


Figure D. 3. HEK293T cells transfected with constructs based on BiFC pair VN210/VC210 with the short linker. Cells were imaged 24 h after transfection.

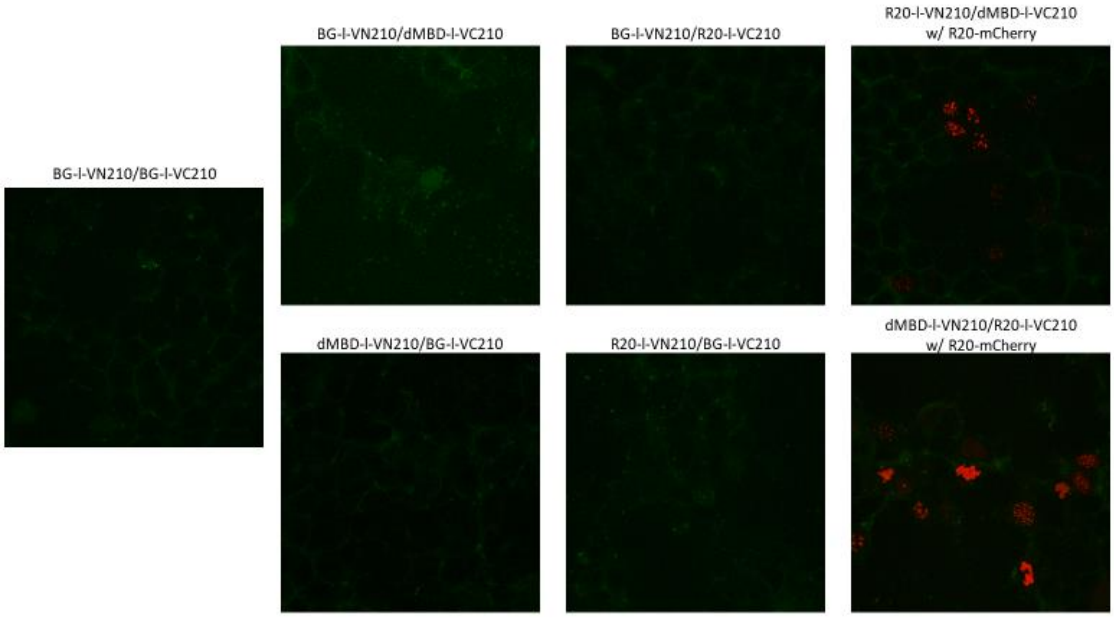


Figure D. 4. HEK293T cells transfected with constructs based on BiFC pair VN210/VC210 with the long linker. Cells were imaged 24 h after transfection. R20-mCherry was used as a positive control for telomere binding (red signal).

Supporting Information

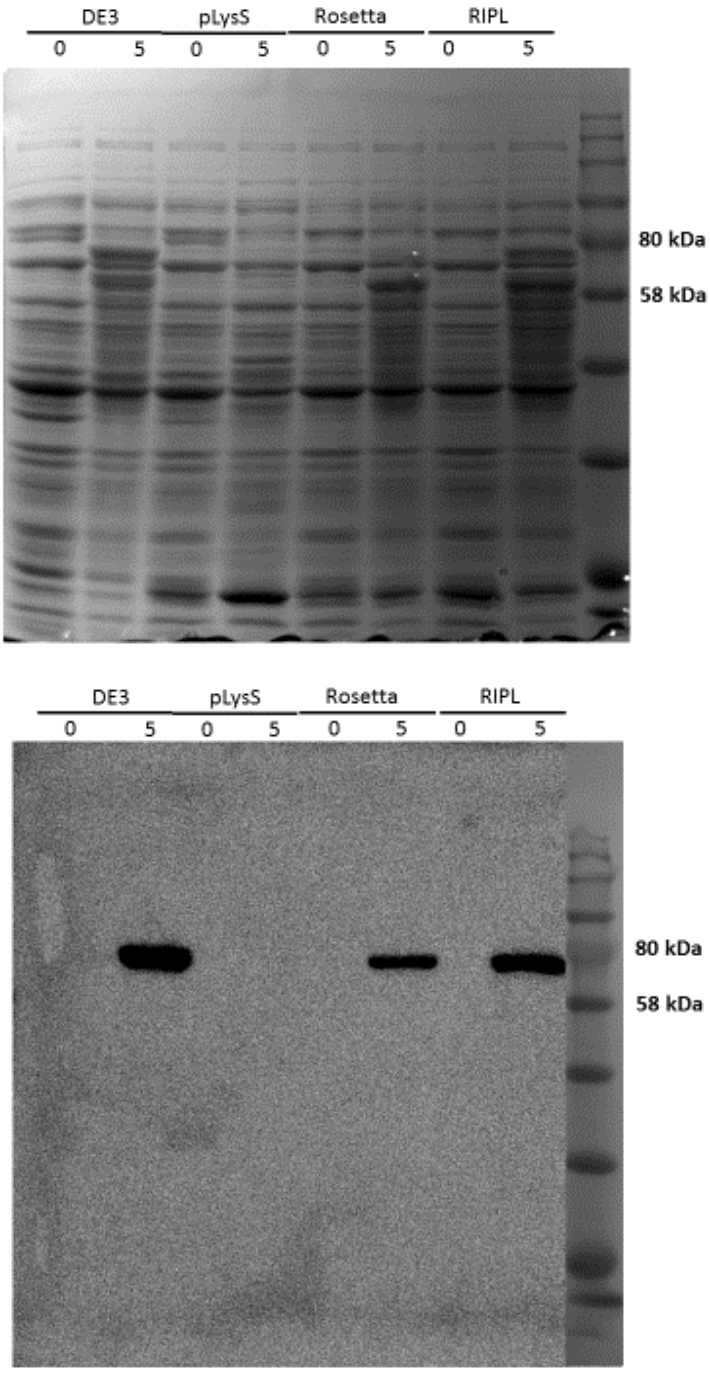


Figure D. 5. Expression of PPM-13 (MW 78.5 kDa). Top: SDS-PAGE gel showing PPM-13 being expressed 5 hours after induction with IPTG. Bottom: The expression was confirmed in three expression *E. coli* hosts except for BL21(DE3)pLysS by anti-His tag Western blot.

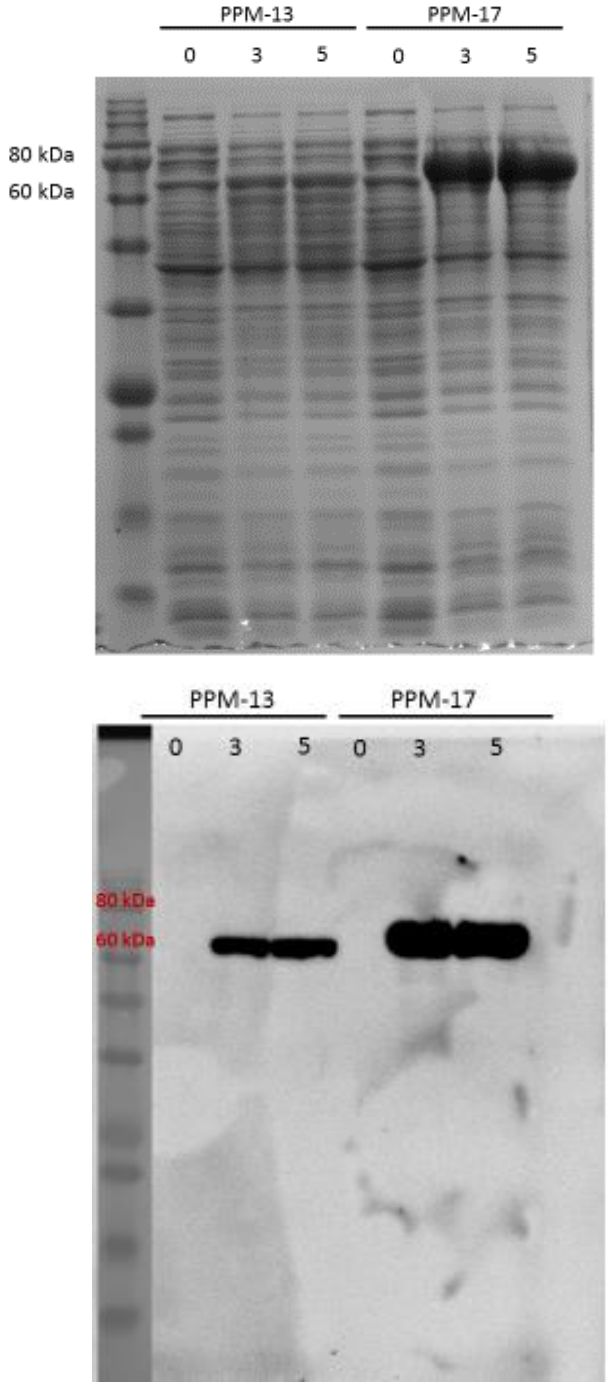


Figure D. 6. Expression of PPM-13 (MW 78.5 kDa) and PPM-17 (MW 95.6 kDa) in BL21(DE3). Top: SDS-PAGE gel showing PPM-17 being expressed 5 hours after induction with IPTG. Bottom: The expression was confirmed in expression *E. coli* host BL21(DE3) by anti-His tag Western blot.

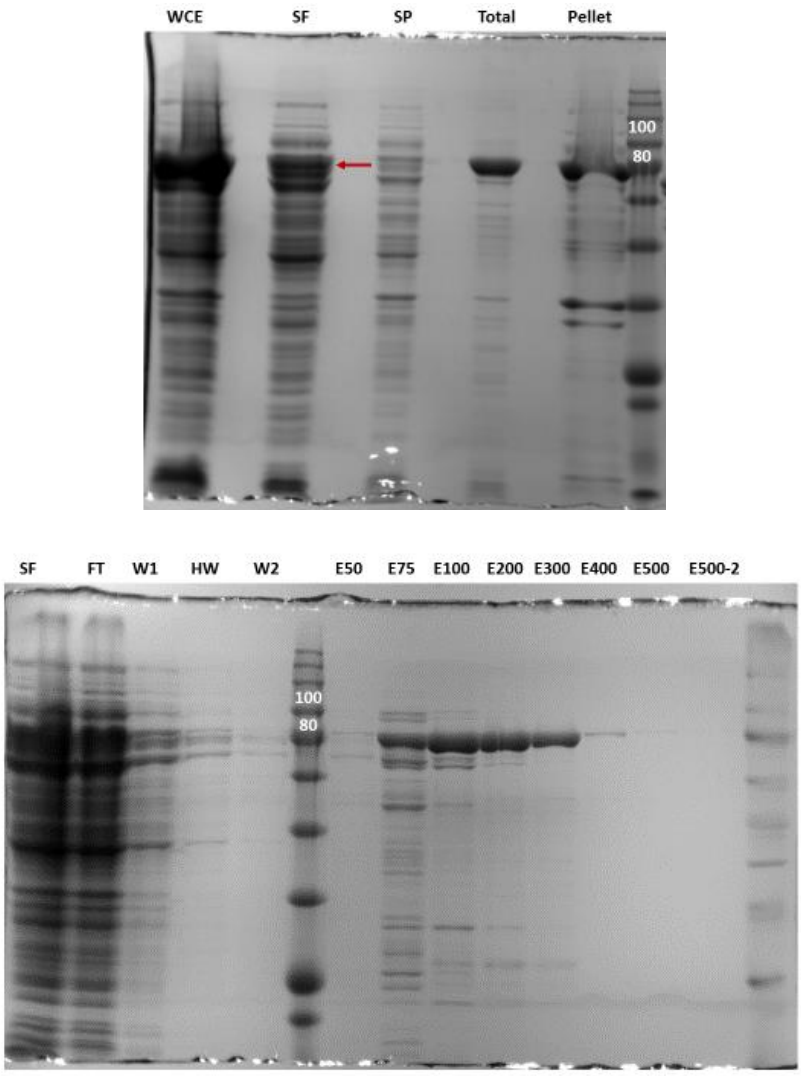


Figure D. 7. PPM-17 purification. Top: Fractionation of cell lysate based on solubility. Red arrow indicates PPM-17 band in soluble fraction (SF). For detailed purification steps, refer to Appendix E. 6. Bottom: SF of PPM-17 purified by a native Ni column. The purity of E300 was found to be 91% by densitometry. SF: soluble fraction; SP: supernatant of Triton wash; FT: flow through; W1 and W2: wash with 500 mM NaCl, 5 mM imidazole, 20 mM Tris-Cl, pH 7.9; HW: wash with 2M NaCl, 5 mM imidazole, 20 mM Tris-Cl, pH 7.9. E50 to E500-2: Elutes with 500 mM NaCl, 50-500 mM imidazole, 20 mM Tris-Cl, pH 7.9.

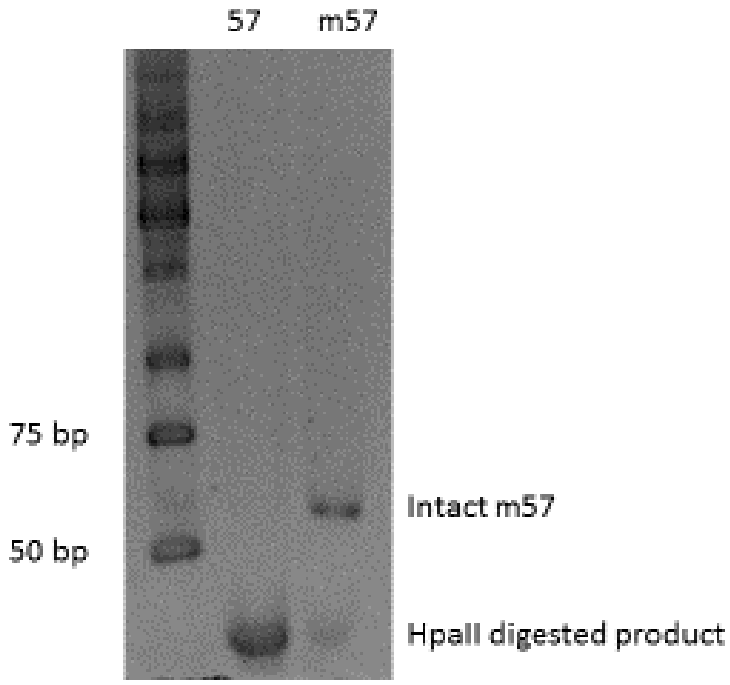


Figure D. 8. HpaII digestion of 57 and m57. The methylation level of m57 was determined to be 93% by densitometry.

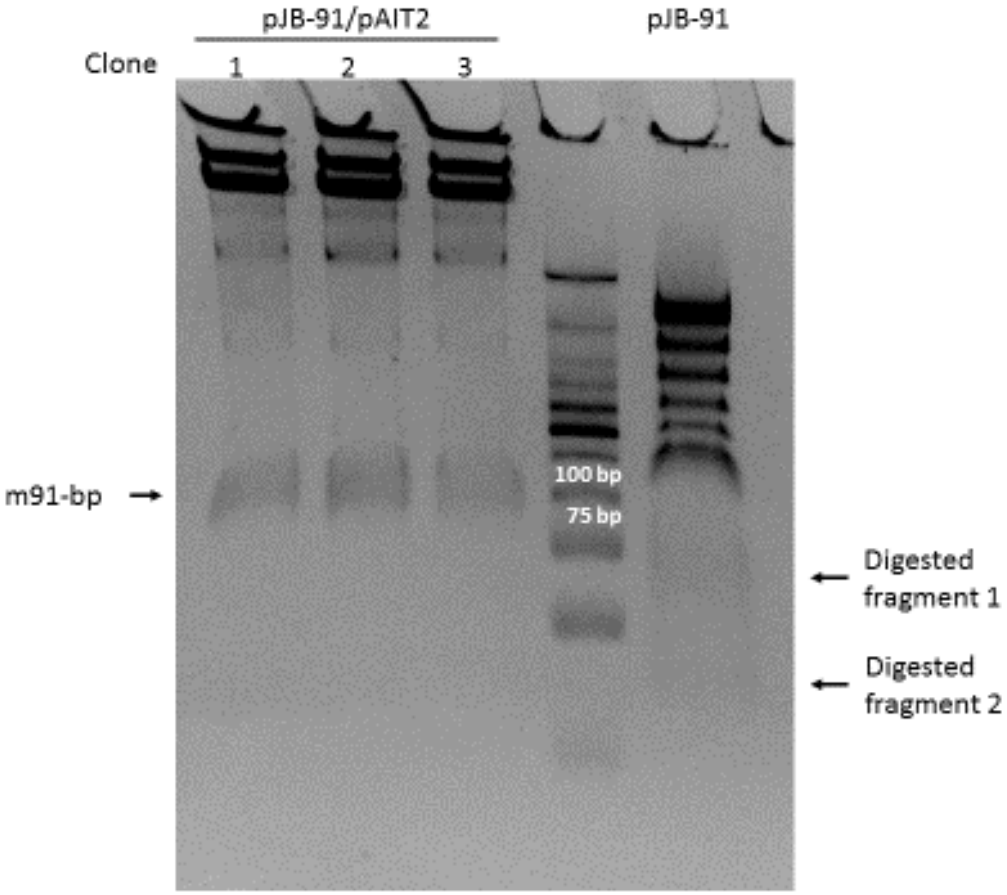


Figure D. 9. HpaII digestion of pJB-91 to check methylation level. Methylation of pJB-91 was accomplished by co-transformation with pAIT2, which encodes methyl transferase M.SssI. Similar methylation was achieved among three clones with pJB-91 and pAIT2.

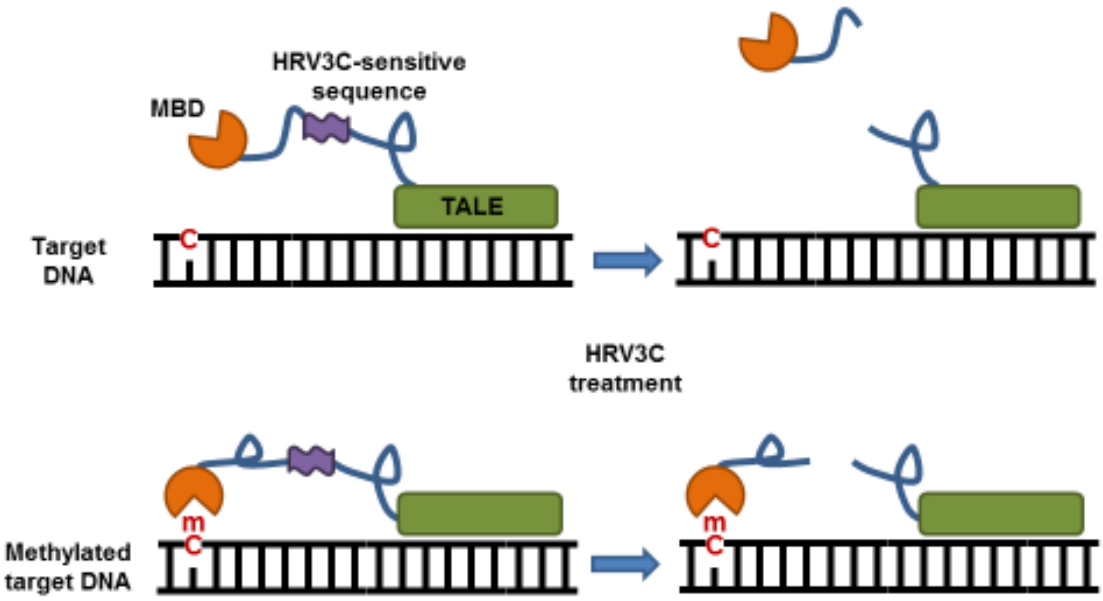


Figure D. 10. Illustration of methylation detection of the one-piece probe (PPM-17).

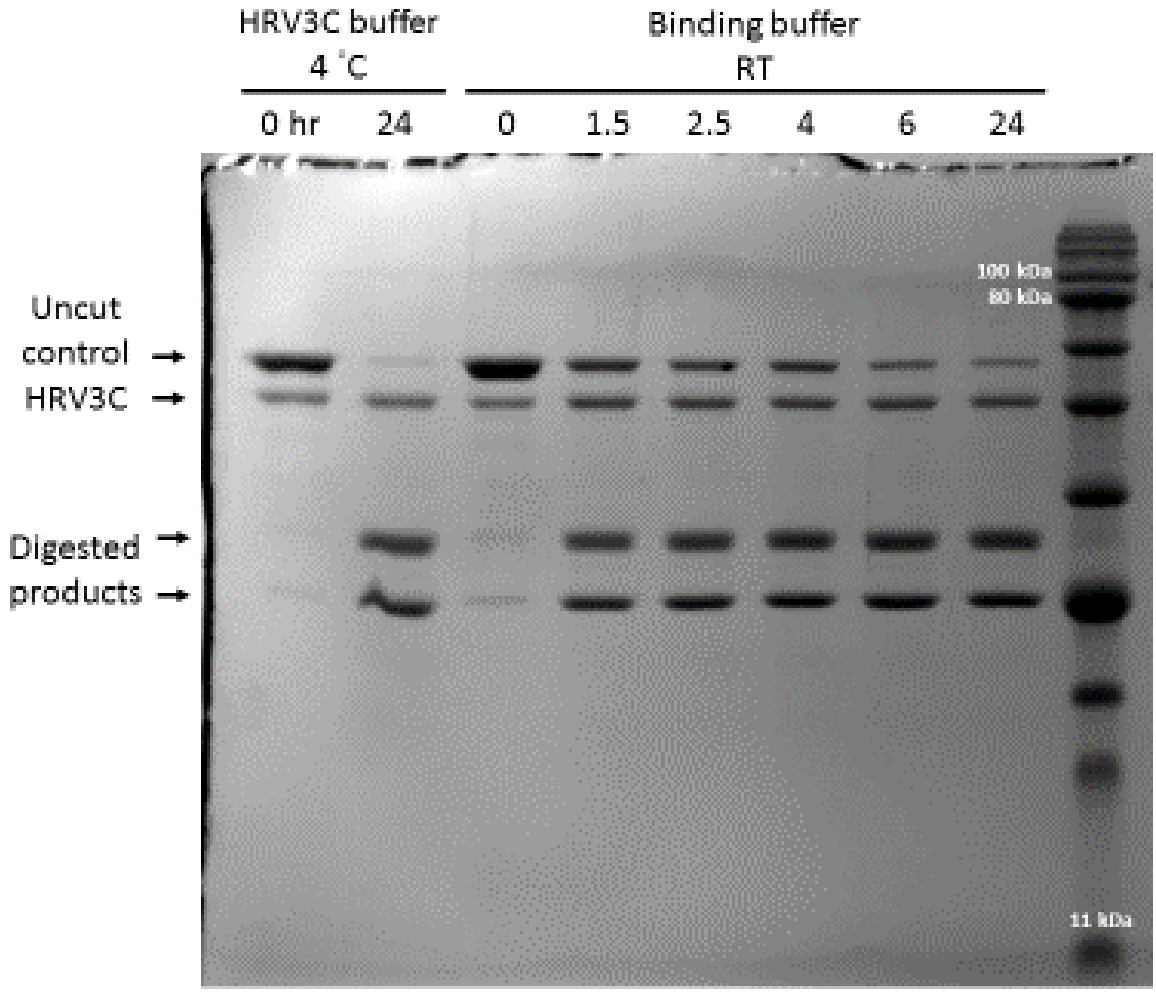


Figure D. 11. Digestion of HRV3C control protein with HRV3C in HRV3C buffer and binding buffer.

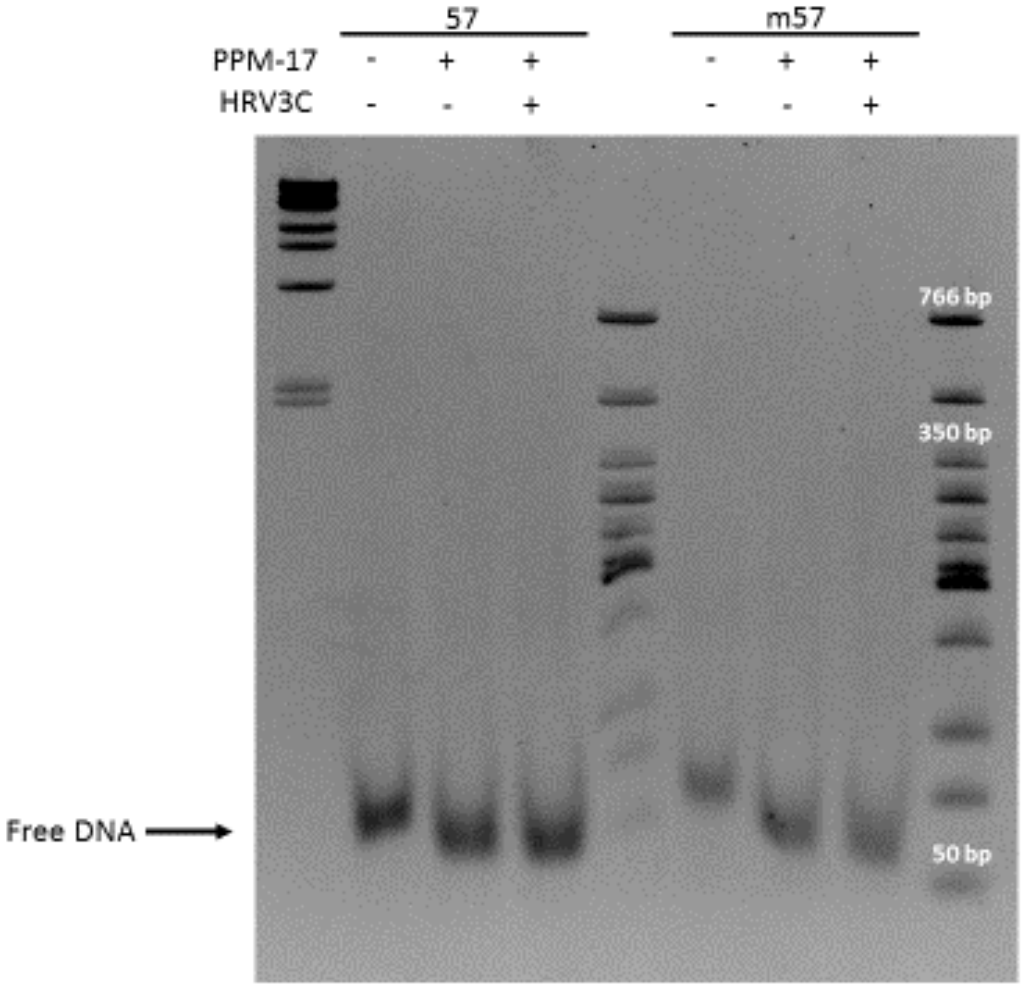


Figure D. 12. Binding assay of PPM-17 in HRV3C buffer with 50 ng/µL dIdC polymer.

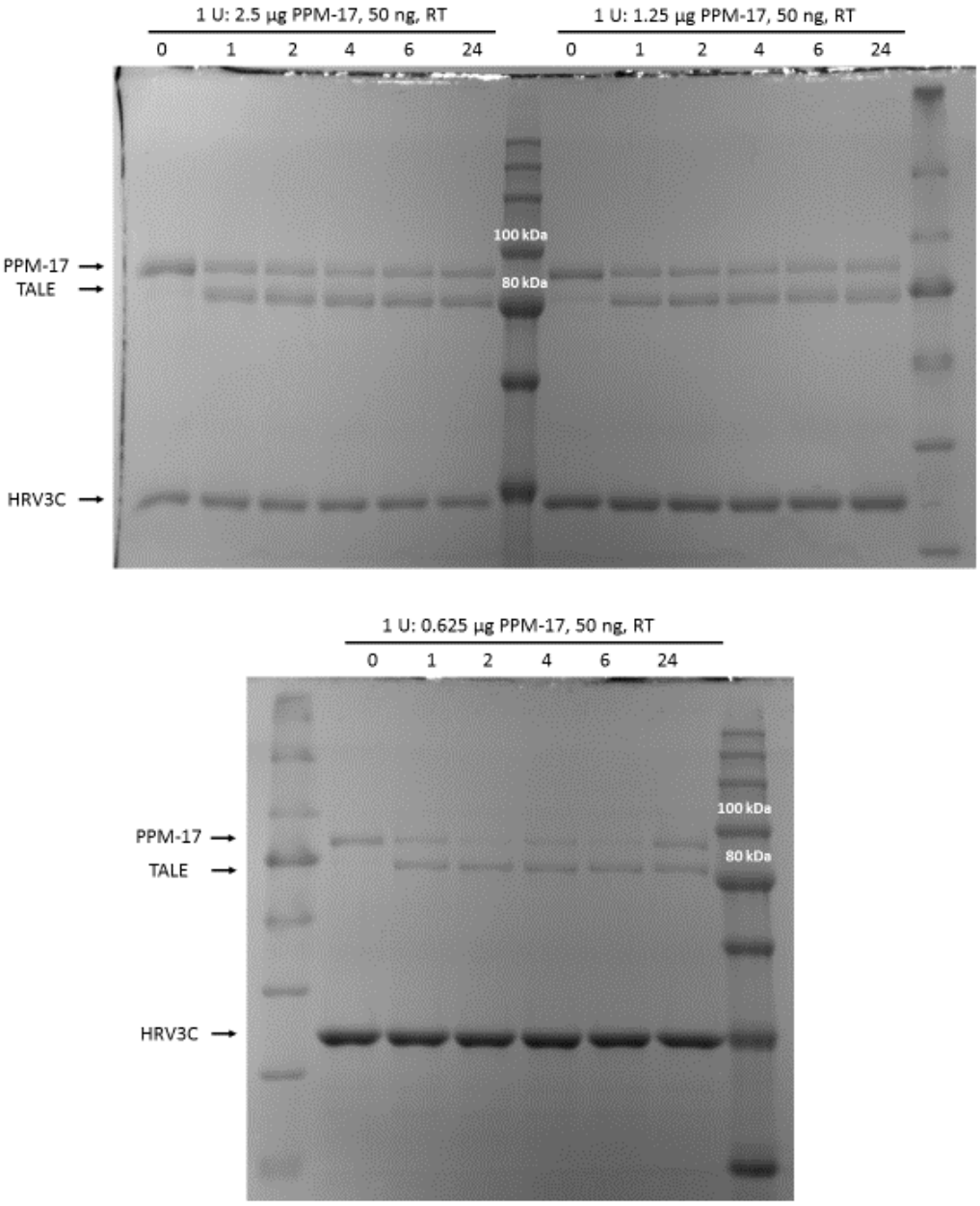
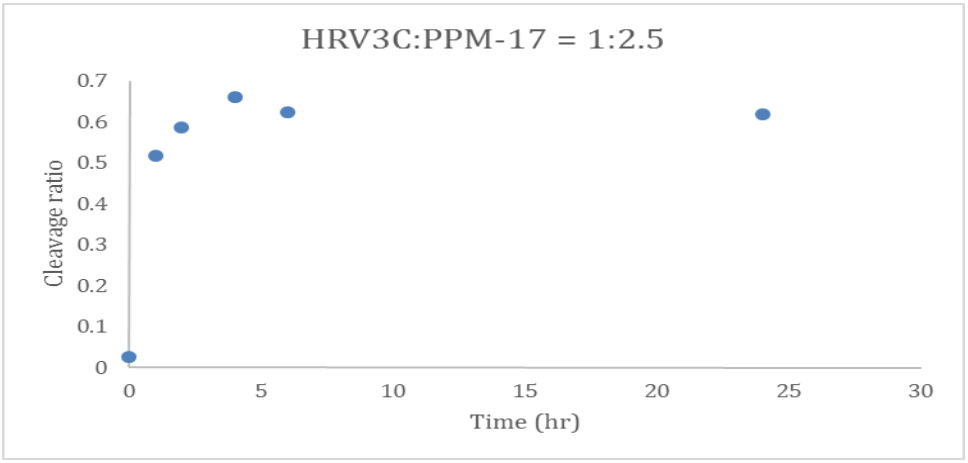
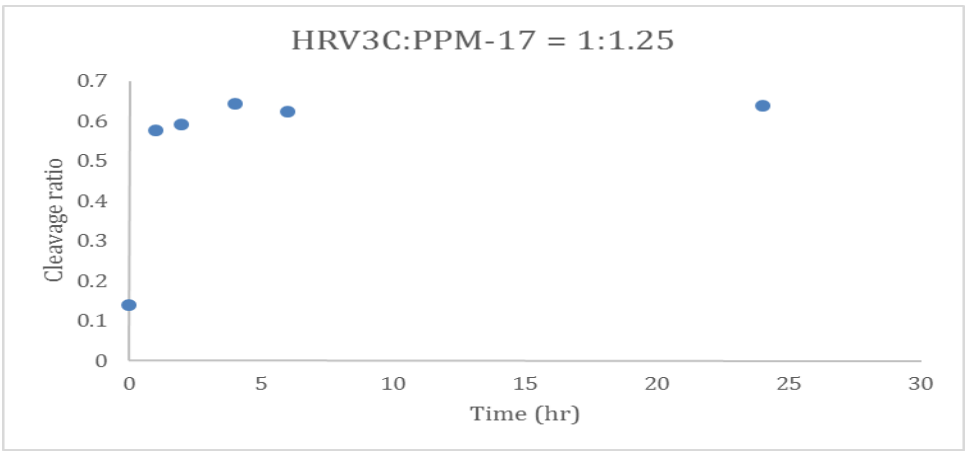


Figure D. 13. HRV3C digestion of PPM-17 with different HRV3C:PPM-17 ratios. PPM-17 arrow indicates undigested PPM-17; TALE arrow indicates TALE domain from digested PPM-17; HRV3C arrow indicates HRV3C protease.





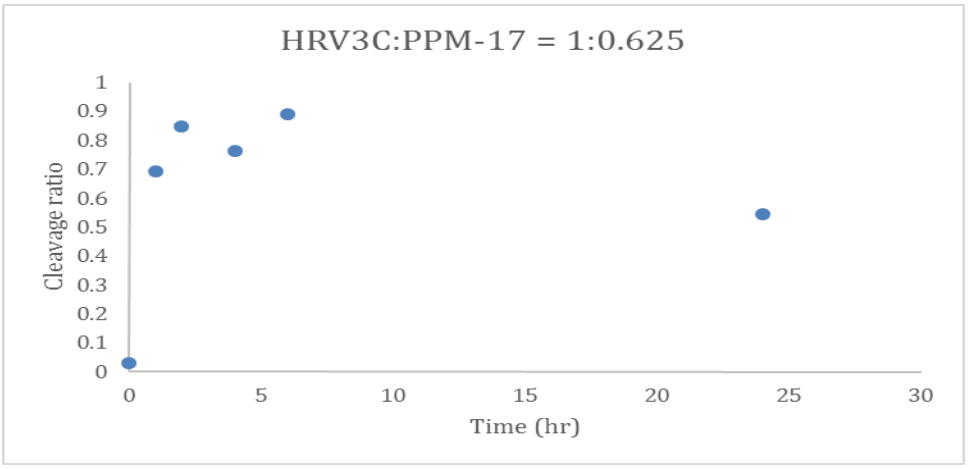


Figure D. 14. PPM-17 digestion efficiency assessed by densitometry. Corresponding SDS-PAGE gels can be found in Figure D. 13.

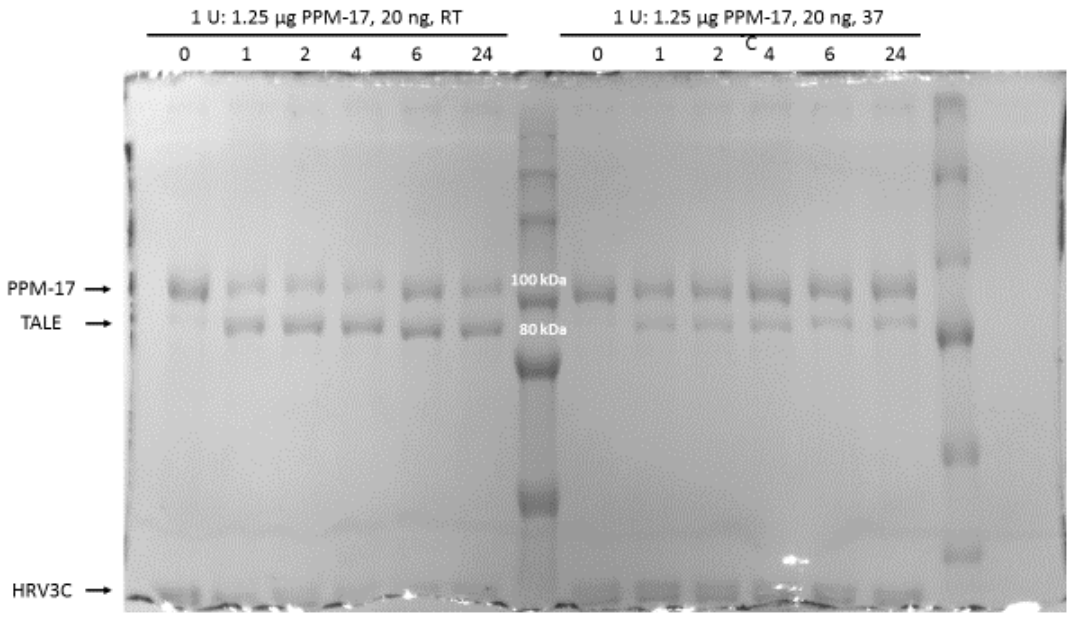


Figure D. 15. HRV3C digestion of PPM-17 at different temperatures. PPM-17 arrow indicates undigested PPM-17; TALE arrow indicates TALE domain from digested PPM-17; HRV3C arrow indicates HRV3C protease.

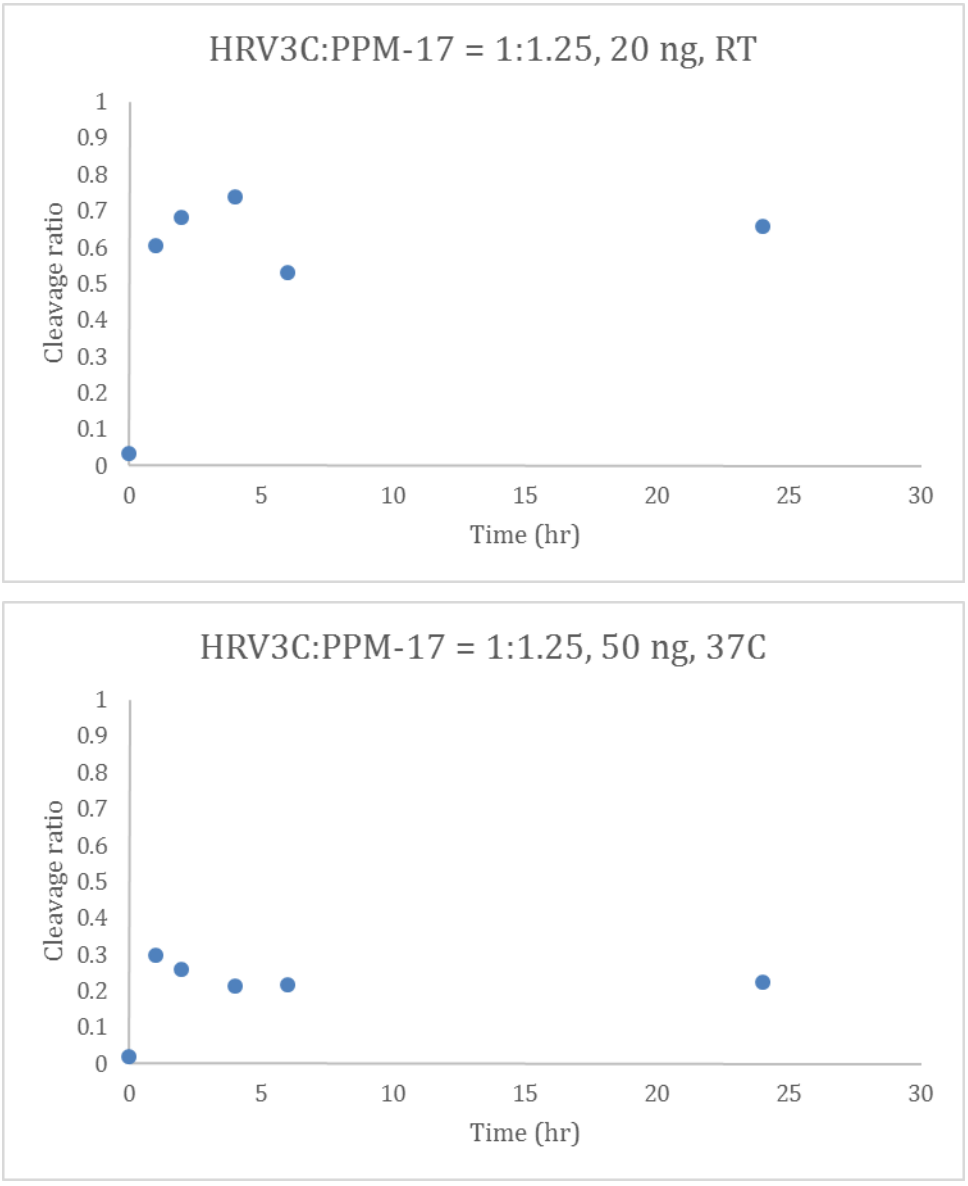


Figure D. 16. PPM-17 digestion efficiency assessed by densitometry. Corresponding SDS-PAGE gels can be found in Figure D. 15.

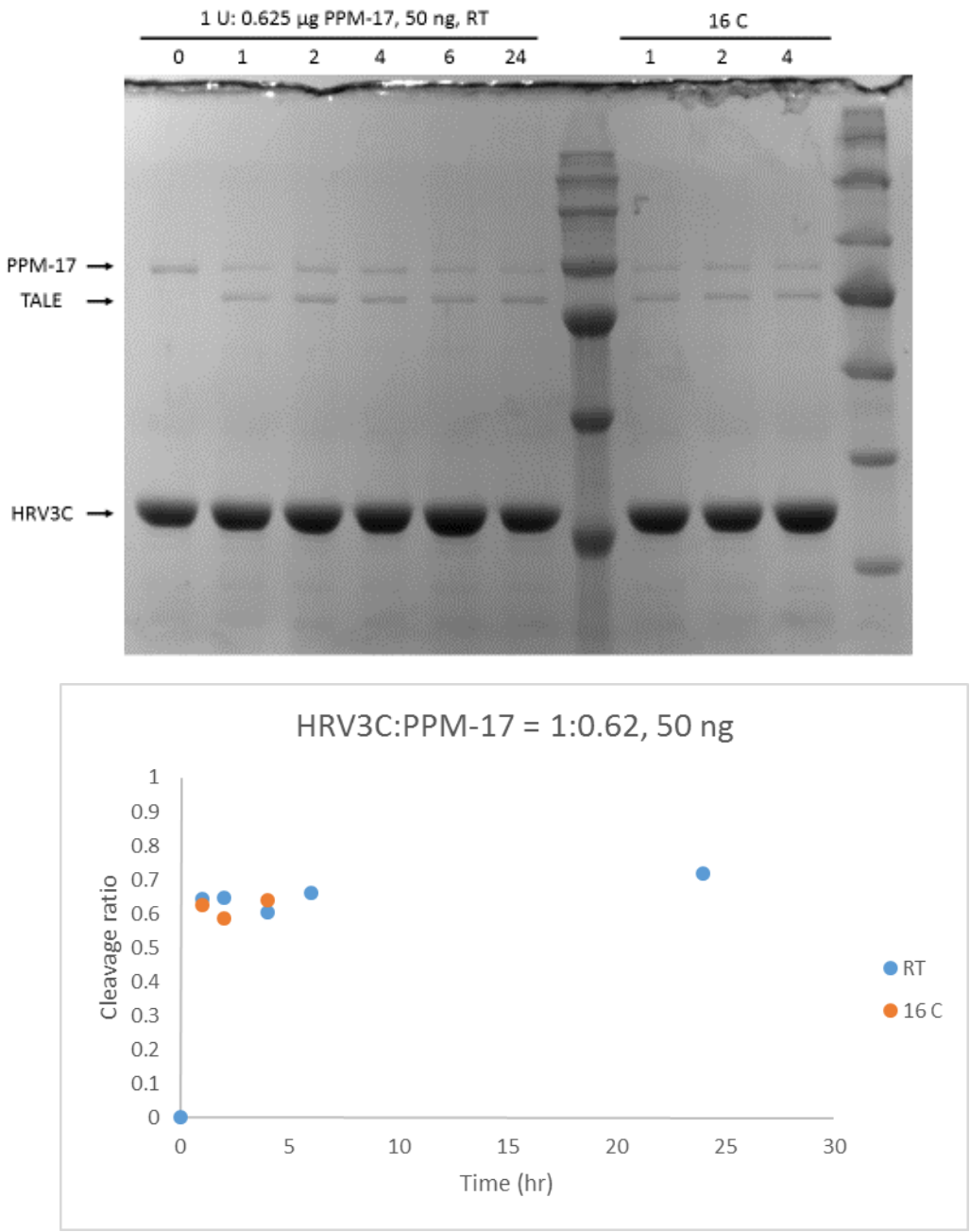


Figure D. 17. PPM-17 digestion by HRV3C at room temperature and 16 °C (top) and PPM-17 digestion efficiency assessed by densitometry (bottom).

REFERENCES

(1) Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003,** *33*, 245-254.

(2) Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **2007,** *447* (7143), 425-432.

(3) Schübeler, D. Function and information content of DNA methylation. *Nature* **2015,** *517* (7534), 321-326.

(4) Fouse, S. D.; Shen, Y.; Pellegrini, M.; Cole, S.; Meissner, A.; Van Neste, L.; Jaenisch, R.; Fan, G. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2008,** *2* (2), 160-169.

(5) Yeo, S.; Jeong, S.; Kim, J.; Han, J.-S.; Han, Y.-M.; Kang, Y.-K. Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells. *Biochem. Biophys. Res. Commun.* **2007,** *359* (3), 536-542.

(6) Choi, M. R.; In, Y.-H.; Park, J.; Park, T.; Jung, K. H.; Chai, J. C.; Chung, M. K.; Lee, Y. S.; Chai, Y. G. Genome-scale DNA methylation pattern profiling of human bone marrow mesenchymal stem cells in long-term culture. *Exp. Mol. Med.* **2012,** *44* (8), 503-512.

(7) Bork, S.; Pfister, S.; Witt, H.; Horn, P.; Korn, B.; Ho, A. D.; Wagner, W. DNA methylation pattern changes upon long- term culture and aging of human mesenchymal stromal cells. *Aging Cell* **2010,** *9* (1), 54-63.

(8) Lee, M.-H.; Kim, Y.-J.; Yoon, W.-J.; Kim, J.-I.; Kim, B.-G.; Hwang, Y.-S.; Wozney, J. M.; Chi, X.-Z.; Bae, S.-C.; Choi, K.-Y. Dlx5 specifically regulates Runx2 type II expression by binding to homeodomain-response elements in the Runx2 distal promoter. *J. Biol. Chem.* **2005,** *280* (42), 35579-35587.

(9) Herman, J. G.; Graff, J. R.; Myöhänen, S.; Nelkin, B. D.; Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U. S. A.* **1996,** *93* (18), 9821-9826.

(10) Eads, C. A.; Danenberg, K. D.; Kawakami, K.; Saltz, L. B.; Blake, C.; Shibata, D.; Danenberg, P. V.; Laird, P. W. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* **2000,** *28* (8), e32-00.

(11) Meissner, A.; Gnirke, A.; Bell, G. W.; Ramsahoye, B.; Lander, E. S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005,** *33* (18), 5868-5877.

(12) Serre, D.; Lee, B. H.; Ting, A. H. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* **2010,** *38* (2), 391-399.

(13) Nan, X.; Meehan, R. R.; Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **1993,** *21* (21), 4886-4892.

(14) Ohki, I.; Shimotake, N.; Fujita, N.; Jee, J.-G.; Ikegami, T.; Nakao, M.; Shirakawa, M. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **2001,** *105* (4), 487-497.

(15) Hendrich, B.; Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **1998,** *18* (11), 6538-6547.

(16) Jørgensen, H. F.; Adie, K.; Chaubert, P.; Bird, A. P. Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res.* **2006,** *34* (13), e96-e96.

(17) Kim, S.-E.; Chang, M.; Yuan, C. One-pot approach for examining the DNA methylation patterns using an engineered methyl-probe. *Biosens. Bioelectron.* **2014,** *58*, 333-337.

(18) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **2001,** *11* (1), 39-46.

(19) Badran, A. H.; Furman, J. L.; Ma, A. S.; Comi, T. J.; Porter, J. R.; Ghosh, I. Evaluating the global CpG methylation status of native DNA utilizing a bipartite split-luciferase sensor. *Anal. Chem.* **2011,** *83* (18), 7151-7157.

(20) Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* **2011**, gkr188.

(21) Streubel, J.; Blücher, C.; Landgraf, A.; Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **2012,** *30* (7), 593-595.

(22) Kim, Y.; Kweon, J.; Kim, A.; Chon, J. K.; Yoo, J. Y.; Kim, H. J.; Kim, S.; Lee, C.; Jeong, E.; Chung, E. A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* **2013,** *31* (3), 251-258.

(23) Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V.; Bogdanove, A. J.; Voytas, D. F. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **2011**, gkr218.

(24) Pabo, C. O.; Peisach, E.; Grant, R. A. Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* **2001,** *70* (1), 313-340.

(25) Meckler, J. F.; Bhakta, M. S.; Kim, M.-S.; Ovadia, R.; Habrian, C. H.; Zykovich, A.; Yu, A.; Lockwood, S. H.; Morbitzer, R.; Elsäesser, J. Quantitative analysis of TALE–DNA interactions suggests polarity effects. *Nucleic Acids Res.* **2013,** *41* (7), 4118-4128.

(26) Lungu, C.; Pinter, S.; Broche, J.; Rathert, P.; Jeltsch, A. Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites. *Nat. Commun.* **2017,** *8* (1), 649.

APPENDIX E. PROTOCOLS

E. 1 Amine Thiolation

Reagents

- EDTA
- Phosphate buffer (PB) (or other non-amine buffer)
- Traut's reagent (Stock solution: 700 mM in PB or in acidic non-amine buffer)
- Fresh urea

Procedure

- Dissolve protein in 8M urea (pH 8.0) with 100 mM PB (with 2-5 mM EDTA, optional)
- Add Traut's stock solution into the protein solution to a 100-fold molar excess
- Incubate the solution for 1 hour at room temperature
- (Optional: add 10 mM final concentration of TCEP and incubate at RT for 15 min)
- Dialyze against MQ or use desalting column to remove unreacted Traut's reagent

Notes

- The reaction requires a slightly alkaline condition
- Traut's reagent could be in 10- to 50-fold molar excess
- The extent of conversion can be quantified by Ellman's assay

E. 2 Ellman's Assay

This assay was used to quantify reduced thiol contents in purified proteins or other thiol-containing samples.

The amounts and volumes below are enough for 10 standards in triplicate

1. Make 40 mL of reaction buffer (0.1 M phosphate buffer with 1 mM EDTA, pH 8)
2. Make Ellman's stock solution: 2 mg of Ellman's reagent in 0.5 mL reaction buffer
3. Prepare Ellman's working solution by adding 200 uL Ellman's **stock** solution to 10.1037 mL of reaction buffer
4. Read blank plate at 412 nm

5. Pipette 15 µL of each standard into each well
6. Pipette 155 µL of Ellman's working solution into each well
7. Mix for 5 min at RT
8. Read at 412 nm
9. To make reaction buffer with 5 mM EDTA, add 11.69 mg EDTA or add 80 µL of 0.5 M EDTA stock solution into 40 mL of reaction buffer

No.	mM	Amount	Volume
1	1.5	1.8174 mg Cys	10 mL
2	1	200 µL sol. 1 + 100 µL blank	
3	0.75	150 µL sol. 1 + 150 µL blank	
4	0.5	100 µL sol. 1 + 200 µL blank	
5	0.25	150 µL sol. 4 + 150 µL blank	
6	0.125	150 µL sol. 5 + 150 µL blank	
7	0.0625	150 µL sol. 6 + 150 µL blank	
8	0.03125	150 µL sol. 7 + 150 µL blank	
9	0.015625	150 µL sol. 8 + 150 µL blank	
10	0	Blank = reaction buffer	

E. 3 Alkylation of Cysteine

This protocol was used to block thiols from forming disulfide bonds or other undesired conjugations.

Reagents

1. Phosphate buffer (PB)
2. Iodoacetamide (IAA)
3. TCEP
4. Urea

Procedure

1. Dissolve proteins in 8M urea (pH 8.0) with 100 mM PB and 10 mM TCEP
2. Incubate the protein solution at 55°C for 15 min
3. Cool down the solution to RT
4. Right before use, prepare IAA stock solution at 375 mM in MQ with 100 mM ammonium bicarbonate at pH 8.0.
5. Add IAA stock solution into the protein solution to a final concentration of 20 mM

6. Incubate the solution for 1-1.5 hr in the dark at room temperature
7. Dialyze against 5% AcOH to remove IAA

Notes

1. The reaction requires a slightly alkaline condition, other pH leads to alkylation of Lys, Met, His, Asp, Glu
2. If using reducing agent other than TCEP (e.g. DTT), incubate the protein solution at 55°C for 30 min. The temperature should not be higher than 60°C to prevent urea-based carbamylation of Lys and protein N-termini
3. IAA is sensitive to light, the stock solution should always be made fresh and kept in dark
4. Prolonged reaction time could lead to alkylation of other residues
5. The extent of alkylation can be quantified by Ellman's assay

E. 4 Amine Acetylation

This protocol was developed to block amines from nucleophilic reactions.

1. Prepare 2 mg/mL protein solution in 50 mM ammonium bicarbonate buffer
2. Make acetylation buffer (25 (v/v)% acetic anhydride with (75 v/v)% methanol)
3. Mix protein solution with 2-5 times acetylation buffer by volume
4. Incubate at room temperature for at least 30 min
5. Stop the reaction by extensive dialysis at 4 °C

E. 5 Carboxyl-Hydrazide Crosslinking

Carboxylic groups on aspartic acid, glutamic acid, or the C-terminus will react with hydrazides on adipic acid dihydrazide (ADH) in the presence of 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) at neutral pH. The chain length of ADH is 11.1 Å, which is close to that of a previously reported chemical crosslinker used in verification of leucine zipper oligomerization.

- ADH stock solution: 100 mg/mL in PBS
- DMTMM stock solution: 144 mg/mL in PBS

- Mix protein solution, ADH stock solution, and DMTMM stock solution to achieve final concentrations as 2 mg/mL protein, 8.3 mg/mL ADH, and 12 mg/mL DMTMM
- Incubate the solution at 37 °C for one hour
- Samples can be analyzed directly by SDS-PAGE

E. 6 Cell Lysate Fractionation

Buffer preparation

Buffer	Composition
Wash Buffer	50 mM Tris Cl pH 7.5 0.1 M NaCl 1mM benzamidine (BZA) 1mM beta mercaptoethanol (BME)
Triton buffer	Wash buffer with 0.01% Triton X-100
His Denature	6 M Gu HCl 10 mM Beta mercaptoethanol 0.1 M NaH2PO4 pH 7.9 0.3 M NaCl 0.1 M Tris-Cl pH 7.9
Elution buffer pH 6.3	His Denature at pH 6.3
Elution buffer pH 4.5	His Denature at pH 4.5
Wash buffer pH 7.2	His Denature at pH 7.2
TALE dialysis and storage buffer	480 mM KCl 1.6 mM EDTA 2 mM DTT 12 mM Tris-Cl pH 7.5

Fractionization

1. Centrifuge culture medium at 2900 x g at 4 °C for 25 min
2. Resuspend the cell pellet at 2 mL Wash buffer/50 mL culture

3. Sonicate the slurry 15 times with 20-second on/off intervals on ice at a moderate power. Collect and label it as WCE.
4. Add four times of resuspension volume of Wash Buffer and centrifuge the slurry at 15,000 x g at 4 °C for 25 min
5. For the supernatant: Take a sample and label it as SF (Soluble Fraction)
6. For the pellet: Resuspend the pellet in 6 to 10 mL of Triton Buffer and pour the resuspended pellet in a tight douncer. Add 4 mL of Triton buffer to wash the falcon tube and add this volume to the douncer to homogenize. Transfer the solution to a 50 mLl falcon tube and use 5 mL of Triton Buffer to wash the douncer. Mix all the liquids together
7. Centrifuge the solution at 15,000 x g at 4 °C for 25 min
8. For the supernatant: Take a sample and label as SP (Supernatant)
9. For the pellet: Resuspend the pellet in 5 mL of His-Denat Buffer and pour the resuspended pellet in a tight douncer. Add 2.5 mL of His-Denat to wash the falcon tube and add this volume to the douncer to homogenize. Transfer the solution to a 50 mL falcon tube and use 2.5 mL of His-Denat buffer to wash the douncer. Mix all the liquids together and stir for one hour at room temperature
10. Centrifuge the solution at 15,000 x g at 4 °C for 25 min
11. For the Pellet: Take a sample and label it as Pel
12. For the rest: Take a sample and label it as Total
13. Run a SDS-PAGE gel to check protein concentration in each fraction
14. For proteins in SF, purify with native Ni column; for proteins in other fractions, purify with denatured Ni column
15. Dialyze purified protein against a desired buffer

E. 7 Protein-DNA Binding Assay (Electrophoretic Mobility Shift Assay, EMSA)

1. Binding buffer: 10 mM Tris pH 7.9, 100 mM KCl, 50 ng/µl of poly dI:dC
2. Use 1 µM methylated DNA or unmethylated DNA containing the TALE binding sites
3. Final protein concentration = 23 µM. Dilute this stock with the TALE storage buffer as needed, such that 1 µl of the diluted protein is added to the 10 µl reaction mixture.

4. Allow the reaction to proceed for 30 min at room temperature
5. Assess the binding on 6 wt% TBE-PAGE gel by mixing samples (5 parts) with 25% sucrose (2 parts)
6. Run the gel at 150 V for 35 min
7. Stain with EtBr solution for at least 30 min and destain in water for at least 30 min

E. 8 Regular Salting-Out and Heating Purification

1. Lyse cells
    a. Add 2mL of lysing buffer (Urea 8M, Tris-Cl 10mM, NaH_2PO_4 100 mM, pH 8) per g of cells and ensure cells are resuspended thoroughly
    b. Freeze at -80 ^oC for at least 1 hr
    c. Defrost, agitate by vortexing for 30 minutes (can be done at room temperature)
    d. Centrifuge (4 ^oC, 20 min, 8000 RCF). Store cleared lysate at -80 ^oC
2. Salt out protein
    a. Thaw cleared lysate on ice
    b. Add 10 w/v% ammonium sulfate (100 mg/mL) to salt out undesired proteins. Vortex 30 s and incubate on ice for 10 min.
    c. Centrifuge (4 ^oC, 20 min, 8000 RCF).
    d. Transfer cleared solution to another tube and add the same amount of ammonium sulfate as in part b for a total of 20 wt% to salt out the desired protein. Vortex30 s and incubate on ice for 10 min.
    e. Centrifuge (4 ^oC, 20 min, 8000 RCF).
    f. Weigh resulting pellet; re-suspend 100 mg/mL in water.
3. Separation based on heat stability
    a. Heat 100 mg/mL protein solution at 75 ^oC for 5 min, vortex, heat again for 5 min, vortex
    b. Cool down to 10 ^oC
    c. Centrifuge, 11000 RCF, 15 min (do it twice if it is hard to separate the solids)
4. Dialysis
    a. Add cleared lysate from heat step to dialysis tubing in water in cold room
    b. Exchange water at 2-3 hours

c. Finish dialysis after 3-5 more hours, centrifuge aggregates 11000 RCF, 15 min
d. Freeze dry overnight, store in desiccant in the refrigerator

E. 9 Denatured Ni-IPA Column for Endotoxin Removal

- Buffer B, pH 8; Buffer C, pH 6.3 (1 L)
  - 100 mM NaH_2PO_4-H_2O
  - 10 mM Tris-HCl
  - 8 M Urea
  - 300 mM NaCl
- Native buffer D, pH 8 (100 mL)
  - 10 mM Tris-HCl
  - 300 mM NaCl
- Buffer D/IPA, pH 8 (100 mL)
  - 10 mM Tris-HCl
  - 40 mL MQ + 60 mL isopropanol (IPA)
  - 300 mM NaCl
- Elute buffer E, pH 8 (100 mL)
  - 10 mM Tris-HCl
  - 500 mM Imidazole
  - 300 mM NaCl

- Equilibrate the column with 2-times bed volume of buffer B
- Incubate for 1 hr at RT
- Flow through of buffer B
- Wash twice with 5-times bed volume of buffer C
- Wash once with 5-times bed volume of native buffer D
- Wash once with 5-times bed volume of buffer D/IPA followed by 5-times bed volume of native buffer D
- Repeat above for 2 more times
- Elute with elute buffer E for 10-times of bed volume

- Notes:
    - For Cys-containing proteins, 10 mM β-mercaptoethanol can be added in all buffers except in elute buffer E
    - Elutes can be reduced with 20 mM DTT before dialysis
    - It is recommended to do a step-down dialysis with 300 mM NaCl, 100 mM NaCL, and then water

E. 10 Purification of YC12 Proteins

Proteins: SK-YC12-His (YC12), SK-YC12-BMP-His (YC12-M), SK-YC12-$(BMP)_2$-His (YC12-D), SK-YC12-$(BMP)_4$-His (YC12-T)

Expression and Harvesting

1. Express proteins in 2xYT with an inoculation ratio of 1:100. Induce at OD = 0.6-0.8 at 37 C, shaking at 300 rpm with 1.25 mM IPTG (298.8 mg/mL) for 3 hr (to 5 hr)
2. Harvest cells at 700 RCF for 15 min at 4 ^{o}C
3. Collect pellets in 50 mL tubes (preweigh). Resuspend pellets **in Buffer B at 2 mL/g wet pellet**
4. Freeze at -80 ^{o}C after resuspension
5. Thaw the resuspension; **add 200 mM BME (14.03 µL/mL).** Sonicate with a tip sonicator at a moderate power with 30 sec on/off cycle on ice. Mix the resuspension every 10 cycles until completing **40 cycles (total sonication time = 20 min)**
6. Centrifuge sonicated resuspension at 11000 rcf for 25 min at 4 ^{o}C
7. Collect supernatant (=cleared lysate = CL). CL can be stored at -20 ^{o}C

Salting out

8. Keep CL on ice **(if stored at -20 ^{o}C, add 200 mM BME after thawing);** record the volume and
    a. For YC12, YC12-M, and YC12-D
        i. Add ammonium sulfate (AS) at 100 mg/mL CL (10 w/v%); vortex to make sure AS is completely dissolved. Incubate on ice for >10 min
        ii. Centrifuge at 11000 RCF for 25 min at 4 ^{o}C

iii. Collect supernatant (10S) in pre-weighed tubes; add the same amount of AS in step i (for example, if used 1 g AS in step i, add another 1 g AS to 10S)
iv. Vortex, incubate, and centrifuge as before
v. Collect pellet (20P)

b. For YC12-T
i. Add AS at 200 mg/mL CL
ii. Vortex, incubate, and centrifuge as before
iii. Collect pellet (20P)

ITC

9. Resuspend 20P pellets
    a. For YC12, at 50 mg/mL in ice-cold PBS with 200 mM BME
    b. For YC12-M, -D, and -T, at 200 mg/mL **in 2 M urea in PBS with 200 mM BME**

    Vigorous mixing is necessary (i.e., long vortexing, vigorous pipetting, or overnight stirring). Keep resuspension on ice
10. Centrifuge at 11000 RCF for 15 min at 4 ^{o}C
11. Collect supernatant (CS0); register the volume and add NaCl to 2 M (e.g., 10 mL CS0 needs 1.2 g NaCl)
12. CS0 should turn cloudy after dissolving NaCl at RT; if not, increase [NaCl] to 3 M
13. Centrifuge cloudy CS0 at 11000 RCF at 28 ^{o}C for 15 min
14. Collect pellet (HP1) and resuspend HP1 **using the same solution and the same volume used in step 9.** Vigorous mixing is necessary to completely resuspend HP1
15. Centrifuge at 11000 RCF for 15 min at 4 C
16. Collect supernatant (CS1). CS1 can be stored in -80 ^{o}C.

Dialysis

17. Record the volume and add DTT to a final concentration of 50 mM
18. Incubate at RT for 10 min
19. Record the volume to calculate the length of dialysis membrane
20. Dialyze CS1 against pre-chilled RO water in the cold room until achieving 10^6 dilution-by-volume

21. Centrifuge dialyzed solution at 11000 RCF for 10 min at 4 ^{o}C
22. Collect supernatant (dia. Sup) and immediately flash-freeze it in liquid N_2
23. Lyophilize the frozen solution immediately

E. 11 Plasmid Isolation from *E. Coli*

This protocol can be used for checking cloning results but not for preparing plasmids for cloning/stock purposes.

1. Resuspend the cell pellet in 250 µL Solution A (25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 0.5 mg/mL RNAse)
2. Mix with 250 µL of Lysis buffer (200 mM NaOH, 1% SDS)
3. Add 300 µL 5 M Potassium acetate, pH 5.5
4. Centrifuge, take supernatant
5. Add 700 µL isopropanol. Incubate on ice for 10 min
6. Centrifuge, take pellet
7. Wash with 500 uL 70% EtOH
8. Centrifuge, dry pellet
9. Resuspend in Tris-EDTA buffer

E. 12 Stem Cell Isolation from Bone Marrow (Mice, Rats, and Rabbits)

Preparation:

Gloves

Isofuorane

cotton or tissue

70% ethanol

Container to anaesthetize mouse

Biohazard bag to discard dead mouse

Sterile scissors (2 pairs)

Sterile tweezers (4 pairs)

Dissection board/Styrofoam lid

Board pins to hold mouse

Ice bucket with ice

Syringe (10 mL) and needles (27-30 gauge)

50 mL conical tubes

Medium: Low-glucose DMEM

Complete medium: Low-glucose DMEM + 20% FBS +P/S + L-glutamine

Sterile Swinney filters

Bone collection:

1. Put animal to sleep with isofluorane and kill by cervical dislocation
2. Disinfect mouse by removing traces of urine and wet thoroughly with 70% ethanol
3. Use sterile tools to cut. Use a pair of scissors and tweezers to dissect the mouse open and a different pair for cutting the bone. Try not to touch the skin when removing bones.
4. Make a cut at the lower abdominal wall
5. Remove femur and tibia together. Try to take the whole femur by going all the way up to vertebral column. Quickly clean off the connective tissues attached to the bone as much as you can and put the bones in cold medium until further use.

Marrow flushing:

1. Have 2 plates filled with medium handy. Put bones into one plate, clean off connective tissue, and separate femur and tibia by cutting the cap in between.
2. Transfer clean bone to new plate with medium. Cut open the ends of bones.
3. Use a 10-mL syringe and needle (27 or 30 gauge) to flush marrow using medium. Make sure to flush all marrow. It may be necessary to inject medium at both the ends to get all the marrow.
4. Break the clumps, if any, by pipetting up and down. Centrifuge the marrow suspension for 10 min at 2000 rpm (~931 xg) and resuspend in 10 mL complete medium. Use complete medium in the following steps.
5. To lyse red blood cells, add 15 mL of sterile water and mix thoroughly by inverting for 30 sec and then make up the volume to 50 mL with medium.
    a. For mouse: 10 mL resuspension media + 15 mL water

b. For Rat: 10 mL resuspension media + 30 mL water
c. For rabbit: 10 mL resuspension media + 30 mL water

6. Centrifuge for 10 min at 2000 rpm (~931 xg), resuspend cells in medium (10-20 mL, depending upon the pellet size, pellet at this point should be white).
7. Count cells and plate (~$1.82x10^5$ cells/cm^2)

(Cells must be in the medium within 15-20 min after the animal has been sacrificed.)

Maintenance:

1. Change medium first time on day 4 with alpha-MEM medium. Subsequent medium changes should be performed every other day.
2. When cells get confluent, add freshly made ascorbic acid (25 μg/mL for first feeding and 50 μg/mL for subsequent feedings) and β-glycerol phosphate for osteoblast differentiation.

E. 13 Acid-Wash Coverslips

1) Heat coverslips in a loosely covered glass beaker in 1M HCl at 50-60 °C for 4-16h
2) Cool to room temperature
3) Rinse out 1M HCl with MQ
4) Fill container with MQ and sonicate in water bath for 30 mins
5) Repeat step 4 two times
6) Fill container with 50% EtOH and 50% MQ and sonicate in water bath for 30 min
7) Fill container with 70% EtOH and 30% MQ and sonicate in water bath for 30 min
8) Fill container with 70% EtOH and store in fridge

APPENDIX F. CLONING SCHEMES AND SEQUENCES

F. 1 Cloning Schemes

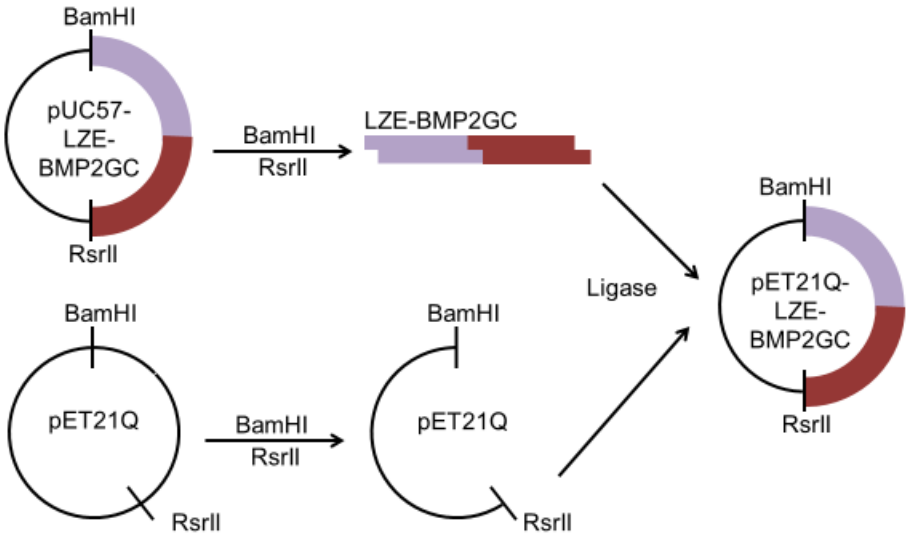


Figure F. 1. Cloning scheme of pET21Q-LZE-BMP2GC

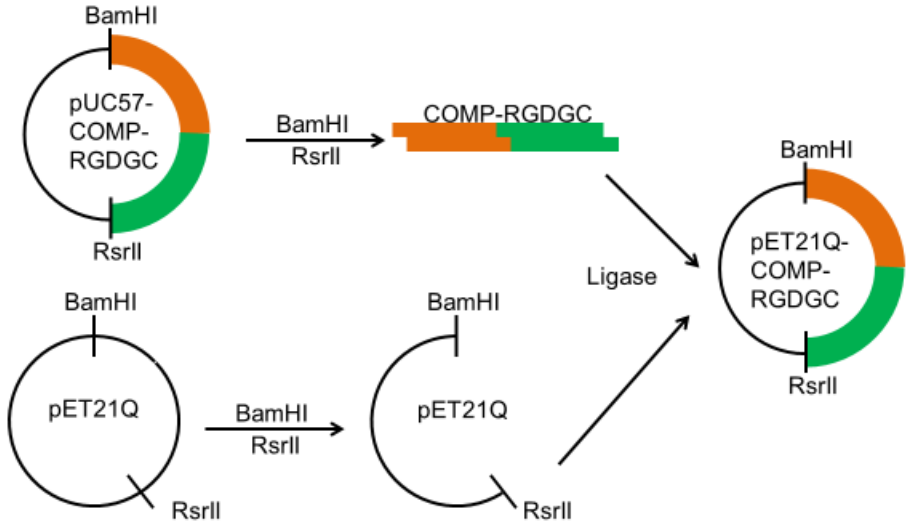


Figure F. 2. Cloning scheme of pET21Q-COMP-RGDGC

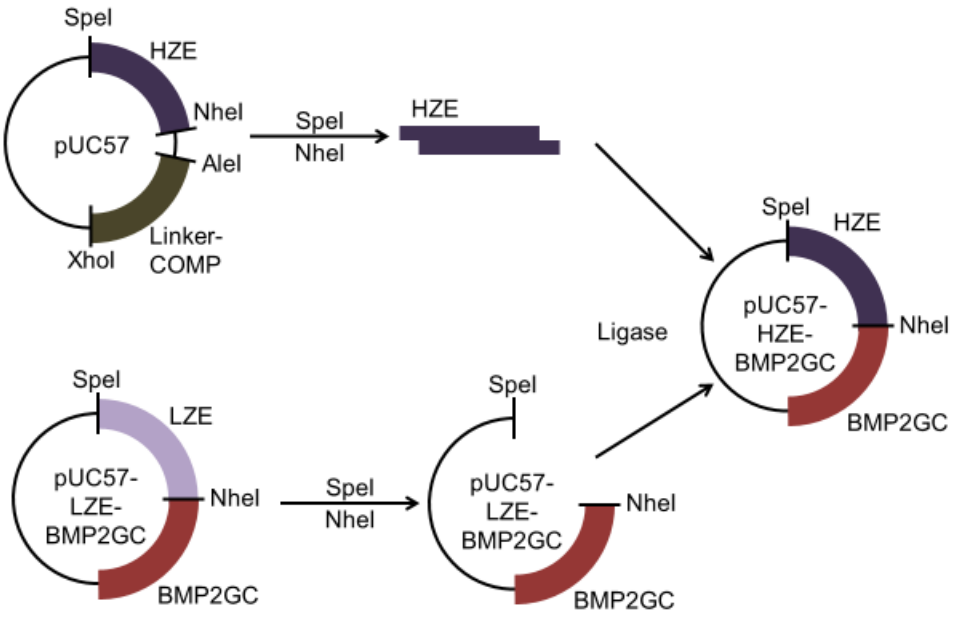


Figure F. 3. Cloning scheme of pUC57-HZE-BMP2GC

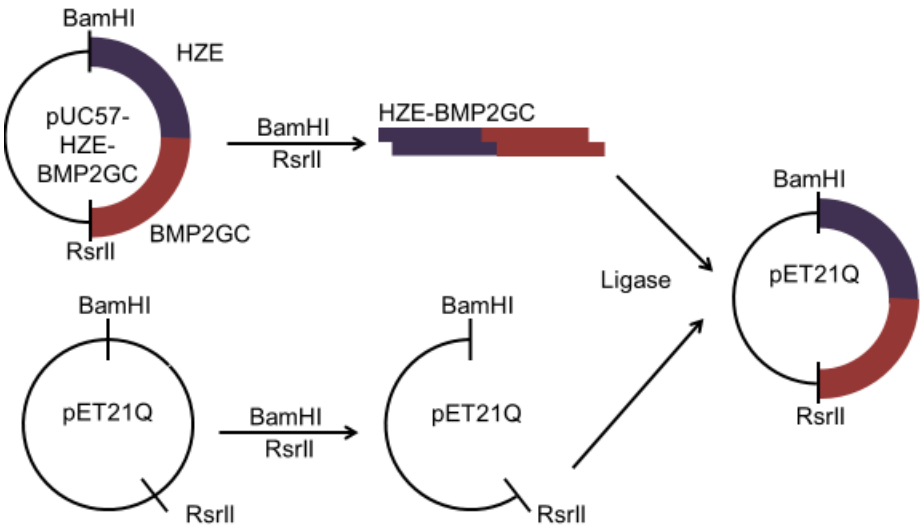


Figure F. 4. Cloning scheme of pET21Q-HZE-BMP2GC

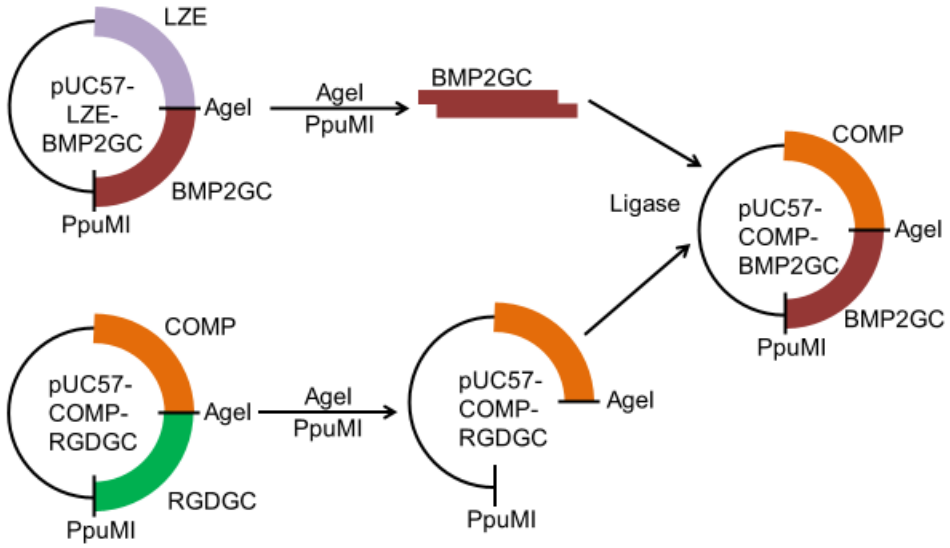


Figure F. 5. Cloning scheme of pUC57-COMP-BMP2GC

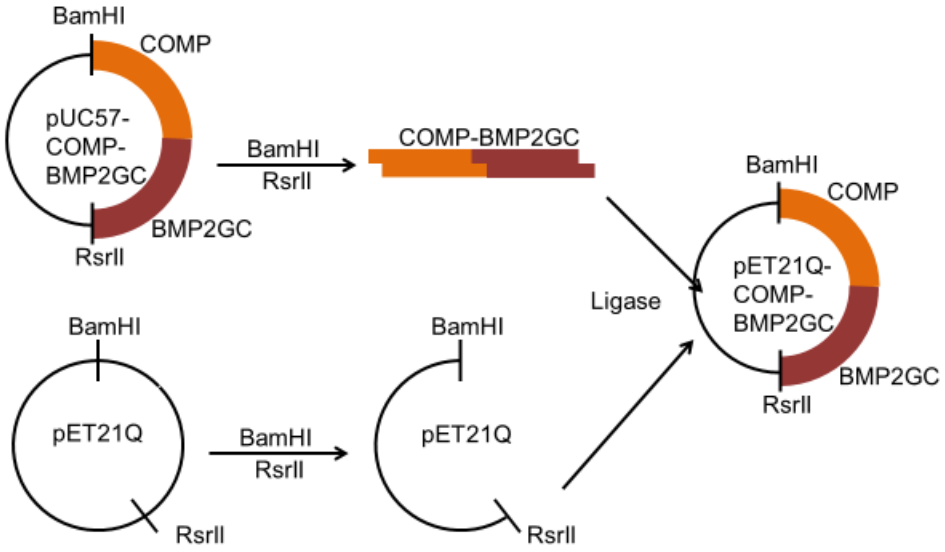


Figure F. 6. Cloning scheme of pET21Q-COMP-BMP2GC

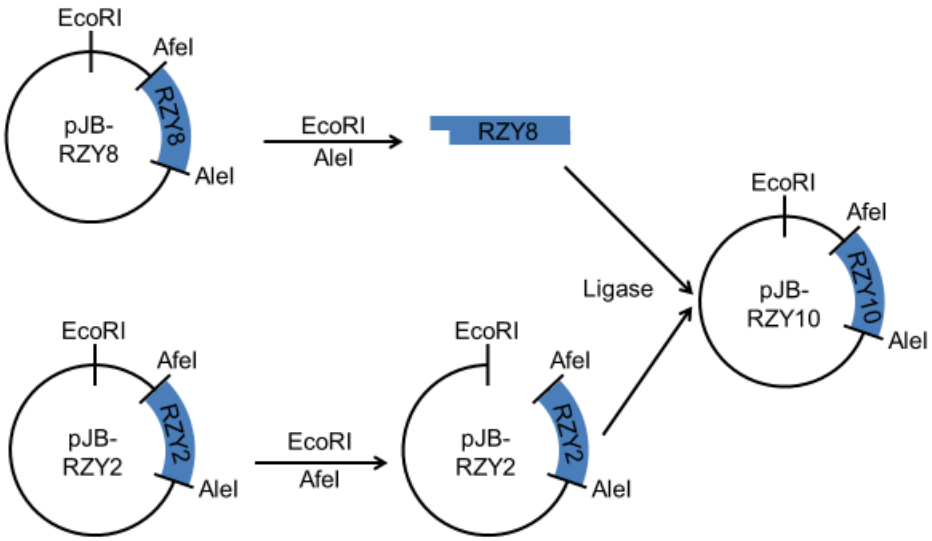


Figure F. 7. Cloning scheme of pJB-RZY10

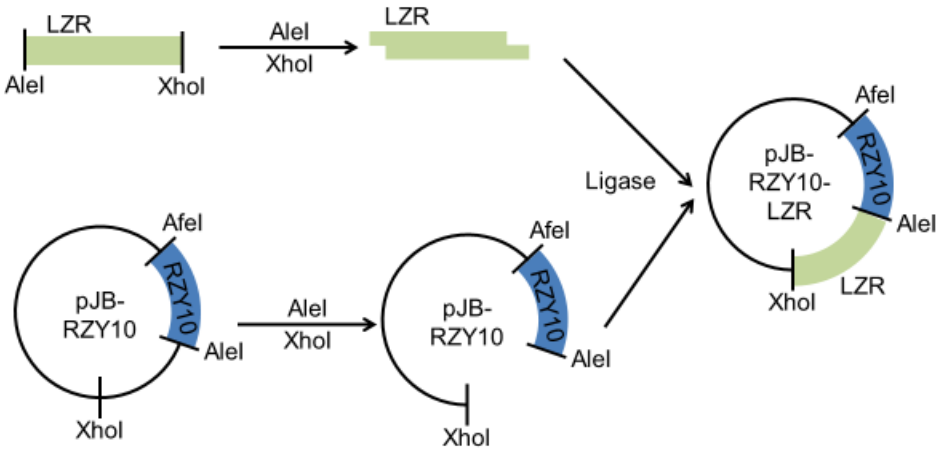


Figure F. 8. Cloning scheme of pJB-RZY10-LZR

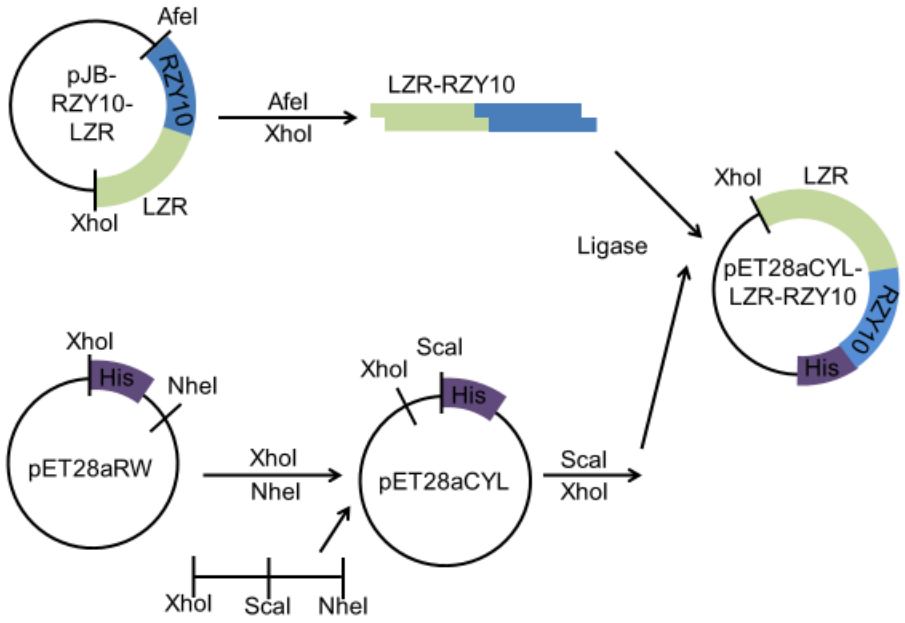


Figure F. 9. Cloning of pET28aCYL-LZR

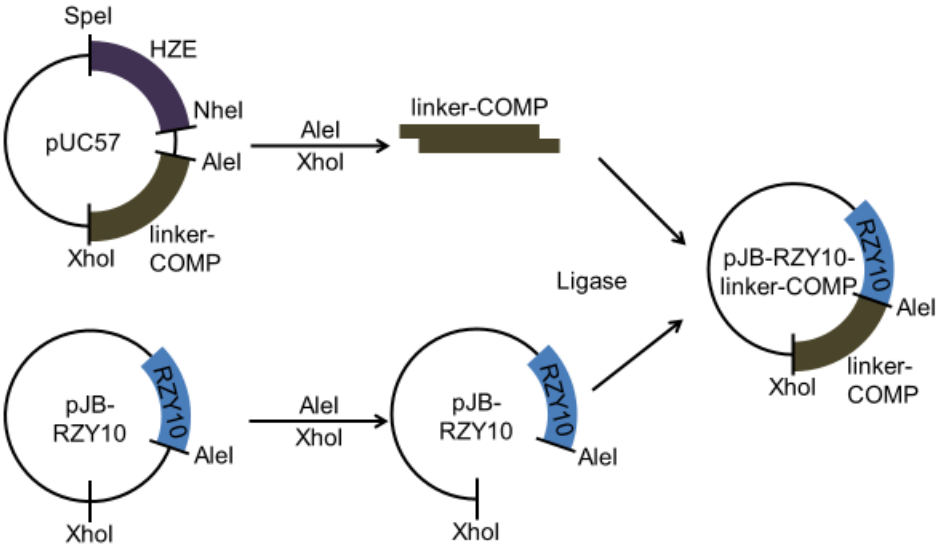


Figure F. 10. Cloning scheme of pJB-RZY10-COMP

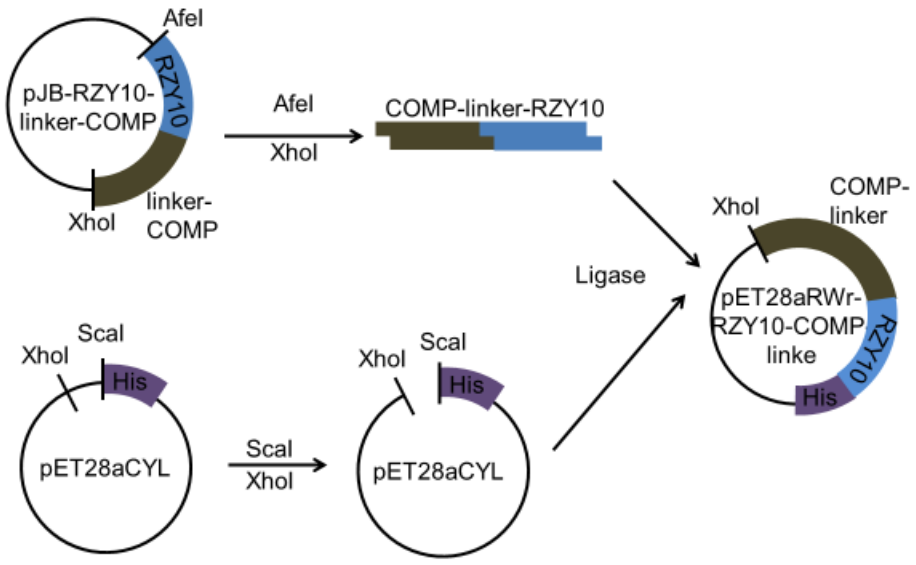


Figure F. 11. Cloning scheme of pET28aRW-RZY10-COMP

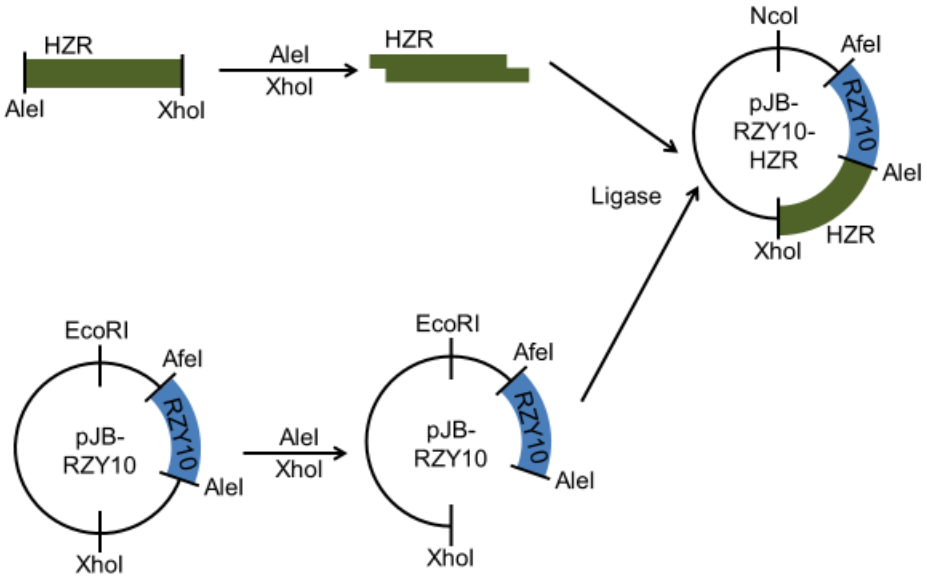


Figure F. 12. Cloning scheme of pJB-HZR

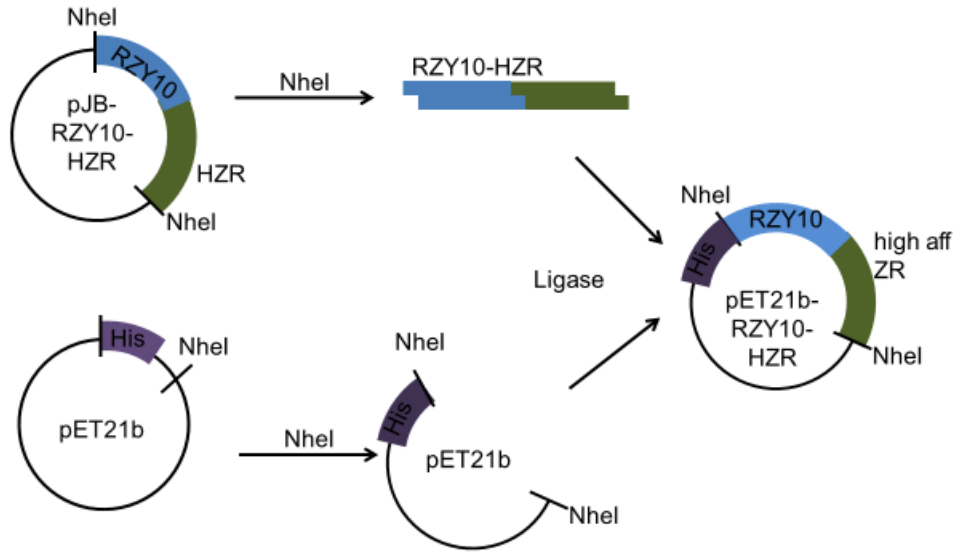


Figure F. 13. Cloning scheme of pET21b-RZY10-HZR

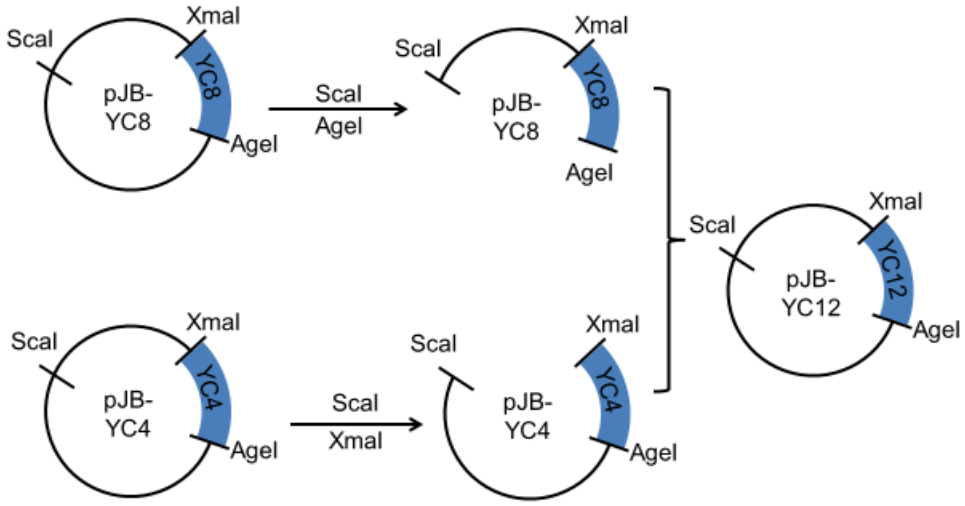


Figure F. 14. Cloning scheme of pJB-YC12 with recursive directional ligation strategy. This scheme can be applied to other pJB constructs as long as the sequence of interest (e.g., YC8) is flanked by compatible ends that are not compatible with ScaI.

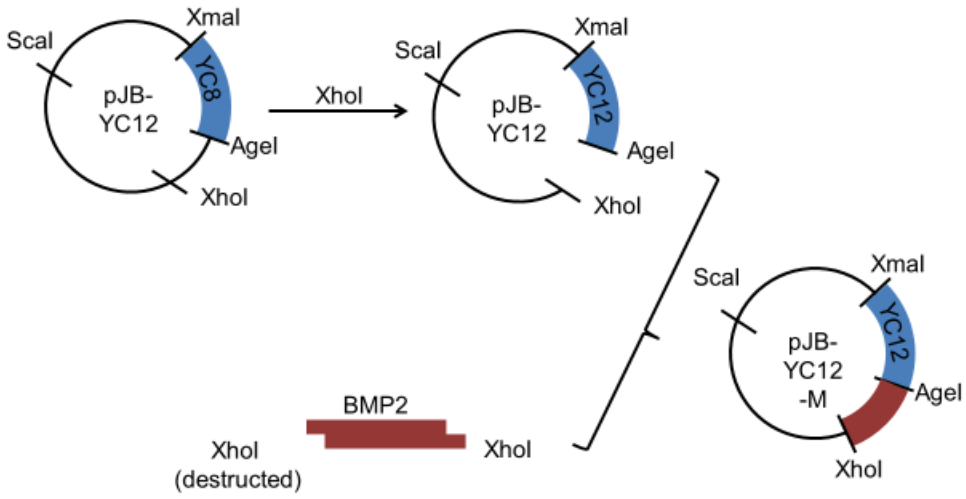


Figure F. 15. Cloning scheme of pJB-YC12-M.

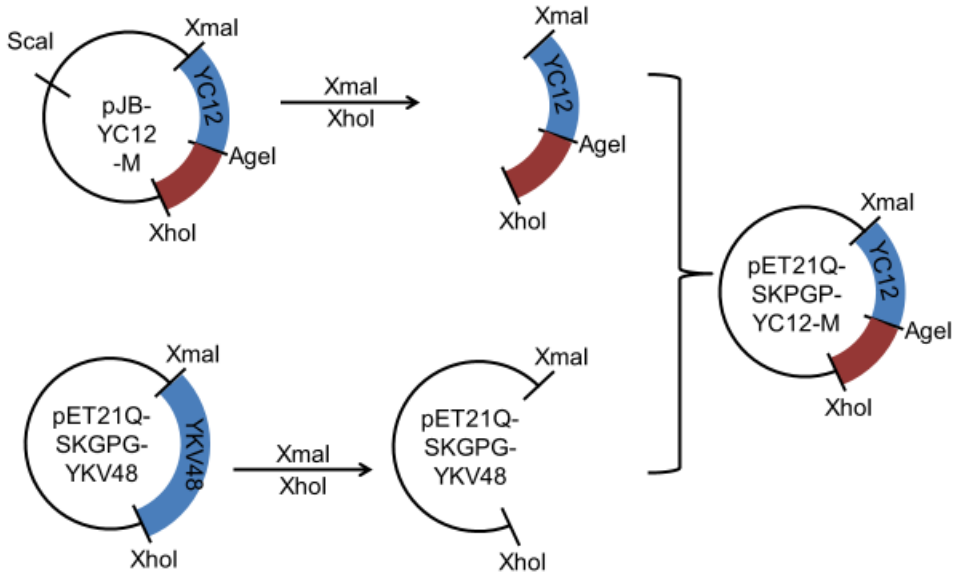


Figure F. 16. Cloning scheme of pET21Q-SKGPG-YC12-M.

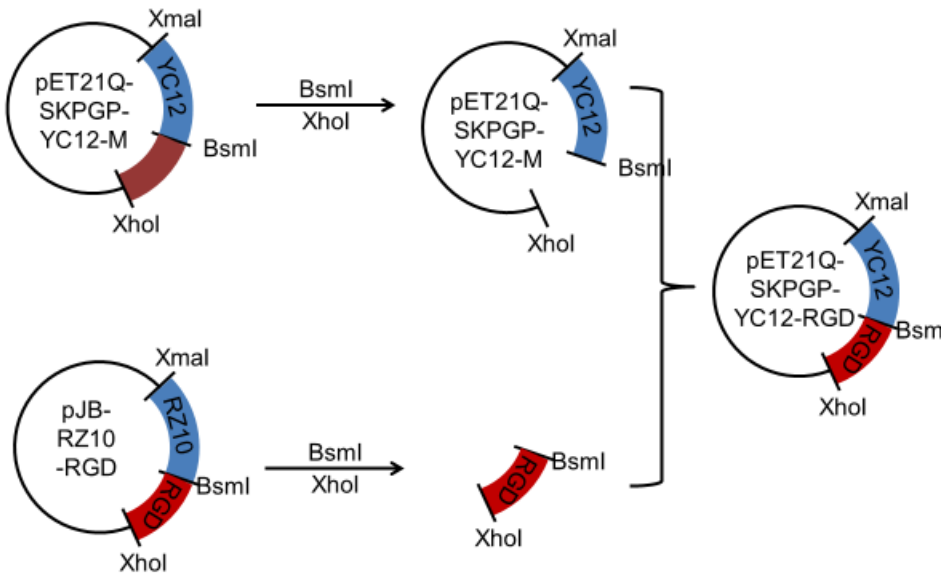


Figure F. 17. Cloning scheme of pET21Q-SKGPG-YC12-RGD.

F. 2 pET28aCYL Expression Vector

```
   1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
  76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
 151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
 226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
 301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
 376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
 451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaat
 526 acattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttat
 601 tcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagt
 676 tccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcc
 751 cctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtt
 826 tatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaac
 901 cgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaa
 976 tcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaata
1051                 XmaI c|ccggg1068
1051                 AvaI c|ycgrg1068
1051 cctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttga
1126 tggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctac
1201 ctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcc
1276 cgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaag
1351 acgtttcccgttgaatatggctcataacacccccttgtattactgtttatgtaagcagacagttttattgttcatg
1426 accaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttga
1501 gatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccg
1576 gatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttcta
1651 gtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtta
1726 ccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg
```

```
1801 cagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac
1876 ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagg
1951 gtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgc
2026 cacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcg
2101 gccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtg
2176 gataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtg
2251         SapI |nnnngaagagc2259
2251 agcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatat
2326 ggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactg
2401 ggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgc
2476 ttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgag
2551 gcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctc
2626 gttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgttt
2701 ggtcactgatgcctccgtgtaaggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgct
2776 cacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggat
2851                                        AfeI agc|gct2892
2851 gcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacaggg
2926 tagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagacttta
3001 cgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgt
3076                                                                PpuMI rg|gwccy3137
3076 tcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacag
3151 gagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggc
3226 gggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgc
3301                                        AfeI agc|gct3341
3301 gctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaa
3376 gaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctct
3451 caagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgc
3526 tttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtat
3601 tgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgag
3676 agagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcggga
3751                                                EcoRV gat|atc3797
3751 tataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcgg
3826 taatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattca
3901 gcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgat
3976 tgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcg
4051 cgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatac
4126 tgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatgg
4201                                          MluI a|cgcgt4243
4201                                BclI t|gatca4229
4201 catcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctt
4276 tacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaa
4351 tcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgc
4426 ccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttt
4501 tcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacat
4576 cgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaa
4651 aggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagccc
4726 agtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccc
4801                                        AfeI agc|gct4842
```

```
4801 ccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttcc
4876                                                SgrAI cr|ccggyg4924
4876 ccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccg
4951              BglII a|gatct4965
4951 gcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaatt
5026      XbaI t|ctaga5031                                  NheI g|ctagc5078
   1                                                M  M  A  S  M  T  G  G  Q  Q
5026 cccctctagaaataattttgtttaactttaagaaggagatataccATGATGGCTAGCATGACTGGTGGACAGCAA
5101                                                                  XhoI c|tcgag5162
5101                                                                  AvaI c|ycgrg5162
5101                                                               EcoRI g|aattc5156
5101                                                          BamHI g|gatcc5150
5101                                                     ScaI agt|act5146
  11 M  G  H  H  H  H  H  H  L  V  P  R  G  S  S  T  G  S  E  F  L  E  *  *
5101 ATGGGTCACCACCACCACCACCACCTGGTTCCGCGTGGTTCTAGTACTGGATCCGAATTCCTCGAGTAATAAagt
5176 cgagtaataaagtcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagt
5251 tggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgc
5326 tgaaaggaggaactatatccggat
```

F. 3 pET21Q Expression Vector

```
   1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
  76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
 151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
 226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
 301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
 376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
 451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaat
 526 acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtat
 601 gagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccaga
 676 aacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacag
 751 cggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
 826 cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
 901      ScaI agt|act906
 901 tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataac
 976 catgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgca
1051 caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcg
1126 tgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttc
1201 ccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
1276 ctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatgg
1351 taagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgc
1426 tgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattt
1501 aaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacg
1576 tgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcg
1651 cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac
1726 tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttagg
1801 ccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag
```

```
1876 tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaac
1951 ggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg
2026 agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg
2101 cacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcg
2176 tcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcct
2251 ggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
2326                                                                  SapI |nnnngaag
agc2392
2326 ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaaga
2401 gcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtaca
2476 atctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgcccc
2551 gacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtga
2626 ccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct
2701 catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaa
2776 gcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccg
2851 tgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactg
2926 atgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaa
3001                         AfeI agc|gct3025
3001 aaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctg
3076 cgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccga
3151 agaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtg
3226                                          PpuMI rg|gwccy3270
3226 attcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgca
3301 cccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaagg
3376 cttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggt
3451                       AfeI agc|gct3474
3451 cctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtg
3526 cggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgag
3601 atcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacc
3676 tgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtt
3751 tttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg
3826 tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtct
3901                              EcoRV gat|atc3930
3901 tcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
3976 cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgt
4051 tgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattta
4126 tgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgaccc
4201 aatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctgg
4276 tcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagc
4351                          MluI a|cgcgt4376
4351            BclI t|gatca4362
4351 ggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccg
4426 cttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgc
4501 gacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
4576 acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctg
4651 gcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggt
4726 ttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
4801 atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggcc
```

```
4876 gttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgc
4951                        AfeI agc|gct4975
4951 caccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggc
5026                              SgrAI cr|ccggyg5057                     BglII a|
gatct5098
5026 gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagat
5101                                                                 XbaI t|ctaga5164
5101 ctcgatcccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataatt
5176                                        NheI g|ctagc5218
5176                                    NcoI c|catgg5213
5176                                  BamHI g|gatcc5209
5176                              NdeI ca|tatg5204
   1                               M  G  S  M  A  S  M  T  G  G  Q  Q  M  G  R  E
5176 ttgtttaactttaagaaggagatatacatATGGGATCCATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGTG
5251                 RsrII cg|gwccg5268                  XhoI c|tcgag5305
5251          PpuMI rg|gwccy5261     SalI g|tcgac5284     AvaI c|ycgrg5305
5251 EcoRI g|aattc5251       SpeI a|ctagt5275      NotI gc|ggccgc5297
  17  F  E  G  S  S  V  R  T  S  S  V  D  K  L  A  A  A  L  E  H  H  H  H  H  H
5251 AATTCGAGGGGTCCTCGGTCCGCACTAGTTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACC
  42  *
5326 ACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcat
5401 aacccctttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
```

F. 4 pET21Q-SKGPG Expression Vector

```
   1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
  76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
 151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
 226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
 301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
 376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
 451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaat
 526 acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtat
 601 gagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccaga
 676 aacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacag
 751 cggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
 826 cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
 901      ScaI agt|act906
 901 tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataac
 976 catgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgca
1051 caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcg
1126 tgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttc
1201 ccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
1276 ctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatgg
1351 taagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgc
1426 tgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattt
1501 aaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacg
```

```
1576 tgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcg
1651 cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac
1726 tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttagg
1801 ccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag
1876 tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaac
1951 gggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg
2026 agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg
2101 cacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcg
2176 tcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcct
2251 ggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
2326                                                                    SapI |nnnngaag
agc2392
2326 ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaaga
2401 gcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtaca
2476 atctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgcccc
2551 gacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtga
2626 ccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct
2701 catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaa
2776 gcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccg
2851 tgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactg
2926 atgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaa
3001                         AfeI agc|gct3025
3001 aaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctg
3076 cgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccga
3151 agaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtg
3226                                              PpuMI rg|gwccy3270
3226 attcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgca
3301 cccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaagg
3376 cttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggt
3451                         AfeI agc|gct3474
3451 cctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtg
3526 cggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgag
3601 atcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacc
3676 tgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtt
3751 tttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg
3826 tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtct
3901                                EcoRV gat|atc3930
3901 tcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
3976 cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgt
4051 tgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattta
4126 tgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgaccc
4201 aatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctgg
4276 tcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagc
4351                           MluI a|cgcgt4376
4351            BclI t|gatca4362
4351 ggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccg
4426 cttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgc
4501 gacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
```

```
4576 acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctg
4651 gcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggt
4726 ttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
4801 atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggcc
4876 gttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgc
4951                        AfeI agc|gct4975
4951 caccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggc
5026                              SgrAI cr|ccggyg5057                    BglII a|
gatct5098
5026 gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagat
5101                                                                  XbaI t|ctaga5164
5101 ctcgatcccgcgaaattaatacgactcactatagggaattgtgagcggataacaattcccctctagaaataatt
5176                                                                  XhoI c|tcgag5239
5176                                                                  AvaI c|ycgrg5239
5176                                                              EcoRI g|aattc5233
5176                           NdeI ca|tatg5204    SalI g|tcgac5224
   1                          H  M  S  K  G  P  G  V  D  K  E  F  L  E  *  *  L
5176 ttgtttaactttaagaaggagatataCATATGAGCAAAGGTCCGGGTGTCGACAAAGAATTCCTCGAGTAATGAC
5251 XhoI c|tcgag5251
5251 AvaI c|ycgrg5251
  18  E  H  H  H  H  H  H  *
5251 TCGAGCACCACCACCACCACCACTGAgatccggctgctaacaaagcccgaaaggaagctgagttggctgctgcca
5326 ccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaa
5401 ctatatccggat
```

F. 5 pET21Q-SKGPG-Cys3 Expression Vector

This vector has Cys residues in between the cloning sites.

```
   1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
  76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
 151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
 226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
 301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
 376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
 451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaat
 526 acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtat
 601 gagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccaga
 676 aacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacag
 751 cggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
 826 cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
 901      ScaI agt|act906
 901 tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataac
 976 catgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgca
1051 caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcg
1126 tgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttc
1201 ccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
```

```
1276 ctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatgg
1351 taagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgc
1426 tgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattt
1501 aaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacg
1576 tgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcg
1651 cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac
1726 tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttagg
1801 ccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag
1876 tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaac
1951 gggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg
2026 agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg
2101 cacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcg
2176 tcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcct
2251 ggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
2326                                                                  SapI |nnnngaag
agc2392
2326 ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaaga
2401 gcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtaca
2476 atctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgcccc
2551 gacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtga
2626 ccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct
2701 catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaa
2776 gcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccg
2851 tgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactg
2926 atgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaa
3001                       AfeI agc|gct3025
3001 aaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctg
3076 cgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccga
3151 agaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtg
3226                                          PpuMI rg|gwccy3270
3226 attcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgca
3301 cccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaagg
3376 cttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggt
3451                       AfeI agc|gct3474
3451 cctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtg
3526 cggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgag
3601 atcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacc
3676 tgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtt
3751 tttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg
3826 tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtct
3901                             EcoRV gat|atc3930
3901 tcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
3976 cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgt
4051 tgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattta
4126 tgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgaccc
4201 aatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctgg
4276 tcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagc
4351                          MluI a|cgcgt4376
```

```
    4351           BclI t|gatca4362
    4351 ggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccg
    4426 cttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgc
    4501 gacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
    4576 acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctg
    4651 gcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggt
    4726 ttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
    4801 atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggcc
    4876 gttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgc
    4951                       AfeI agc|gct4975
    4951 caccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggc
    5026                             SgrAI cr|ccggyg5057                    BglII a|
gatct5098
    5026 gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagat
    5101                                                                XbaI t|ctaga5164
    5101 ctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatt
    5176                                                                  EcoRI g|aattc5
242
    5176                                                  XmaI c|ccggg5232
    5176                          NdeI ca|tatg5204          AvaI c|ycgrg5232
       1                           M  S  K  G  P  G  V  D  C  P  G  C  E  F  K  K
    5176 ttgtttaactttaagaaggagatatacatATGAGCAAAGGTCCGGGTGTCGATTGCCCGGGCTGTGAATTCAAAA
    5251   AgeI a|ccggt5253        SpeI a|ctagt5278
      17   P  V  C  G  G  G  G  S  T  S  *
    5251 AACCGGTGTGCGGCGGCGGCGGTTCTACTAGTTAAttcgagcaccaccaccaccaccactgagatccggctgcta
    5326 acaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctcta
    5401 aacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
```

F. 6 pET28F Expression Vector

This vector has a 3xFLAG tag at the N-terminus.

```
       1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
      76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
     151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
     226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
     301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
     376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
     451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaat
     526 acattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttat
     601 tcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagt
     676 tccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcc
     751 cctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtt
     826 tatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaac
     901 cgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaa
     976 tcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaata
    1051                 XmaI c|ccggg1068
```

```
1051                AvaI c|ycgrg1068
1051 cctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttga
1126 tggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctac
1201 ctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcc
1276 cgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaag
1351 acgtttcccgttgaatatggctcataacacccccttgtattactgtttatgtaagcagacagttttattgttcatg
1426 accaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttga
1501 gatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccg
1576 gatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttcta
1651 gtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtta
1726 ccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg
1801 cagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac
1876 ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagg
1951 gtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgc
2026 cacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcg
2101 gccttttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtg
2176 gataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtg
2251         SapI |nnnngaagagc2259
2251 agcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatat
2326 ggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactg
2401 ggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgc
2476 ttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgag
2551 gcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctc
2626 gttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgttt
2701 ggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgct
2776 cacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggat
2851                                      AfeI agc|gct2892
2851 gcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacaggg
2926 tagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagacttta
3001 cgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgt
3076                                                              PpuMI rg|gwccy3137
3076 tcgctcgcgtatcggtgattcattctgctaaccagtaaggcaacccgccagcctagccgggtcctcaacgacag
3151 gagcacgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggc
3226 gggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgc
3301                                      AfeI agc|gct3341
3301 gctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaa
3376 gaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctct
3451 caagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgc
3526 tttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtat
3601 tgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgag
3676 agagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcggga
3751                                            EcoRV gat|atc3797
3751 tataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcgg
3826 taatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattca
3901 gcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgat
3976 tgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcg
4051 cgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatac
4126 tgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatgg
```

```
    4201                                          MluI a|cgcgt4243
    4201                              BclI t|gatca4229
    4201 catcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctt
    4276 tacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaa
    4351 tcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgc
    4426 ccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttt
    4501 tcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacat
    4576 cgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaa
    4651 aggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagccc
    4726 agtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtccc
    4801                                       AfeI agc|gct4842
    4801 ccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttcc
    4876                                                 SgrAI cr|ccggyg4924
    4876 ccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccg
    4951              BglII a|gatct4965
    4951 gcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaatt
    5026      XbaI t|ctaga5031
       1                                              M  M  A  S  H  H  H  H  H  H
    5026 cccctctagaaataattttgtttaactttaagaaggagatataccATGATGGCTAGTCATCACCACCACCACCAC
    5101                                       EcoRV gat|atc5140            ScaI agt|ac
t5170
      11 D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  S  T  H
    5101 GACTACAAAGACCATGATGGCGATTATAAAGATCACGATATCGATTATAAAGATGACGATGACAAAAGTACTCAT
    5176                XhoI c|tcgag5192
    5176                AvaI c|ycgrg5192
    5176          EcoRI g|aattc5186
    5176     SpeI a|ctagt5180
      36 K  T  S  E  F  L  E  *  *
    5176 AAAACTAGTGAATTCCTCGAGTAATAAagtcgagtaataaagtcgagcaccaccaccaccaccactgagatccgg
    5251 ctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttgggg
    5326 cctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat
```

F. 7 pJB-SgrAI-CYL Cloning Vector

This vector was a modification from pJB with different sites and frames.

```
       1 tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaag
      76 cggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatg
     151                                  NdeI ca|tatg185
     151 cggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaat
     226 accgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctat
     301 tacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgac
     376                                  PpuMI rg|gwccy410   AgeI a|ccggt430  XhoI c|tc
gag447
     376                           NheI g|ctagc403     RsrII cg|gwccg423       AvaI c|yc
grg447
     376                     EcoRI g|aattc397   SalI g|tcgac416      SgrAI cr|ccggyg437
```

```
   1                    E  F  A  S  G  T  R  R  P  D  R  T  G  H  R  R  G  S  R
 376 gttgtaaaacgacggccagtGAATTCGCTAGCGGGACCCGTCGACCGGACCGTACCGGTCACCGGCGGGGCTCGA
 451               SalI g|tcgac465
 451         XbaI t|ctaga459
 451   BamHI g|gatcc453
  20   D  P  L  E  S  T  C  R  H  A  S  L  A  *
 451 GGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAAtcatggtcatagctgtttcctgtgtgaaatt
 526 gttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtga
 601 gctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaat
 676                                                  SapI gctcttcn|nnn726
 676 gaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgc
 751 gctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg
 826 ataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgt
 901 ttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacag
 976 gactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccg
1051 gatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgg
1126 tgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccccgttcagcccgaccgctgcgccttatccggta
1201 actatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagca
1276 gagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtat
1351 ttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacca
1426 ccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctt
1501 tgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaa
1576 aaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttt
1651 ggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttg
1726 cctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgc
1801 gagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtc
1876 ctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaata
1951 gtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagct
2026 ccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctc
2101 cgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactg
2176                                        ScaI agt|act2215
2176 tcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc
2251 gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatca
2326 ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactc
2401 gtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatg
2476 ccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagca
2551 tttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgc
2626 gcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggc
2701 gtatcacgaggccctttcgtc
```

F. 8 pET21b Expression Vector

```
   1 tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctac
  76 acttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccg
 151 tcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
 226 ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
 301 ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagg
```

```
 376 gattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaat
 451 attaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaat
 526 acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtat
 601 gagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccaga
 676 aacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacag
 751 cggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
 826 cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
 901      ScaI agt|act906
 901 tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataac
 976 catgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgca
1051 caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcg
1126 tgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttc
1201 ccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctgg
1276 ctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatgg
1351 taagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgc
1426 tgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattt
1501 aaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacg
1576 tgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcg
1651 cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaac
1726 tctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttagg
1801 ccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccag
1876 tggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaac
1951 gggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatg
2026 agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg
2101 cacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcg
2176 tcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcct
2251 ggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgc
2326                                                                    SapI |nnnngaag
agc2392
2326 ctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaaga
2401 gcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtaca
2476 atctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgcccc
2551 gacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtga
2626 ccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct
2701 catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaa
2776 gcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccg
2851 tgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactg
2926 atgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaa
3001                       AfeI agc|gct3025
3001 aaatcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctg
3076 cgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccga
3151 agaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtg
3226                                          PpuMI rg|gwccy3270
3226 attcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgca
3301 cccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaagg
3376 cttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggt
3451                       AfeI agc|gct3474
3451 cctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataagtg
```

```
    3526 cggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgag
    3601 atcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacc
    3676 tgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtt
    3751 tttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg
    3826 tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtct
    3901                              EcoRV gat|atc3930
    3901 tcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
    3976 cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgt
    4051 tgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattta
    4126 tgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgaccc
    4201 aatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctgg
    4276 tcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagc
    4351                         MluI a|cgcgt4376
    4351              BclI t|gatca4362
    4351 ggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccg
    4426 cttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgc
    4501 gacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
    4576 acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctg
    4651 gcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggt
    4726 ttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
    4801 atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggcc
    4876 gttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgc
    4951                        AfeI agc|gct4975
    4951 caccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggc
    5026                              SgrAI cr|ccggyg5057                    BglII a|
gatct5098
    5026 gatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagat
    5101                                                                XbaI t|ctaga5164
    5101 ctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatt
    5176                                NheI g|ctagc5209                    EcoRI g|a
attc5247
    5176                           NdeI ca|tatg5204                   BamHI g|gatcc52
41
       1                            M  A  S  M  T  G  G  Q  Q  M  G  R  D  P  N  S
    5176 ttgtttaactttaagaaggagatatacatATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGGGATCCGAATT
    5251                              XhoI c|tcgag5281
    5251                       NotI gc|ggccgc5273
    5251          SalI g|tcgac5260     AvaI c|ycgrg5281
      17   S  S  V  D  K  L  A  A  A  L  E  H  H  H  H  H  H  *
    5251 CGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAgatccggctgctaacaaag
    5326 cccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacggg
    5401 tcttgaggggttttttgctgaaag
```

The sequences below are only the encoding part of the indicated plasmids. For their backbone sequences, please refer to their corresponding empty vectors listed above.

F. 9 pUC57Simple-SpeI-high aff ZE-NheI-AleI-linker-COMP-XhoI

This plasmid was synthesized by Genscript. The whole backbone sequence of PUC57Simple can be found on the Genscript website. Note that the sequence of SpeI-high aff Ze-NheI and the sequence of AleI-linker COMP-XhoI are in different frames.

```
      1  SpeI a|ctagt2
      1   *  S  G  N  R  G  S  C  S  G  T  G  K  H  G  A  G  D  R  S  C  G  T  G  T
      1  L  V  W  K  S  R  Q  L  L  W  N  R  K  T  R  R  W  R  P  K  L  R  N  W  N
      1 T  S  L  E  I  E  A  A  A  L  E  Q  E  N  T  A  L  E  T  E  V  A  E  L  E
      1 ACTAGTCTGGAAATCGAGGCAGCTGCTCTGGAACAGGAAAACACGGCGCTGGAGACCGAAGTTGCGGAACTGGAA
     76                                                                        AleI cacn
n|nngtg147
     76                                                                NheI g|ctagc137
     26   G  S  A  A  S  G  E  H  R  V  S  V  Q  N  P  L  W  S  A  G  *  P  P  Q  W
     26  R  K  C  S  V  W  R  T  S  C  L  S  T  K  P  V  M  V  R  W  L  A  T  A  V
     26 Q  E  V  Q  R  L  E  N  I  V  S  Q  Y  K  T  R  Y  G  P  L  A  S  H  R  S
     76 CAGGAAGTGCAGCGTCTGGAGAACATCGTGTCTCAGTACAAAACCCGTTATGGTCCGCTGGCTAGCCACCGCAGT
     51   G  G  G  G  S  G  G  G  G  S  A  P  Q  M  L  R  E  L  Q  E  T  N  A  A  L
     51  G  R  R  R  F  W  R  R  W  F  C  S  A  D  A  A  *  T  A  G  N  Q  C  C  S
     51 G  A  A  A  V  L  A  A  V  V  L  L  R  R  C  C  V  N  C  R  K  P  M  L  L
    151 GGGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTGCTCCGCAGATGCTGCGTGAACTGCAGGAAACCAATGCTGCTC
     76   Q  D  V  R  E  L  L  R  Q  A  V  E  Q  E  I  T  F  L  Q  N  T  V  M  E  S
     76  A  G  R  S  *  T  A  A  S  G  S  G  T  G  N  H  L  P  A  K  H  R  Y  G  I
     76 C  R  T  F  V  N  C  C  V  R  Q  W  N  R  K  S  P  S  C  K  T  P  L  W  N
    226 TGCAGGACGTTCGTGAACTGCTGCGTCAGGCAGTGGAACAGGAAATCACCTTCCTGCAAAACACCGTTATGGAAT
    301                 XhoI c|tcgag319
    301                 AvaI c|ycgrg319
    301    MluI a|cgcgt305
    101   D  A  S  *  *  L  E
    101  *  R  V  L  I  T  R
    101 L  T  R  L  N  N  S
    301 CTGACGCGTCTTAATAACTCGAG
```

F. 10 pUC57Simple-LZE-BMP2GC

This plasmid was synthesized by Genscript.

```
      1        BamHI g|gatcc8
      1  EcoRI g|aattc2                                SpeI a|ctagt50
      1 E  F  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  I  T  I  R  A  A  F
      1 GAATTCGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTATCACCATCCGTGCCGCGTTC
     26 L  E  Q  E  N  T  A  L  R  T  E  A  A  E  L  E  Q  E  V  G  R  L  E  N  I
     76 CTGGAACAGGAAAACACCGCCCTGCGTACCGAAGCCGCTGAGCTCGAACAGGAGGTCGGCCGTCTGGAAAACATT
    151                                 NheI g|ctagc185                  AgeI a|ccg
gt221
     51 V  S  Q  Y  K  T  R  Y  G  P  L  A  S  G  G  G  G  S  G  G  G  G  S  T  G
```

```
151 GTTAGCCAGTACAAAACCCGTTATGGTCCGCTGGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGT
226                                                                 AvaI c|ycgrg284
 76 K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  G  G  G  G  S
226 AAGATCCCAAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTCGGGGGTGGCGGTTCT
301                        XbaI t|ctaga325
301                  RsrII cg|gwccg319
301            PpuMI rg|gwccy312
101 C  *  *  G  S  S  V  R  L
301 TGCTAATAAGGGTCCTCGGTCCGTCTAGA
```

F. 11 pUC57Simple-COMPm-RGDGC

This plasmid was synthesized by Genscript.

```
  1        BamHI g|gatcc8
  1  EcoRI g|aattc2                                  SpeI a|ctagt50
  1 E  F  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  A  P  Q  M  L  R  E
  1 GAATTCGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTGCTCCGCAGATGCTGCGTGAA
 26 L  Q  E  T  N  A  A  L  Q  D  V  R  E  L  L  R  Q  A  V  Q  E  I  T  F  L
 76 CTGCAGGAAACCAATGCTGCTCTGCAGGACGTTCGTGAACTGCTGCGTCAGGCAGTGCAGGAAATCACCTTCCTG
151                              NheI g|ctagc182
151                       MluI a|cgcgt174                              AgeI a|ccggt2
18
 51 Q  N  T  V  M  E  S  D  A  S  A  S  G  G  G  G  S  G  G  G  G  S  T  G  T
151 CAAAACACCGTTATGGAATCTGACGCGTCTGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTACC
 76 I  T  V  Y  A  V  T  G  R  G  D  S  P  A  S  S  K  P  I  G  G  G  G  S  C
226 ATCACCGTGTATGCGGTTACTGGTCGTGGCGACTCTCCGGCATCTTCTAAACCGATTGGTGGCGGTGGCTCTTGT
301                      XbaI t|ctaga322
301                 RsrII cg|gwccg316
301          PpuMI rg|gwccy309
101 *  *  G  S  S  V  R  L
301 TAATAAGGGTCCTCGGTCCGTCTAGA
```

F. 12 pJB-RZY10

```
  1 tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaag
 76 cggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatg
151                                  NdeI ca|tatg185
151 cggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaat
226 accgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctat
301 tacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgac
376                                  PpuMI rg|gwccy410
376                             SalI g|tcgac403
376                       EcoRI g|aattc397       AfeI agc|gct420
  1                     E  F  V  D  G  T  L  S  A  Q  T  P  S  S  K  Q  Y  G  A
376 gttgtaaaacgacggccagtGAATTCGTCGACGGGACCCTCAGCGCTCAGACCCCTTCTTCCAAGCAGTATGGCG
 20   P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A
```

```
451 CTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGG
 45   Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T
526 CGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGA
 70   P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S
601 CACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGA
 95   S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q
676 GCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCC
751                                                  XbaI t|ctaga802
751                                        XhoI c|tcgag790
751                          RsrII cg|gwccg774     BamHI g|gatcc796
751               AleI cacnn|nngtg765      AvaI c|ycgrg790   SalI g|tcgac808
120   Y  G  A  P  K  W  A  D  R  G  M  R  L  E  G  S  S  R  V  D
751 AGTACGGTGCACCGAAGTGGGCGGACCGTGGAATGCGGCTCGAGGGATCCTCTAGAGTCGACctgcaggcatgca
826 agcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacga
901 gccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactg
976 cccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttg
1051                  SapI gctcttcn|nnn1069
1051 cgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatca
1126 gctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc
1201 cagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcat
1276 cacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctgga
1351 agctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagc
1426 gtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtg
1501 cacgaacccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacac
1576 gacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttc
1651 ttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacc
1726 ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaag
1801 cagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtgg
1876 aacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaa
1951 aaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgag
2026 gcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgata
2101 cgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca
2176 gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctatt
2251 aattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggc
2326 atcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatga
2401 tcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtg
2476 ttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgact
2551        ScaI agt|act2558
2551 ggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgg
2626 gataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctca
2701 aggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttact
2776 ttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaa
2851 tgttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatac
2926 atatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtc
3001 taagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc
```

F. 13 pJB-SgrAI-CYL-QEQV12

```
   1 tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaag
  76 cggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatg
 151                                  NdeI ca|tatg185
 151 cggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaat
 226 accgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctat
 301 tacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgac
 376                                   XmaI c|ccggg411
 376                                   AvaI c|ycgrg411
 376                             NheI g|ctagc403
 376                       EcoRI g|aattc397
   1                     E  F  A  S  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G
 376 gttgtaaaacgacggccagtGAATTCGCTAGCGTCCCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGG
  20   V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G
 451 GCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTG
  45   V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G
 526 GCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGG
  70   V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G
 601 GTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAG
  95   V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G
 676 GTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGG
 120   V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G
 751 GCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTG
 145   V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G
 826 GCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGG
 170   V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G
 901 GTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAG
 195   V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G
 976 GTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGG
 220   V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G
1051 GCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTG
1126                              BamHI g|gatcc1154
1126                        XhoI c|tcgag1148
1126              SgrAI cr|ccggyg1138   XbaI t|ctaga1160
1126       AgeI a|ccggt1131 AvaI c|ycgrg1148  SalI g|tcgac1166
 245   V  P  V  T  G  G  A  R  G  I  L  *  S  R  P
1126 GCGTACCGGTCACCGGCGGGGCTCGAGGGATCCTCTAGAGTCGACCTgcaggcatgcaagcttggcgtaatcatg
1201 gtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtg
1276 taaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcggg
1351 aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttc
1426  SapI gctcttcn|nnn1427
1426 cgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggt
1501 aatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaa
1576 ccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctc
1651 aagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctc
1726 tcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatag
1801 ctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttca
```

```
1876 gcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggc
1951 agcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaa
2026 ctacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttgg
2101 tagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgcaagcagcagattacgcgcag
2176 aaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgtta
2251 agggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatc
2326 aatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgat
2401 ctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatc
2476 tggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagc
2551 cggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagc
2626 tagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctc
2701 gtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaa
2776 aaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttat
2851                                                                    ScaI agt|act291
6

2851 ggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaa
2926 gtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccaca
3001 tagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgtt
3076 gagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgg
3151 gtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatact
3226 cttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtattta
3301 gaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattat
3376 catgacattaacctataaaaataggcgtatcacgaggccctttcgtc
```

F. 14 pJB-YC12-Tetra

```
   1 tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaag
  76 cggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatg
 151                                  NdeI ca|tatg185
 151 cggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaat
 226 accgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctat
 301 tacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgac
 376                                       XmaI c|ccggg417
 376                                  PpuMI rg|gwccy410
 376                            SalI g|tcgac403
 376                      EcoRI g|aattc397    AvaI c|ycgrg417
   1                    E  F  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G
 376 gttgtaaaacgacggccagtGAATTCGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGG
  20   V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G
 451 GTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGG
  45   K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G
 526 GTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGG
  70   Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G
 601 GCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGG
  95   V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G
 676 GTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGG
 120   K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G
```

```
 751 GTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGG
 145   C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G
 826 GTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGG
 170   V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G
 901 GTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGG
 195   K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G
 976 GTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGG
 220   Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G
1051 GCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGG
 245   V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G
1126 GTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGG
 270   K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G
1201 GTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGG
 295   C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G
1276 GTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGG
 320   V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G
1351 GTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGG
1426                                                                    AgeI a|cc
ggt1497
 345   K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V
1426 GTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGG
1501      RsrII cg|gwccg1506
 370   A  D  R  G  M  R  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S
1501 TAGCGGACCGTGGAATGCGGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCT
 395   T  L  Y  L  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L
1576 CTACGCTGTACCTGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGC
 420   Y  L  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L
1651 TGTACCTGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACC
1726                                                                    XhoI c|t
cgag1798
1726                                                                    AvaI c|y
cgrg1798
 445   L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  L  E
1726 TGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTGCTCG
1801               SalI g|tcgac1816
1801         XbaI t|ctaga1810
1801    BamHI g|gatcc1804
 470   G  S  S  R  V  D
1801 AGGGATCCTCTAGAGTCGACctgcaggcatgcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaat
1876 tgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtg
1951 agctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaa
2026                                                  SapI gctcttcn|nnn2077
2026 tgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctg
2101 cgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggg
2176 gataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcg
2251 tttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgaca
2326 ggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttacc
2401 ggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcg
```

```
2476 gtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggt
2551 aactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc
2626 agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagta
2701 tttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacc
2776 accgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcct
2851 ttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatca
2926 aaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaact
3001 tggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagtt
3076 gcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccg
3151 cgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggt
3226 cctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaat
3301 agtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagc
3376 tccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcct
3451 ccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttact
3526                                        ScaI agt|act3566
3526 gtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcgg
3601 cgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatc
3676 attggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccact
3751 cgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaat
3826 gccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagc
3901 atttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggggttccg
3976 cgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaatagg
4051 cgtatcacgaggccctttcgtc
```

F. 15 pET28aCYL-RZY10

```
   1        NheI g|ctagc8
   1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  L  V  P  R  G  S  S
   1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCTGGTTCCGCGTGGTTCTAGT
  26 A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q
  76 GCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAG
  51 T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P
 151 ACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCT
  76 S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S
 226 TCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCC
 101 K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q
 301 AAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAG
 376                                                                    XhoI c|tcg
ag446
 376                                                        RsrII cg|gwccg430
 376                                                AleI cacnn|nngtg421     AvaI c|ycg
rg446
 126 Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  K  W  A  D  R  G  M  R  L  E
 376 TATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGAAGTGGGCGGACCGTGGAATGCGGCTCGAG
 151 *  *
 451 TAATAA
```

F. 16 pET21Q-SKGPG-RZY10

```
   1                          PpuMI rg|gwccy27
   1                  SalI g|tcgac20   AfeI agc|gct37
   1 M  S  K  G  P  G  V  D  G  T  L  S  A  Q  T  P  S  S  K  Q  Y  G  A  P  A
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCAGCGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCG
  26 Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T
  76 CAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACA
  51 P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S
 151 CCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGC
  76 S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q
 226 AGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAG
 101 Y  G  A  P  A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G
 301 TACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGT
 376                              XhoI c|tcgag407
 376               RsrII cg|gwccg391
 376       AleI cacnn|nngtg382      AvaI c|ycgrg407
 126 A  P  K  W  A  D  R  G  M  R  L  E  *  *
 376 GCACCGAAGTGGGCGGACCGTGGAATGCGGCTCGAGTAATGA
```

F. 17 pET21b-RZY10-HZR_HZE

```
   1     NheI g|ctagc5                                                    AfeI agc
|gct73
   1 M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  L  V  P  R  G  S  S  A
   1 ATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCTGGTTCCGCGTGGTTCTAGCGCT
  26 Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T
  76 CAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACC
  51 P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S
 151 CCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCT
  76 S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K
 226 TCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAG
 101 Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q  Y
 301 CAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAGTAT
 376                                        AleI cacnn|nngtg418
 126 G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  Q  W  G  G  G  G  S  G  G  G  G
 376 GGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGCAGTGGGGCGGCGGCGGAAGCGGCGGCGGCGGG
 151 S  L  E  I  R  A  A  A  L  R  R  R  N  T  A  L  R  T  R  V  A  E  L  R  Q
 451 AGCCTGGAAATCCGTGCAGCCGCGCTCCGCCGTCGCAACACAGCCCTGCGGACCCGTGTTGCGGAGCTGCGCCAA
 526 MluI a|cgcgt526                                                  BglII a|gatct590
 176 R  V  Q  R  L  R  N  E  V  S  Q  Y  E  T  R  Y  G  P  L  *  *
 526 CGCGTTCAGCGTCTGCGCAATGAAGTAAGCCAGTACGAAACCCGCTATGGCCCGTTGTAATAAagatctcatcac
 601 catcaccatcactaaccttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccat
 676                                                        NheI g|ctagc735
 676 ctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcatgactggtgg
 751                       EcoRI g|aattc773
 751                BamHI g|gatcc767                                  BamHI g|gatcc815
```

```
197                                                 M  R  G  G  G  G  S  L  E
751 acagcaaatgggtcgggatccgaattcattaaagaggagaaattaactATGAGAGGAGGTGGCGGATCCCTGGAA
206 I  E  A  A  A  L  E  Q  E  N  T  A  L  E  T  E  V  A  E  L  E  Q  E  V  Q
826 ATCGAAGCGGCGGCGCTGGAACAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGTTCAG
231 R  L  E  N  I  V  S  Q  Y  R  T  R  Y  G  P  L  *  *
901 CGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTATGGTCCGCTGTAATAA
```

F. 18 pET28aCYL-RZY10-LZR

```
  1        NheI g|ctagc8
  1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  L  V  P  R  G  S  S
  1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCTGGTTCCGCGTGGTTCTAGT
 26 A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q
 76 GCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAG
 51 T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P
151 ACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCT
 76 S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S
226 TCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCC
101 K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q
301 AAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAG
376                                          AleI cacnn|nngtg421
126 Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  Q  W  G  G  G  G  S  G  G  G
376 TATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGCAGTGGGGCGGAGGGGGCAGCGGCGGCGGC
151 G  S  L  E  I  R  A  A  F  L  E  Q  E  N  T  A  L  R  T  R  A  A  E  L  R
451 GGCAGCTTAGAGATTCGTGCCGCCTTTTTGGAACAGGAAAATACGGCCCTGCGCACTCGCGCAGCGGAACTGCGC
176 Q  R  V  G  R  L  R  N  I  V  S  Q  Y  E  T  R  Y  G  P  L  *  *
526 CAGCGCGTTGGTCGTCTGCGCAACATCGTCTCCCAATATGAAACCCGCTACGGCCCGCTGTAATAA
```

F. 19 pET28aCYL-RZY10-COMPm

```
  1        NheI g|ctagc8
  1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  L  V  P  R  G  S  S
  1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCTGGTTCCGCGTGGTTCTAGT
 26 A  Q  T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q
 76 GCTCAGACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAG
 51 T  P  S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P
151 ACCCCTTCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCT
 76 S  S  K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S
226 TCTTCCAAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCC
101 K  Q  Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  A  Q  T  P  S  S  K  Q
301 AAGCAGTATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGGCTCAGACCCCTTCTTCCAAGCAG
376                                          AleI cacnn|nngtg421
126 Y  G  A  P  A  Q  T  P  S  S  Q  Y  G  A  P  Q  W  G  G  G  G  S  G  G  G
376 TATGGCGCTCCGGCGCAGACACCGAGCAGCCAGTACGGTGCACCGCAGTGGGGCGGCGGCGGTTCTGGCGGCGGT
151 G  S  A  P  Q  M  L  R  E  L  Q  E  T  N  A  A  L  Q  D  V  R  E  L  L  R
451 GGTTCTGCTCCGCAGATGCTGCGTGAACTGCAGGAAACCAATGCTGCTCTGCAGGACGTTCGTGAACTGCTGCGT
```

```
526                                                MluI a|cgcgt576
176 Q  A  V  Q  E  I  T  F  L  Q  N  T  V  M  E  S  D  A  S  *  *
526 CAGGCAGTGCAGGAAATCACCTTCCTGCAAAACACCGTTATGGAATCTGACGCGTCTTAATAA
```

F. 20 pET21Q-HZE-BMP2GC

```
  1     BamHI g|gatcc5                          SpeI a|ctagt47
  1 M  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  L  E  I  E  A  A  A  L
  1 ATGGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTCTGGAAATCGAGGCAGCTGCTCTG
 26 E  Q  E  N  T  A  L  E  T  E  V  A  E  L  E  Q  E  V  Q  R  L  E  N  I  V
 76 GAACAGGAAAACACGGCGCTGGAGACCGAAGTTGCGGAACTGGAACAGGAAGTGCAGCGTCTGGAGAACATCGTG
151                              NheI g|ctagc182                   AgeI a|ccggt2
18
 51 S  Q  Y  K  T  R  Y  G  P  L  A  S  G  G  G  G  S  G  G  G  G  S  T  G  K
151 TCTCAGTACAAAACCCGTTATGGTCCGCTGGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTAAG
226                                                          AvaI c|ycgrg281
 76 I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  G  G  G  G  S  C
226 ATCCCAAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTCGGGGGTGGCGGTTCTTGC
101 *  *
301 TAATAA
```

F. 21 pET21Q-HZE-RGDGC

```
  1     BamHI g|gatcc5                          SpeI a|ctagt47
  1 M  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  L  E  I  E  A  A  A  L
  1 ATGGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTCTGGAAATCGAGGCAGCTGCTCTG
 26 E  Q  E  N  T  A  L  E  T  E  V  A  E  L  E  Q  E  V  Q  R  L  E  N  I  V
 76 GAACAGGAAAACACGGCGCTGGAGACCGAAGTTGCGGAACTGGAACAGGAAGTGCAGCGTCTGGAGAACATCGTG
151                              NheI g|ctagc182                   AgeI a|ccggt2
18
 51 S  Q  Y  K  T  R  Y  G  P  L  A  S  G  G  G  G  S  G  G  G  G  S  T  G  T
151 TCTCAGTACAAAACCCGTTATGGTCCGCTGGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTACC
 76 I  T  V  Y  A  V  T  G  R  G  D  S  P  A  S  S  K  P  I  G  G  G  G  S  C
226 ATCACCGTGTATGCGGTTACTGGTCGTGGCGACTCTCCGGCATCTTCTAAACCGATTGGTGGCGGTGGCTCTTGT
101 *  *
301 TAATAA
```

F. 22 pET21Q-LZE-BMP2GC

```
  1     BamHI g|gatcc5                          SpeI a|ctagt47
  1 M  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  I  T  I  R  A  A  F  L
  1 ATGGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTATCACCATCCGTGCCGCGTTCCTG
 26 E  Q  E  N  T  A  L  R  T  E  A  A  E  L  E  Q  E  V  G  R  L  E  N  I  V
 76 GAACAGGAAAACACCGCCCTGCGTACCGAAGCCGCTGAGCTCGAACAGGAGGTCGGCCGTCTGGAAAACATTGTT
151                              NheI g|ctagc182                   AgeI a|ccggt2
18
```

```
 51 S  Q  Y  K  T  R  Y  G  P  L  A  S  G  G  G  G  S  G  G  G  G  S  T  G  K
151 AGCCAGTACAAAACCCGTTATGGTCCGCTGGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTAAG
226                                                          AvaI c|ycgrg281
 76 I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  G  G  G  G  S  C
226 ATCCCAAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTCGGGGGTGGCGGTTCTTGC
101 *  *
301 TAATAA
```

F. 23 pET21Q-COMPm-BMP2GC

```
  1     BamHI g|gatcc5                              SpeI a|ctagt47
  1 M  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  A  P  Q  M  L  R  E  L
  1 ATGGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTGCTCCGCAGATGCTGCGTGAACTG
 26 Q  E  T  N  A  A  L  Q  D  V  R  E  L  L  R  Q  A  V  Q  E  I  T  F  L  Q
 76 CAGGAAACCAATGCTGCTCTGCAGGACGTTCGTGAACTGCTGCGTCAGGCAGTGCAGGAAATCACCTTCCTGCAA
151                             NheI g|ctagc179
151                     MluI a|cgcgt171                              AgeI a|ccggt215
 51 N  T  V  M  E  S  D  A  S  A  S  G  G  G  G  S  G  G  G  G  S  T  G  K  I
151 AACACCGTTATGGAATCTGACGCGTCTGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTAAGATC
226                                                        AvaI c|ycgrg278
 76 P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  G  G  G  G  S  C  *
226 CCAAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTCGGGGGTGGCGGTTCTTGCTAA
101 *
301 TAA
```

F. 24 pET21Q-COMPm-RGDGC

```
  1     BamHI g|gatcc5                              SpeI a|ctagt47
  1 M  G  S  H  H  H  H  H  H  L  V  P  R  G  S  T  S  A  P  Q  M  L  R  E  L
  1 ATGGGATCCCATCACCACCACCACCACCTGGTTCCGCGTGGTTCTACTAGTGCTCCGCAGATGCTGCGTGAACTG
 26 Q  E  T  N  A  A  L  Q  D  V  R  E  L  L  R  Q  A  V  Q  E  I  T  F  L  Q
 76 CAGGAAACCAATGCTGCTCTGCAGGACGTTCGTGAACTGCTGCGTCAGGCAGTGCAGGAAATCACCTTCCTGCAA
151                             NheI g|ctagc179
151                     MluI a|cgcgt171                              AgeI a|ccggt215
 51 N  T  V  M  E  S  D  A  S  A  S  G  G  G  G  S  G  G  G  G  S  T  G  T  I
151 AACACCGTTATGGAATCTGACGCGTCTGCTAGCGGCGGCGGCGGTTCTGGCGGCGGTGGTTCTACCGGTACCATC
 76 T  V  Y  A  V  T  G  R  G  D  S  P  A  S  S  K  P  I  G  G  G  G  S  C  *
226 ACCGTGTATGCGGTTACTGGTCGTGGCGACTCTCCGGCATCTTCTAAACCGATTGGTGGCGGTGGCTCTTGTTAA
101 *
301 TAA
```

F. 25 pET28aRW-YKV72

```
  1        NheI g|ctagc8
  1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  D  D  D  D  K  L
  1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCATGATGATGATGATAAACTC
```

```
  76            XmaI c|ccggg88
  76            AvaI c|ycgrg88
  76      PpuMI rg|gwccy81
  26 D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
  76 GACGGGACCCTCCCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
  51 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 151 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
  76 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 226 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
 101 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 301 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
 126 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 376 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 151 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 451 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 176 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 526 GGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
 201 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 601 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
 226 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 676 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
 251 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 751 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
 276 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 826 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 301 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 901 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 326 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 976 GGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
 351 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
1051 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
1126                                                                  XhoI c|tcgag1
193
1126                                                  RsrII cg|gwccg1177
1126                                        AgeI a|ccggt1168        AvaI c|ycgrg1
193
 376 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D  R  G  M  R  L  E  *
1126 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGTAGCGGACCGTGGAATGCGGCTCGAGTAA
 401 *
1201 TAA
```

F. 26 pET28aRW-YKV96

```
   1        NheI g|ctagc8
   1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  H  D  D  D  D  K  L
   1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCATGATGATGATGATAAACTC
  76            XmaI c|ccggg88
```

```
  76            AvaI c|ycgrg88
  76      PpuMI rg|gwccy81
  26 D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
  76 GACGGGACCCTCCCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
  51 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 151 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
  76 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 226 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
 101 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 301 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
 126 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 376 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 151 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 451 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 176 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 526 GGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
 201 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 601 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
 226 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 676 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
 251 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 751 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
 276 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 826 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 301 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 901 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 326 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 976 GGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
 351 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
1051 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
 376 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
1126 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
 401 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
1201 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
 426 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
1276 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 451 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
1351 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 476 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
1426 GGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCG
1501                                                  XhoI c|tcgag1553
1501                                  RsrII cg|gwccg1537
1501                          AgeI a|ccggt1528        AvaI c|ycgrg1553
 501 G  K  G  V  P  G  V  G  V  P  V  A  D  R  G  M  R  L  E  *  *
1501 GGTAAAGGGGTTCCAGGCGTGGGTGTACCGGTAGCGGACCGTGGAATGCGGCTCGAGTAATAA
```

F. 27 pET21Q-SKGPG-YKV48

```
   1                                  XmaI c|ccggg34
   1                                  AvaI c|ycgrg34
   1                             PpuMI rg|gwccy27
   1                      SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
  26 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGC
  51 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGG
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGT
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 301 GTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGG
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 376 GTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 176 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGC
 201 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGG
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGT
 751                                          XhoI c|tcgag794
 751            RsrII cg|gwccg763             AvaI c|ycgrg794
 751    AgeI a|ccggt754                  EcoRI g|aattc788
 251 V  P  V  A  D  R  G  M  R  L  D  K  E  F  L  E  *  *
 751 GTACCGGTAGCGGACCGTGGAATGCGGCTCGACAAAGAATTCCTCGAGTAATGA
```

F. 28 pET28aRW-QEQV12 (aka QEQV48)

```
   1        NheI g|ctagc8
   1 M  M  A  S  M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  D  D  D  D  K  L
   1 ATGATGGCTAGCATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCATGATGATGATGATAAACTC
  76            XmaI c|ccggg88
  76            AvaI c|ycgrg88
  76      PpuMI rg|gwccy81
  26 D  G  T  L  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P
  76 GACGGGACCCTCCCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCG
  51 G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P
 151 GGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCG
  76 G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P
 226 GGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCG
 101 G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P
```

```
301 GGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCG
126 G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P
376 GGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCG
151 G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P
451 GGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCG
176 G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P
526 GGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCG
201 G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P
601 GGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCG
226 G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P
676 GGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCG
751                                                                  RsrII cg|gwccg
817

751                                                          AgeI a|ccggt808
251 G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  V  A  D  R  G
751 GGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGTAGCGGACCGTGGA
826        XhoI c|tcgag833
826        AvaI c|ycgrg833
276 M  R  L  E  *  *
826 ATGCGGCTCGAGTAATAA
```

F. 29 pET21Q-SKGPG-QEQV12 (aka SKGPG-QEQV48)

```
  1                                        XmaI c|ccggg43
  1                                        AvaI c|ycgrg43
  1                            EcoRI g|aattc29
  1                    SalI g|tcgac20 NheI g|ctagc35
  1 M  S  K  G  P  G  V  D  K  E  F  A  S  V  P  G  Q  G  V  P  G  E  G  V  P
  1 ATGAGCAAAGGTCCGGGTGTCGACAAAGAATTCGCTAGCGTCCCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCG
 26 G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P
 76 GGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCG
 51 G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P
151 GGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCG
 76 G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P
226 GGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCG
101 G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P
301 GGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCG
126 G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P
376 GGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCG
151 G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P
451 GGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCG
176 G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P
526 GGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCG
201 G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P
601 GGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCG
226 G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P
676 GGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCG
```

```
 751                                    BamHI g|gatcc786
 751                                XhoI c|tcgag780
 751                     SgrAI cr|ccggyg770
 751             AgeI a|ccggt763  AvaI c|ycgrg780
 251 G  V  G  V  P  V  T  G  G  A  R  G  I  L  *
 751 GGTGTTGGCGTACCGGTCACCGGCGGGGCTCGAGGGATCCTCTAG
```

F. 30 pET21Q-SKGPG-YC12-His

```
   1                                   XmaI c|ccggg34
   1                                   AvaI c|ycgrg34
   1                             PpuMI rg|gwccy27
   1                     SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
  26 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
  51 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 301 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 376 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 176 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
 201 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 251 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 751 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 276 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 826 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 301 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 901 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 326 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 976 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
1051                                                                        RsrII cg
|gwccg1123
1051                                                                  AgeI a|ccggt1114
 351 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D
1051 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGTAGCGGAC
1126             XhoI c|tcgag1139
```

```
1126              AvaI c|ycgrg1139
 376 R  G  M  R  L  E  H  H  H  H  H  H  *
1126 CGTGGAATGCGGCTCGAGCACCACCACCACCACCACTGA
```

F. 31 pET21Q-SKGPG-YC12-M-His

```
   1                                  XmaI c|ccggg34
   1                                  AvaI c|ycgrg34
   1                            PpuMI rg|gwccy27
   1                    SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
  26 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
  51 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 301 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 376 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 176 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
 201 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 251 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 751 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 276 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 826 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 301 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 901 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 326 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 976 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
1051                                                                    RsrII cg
|gwccg1123
1051                                                          AgeI a|ccggt1114
 351 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D
1051 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGTAGCGGAC
 376 R  G  M  R  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L
1126 CGTGGAATGCGGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTG
1201        XhoI c|tcgag1208
1201        AvaI c|ycgrg1208
```

```
 401 Y  L  L  E  H  H  H  H  H  H  *
1201 TACCTGCTCGAGCACCACCACCACCACCACTGA
```

F. 32 pET21Q-SKGPG-YC12-D-His

```
   1                                XmaI c|ccggg34
   1                                AvaI c|ycgrg34
   1                           PpuMI rg|gwccy27
   1                     SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
  26 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
  51 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 301 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 376 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 176 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
 201 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 251 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 751 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 276 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 826 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 301 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 901 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 326 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 976 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
1051                                                                  RsrII cg
|gwccg1123
1051                                                              AgeI a|ccggt1114
 351 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D
1051 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGTAGCGGAC
 376 R  G  M  R  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L
1126 CGTGGAATGCGGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTG
 401 Y  L  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L
1201 TACCTGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTG
1276  XhoI c|tcgag1277
```

```
1276  AvaI c|ycgrg1277
 426 L  E  H  H  H  H  H  H  *
1276 CTCGAGCACCACCACCACCACCACTGA
```

F. 33 pET21Q-SKGPG-YC12-T-His

```
   1                                  XmaI c|ccggg34
   1                                  AvaI c|ycgrg34
   1                             PpuMI rg|gwccy27
   1                     SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
  26 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
  51 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 301 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 376 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 176 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
 201 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 251 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 751 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 276 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 826 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 301 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 901 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 326 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 976 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
1051                                                                    RsrII cg
|gwccg1123
1051                                                          AgeI a|ccggt1114
 351 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D
1051 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGTAGCGGAC
 376 R  G  M  R  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L
1126 CGTGGAATGCGGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTG
 401 Y  L  L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L
1201 TACCTGCTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTG
```

```
 426 L  D  S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  L  D
1276 CTCGATAGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTGCTCGAT
1351                                                                 XhoI c|tcgag1415
1351                                                                 AvaI c|ycgrg1415
 451 S  K  I  P  K  A  S  S  V  P  T  E  L  S  A  I  S  T  L  Y  L  L  E  H  H
1351 AGCAAAATCCCGAAAGCCAGCTCCGTGCCGACCGAACTGTCAGCAATCTCTACGCTGTACCTGCTCGAGCACCAC
 476 H  H  H  H  *
1426 CACCACCACCACTGA
```

F. 34 pET21Q-SKGPG-YC12-RGD

```
   1                                XmaI c|ccggg34
   1                                AvaI c|ycgrg34
   1                          PpuMI rg|gwccy27
   1                   SalI g|tcgac20
   1 M  S  K  G  P  G  V  D  G  T  L  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
   1 ATGAGCAAAGGTCCGGGTGTCGACGGGACCCTCCCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
  26 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
  76 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
  51 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 151 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
  76 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 226 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 101 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 301 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 126 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 376 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 151 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 451 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 176 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 526 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
 201 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G
 601 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGT
 226 V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G
 676 GTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGC
 251 V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G  V  P  G  K  G
 751 GTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGC
 276 V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  C  G
 826 GTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGCGTTCCGGGTTGCGGC
 301 V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G
 901 GTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGCGTTCCGGGTGTGGGC
 326 V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G
 976 GTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGCTACGGTGTGCCGGGTAAGGGC
1051                                                                 RsrII cg
|gwccg1123
1051                                                               AgeI a|ccggt1114
 351 V  P  G  V  G  V  P  G  C  G  V  P  G  K  G  V  P  G  V  G  V  P  V  A  D
```

```
1051 GTTCCGGGTGTGGGCGTTCCGGGTTGCGGCGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGTAGCGGAC
1126                                                                        XhoI c|
tcgag1199
1126                                                                        AvaI c|
ycgrg1199
 376 R  G  M  R  G  G  T  V  Y  A  V  T  G  R  G  D  S  P  A  S  S  G  G  G  L
1126 CGTGGAATGCGGGGGGGCACTGTGTACGCGGTAACGGGCCGCGGTGACTCCCCGGCATCAAGCGGTGGTGGTCTC
 401 E  H  H  H  H  H  H  *
1201 GAGCACCACCACCACCACCACTGA
```

F. 35 pET21Q-SKGPG-YKV24-QEQV24-BMP2

```
   1                          XmaI c|ccggg28
   1                          AvaI c|ycgrg28
   1 M  S  K  G  P  G  V  D  C  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
   1 ATGAGCAAAGGTCCGGGTGTCGATTGCCCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCG
  26 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
  76 GGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCGGGTAAAGGCGTTCCG
  51 G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P
 151 GGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCGGGCTATGGGGTGCCG
  76 G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P
 226 GGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCAGGCGTGGGTGTACCG
 101 G  Y  G  V  P  G  K  G  V  P  G  V  G  V  P  G  Y  G  V  P  G  K  G  V  P
 301 GGCTATGGGGTGCCGGGTAAAGGCGTTCCGGGTGTGGGCGTACCGGGTTACGGCGTACCGGGTAAAGGGGTTCCA
 376                                   XmaI c|ccggg412
 376                                   AvaI c|ycgrg412
 376                               NheI g|ctagc404
 376                         EcoRI g|aattc398
 126 G  V  G  V  P  G  C  E  F  A  S  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q
 376 GGCGTGGGTGTACCGGGCTGTGAATTCGCTAGCGTCCCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAG
 151 G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V
 451 GGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTT
 176 G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q
 526 GGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAG
 201 G  V  P  G  E  G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E
 601 GGTGTGCCGGGCGAAGGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAA
 226 G  V  P  G  Q  G  V  P  G  V  G  V  P  G  Q  G  V  P  G  E  G  V  P  G  Q
 676 GGTGTTCCGGGTCAGGGCGTGCCGGGTGTTGGCGTACCGGGCCAGGGTGTGCCGGGCGAAGGTGTTCCGGGTCAG
 751                         AgeI a|ccggt772
 251 G  V  P  G  V  G  V  P  V  C  G  G  G  G  S  T  S  K  I  P  K  A  S  S  V
 751 GGCGTGCCGGGTGTTGGCGTACCGGTGTGCGGCGGCGGCGGTTCTACTAGCAAGATCCCAAAAGCCAGCTCCGTG
 826                                     SpeI a|ctagt863
 276 P  T  E  L  S  A  I  S  T  L  Y  L  T  S  *  F  E  H  H  H  H  H  H  *
 826 CCGACCGAACTGTCAGCAATCTCTACGCTGTACCTCACTAGTTAATTCGAGCACCACCACCACCACCACTGA
```

F. 36 pJB-91x3

```
       1                                                    GTCAGCCAAAGGCTCCA-5' TALE-
CDKN2A reverse53
       1 GAATTCACTAGTGTGGTGGGGAGCAGCATCGAGCCGGTGGCGGGGAGCAGCAGTCAGCCAAAGGCTCCATGGCTG
     260 CTTAAGTGATCACACCACCCCTCGTCGTAGCTCGGCCACCGCCCCTCGTCGTCAGTCGGTTTCCGAGGTACCGAC
      76                                                                GTCAGCCAAAGGCTCC
A-5' TALE-CDKN2A reverse140
      76 GCCACGCTAGCTGAATTCACTAGTGTGGTGGGGAGCAGCATCGAGCCGGTGGCGGGGAGCAGCAGTCAGCCAAAG
     185 CGGTGCGATCGACTTAAGTGATCACACCACCCCTCGTCGTAGCTCGGCCACCGCCCCTCGTCGTCAGTCGGTTTC
     151 GCTCCATGGCTGGCCACGCTAGCTGAATTCACTAGTGTGGTGGGGAGCAGCATCGAGCCGGTGGCGGGGAGCAGC
     110 CGAGGTACCGACCGGTGCGATCGACTTAAGTGATCACACCACCCCTCGTCGTAGCTCGGCCACCGCCCCTCGTCG
     226  GTCAGCCAAAGGCTCCA-5' TALE-CDKN2A reverse227
     226 AGTCAGCCAAAGGCTCCATGGCTGGCCACGCTAGC
      35 TCAGTCGGTTTCCGAGGTACCGACCGGTGCGATCG
```

F. 37 pJB-57

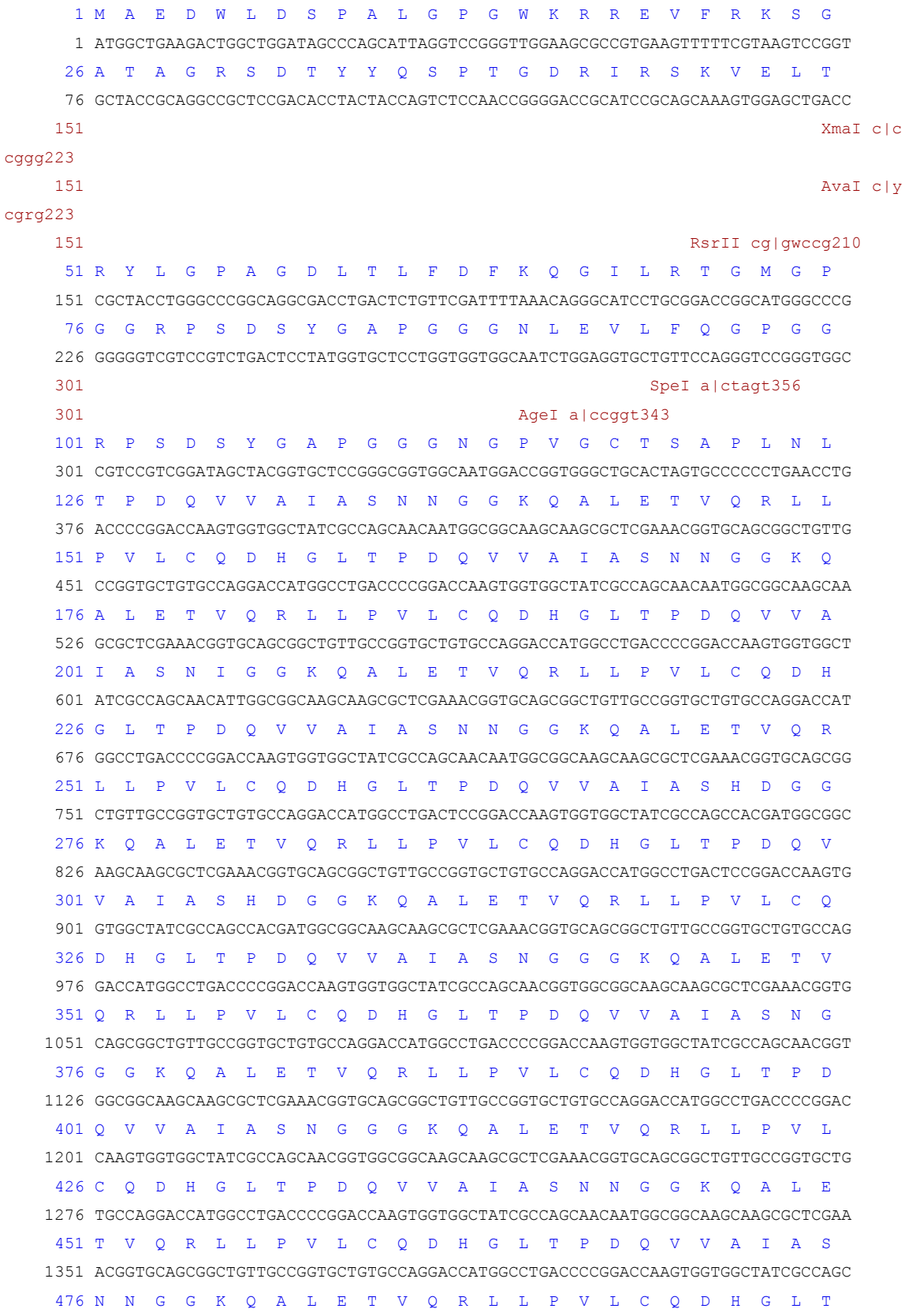
```
       1                                                        TGGAGCCTTCGGCTGAC-
3' TALE-CDKN2A60
       1 GAATTCGTCGACGGGACCCTCAGCGGGGAGCAGCATGGAGCCGGTGGTGGGGAGCAGCATGGAGCCTTCGGCTGA
      95 CTTAAGCAGCTGCCCTGGGAGTCGCCCCTCGTCGTACCTCGGCCACCACCCCTCGTCGTACCTCGGAAGCCGACT
      76 CTGGCTAAGTGGGCGGACCG
      20 GACCGATTCACCCGCCTGGC
```

F. 38 pET21b-MBD-RZ2-TALE13-His

```
       1 M  A  E  D  W  L  D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G
       1 ATGGCTGAAGACTGGCTGGATAGCCCAGCATTAGGTCCGGGTTGGAAGCGCCGTGAAGTTTTTCGTAAGTCCGGT
      26 A  T  A  G  R  S  D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T
      76 GCTACCGCAGGCCGCTCCGACACCTACTACCAGTCTCCAACCGGGGACCGCATCCGCAGCAAAGTGGAGCTGACC
     151                                                                    XmaI c|c
cggg223
     151                                                                    AvaI c|y
cgrg223
     151                                                        RsrII cg|gwccg210
      51 R  Y  L  G  P  A  G  D  L  T  L  F  D  F  K  Q  G  I  L  R  T  G  M  G  P
     151 CGCTACCTGGGCCCGGCAGGCGACCTGACTCTGTTCGATTTTAAACAGGGCATCCTGCGGACCGGCATGGGCCCG
      76 G  G  R  P  S  D  S  Y  G  A  P  G  G  G  N  L  E  V  L  F  Q  G  P  G  G
     226 GGGGGTCGTCCGTCTGACTCCTATGGTGCTCCTGGTGGTGGCAATCTGGAGGTGCTGTTCCAGGGTCCGGGTGGC
     301                                          AgeI a|ccggt343
     101 R  P  S  D  S  Y  G  A  P  G  G  G  N  G  P  V  G  C  T  S  M  T  G  G  Q
     301 CGTCCGTCGGATAGCTACGGTGCTCCGGGCGGTGGCAATGGACCGGTGGGCTGCACTAGCATGACTGGTGGACAG
     376                        SalI g|tcgac400
     376                  XbaI t|ctaga394
     376            BamHI g|gatcc388
```

```
 126 Q  M  G  R  D  P  L  E  S  T  C  R  H  A  S  R  N  A  L  T  G  A  P  L  N
 376 CAAATGGGTCGGGATCCTCTAGAGTCGACCTGCAGGCATGCATCGCGCAATGCACTGACGGGTGCCCCCCTGAAC
 151 L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L
 451 CTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTG
 176 L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K
 526 TTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAG
 201 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V
 601 CAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTG
 226 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D
 676 GCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGAC
 251 H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
 751 CATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
 276 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G
 826 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC
 301 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
 901 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 326 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
 976 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
 351 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
1051 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACG
 376 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N
1126 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAC
 401 G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
1201 GGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCG
 426 D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V
1276 GACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
 451 L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L
1351 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTC
 476 E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A
1426 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
 501 S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L
1501 AGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
 526 T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L
1576 ACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
 551 P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q
1651 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAA
 576 A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T  N  D  H  L  V
1726 GCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGTC
 601 A  L  A  C  L  G  G  R  P  A  M  D  A  V  K  K  G  L  P  H  A  P  E  L  I
1801 GCCTTGGCCTGCCTCGGCGGACGTCCTGCCATGGATGCAGTGAAAAAGGGATTGCCGCACGCGCCGGAATTGATC
 626 R  R  V  N  R  R  I  G  E  R  T  S  H  R  V  A  D  Y  A  Q  V  V  R  V  L
1876 AGAAGAGTCAATCGCCGTATTGGCGAACGCACGTCCCATCGCGTTGCCGACTACGCGCAAGTGGTTCGCGTGCTG
 651 E  F  F  Q  C  H  S  H  P  A  Y  A  F  D  E  A  M  T  Q  F  G  M  S  R  N
1951 GAGTTTTTCCAGTGCCACTCCCACCCAGCGTACGCATTTGATGAGGCCATGACGCAGTTCGGGATGAGCAGGAAC
 676 G  L  V  Q  L  F  R  R  V  G  V  T  E  L  E  A  R  G  H  H  H  H  H  H  *
2026 GGGTTGGTACAGCTCTTTCGCAGAGTGGGCGTCACCGAACTCGAAGCCCGCGGTCACCACCACCACCACCACTGA
 701 *  *
2101 TAATAA
```

F. 39 pET21b-MBD-RZ2-TALE17 via pTAL2m



```
1426 AACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACT
 501 P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1501 CCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCG
 526 V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A
1576 GTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCG
 551 L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I
1651 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATC
 576 A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G
1726 GCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
 601 L  T  P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L
1801 CTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTG
 626 L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K
1876 TTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAG
 651 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V
1951 CAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTG
 676 A  I  A  S  N  G  G  G  K  Q  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A
2026 GCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCG
2101                                XhoI c|tcgag2135
2101                                AvaI c|ycgrg2135
2101                             NheI g|ctagc2129
 701 L  A  A  L  T  N  D  H  L  A  S  L  E  H  H  H  H  H  H  *
2101 TTGGCCGCGTTGACCAACGACCACCTCGCTAGCCTCGAGCACCACCACCACCACCACTGA
```

F. 40 PPM-13

```
   1 M  A  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M
   1 ATGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
  26 A  P  K  K  K  R  K  V  G  I  H  R  G  V  P  M  V  D  L  R  T  L  G  Y  S
  76 GCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCG
  51 Q  Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H
 151 CAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
  76 G  F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q
 226 GGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
 101 D  M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R
 301 GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
 126 A  L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L
 376 GCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
 451                                                                SgrAI cr|ccggyg5
15
 451                                                                AgeI a|ccggt515
 151 K  I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P
 451 AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGTGCCCCC
 176 L  N  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
 526 CTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
 201 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G
 601 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGC
 226 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
```

```
 676 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAA
 251 V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
 751 GTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
 276 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T
 826 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACG
 301 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H
 901 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCAC
 326 D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
 976 GATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCG
 351 D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V
1051 GACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
 376 L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L
1126 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTC
 401 E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A
1201 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
 426 S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L
1276 AGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
 451 T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L
1351 ACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
 476 P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q
1426 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAA
 501 A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A
1501 GCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCT
 526 I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H
1576 ATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCAT
 551 G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R
1651 GGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGG
 576 L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G
1726 CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGC
 601 K  Q  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T  N  D  H
1801 AAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCAT
 626 L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H  A  P  A
1876 CTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCA
1951                            AvaI c|ycgrg1979          BamHI g|gatcc2006
 651 L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  S  L  E  V  L  F
1951 TTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCTGGAGGTGCTGTTC
2026           NdeI ca|tatg2037
 676 Q  G  P  H  M  A  E  D  W  L  D  S  P  A  L  G  P  G  W  K  R  R  E  V  F
2026 CAGGGCCCCCATATGGCTGAGGACTGGCTGGACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTT
2101                                                                BamHI g|gatcc2163
 701 R  K  S  G  A  T  A  G  R  S  D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K
2101 CGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAA
 726 V  E  L  T  R  Y  L  G  P  A  G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G
2176 GTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGT
2251                         XhoI c|tcgag2276
2251                         AvaI c|ycgrg2276
 751 G  G  T  G  G  G  G  S  L  E  H  H  H  H  H  H  C  *
2251 GGCGGTACCGGCGGTGGCGGTAGCCTCGAGCATCACCATCACCATCACTGCTAA
```

F. 41 PPM-17

```
     1 M  A  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M
     1 ATGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
    26 A  P  K  K  K  R  K  V  G  I  H  R  G  V  P  M  V  D  L  R  T  L  G  Y  S
    76 GCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCG
    51 Q  Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H
   151 CAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
    76 G  F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q
   226 GGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
   101 D  M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R
   301 GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
   126 A  L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L
   376 GCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
   451                                                                 SgrAI cr|ccggyg5
15
   451                                                                   AgeI a|ccggt515
   151 K  I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P
   451 AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGTGCCCCC
   176 L  N  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
   526 CTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
   201 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G
   601 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGC
   226 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
   676 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAA
   251 V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
   751 GTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
   276 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T
   826 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACG
   301 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H
   901 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCAC
   326 D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
   976 GATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCG
   351 D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V
  1051 GACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
   376 L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L
  1126 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTC
   401 E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A
  1201 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
   426 S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L
  1276 AGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
   451 T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L
  1351 ACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
   476 P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q
  1426 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAA
   501 A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A
  1501 GCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCT
   526 I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H
```

```
1576 ATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCAT
 551 G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R
1651 GGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGG
 576 L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G
1726 CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGC
 601 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V
1801 AAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTG
 626 V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
1876 GTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAG
 651 D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V
1951 GACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTG
 676 Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D
2026 CAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGAT
 701 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D
2101 GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGAC
 726 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  S  I  V  A  Q  L  S  R  P
2176 CAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCT
 751 D  P  A  L  A  A  L  T  N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D
2251 GATCCGGCGTTGGCCGCGTTGACCAACGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGAT
2326                                                                 AvaI c|ycgrg2387
 776 A  V  K  K  G  L  P  H  A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S
2326 GCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCC
2401             BamHI g|gatcc2414             NdeI ca|tatg2445
 801 H  R  V  A  G  S  L  E  V  L  F  Q  G  P  H  M  A  E  D  W  L  D  S  P  A
2401 CATCGAGTCGCGGGATCCCTGGAGGTGCTGTTCCAGGGCCCCCATATGGCTGAGGACTGGCTGGACAGCCCGGCC
 826 L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S  D  T  Y  Y
2476 CTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACCTATTAC
2551                   BamHI g|gatcc2571
 851 Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A  G  D  L  T
2551 CAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGATCTCACC
2626                                                             XhoI c|tcgag2684
2626                                                             AvaI c|ycgrg2684
 876 L  F  D  F  K  Q  G  I  L  G  G  G  G  T  G  G  G  G  S  L  E  H  H  H  H
2626 CTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCGGTACCGGCGGTGGCGGTAGCCTCGAGCATCACCATCAC
 901 H  H  C  *
2701 CATCACTGCTAA
```

F. 42 PPMG-17

```
   1 M  A  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M
   1 ATGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
  26 A  P  K  K  K  R  K  V  G  I  H  R  G  V  P  M  V  D  L  R  T  L  G  Y  S
  76 GCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCG
  51 Q  Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H
 151 CAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
  76 G  F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q
 226 GGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
```

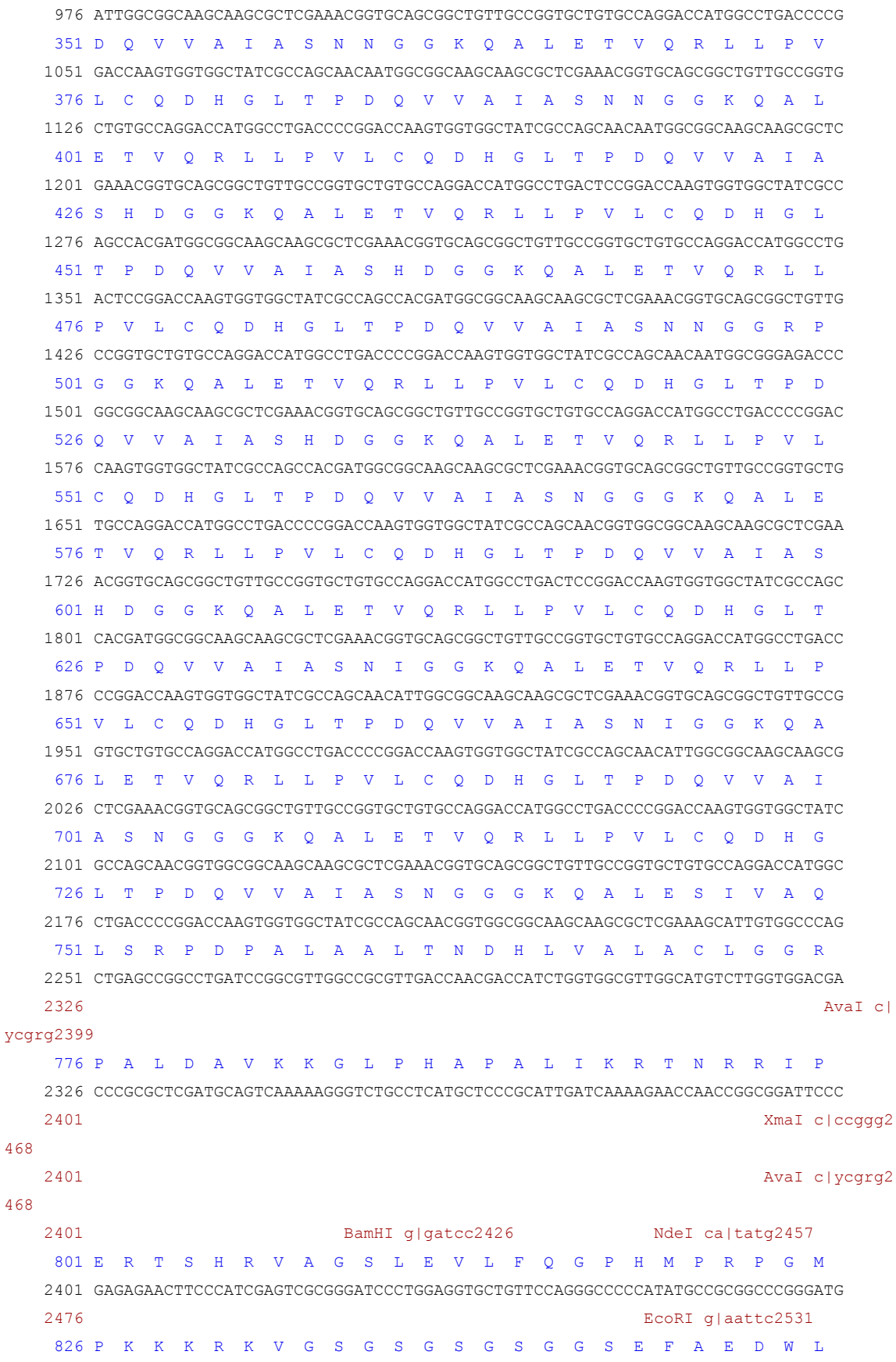
```
 101 D  M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R
 301 GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
 126 A  L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L
 376 GCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
 451                                                                  SgrAI cr|ccggyg5
15
 451                                                                  AgeI a|ccggt515
 151 K  I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P
 451 AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGTGCCCCC
 176 L  N  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
 526 CTGAACCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
 201 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G
 601 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC
 226 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
 676 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 251 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
 751 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
 276 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T
 826 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACG
 301 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N
 901 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAC
 326 I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
 976 ATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCG
 351 D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V
1051 GACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
 376 L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L
1126 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTC
 401 E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A
1201 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
 426 S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L
1276 AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
 451 T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L
1351 ACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
 476 P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  R  P
1426 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGGAGACCC
 501 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D
1501 GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGAC
 526 Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1576 CAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTG
 551 C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E
1651 TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAA
 576 T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S
1726 ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGC
 601 H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T
1801 CACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACC
 626 P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1876 CCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCG
 651 V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A
```

```
    1951 GTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCG
     676 L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I
    2026 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATC
     701 A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G
    2101 GCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
     726 L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  S  I  V  A  Q
    2176 CTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAG
     751 L  S  R  P  D  P  A  L  A  A  L  T  N  D  H  L  V  A  L  A  C  L  G  G  R
    2251 CTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGA
    2326                                                                    AvaI c|
ycgrg2399
     776 P  A  L  D  A  V  K  K  G  L  P  H  A  P  A  L  I  K  R  T  N  R  R  I  P
    2326 CCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCC
    2401                                                                  XmaI c|ccggg2
468
    2401                                                                  AvaI c|ycgrg2
468
    2401                        BamHI g|gatcc2426            NdeI ca|tatg2457
     801 E  R  T  S  H  R  V  A  G  S  L  E  V  L  F  Q  G  P  H  M  P  R  P  G  M
    2401 GAGAGAACTTCCCATCGAGTCGCGGGATCCCTGGAGGTGCTGTTCCAGGGCCCCCATATGCCGCGGCCCGGGATG
    2476                                                  EcoRI g|aattc2531
     826 P  K  K  K  R  K  V  G  S  G  S  G  S  G  S  G  G  S  E  F  A  E  D  W  L
    2476 CCAAAAAAGAAGAGAAAGGTAGGTAGCGGCAGCGGTAGCGGTTCCGGAGGCAGCGAATTCGCTGAGGACTGGCTG
     851 D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S
    2551 GACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCA
    2626                                BamHI g|gatcc2658
     876 D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A
    2626 GACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCG
    2701                                                                    AgeI a|c
cggt2773
     901 G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  G  T  G  G  G  Q  F  V  P
    2701 GGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCGGTACCGGCGGTGGCCAATTCGTACCG
    2776          AleI cacnn|nngtg2786          AleI cacnn|nngtg2816
     926 V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D
    2776 GTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
     951 V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F
    2851 GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
    2926                                    AleI cacnn|nngtg2966
     976 I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V  Q  C  F
    2926 ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC
    1001 S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R
    3001 AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
    3076                                                                  AleI cacnn|nn
gtg3143
    1026 T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N
    3076 ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
    1051 R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N
    3151 CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC
```

```
1076 S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R  H  N  I
3226 AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATC
1101 E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P
3301 GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
1126 D  N  H  Y  L  S  T  Q  S  K  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L
3376 GACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
3451                                                                 SalI g|tcgac3
518
3451                                                                 XbaI t|ctaga3512
1151 E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  Q  L  S  S  R  V  D  S
3451 GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCAATTGTCCTCTAGAGTCGACAGC
3526       AgeI a|ccggt3532
1176 C  L  P  V  R  H  H  H  H  H  H  *  *
3526 TGCCTACCGGTCAGGCATCACCATCACCATCACTAATAA
```

F. 43 pET28aRW-PPMG-17

```
   1 M  A  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M
   1 ATGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
  26 A  P  K  K  K  R  K  V  G  I  H  R  G  V  P  M  V  D  L  R  T  L  G  Y  S
  76 GCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCG
  51 Q  Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H
 151 CAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
  76 G  F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q
 226 GGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
 101 D  M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R
 301 GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
 126 A  L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L
 376 GCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
 451                                                                 SgrAI cr|ccggyg5
15
 451                                                                 AgeI a|ccggt515
 151 K  I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P
 451 AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGTGCCCCC
 176 L  N  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
 526 CTGAACCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
 201 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G
 601 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC
 226 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
 676 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 251 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
 751 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
 276 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T
 826 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACG
 301 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N
 901 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAC
 326 I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
```


976 ATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCG
351 D Q V V A I A S N N G G K Q A L E T V Q R L L P V
1051 GACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
376 L C Q D H G L T P D Q V V A I A S N N G G K Q A L
1126 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTC
401 E T V Q R L L P V L C Q D H G L T P D Q V V A I A
1201 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
426 S H D G G K Q A L E T V Q R L L P V L C Q D H G L
1276 AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
451 T P D Q V V A I A S H D G G K Q A L E T V Q R L L
1351 ACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
476 P V L C Q D H G L T P D Q V V A I A S N N G G R P
1426 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGGAGACCC
501 G G K Q A L E T V Q R L L P V L C Q D H G L T P D
1501 GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGAC
526 Q V V A I A S H D G G K Q A L E T V Q R L L P V L
1576 CAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTG
551 C Q D H G L T P D Q V V A I A S N G G G K Q A L E
1651 TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAA
576 T V Q R L L P V L C Q D H G L T P D Q V V A I A S
1726 ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGC
601 H D G G K Q A L E T V Q R L L P V L C Q D H G L T
1801 CACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACC
626 P D Q V V A I A S N I G G K Q A L E T V Q R L L P
1876 CCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCG
651 V L C Q D H G L T P D Q V V A I A S N I G G K Q A
1951 GTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCG
676 L E T V Q R L L P V L C Q D H G L T P D Q V V A I
2026 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATC
701 A S N G G G K Q A L E T V Q R L L P V L C Q D H G
2101 GCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
726 L T P D Q V V A I A S N G G G K Q A L E S I V A Q
2176 CTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAG
751 L S R P D P A L A A L T N D H L V A L A C L G G R
2251 CTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGA
2326
AvaI c|
ycgrg2399
776 P A L D A V K K G L P H A P A L I K R T N R R I P
2326 CCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCC
2401
XmaI c|ccggg2
468
2401
AvaI c|ycgrg2
468
2401
BamHI g|gatcc2426
NdeI ca|tatg2457
801 E R T S H R V A G S L E V L F Q G P H M P R P G M
2401 GAGAGAACTTCCCATCGAGTCGCGGGATCCCTGGAGGTGCTGTTCCAGGGCCCCCATATGCCGCGGCCCGGGATG
2476
EcoRI g|aattc2531
826 P K K K R K V G S G S G S G S G G S E F A E D W L

```
2476 CCAAAAAAGAAGAGAAAGGTAGGTAGCGGCAGCGGTAGCGGTTCCGGAGGCAGCGAATTCGCTGAGGACTGGCTG
 851 D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S
2551 GACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCA
2626                              BamHI g|gatcc2658
 876 D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A
2626 GACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCG
2701                                                                    AgeI a|c
cggt2773
 901 G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  G  T  G  G  G  Q  F  V  P
2701 GGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCGGTACCGGCGGTGGCCAATTCGTACCG
2776          AleI cacnn|nngtg2786          AleI cacnn|nngtg2816
 926 V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D
2776 GTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC
 951 V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F
2851 GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC
2926                                     AleI cacnn|nngtg2966
 976 I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V  Q  C  F
2926 ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC
1001 S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R
3001 AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
3076                                                                  AleI cacnn|nn
gtg3143
1026 T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N
3076 ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
1051 R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N
3151 CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC
1076 S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R  H  N  I
3226 AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATC
1101 E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P
3301 GAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC
1126 D  N  H  Y  L  S  T  Q  S  K  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L
3376 GACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG
1151 E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  Q  L  S  S  S  M  T  G
3451 GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCAATTGTCCTCTAGCATGACTGGT
3526                                                    XhoI c|tcgag3578
3526                                                    AvaI c|ycgrg3578
1176 G  Q  Q  M  G  H  H  H  H  H  H  H  D  D  D  D  K  L  E  *  *
3526 GGACAGCAAATGGGTCACCACCACCACCACCACCATGATGATGATGATAAACTCGAGTAATAA
```

F. 44 pET28aRW-PPdMG-17

```
   1 M  A  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M
   1 ATGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATG
  26 A  P  K  K  K  R  K  V  G  I  H  R  G  V  P  M  V  D  L  R  T  L  G  Y  S
  76 GCCCCCAAGAAGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTCG
  51 Q  Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H
 151 CAACAGCAACAGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCAT
```

```
  76 G  F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q
 226 GGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAA
 101 D  M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R
 301 GATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGA
 126 A  L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L
 376 GCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTG
 451                                                                 SgrAI cr|ccggyg5
15

 451                                                                 AgeI a|ccggt515
 151 K  I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P
 451 AAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGTGCCCCC
 176 L  N  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q
 526 CTGAACCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAG
 201 R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  H  D  G
 601 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGC
 226 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q
 676 GGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 251 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
 751 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGC
 276 Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T
 826 CAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACG
 301 V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N
 901 GTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAC
 326 I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P
 976 ATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCG
 351 D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V
1051 GACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
 376 L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  K  Q  A  L
1126 CTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAAGCGCTC
 401 E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A
1201 GAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
 426 S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L
1276 AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTG
 451 T  P  D  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L
1351 ACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTG
 476 P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  N  G  G  R  P
1426 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGGAGACCC
 501 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D
1501 GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGAC
 526 Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1576 CAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTG
 551 C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E
1651 TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAA
 576 T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S
1726 ACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGC
 601 H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T
1801 CACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACC
 626 P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
```

```
1876 CCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCG
 651 V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A
1951 GTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCG
 676 L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G  L  T  P  D  Q  V  V  A  I
2026 CTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATC
 701 A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  D  H  G
2101 GCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
 726 L  T  P  D  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  S  I  V  A  Q
2176 CTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAG
 751 L  S  R  P  D  P  A  L  A  A  L  T  N  D  H  L  V  A  L  A  C  L  G  G  R
2251 CTGAGCCGGCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGA
2326                                                                    AvaI c|
ycgrg2399
 776 P  A  L  D  A  V  K  K  G  L  P  H  A  P  A  L  I  K  R  T  N  R  R  I  P
2326 CCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCC
2401                                                                  XmaI c|ccggg2
468
2401                                                                  AvaI c|ycgrg2
468
2401                          BamHI g|gatcc2426             NdeI ca|tatg2457
 801 E  R  T  S  H  R  V  A  G  S  L  E  V  L  F  Q  G  P  H  M  P  R  P  G  M
2401 GAGAGAACTTCCCATCGAGTCGCGGGATCCCTGGAGGTGCTGTTCCAGGGCCCCCATATGCCGCGGCCCGGGATG
2476                                                  EcoRI g|aattc2531
 826 P  K  K  K  R  K  V  G  S  G  S  G  S  G  S  G  G  S  E  F  A  E  D  W  L
2476 CCAAAAAAGAAGAGAAAGGTAGGTAGCGGCAGCGGTAGCGGTTCCGGAGGCAGCGAATTCGCTGAGGACTGGCTG
 851 D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S
2551 GACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCA
2626                               BamHI g|gatcc2658
 876 D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A
2626 GACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCG
 901 G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  G  T  G  G  G  Q  F  A  E
2701 GGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCGGTACCGGCGGTGGCCAATTCGCTGAG
 926 D  W  L  D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A
2776 GACTGGCTGGACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCG
2851                                     BamHI g|gatcc2892
 951 G  R  S  D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L
2851 GGACGCTCAGACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTG
 976 G  P  A  G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  G  T  G  G  G  Q
2926 GGCCCTGCGGGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCGGTACCGGCGGTGGCCAA
3001                  AleI cacnn|nngtg3020
3001       AgeI a|ccggt3007                        AleI cacnn|nngtg3050
1001 F  V  P  V  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L
3001 TTCGTACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
1026 D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T
3076 GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
3151                                                 AleI cacnn|nngtg3200
1051 L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V
3151 CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTG
```

```
1076 Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V
3226 CAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
1101 Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T
3301 CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
3376  AleI cacnn|nngtg3377
1126 L  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y
3376 CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
1151 N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R
3451 AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGC
1176 H  N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V
3526 CACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
1201 L  L  P  D  N  H  Y  L  S  T  Q  S  K  L  S  K  D  P  N  E  K  R  D  H  M
3601 CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
1226 V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  Q  L  S  S  S
3676 GTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGCAATTGTCCTCTAGC
3751                                                                  XhoI c|tcgag3812
3751                                                                  AvaI c|ycgrg3812
1251 M  T  G  G  Q  Q  M  G  H  H  H  H  H  H  H  D  D  D  D  K  L  E  *  *
3751 ATGACTGGTGGACAGCAAATGGGTCACCACCACCACCACCACCATGATGATGATGATAAACTCGAGTAATAA
```

F. 45 pTH-R20-VN173

For the whole backbone sequence of pTH-R20, refer to pTH-TelR20-mCherry from Addgene (Plasmid #49637)

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369

 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
```

```
 251 V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGAC
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
```

```
2551                                    AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
2626                              EcoRI g|aattc2656   EcoRV gat|atc2676
2626                       NotI gc|ggccgc2649 BglII a|gatct2668       XbaI t|ctaga2
693
 876 G  G  G  G  S  K  L  A  A  A  N  S  S  I  D  L  I  S  V  P  V  D  S  R  R
2626 GGCGGAGGTGGCTCCAAGCTTGCGGCCGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTCTAGAAGA
 901 S  I  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G
2701 TCCATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
 926 D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K
2776 GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 951 L  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  G  Y  G  L  Q  C
2851 CTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCCTGCAGTGC
 976 F  A  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E
2926 TTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
1001 R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
3001 CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
1026 N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y
3076 AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC
1051 N  S  H  N  V  Y  I  T  A  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N
3151 AACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAAC
1076 I  E  *
3226 ATCGAGTAG
```

F. 46 pTH-R20-VC155

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                AvaI c|ycgrg
369
 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGA
```

```
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGAC
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
```

```
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                    AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
2626                       XhoI c|tcgag2649
2626                       AvaI c|ycgrg2649
 876 G  G  G  G  S  K  L  S  R  G  T  R  P  A  C  K  I  P  N  D  L  K  Q  K  V
2626 GGCGGAGGTGGCTCCAAGCTTTCTCGAGGTACCCGTCCGGCGTGCAAAATCCCGAACGACCTGAAACAGAAAGTC
 901 M  N  H  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  G  V
2701 ATGAACCACGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
 926 Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L
2776 CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
 951 S  Y  Q  S  K  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A
2851 AGCTACCAGTCCAAACTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
 976 A  G  I  T  L  G  M  D  E  L  Y  K  *
2926 GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG
```

F. 47 pTH-L20-VN173

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369
 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G
```

```
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACGGA
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                    AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
2626                              EcoRI g|aattc2656   EcoRV gat|atc2676
2626                        NotI gc|ggccgc2649 BglII a|gatct2668       XbaI t|ctaga2
```

```
693
 876 G  G  G  G  S  K  L  A  A  A  N  S  S  I  D  L  I  S  V  P  V  D  S  R  R
2626 GGCGGAGGTGGCTCCAAGCTTGCGGCCGCGAATTCATCGATAGATCTGATATCGGTACCAGTCGACTCTAGAAGA
 901 S  I  A  T  M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G
2701 TCCATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
 926 D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K
2776 GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAG
 951 L  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  G  Y  G  L  Q  C
2851 CTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCCTGCAGTGC
 976 F  A  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E
2926 TTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
1001 R  T  I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
3001 CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTG
1026 N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y
3076 AACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTAC
1051 N  S  H  N  V  Y  I  T  A  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N
3151 AACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAAC
1076 I  E  *
3226 ATCGAGTAG
```

F. 48 pTH-L20-VC155

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369
 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V
```

```
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACGGA
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  H  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACCACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                    AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
```

```
2626                        XhoI c|tcgag2649
2626                        AvaI c|ycgrg2649
 876 G  G  G  G  S  K  L  S  R  G  T  R  P  A  C  K  I  P  N  D  L  K  Q  K  V
2626 GGCGGAGGTGGCTCCAAGCTTTCTCGAGGTACCCGTCCGGCGTGCAAAATCCCGAACGACCTGAAACAGAAAGTC
 901 M  N  H  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  G  V
2701 ATGAACCACGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTG
 926 Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L
2776 CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTG
 951 S  Y  Q  S  K  L  S  K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A
2851 AGCTACCAGTCCAAACTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
 976 A  G  I  T  L  G  M  D  E  L  Y  K  *
2926 GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAG
```

F. 49 pTH-R20-VN210

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369
 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGAC
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
```

```
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                  AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
 876 S  G  S  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V
2626 AGTGGAAGTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
 901 N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  L  I
2701 AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATC
 926 C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  G  Y  G  L  Q  C  F  A
2776 TGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCCTGCAGTGCTTCGCC
 951 R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T
```

```
2851 CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
 976 I  F  F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R
2926 ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
1001 I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S
3001 ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC
1026 H  N  V  Y  I  T  A  D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N  I  E
3076 CACAACGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAG
1051 D  G  G  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D
3151 GACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
1076 N  H  Y  L  S  Y  Q  S  K  L  S  K  D  *
3226 AACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACTAA
```

F. 50 pTH-R20-linker-VN210

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369

 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGAC
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
```

```
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                  AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
 876 S  G  S  G  S  G  S  G  S  G  L  Q  G  S  G  S  V  S  K  G  E  E  L  F  T
2626 TCAGGTTCTGGAAGTGGAAGTGGGTCTGGACTGCAGGGAAGTGGAAGTGTGAGCAAGGGCGAGGAGCTGTTCACC
 901 G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E
2701 GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
 926 G  D  A  T  Y  G  K  L  T  L  K  L  I  C  T  T  G  K  L  P  V  P  W  P  T
2776 GGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
 951 L  V  T  T  L  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K  Q  H  D  F  F
2851 CTCGTGACCACCCTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTC
```

```
 976 K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R
2926 AAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
1001 A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G
3001 GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
1026 N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  T  A  D  K  Q  K  N
3076 AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAC
1051 G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  G  V  Q  L  A  D  H  Y  Q  Q
3151 GGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAGCAG
1076 N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  Y  Q  S  K  L  S  K
3226 AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAA
1101 D  *
3301 GACTAA
```

F. 51 pTH-R20-VC210

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
  51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  V  G  H  G
 151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGTGGGTCATGGG
  76 F  T  H  A  H  I  V  A  L  S  Q  H  P  A  A  L  G  T  V  A  V  K  Y  Q  D
 226 TTTACACATGCCCACATCGTAGCCTTGTCGCAGCACCCTGCAGCCCTTGGCACGGTCGCCGTCAAGTACCAGGAC
 301                                                                  AvaI c|ycgrg
369

 101 M  I  A  A  L  P  E  A  T  H  E  A  I  V  G  V  G  K  Q  W  S  G  A  R  A
 301 ATGATTGCGGCGTTGCCGGAAGCCACACATGAGGCGATCGTCGGTGTGGGGAAACAGTGGAGCGGAGCCCGAGCG
 126 L  E  A  L  L  T  V  A  G  E  L  R  G  P  P  L  Q  L  D  T  G  Q  L  L  K
 376 CTTGAGGCCCTGTTGACGGTCGCGGGAGAGCTGAGAGGGCCTCCCCTTCAGCTGGACACGGGCCAGTTGCTGAAG
 151 I  A  K  R  G  G  V  T  A  V  E  A  V  H  A  W  R  N  A  L  T  G  A  P  L
 451 ATCGCGAAGCGGGGAGGAGTCACGGCGGTCGAGGCGGTGCACGCGTGGCGCAATGCGCTCACGGGAGCACCCCTC
 176 N  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R
 526 AACCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGG
 201 L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G
 601 TTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGA
 226 K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V
 676 AAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACCCCAGAGCAGGTC
 251 V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q
 751 GTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAA
 276 A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V
 826 GCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTC
 301 Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D
 901 CAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGAC
 326 G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E
 976 GGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTGACACCAGAG
 351 Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L
1051 CAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTG
```

```
 376 C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K  Q  A  L  E
1126 TGCCAAGCGCACGGACTTACACCCGAACAAGTCGTGGCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAA
 401 T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S
1201 ACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTGGCAATTGCGAGC
 426 N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T
1276 AACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTAACC
 451 P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P
1351 CCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCT
 476 V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A
1426 GTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCA
 501 L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I
1501 CTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAGGTCGTGGCAATT
 526 A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G
1576 GCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGA
 551 L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T  V  Q  R  L
1651 CTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTG
 576 L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  I  G  G  K
1726 CTGCCTGTGCTGTGCCAAGCGCACGGCCTCACCCCAGAGCAGGTCGTGGCAATTGCGAGCAACATCGGGGGAAAG
 601 Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V
1801 CAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTTACGCCAGAGCAGGTCGTG
 626 A  I  A  S  N  I  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A
1876 GCAATTGCGAGCAACATCGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCG
 651 H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q
1951 CACGGACTAACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAG
 676 R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  H  D  G
2026 AGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGGTTGACCCCAGAGCAGGTCGTGGCAATTGCGAGCCATGACGGG
 701 G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q
2101 GGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGCCTGACCCCAGAGCAG
 726 V  V  A  I  A  S  H  D  G  G  K  Q  A  L  E  T  V  Q  R  L  L  P  V  L  C
2176 GTCGTGGCAATTGCGAGCCATGACGGGGGAAAGCAGGCACTCGAAACCGTCCAGAGGTTGCTGCCTGTGCTGTGC
 751 Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N  G  G  G  K  Q  A  L  E  T
2251 CAAGCGCACGGACTGACACCAGAGCAGGTCGTGGCAATTGCGAGCAACGGAGGGGGAAAGCAGGCACTCGAAACC
 776 V  Q  R  L  L  P  V  L  C  Q  A  H  G  L  T  P  E  Q  V  V  A  I  A  S  N
2326 GTCCAGAGGTTGCTGCCTGTGCTGTGCCAAGCGCACGGACTCACGCCTGAGCAGGTAGTGGCTATTGCATCCAAC
 801 I  G  G  R  P  A  L  E  S  I  V  A  Q  L  S  R  P  D  P  A  L  A  A  L  T
2401 ATCGGGGGCAGACCCGCACTGGAGTCAATCGTGGCCCAGCTTTCGAGGCCGGACCCCGCGCTGGCCGCACTCACT
 826 N  D  H  L  V  A  L  A  C  L  G  G  R  P  A  L  D  A  V  K  K  G  L  P  H
2476 AATGATCATCTTGTAGCGCTGGCCTGCCTCGGCGGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCAC
2551                                  AvaI c|ycgrg2588
 851 A  P  A  L  I  K  R  T  N  R  R  I  P  E  R  T  S  H  R  V  A  G  L  Q  G
2551 GCGCCTGCATTGATTAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGA
 876 S  G  S  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L
2626 AGTGGAAGTGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
 901 G  M  D  E  L  Y  K  *
2701 GGCATGGACGAGCTGTACAAGTAG
```

F. 52 pTH-tMBD-VN210

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  76                                                                 EcoRI g|aattc137
  76                                                    BspEI t|ccgga125
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  G  S  G  G  S  E  F  A  E  D
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGGTTCCGGAGGCAGCGAATTCGCTGAGGAC
  51 W  L  D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G
 151 TGGCTGGACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGA
 226                                        BamHI g|gatcc264
  76 R  S  D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G
 226 CGCTCAGACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGC
 101 P  A  G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L
 301 CCTGCGGGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTG
 126 D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S
 376 GACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCA
 451                                    BamHI g|gatcc483
 151 D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A
 451 GACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCG
 176 G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L  D  S
 526 GGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTGGACAGC
 201 P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S  D  T
 601 CCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACC
 676                            BamHI g|gatcc702
 226 Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A  G  D
 676 TATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGAT
 251 L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L  D  S  P  A
 751 CTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTGGACAGCCCGGCC
 276 L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S  D  T  Y  Y
 826 CTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACCTATTAC
 901                       BamHI g|gatcc921
 301 Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A  G  D  L  T
 901 CAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGATCTCACC
 326 L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  G  L  Q  G  S  G  S  V  S  K  G
 976 CTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGGCCTGCAGGGAAGTGGAAGTGTGAGCAAGGGC
 351 E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V
1051 GAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG
 376 S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  L  I  C  T  T  G  K  L  P
1126 TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCC
 401 V  P  W  P  T  L  V  T  T  L  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K
1201 GTGCCCTGGCCCACCCTCGTGACCACCCTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAG
 426 Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G
1276 CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGC
 451 N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D
1351 AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
 476 F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  T  A
1426 TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCC
```

```
 501 D  K  Q  K  N  G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  G  V  Q  L  A
1501 GACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCC
 526 D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  Y  Q
1576 GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAG
1651                     XbaI t|ctaga1673
 551 S  K  L  S  K  D  *  S  R
1651 TCCAAGCTGAGCAAAGACTAATCTAGA
```

F. 53 pTH-tMBD-VC210

```
   1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
   1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
  76                                                                EcoRI g|aattc137
  76                                                BspEI t|ccgga125
  26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  G  S  G  G  S  E  F  A  E  D
  76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGGTTCCGGAGGCAGCGAATTCGCTGAGGAC
  51 W  L  D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G
 151 TGGCTGGACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGA
 226                                  BamHI g|gatcc264
  76 R  S  D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G
 226 CGCTCAGACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGC
 101 P  A  G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L
 301 CCTGCGGGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTG
 126 D  S  P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S
 376 GACAGCCCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCA
 451                               BamHI g|gatcc483
 151 D  T  Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A
 451 GACACCTATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCG
 176 G  D  L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L  D  S
 526 GGCGATCTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTGGACAGC
 201 P  A  L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S  D  T
 601 CCGGCCCTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACC
 676                          BamHI g|gatcc702
 226 Y  Y  Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A  G  D
 676 TATTACCAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGAT
 251 L  T  L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  A  E  D  W  L  D  S  P  A
 751 CTCACCCTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGCTGAGGACTGGCTGGACAGCCCGGCC
 276 L  G  P  G  W  K  R  R  E  V  F  R  K  S  G  A  T  A  G  R  S  D  T  Y  Y
 826 CTGGGCCCTGGCTGGAAGCGCCGCGAAGTCTTTCGCAAGTCAGGGGCCACCGCGGGACGCTCAGACACCTATTAC
 901                      BamHI g|gatcc921
 301 Q  S  P  T  G  D  R  I  R  S  K  V  E  L  T  R  Y  L  G  P  A  G  D  L  T
 901 CAGAGCCCCACAGGAGACAGGATCCGAAGCAAAGTTGAGCTGACTCGATACCTGGGCCCTGCGGGCGATCTCACC
 326 L  F  D  F  K  Q  G  I  L  G  G  G  Q  F  G  L  Q  G  S  G  S  D  P  N  E
 976 CTCTTCGACTTCAAACAAGGCATCTTGGGCGGTGGCCAATTCGGCCTGCAGGGAAGTGGAAGTGACCCCAACGAG
 351 K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
1051 AAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
1126     XbaI t|ctaga1130
```

```
376 *  S  R
1126 TAGTCTAGA
```

F. 54 pTH-BG-VN210

```
  1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
  1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
 26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
 76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
151                                SalI g|tcgac184
 51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  A  C  L  G
151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGCCTGCCTCGGC
 76 G  R  P  A  L  D  A  V  K  K  G  L  P  H  A  P  A  L  I  K  R  T  N  R  R
226 GGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCACGCGCCTGCATTGATTAAGCGGACCAACAGAAGG
301     AvaI c|ycgrg305
101 I  P  E  R  T  S  H  R  V  A  G  L  Q  G  S  G  S  V  S  K  G  E  E  L  F
301 ATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGAAGTGGAAGTGTGAGCAAGGGCGAGGAGCTGTTC
126 T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G
376 ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC
151 E  G  D  A  T  Y  G  K  L  T  L  K  L  I  C  T  T  G  K  L  P  V  P  W  P
451 GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
176 T  L  V  T  T  L  G  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K  Q  H  D  F
526 ACCCTCGTGACCACCCTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTC
201 F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T
601 TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
226 R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E  D
676 CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGAC
251 G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N  V  Y  I  T  A  D  K  Q  K
751 GGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCACCGCCGACAAGCAGAAG
276 N  G  I  K  A  N  F  K  I  R  H  N  I  E  D  G  G  V  Q  L  A  D  H  Y  Q
826 AACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTCGCCGACCACTACCAG
301 Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  Y  Q  S  K  L  S
901 CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGC
976          XbaI t|ctaga986
326 K  D  *  S  R
976 AAAGACTAATCTAGA
```

F. 55 pTH-BG-VC210

```
  1 M  D  Y  K  D  H  D  G  D  Y  K  D  H  D  I  D  Y  K  D  D  D  D  K  M  A
  1 ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCC
 26 P  K  K  K  R  K  V  G  I  H  G  V  P  A  A  V  D  L  R  T  L  G  Y  S  Q
 76 CCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGTAGATTTGAGAACTTTGGGATATTCACAG
151                                SalI g|tcgac184
 51 Q  Q  Q  E  K  I  K  P  K  V  R  S  T  V  A  Q  H  H  E  A  L  A  C  L  G
151 CAGCAGCAGGAAAAGATCAAGCCCAAAGTGAGGTCGACAGTCGCGCAGCATCACGAAGCGCTGGCCTGCCTCGGC
```

```
 76 G  R  P  A  L  D  A  V  K  K  G  L  P  H  A  P  A  L  I  K  R  T  N  R  R
226 GGACGACCCGCCTTGGATGCGGTGAAGAAGGGGCTCCCGCACGCGCCTGCATTGATTAAGCGGACCAACAGAAGG
301     AvaI c|ycgrg305
101 I  P  E  R  T  S  H  R  V  A  G  L  Q  G  S  G  S  D  P  N  E  K  R  D  H
301 ATTCCCGAGAGGACATCACATCGAGTGGCAGGCCTGCAGGGAAGTGGAAGTGACCCCAACGAGAAGCGCGATCAC
376                                                                  XbaI t|ctaga4
43

126 M  V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  *  S  R
376 ATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAGTCTAGA
```

VITA

Charng-Yu Lin

Education

Purdue University, West Lafayette, IN, USA — Aug 2013 – Feb 2019
Ph.D. in Chemical Engineering

National Taiwan University, Taipei, Taiwan — Sep 2011 – Jul 2013
M.S. in Chemical Engineering

National Taiwan University, Taipei, Taiwan — Sep 2006 – Jun 2010
B.S. in Chemical Engineering

Honors

Purdue Graduate Student Government Travel Grant (Fall 2016)
Eastman Travel Grant (Fall 2016)

Publications

CY Lin and JC Liu, "Protein Adhesives Based on Elastin-Like Polypeptides: A Comparison between Catechol- and Thiol-Based Adhesion," in preparation for submission

CY Lin and JC Liu, "Redox-Responsive Hydrogel via Simultaneous Thiol-Ene Michael Addition and Thiol Oxidation," in preparation

CY Lin and JC Liu, "Tailoring the Temperature-Responsiveness of Positively Charged Elastin-Like Polypeptides with External Domains," submitted

SE Hollingshead, CY Lin, JC Liu. "Designing Smart Materials with Recombinant Proteins," *Macromolecular Bioscience*, 2017.

JC Liu, JJ Wilker, MJ Brennan, CY Lin, "Protein-based adhesives." US paten app. 15/212, 152, 2016

CY Lin and JC Liu, "Modular Protein Domains: An Engineering Approach toward Functional Biomaterials," *Current opinion in biotechnology*, 2016

CY Lin and L Chao, "Tunable Nucleation Time of Functional Sphingomyelinase-Lipid Features Studied by Membrane Array Statistic Tool," *Langmuir*, 2013