ENGINEERING PROTEIN-BASED MATERIALS THROUGH COILED-COIL MOTIFS

Thesis by

Kechun Zhang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2007

(Defended January 23, 2007)

ii

©2007

Kechun Zhang

All Rights Reserved

ACKNOWLEDGEMENTS

First, I thank David Tirrell, for being a great mentor and advisor, and for his constant guidance and support. I still remember the exciting moment five years ago when he welcomed me to be part of his research group. Dave's patience helped me safely struggle through those early rough years. His insight and open-mindedness are a great inspiration for me to pursue science with passion. I have been given a lot of opportunities to work on various projects and interact with people from diverse scientific backgrounds. The interdisciplinary nature of research in Dave's group is one of the things that have benefited me most these past several years. In addition, I am grateful to Dave for advising me on my career choices and painstakingly teaching me how to write papers.

I thank my thesis committee, Dennis Dougherty, Linda Hsieh-Wilson and Anand Asthagiri, for their valuable advice and comments on my thesis. I appreciate the way my proposals were challenged by these scientists during the candidacy and proposal examinations. I would also like to thank Steve Quake, Scott Fraser and Alireza Ghaffari for their generosity to allow me to use their facilities.

I also thank the former and current group members. When I first joined the group, Kent Kirshenbaum taught me the basic molecular biology techniques. Julie Liu shared her experience on culturing of mammalian cells with me. In particular, I feel the most profound gratitude towards Michael Diehl. As we worked with each other, joked with each other, and occasionally argued with each other, my English and communication

skills were greatly improved. We had fruitful collaborations that led to two chapters of this thesis. I thank Wei Shen, who initially proposed the idea on the work described in chapter 5. This project exposed me to a higher-dimensional thinking about materials. I would also like to express my heartful thanks to Rebecca Connor and Stacey Maskarinec for their proofreading and comments on my thesis. They helped me a lot in improving my writing in English. I enjoyed working with Shawn, Sunny and Mariam. Interactions with these undergraduate students gave me memorable tastes of mentoring. Also, discussions with other group members such as James Link, Tae Hyeon Yoo, Ayae Sugawara, Inchan Kwon, Charles Liu, Katherine Poulin, Kimberly Beatty, Jim Van Deventer, Eileen Fong, Nick Fisk, John Ngo, Caglar Tanrikulu, and Chihoon Ahn were very helpful.

Life at Caltech would be less enjoyable and colorful without my good friends here. Tingwei Mu was extremely helpful when I first arrived at Caltech. Friendship and support from Xiaoli Feng, Chengzhong Zhang, Zhipu Jin, Bolin Lin, Yougen Li, Robert Bao, etc. are memorable. It has been a lot of fun to play soccer and basketball with so many nice people. Those weekly games are very refreshing and keep me energetic.

Finally, I thank my family for encouraging me to pursue education and knowledge since my childhood. Without this driving force, I would not be able to finish the long journey from a small town in China to a PhD at Caltech. I truly thank my wife, Xiao Feng, for her deep love, for taking care of everything at home and putting up with my midnight working hours.

ABSTRACT

Natural biomaterials are highly organized from the molecular to the macroscopic level in a hierarchical manner, requiring synthetic technologies to achieve this level of complexity. A biosynthetic approach to material design has emerged as an attractive option. In particular, proteins represent a promising class of molecules for creating new materials due to their determined sequence and structure. The research described in this thesis focuses on engineering protein-based materials using coiled-coil motifs. The coiled coil is a common protein architecture consisting of two or more α-helices wrapped around one another to form a supercoil. Despite its simple conformation, the coiled-coil motif plays diverse roles in biological systems functioning as sensors, recognition elements, scaffolds, levers, rotating arms and springs.

First, a designed parallel heterodimeric leucine zipper pair was used as the protein capture domain to construct an artificial polypeptide scaffold for surface functionalization. By using a mutant *E. coli* phenylalanyl-tRNA synthetase, the photoreactive amino acid *para*-azidophenylalanine was incorporated. This protein polymer was spin-coated and photocrosslinked to octyltrichlorosilane-treated surfaces. The resulting protein films were shown to immobilize recombinant proteins through association of coiled coil heterodimer. Furthermore, in conjunction with microfluidic chips that were specifically designed for on-chip mixing using laminar flow, gradients of leucine zipper tagged proteins were formed in the microchannels and immobilized on the engineered protein films. This provides a general technique for producing surface-bound multicomponent gradients. The adhesion of human umbilical vein endothelial cells cultured on a surface-bound gradient

of cell binding ligands generated by this technique was examined. In addition, to generate protein walkers that have different lateral mobility rates on a surface, several variants of the leucine zipper pair with tunable heterodimerization affinities were designed and synthesized to allow diversity in the association strength of proteins linked to a surface.

The coiled-coil motif was also used to construct protein hydrogels. Hydrogels formed from a triblock artificial protein bearing dissimilar helical coiled-coil end domains (P and A) erode more than one hundred fold slower than hydrogels formed from those bearing the same end domains (either P or A). The reduced erosion rate is a consequence of the fact that looped chains are suppressed because P and A tend not to associate with each other. Thus, by harnessing selective molecular recognition, discrete aggregation number and orientational discrimination of coiled-coil protein domains, the erosion rate of hydrogels can be tuned over several orders of magnitude.

Finally, a biosynthetic approach was developed to control and probe cooperativity in multiunit biomotor assemblies by linking molecular motors to artificial protein scaffolds using the heterodimeric leucine zipper pair. This approach provides precise control over spatial and elastic coupling between motors. Cooperative interactions between monomeric kinesin-1 motors attached to protein scaffolds enhance hydrolysis activity and microtubule gliding velocity. However, these interactions are not influenced by changes in the elastic properties of the scaffold, distinguishing multimotor transport from that powered by unorganized monomeric motors. These results highlight the role of supramolecular architecture in determining mechanisms of collective transport.

TABLE OF CONTENTS

List of Schemes and Tables

List of Figures

CHAPTER I

Introduction

1.1 Introduction to artificial protein polymers

1.1.1 Background

Nature is a master in building exquisite biological materials with unique organization and function. Despite a limited set of starting materials that include amino acids, nucleotides, lipids and sugars, diverse products have been created and perfected over billions of years of evolution. These material systems support many important functions in living organisms. For instance, protein and RNA scaffolds help assemble polymerase and ribosome complexes, two essential self-replicating factories in the biological world. Furthermore, fibrous proteins, such as collagens, elastins and silks, are important structural proteins due to their superior mechanical properties. Collagens are major constituents of ligaments, cartilage and bone. Elastins are highly resilient and stretchy proteins found in connective tissues. Silks are nature's high-performance fibers having a remarkable combination of strength, stiffness and extensibility.

In many respects, the sophistication of nature provides a limitless source of templates for material scientists to utilize to generate new biomaterials. However, the construction of functional biomimetic materials still remains a great challenge. A major obstacle stems from the limitation of traditional synthetic approaches, which are unable to control the precise length and structure of polymer products. In contrast,

natural biomaterials are highly organized from the molecular to the macroscopic level in a hierarchical manner, requiring a synthetic technology that achieves this level of complexity. With the development of recombinant DNA technology, a biosynthetic approach to material design has emerged as an attractive option. In particular, proteins represent a promising class of molecules for creating new materials.[1,2] Genetic engineering allows artificially designed polypeptides to be synthesized in host organisms such as *Escherichia coli* (*E. coli*) and *Saccharomyces cerevisiae* (yeast) by utilizing the natural translational machinery. Compared to synthetic polymers, these bio-derived protein polymers have determined amino acid sequences, which dictate their folding into secondary structures such as alpha helix or beta sheet. Further organization of these secondary structures allows the formation of well defined three-dimensional structures. Nature has provided a diverse repertoire of both structural elements and bioactive elements encoded within protein sequences that can be used as building blocks and arranged to create a variety of macromolecules with interesting properties.

1.1.2 Design and application of protein polymers

Over the past decade, various protein polymers have been synthesized and characterized, demonstrating a high degree of control of macromolecular architectures. In 1994, Krejchi et al. showed that β-sheet assemblies from periodic polypeptides formed needle-shaped lamellar crystals, where the lamella thickness was determined

by the sequence periodicity.[3] In addition to sequence control, molecular weight distribution is another important factor in polymer chemistry, and its effect on material properties can not be easily investigated with synthetic polymers. For example, a polydisperse sample of the polypeptide polybenzyl-L-glutamate (pBLG) prepared by N-carboxy-α-amino acid anhydride (NCA) polymerization may exhibit either cholesteric or nematic liquid crystal behavior. However, monodisperse pBLG obtained from post-translational modification of a biosynthetically produced sample of poly L-α-glutamic acid adopted a twisted smetic liquid crystal phase with precise control of layer spacing on the scale of tens of nanometers.[4]

The integration of structural elements and bioactive elements into a single protein has also become a new paradigm for designing protein-based materials. There exist many functional domains in nature that are useful for constructing advanced materials. These include architectural motifs for the creation of fibers, hydrogels, elastic materials, and capsules as well as enzymatic motifs for catalysis, binding and signaling. In principle, artificial protein polymers may have any combination of these diverse functions. Such potential makes protein-based materials ideal candidates for tissue engineering. For example, artificial extracellular matrix proteins have been engineered and investigated as a potential replacement for damaged blood vessels.[5,6] These materials contain cell binding domains such as RGD or REDV, which are known to support the adhesion of endothelial cells,[7] and elastin-mimetic domains based on the pentapeptide sequence VPGVG, which upon crosslinking forms a

homogeneous film and exhibits mechanical properties similar to those of the arterial wall.[8]

More recently, by incorporating the active domain of Delta/Serrate/Lag2 (DSL) ligand into the elastin-based backbone of this protein, it is possible to create an artificial niche to guide the differentiation of neural stem cells.[9] These materials have the potential to be used as implants for neuroregneration. Another possible application of this type of elastic protein polymers is for surgical refractive correction of the cornea. Rather than ablative vision surgeries, synthetic protein-based corneal onlays can be implanted to promote the adhesion of epithelial cells critical to restore normal function of the cornea.[10]

Even though protein engineering has achieved great success in material design, one drawback of using protein polymers is that the choice of monomers is limited. Whereas organic chemists can use thousands of different monomeric building blocks to make synthetic polymers, biosynthetic protein polymers are limited to the twenty natural amino acids. Extensive efforts have been made to expand the amino acid repertoire from canonical amino acids to unnatural analogs with diverse chemical functionalities. From a materials science perspective, reassignment of sense codons to noncanonical amino acids *in vivo* is particularly relevant because global incorporation of amino acid analogs will not only affect bulk material characteristics, but also result in higher protein yields compared to site-specific incorporation.[11,12] Such unnatural replacement is generally achieved by forcing auxotrophic bacterial expression hosts

that are deficient in synthesizing particular amino acids to use chemically similar amino acid analogs supplied in the culture media. For the past several years, the introduction of noncanonical amino acids has brought new chemical, physical and biological properties into engineered protein polymers. For example, incorporation of fluorinated amino acid analogs could dramatically increase the melting temperature of coiled-coil domains[13,14] and collagen-like triple-helices[15]. Replacement with a phenylalanine analog 3-thienylalanine turned the repetitive beta-sheet forming polypeptide, $([Ala\text{-}Gly]_3Phe\text{-}Gly)_{13,}$ into precursors for synthesizing conducting polymers.[16] Lastly, the reactive azide group offers a simple way to post-translationally modify newly synthesized proteins containing azidophenylalanine[17] or azidohomoalanine[18-21].

1.2 The coiled-coil motif

The research described in this thesis focuses on engineering protein-based materials using coiled-coil motifs. The coiled coil is a common protein architecture consisting of two or more α-helices wrapped around one another to form a supercoil. Using a structure-prediction algorithm, it was estimated that roughly 10% of the eukaryotic proteins contain coiled-coil domains.[22] This motif was first discovered in the intermediate filament protein α-keratin, a major structural component of cytoskeleton.[23] The elementary building block of intermediate filaments is a highly elongated coiled-coil dimer, which initiates the assembly of mature filaments.[24] In

addition to the cytoskeleton, motor proteins such as myosin and kinesin contain coiled-coil domains that are important for transduction of mechanical force; SMC (structural maintenance of chromosomes) proteins use coiled-coil motifs to regulate the organization of chromatin; membrane bound coiled-coil proteins such as golgins support the membrane structure of the cell; and transcription factors including Fos/Jun and GCN4 modulate the transcription activities of target genes through specific coiled coil pairing.[25,26] Interestingly, the oligomerization of certain coiled-coil proteins is regulated by environmental conditions such as pH, temperature, hydration and solute availability.[25] Thus, despite its simple conformation, the coiled-coil motif plays diverse roles in biological systems functioning as sensors, recognition elements, scaffolds, levers, rotating arms and springs.

1.2.1 The structure of coiled coils

The structure of coiled coils was first described by Crick to explain the X-ray diffraction pattern of α-keratin.[27] He suggested that α-helices pack together with a rotation angle of 20° from parallel. A “knobs-into-holes” packing mode was then proposed and considered to be the central determinant of the coiled-coil structure, where apolar side chains from one helix pack into cavities formed by the surrounding helices. This theory was not verified until Alber and coworkers solved the X-ray structure of a peptide corresponding to the leucine zipper of the yeast transcription factor GCN4 in 1991.[28] The leucine zipper motif is a special type of coiled-coil

protein. Since then, high resolution structures of other proteins containing two-stranded, three-stranded, four-stranded or even five-stranded coiled coils have also been determined.[29,30] These structures, combining studies of mutant and designed coiled-coil peptides,[31-34] have provided detailed information on the molecular structure and interaction of coiled coils.

The general amino acid sequence of a coiled coil is characterized by a seven-residue repeat, $(\mathbf{abcdefg})_n$, with the first (**a**) and fourth (**d**) positions frequently occupied by apolar amino acids that form the hydrophobic core. Amino acids in the remaining positions are hydrophilic and form the solvent-exposed part of the coiled coil. In particular, core-flanking residues at the **e** and **g** positions are populated with charged amino acids such as glutamic acid and lysine, conferring electrostatic interactions between helices.[34-37] The hydrophobicity of the core residues plays an important role in determining the orientation and number of helices in a coiled coil. In GCN4, there is an asparagine located at a core **a** position; mutation of this residue to valine leads to the formation of a mixture of dimers and trimers.[32] While mutation of this residue to leucine changes the original parallel dimer into a mixture of parallel and antiparallel tetramers.[33] On the other hand, the pairing specificity of coiled coils is governed by the attractive or repulsive electrostatic interactions between the **e** and **g** positions on opposing strands. The charge pattern in these positions dictates the preference for homo- or hetero-oligomeric association of helices into a coiled coil. Based on the number of ionic interactions between **g**/**e'** (prime indicates residues from

an adjacent chain) pairs, the heterodimerization preferences of bZip proteins can be predicted using an interhelical salt bridge rule.[38] This rule has been used to design synthetic peptides that specifically form heterodimers[39], -trimers[40], and –tetramers[41].

1.2.2 Model coiled-coil peptides

A variant of the coiled-coil motif, the leucine zipper, has leucine residues occupying 80% or more of the **d** positions. The name was coined by McKnight because he thought leucine residues lining the α-helix interlocked like the teeth of a zipper.[42] Later, it was discovered that the leucine residues actually did not zip together, instead they snapped into the holes formed by the neighboring helix as proposed in Crick's model.[28] The leucine zipper is a small and well-foldable protein motif, which can be engineered to produce long fibers through 'sticky-end' extension,[43] or act as multimerization domains for assembly of nanoparticles[44] and high-affinity multivalent antibodies[45]. The following leucine zipper peptides, either naturally occurring or artificially designed, have been utilized to construct functional materials.

(a) A1

The A1 leucine zipper domain was originally designed by McGrath et al.[46] It is composed of six heptad repeats, and has a total of 42 amino acids (Figure 1.1). Residues in the hydrophobic core were selected based on the distribution pattern of **a/d** residues in the *Jun* oncogene product.[47] An algorithm developed by Lupas et al.[48]

was used to choose the amino acid residues occupying the **b**, **c**, and **f** positions. In order to control the association of coiled coil structure, glutamic acid residues were placed at nine of the twelve **e** and **g** positions. Under basic conditions, deprotonation of glutamic acid residues introduces electrostatic repulsion between the parallel strands and leads to dissociation of the coiled coil assembly.

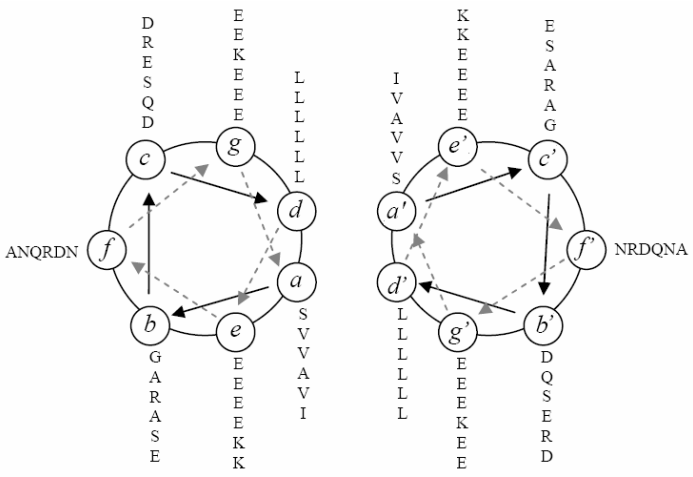


Figure 1.1 Helical wheel representation of the parallel A1 homodimer.

(b) COMPcc pentamer

Cartilage oligomeric matrix protein (COMP), a non-collageneous glycoprotein, was found first in cartilage[49] and later in tendon[50] and ligaments[51]. COMP is a 524 KDa bouquet-like complex with five identical subunits, which consist of a C-terminal globular cell binding domain, followed by seven calcium-binding domains, four epidermal growth factor (EGF)-like domains, and an N-terminal coiled-coil region (COMPcc). The assembly of COMP is initiated by organizing COMPcc into a

five-stranded bundle.[52] The high resolution X-ray crystal structure[30] shows that COMPcc is a 7.3 nm long coiled coil with a diameter of 3 nm. In the pentameric complex, there are four types of 'knobs-into-holes' interactions (Figure 1.2): residues at **a**, **d**, **e** and **g** positions pack into holes formed by residues at **a'-g'**, **e'-d'**, **c-d'** and **a'-b'** respectively. Another distinguishing feature of the hydrophobic core is that five glutamine (Gln[54]) side chains form a ring of hydrogen bonds. Interchain salt bridges between Asp[46] and Arg[48'] (**c**/**e'** interaction) and between Glu[57] and Lys[62'] (**g**/**e'** interaction) also contribute to the stability of the complex. It is believed that Asn[41] is the key to specific formation of the pentamer; mutation of this residue to leucine favors the formation of a tetramer.[30]

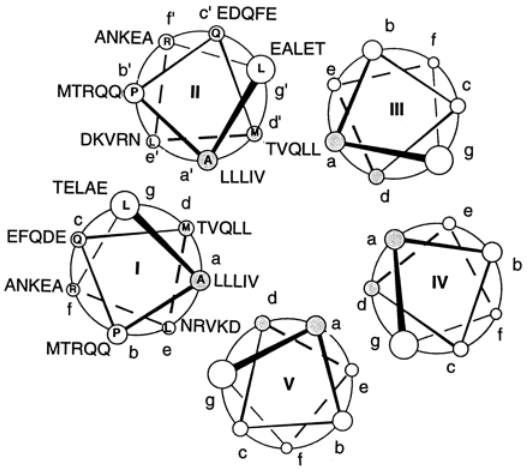


Figure 1.2 Helical wheel representation of the COMPcc pentamer.

(adapted from reference 30)

(c) ZE/ZR heterodimer

The ZE/ZR heterodimeric leucine zipper pair was derived from

vitellogenin-binding protein (VBP), a bZIP homodimer.[53] An asparagine was placed at the second **a** position to limit oligomerization to dimers,[32] and to direct the parallel orientation of helices in the coiled-coil complex.[54] All residues at the **e** and **g** positions of the first four heptads were changed to glutamic acid or arginine, respectively, to produce acidic peptide ZE and basic peptide ZR. Thus, both ZE and ZR homodimers have four pairs of electrostatic repulsions, while the ZE/ZR heterodimer contains four pairs of attractive salt bridges. Moll et al. showed that this leucine zipper system has a heterodimerization affinity of 10^{-15} M, similar to that reported for the streptavidin/biotin system, while homodimerization affinities are in the micromolar range.[55]

def gabcdef gabcdef gabcdef gabcdef gabcdef gabcd

ZR: LEI RAAFLRR RNTALRT RVAELRQ RVQRLRN IVSQYET RYGPL

ZE: LEI EAAFLEQ ENTALET EVAELEQ EVQRLEN IVSQYET RYGPL

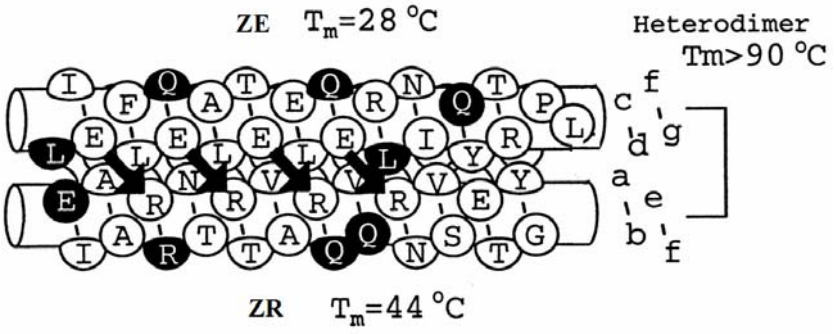


Figure 1.3 Schematic of the ZE/ZR heterodimer.

(adapted from reference 54)

1.3 Thesis outline

Chapter 2 focuses on the design and synthesis of an artificial polypeptide scaffold for surface functionalization and its application to immobilization of leucine zipper tagged proteins. Chapter 3 discusses the use of this protein immobilization technique to generate surface-bound multicomponent protein gradients through microfluidics technology. The adhesion of human umbilical vein endothelial cells (HUVEC) cultured on these surface-bound protein gradients was investigated. Chapter 4 focuses on progress towards the creation of protein walkers. Engineered variants of leucine zipper pairs that have tunable heterodimerization affinities were created for this purpose. Chapter 5 introduces the design of a new hydrogel material composed of an artificial triblock protein polymer bearing dissimilar helical coiled-coil end domains, which exhibits greatly improved mechanical properties and stability in open systems. Chapter 6 presents a biosynthetic approach to control and probe cooperativity in multiunit biomotor assemblies by linking molecular motors to artificial protein scaffolds.

1.4 References

(1) van Hest, J. C. M.; Tirrell, D. A. *Chem. Commun.* **2001**, *7*, 1897.

(2) Maskarinec, S. A.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2005**, *16*, 422.

(3) Krejchi, M. T.; Atkins, E. D.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science.* **1994**, *265*, 1427.

(4) Yu, S. M.; Conticello, V. P.; Zhang, G.; Kayser, C.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Nature* **1997**, *389*, 167.

(5) Nicol, A.; Gowda, D. C.; Urry, D. W. *J. Biomed. Mater. Res.* **1992**, *26*, 393.

(6) Panitch, A.; Yamaoka, T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1999**, *32*, 1701.

(7) Humphries, M. J.; Akiyama, S. K.; Komoriya, A.; Olden, K.; Yamada, K. M. *J. Cell Biol.* **1986**, *103*, 2637.

(8) Nowatzki, P. J.; Tirrell, D. A. *Biomaterials.* **2004**, *25*, 1261.

(9) Liu, C. Y.; Westerlund, U.; Svensson, M.; Moe, M. C.; Varghese, M.; Berg-Johnsen, J.; Apuzzo, M. L.; Tirrell, D. A.; Langmoen, I. A. *J. Hematother. Stem Cell Res.* **2003**, *12*, 689.

(10) Tirrell, D. A.; Schwartz, D. M.; Nowatzki, P. J.; Grubbs, R. H. *U.S. Patent application No. 11/040,130* **2005**.

(11) Link, A. J.; Tirrell, D. A. *Methods* **2005**, *36*, 291.

(12) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.

(13) Tang, Y.; Ghirlanda, G.; Petka, W. A.; Nakajima, T.; DeGrado, W. F.; Tirrell, D. A. *Angew Chem. Int. Ed. Engl.* **2001**, *40*, 1494.

(14) Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.

(15) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666.

(16) Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *J. Am. Chem. Soc.*

1995, *117*, 536.

(17) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235.

(18) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 19.

(19) Link, A. J.; Tirrell, D. A. *J. Am. Chem. Soc.* **2003**, *125*, 11164.

(20) Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 14150.

(21) Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9482.

(22) Liu, J.; Rost, B. *Protein Sci.* **2001**, *10*, 1970.

(23) McArthur, I. *Nature* **1943**, *152*, 38.

(24) Strelkov, S. V.; Herrmann, H.; Aebi, U. *Bioessays* **2003**, *25*, 243.

(25) Rose, A.; Meier, I. *Cell Mol. Life Sci.* **2004**, *61*, 1996.

(26) Burkhard, P.; Stetefeld, J.; Strelkov, S. V. *Trends Cell Biol.* **2001**, *11*, 82.

(27) Crick, F. H. C. *Acta Crystallogr. Sect. A* **1953**, *6*, 689.

(28) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539.

(29) Lupas, A. *Trends Biochem. Sci.* **1996**, *21*, 375.

(30) Malashkevich, V. N.; Kammerer, R. A.; Efimov, V. P.; Schulthess, T.; Engel, J. *Science* **1996**, *274*, 761.

(31) Lovejoy, B.; Choe, S.; Cascio, D.; McRorie, D. K.; DeGrado, W. F.; Eisenberg, D. *Science* **1993**, *259*, 1288.

(32) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.

(33) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642.

(34) Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849.

(35) O'Shea, E. K.; Lumb, K. J.; Kim, P. S. *Curr. Biol.* **1993**, *3*, 658.

(36) Graddis, T. J.; Myszka, D. G.; Chaiken, I. M. *Biochemistry* **1993**, *32*, 12664.

(37) Yu, Y.; Monera, O. D.; Hodges, R. S.; Privalov, P. L. *J. Mol. Biol.* **1996**, *255*, 367.

(38) Vinson, C. R.; Hai, T.; Boyd, S. M. *Genes Dev.* **1993**, *7*, 1047.

(39) Zhou, N. E.; Kay, C. M.; Hodges, R. S. *J. Mol. Biol.* **1994**, *237*, 500.

(40) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645.

(41) Fairman, R.; Chao, H. G.; Lavoie, T. B.; Villafranca, J. J.; Matsueda, G. R.; Novotny, J. *Biochemistry* **1996**, *35*, 2824.

(42) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759.

(43) Ryadnov, M. G.; Woolfson, D. N. *Nat. Mater.* **2003**, *2*, 329.

(44) Stevens, M. M.; Flynn, N. T.; Wang, C.; Tirrell, D. A.; Langer, R. *Adv. Mater.* **2004**, *16*, 915.

(45) Terskikh, A. V.; Le Doussal, J. M.; Crameri, R.; Fisch, I.; Mach, J. P.; Kajava, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1663.

(46) McGrath, K. P.; Butler, M. M.; DiGirolamo, C. M.; Kaplan, D. L.; Petka, W.

A.; Laue, T. M. *J. Bioact. Compat. Polym.* **2000**, *15*, 334.

(47) O'Shea, E. K.; Rutkowski, R.; Kim, P. S. *Cell* **1992**, *68*, 699.

(48) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.

(49) Hedbom, E.; Antonsson, P.; Hjerpe, A.; Aeschlimann, D.; Paulsson, M.; Rosa-Pimentel, E.; Sommarin, Y.; Wendel, M.; Oldberg, A.; Heinegard, D. *J. Biol. Chem.* **1992**, *267*, 6132.

(50) DiCesare, P.; Hauser, N.; Lehman, D.; Pasumarti, S.; Paulsson, M. *FEBS Lett.* **1994**, *354*, 237.

(51) Muller, G.; Michel, A.; Altenburg, E. *Connect Tissue Res.* **1998**, *39*, 233.

(52) Efimov, V. P.; Lustig, A.; Engel, J. *FEBS Lett.* **1994**, *341*, 54.

(53) Krylov, D.; Barchi, J.; Vinson, C. *J. Mol. Biol.* **1998**, *279*, 959.

(54) Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603.

(55) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, *10*, 649.

CHAPTER II

Artificial Polypeptide Scaffold for Protein Immobilization

Abstract

An artificial polypeptide scaffold composed of surface anchor and protein capture domains was designed and expressed *in vivo*. By using a mutant *E. coli* phenylalanyl-tRNA synthetase, the photoreactive amino acid *para*-azidophenylalanine was incorporated into the surface anchor domain. Octyltrichlorosilane-treated surfaces were functionalized with this polypeptide by spin coating and photocrosslinking. The resulting protein films were shown to immobilize recombinant proteins through formation of coiled coil heterodimers.

The text in this chapter is reprinted with permission from Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 10136-10137. Copyright 2005, American Chemical Society.

2.1 Introduction

Protein microarray technologies are beginning to advance the field of proteomics by providing miniaturized platforms to probe the interactions and functions of proteins.[1] Presenting proteins in dense arrays enables the rapid screening of thousands of molecular events in a single experiment. This capability should facilitate the elucidation of protein profiles in organisms, the discovery of novel protein functions, and the development of systems-level understanding of biological phenomena.[2] The utility of microarrays to probe protein-protein interactions has been demonstrated by Zhu et al., where new calmodulin- and phospholipid-binding proteins were identified by screening a full scale yeast proteome microarray.[3] More recently, Nielsen et al. used antibody microarrays to profile the activation of receptor tyrosine kinases and analyze signal transduction networks in mammalian cells.[4]

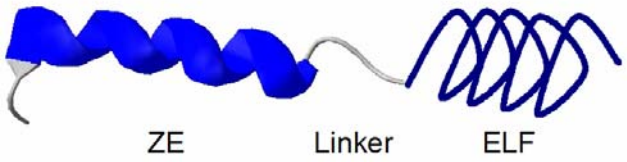
Despite the growing success of protein microarrays, it remains a central challenge to develop simple and general techniques to immobilize functional proteins onto solid supports. In certain cases, conventional immobilization methods (such as physical adsorption or covalent binding through lysine and cysteine residues) render active sites inaccessible or even denature proteins.[5] This difficulty may be overcome by engineering site specific attachment to a substrate through expression of recombinant fusion proteins bearing affinity tags. For example, the immobilization of his tag fusion proteins onto Ni-NTA functionalized slides has been shown to maintain higher protein activity than direct attachment to aldehyde slides.[3] Nevertheless, his tags do

not necessarily provide a general approach to array fabrication, because the binding interaction is sensitive to pH and to some common buffer components.[6] To solve this problem, a variety of alternative strategies are being developed. One such approach uses the strong interaction between avidin and biotin to immobilize proteins in combination with *in vivo* or *in vitro* biotinylation.[7] Other methods introduce polypeptide tags to effect selective and covalent attachment.[8] Optimal immobilization schemes should be characterized by simple cloning schemes, efficient protein expression, selective affinity and simple surface chemistry. This formidable challenge requires the design of new biomaterials that maintain protein architecture and allow specific chemistries to be utilized for immobilization.[9]

2.2 Results and discussion

We have approached this problem by creating an artificial polypeptide scaffold **1** that can be used to immobilize recombinant proteins on substrates (Figure 2.1). The polypeptide contains separate surface anchor and protein capture domains, and uses an artificial amino acid to covalently crosslink the polypeptide to surfaces. The protein capture domain functions through coiled coil association of a designed parallel heterodimeric leucine zipper pair, designated ZE and ZR. These structures are based on the sequences developed by Vinson et al.[10] with minor modifications (see materials and methods). Vinson et al. showed that this leucine zipper system has a heterodimerization affinity of 10^{-15} M, while homodimerization affinities are in the

micromolar range. The acidic component ZE is incorporated into **1** as the protein capture domain and the basic portion ZR is fused to target proteins as an affinity tag.



def gabcdef gabcdef gabcdef gabcdef gabcdef gabcd

ZE LEI EAAALEQ ENTALET EVAELEQ EVQRLEN IVSQYRT RYGPL

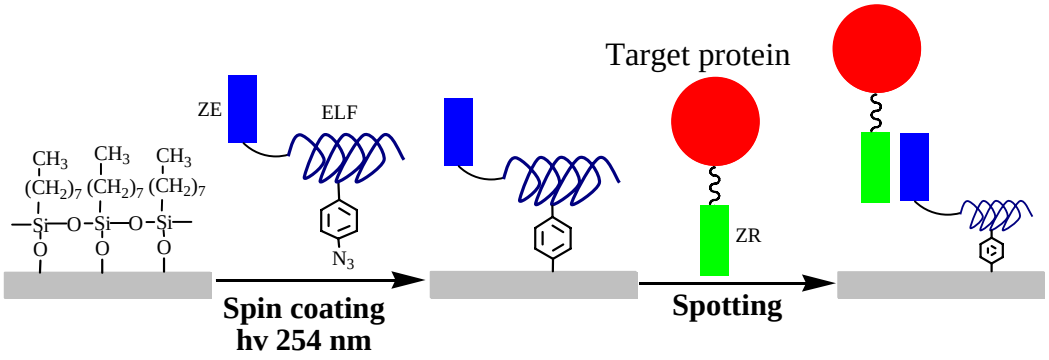
ZR LEI RAAALRR RNTALRT RVAELRQ RVQRLRN EVSQYET RYGPL

Linker $G(GS)_6G$ **ELF** $[(VPGVG)_2VPGFG(VPGVG)_2]_5VPGC$

Figure 2.1 Design of the artificial polypeptide scaffold **1** and related amino acid sequence.

An important part of the scaffold design is introduction of an elastin mimetic domain ELF for surface anchorage. ELF consists of five repeats of 25 amino acids with the sequence $(VPGVG)_2VPG\mathit{F}G(VPGVG)_2$. Because of its hydrophobic character, ELF provides strong adhesion to hydrophobic surfaces.[11] Moreover, because **1** is expressed in a bacterial host harboring a mutant *E. coli* phenylalanyl-tRNA synthetase (A294G), the phenylalanine residues in the ELF domain are partially replaced by a photoreactive nonnatural amino acid, *para*-azidophenylalanine.[12] This moiety can be used to generate covalent linkages to substrates upon UV irradiation. In order to reduce possible steric hindrance, the ZE and ELF domains are linked by a flexible spacer of 14 amino acids. The designed protein sequence was reverse-translated based

on the codons most often used in *E. coli* and expressed in the phenylalanine auxotroph strain AF-IQ.[13] Typical yields were 50 mg/L and the rate of incorporation of *para*-azidophenylalanine was approximately 45% as determined by amino acid analysis.



Scheme 2.1 Surface functionalization and coiled-coil mediated immobilization of proteins.

The successful design, *in vivo* expression and purification of **1** allowed us to prepare a functionalized surface for protein immobilization (Scheme 2.1). In this procedure, a solution of **1** (0.8 mg/ml in 50% trifluoroethanol) was spin-coated on glass slides that were pretreated with octyltrichlorosilane (OTS) to make them hydrophobic. Once dry, the protein films were irradiated with UV light.[14] Irradiation of the films covalently crosslinked the protein to the substrate through photodecomposition of the arylazide groups[15] of *para*-azidophenylalanine. Any noncovalently bound protein was removed by sonicating in 80% DMSO for 20 minutes. Measurements of the water static contact angle indicated marked changes in

wettability upon formation of the protein film; the contact angle was 60^o after photocrosslinking and sonication as compared to 107^o for the initial OTS substrates. In a final step, films were blocked with 1% casein solution to reduce nonspecific protein adsorption.

Green fluorescent protein (GFP) and glutathione-S-transferase (GST) were chosen as model systems to test the efficacy of the functionalized surface for protein immobilization. These proteins were expressed *in vivo* with the ZR tag fused to their C-termini. Proteins lacking the fusion tag were expressed as controls. Purified proteins at a concentration of 5 μM were spotted onto the surface to generate protein microarrays (Fig 2.2a). The arrays were incubated in a humid chamber for 1 hour and then thoroughly washed twice with PBS-Tween buffer (PBS plus 0.5% Tween-20) to remove nonspecifically bound protein. Each array was probed with a mixture of cy3-anti-GST and alexa647-anti-GFP (4 μg/ml each), washed, and scanned with a Genepix microarray scanner. As expected, spots containing fusion proteins showed much stronger signals than control proteins. The average signal to noise ratio SNR[16] of GST-ZR was 196 ± 20 and that of GFP-ZR was 43 ± 4. Without the ZR fusion, GST spots yielded weaker detectable signals (SNR 15 ± 3[17]), and GFP spots could not be distinguished from background. The sensitivity of the method is high; SNR ratios of 4 or 2 (GST-ZR or GFP-ZR) were obtained when proteins were spotted at concentrations as low as 50 nM. These qualities encouraged us to examine immobilization of ZR tagged proteins directly from crude cell lysates. Cell lysates containing overexpressed fusion or control proteins were spotted onto functionalized surface and detected by the procedure described above. As shown in Fig 2.2b, significant protein attachment occurred only when the complementary zipper fusion

tag was present.

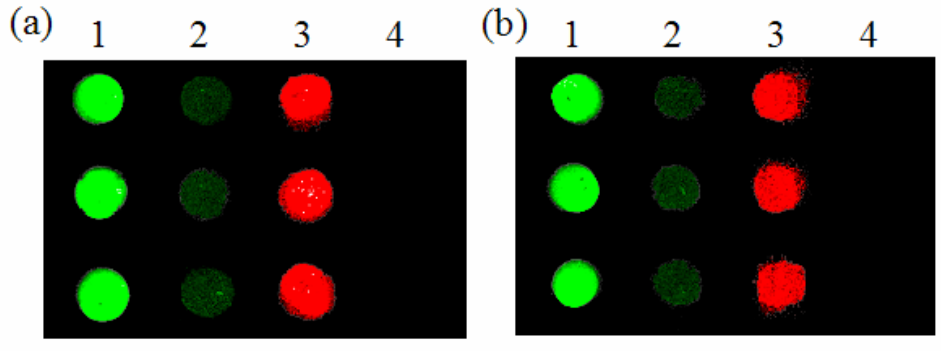


Figure 2.2 Immobilization and detection of proteins on polypeptide functionalized surfaces. Printed spots were detected with a mixture of cy3-anti-GST and alexa647-anti-GFP. (a) 5 μM purified proteins (1) GST-ZR, (2) GST, (3) GFP-ZR, (4) GFP. & (b) Cell lysates (1) GST-ZR, (2) GST, (3) GFP-ZR, (4) GFP. The spots are 200 μm in diameter.

The protein immobilization method presented here has several advantages over traditional methods. First, spin coating plus photoimmobilization provides a simple and convenient route to uniform protein films. This procedure requires a fabrication time of minutes and yields dense surface coverage. Second, the heterodimeric association of this leucine zipper system is highly specific and stable. In fact, we have found that the heterodimer forms even in 8 M urea solution over a wide range of pH values (tested from pH 4.0 to 8.0). This stability expands the range of working conditions to stringent situations where other methods are not applicable. Third, considering the relatively small size of the zipper tag (43 amino acids), it is unlikely

that the function of the fusion proteins will be compromised. Finally, direct immobilization of fusion proteins from crude cell lysates makes it feasible to fabricate protein arrays in high throughput fashion by eliminating time consuming and costly purification steps.

2.3 Materials and methods

(This section appeared as supporting information for Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. J. Am. Chem. Soc. ***2005****, 127, 10136-10137.)*

2.3.1 Cloning and expression of the polypeptide scaffold

(a) Sequence optimization

def gabcdef gabcdef gabcdef gabcdef gabcdef gabcd

ZE LEI EAAALEQ ENTALET EVAELEQ EVQRLEN IVSQYRT RYGPL

ZR LEI RAAALRR RNTALRT RVAELRQ RVQRLRN EVSQYET RYGPL

Figure 2.3 Sequence modifications of the heterodimeric leucine zipper.

Compared to the leucine zipper sequences reported by Vinson et al.,[10] some modifications (shown in red) were made to enhance the application described here. Phenylalanine residues in the first heptads of both coils were replaced by alanine. In order to provide an additional salt bridge, glutamic acid in the fifth heptad of ZE was changed to arginine and isoleucine in ZR was changed to glutamic acid.

(b) Vector construction

The genes for ZE and ZR were constructed by assembling synthetic oligonucleotides through PCR. A *BamH*I fragment encoding ZE with a C-terminal flexible linker $(Gly\text{-}Ser)_6$ was ligated in frame with a synthetic gene encoding ELF in a pQE60 (Qiagen) plasmid developed in this laboratory. To avoid *in vivo* degradation at 37 ^{o}C and enhance the expression yield, a dicistronic construct (Figure 2.4) was used to express ZE$(gs)_6$ELF. First, a *Nhe*I fragment encoding mutant *E. coli* PheRS (A294G) was inserted into expression plasmid pQE60 to yield pQE-FS*. Then the ZR gene was ligated into the *Bam*HI site of pQE-FS* to yield ZR-pQE-FS*. Finally, a DNA fragment encoding ZE$(gs)_6$ELF and containing a ribosome binding site was amplified and ligated into the *Bpu*1102I site of ZR-pQE-FS*. The resulting plasmid (ZRhis-ZE$(gs)_6$ELF) was transformed into phenylalanine auxotroph *E. coli* strain AF-IQ.

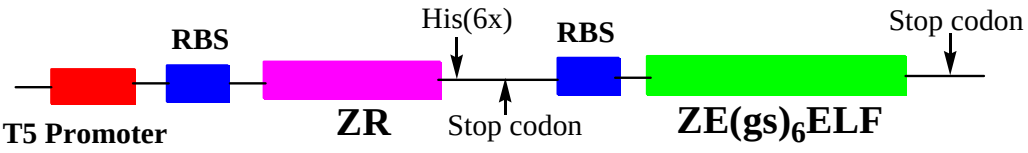


Figure 2.4 Dicistronic expression construct for the polypeptide scaffold.

(c) Protein expression and purification

Cultures were grown at 37 ^{o}C in M9 minimal medium supplemented with glucose (0.4 wt %), thiamine (5 mg L^{-1}), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), 20 amino

acids (15 mg L^{-1} Phe, 40 mg L^{-1} other amino acids), and antibiotics (200 μg/ml ampicillin and 35 μg/ml chloramphenicol). At an optical density of 1.0 at 600 nm (OD_{600}), the culture was supplemented with *para*-azidophenylalanine (0.3 g L^{-1}). After incubation for 15 minutes, IPTG (2 mM) was added to induce protein expression at 37°C for 5 hours. The cell pellet was lysed in 8 M urea solution, followed by freeze-thawing and sonication. The cell lysate was passed through a Ni-NTA column and washed with 8 M urea solution (pH=6.3). $ZE(gs)_6ELF$ was eluted with 6 M GuHCl solution (pH=7.0) and collected. The collected protein sample was subjected to dialysis against ddH_2O and freeze drying.

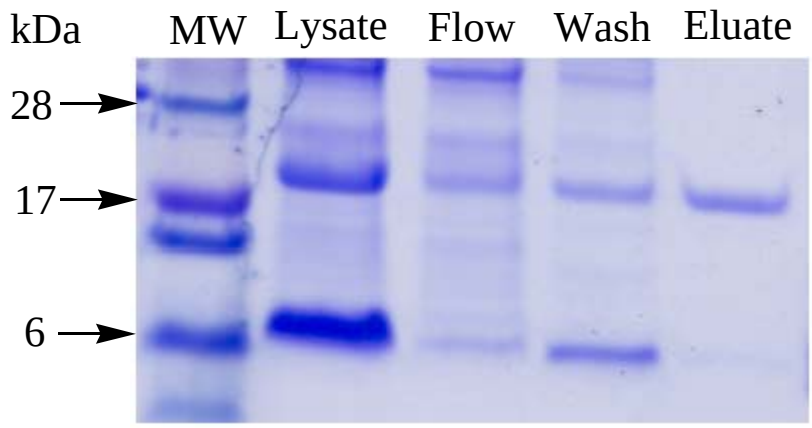


Figure 2.5 Purification of $ZE(gs)_6ELF$ by Ni-NTA affinity chromatography.

2.3.2 Cloning and expression of the fusion proteins

A GFP gene fragment was PCR amplified from pGFPuv (Clontech) and cloned into the *Pst*I site of pQE9 (Qiagen) to yield pQE9-GFP. The ZR gene assembled from synthetic oligonucleotides was digested by *Hind*III and ligated into pQE9-GFP to

yield the expression vector pQE9-GFPZR. To construct the GST-ZR fusion protein, the ZR fragment was inserted into the *Bam*HI site of expression vector pGEX-2TK (Amersham Pharmacia Biotech). The resulting expression plasmids were transformed into *E. coli* strain BL-21 and protein expression was performed in 2XYT rich medium induced by 1 mM IPTG. GFP and GFP-ZR, both of which carry N-terminal hexahistidine tags, were purified on Ni-NTA spin columns, while GST and GST-ZR were purified on glutathione columns.

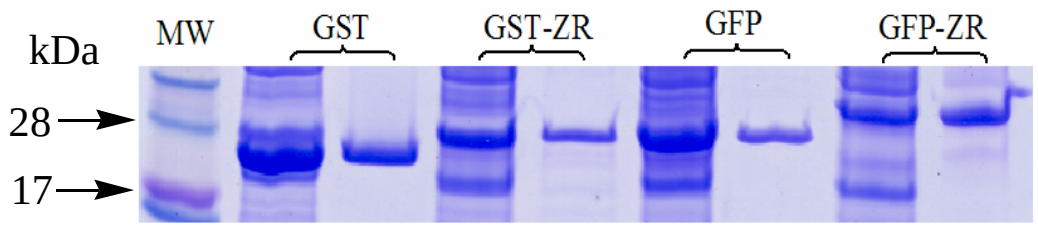


Figure 2.6 Expression and purification of model proteins.

2.3.3 Protein array procedure

(a) Substrate preparation

Standard glass slides (Corning) were immersed into concentrated H_2SO_4 for 1 hour. After washing thoroughly with water, they were immersed into a boiling solution of 1/1/5 (v/v) H_2O_2 (30%) /NH_4OH (30%)/H_2O for 30 minutes. Then the slides were gently shaken in 1% octyltrichlorosilane in toluene for 30 minutes. Finally they were cleaned twice in methanol and twice in DI water. The functionalized slides were cured at 110^{o}C for 30 minutes.

(b) Polypeptide film preparation

A solution of $ZE(gs)_6ELF$ (80 µl, 0.8 mg/ml) in 50% trifluoroethanol was applied to a glass slide and spun at 1,500 RPM for 45 seconds. The resulting film was irradiated in a photochemical reactor equipped with 254 nm wavelength UV lamps for 5 minutes. After thorough washing with 80% DMSO and ddH_2O, the slides were dried for microarray experiments. They were blocked with 1% casein for 1 hour and then stored at 4^oC for later use.

(c) Protein spotting and detection

GFP and GST fusion or control proteins diluted in printing buffer (0.5% casein, 0.5% polyvinylpyrrolidone, 15% sorbitol, 0.05% sarkosyl, 1X PBS) to a final concentration of 5 µM were spotted on slides with a custom-built microarrayer equipped with a MicroQuill® 2000 Array Pin (Majer Precision). After incubation for 1 hour in a humid chamber (saturated K_2SO_4), the slides were washed twice, each time for 15 minutes with PBS-Tween buffer (PBS plus 0.5% Tween-20). Finally, the immobilized proteins were probed with a mixture of cy3-antiGST and alexa647-antiGFP (4 µg/ml each) for 2 hours. After thorough washing with PBS-Tween buffer, fluorescence scans were acquired on a Genepix 4200A microarray scanner .

For cell lysate experiments, cell pellets from 100 ml 2XYT expression cultures were lysed with 5 ml 8 M urea solution and centrifuged to obtain clear supernatants.

Printing solutions were prepared by diluting the supernatants 20-fold into printing buffer. The protocol for microarray preparation was the same as that described above.

Acknowledgment. This work was supported by the NSF-sponsored Center for the Science and Engineering of Materials at the California Institute of Technology.

2.4 References

(1) (a) MacBeath, G.; Schreiber, S. L. *Science* **2000**, *289*, 1760. (b) Zhu, H.; Klemic, J. F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K. G.; Smith, D.; Gerstein, M.; Reed, M. A.; Snyder, M. *Nat. Genet.* **2000**, *26*, 283. (c) Haab, B. B.; Dunham, M. J.; Brown, P. O. *Genome Biol.* **2001**, *2*, 1.

(2) (a) MacBeath, G. *Nat. Genet.* **2002**, *32(suppl.)*, 526. (b) Phizicky, E.; Bastiaens, P. I.; Zhu, H.; Snyder, M.; Fields, S. *Nature* **2003**, *422*, 208.

(3) Zhu, H.; Bigin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. *Science* **2001**, *293*, 2101.

(4) Nielsen, U. B.; Cardone, M. H.; Sinskey, A. J.; MacBeath, G.; Sorger, P. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9330.

(5) Zhu, H.; Snyder, M. *Curr. Opin. Chem. Biol.* **2003**, *7*, 55.

(6) Paborsky, L. R.; Dunn, K. E.; Gibbs, C. S.; Dougherty, J. P. *Anal. Biochem.***1996**, *234*, 60.

(7) (a) Boer, E.; Rodriguez, P.; Bonte, E.; Krijgsveld, J.; Katsantoni, E.; Heck, A.; Grosveld, F.; Strouboulis, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7480. (b) Lesaicherre, M. L.; Lue, R. Y.; Chen, G. Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2002**, *124*, 8768. (c) Lue, R. Y.; Chen, G. Y.; Hu, Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2004**, *126*, 1055. (d) Yin, J.; Liu, F.; Li, X.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, 126, 7754.

(8) (a) Hodneland, C. D.; Lee, Y. S.; Min, D. H.; Mrksich, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5048. (b) Kindermann, M.; George, N.; Johnsson, N.; Johnsson, K. *J. Am. Chem. Soc.* **2003**, *125*, 7810. (c) Camarero, J. A.; Kwon, Y.; Coleman, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 14730.

(9) Langer, R.; Tirrell, D. A. *Nature* **2004**, *428*, 487.

(10) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, *10*, 649.

(11) (a) Shimazu, M.; Mulchandani, A.; Chen, W.; *Biotechnol. Bioeng.* **2003**, *81*, 74. (b) Nath, N.; Chilkoti, A. *Anal. Chem.* **2003**, *75*, 709. (c) Wada, A.; Mie, M.; Aizawa, M.; Lahoud, P.; Cass, A. E.; Kobatake, E. *J. Am. Chem. Soc*. **2003**, *125*, 16228. (d) Hyun, J.; Lee, W. K.; Nath, N.; Chilkoti, A.; Zauscher, S. *J. Am. Chem. Soc*. **2004**, 126, 7330.

(12) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem.* **2002**, *3*, 235.

(13) Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. *FEBS. Lett.* **2000**, *467*, 37.

(14) Photocrosslinking was done in a Rayonet photoreactor equipped with 254-nm lamps for 5 minutes at room temperature.

(15) Borden, W. T.; Gritsan, N. P.; Hadad, C. M.; Karney, W. L.; Kemnitz, C. R; Platz, M. S. *Acc. Chem. Res.* **2000**, 33, 765.

(16) SNR= (signal-background)/ (standard deviation of background).

(17) The nonspecific adsorption signal was caused by rapid drying of microspots during printing. We didn't observe signals from GST spots if we used a manual pipette to spot 1 μL protein solutions.

CHAPTER III

Generation of Surface-Bound Multicomponent Gradients

Abstract

Cell migration and differentiation are regulated by soluble or matrix-bound gradients in biological systems. Usually, multiple signals are involved in these processes. Herein, we have developed a new method for the generation of protein films with precise control over the spatial density of several different proteins. First, an artificial polypeptide scaffold composed of separate protein capture and surface anchor domains was designed and synthesized. These domains contain a heterodimeric leucine zipper system and an elastin mimetic motif. By using a mutant *E. coli* phenylalanyl-tRNA synthetase, the photoreactive amino acid *para*-azidophenylalanine was incorporated. Glass slides were then functionalized with this polypeptide by spin coating and photocrosslinking. The resulting protein-coated slides were subsequently secured underneath microfluidic chips specifically designed for on-chip mixing using laminar flow. Gradients of leucine zipper tagged proteins were formed in the microchannels and immobilized on the engineered protein films through association of the coiled-coil heterodimer. The dissociation kinetics of the immobilized proteins was also investigated. Lastly, the adhesion of human umbilical vein endothelial cells cultured on the surface-bound gradients of cell binding ligands generated by this technique were examined.

3.1 Introduction

In biological systems, gradients of soluble or matrix-bound stimuli are crucial to many important cellular processes. For instance, under the guidance of specific signaling protein gradients, stems cells move to the right positions to form organs,[1] axons extend their growth cones over long distances to establish the correct synaptic connections,[2] and endothelial cells migrate into surrounding tissues to initiate angiogenesis.[3] In addition to directing cell migration, gradients play important roles in embryonic development, where the spatial and temporal presentation of morphogens specifies cell fate and results in the formation of patterned tissues.[4] Therefore, it is of general interest to engineer microenvironments with a spatially controlled presentation of extracellular cues either for studying gradient-sensing mechanisms *in vitro*, or for developing instructive materials that elicit desired cellular responses.

One of the most widely used techniques to study chemotaxis is the Boyden chamber assay.[5] Chemotatic gradients are generated by placing different concentrations of chemo-attractants in the upper and lower compartments separated by a porous filter through which cells can migrate. This assay does not allow direct visualization of cell locomotion, and gradient shapes are not well defined.[6] To produce gradients on the biological length scale with single-cell resolution (10-100 μm), advanced techniques such as microfabrication are needed. Recently, Jeon et al. proposed the design of a microfluidic gradient generator that harnessed laminar flow of fluids in poly(dimethylsiloxane) (PDMS) channels to create concentration

gradients of molecules in solution.[7] From a materials design perspective, it is more practical to generate surface-bound rather than solution-phase gradients because smaller volumes of reagents are consumed during the experiments, and the modified materials may be readily used for tissue engineering applications. Currently, there are two major ways to immobilize protein gradients on material surfaces. Nonspecific adsorption is the simplest way to produce such gradients,[8,9] but the resulting structures are quite unstable due to fast surface dissociation kinetics[10] and cell remodeling[11]. Alternatively, immobilization via covalent crosslinking through lysine and cysteine residues provides increased stability of surface-bound gradients.[10] Unfortunately, both methods result in the random orientation and altered conformation of proteins, and the biological activities of proteins are greatly reduced. Moreover, these methods are not generally applicable to generating multi-ligand gradients that usually encountered *in vivo* by cells since some proteins (such as small peptides) have limited surface-exposed reactive residues for conjugation or are too hydrophilic to adsorb stably onto hydrophobic substrates.

These difficulties can be overcome by engineering site specific attachment of ligands through expression of recombinant fusion proteins bearing affinity tags. Previously, we designed an artificial polypeptide scaffold for protein immobilization onto solid substrates.[12] By combining this surface functionalization method with microfluidics technology, we developed a new approach to generate surface-bound multicomponent protein gradients.

3.2 Materials and methods

3.2.1 Cloning, expression and purification of proteins

The genes encoding the elastin mimetic domain ELF and the heterodimeric leucine zipper pair, ZE and ZR, have been previously constructed.[12] Since ZRELF is more soluble than $ZE(gs)_6ELF$, we use this newly designed protein for surface functionalization. ZRELF containing the photoreactive unnatural amino acid, *para*-azidophenylalanine was biosynthesized using the phenylalanine auxotroph *E. coli* strain AF-IQ.[13] Two proteins, a mutant tenth fibronectin type III domain[14] (FN) with specificity to $\alpha_v\beta_3$ and a *de novo* designed VEGF mimetic peptide[15] (QK), were selected as the model ligands for gradient generation and their genes were obtained by assembling synthetic oligonucleotides through PCR. Recombinant proteins FNZE, QKZE and ZE were expressed in *E. coli* strain BL-21 at 22 ^oC. All proteins were purified on Ni-NTA columns and their purity was assessed by SDS-PAGE and MALDI-TOF mass spectrometry.

3.2.2 Surface functionalization

Standard glass slides (Corning) were immersed into concentrated H_2SO_4 for one hour. After washing thoroughly with water, they were immersed into a boiling solution of 1/1/5 (v/v) H_2O_2 (30%) /NH_4OH (30%)/H_2O for 30 minutes. Then the slides were gently shaken in 1% octyltrichlorosilane in toluene for three hours. Finally they were cleaned twice in methanol and twice in DI water. The functionalized slides

were cured at 110^{o}C for 30 minutes.

A solution of ZRELF (50 µl, 2.5 mg/ml) in 50% n-propanol was applied to a small part of a glass slide and spun at 1,500 rpm for 45 seconds. The resulting protein film was irradiated with Hg-arc lamps (I-line and H-line, 4.5 mW cm^{-2}) in a Karl Suss mask aligner for 2 minutes. Exposed protein-coated slides were thoroughly washed with 50% isopropanol and ddH_2O to remove uncrosslinked protein.

3.2.3 Fabrication of microfluidic chips

The microfluidic gradient generator was fabricated using rapid prototyping and soft lithography as originally described by Jeon et al.[7] Briefly, a high-resolution printer was used to generate a mask with a minimum feature size of 30 µm from a CAD file (CAD/Art Services, Poway, CA). A SU-8 2100 photoresist (Microchem, Newton, MA) layer was spin-coated onto a silicon wafer at 3500 rpm for 30 seconds and exposed to ultraviolet light for 150 seconds through the mask with the pre-printed pattern in a Karl Suss mask aligner. The wafer was then immersed into SU-8 developer, and the unexposed photoresist was dissolved into solution, leaving behind a master mold composed of 100-µm high crosslinked photoresist structures. Poly(dimethylsiloxane) (PDMS) chips were formed by curing prepolymer solution (Sylgard 184, Dow Corning) on silicon masters. Inlet and outlet ports were punched out of the PDMS using a sharpened needle. Polyethylene tubing was inserted into these holes to enable fluid flow in and out of the microchannels.

3.2.4 Quantitative measurement of surface density

Lyophilized samples of FNZE or QKZE (1 mg each) were dissolved in 0.5 ml of $NaHCO_3$ buffer (100 mM, pH~9, adding SDS until dissolution) and reacted with 0.5 mg of Cy3 (Amersham) or Alexa 647 (Invitrogen) NHS ester for 5 hours at room temperature. Dialysis was used to remove unconjugated dye molecules. Then the dialyzed samples were subjected to gel electrophoresis, and the ratio of conjugated to unconjugated dye molecules was determined by fluorescence imaging. Protein solutions (2 μl each) at different concentrations (0.2, 0.1 and 0.05μM of FNZE; 0.4, 0.2 and 0.1μM of QKZE) were spotted onto ZRELF-coated glass slides and dried in air over night. The fluorescence intensity and the area of each spot were measured by a GenePix 4200A chip reader. The surface density at each spot was calculated and plotted against fluorescence intensity. The final curves were linearly fitted with intercepts set to zero (the scanning setting was adjusted to minimize the background fluorescence to zero arbitrary unit).

3.2.5 Determination of dissociation kinetics

FNZE and QKZE solutions (0.5 μM) were spotted onto ZRELF-coated glass slides and incubated for one hour. The slides were then sonicated in an excess of PBS for half an hour. Keeping the samples hydrated, the slides were transferred to glass beakers containing 50 ml PBS or Dulbecco's modified Eagle's cell culture medium (CO_2-independent pH equilibration, Invitrogen). The beakers were placed in a 37 ^oC

incubator, and fluorescence intensities of immobilized proteins were measured over time using a Carl Zeiss microscope.

3.2.6 Cell culture

Human umbilical vein endothelial cells (HUVECs, Clonetics) were maintained in a 37 °C, 5% CO_2 humidified environmental chamber. The cells were grown in endothelial cell basal medium (EBM-2, Clonetics) supplemented with the supplied Bulletkit, which was replaced every 2 days. Near confluent HUVEC cultures were passaged nonenzymatically by treatment with 0.61 mM EDTA (Gibco). Passages 3–6 were used.

3.2.7 Generation of immobilized protein gradients

The PDMS chip and glass substrate (ZRELF-coated region was covered with a protection box[9]) were activated with oxygen plasma (200 millitorr, 35 sec, 80 W, Anatech) and coupled immediately to form an irreversible seal in order to prevent leakage when injecting fluids. To remove trapped bubbles and block nonspecific protein adsorption, microfluidic channels were flushed with 2% BSA solution for half an hour using a PHD 2000 syringe pump (Harvard Apparatus). Afterwards, relevant protein mixtures diluted in 2% BSA solutions were individually injected into the two inlets at a rate of 0.5 μl/min for one hour to generate immobilized protein gradients. The whole assembly was then soaked in water and the PDMS portion above the

gradient region was peeled away from the remaining chip using a razor blade. This resulted in the formation of a PDMS well on the glass slide. The sample was sonicated in PBS for half an hour and blown dry, and the gradient profile was scanned by a GenePix 4200A chip reader. For substrates used for cell studies, PBS was left in the PDMS well to maintain the hydrated gradient surfaces.

3.2.8 Cell studies

The gradient region containing the PDMS well was cut out from the glass slide using a diamond pen and placed into a 6-well tissue culture plate. The whole set was sterilized upon ultraviolet light exposure for five minutes in a laminar flow hood. Passaged HUVECs resuspended in 5 ml serum-free EBM-2 containing 2% BSA were added to the samples at a density of 10,000 cells /cm^2. After 2 hours, the plates were removed from the incubation chamber, gently washed twice with EBM-2 containing 2% BSA, and imaged using a 10× phase contrast objective on a Nikon Eclipse TE 300 inverted microscope. Images were captured on a Sony CCD color video camera (model DXC-151A) equipped with Metamorph software. Fifteen images randomly taken from three gradient substrates were used to quantify the cell attachment.

3.3 Results and discussion

3.3.1 Design and biosynthesis of relevant proteins

The ZE/ZR heterodimer leucine zipper pair was derived from vitellogenin-binding protein (VBP), a bZIP homodimer.[16,17] All residues at the **e** and **g**

positions of the first four heptads were changed to glutamic acid or arginine, respectively, to produce acidic peptide ZE and basic peptide ZR.[12] In our hands, ZE can be stably expressed in *E. coli*, while ZR cannot be produced unless it is fused to a well-folded protein or co-translated with ZE. Therefore, ZE was chosen as a universal affinity tag to make recombinant FNZE and QKZE fusion proteins. Meanwhile, ZR was fused to elastin mimetic domain ELF to make ZRELF peptide for surface functionalization. ZRELF was co-expressed with ZE using a dicistronic construct as described previously.[12] After purification, dialysis and lyophilization, the protein yields of ZRELF, QKZE, FNZE and ZE were 75 mg, 80 mg, 150 mg and 30 mg per liter of culture, respectively. Sample purity was analyzed by SDS-PAGE (Figure 3.1). According to amino acid analysis, the incorporation rate of *p*-azidophenylalanine in ZRELF is 40%.

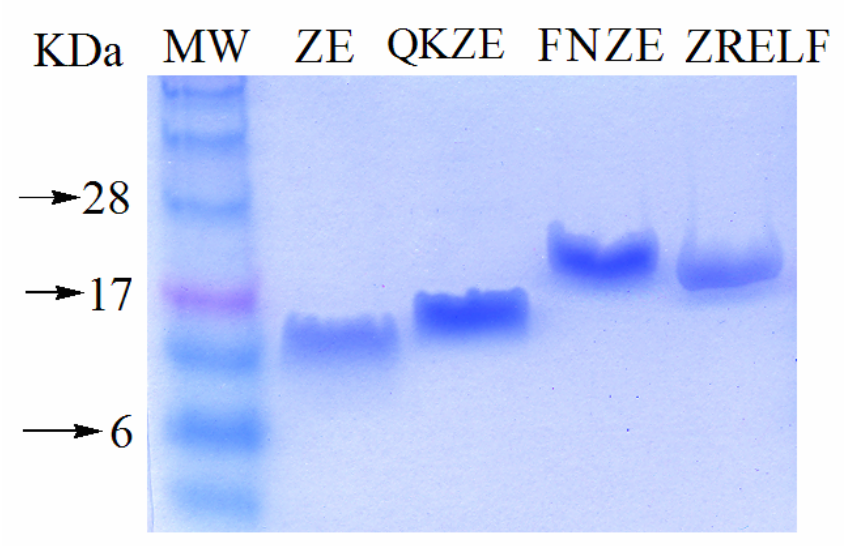
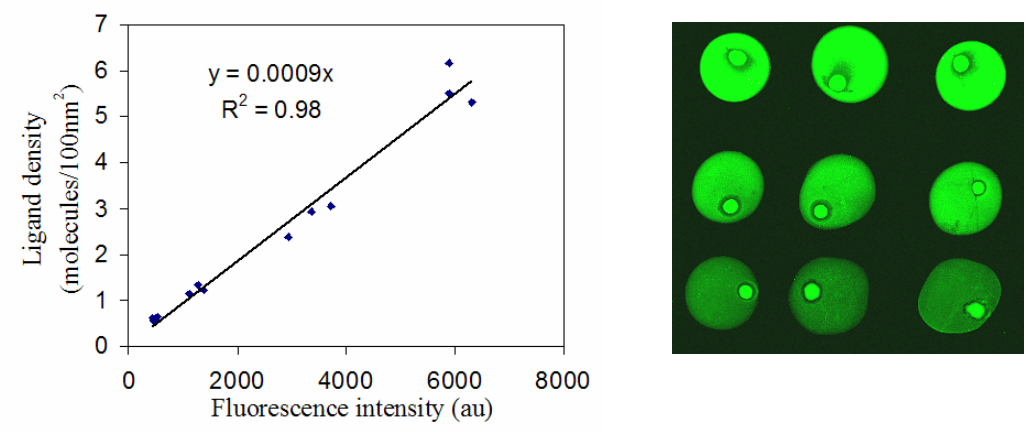


Figure 3.1 SDS-PAGE of purified proteins.

3.3.2 Generation and characterization of immobilized protein gradients

To quantify the surface density of proteins, we labeled FNZE and QKZE with Cy3 and Alexa 647 fluorophores, respectively. Aliquots at different protein concentrations were spotted onto the surface of ZRELF-coated glass slides and dried (right panel of Figure 3.2). After measuring the fluorescence intensity and area of

(a)



(b)

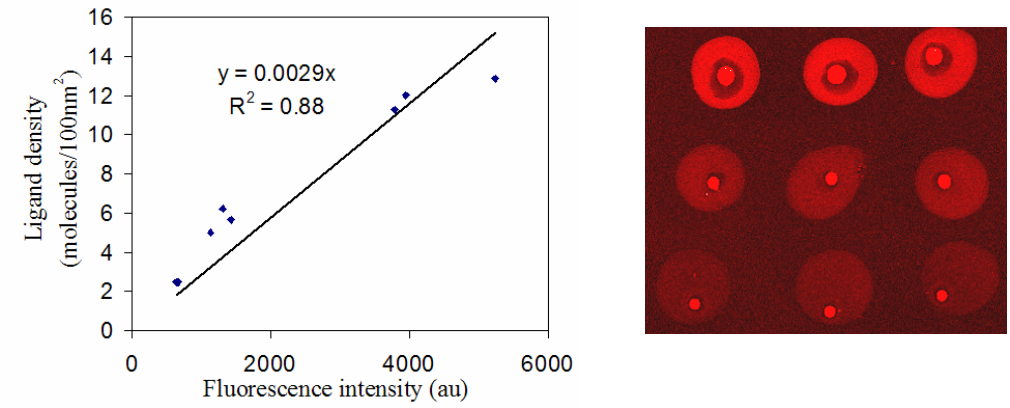
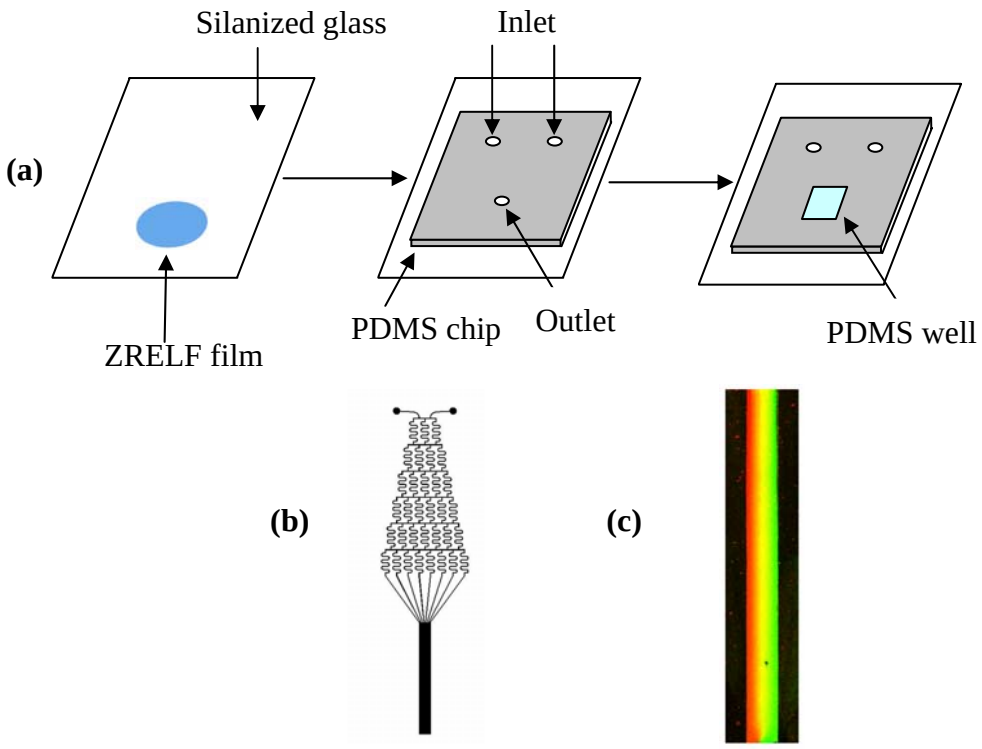


Figure 3.2 Calibration curve for quantification of ligand density. (a) Cy3 conjugated FNZE, (b) Alexa 647 conjugated QKZE. Protein solutions (2 μl) were spotted and dried. The fluorescence intensity and the area of each spot were measured by a GenePix 4200A chip reader. The final curves were linearly fitted.

these spots, standard curves correlating surface density (molecules/100nm^2) and pixel intensity (au) were constructed (left panel of Figure 3.2). Linear relationships were obtained when the protein solutions were spotted on ZRELF films since strong coiled-coil association allowed homogenous distribution of protein molecules during drying. If blank glass slides were used, most of the protein molecules were concentrated in the center of the drying spots and the linear relationship was lost. Calibration curves from linear regression were used to estimate the surface concentration of immobilized protein gradients.

As illustrated in Scheme 3.1a, gradients of surface-bound protein ligands were obtained by combining surface functionalization and microfluidic delivery. Glass slides pretreated with OTS were spin-coated with a solution of ZRELF dissolved in 50% propanol (the addition of organic solvent helped wet the hydrophobic surface). The protein films were covalently crosslinked to the substrates through photodecomposition of the arylazides upon UV irradiation. A microfluidic gradient generator (Scheme 3.1b) fabricated from PDMS was sealed to the functionalized glass slide after oxygen plasma activation. Different solutions containing ZE tagged proteins were continuously injected into the microfluidic network using a syringe pump. The fluid streams introduced through the two inlets were combined and mixed in the serpentine channels and eventually established a concentration gradient across the output channel. There the protein gradients from solution were immobilized on the ZRELF film coated underneath the microchannel

via coiled-coil association between ZE and ZR. After generation of surface-bound gradients, the device was soaked in water and the PDMS portion above the gradient region was gently removed. The resulting well kept the proteins in solution (drying will denature proteins and remove bound proteins from surface) and facilitated studies of cells to be cultured on this surface. Scheme 3.1c shows a fluorescence image representing the immobilized counter-gradient of QKZE and FNZE across the entire output channels (800 μm wide and 1 mm long).



Scheme 3.1 The assembly of microfludic device and gradient generation. (a) Steps to assemble the setup to generate gradients. (b) Schematic design of the gradient generator. (c) Fluorescent image of the counter-gradient of QKZE and FNZE across the output channel.

FN (94 residues) and QK (15 residues) were chosen as model proteins and C-terminally fused with the affinity tag ZE to generate FNZE and QKZE recombinant proteins. The ZE peptide was also required as an "inert" density controller to adjust the available binding sites. To form smooth linear gradients, a flow rate of 0.5 μl/min and a binding period of 1 hour were empirically determined. First, we generated a surface-bound gradient of FNZE by injecting ZE and FNZE (both at a concentration of 100 nM) through the two inlets. In the microfluidic network, solution with increasing concentrations of FNZE was accompanied by decreasing concentrations of ZE. Both proteins competed for the same binding sites on the surface, thus the gradient in solution was translated into the immobilized gradient. Based on the calibration curve (Figure 3.2) and fluorescence image of the gradient (left panel of Figure 3.3a), the corresponding density profile of FNZE was obtained (right panel of Figure 3.3a). For these substrates, the protein density varied from zero to 4.5 molecules/100nm^2 across the 800 μm channel with a slope of 56.25 molecules/μm^3.

To investigate the possibility of generating gradients containing multiple species, QKZE was mixed with FNZE and both were introduced through the same inlet. A solution of ZE was introduced through the other inlet. All protein solutions were at the same concentration (100 nM). As shown in Figure 3.3b, an overlapping gradient of FNZE and QKZE was created. The density of FNZE was increased from zero to 1.8 molecules/100nm^2 across the channel, while that of QKZE was increased from zero to 6.4 molecules/100nm^2. This result demonstrated that even though FNZE

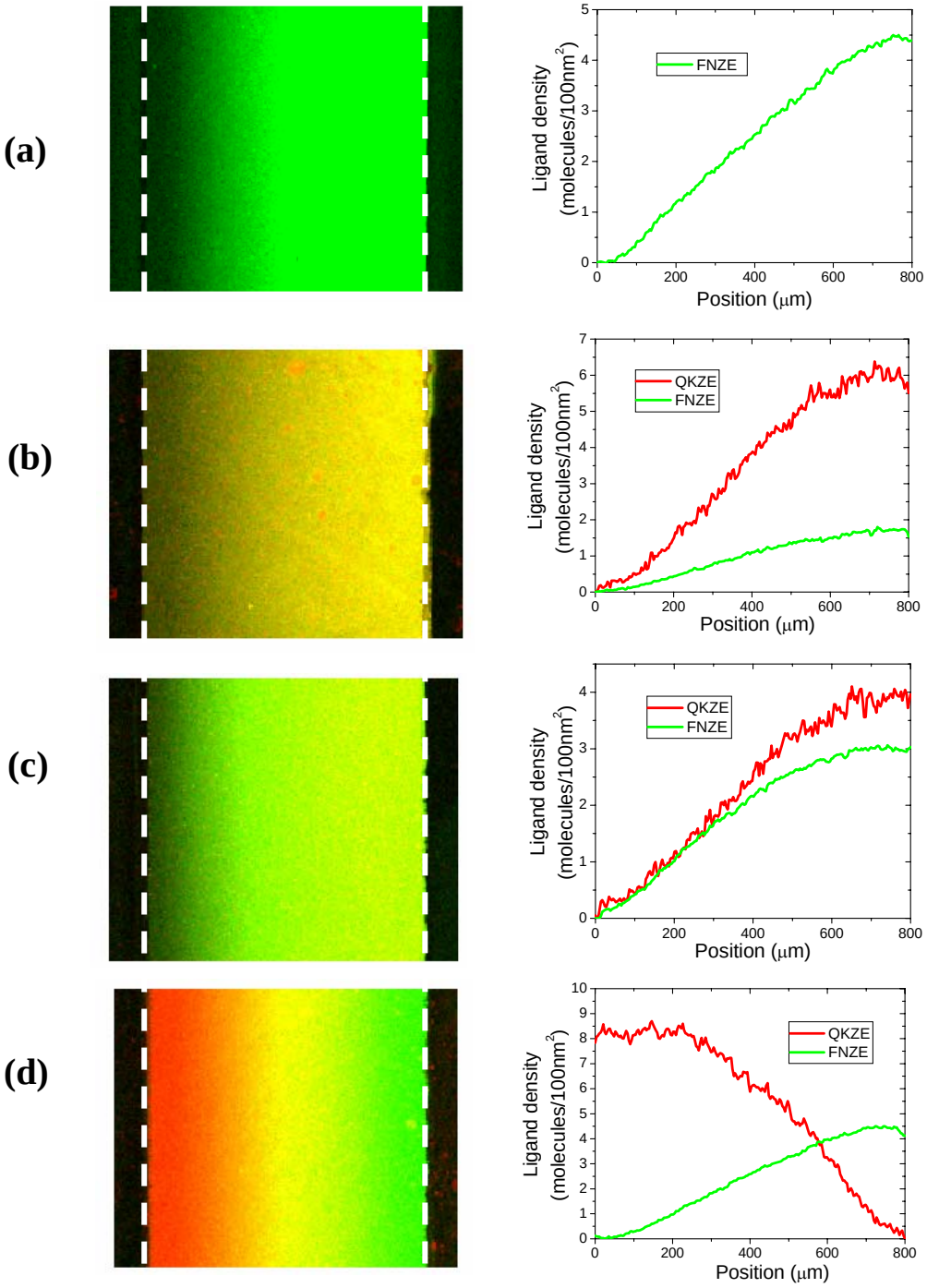


Figure 3.3 Fluorescence images of immobilized gradients and their corresponding density profiles. Cy3 (FNZE) was pseudocolored green, Alexa 647(QKZE) red and overlay yellow. (a) ZE(100 nM) ↔ FNZE(100 nM); (b) ZE(100 nM) ↔ FNZE(100 nM), QKZE(100 nM); (c) ZE(100 nM) ↔ FNZE(100 nM), QKZE(50 nM); (d) QKZE(100 nM) ↔ FNZE(100 nM).

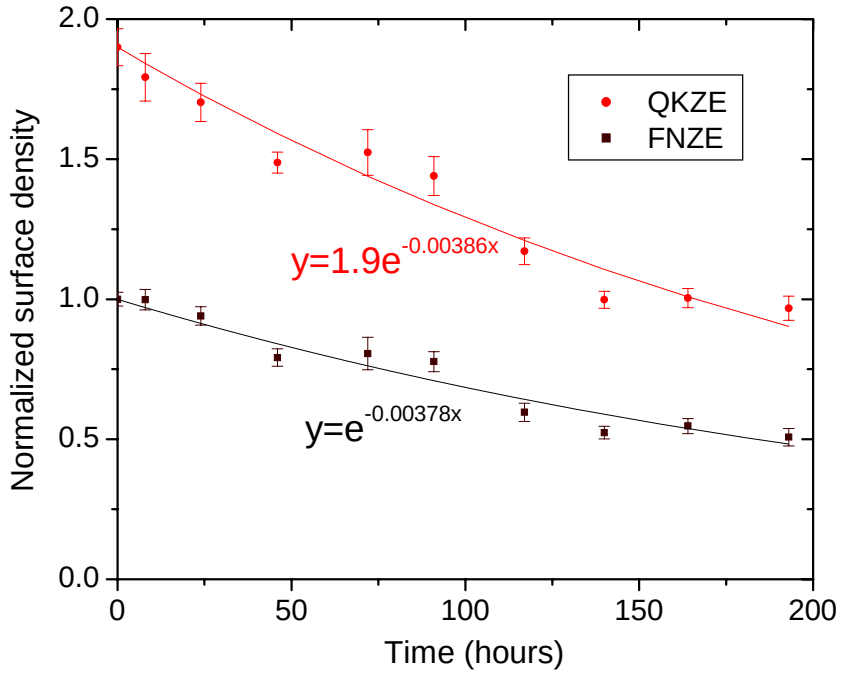
and QKZE were at the same concentration in the solution, the resulting surface densities were significantly different. A possible explanation for this result may be that since the size of QKZE is smaller than that of FNZE, QKZE has faster binding kinetics and occupies more sites per unit area because of less steric hindrance. The densities of each protein can be controlled by changing the ratio of FNZE to QKZE in the injection solution. For example, when the concentration of FNZE was maintained at 100 nM and the concentration of QKZE was reduced to 50 nM, a gradient with different profile was produced (Figure 3.3c). In this case, the density profile of FNZE spanned from zero to 3 molecules/100nm^2 compared to that of QKZE, which varied from zero to 4 molecules/100nm^2. We also generated immobilized counter-gradients as demonstrated in Figure 3.3d by inputting QKZE and FNZE from different inlets. The density of FNZE gradually decreased from 4.5 molecules/100nm^2 to zero while the density of QKZE increased from zero to 8.6 molecules/100nm^2.

3.3.3 Dissociation kinetics of immobilized proteins

For cell studies and tissue engineering, it is important to maintain the long-term stability of the surface characteristics of engineered materials. We soaked the protein-bound surfaces in excess buffer and monitored the dissociation kinetics through repeated fluorescence imaging. The desorption curves are shown in Figure 3.4. The density at time zero corresponds to the amount of proteins remaining on the surface after sonication in PBS buffer for half an hour, which is a critical step to

remove weakly bound molecules. When soaked in PBS, following the first-order kinetics law, the proteins gradually diffused away from the surface (Figure 3.4a). Both QKZE and FNZE had a very slow release rate. The dissociation rate constant k_{off} was around $1.1x10^{-6}$ s^{-1}, approximately two-fold lower than that observed for biotin/streptavidin complex ($2.4x10^{-6}$ s^{-1}).[18] This meant that half of the proteins still remained attached after 175 hours. Interestingly, QKZE and FNZE have very similar dissociation kinetics despite difference in both their molecular weights and three dimensional structures. The phenomenon indicates that after immobilization, surface dissociation mainly depends on the coiled-coil interaction between ZE and ZR, irrespective of their fusion proteins. However, these proteins dissociated faster in cell culture medium (Figure 3.4b). Dissociation of FNZE could still be described by the first-order rate law, and the rate constant was two-fold higher than that in PBS. QKZE showed an initial burst release in the first 24 hours. Beyond 24 hours, the dissociation of QKZE followed a slow first-order process comparable to that of FNZE. The faster dissociation in cell culture medium may relate to the higher ionic strength, which shields the electrostatic interactions between arginine and glutamic acid residues.[16]

(a)



(b)

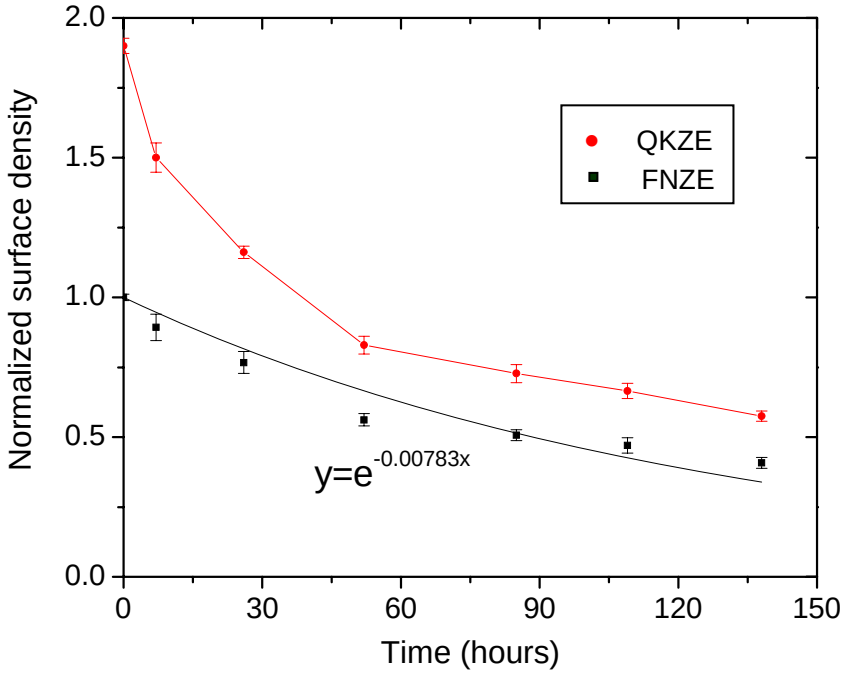


Figure 3.4 Dissociation curves of immobilized proteins monitored by changes in surface fluorescence. (a) Slides soaked in PBS. (b) Slides soaked in CO_2-independent Dulbecco's modified Eagle's cell culture medium.

3.3.4 Cellular response to surface-bound gradients

Finally, we subjected our substrates to cell studies. On a control substrate with only ZE peptide, very few HUVECs adhered, and these cells did not spread and showed a round morphology (Figure 3.5a). In contrast, when FNZE was immobilized on the surface, significant cell adhesion was observed (Figure 3.5b), demonstrating that cells recognized the adhesion ligand within the recombinant protein FNZE. The attached cells distributed homogeneously across the entire substrate, and most of them exhibited well-spread morphologies.

Alternatively, a gradient shown in Figure 3.3a had an increasing density of FNZE as well as a decreasing density of ZE across the substrate. To this surface HUVECs were seeded (Figure 3.5c), and as expected, the number of adherent cells varied spatially across the gradient. It was discovered that more cells attached to the region with a higher density of cell adhesion ligands. To quantify the cell attachment, each image (800 μm wide by 600 μm long) was split into three regions, and the fraction of adherent cells was determined through dividing the number of cells per region by the total number of cells attached (Figure 3.5d). On the left third of the gradient, the fraction of attachment was 0.13±0.05; the fraction increased to 0.36±0.06 on the middle third of the gradient; and on the right third, the fraction reached the highest value, 0.51±0.06.

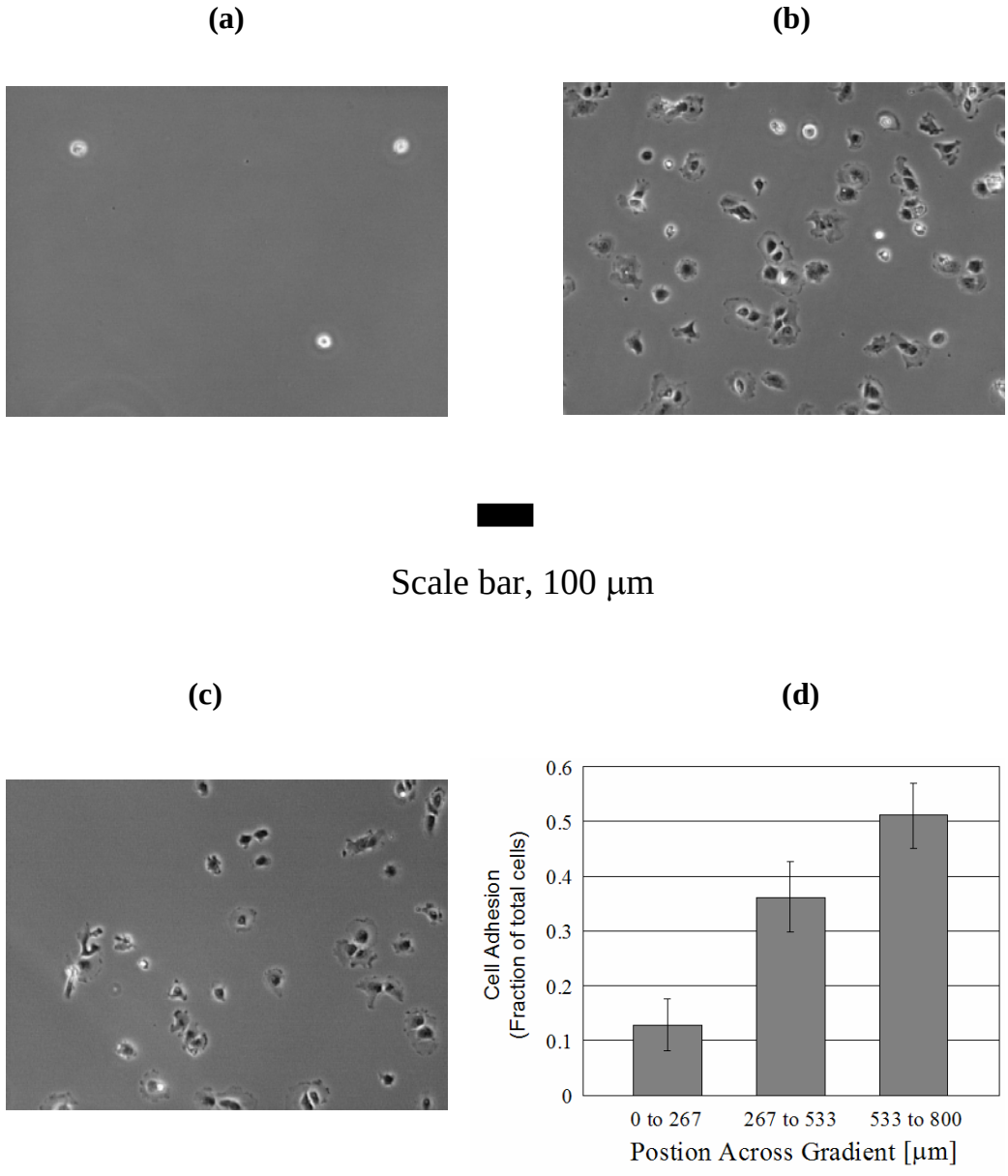


Figure 3.5 The attachment of HUVECs. Phase contrast images of HUVECs attached to the surface of (a) ZE, (b) FNZE, (c) ZE◂▸FNZE gradient. (d) Quantification of HUVECs adhesion.

3.4 Conclusion and future work

We have demonstrated that the integration of microfluidics, protein engineering and surface functionalization provides a simple and flexible approach to generating surface-bound multicomponent gradients. We have also showed that gradient of cell adhesion ligands generated by this technique modulates the attachment and spreading of endothelial cells. In principle, any combination of proteins or peptides can be introduced into this system and their cooperative effects on cellular response could be investigated. Moreover, the relative ratio of different proteins or steepness of gradients can be easily and precisely controlled. Since it is relatively straightforward to engineer leucine zippers with various affinities by changing the number of heptad repeats[17], the surface stability can be tuned for specific proteins, which may be harnessed to construct dynamic surfaces that allow spatiotemporal control of protein ligands. We can also incorporate this leucine zipper system into hydrogel-forming protein polymers[19] and the generation of gradients in such three-dimensional scaffolds may be used to guide cell behaviors such as neuron regeneration or angiogenesis.

3.5 References

(1) Christiansen, J. H.; Goles, E. G.; Wilkinson, D. G. *Curr. Opin. Cell Biol.* **2000**, *12*, 719.

(2) Dickson, B. J. *Science* **2002**, *298*, 1959.

(3) Carmeliet, P. M. J. *Nat. Med.* **2000**, *6*, 389.

(4) Ashe, H. L.; Briscoe, J. *Development* **2006**, *133*, 385.

(5) Boyden, S. *J. Exp. Med.* **1962**, *115*, 453.

(6) Entschladen, F.; Drell, T. L.; Lang, K.; Masur, K.; Palm, D.; Bastian, P.; Niggemann, B.; Zaenker, K. S. *Exp. Cell Res.* **2005**, *307*, 418.

(7) Jeon, N. L.; Dertinger, S. K. W.; Chiu, D. T.; Choi, I. S.; Stroock, A. D.; Whitesides, G. M. *Langmuir* **2000**, *16*, 8311.

(8) Dertinger, S. K.; Jiang, X.; Li, Z.; Murthy, V. N.; Whitesides, G. M. *Proc. Natl. Acad. Sci. U S A.* **2002**, *99*, 12542.

(9) Jiang, X. Y.; Xu, Q. B.; Dertinger, S. K. W.; Stroock, A. D.; Fu, T. M.; Whitesides, G. M. *Anal. Chem.* **2005**, *77*, 2338.

(10) Gunawan, R. C.; Choban, E. R.; Conour, J. E.; Silvestre, J.; Schook, L. B.; Gaskins, H. R.; Leckband, D. E.; Kenis, P. J. *Langmuir.* **2005**, *21*, 3061.

(11) Katz, B. Z.; Zamir, E.; Bershadsky, A.; Kam, Z.; Yamada, K. M.; Geiger, B. *Mol. Biol. Cell* **2000**, *11*, 1047.

(12) Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 10136.

(13) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235.

(14) Richards, J.; Miller, M.; Abend, J.; Koide, A.; Koide, S.; Dewhurst, S. *J. Mol. Biol.* **2003**, *326*, 1475.

(15) D'Andrea, L. D.; Iaccarino, G.; Fattorusso, R.; Sorriento, D.; Carannante, C.;

Capasso, D.; Trimarco, B.; Pedone, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14215.

(16) Krylov, D.; Barchi, J.; Vinson, C. *J. Mol. Biol.* **1998**, *279*, 959.

(17) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, *10*, 649.

(18) Piran, U.; Riordan, W. J. *J. Immunol. Methods.* **1990**, *133*, 141.

(19) Shen, W.; Zhang, K.; Kornfield, J. A.; Tirrell, D. A. *Nat. Mater.* **2006**, *5*, 153.

CHAPTER IV

Progress towards the Creation of Protein Walkers

Abstract

The extracellular matrix serves many important functions such as maintaining the structural organization of tissues, providing support and anchorage for cells, and regulating cellular physiology. Cell-ECM interactions are mediated through specific cell surface receptors known as integrins. Clustering of integrins into focal adhesions plays a central role in determining cell behaviors. Our goal is to generate protein walkers that have different lateral mobility rates on a surface, which, if presenting cell adhesion ligands, may influence the clustering of integrin receptors. Herein, an artificial polypeptide scaffold was developed for protein immobilization onto solid substrates via heterodimeric coiled-coil association. Several variants of a leucine zipper pair that have tunable heterodimerization affinities were designed and synthesized to allow diversity in the association strength of proteins linked to a surface. Then four proteins containing either two or three of these binding motifs were created. However, FRAP experiments did not provide evidence of controlled lateral mobility of these proteins on a surface modified with the designed polypeptide scaffold.

4.1 Introduction

Most mammalian cells in tissues are surrounded by an insoluble network of fibrous proteins and polysaccharides that comprise the extracellular matrix (ECM).[1] The ECM serves many important functions such as maintaining the structural organization of tissues, providing support and anchorage for cells, and regulating cellular physiology.[2,3] The ability of cells to proliferate, differentiate, and migrate largely depends on their capacity to bind to certain components of the ECM. Cell-ECM interactions are mediated through specific cell surface receptors known as integrins. An integrin is a transmembrane heterodimeric complex composed of noncovalently associated α and β subunits. Upon binding to recognition sites (such as the Arg-Gly-Asp domain) embedded in ECM proteins, integrins can cluster and rearrange cytoskeletal structures, leading to the assembly of focal adhesions and actin stress fibers.[4] Focal adhesions serve as communication bridges, activating numerous signal transduction pathways and mediating the crosstalk between the actin cytoskeleton and ECM.[5]

Since focal adhesions play a central role in determining cell behaviors, it is interesting to fabricate extracellular-mimetic substrates that can be used to probe and better understand the influence of environmental cues on the formation of integrin clusters. To investigate the effect of ligand affinity on the number and size of focal adhesions, Kato and Mrksich immobilized a low-affinity, linear RGD ligand and a high-affinity, cyclic RGD ligand at identical densities on self-assembled monolayers

of alkanethiolates. 3T3 Swiss fibroblasts deposited on the surface modified with the cyclic RGD ligand developed twice the number of focal adhesions compared to the linear RGD ligand, while the average adhesion size was smaller.[6] In addition to composition, another study showed that the physical state of the ECM could regulate the differential assembly of adhesion sites. More specifically, it was demonstrated that physisorbed fibronectin could be reorganized into fibrils by attached cells. This behavior was mainly associated with $\alpha_5\beta_1$ integrin and led to the formation of elongated fibrillar adhesion structures. In contrast, covalently immobilized fibronectin complexed predominantly with $\alpha_v\beta_3$ integrin and the cells formed classic focal adhesions.[7] These observations have stimulated the design and synthesis of polymer materials with controlled physicochemical surface characteristics for regulating the proliferation and differentiation of endothelial cells into vascular-like structures.[8]

An appealing goal for material scientists is to develop instructive materials with fine-tuned physical properties and biochemical functionalities that could provide precise extracellular cues for regulating cellular behaviors and functions.[9,10] In this context, protein engineering provides a promising approach that allows extraordinary control over material properties and cellular responses.[11] Previously, an artificial polypeptide scaffold for protein immobilization onto solid substrates via heterodimeric coiled-coil association was developed.[12] Since it is straightforward to engineer leucine zippers with various affinities,[13,14] this technique permits diversity in the association strength of proteins linked to a surface. A step further is to generate

protein walkers that have different lateral mobility rates on a surface, which, if presenting cell adhesion ligands, may influence the clustering of integrin receptors.

4.2 Materials and methods

4.2.1 Cloning strategy

Recombinant genes encoding leucine zipper variants (Table 4.1) were created through polymerase chain reaction (PCR), assembling a mixture of multiple overlapping synthetic oligonucleotides designed by DNAWorks software. As indicated in Figure 4.1, proteins containing dimeric or trimeric leucine zipper variants were cloned. For the zipper domain ZE, for example, a previously constructed plasmid (pQE60_FNZE-FN-ZE) was digested with RsrII to obtain the $FNC_{10}ZE$ fragment. This fragment was multimerized by ligating monomers with T4 DNA ligase. Both the monomer and the dimer were ligated into pQE60_ZE to yield pQE60_ZE-FN-ZE and pQE60_ZE-FN-ZE-FN-ZE. All the resulting plasmids were transformed into the BL21 strain of *E. coli* by heat shock techniques.

Table 4.1 **Amino acid sequences of synthesized leucine zipper variants.**

Peptide	*def gabcdef gabcdef gabcdef gabcdef gabcdef gabcd*
ZR	LEI RAAALRR RNTALRT RVAELRQ RVQRLRN EVSQYET RYGPL
RE_3R	LEI RAAALRR ENTALET EVAELEQ EVQRLEN RVSQYRT RYGPL
RE_2RE	LEI RAAALRR ENTALET EVAELEQ RVQRLRN EVSQYET RYGPL
ZE	LEI EAAALEQ ENTALET EVAELEQ EVQRLEN RVSQYRT RYGPL

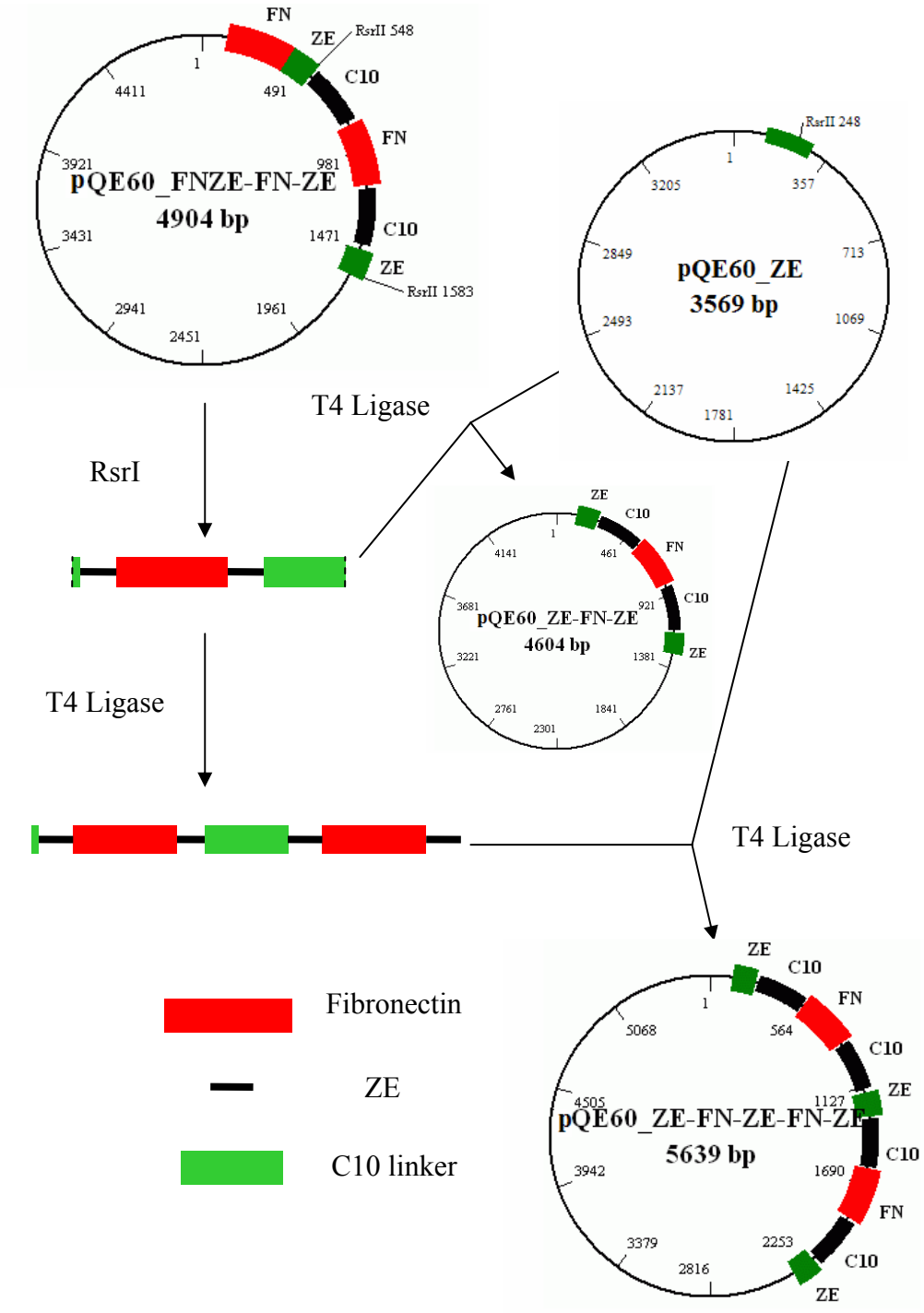


Figure 4.1 Cloning strategy for proteins containing multimeric leucine zipper variants.

4.2.2 Expression and purification of proteins

Transformed *E. coli* cells were grown in one liter 2XYT medium containing ampicillin and chloramphenicol at 37°C to an OD_{600} of 1.0. Overnight expression was induced upon the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). All the proteins were purified using Ni-NTA columns, and sample purity was assessed by SDS-PAGE and MALDI-TOF mass spectrometry.

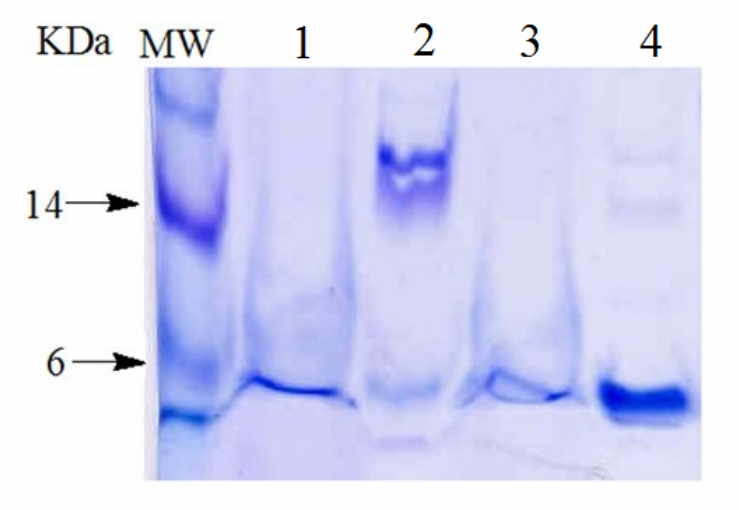


Figure 4.2 SDS-PAGE analysis of purified RE_3R and RE_2RE peptides (both contain His-tag). Proteins were co-expressed with ZRELF (without His-tag) to avoid *in vivo* degradation. Lane 1: RE_3R, Lane 2: a mixture of RE_3R and ZRELF was applied to the Ni-NTA agarose affinity column, washed with urea (pH=7.0) and eluted with urea (pH=4.5), Lane 3: RE_2RE, Lane 4: a mixture of RE_2RE and ZRELF was applied to the Ni-NTA agarose affinity column, washed with urea (pH=7.0) and eluted with urea (pH=4.5). Note that even after urea wash, ZRELF still associated with RE_3R in the column. The relative intensities of ZRELF bands (17 kDa) suggest that RE_3R binds more strongly than RE_2RE with ZR.

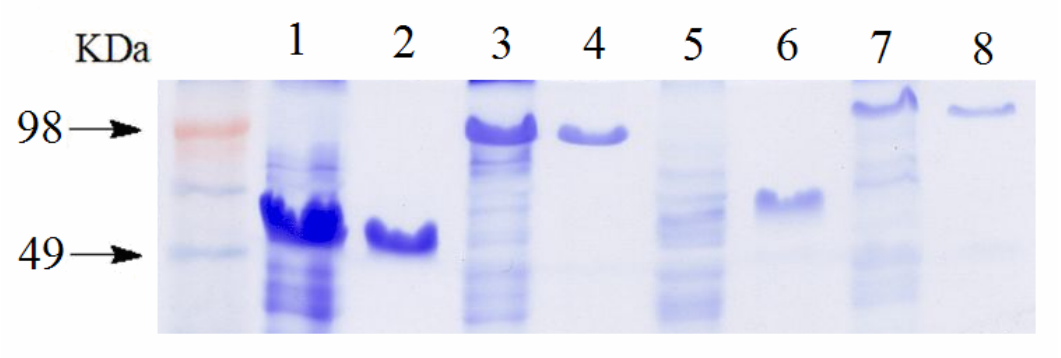


Figure 4.3 SDS-PAGE analysis showing expression and purification of proteins containing multimeric leucine zipper variants. Lane 1:cell lysate (RE_2RE dimer), Lane 2: Purified RE_2RE dimer, Lane 3: cell lysate(RE_2RE trimer), Lane 4: Purified RE_2RE trimer, Lane 5: cell lysate (ZE dimer), Lane 6: Purified ZE dimer, Lane 7: cell lysate(ZE trimer), Lane 8: Purified ZE trimer.

4.2.3 Circular dichroism (CD) spectrometry

CD spectra were obtained using a Model 62A DS Aviv Circular Dichroism Spectrometer (Lakewood, NJ). All samples were dissolved in 12.5 mM potassium phosphate (pH 7.4), 150 mM KCI, 1 mM EDTA; a 300 μL aliquot of these samples was used in the experiments. Homodimeric proteins (ZE, RE_2RE, RE_3R, and ZR) were prepared at a concentration of 20 μM. In addition, 20 μM ZE, RE_2RE, and RE_3R were individually mixed with 20 μM ZR to form heterodimers at a final concentration of 10 μM. Wavelength scans were obtained with the monochromator set to scan from 300 nm to 190 nm. Temperature scans were performed by scanning continuously from 10 °C to 90 °C at 222 nm in 2 °C increments.

4.2.4 Fluorescence recovery after photobleaching (FRAP)

To explore fluorescence recovery after photobleaching, proteins of interest were conjugated to Cy3 NHS ester supplied by Amersham Biosciences (Piscataway, NJ). The dye-labeled proteins were separated from excess dyes by dialysis against water for several days. The fluorescent proteins were immobilized onto a ZR-ELF functionalized surface as described previously and a DPSS 532-75 diode was used to excite the Cy3 fluorophore. Photobleaching was accomplished upon irradiation with 100% laser power for 2 minutes. Subsequent fluorescent recovery was monitored using the Achroplan 20x/0.5 W objectives of the Carl Zeiss LSM 5 Live microscope (Thornwood, NY).

4.3 Results and discussion

4.3.1 Design and characterization of heterodimeric leucine zipper pairs

The ZE/ZR heterodimer leucine zipper pair was derived from the vitellogenin-binding protein (VBP), a bZIP homodimer.[15] All residues at the **e** and **g** positions of the first four heptads were changed to glutamic acid or arginine, respectively, to produce acidic peptide ZE and basic peptide ZR.[16] Thus, both ZE and ZR homodimers have four pairs of repulsive electrostatic interactions, whereas the ZE/ZR heterodimer contains four pairs of attractive salt bridges. To tune the heterodimerization affinities of the leucine zipper pairs, we also synthesized two modified acidic peptides, RE_3R and RE_2RE (Table 4.1), which form a decreasing

number of attractive salt bridges with ZR. The conformations of the synthetic peptides in solution were studied by CD (Figure 4.4). All three heterodimers exhibited an α-helix-like CD spectrum with large molar ellipticity minima values at 222 and 208 nm. The greatest negative ellipticity of the ZE/ZR heterodimer suggests the strongest helix-helix interactions. In addition, the stability of the coiled coils was determined by thermal denaturation experiments, where the changes in ellipticity at 222 nm were monitored (Figure 4.5). The data indicated a melting point of 74°C for RE_2RE/ZR. Since the other complexes were not completely melted at 90°C, their melting points could not be determined. However, the CD spectrum still showed that ZE/ZR pair was more stable than RE_3R/ZR pair, which in turn was more stable than the RE_2RE/ZR pair.

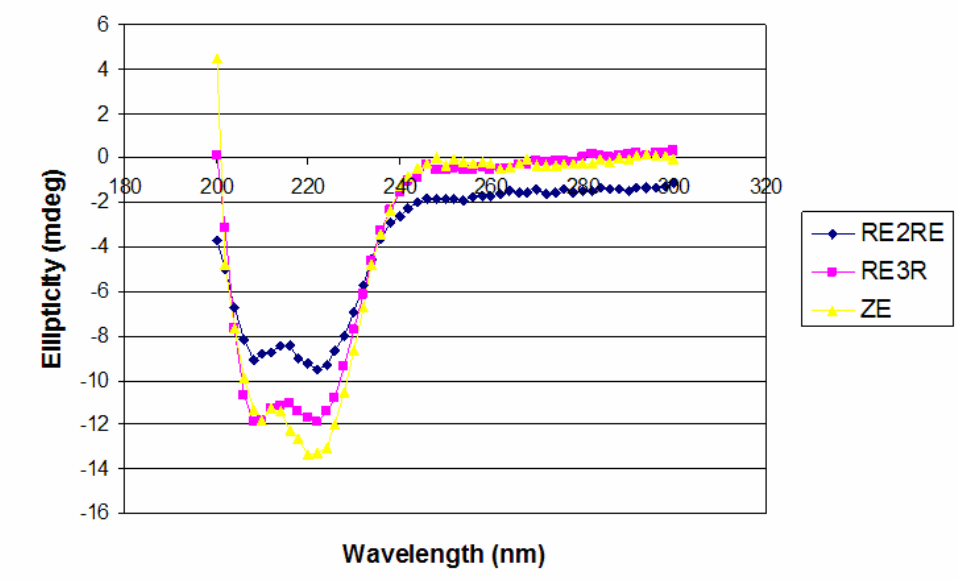


Figure 4.4 CD spectral data of RE_2RE, RE_3R, and ZE complexes to the basic ZR. All three heterodimers exhibited an α-helix-like CD spectrum with large molar ellipticity minima values at 222 and 208 nm. The greater negative ellipticity suggests stronger helix-helix interactions.

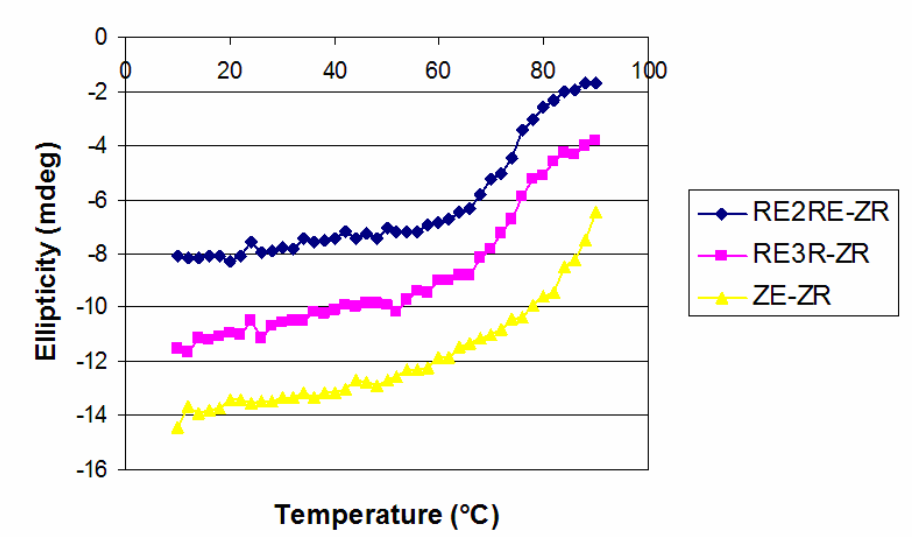


Figure 4.5 Thermal denaturation curves of RE_2RE, RE_3R, and ZE complexes to ZR. Temperature scans were performed by scanning continuously from 10°C to 90°C at 222 nm in 2 °C increments. The data suggested a melting point of 74°C for RE_2RE/ZR. The other complexes were not completely melted at 90°C, so melting points could not be determined. However, the spectrum showed that ZE/ZR was more stable than RE_3R/ZR, which in turn was more stable that RE_2RE/ZR.

4.3.2 Synthesis of protein walkers and measurement of lateral mobility

Utilizing the newly created leucine zipper pair variants, protein walkers were constructed to contain multiple sites for binding to ZRELF functionalized surfaces. Figure 4.6 illustrates a hypothetical model for lateral mobility using artificial protein walkers. In this scheme, one strand of the protein walker dissociates from the surface and then reattaches to the surface at a new binding site. The other strands may remain fixed or participate in strand exchange, allowing the protein walkers to move

laterally across the surface. The attachment of the protein walkers to the surface makes this movement stepwise, preventing diffusion of protein walkers into the solution.

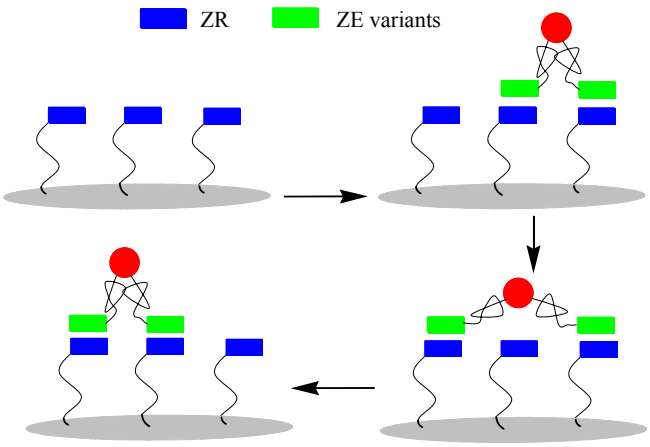
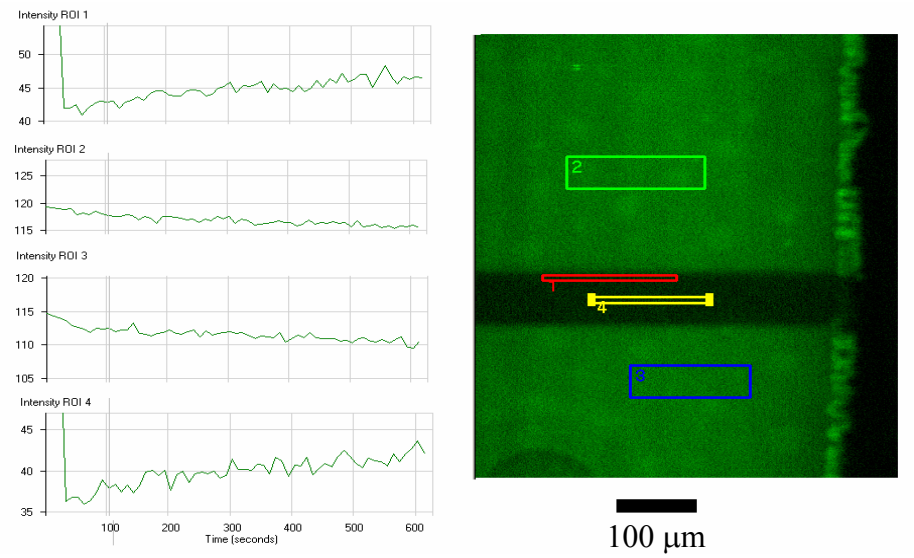


Figure 4.6 A hypothetical model: proteins containing multimeric binding motifs "walk" across the surface through strand exchange.

Based on the hypothetical model, we synthesized four proteins containing either two or three binding sites (either ZE or RE_2RE). These proteins were labeled with the fluorescent dye Cy3. After immobilization of these proteins on ZRELF functionalized glass substrates, the fluorescence intensity of four regions was monitored continuously over a period of ten minutes after photobleaching (Figure 4.7). The non-photobleached regions {ROI (region of interest) 2 & ROI 3} and the center of the photobleached region (ROI 4) were monitored as controls. Both ROI 2 and ROI 3 regions showed a steady decrease in intensity over time, suggesting a gradual dissociation of immobilized proteins from the surface. The ROI 4 region showed a

(a)



(b)

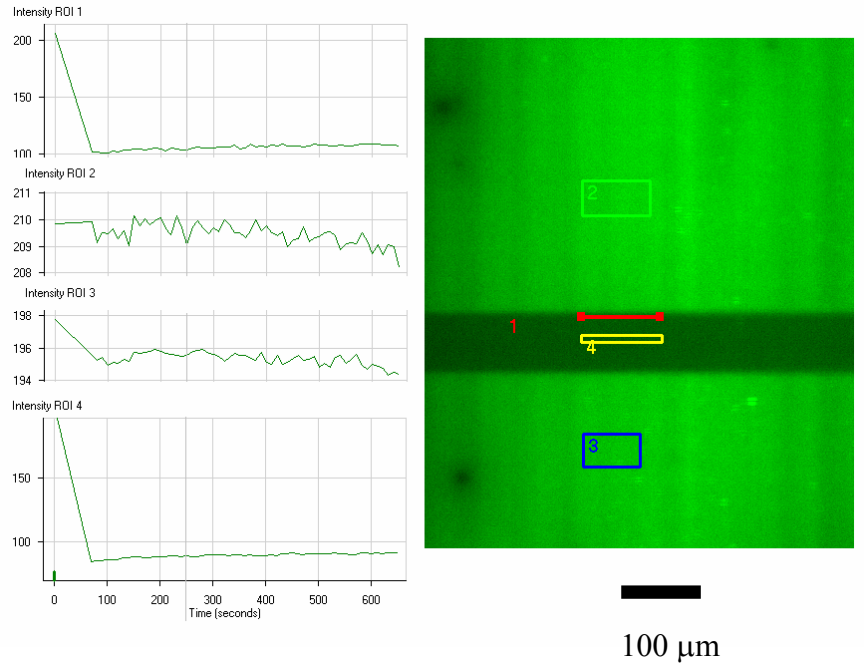


Figure 4.7 FRAP monitoring of protein walkers. (a) RE_2RE-FN-RE_2RE, (b) RE_2RE-FN-RE_2RE-FN-RE_2RE. The non-photobleached region (ROI 2 & ROI 3) and the center of the photobleached region (ROI 4) were monitored as controls. Both ROI 2 and ROI 3 regions showed a steady decrease in intensity over time. The ROI 4 region showed a steady increase in intensity over time. The fluorescence intensity of the interface between the photobleached region and the non-photobleached region (ROI 1) recovered at a similar level, indicating that self-recovery of photobleached dyes was the dominant factor in fluorescence increase rather than cross-diffusion of proteins between the photobleached region and the non-photobleached region.

steady increase in intensity over time, which suggested gradual self-recovery of the photobleached Cy3 fluorophores. However, the fluorescence intensity of the interface between the photobleached region and the non-photobleached region (ROI 1) recovered at a similar level, indicating that self-recovery of photobleached dyes was the dominant factor in the fluorescence increase rather than diffusion of proteins between the photobleached and non-photobleached regions. Thus, FRAP measurements did not provide evidence of lateral mobility of either dimeric or trimeric proteins.

Several possibilities may account for the lack of evidence for lateral mobility. First, in order to achieve efficient movement, leucine zipper pairs should have extremely fast association-dissociation kinetics so that the legs of protein walkers could effectively probe the surface without total detachment of the entire protein from the surface. Second, the expected lateral mobility might occur on a smaller scale than could be observed with the LSM 5 Live microscope (approximately 1 μm resolution). High resolution experimental setups such as single molecule spectroscopy may be required to observe the movement of such protein walkers. Finally, it is possible that lateral mobility was not occurring due to surface saturation. For example, when a leucine zipper became detached, it might reattach to the same ZR site and there was no net movement.

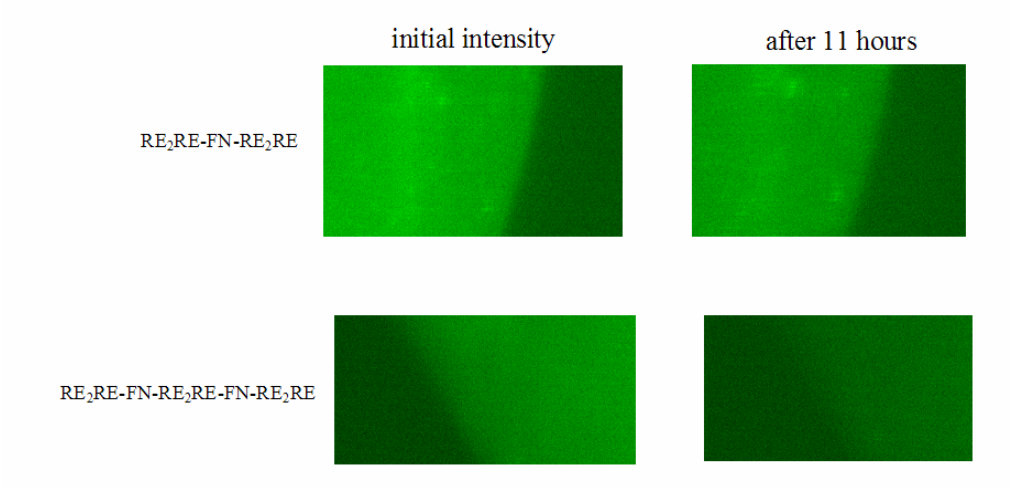


Figure 4.8 Spotting proteins on ZRELF surface and monitoring fluorescence change in PBS buffer after 11 hours.

To address this concern, an additional experiment was performed to examine the potential saturation effect. Both RE_2RE-FN-RE_2RE and RE_2RE-FN-RE_2RE-FN-RE_2RE were spotted onto a ZRELF functionalized surface. The fluorescence of immobilized protein spots was monitored over a period of eleven hours (Figure 4.8). The highly fluorescent regions were saturated with proteins, while dark regions were devoid of surface-bound ligands. Therefore, if proteins could move from the saturated regions to empty regions, the edge should become more fluorescent over time. This result was not observed; instead, there was a sustained dissociation of proteins from the surface into the medium as indicated by the decrease of fluorescence. Thus, the saturation effect does not account for the lack of lateral mobility of the protein walkers.

4.4 Conclusion

We have designed and synthesized several variants of leucine zipper pair that have tunable heterodimerization affinities. Four proteins containing either two or three of these binding motifs were created. However, we did not observe lateral mobility of either dimeric or trimeric proteins.

4.5 References

(1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell. 4th ed.* New York: Garland Science, 2002.

(2) Boudreau, N.; Bissell, M. J. *Curr. Opin. Cell Biol.* **1998**, *10*, 640.

(3) Buckley, C. D.; Rainger, G. E.; Bradfield, P. F.; Nash, G. B.; Simmons, D. L. *Mol. Membr. Biol.* **1998**, *15*, 167.

(4) Burridge, K.; Chrzanowska-Wodnicka, M. *Annu. Rev. Cell Dev. Biol.* **1996**, *12*, 463.

(5) Romer, L. H.; Birukov, K. G.; Garcia, J. G. *Circ. Res.* **2006**, *98*, 606.

(6) Kato, M.; Mrksich, M. *Biochemistry* **2004**, *43*, 2699.

(7) Katz, B. Z.; Zamir, E.; Bershadsky, A.; Kam, Z.; Yamada, K. M.; Geiger, B. *Mol. Biol. Cell* **2000**, *11*, 1047.

(8) Pompe, T.; Markowski, M.; Werner, C. *Tissue Eng.* **2004**, *10*, 841.

(9) Langer, R.; Tirrell, D. A. *Nature* **2004**, *428*, 487.

(10) Lutolf, M. P.; Hubbell, J. A. *Nat. Biotechnol.* **2005**, *23*, 47.

(11) Maskarinec, S. A.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2005**, *16*, 422.

(12) Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 10136.

(13) Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849.

(14) Mason, J. M.; Arndt, K. M. *ChemBioChem.* **2004**, *5*, 170.

(15) Krylov, D.; Barchi, J.; Vinson, C. *J. Mol. Biol.* **1998**, *279*, 959.

(16) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, *10*, 649.

CHAPTER V

Tuning the Erosion Rate of Artificial Protein Hydrogels through Control of Network Topology

Abstract

Erosion behaviour governs the use of physical hydrogels in biomedical applications ranging from controlled release to cell encapsulation. Genetically engineered protein hydrogels offer unique means of controlling the erosion rate by engineering their amino acid sequences and network topology. Here, we show that the erosion rate of such materials can be tuned by harnessing selective molecular recognition, discrete aggregation number and orientational discrimination of coiled-coil protein domains. Hydrogels formed from a triblock artificial protein bearing dissimilar helical coiled-coil end domains (P and A) erode more than one hundredfold slower than hydrogels formed from those bearing the same end domains (either P or A). The reduced erosion rate is a consequence of the fact that looped chains are suppressed because P and A tend not to associate with each other. Thus, the erosion rate can be tuned over several orders of magnitude in artificial protein hydrogels, opening the door to diverse biomedical applications.

The text in this chapter is reprinted with permission from Shen W.; Zhang, K. C.; Kornfield J. A.; Tirrell, D. A. *Nat. Mater.* **2006**, *5*, 153-158. Copyright 2006. Nature Publishing Group.

5.1 Introduction

Artificial protein hydrogels assembled through aggregation of leucine zipper domains have the capacity for self-assembly encoded in the protein sequence[1]. Gelation does not require chemical crosslinking reagents that can compromise material safety in biomedical applications. The modular nature and fidelity of the biosynthetic method used to create these artificial proteins allow different biological determinants—including cell binding domains and enzyme recognition sites—to be incorporated readily in precisely controlled fashion. These advantages make genetically engineered, physically crosslinked hydrogels promising candidates for applications in biomedical fields such as controlled release and tissue engineering.

Control of erosion rate is a critical design objective for biomedical hydrogels[2-6]. For chemically crosslinked hydrogels, erosion behavior is controlled by introducing linkages susceptible to hydrolysis[6] or enzymatic cleavage[5]. On the other hand, many physical hydrogels exhibit undesirably rapid erosion when placed in open aqueous environments[2,7-12]. For hydrogels assembled from hydrophobically modified ethylene oxide polymers (PEOs), this issue has been addressed by controlling phase separation behavior[13-15]; molecular structure is adjusted to produce a transient network that coexists with a dilute sol phase in order to confer slow surface erosion. For ionically crosslinked alginate hydrogels, covalent cross-linking has been used to improve hydrogel stability[2]. Other examples of rapidly-eroding physical hydrogels include those formed by PEO-PPO-PEO block copolymers[11] and by mixtures of PEOs and

α-cyclodextrin[10]. For the physically crosslinked leucine zipper hydrogels reported to date (e.g. $AC_{10}A$, Fig. 5.1), rapid erosion was observed in open solutions. For example, a 1-mm-thick $AC_{10}A$ hydrogel (7% w/v) dissolves completely within 2.9 hours in 100 mM, pH 7.6 phosphate buffer. Such rapid dissolution precludes applications in which the gel must persist while immersed in excess fluid.

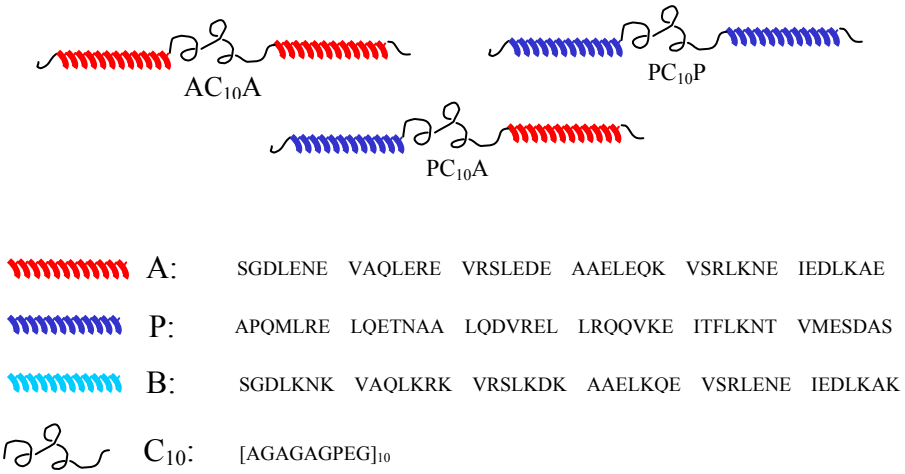


Figure 5.1 Schematic representations of triblock proteins and the amino acid sequences of major domains. The major domains of each triblock protein are joined by short sequences of amino acids introduced in the construction of the cloning and expression vectors. Each protein carries an N-terminal hexahistidine tag to allow the protein to be purified by affinity chromatography on a nickel nitrilotriacetic acid.

Our previous studies of the structural and dynamic properties of $AC_{10}A$ hydrogels in closed systems revealed that these multi-domain protein chains have a strong tendency to form intramolecular loops[16]. The aggregation number of the associative

domain (A) is small (4)[16,17], and association is transient[16]. These three factors all contribute to the fast erosion of $AC_{10}A$ networks. Disengaged clusters form readily because intramolecular loops are favored and the aggregation number of the leucine zipper domain is small (Fig. 5.2a). Since the strand exchange time of the leucine zipper domain is on the order of 100 to 1000 seconds[16] near physiological pH, the

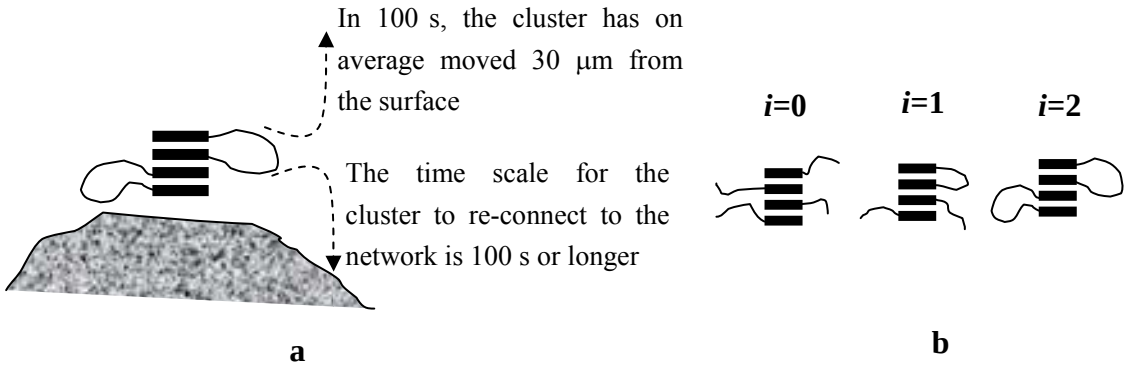


Figure 5.2 Structural and dynamic properties underlying the fast erosion of $AC_{10}A$ hydrogels. a, Disengaged clusters form readily in the system because of the strong tendency toward intramolecular association and the small aggregation number of the associative domain. They are lost from the surface through diffusion before re-connecting to the network. b, Three possible states of tetrameric aggregates designated by the number of loops *i*.

time scale for disengaged clusters to re-bind to the network is 100 seconds or longer. In 100 seconds, a cluster with diffusivity of 10^{-7} cm^2/s has on average moved 30 μm away from the surface of the network, and is lost to the surrounding buffer before it

can re-connect to the network. The transient nature of the association continuously releases disengaged clusters, leading to a quasi-steady concentration of free clusters at the surface such that their diffusive flux balances their rate of creation at the network surface. Consequently, the network erodes at a constant rate from its surface. The rate of cluster release is governed by the probability of a cluster simultaneously losing all of its connections to the network. We infer that among the structural and dynamic properties that cause the fast erosion of $AC_{10}A$ networks, an essential factor is facile intramolecular association. At any given moment, many tetrameric aggregates of the A domain have no connection to the network (Fig. 5.2b i=2), and many adopt configurations in which exchange of a single leucine zipper can liberate the cluster (Fig. 5.2b i=1).

We speculated that suppressing intramolecular loops in the networks would substantially reduce the erosion rate, and that intramolecular loops could be suppressed by engineering two dissimilar endblocks that do not associate with each other. Fidelity of molecular recognition between protein domains, which is the basis for many aspects of biological function, provides us the opportunity to control network structure and reduce the erosion rate. In particular, we speculated that the coiled-coil domain derived from the N-terminal fragment of rat cartilage oligomeric matrix protein (COMP)[18] would be likely to prefer homo-oligomerization (self-aggregation of identical protein domains) rather than hetero-oligomerization (aggregation of different protein domains) with leucine zipper A. The coiled-coil

domain from COMP assembles into five-stranded bundles[19,20]; in contrast, the A domain assembles into tetramers. The different packing structures of these two domains should suppress hetero-oligomerization.

5.2 Results and discussion

This expectation was confirmed experimentally. To create the coiled-coil domain (designated P) for physical association in a reversible gel, the amino acid sequence of the wild-type COMP domain was modified slightly to avoid chemical crosslinking: the two cysteine residues were mutated into serine residues. A DNA fragment encoding the P domain was inserted into pQE9 and expressed in *Escherichia coli*. The resulting protein was expressed, and its expected molar mass (6942 Da) was confirmed by mass spectral analysis. Multi-angle light scattering measurements for 30 μM and 107 μM P solutions revealed average molecular weights of 34310±380 and 35260±160, respectively, suggesting that the cysteine-free variant retains pentameric association. Native electrophoresis of a solution containing AC_{10} and P (100 μM in each protein, incubated at pH 7.6 and room temperature overnight) yielded two separate bands in which proteins migrated at the same rates as AC_{10}–aggregates and P-aggregates, respectively (Fig. 5.3a). As a control experiment in which hetero-oligomerization is dominant, retardation in migration of AC_{10} due to its strong association with leucine zipper B[21] (sequence shown in Fig. 5.1) was observed on the same gel. The results of these experiments suggest that the A and P domains

discriminate against each other in mixtures. Mass spectral analysis of the trypsin digests of the resolved protein bands provided further confirmation that each band contains only one species: each band yielded signals corresponding either solely to P or solely to AC_{10} (Fig. 5.3b).

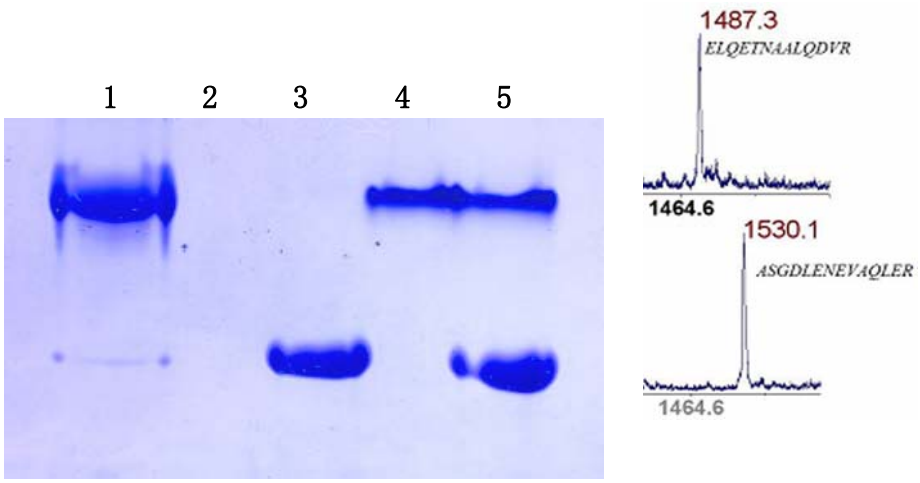


Figure 5.3 Coiled-coil domains A and P do not associate with each other. a, Native electrophoresis of recombinant proteins. Lane 1. AC_{10} + B; 2. B; 3. AC_{10}; 4. P; 5. AC_{10} + P. Protein B by itself does not migrate into the gel due to its net positive charge. b, Mass spectral analysis of trypsin digests of the proteins in the two bands excised from lane 5. ELQETNAALQDVR and ASGDLENEVAQLER are fragments from P and AC_{10}, respectively.

New triblock proteins $PC_{10}A$ and $PC_{10}P$ (Fig. 5.1) were then expressed and characterized. The molar masses of $PC_{10}A$, $PC_{10}P$ and $AC_{10}A$ are 20860 Da, 20486

Da and 22105 Da, respectively. All three proteins have nearly identical midblocks and their coiled-coil domains are the same length (six heptad repeats). They all assemble into hydrogels in aqueous solutions. Rheological oscillatory shear measurements

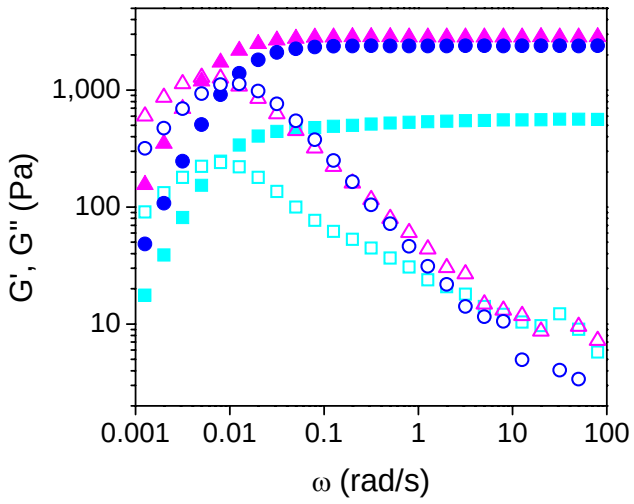


Figure 5.4 Dynamic moduli (closed symbols for storage moduli; open symbols for loss moduli) of $AC_{10}A$ (■, □); $PC_{10}P$ (●, ○); and $PC_{10}A$ (▲, △) hydrogels. (7% w/v, 100 mM phosphate buffer, pH 7.6, 22 °C)

revealed increased rigidity of the new materials compared to $AC_{10}A$ hydrogels (Fig. 5.4). The normalized plateau storage moduli G'_{∞}/nkT (G'_{∞}, plateau storage modulus; n, chain number density) were 0.35±0.01, 0.29±0.02 and 0.07±0.00 for $PC_{10}A$, $PC_{10}P$ and $AC_{10}A$, respectively, suggesting that loops are suppressed in the new materials.

Suppression of looped chains in $PC_{10}A$ gels was expected because the A domain and the P domain do not associate with each other. A decrease in the fraction

of looped chains in $PC_{10}P$ gels relative to $AC_{10}A$ might result from two structural features of the associative domains. First, the odd aggregation number of the pentameric P domain limits the maximum loop fraction in $PC_{10}P$ networks to 80%, while there is no such constraint for $AC_{10}A$ networks with tetrameric junctions. Another possible source of this different behavior may be the orientation of the A and P peptide strands in their aggregates. The isolated COMP pentamerization domain forms exclusively parallel aggregates[18,22], while our previous studies have shown that the isolated A domain can adopt an antiparallel orientation[16]. The length of the helical domains (A or P) with six heptad repeats is 65 Å[23], while the average hydrodynamic diameter of midblock chains is 40 Å as determined by quasi-elastic light scattering measurements[16]. Therefore, the midblock would have to stretch for a loop to form with parallel association of the end domains. Loops form readily in $AC_{10}A$ networks because A can adopt antiparallel association. In contrast, formation of loops in $PC_{10}P$ networks costs energy either to stretch the midblock (if the endgroups are parallel) or to adopt a thermodynamically unfavorable orientation (if the endgroups are antiparallel).

Network relaxation dynamics of $PC_{10}A$, $PC_{10}P$, and $AC_{10}A$ hydrogels are revealed from the frequencies of the maxima in their loss moduli (Fig. 5.4). The dominant stress relaxation time (the reciprocal of the frequency at which the loss modulus peaks) of the $PC_{10}P$ gel (ca. 80 s) is shorter than those of the $AC_{10}A$ gel (ca. 130 s) and $PC_{10}A$ gel (ca. 200 s). Our previous studies showed that the dominant

stress relaxation time of an $AC_{10}A$ hydrogel is strongly correlated with the strand exchange time of the leucine zipper domain[16], but systematically shorter than the strand exchange time by a factor of 3~4 due to the formation of looped chains[16,24]. The higher storage modulus of the $PC_{10}P$ gel relative to the $AC_{10}A$ gel suggests fewer loops and therefore a closer correspondence of the stress relaxation and strand exchange times. Therefore, the rate of exchange of the P domain must be greater than that of the A domain to account for the faster relaxation of $PC_{10}P$ relative to $AC_{10}A$.

Despite the fact that P domains undergo more rapid strand exchange, introduction of P into multi-domain proteins results in materials characterized by slower erosion rates in open aqueous solutions. The erosion profiles of 7% w/v $AC_{10}A$, $PC_{10}P$, and

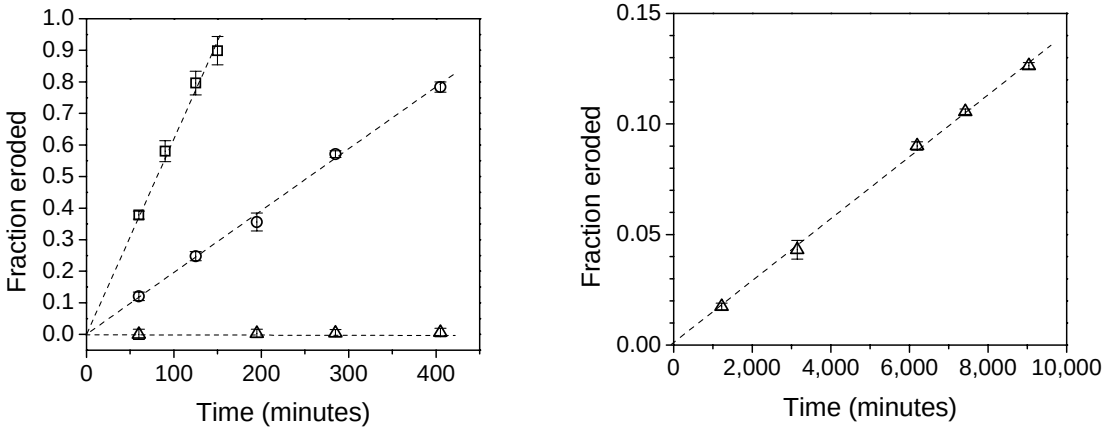


Figure 5.5 Erosion profiles of $AC_{10}A$ (□); $PC_{10}P$ (○); and $PC_{10}A$ (△) hydrogels. (7% w/v, 100 mM phosphate buffer, pH 7.6, room temperature) The surface area of each gel is 0.5672 cm^2. The total mass of each gel is 60 mg. Erosion rates: 4.3×10^{-2} mg/cm^2min for $AC_{10}A$, 1.3×10^{-2} mg/cm^2min for $PC_{10}P$ and 9.6×10^{-5} mg/cm^2min for $PC_{10}A$.

$PC_{10}A$ hydrogels (Fig. 5.5) all show linear mass-loss vs. time profiles, indicating that erosion is occurring at the surface rather than in the bulk. The erosion rates are 4.3×10^{-2} mg/cm^{2}min, 1.3×10^{-2} mg/cm^{2}min, and 9.6×10^{-5} mg/cm^{2}min for $AC_{10}A$, $PC_{10}P$, and $PC_{10}A$ hydrogels, respectively. A $PC_{10}A$ gel erodes ca. 500 times more slowly than an $AC_{10}A$ gel, and ca. 135 times more slowly than a $PC_{10}P$ gel. For 1-mm-thick samples, an $AC_{10}A$ gel dissolves completely within 2.9 hrs, while a $PC_{10}A$ gel erodes in 50 days.

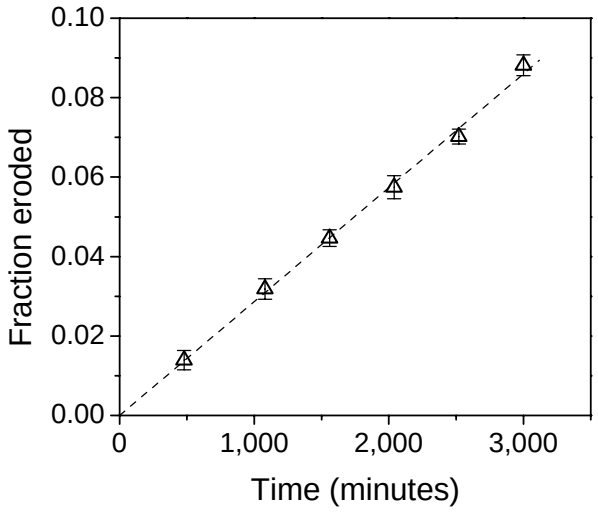


Figure 5.6 The erosion profile of $PC_{10}A$ hydrogels (7% w/v) at 37 °C in Dulbecco's Phosphate Buffered Saline (1x, pH 7.4). The surface area of each gel is 0.5672 cm^2. The total mass of each gel is 60 mg. The erosion rate is 1.9×10^{-4} mg/cm^{2}min.

The gradual erosion of $PC_{10}A$ at 22 °C into 100 mM phosphate buffer also holds for erosion under physiologically relevant conditions. The erosion rate of 7% w/v $PC_{10}A$ gels at 37 °C in Dulbecco's Phosphate Buffered Saline (PBS) (1x, pH 7.4) is 1.9×10^{-4} mg/cm^{2}min, only two-fold greater than that at room temperature(Fig. 5.6).

The erosion of $PC_{10}A$ at 37 °C is still more than 200 times slower than that of $AC_{10}A$ at room temperature (a 1-mm-thick sample erodes in 25 days). Erosion of $PC_{10}A$ at 37 °C in Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum added was monitored by gel electrophoresis of the supernatant followed by densitometry analysis of the $PC_{10}A$ protein bands using NIH ImageJ. The erosion rate is consistent with that in PBS (Fig. 5.7).

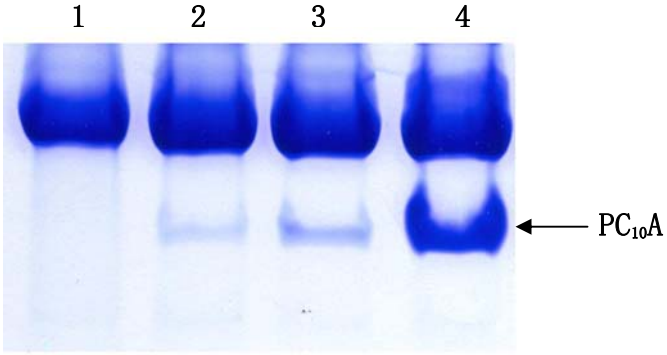


Figure 5.7 The erosion of $PC_{10}A$ hydrogels at 37 °C in 3 ml of Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum added monitored over 4 days.Lane 1. medium; 2. supernatant collected after the gel eroded for 48 hrs; 3. supernatant collected after the gel eroded for 96 hrs; 4. $PC_{10}A$ solution at a concentration corresponding to 100% dissolved gel in the medium.

The erosion rate of $PC_{10}A$ networks is reduced by 2~3 orders of magnitude relative to $AC_{10}A$ and $PC_{10}P$ networks. Since $PC_{10}A$ is constituted from the same associative domains as $AC_{10}A$ and $PC_{10}P$, the significant decrease in erosion rate cannot originate from the strand exchange kinetics of the associative domains. The

slow erosion of $PC_{10}A$ networks is not a consequence of phase separation behavior either—a constant erosion rate for $PC_{10}A$ gels was observed over a period of 7 days (Fig. 5.5b) even though the supernatant was not refreshed, indicating that the surrounding solution was far from saturation. It is control of network topology that causes the slow erosion of $PC_{10}A$ networks. Among the three gels, $PC_{10}A$ exhibits the highest G'_∞/nkT value, suggesting that it has the fewest intramolecular loops. The concomitant decrease in erosion rate supports the design concept proposed here: controlling network structure to suppress loops reduces erosion rates of transient networks formed from artificial proteins. The erosion rate decreases much more strongly than the modulus increases due to the different physics involved. Erosion requires that free species be present at the surface of the gel. In the extreme that loops are forbidden, the concentration of free species at the surface of the gel can be reduced by orders of magnitude. On the other hand, the upper bound on the modulus is nkT when all strands form bridges. The modulus lies between nkT and its minimum value at the percolation threshold. The range of the modulus is therefore constrained to vary by less than one order of magnitude.

The small, discrete number of endblocks per aggregate (m=4 for A and m=5 for P) is conducive to a simple statistical analysis of the relationship between the fraction of looped chains q and the number of effective crosslinks in the network. The model also gives qualitative insight into the effects of q on the erosion rate. Consider a solution of n chains having tetrameric aggregates of the endblocks. Each aggregate has three

possible states (Fig. 5.2b), with i=0, 1, or 2 loops, respectively. Of the $2n/m$ aggregates, the fraction of the aggregates having i loops (f_i) is related to the overall fraction of looped chains q through balances on the number of looped chains and the total number of aggregates (n/2):

$$\frac{2n}{m}(f_1 + 2f_2) = nq \qquad (1)$$

$$\sum_{i=1}^{3} f_i = 1 \qquad (2)$$

These equations hold for both tetrameric and pentameric aggregates. To solve for the f_i in terms of q, it is useful to define the relative probability that an aggregate has a single loop:

$$f_1 / f_0 \equiv s \qquad (3)$$

Since the relative probability that an aggregate has two loops is:

$$f_2 / f_0 = \frac{1}{2}s^2 \quad \text{(two loops in state } i\text{=2 are indistinguishable)} \qquad (4)$$

The conservation equations (1) and (2) yield a quadratic equation for s in terms of q. The solution is:

$$s = \frac{\frac{2}{m} - q - \sqrt{(\frac{2}{m} - q)^2 - 2q(q - \frac{4}{m})}}{q - \frac{4}{m}} \qquad (5)$$

In terms of s, the fraction of aggregates in each state is:

$$f_i = \frac{s^i}{i!} \cdot \frac{1}{\sum_{i=0}^{2} \frac{s^i}{i!}} \quad (i\text{=0, 1, 2}) \qquad (6)$$

An aggregate may form a network junction if it has three or more bridges. For

tetrameric aggregates, this includes only i=0 aggregates. So its modulus relative to a loop-free network (q=0) is simply:

$$\frac{G'_{\infty}}{G'_{\infty(q=0)}} = f_0 \qquad (m=4)$$

The fraction of tetrameric aggregates that are not connected to the network is:

$$\phi_{free} = f_2 \qquad (m=4)$$

If we use the observed modulus of a 7% w/v $PC_{10}A$ as a q = 0 value, then a 7% w/v $AC_{10}A$ gel has $f_0 \approx 0.2$, which can be explained in terms of loop fraction $q \approx 0.6$ (Table 1). For this value of q, 40% of the aggregates are free (f_2 = 0.4), consistent with the observation of rapid erosion.

In the case of $PC_{10}P$, which has pentameric aggregates of the endblocks, the normalized modulus reflects aggregates with i=0 and i=1 because both have ≥3 bridges: $\frac{G'_{\infty}}{G'_{\infty(q=0)}} = f_0 + f_1$. Again using the modulus of a $PC_{10}A$ gel as a q = 0 value, we infer that a 7% w/v $PC_{10}P$ gel has $f_0+f_1 \approx 0.8$, corresponding to $q \approx 0.3$. For this value of q, 20% of the aggregates are in the state i=2. However, these aggregates do not represent free species, since there is still a bridging chain. An upper bound on the erosion rate of $PC_{10}P$ is provided by considering a single attachment to be released so frequently that it does not limit the erosion rate. Then a $PC_{10}P$ gel would have an erosion rate that is half that of an $AC_{10}A$ gel ($f_{2(m=5)}/f_{2(m=4)} \approx 0.5$). The greater the effect of the attached arm, the slower $PC_{10}P$ would erode. The observed erosion rate of $PC_{10}P$ is 1/4~1/3 that of $AC_{10}A$—surprisingly close to the upper bound.

In contrast to $AC_{10}A$ and $PC_{10}P$, it is not possible for a single oligomer of endblocks to leave a $PC_{10}A$ gel due to the strong suppression of hetero-oligomerization. Instead, an A tetramer cannot be liberated without bringing four P domains with it. Compared to larger clusters, a pair of P and A aggregates ("P-A pair") with four bridges to each other is the most probable species that can be released, although liberation of such species is associated with some free energy penalty (for example, for a P tetramer relative to a P pentamer). The erosion rate of $PC_{10}A$ relative to that of $AC_{10}A$ reflects, in part, the small probability of free P-A pairs in $PC_{10}A$. If an aggregate has *m* coiled-coils in it and *t* neighboring aggregates, then the probability that it forms all of its *m* bridges with just one neighbor is t^{1-m}. For example, for a tetrameric aggregate having 6 neighbors, the probability of the occurrence of such species is ~1/200. In contrast, ca. 40% of the aggregates have two loops and are free in an $AC_{10}A$ gel. An additional reduction in erosion rate of $PC_{10}A$ is due to the energy penalty noted above. This analysis is consistent with the observation that a $PC_{10}A$ gel erodes 2~3 orders of magnitude more slowly than an $AC_{10}A$ gel.

Control of network topology substantially expands the range of material properties that can be achieved in artificial protein hydrogels. Previously established principles for engineering thermal and pH responsiveness of coiled-coil domains[25-30] can now be coupled with molecular design for desired network structure to confer properties that are otherwise inaccessible. Design principles based on network

topology proved effective under physiologically relevant conditions. Prior results demonstrated that these gels are non-toxic (by viability of mammalian 3T3 fibroblast cells cultured in the presence of $AC_{10}A$[16]). Thus, the approach presented here can be used to optimize systems for a broad range of applications in biology and medicine.

5.3 Materials and methods

The DNA segment encoding the P domain was created by PCR assembly. Expression vectors pQE9PC_{10}P, pQE9PC_{10}A, and pQE9P were constructed by standard recombinant DNA techniques. Proteins were expressed and purified as described previously[1]. To determine the oligomerization state of the P domain, multi-angle static light scattering measurements were carried out on a DAWN EOS light scattering instrument (Wyatt Technology Corporation, CA) and the data were analyzed with Debye plots by using a dn/dc value of 0.185[31]. To examine whether P and A domains tend to associate with each other, native electrophoresis was performed on 12% polyacrylamide gels using the standard protocol with SDS and reducing agents omitted from all solutions. The resolved protein bands were cut from the gel. After the Coomassie stain was removed from each protein band[32], digestion with 0.02 mg/mL trypsin (Promega) was allowed to proceed at 37 °C overnight. Mass spectral analysis of the trypsin digests was performed on an Applied Biosystems Voyager mass spectrometer using MALDI matrix α-cyano-β-hydroxycinnamic acid (10 mg/mL in 50% CH_3CN). Rheological oscillatory shear measurements for $PC_{10}A$,

$PC_{10}P$, and $AC_{10}A$ gels were carried out on an RFS III rheometer (TA Instruments, New Castle, Delaware) with a cone-and-plate geometry (0.04 rad cone angle and 25-mm diameter). To measure the erosion rates of hydrogels, 1-mm-thick flat gel films were made in cylindrical plastic containers of 8.5 mm diameter and 3 mm height, which were then placed in 3 mL phosphate buffer (100 mM, pH 7.6) in scintillation vials with the gel surfaces facing down. The erosion profiles were determined by measurement of the protein concentration in the supernatant at successive time points. Protein concentrations were determined by measuring the absorbance at 280 nm on a Cary 50 Bio UV-vis spectrophotometer (Varian, Palo Alto, CA). The experiments were performed in triplicate.

To characterize erosion of $PC_{10}A$ in the presence of serum proteins, erosion from one face of an 8.5 mm diameter × 1.06 mm thick gel sample at 37 °C into 3 ml of Dulbecco's Modified Eagle's Medium (DMEM) with 10% fetal bovine serum added was monitored over 4 days. To distinguish $PC_{10}A$ from the serum protein background, the medium was sampled, mixed with loading buffer and separated using SDS-polyacrylamide gel electrophoresis. The volume of medium sample loaded onto each lane was kept constant (10 μl). Protein bands were visualized by staining with Coomassie Blue. The gel was digitally imaged and densitometry analysis of $PC_{10}A$ bands was performed using NIH ImageJ. Interpretation of $PC_{10}A$ concentrations in medium was based on calibration $PC_{10}A$ solutions of known concentrations. Densitometry analysis indicates that at 48 hrs and 96 hrs 6% and 15%, respectively, of

the initial gel had dissolved. Although there is certain uncertainty in the present method, it shows that the erosion time in the presence of serum proteins is on the same order as the time for erosion in PBS.

Acknowledgment. The authors acknowledge the NSF Center for the Science and Engineering of Materials for financial support.

5.4 References

(1) Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, *281*, 389.

(2) Lee, K. Y.; Mooney, D. J. *Chem. Rev.* **2001**, *101*, 1869.

(3) Anseth, K. S.; Metters, A. T.; Bryant, S. J.; Martens, P. J.; Elisseeff, J. H.; Bowman, C. N. *J. Control. Rel.* **2002**, *78*, 199.

(4) Yamaguchi, N.; Kiick, K. L. *Biomacromolecules* **2005**, *6*, 1921.

(5) Rizzi, S. C.; Hubbell, J. A. *Biomacromolecules* **2005**, *6*, 1226.

(6) Kim, B. S.; Hrkach, J. S.; Langer, R. *Journal of Polymer Science Part a-Polymer Chemistry* **2000**, *38*, 1277.

(7) Kaczmarski, J. P.; Glass, J. E. *Macromolecules* **1993**, *26*, 5149.

(8) Lundberg, D. J.; Brown, R. G.; Glass, J. E.; Eley, R. R. *Langmuir* **1994**, *10*, 3027.

(9) Tam, K. C.; Jenkins, R. D.; Winnik, M. A.; Bassett, D. R. *Macromolecules*

1998, *31*, 4149.

(10) Li, J.; Ni, X. P.; Leong, K. W. *Journal of Biomedical Materials Research Part A* **2003**, *65A*, 196.

(11) Anderson, B. C.; Pandit, N. K.; Mallapragada, S. K. *Journal of Controlled Release* **2001**, *70*, 157.

(12) Wang, C.; Stewart, R. J.; Kopecek, J. *Nature* **1999**, *397*, 417.

(13) Pham, Q. T.; Russel, W. B.; Thibeault, J. C.; Lau, W. *Macromolecules* **1999**, *32*, 2996.

(14) Tae, G.; Kornfield, J. A.; Hubbell, J. A.; Johannsmann, D.; Hogen-Esch, T. E. *Macromolecules* **2001**, *34*, 6409.

(15) Francois, J.; Beaudoin, E.; Borisov, O. *Langmuir* **2003**, *19*, 10011.

(16) Shen, W., Ph.D. Dissertation; California Institute of Technology, CA 2005.

(17) Kennedy, S. B. Ph.D. Dissertation, Ph.D. Dissertation; University of Massachusetts Amherst, MA 2001.

(18) Malashkevich, V. N.; Kammerer, R. A.; Efimov, V. P.; Schulthess, T.; Engel, J. *Science* **1996**, *274*, 761.

(19) Efimov, V. P.; Engel, J.; Malashkevich, V. N. *Proteins-Structure Function and Genetics* **1996**, *24*, 259.

(20) Efimov, V. P.; Lustig, A.; Engel, J. *Febs Letters* **1994**, *341*, 54.

(21) Petka, W. A., Ph.D. Dissertation; University of Massachusetts Amherst, MA 1997.

(22) Ozbek, S.; Engel, J.; Stetefeld, J. *Embo Journal* **2002**, *21*, 5960.

(23) Hodges, R. S. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **1996**, *74*, 133.

(24) Annable, T.; Buscall, R.; Ettelaie, R.; Whittlestone, D. *Journal of Rheology* **1993**, *37*, 695.

(25) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401.

(26) Kohn, W. D.; Hodges, R. S. *Trends in Biotechnology* **1998**, *16*, 379-389.

(27) Litowski, J. R.; Hodges, R. S. *Journal of Biological Chemistry* **2002**, *277*, 37272.

(28) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Science* **2001**, *10*, 649.

(29) Wagschal, K.; Tripet, B.; Hodges, R. S. *Journal of Molecular Biology* **1999**, *285*, 785.

(30) Yu, Y. B. *Advanced Drug Delivery Reviews* **2002**, *54*, 1113.

(31) Huglin, M. B. *Light scattering from polymer solutions*; Academic Press: London, New York, 1972.

(32) Hellman, U.; Wernstedt, C.; Gonez, J.; Heldin, C. H. *Analytical Biochemistry* **1995**, *224*, 451.

CHAPTER VI

Engineering Cooperativity in Biomotor-Protein Assemblies

Abstract

A biosynthetic approach was developed to control and probe cooperativity in multiunit biomotor assemblies by linking molecular motors to artificial protein scaffolds. This approach provides precise control over spatial and elastic coupling between motors. Cooperative interactions between monomeric kinesin-1 motors attached to protein scaffolds enhance hydrolysis activity and microtubule gliding velocity. However, these interactions are not influenced by changes in the elastic properties of the scaffold, distinguishing multimotor transport from that powered by unorganized monomeric motors. These results highlight the role of supramolecular architecture in determining mechanisms of collective transport.

The text in this chapter is reprinted with permission from Diehl, M. R.; Zhang, K. C.; Lee, H. J.; Tirrell, D. A. *Science* 2006, *311*, 1468-1471. Copyright 2006. American Association for the Advancement of Science (AAAS).

6.1 Introduction

Protein cooperativity allows systems of biomotor assemblies to operate with greater determinism and efficiency and often provides physiological functionality that cannot be achieved by single molecules.[1-9] For example, cooperation between RNA polymerase molecules can result in increased rates of transcription[2], and increased transport velocities have been observed with groups of monomeric kinesin motors[3,4] and with multimotor assemblies comprising dimeric kinesins and dyneins[5]. In the latter case, assemblies traveled in vivo with velocities up to 10 times as high as those observed in vitro. This result implies the presence of intermotor interactions *in vivo* that are not reproduced in *in vitro* assays. Although models of biomotor cooperativity[1,6] can explain generic aspects of multimotor transport and predict new modes of transport such as spontaneous oscillations[7-11], the molecular details that give rise to these cooperative effects remain elusive.

Many systems of motors are arranged in highly organized and hierarchical architectures *in vivo*[12,13], but it is not clear how features such as the mechanical compliance of motor-motor linkages and intermotor spacing influence collective dynamics. Because the mechanochemistry of biomotors is strongly dependent on strain and hence on the mechanical coupling between motors[14,15], developing a more complete picture of collective motility requires a better understanding of the relations between architecture and function in multimotor assemblies.

6.2 Results and discussion

To investigate the influence of supramolecular architecture on biomotor cooperativity, we have engineered a model multimotor system that allows us to precisely regulate intermotor coupling. We have synthesized a series of modular polymeric scaffolds (Fig 6.1) in which molecular properties such as length, sequence, and secondary structure are specified by artificial genes that encode alternating rigid and elastic protein motifs. The rigid block is comprised of strongly associated acidic and basic leucine zipper domains that anchor motor proteins at specific distances along the polymer backbone. Based on amino acid sequences developed by Vinson et al.[16], these zippers form strong heterodimeric complexes ($K_D \sim 10^{-15}$ M) and much weaker homodimers ($K_D \sim 10^{-6}$ to 10^{-3} M). The artificial protein scaffolds incorporate the basic zipper (Z_R) into the polymer backbone, whereas the complementary acidic zipper (Z_E) is fused to the C terminus of a truncated kinesin-1 motor (designated K350-Z_E). The flexible polymer block is derived from the elastomeric poly(VPGV_aG) structural motif of the protein elastin (EL) and confers well-characterized mechanical compliance on the assembly[17-19]. Every fifth α-valine (V_a) residue is replaced by a phenylalanine (F) residue in the EL sequence used here, yielding the designation ELF. This substitution provides a means to control the thermoresponsive behavior of the polymers, as discussed in more detail below. Variation in the number of diblock repeats in the polymer provides discrete control over the number of coupled motors, which in the present series of experiments ranges from one to three. The C terminus of

each scaffold is labeled with biotin to allow the motor assemblies to be tethered to streptavidin-coated surfaces.

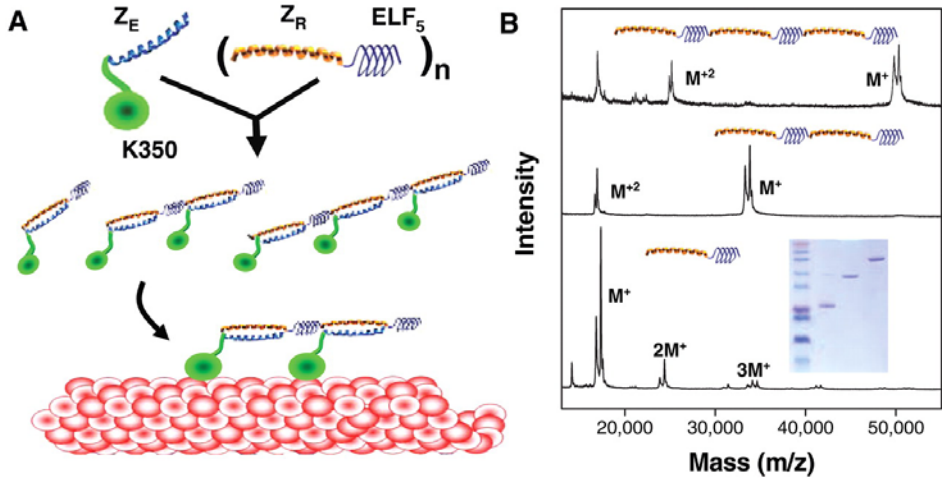


Figure 6.1 Engineered multimotor assemblies. (**A**) Schematic representation of the synthesis of the engineered multimotor assemblies. As a complex, the zippers (Z_E and Z_R) form a rigid linker approximately 6.5-nm long (assuming 6.3 heptad repeats in a zipper and 1.03 nm per heptad).[30] The length of the flexible ELF block can be approximated by assuming a ß-spiral conformation. In this conformational state, elastin proteins possess a spiral pitch of 1 nm, where each turn contains three VPGV G pentapeptide units. Repeating this ELF motif, $(VPGVG)_2VPGFG(VPGVG)_2$, five times gives a length of 8 nm for the ELF_5 domain. Considering that four amino acids (KASK) form linkers between adjacent Z_R-ELF_5 diblock units, the total intermotor spacing set by the polymer is approximately 16 nm when bound to a microtubule (shown in red). (**B**) Matrix-assisted laser desorption/ionization mass spectra of polymer scaffolds containing one, two, and three repeats of the Z_R-ELF_5 diblock. The

splitting of the main peaks (M+) is due to a 525-Da shift in mass that arises from biotin functionalization at the C-terminal cysteine positions of the polymers. A tris-tricine gel of all three polymers is shown in the inset.

The monomeric kinesin-1 construct contains a Z_E-fusion to the motor's catalytic domain and neck linker. Monomeric truncations of kinesin-1 that contain the neck linker are nonprocessive and maintain plus-end directionality[20]. As a result, the motions of individual K350-Z_E motors along microtubules can be described by Brownian diffusion models[21,22]. When anchored to the Z_R blocks of the artificial protein scaffolds $(Z_R\text{-}ELF_5)_n$, motors are separated by approximately 16 nm, or two microtubule lattice sites (Fig 6.1).

To examine the consequences of clustering multiple motors, we measured the microtubule-stimulated adenosine triphosphatase (ATPase) rates of the polymer-motor complexes (Fig 6.2A). Motor assemblies were preformed in solution by incubating the $(Z_R\text{-}ELF_5)_n$ polymers and K350-Z_E motors.[23] When assembled on dimeric $(Z_R\text{-}ELF_5)_2$ and trimeric $(Z_R\text{-}ELF_5)_3$ scaffolds, the K350-Z_E motors exhibit roughly a 60% increase in the maximum microtubule-stimulated ATPase rate (k_{cat}) accompanied by a decrease by a factor of 2.6 to 3.1 in the Michaelis-Menten constant ($K_{0.5MT}$) when compared with complexes formed on monomeric ($Z_R\text{-}ELF_5$) scaffolds (Fig 6.2). Similar results were obtained by using either microtubule affinity–purified or Ni-NTA (nickel nitrilotriacetic acid)–purified K350-Z_E motors (Table 6.1).

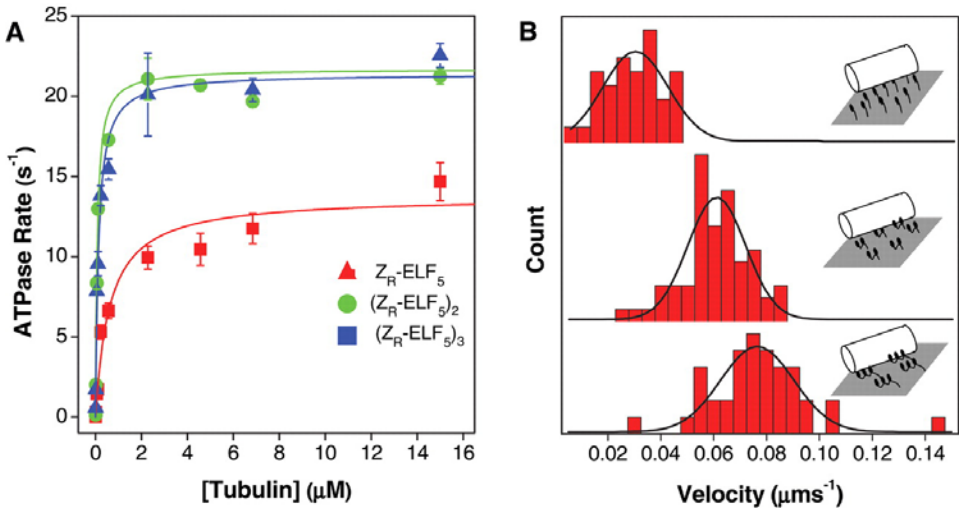


Figure 6.2 Enhancement of ATPase activity and motility when multiple K350-Z_E motors are coupled to (Z_R-ELF$_5$)$_n$ polymer scaffolds. (A) Microtubule-dependent ATPase activity of [K350-Z_E]$_n$/(Z_R-ELF$_5$)$_n$ assemblies. The multimotor assemblies, [K350-Z_E]$_2$/(Z_R-ELF$_5$)$_2$ and [K350-Z_E]$_3$/(Z_R-ELF$_5$)$_3$, exhibit a 52% to 67% increase in the maximum ATPase rate (k_{cat}) compared with [K350-Z_E]/(Z_R-ELF$_5$). The lines are Michaelis-Menten fits yielding $K_{0.5MT}$ (0.62 ± 0.20; 0.08 ± 0.01; 0.14 ± 0.20) and k_{cat} (13.7 ± 1.0; 21.7 ± 0.8; 21.4 ± 0.7) values for the monomer, dimer, and trimer complexes, respectively. (B) Velocity histograms for microtubules gliding over films prepared using preassembled motor/polymer complexes at 17°C for monomeric (top), dimeric (middle), and trimeric (bottom) complexes.

Assembly	MT Affinity-purified K350-Z_E			Ni-NTA purified K350-Z		
	$K_{0.5MT}$* (μM)	k_{cat}* (s^{-1})	$k_{cat}/K_{0.5MT}$* (μM^{-1} s^{-1})	V† (nm s^{-1})	ΔE (T < TC) (kJ mol^{-1})	ΔE (T > TC) (kJ mol^{-1})
[K350-Z_E]/(Z_R-ELF$_5$)	0.75 ± 0.30	15.4 ± 2.3	20	29 ± 22	150	105
[K350-Z_E]/(Z_R-ELF$_5$)$_2$‡	—	—	—	—	163	133
[K350-Z_E]$_2$/(Z_R-ELF$_5$)$_2$	0.29 ± 0.08	23.4 ± 1.7	81	61 ± 22	——58——	
[K350-Z_E]$_3$/(Z_R-ELF$_5$)$_3$	0.24 ± 0.04	25.7 ± 0.9	104	76 ± 27	——69——	

*Measured at 20°C †MT gliding at 17°C ‡Monomeric complex using dimeric polymer

Table 6.1 Summary of kinetics and temperature-dependent gliding assays.

Whereas the apparent bimolecular reaction rate ($k_{cat}/K_{0.5MT}$) of the monomeric motor complex falls within values predicted for diffusion-limited reactions (20 to 30 $\mu M^{-1}\ s^{-1}$) [24], the dimeric and trimeric complexes exhibit an increase by a factor of 4 or more in $k_{cat}/K_{0.5MT}$. This result suggests that the multiprotein complexes are processive. However, $k_{cat}/K_{0.5MT}$ values are smaller than those of native kinesin ($\geq$1000 $\mu M^{-1}\ s^{-1}$)[26], implying (as expected) a distinct transport mechanism.

In microtubule gliding assays, microtubule velocities of two- and three-headed multimotors are about twice that of the monomer constructs (Fig 6.2B). In these experiments, polymer and motor concentrations were chosen to ensure that the coverslip surface was saturated with motor assemblies. In each case, microtubules exhibited smooth gliding motions across the surface. Thus, microtubule gliding is occurring in a regime where transport is characterized by multiprotein suppression of individual motor fluctuations and is independent of the number of motors involved in motility.[6,26] As a result, multimotor complexes must use an additional mechanism that enhances the velocity-determining step of the K350-Z_E motors. This mechanism should be linked to the enhanced ATPase activity observed in solution phase experiments and is likely the result of specific motor-motor coupling that occurs when several motors are anchored along a single polymer chain. Interestingly, trimeric multimotor complexes produced gliding velocities similar to those of dimeric assemblies.

Insight into the nature of the intermotor coupling in the multimotor assemblies can be gained by tuning the elastic properties of the scaffold's ELF motif. Elastin-like polypeptides (ELPs) undergo a phase transition in which hydrophobic folding of the chain drives a condensation process, forming a denser viscoelastic phase when the temperature is raised above the lower critical solution temperature (LCST) of the protein[17]. Single-molecule atomic force micoscopy pulling experiments demonstrate that the LCST transition results in a decrease in polymer length.[27]

In concentrated polymer solutions, the condensation of ELPs can be monitored by measuring changes in turbidity with temperature (Fig. 6.3A). The transition temperature of the polymers increases with decreasing polymer length[28]. In temperature-dependent microtubule gliding assays (Fig. 6.3B), elastin condensation results in a decreased microtubule velocity. Similar behavior is observed when a stoichometric excess of the $(Z_R\text{-}ELF_5)_2$ polymer is used to produce longer monomeric motors. In each case, the microtubule gliding velocity increases with increasing temperature in accordance with a standard Arrhenius-like temperature dependence above and below the transition. However, the condensation of the ELF units is accompanied by a decrease in the slope, yielding 20 to 30% lower activation barriers above the LCST (Table 6.1).

The attenuation of microtubule velocity upon condensation of the ELF domains when transport is powered by monomeric complexes can be explained by previous models of cooperating motors. These models predict enhancements of

multimotor efficiency, velocity, and force generation when motor anchorages are stiff[1,3]. The observed decreases in both the velocities and activation barriers of microtubule transport are consistent with these theories if the simultaneous decrease in ELF length and stiffening of the mechanical linkage to the cover slip are taken into account. In microtubule gliding assays, the decrease in polymer length influences transport by producing a monomeric motor with a shorter "lever arm" and, consequently, a smaller working stroke.

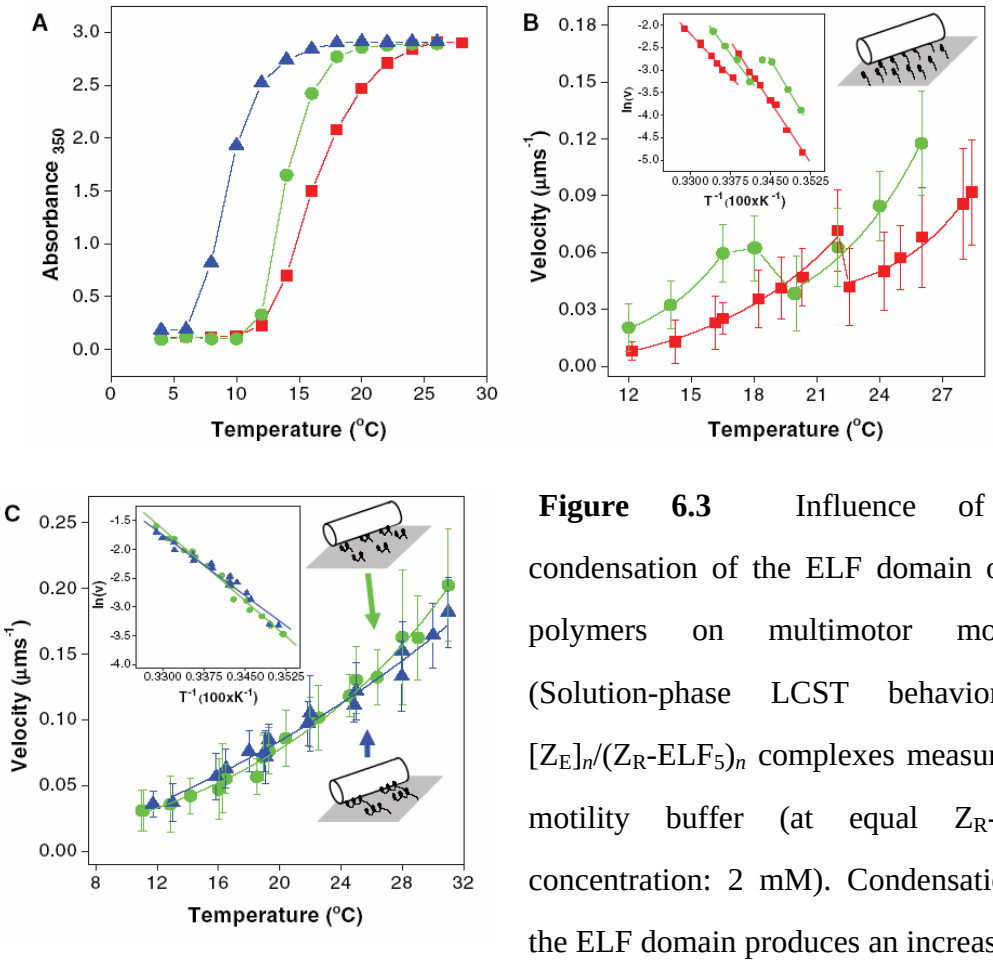


Figure 6.3 Influence of the condensation of the ELF domain of the polymers on multimotor motility. (Solution-phase LCST behavior of $[Z_E]_n/(Z_R\text{-}ELF_5)_n$ complexes measured in motility buffer (at equal $Z_R\text{-}ELF_5$ concentration: 2 mM). Condensation of the ELF domain produces an increase in turbidity of the solution as the temperature rises. (B) and (C) Temperature-dependent microtubule gliding velocities of monomeric (B) and multimeric (C) polymer/motor complexes. (B) Monomeric complexes were formed using either the $Z_R\text{-}ELF_5$ (red) or

the dimeric $(Z_R\text{-}ELF_5)_2$ (green) polymers. Monomeric assemblies were prepared from the dimeric $(Z_R\text{-}ELF_5)_2$ polymers by using a threefold stochiometric excess of the total Z_R attachment sites relative to the motor. (C) Temperature-dependent gliding of microtubules using fully functionalized dimeric $[K350\text{-}Z_E]_2/(Z_R\text{-}ELF_5)_2$ (green) and trimeric $[K350\text{-}Z_E]_3/(Z_R\text{-}ELF_5)_3$ (blue) complexes.

Multimotor assemblies consisting of fully functionalized dimeric $[K350Z_E]_2/(Z_R\text{-}ELF_5)_2$ and trimeric $[K350\text{-}Z_E]_3/(Z_R\text{-}ELF_5)_3$ complexes exhibit simple Arrhenius-like temperature dependence throughout the ELF phase transition region (Fig. 6.3C). The activation barriers obtained from these measurements are substantially lower than those of the monomeric motor assemblies (Table 6.1). Here, the insensitivity of multimotor transport to ELF condensation suggests a mechanism that relies on processes distinct from those that dominate transport driven by teams of unorganized monomeric complexes. One explanation for the difference is that mechanochemical coupling between neighboring motors is enhanced by the stiffening of the ELF linkages, compensating for a decrease in motor working distance. However, this possibility requires that these two competing factors are in near-perfect balance for both dimeric and trimeric assemblies. Alternatively, the velocity determining step of multimotor transport may not be dependent on the motions of lever arms that contain the ELF linkers as structural elements. Instead, engineered assemblies may use a multistep mechanism that is rate-limited by other mechanical processes, such as motions where the motor's neck linkers alone serve as lever arms or a diffusive search by a motor domain for its next binding site.

Although models involving diffusion-to-capture processes and/or conformational changes in neck linkers have been used to describe the stepping mechanics of kinesins[29], the results reported here indicate that the artificial proteins provide a structural framework that allows motors to push and pull on one another to enhance activity. Such cooperative interactions should lead to "inchworm-like" stepping motions that are influenced by weak mechanochemical coupling and, possibly, by coordinated displacements along the microtubule. The mechanism of movement should be influenced by architectural features of the assembly, including the large intermotor distance, the flexibility of the ELF linkages, the ability of an assembly to bind multiple microtubule sites, and the asymmetric anchoring to the surface at one end of the scaffold. These features create a structural framework where motors are attached to the microtubule and to the surface through different mechanical linkages and where strain is unequally distributed across the motors in an assembly. These factors should constrain the mechanism of multimotor transport by determining the local reference frame for the displacement of a motor within an assembly and should influence cooperative interactions by tuning both the strain-dependent detachment of, and the mechanochemical coupling between, neighboring motors. Although further experiments to investigate the details of multimotor transport are underway, the results described here clearly demonstrate that controlling the supramolecular architecture of multimotor assemblies provides a means to reconfigure mechanisms of collective biomotor transport.

Acknowledgement. We thank A. J. Link and I. Fushman their help during the early stages of this project; L. Wade and D. Pearson for use of the temperature controller; and P. Wiggins, R. Bao, T. Squires, and S. Quake for valuable discussions. This work was supported by the Beckman Foundation through a Beckman Senior Research Fellowship (to M.R.D.) and by a grant from the National Science Foundation.

6.3 References

(1) Julicher, F.; Ajdari, A.; Prost, J. *Rev. Mod. Phys.* **1997**, *69*, 1269.

(2) Epshtein, V.; Nudler, E. *Science* **2003**, *300*, 801.

(3) Okada, Y.; Hirokowa, N. *Nature* **2003**, *424*, 574.

(4) Klopfenstein, D. R.;Tomishige, M.; Stuurman, N.; Vale, R. D. *Cell* **2002**, *109*, 347.

(5) Kural, C.; Kim, H.; Syed, S.; Goshima, G.; Gelfand, V. I.; Selvin, P. R**.** *Science* **2005**,*308*, 1469

(6) Leibler, S.; Huse, D. A. *J. Cell Biol.* **1993**, *121*, 1357.

(7) Endow, S. A.; Higuchi, H. *Nature* **2000**, *406*, 913.

(8) Badoual, M.; Julicher, F.; Prost, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6696.

(9) Jülicher, F.; Prost, J.; *Phys. Rev. Lett.* **1997**, *78*, 4510.

(10) Camalet, S.; Julicher, F.; Prost, J. *J. Phys. Rev. Lett.* **1999**, *82*, 1590.

(11) Camalet, S.; Duke, T.; Julicher, F.; Prost, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 318.

(12) Burgess, S. A. *J. Mol. Biol.* **1995**, *250*, 52.

(13) Gross, S. P.; Guo, Y.; Martinez, J. E.; Welte, M. A. *Curr. Biol.* **2003**, *13*, 1660.

(14) Duke, T. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2000**, *355*, 529.

(15) Vilfan, A.; Frey, E.; Schwabl, F. *Eur. Phys. J. B* **1998**, *3*, 535.

(16) Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, *10*, 649.

(17) Urry, D. W. *J. Phys. Chem. B* **1997**, *101*, 11007.

(18) Urry, D. W.; Hugel, T.; Seitz, M.; Gaub, H. E.; Sheiba, L.; Dea, J.; Xu, J.; Parker, T. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2002**, *357*, 169.

(19) Polymer flexibility can be estimated by calculating the elastic spring constant (**κ**) of a single ELF_5 linker using data from single molecule force-extension measurements. These experiments yield persistence lengths (L_p) for elastins of ~0.4 nm in the low force regime. Assuming a wormlike chain model for the polymer at low extension, **κ** can be estimated from **κ** $= 3k_bT/2L_pL_c$, where L_c is the contour length of the polymer, k_b is the Boltzman constant, and T is temperature. A single ELF_5 linker contains ~125 amino acids, yielding L_c = 36.35 nm, assuming 0.29 nm per residue, and **κ** = 0.4 pN/nm.

(20) Case, R. B.; Pierce, D. W.; Hom-Booher, N.; Hart, C. L.; Vale, R. D. *Cell* **1997**, *90*, 959.

(21) Inoue, Y.; Iwane, A. H.; Miyai, T.; Muto, E.; Yanagida, T. *Biophys. J.* **2001**, *81*, 2838.

(22) Okada, Y.; Hirokawa, N. *Science* **1999**, *283*, 1152.

(23) Polymer-motor assemblies were formed by making a master mix of K350-Z_E and $(Z_R\text{-}ELF_5)_n$ polymers and incubating for at least 20 min at 4°C before addition to the reaction. Concentrations of polymer solutions were determined from their A_{280} values, using an extinction coefficient of 1480 cm^{-1} mol^{-1} for each Z_R-ELF_5 repeat in the polymers. The fidelity of the assembly process was examined for the trimeric $[K350\text{-}Z_E]_3/(Z_R\text{-}ELF_5)_n$ complex using selectively radiolabeled proteins and multichannel scintillation counting. In these experiments, a K350-Z_E motor was radiolabeled with ^{35}S (77,104 cpm/nmol) by expressing the motor in 1 liter of LB medium supplemented with L-[^{35}S]cysteine (5 mCi). Similarly, a ^3H labeled $(Z_R\text{-}ELF_5)_3$ polymer (33,046 cpm/nmol) was prepared by expression in 0.1 L of LB medium supplemented with L-[3,4(n)-^3H]valine (2.5 mCi). After purification, the polymer was functionalized with a PEO-biotinmaleimide (Pierce) using standard maleimide labeling protocols. Then, radiolabeled polymers and motors were mixed in a 1:1.5 ratio with respect to the Z_R sites of the polymer and the motor. After incubation, the [motor]/(polymer) complex was selectively bound to a neutravidin resin. Excess motor was washed from the resin, and the sample, including the resin, was transferred to a scintillation vial. Comparison of the signals from ^{35}S and ^3H channels yielded a motor/polymer ratio of 2.9 ± 0.4.

Control experiments were performed with the polymers omitted from solution; results indicated nonspecific binding of the K350-Z_E motors to the resin did not influence our measurements.

(24) Hackney, D. D. *Biophys. J.* **1995**, *68*, 267.

(25) Hackney, D. D. *Nature* **1995**, *377*, 448.

(26) Hancock, W. O.; Howard, J. *J. Cell Biol.* **1998**, *140*, 1395.

(27) Hyun, J.; Lee, W. K.; Nath, N.; Chilkoti, A.; Zauscher, S. *J. Am. Chem. Soc.* **2004**, *126*, 7330.

(28) Meyer, D. E.; Chilkoti, A. *Biomacromolecules* **2004**, *5*, 846.

(29) Carter, N. J.; Cross, R. A. *Nature* **2005**, *435*, 308.

(30) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 689.

APPENDIX

DNA sequences and plasmid maps

A.1 pQE60_(ZRhis-ZE(gs)$_6$ELF)

Submitted by Kechun Zhang **Date** 8 January 2007

Strain name AF-IQ

Vector pQE60_(ZRhis-ZE(gs)$_6$ELF)

Construction

A *Nhe*I fragment encoding mutant *E. coli* PheRS (A294G) was inserted into expression plasmid pQE60 to yield pQE-FS*. Then the ZR gene was ligated into the *Bam*HI site of pQE-FS* to yield ZR-pQE-FS*. Finally, a DNA fragment encoding ZE(gs)$_6$ELF and containing a ribosome binding site was amplified and ligated into the *Bpu*1102I site of ZR-pQE-FS*. The resulting plasmid (ZRhis-ZE(gs)$_6$ELF) was transformed into phenylalanine auxotroph *E. coli* strain AF-IQ.

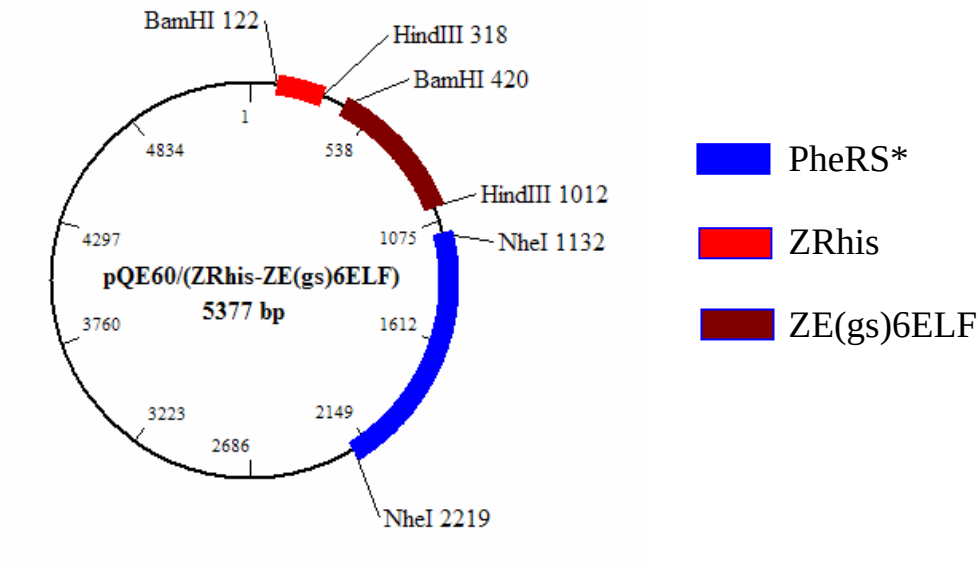
Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Reference

Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 10136-10137.

Plamsid map



Gene sequence

atgaaaggatccctggaaatccgtgcggcggcgctgcgtcgtcgtaacaccgcgctgcgt
M K G S L E I R A A A L R R R N T A L R
acccgtgttgcggaactgcgtcagcgtgttcagcgtctgcgtaacgaagtttctcagtac
T R V A E L R Q R V Q R L R N E V S Q Y
gaaacccgttacggtccgctgggtggttctggttctggttctggttctggtagcttcaga
E T R Y G P L G G S G S G S G S G S F R
tctcatcaccatcaccatcactaagcttaattagctgagccggataacaattataatagat
S H H H H H H -

tcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattaactatg
M
aaaggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcgctggaaacc
K G S L E I E A A A L E Q E N T A L E T
gaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttctcagtaccgt
E V A E L E Q E V Q R L E N I V S Q Y R
acccgttacggtccgctgggtggttctggttctggttctggttctggtagcggatccggg
T R Y G P L G G S G S G S G S G S G S G
gtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggt
V P G V G V P G V G V P G F G V P G V G
gttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggt
V P G V G V P G V G V P G V G V P G F G
gtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggt
V P G V G V P G V G V P G V G V P G V G
gtaccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggc
V P G F G V P G V G V P G V G V P G V G
gttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggg
V P G V G V P G F G V P G V G V P G V G
gtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggt
V P G V G V P G V G V P G F G V P G V G
gttccgggtgtcggggtgcccgggtgctaaagatctcatcaccatcaccatcactaagct
V P G V G V P G C -

Full plasmid sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACTATGAAAGGATCCCTGGAAATCCGTGCGGCGGCGCTGC
GTCGTCGTAACACCGCGCTGCGTACCCGTGTTGCGGAACTGCGTCAGCGT
GTTCAGCGTCTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCC

GCTGGGTGGTTCTGGTTCTGGTTCTGGTTCTGGTAGCTTCAGATCTCATCA
CCATCACCATCACTAAGCTTAATTAGCTGAGCCGGATAACAATTATAATA
GATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGA
GAAATTAACTATGAAAGGATCCCTGGAAATCGAAGCGGCGGCGCTGGAA
CAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAG
TTCAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGC
TGGGTGGTTCTGGTTCTGGTTCTGGTTCTGGTAGCGGATCCGGGGTGCCGG
GTGTGGGCGTTCCGGGCGTGGGTGTACCGGGCTTCGGTGTCCCGGGCGTA
GGTGTTCCGGGTGTCGGGGTGCCGGGTGTGGGCGTTCCGGGCGTGGGTGT
ACCGGGCTTCGGTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGCCGG
GTGTGGGCGTTCCGGGCGTGGGTGTACCGGGCTTCGGTGTCCCGGGCGTA
GGTGTTCCGGGTGTCGGGGTGCCGGGTGTGGGCGTTCCGGGCGTGGGTGT
ACCGGGCTTCGGTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGCCGG
GTGTGGGCGTTCCGGGCGTGGGTGTACCGGGCTTCGGTGTCCCGGGCGTA
GGTGTTCCGGGTGTCGGGGTGCCCGGGTGCTAAAGATCTCATCACCATCA
CCATCACTAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGT
AATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCG
GGCGTTTTTTATTGGTGAGAATCCAAGCTAGCAGTTGACAATTAATCATCG
GCTCGTATAATGGATCGAATTGTGAGCGGAATCGATTTTCACACAGGAAA
CAGACCATGGCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCAT
TAGCCAGGCGTCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATT
TGGGTAAAAAAGGGCACTTAACCCTTCAGATGACGACCCTGCGTGAGCTG
CCGCCAGAAGAGCGTCCGGCAGCTGGTGCGGTTATCAACGAAGCGAAAG
AGCAGGTTCAGCAGGCGCTGAATGCGCGTAAAGCGGAACTGGAAAGCGC
TGCACTGAATGCGCGTCTGGCGGCGGAAACGATTGATGTCTCTCTGCCAG
GTCGTCGCATTGAAAACGGCGGTCTGCATCCGGTTACCCGTACCATCGAC
CGTATCGAAAGTTTCTTCGGTGAGCTTGGCTTTACCGTGGCAACCGGGCCG
GAAATCGAAGACGATTATCATAACTTCGATGCTCTGAACATTCCTGGTCA
CCACCCGGCGCGCGCTGACCACGACACTTTCTGGTTTGACACTACCCGCCT
GCTGCGTACCCAGACCTCTGGCGTACAGATCCGCACCATGAAAGCCCAGC
AGCCACCGATTCGTATCATCGCGCCTGGCCGTGTTTATCGTAACGACTACG
ACCAGACTCACACGCCGATGTTCCATCAGATGGAAGGTCTGATTGTTGAT
ACCAACATCAGCTTTACCAACCTGAAAGGCACGCTGCACGACTTCCTGCG
TAACTTCTTTGAGGAAGATTTGCAGATTCGCTTCCGTCCTTCCTACTTCCC
GTTTACCGAACCTTCTGCAGAAGTGGACGTCATGGGTAAAAACGGTAAAT
GGCTGGAAGTGCTGGGCTGCGGGATGGTGCATCCGAACGTGTTGCGTAAC
GTTGGCATCGACCCGGAAGTTTACTCTGGTTTCGGCTTCGGGATGGGGAT
GGAGCGTCTGACTATGTTGCGTTACGGCGTCACCGACCTGCGTTCATTCTT
CGAAAACGATCTGCGTTTCCTCAAACAGTTTAAATAAGGCAGGAATAGAT
TATAAGCTTGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAAT
GGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATC
GTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACC
AGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAAT

AAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAAT
GCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATG
GGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTT
TTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACAT
ATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAA
AGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTT
CACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGT
TTTCACCATGCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATG
CCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGC
AGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGC
GTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGAC
TCTCTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGG
GCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACA
AATCCGCCGCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAA
CCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGG
ATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGG
GTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATA
CTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATA
TGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCT
GCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAA
GGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCC
GCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGA
AACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCT
CGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTT
TCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCT
CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG
ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
GCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCT
GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC
AAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
ATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTT
AATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGT
TGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCA
GATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTG
GTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG

CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTG
CTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAA
AAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA
TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAAT
ACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTG
GAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGA
TCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGC
AAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCC
TTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGAT
ACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACA
TTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGAC
ATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.2 pQE60_(ZEhis-ZRELF)

Submitted by Kechun Zhang **Date** 8 January 2007

Strain name AF-IQ

Vector pQE60_(ZEhis-ZRELF)

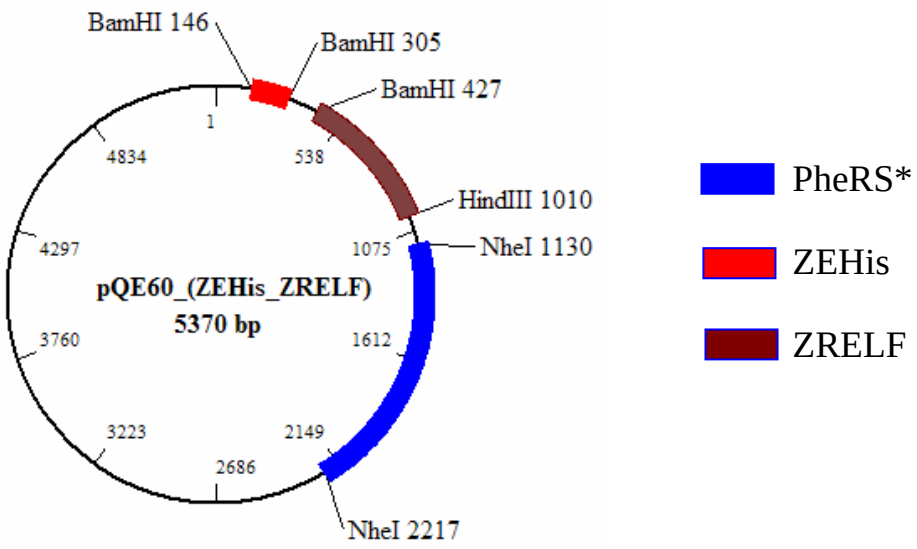
Construction

A *Nhe*I fragment encoding mutant *E. coli* PheRS (A294G) was inserted into expression plasmid pQE60 to yield pQE-FS*. Then the ZE gene was ligated into the *Bam*HI site of pQE-FS* to yield ZE-pQE-FS*. Finally, a DNA fragment encoding ZRELF and containing a ribosome binding site was amplified and ligated into the *Bpu*1102I site of ZE-pQE-FS*. The resulting plasmid (ZEhis-ZRELF) was transformed into phenylalanine auxotroph *E. coli* strain AF-IQ.

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

atgagaggatcgcatcaccatcaccatcacggatccctggaaatcgaagcggcggcgctg
M R G S H H H H H H G S L E I E A A A L
gaacaggaaaacaccgcgctggaaaccgaagttgcggaactggaacaggaagttcagcgt
E Q E N T A L E T E V A E L E Q E V Q R
ctggaaaacatcgtttctcagtaccgtacccgttacggtccgctgggtggtggtttcttc
L E N I V S Q Y R T R Y G P L G G G F F
ttcccgtaaggatccgtcgacctgcagccaagcttaattagctgagccggataacaatta
F P -

taatagattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaatta
actatgaaaggatccctggaaatccgtgcggcggcgctgcgtcgtcgtaacaccgcgctg
M K G S L E I R A A A L R R R N T A L
cgtacccgtgttgcggaactgcgtcagcgtgttcagcgtctgcgtaacgaagtttctcag
R T R V A E L R Q R V Q R L R N E V S Q
tacgaaacccgttacggtccgctgggtggtggcggttctggtggcggtggatccggggtg
Y E T R Y G P L G G G G S G G G G S G V
ccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggtgtt
P G V G V P G V G V P G F G V P G V G V
ccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtc
P G V G V P G V G V P G V G V P G F G V
ccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgttccgggcgtgggtgta
P G V G V P G V G V P G V G V P G V G V
ccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtgccgggtgtgggcgtt
P G F G V P G V G V P G V G V P G V G V
ccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggtgttccgggtgtcggggtg
P G V G V P G F G V P G V G V P G V G V
ccgggtgtgggcgttccgggcgtgggtgtaccgggcttcggtgtcccgggcgtaggtgtt
P G V G V P G V G V P G F G V P G V G V
ccgggtgtcggggtgcccgggtgctaaagatctcatcaccatcaccatcactaagcttaa
P G V G V P G C -

Full plasmid sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATT
CAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGAGGAGAAATTAACTA
TGAGAGGATCGCATCACCATCACCATCACGGATCCCTGGAAATCGAAGCGGCGGCGC
TGGAACAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGTT
CAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGCTGGGTGGTG
GTTTCTTCTTCCCGTAAGGATCCGTCGACCTGCAGCCAAGCTTAATTAGCTGAGCCGG
ATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTA
AAGAGGAGAAATTAACTATGAAAGGATCCCTGGAAATCCGTGCGGCGGCGCTGCGTC

GTCGTAACACCGCGCTGCGTACCCGTGTTGCGGAACTGCGTCAGCGTGTTCAGCGTCT
GCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCCGCTGGGTGGTGGCGGTTCT
GGTGGCGGTGGATCCGGGGTGCCGGGTGTGGGCGTTCCGGGCGTGGGTGTACCGGGC
TTCGGTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGCCGGGTGTGGGCGTTCCGG
GCGTGGGTGTACCGGGCTTCGGTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGC
CGGGTGTGGGCGTTCCGGGCGTGGGTGTACCGGGCTTCGGTGTCCCGGGCGTAGGTG
TTCCGGGTGTCGGGGTGCCGGGTGTGGGCGTTCCGGGCGTGGGTGTACCGGGCTTCG
GTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGCCGGGTGTGGGCGTTCCGGGCG
TGGGTGTACCGGGCTTCGGTGTCCCGGGCGTAGGTGTTCCGGGTGTCGGGGTGCCCG
GGTGCTAAAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGAGCTTGGACT
CCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTC
GGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCTAGCAGTTGACAATTAATC
ATCGGCTCGTATAATGGATCGAATTGTGAGCGGAATCGATTTTCACACAGGAAACAG
ACCATGGCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCAGGCG
TCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATTTGGGTAAAAAAGGGCAC
TTAACCCTTCAGATGACGACCCTGCGTGAGCTGCCGCCAGAAGAGCGTCCGGCAGCT
GGTGCGGTTATCAACGAAGCGAAAGAGCAGGTTCAGCAGGCGCTGAATGCGCGTAA
AGCGGAACTGGAAAGCGCTGCACTGAATGCGCGTCTGGCGGCGGAAACGATTGATGT
CTCTCTGCCAGGTCGTCGCATTGAAAACGGCGGTCTGCATCCGGTTACCCGTACCATC
GACCGTATCGAAAGTTTCTTCGGTGAGCTTGGCTTTACCGTGGCAACCGGGCCGGAA
ATCGAAGACGATTATCATAACTTCGATGCTCTGAACATTCCTGGTCACCACCCGGCGC
GCGCTGACCACGACACTTTCTGGTTTGACACTACCCGCCTGCTGCGTACCCAGACCTC
TGGCGTACAGATCCGCACCATGAAAGCCCAGCAGCCACCGATTCGTATCATCGCGCC
TGGCCGTGTTTATCGTAACGACTACGACCAGACTCACACGCCGATGTTCCATCAGATG
GAAGGTCTGATTGTTGATACCAACATCAGCTTTACCAACCTGAAAGGCACGCTGCAC
GACTTCCTGCGTAACTTCTTTGAGGAAGATTTGCAGATTCGCTTCCGTCCTTCCTACTT
CCCGTTTACCGAACCTTCTGCAGAAGTGGACGTCATGGGTAAAAACGGTAAATGGCT
GGAAGTGCTGGGCTGCGGGATGGTGCATCCGAACGTGTTGCGTAACGTTGGCATCGA
CCCGGAAGTTTACTCTGGTTTCGGCTTCGGGATGGGGATGGAGCGTCTGACTATGTTG
CGTTACGGCGTCACCGACCTGCGTTCATTCTTCGAAAACGATCTGCGTTTCCTCAAAC
AGTTTAAATAAGGCAGGAATAGATTATAAGCTTGCTAGCTTGGCGAGATTTTCAGGA
GCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCC
CAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATA
ACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGC
ACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGA
ATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGT
TACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACG
ACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAA
CCTGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCT
GGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCC
CGTTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGC
GATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAA
TTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATT

GGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGA
AAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTC
TCCTGAGTAGGACAAATCCGCCCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGT
GAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGAT
GCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCG
GCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGA
TGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGC
TGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATAC
GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAG
CAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTC
CGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCT
TTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG
CGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAA
GGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAA
AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCA
ATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAA
CCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTG
CGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGG
CTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTG
CAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGT
AAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATG
CGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGC
AGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGG
ATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATG
CCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTT
TTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCC
ACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATC
ACGAGGCCCTTTCGTCTTCAC

A.3 pQE60_ZE

Submitted by Kechun Zhang **Date** 8 January 2007

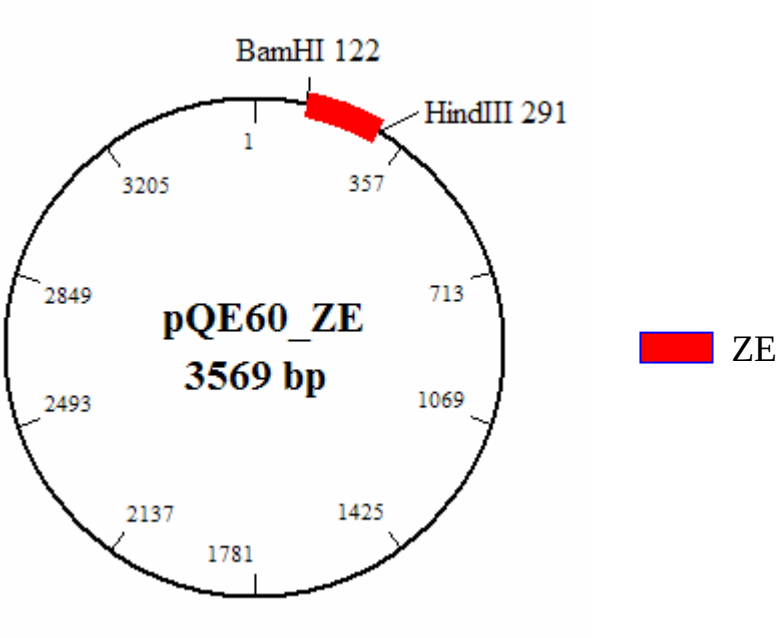
Strain name AF-IQ

Vector pQE60_ZE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

atgaaaggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcgctggaa
M K G S L E I E A A A L E Q E N T A L E
accgaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttctcagtac
T E V A E L E Q E V Q R L E N I V S Q Y
cgtacccgttacggtccgctgggtggtggtagatctcatcaccatcaccatcactaa
R T R Y G P L G G G R S H H H H H H -

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACtATGaaAggatccCTGGAAATCGAAGCGGCGGCGCTGGAAC
AGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGT
TCAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGCT
GGGTGGTGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTG
AGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGG
ATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATC
CAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGA
AAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAA
GAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACC
GTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCA
CAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCA
TCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATA
GTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCAT
CGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATT
CGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGG
TTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACC
AGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTC
ACCATGCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGC
TGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAA
TGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAA
TTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTC
TAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCT
TTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATC
CGCCGCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTC
TGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGT
CGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGG
CTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCG

GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATC
AAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC
CTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA
GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCG
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAA
GCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATG
CCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACAT
TAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.4 pQE60_QKZE

Submitted by Kechun Zhang **Date** 8 January 2007

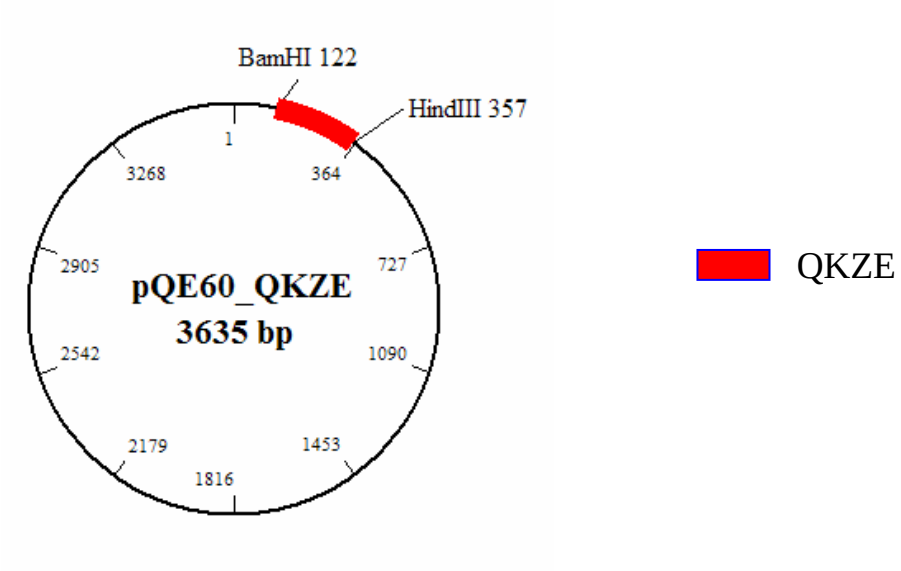
Strain name AF-IQ

Vector pQE60_QKZE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgaaaggatccaagctgacctggcaggagctgtaccagctgaagtataaaggcattggc
 M  K  G  S  K  L  T  W  Q  E  L  Y  Q  L  K  Y  K  G  I  G
ggaggcggcagcggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcg
 G  G  G  S  G  S  L  E  I  E  A  A  A  L  E  Q  E  N  T  A
ctggaaaccgaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttct
 L  E  T  E  V  A  E  L  E  Q  E  V  Q  R  L  E  N  I  V  S
cagtaccgtacccgttacggtccgctgggtggtggtagatctcatcaccatcaccatcac
 Q  Y  R  T  R  Y  G  P  L  G  G  G  R  S  H  H  H  H  H  H
taagcttaa
```

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACTATGAAAGGATCCAAGCTGACCTGGCAGGAGCTGTACC
AGCTGAAGTATAAAGGCATTGGCGGAGGCGGCAGCGGATCCCTGGAAAT
CGAAGCGGCGGCGCTGGAACAGGAAAACACCGCGCTGGAAACCGAAGTT
GCGGAACTGGAACAGGAAGTTCAGCGTCTGGAAAACATCGTTTCTCAGTA
CCGTACCCGTTACGGTCCGCTGGGTGGTGGTAGATCTCATCACCATCACCA
TCACTAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAAT
GACCTCAGAACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGG
CGTTTTTTATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCT
AAGGAAGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATA
TATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTC
AATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGA
CCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTG
CCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGT
GAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAG
CAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCG
GCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACC
TGGCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCA
ATCCCTGGGTGAGTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACA
ACTTCTTCGCCCCCGTTTTCACCATGCATGGGCAAATATTATACGCAAGGC
GACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGA
TGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATG
AGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGTGCCCTTAA
ACGCCTGGGGTAATGACTCTCTAGCTTGAGGCATCAAATAAAACGAAAGG
CTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACG
CTCTCCTGAGTAGGACAAATCCGCCGCTCTAGAGCTGCCTCGCGCGTTTCG
GTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACA
GCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTC
AGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATGACCCAGTCACGTAGCG
ATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTAC
TGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAG
AAAATACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCG
CTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAA
TACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGC
AAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCG
TTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCA
AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCC
CCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGG
ATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC
ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTG

TGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTA
TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAG
CCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGG
ATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAAT
TAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA
CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGG
CCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTA
ATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGC
AACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGT
ATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCC
CCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCA
TAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA
GTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGAT
CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGG
AAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGA
ATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATA
GGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTT
TCGTCTTCAC

A.5 pQE60_FNZE

Submitted by Kechun Zhang **Date** 8 January 2007

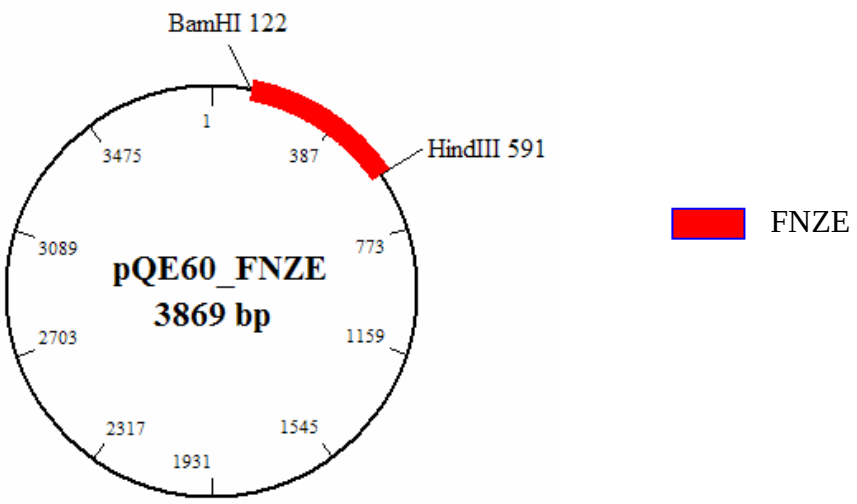
Strain name AF-IQ

Vector pQE60_FNZE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgaaaggatccaaaaaagttagcgatgttccacgtgatctggaagttgtggcggccacc
 M  K  G  S  K  K  V  S  D  V  P  R  D  L  E  V  V  A  A  T
ccgacgtctctgttaatcagctgggatgccccggcggtgacggttcgctattatcgtatc
 P  T  S  L  L  I  S  W  D  A  P  A  V  T  V  R  Y  Y  R  I
acctatggcgaaacgggcggcaactctccggtgcaagaatttacggtgccgggttctaag
 T  Y  G  E  T  G  G  N  S  P  V  Q  E  F  T  V  P  G  S  K
agcaccgcgaccatttctggcttaaaaccgggcgttgattacacgattaccgtgtatgcg
 S  T  A  T  I  S  G  L  K  P  G  V  D  Y  T  I  T  V  Y  A
gttaccccgcgtggcgattggaatgagggctctaaaccaatcagcatcaactatcgcacg
 V  T  P  R  G  D  W  N  E  G  S  K  P  I  S  I  N  Y  R  T
actagtggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcgctggaa
 T  S  G  S  L  E  I  E  A  A  A  L  E  Q  E  N  T  A  L  E
```

accgaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttctcagtac
T E V A E L E Q E V Q R L E N I V S Q Y
cgtacccgttacggtccgctgggtggtggtagatctcatcaccatcaccatcactaagct
R T R Y G P L G G G R S H H H H H H –

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACTATGAAAGGATCCAAAAAAGTTAGCGATGTTCCACGTG
ATCTGGAAGTTGTGGCGGCCACCCCGACGTCTCTGTTAATCAGCTGGGAT
GCCCCGGCGGTGACGGTTCGCTATTATCGTATCACCTATGGCGAAACGGG
CGGCAACTCTCCGGTGCAAGAATTTACGGTGCCGGGTTCTAAGAGCACCG
CGACCATTTCTGGCTTAAAACCGGGCGTTGATTACACGATTACCGTGTATG
CGGTTACCCCGCGTGGCGATTGGAATGAGGGCTCTAAACCAATCAGCATC
AACTATCGCACGACTAGTGGATCCCTGGAAATCGAAGCGGCGGCGCTGGA
ACAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAA
GTTCAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCG
CTGGGTGGTGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCT
GAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTG
GATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAAT
CCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGA
AAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAA
GAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACC
GTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCA
CAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCA
TCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATA
GTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCAT
CGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATT
CGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGG
TTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACC
AGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTC
ACCATGCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGC
TGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAA
TGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAA
TTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTC
TAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCT
TTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATC
CGCCGCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTC
TGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGT
CGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGG
CTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCG

GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATC
AAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC
CTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA
GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCG
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAA
GCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATG
CCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACAT
TAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.6 pQE60_ ZE-FN-ZE

Submitted by Kechun Zhang **Date** 8 January 2007

Strain name AF-IQ

Vector pQE60_ ZE-FN-ZE

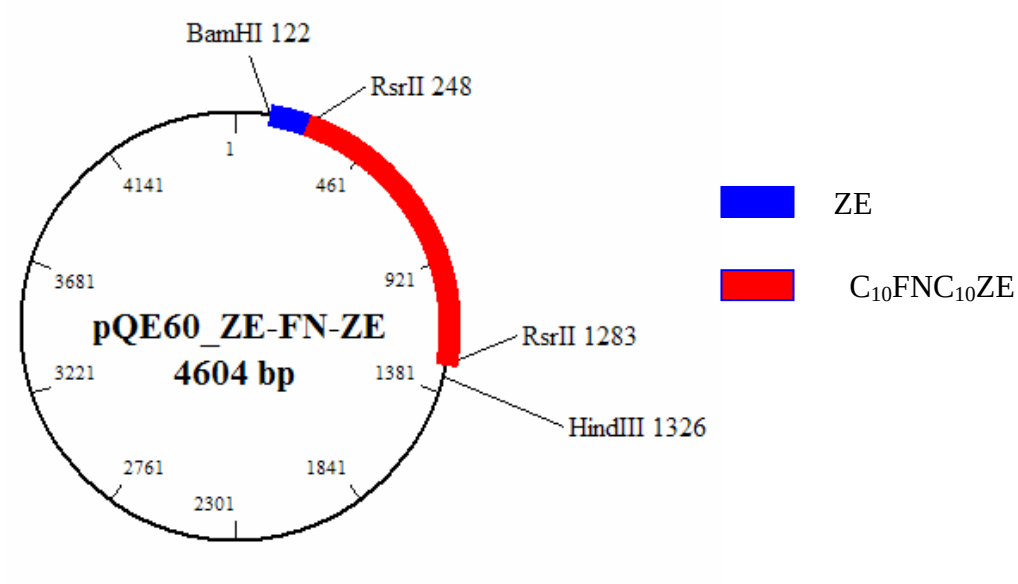
Construction

A previously constructed plasmid (pQE60_FNZE-FN-ZE) was digested with RsrII to obtain the $FNC_{10}ZE$ fragment. This fragment was ligated into pQE60_ZE to yield pQE60_ZE-FN-ZE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgaaaggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcgctggaa
 M  K  G  S  L  E  I  E  A  A  A  L  E  Q  E  N  T  A  L  E
accgaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttctcagtac
 T  E  V  A  E  L  E  Q  E  V  Q  R  L  E  N  I  V  S  Q  Y
cgtacccgttacggtccgctgtatcgcgagccgatgggtgccggcgctggtgcgggcccg
 R  T  R  Y  G  P  L  Y  R  E  P  M  G  A  G  A  G  A  G  P
```

gaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggt
E G A G A G A G P E G A G A G A G P E G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
A G A G P E G A G A G A G P E G A G A G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgtgatctggaagttgtg
P E G A R M P K K V S D V P R D L E V V
gcggccaccccgacgtctctgttaatcagctgggatgccccggcggtgacggttcgctat
A A T P T S L L I S W D A P A V T V R Y
tatcgtatcacctatggcgaaacgggcggcaactctccggtgcaagaatttacggtgccg
Y R I T Y G E T G G N S P V Q E F T V P
ggttctaagagcaccgcgaccatttctggcttaaaaccgggcgttgattacacgattacc
G S K S T A T I S G L K P G V D Y T I T
gtgtatgcggttaccccgcgtggcgattggaatgagggctctaaaccaatcagcatcaac
V Y A V T P R G D W N E G S K P I S I N
tatcgcacgactagctatcgcgatccgatgggtgccggcgctggtgcgggcccggaaggt
Y R T T S Y R D P M G A G A G A G P E G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
A G A G P E G A G A G A G P E G A G A G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcccgcatgccgactagtggatccctggaaatcgaagcggcggcgctggaacaggaa
G A R M P T S G S L E I E A A A L E Q E
aacaccgcgctggaaaccgaagttgcggaactggaacaggaagttcagcgtctggaaaac
N T A L E T E V A E L E Q E V Q R L E N
atcgtttctcagtaccgtacccgttacggtccgctgggtggtggtagatctcatcaccat
I V S Q Y R T R Y G P L G G G R S H H H
caccatcactaagcttaattagctgagcttggactcctgttgatagatccagtaatgacc
H H H -

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACTATGAAAGGATCCCTGGAAATCGAAGCGGCGGCGCTG
GAACAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGG
AAGTTCAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTC
CGCTGTATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGT

GCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGG
AAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGG
CCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGT
GCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCG
CTGGTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGC
CCGCATGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGG
CGGCCACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACG
GTTCGCTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGT
GCAAGAATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCT
TAAAACCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGT
GGCGATTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGAC
TAGCTATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTG
CAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGA
AGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGC
CCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTG
CGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGC
TGGTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCC
CGCATGCCGACTAGTGGATCCCTGGAAATCGAAGCGGCGGCGCTGGAACA
GGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGTT
CAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGCTG
GGTGGTGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGA
GCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGA
TTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCC
AAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAA
AAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAG
AACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCG
TTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCAC
AAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCAT
CCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAG
TGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATC
GCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTC
GCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGT
TTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCA
GTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCA
CCATGCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCT
GGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAAT
GCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAT
TTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCT
AGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTT
TCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCC
GCCGCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCT
GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCC
GGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTC

GGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGC
TTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGG
TGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTC
TTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCG
AGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAG
GGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAG
GAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
CTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCC
TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTC
AGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG
AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATC
AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCC
TGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGG
CCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCC
TGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAG
AGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTAC
AGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGG
TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAG
CGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGG
GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAA
ACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCA
GTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAA
AAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTT
TTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
TATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTT
CCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATT
AACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.7 pQE60_ ZE-FN-ZE-FN-ZE

Submitted by Kechun Zhang **Date** 8 January 2007

Strain name AF-IQ

Vector pQE60_ ZE-FN-ZE-FN-ZE

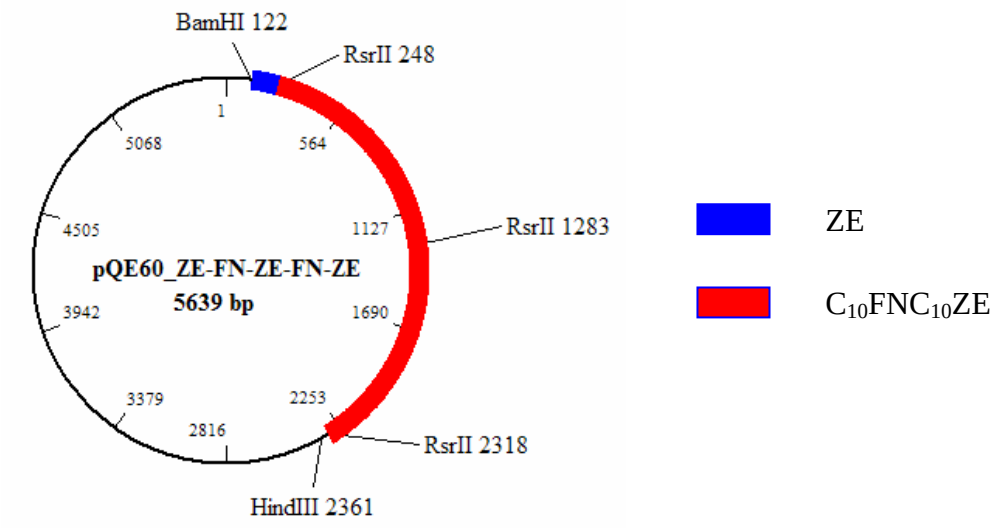
Construction

A previously constructed plasmid (pQE60_FNZE-FN-ZE) was digested with RsrII to obtain the $FNC_{10}ZE$ fragment. This fragment was multimerized by ligating monomers with T4 DNA ligase. The resulting dimer was ligated into pQE60_ZE to yield pQE60_ZE-FN-ZE-FN-ZE.

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

atgaaaggatccctggaaatcgaagcggcggcgctggaacaggaaaacaccgcgctggaa
M K G S L E I E A A A L E Q E N T A L E
accgaagttgcggaactggaacaggaagttcagcgtctggaaaacatcgtttctcagtac
T E V A E L E Q E V Q R L E N I V S Q Y
cgtacccgttacggtccgctgtatcgcgagccgatgggtgccggcgctggtgcgggcccg
R T R Y G P L Y R E P M G A G A G A G P
gaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggt
E G A G A G A G P E G A G A G A G P E G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
A G A G P E G A G A G A G P E G A G A G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgtgatctggaagttgtg
P E G A R M P K K V S D V P R D L E V V
gcggccaccccgacgtctctgttaatcagctgggatgccccggcggtgacggttcgctat
A A T P T S L L I S W D A P A V T V R Y
tatcgtatcacctatggcgaaacgggcggcaactctccggtgcaagaatttacggtgccg
Y R I T Y G E T G G N S P V Q E F T V P
ggttctaagagcaccgcgaccatttctggcttaaaaccgggcgttgattacacgattacc
G S K S T A T I S G L K P G V D Y T I T
gtgtatgcggttaccccgcgtggcgattggaatgagggctctaaaccaatcagcatcaac
V Y A V T P R G D W N E G S K P I S I N
tatcgcacgactagctatcgcgatccgatgggtgccggcgctggtgcgggcccggaaggt
Y R T T S Y R D P M G A G A G A G P E G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
A G A G P E G A G A G A G P E G A G A G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcccgcatgccgactagtggatccctggaaatcgaagcggcggcgctggaacaggaa
G A R M P T S G S L E I E A A A L E Q E
aacaccgcgctggaaaccgaagttgcggaactggaacaggaagttcagcgtctggaaaac
N T A L E T E V A E L E Q E V Q R L E N
atcgtttctcagtaccgtacccgttacggtccgctgtatcgcgagccgatgggtgccggc
I V S Q Y R T R Y G P L Y R E P M G A G
gctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggt
A G A G P E G A G A G A G P E G A G A G

gcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgca
G A G A G A G P E G A G A G A G P E G A
ggcgctggtgcgggcccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgt
G A G A G P E G A R M P K K V S D V P R
gatctggaagttgtggcggccaccccgacgtctctgttaatcagctgggatgccccggcg
D L E V V A A T P T S L L I S W D A P A
gtgacggttcgctattatcgtatcacctatggcgaaacgggcggcaactctccggtgcaa
V T V R Y Y R I T Y G E T G G N S P V Q
gaatttacggtgccgggttctaagagcaccgcgaccatttctggcttaaaaccgggcgtt
E F T V P G S K S T A T I S G L K P G V
gattacacgattaccgtgtatgcggttaccccgcgtggcgattggaatgagggctctaaa
D Y T I T V Y A V T P R G D W N E G S K
ccaatcagcatcaactatcgcacgactagctatcgcgatccgatgggtgccggcgctggt
P I S I N Y R T T S Y R D P M G A G A G
gcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgca
G A G A G A G P E G A G A G A G P E G A
ggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgct
G A G A G P E G A G A G A G P E G A G A
ggtgcgggcccggaaggtgcccgcatgccgactagtggatccctggaaatcgaagcggcg
G A G P E G A R M P T S G S L E I E A A
gcgctggaacaggaaaacaccgcgctggaaaccgaagttgcggaactggaacaggaagtt
A L E Q E N T A L E T E V A E L E Q E V
cagcgtctggaaaacatcgtttctcagtaccgtacccgttacggtccgctgggtggtggt
Q R L E N I V S Q Y R T R Y G P L G G G
agatctcatcaccatcaccatcactaagcttaattagctgagcttggactcctgttgata
R S H H H H H H –

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACtATGaaAGGATCCCTGGAAATCGAAGCGGCGGCGCTGGA
ACAGGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAA
GTTCAGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCG
CTGTATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGC
AGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAA

GGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCC
CGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGC
GGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCT
GGTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCC
GCATGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGGCG
GCCACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACGGT
TCGCTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGTGC
AAGAATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCTTA
AAACCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGTGG
CGATTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGACTA
GCTATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCA
GGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAG
GTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCC
GGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCG
GGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTG
GTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCG
CATGCCGACTAGTGGATCCCTGGAAATCGAAGCGGCGGCGCTGGAACAG
GAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGTTC
AGCGTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGCTGT
ATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGG
CGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGT
GCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGG
AAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGG
CCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGT
GCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCGCA
TGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGGCGGCC
ACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACGGTTCG
CTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGTGCAAG
AATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCTTAAAA
CCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGTGGCGA
TTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGACTAGCT
ATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGGC
GCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTG
CAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGA
AGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGC
CCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTG
CGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCGCAT
GCCGACTAGTGGATCCCTGGAAATCGAAGCGGCGGCGCTGGAACAGGAA
AACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGGAAGTTCAGC
GTCTGGAAAACATCGTTTCTCAGTACCGTACCCGTTACGGTCCGCTGGGTG
GTGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGAGCTTG
GACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGT
TCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCT

AGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAA
TCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACAT
TTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAG
CTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTT
TTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGA
ATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTC
ACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCT
GGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAA
GATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATT
GAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTT
GATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATG
CATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGA
TTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTA
ATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTT
TAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTT
GAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTT
TTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGC
TCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACAC
ATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAG
CAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC
GCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAAC
TATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGA
AATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACC
GTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG
AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
CCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG
GAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC

CCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAG
CAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAAC
TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTA
GCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTAT
CACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT
ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTC
TTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCA
GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTAT
AAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.8 pQE60_ RE_2RE -FN- RE_2RE

Submitted by Kechun Zhang **Date** 8 January 2007

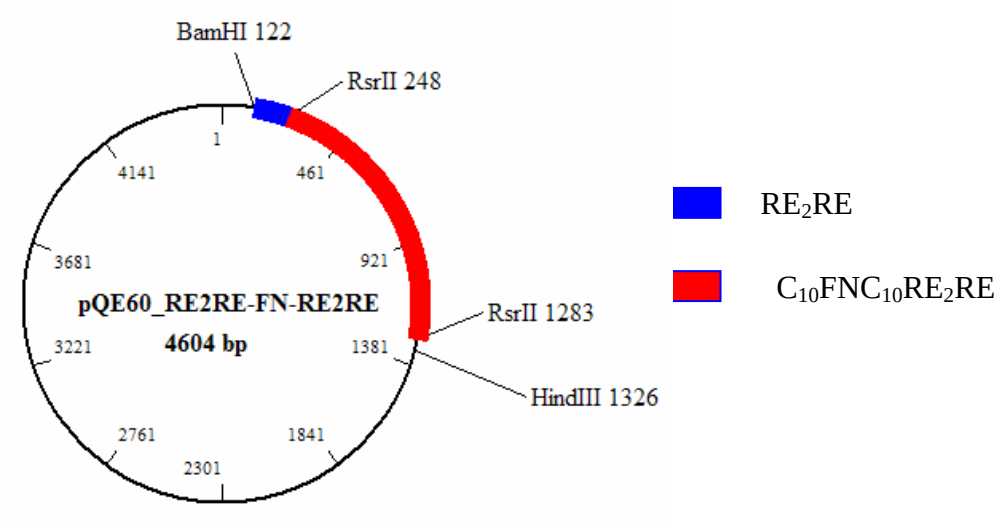
Strain name AF-IQ

Vector pQE60_ RE_2RE -FN- RE_2RE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgaaaggatccctggaaatccgtgcggcggcgctgcgtcgtgaaaacaccgcgctggaa
 M  K  G  S  L  E  I  R  A  A  A  L  R  R  E  N  T  A  L  E
accgaagttgcggaactggaacagcgtgttcagcgtctgcgtaacgaagtttctcagtac
 T  E  V  A  E  L  E  Q  R  V  Q  R  L  R  N  E  V  S  Q  Y
gaaacccgttacggtccgctgtatcgcgagccgatgggtgccggcgctggtgcgggcccg
 E  T  R  Y  G  P  L  Y  R  E  P  M  G  A  G  A  G  A  G  P
gaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggt
 E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
 A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
 A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G
```

gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgtgatctggaagttgtg
P E G A R M P K K V S D V P R D L E V V
gcggccaccccgacgtctctgttaatcagctgggatgccccggcggtgacggttcgctat
A A T P T S L L I S W D A P A V T V R Y
tatcgtatcacctatggcgaaacgggcggcaactctccggtgcaagaatttacggtgccg
Y R I T Y G E T G G N S P V Q E F T V P
ggttctaagagcaccgcgaccatttctggcttaaaaccgggcgttgattacacgattacc
G S K S T A T I S G L K P G V D Y T I T
gtgtatgcggttaccccgcgtggcgattggaatgagggctctaaaccaatcagcatcaac
V Y A V T P R G D W N E G S K P I S I N
tatcgcacgactagctatcgcgatccgatgggtgccggcgctggtgcgggcccggaaggt
Y R T T S Y R D P M G A G A G A G P E G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
A G A G P E G A G A G A G P E G A G A G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcccgcatgccgactagtggatccctggaaatccgtgcggcggcgctgcgtcgtgaa
G A R M P T S G S L E I R A A A L R R E
aacaccgcgctggaaaccgaagttgcggaactggaacagcgtgttcagcgtctgcgtaac
N T A L E T E V A E L E Q R V Q R L R N
gaagtttctcagtacgaaacccgttacggtccgctgggtggtggtagatctcatcaccat
E V S Q Y E T R Y G P L G G G R S H H H
caccatcactaagcttaattagctgagcttggactcctgttgatagatccagtaatgacc
H H H -

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACtATGaaAGGATCCCTGGAAATCCGTGCGGCGGCGCTGCG
TCGTGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGCGT
GTTCAGCGTCTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCC
GCTGTATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTG
CAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGA
AGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGC
CCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTG
CGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGC
TGGTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCC

CGCATGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGGC
GGCCACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACGG
TTCGCTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGTGC
AAGAATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCTTA
AAACCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGTGG
CGATTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGACTA
GCTATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCA
GGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAG
GTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCC
GGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCG
GGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTG
GTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCG
CATGCCGACTAGTGGATCCCTGGAAATCCGTGCGGCGGCGCTGCGTCGTG
AAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGCGTGTTCA
GCGTCTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCCGCTGG
GTGGTGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGAGC
TTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATT
TGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAA
GCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAA
AAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAA
CATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTT
CAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAA
GTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCG
GAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGT
TCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCT
CTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCA
AGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTAT
TGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTT
TGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCAT
GCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCG
ATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTT
AATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTT
TTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCT
TGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT
TTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCG
CTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACA
CATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGA
GCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGG
CGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAA
CTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTG
AAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCC
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGA

TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
CGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC
GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC
CTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGT
AGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTAC
GGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGT
TACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAA
AAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAG
TGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTA
AAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTG
AGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCC
AGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATC
AGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA
ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTA
AGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGC
ATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCC
CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTT
ATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATC
CGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGA
ATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATA
ATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGT
TCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCG
ATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGG
GAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTA
TAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.9 pQE60_ RE_2RE-FN-RE_2RE-FN-RE_2RE

Submitted by Kechun Zhang **Date** 8 January 2007

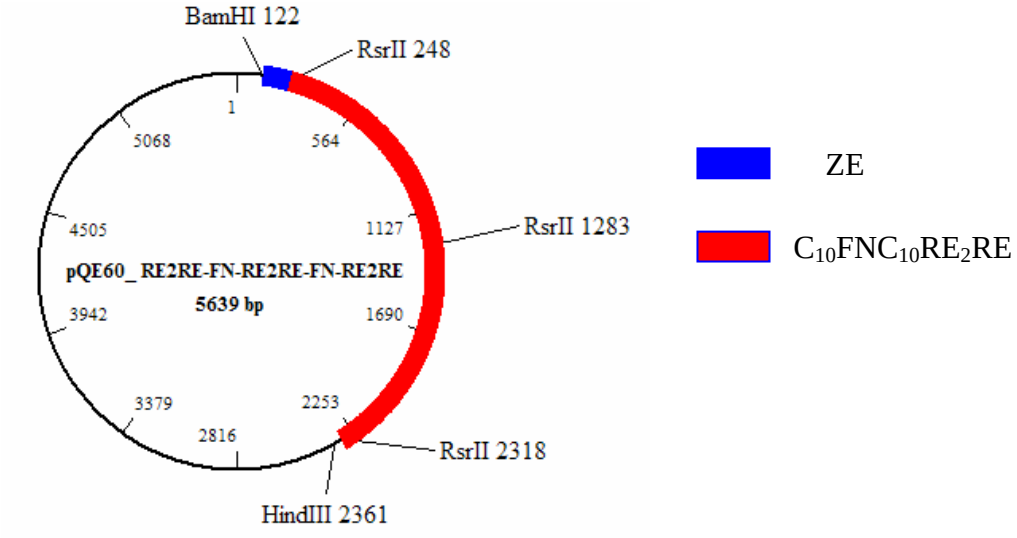
Strain name AF-IQ

Vector pQE60_ RE_2RE-FN-RE_2RE-FN-RE_2RE

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgaaaggatccctggaaatccgtgcggcggcgctgcgtcgtgaaaacaccgcgctggaa
 M  K  G  S  L  E  I  R  A  A  A  L  R  R  E  N  T  A  L  E
accgaagttgcggaactggaacagcgtgttcagcgtctgcgtaacgaagtttctcagtac
 T  E  V  A  E  L  E  Q  R  V  Q  R  L  R  N  E  V  S  Q  Y
gaaacccgttacggtccgctgtatcgcgagccgatgggtgccggcgctggtgcgggcccg
 E  T  R  Y  G  P  L  Y  R  E  P  M  G  A  G  A  G  A  G  P
gaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggt
 E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G
```

```
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
 A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
 A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
 A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G
ccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgtgatctggaagttgtg
 P  E  G  A  R  M  P  K  K  V  S  D  V  P  R  D  L  E  V  V
gcggccaccccgacgtctctgttaatcagctgggatgccccggcggtgacggttcgctat
 A  A  T  P  T  S  L  L  I  S  W  D  A  P  A  V  T  V  R  Y
tatcgtatcacctatggcgaaacgggcggcaactctccggtgcaagaatttacggtgccg
 Y  R  I  T  Y  G  E  T  G  G  N  S  P  V  Q  E  F  T  V  P
ggttctaagagcaccgcgaccatttctggcttaaaaccgggcgttgattacacgattacc
 G  S  K  S  T  A  T  I  S  G  L  K  P  G  V  D  Y  T  I  T
gtgtatgcggttaccccgcgtggcgattggaatgagggctctaaaccaatcagcatcaac
 V  Y  A  V  T  P  R  G  D  W  N  E  G  S  K  P  I  S  I  N
tatcgcacgactagctatcgcgatccgatgggtgccggcgctggtgcgggcccggaaggt
 Y  R  T  T  S  Y  R  D  P  M  G  A  G  A  G  A  G  P  E  G
gcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggc
 A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G
gctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggt
 A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G
gcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggc
 A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G
ccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaa
 P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E
ggtgcccgcatgccgactagtggatccctggaaatccgtgcggcggcgctgcgtcgtgaa
 G  A  R  M  P  T  S  G  S  L  E  I  R  A  A  A  L  R  R  E
aacaccgcgctggaaaccgaagttgcggaactggaacagcgtgttcagcgtctgcgtaac
 N  T  A  L  E  T  E  V  A  E  L  E  Q  R  V  Q  R  L  R  N
gaagtttctcagtacgaaacccgttacggtccgctgtatcgcgagccgatgggtgccggc
 E  V  S  Q  Y  E  T  R  Y  G  P  L  Y  R  E  P  M  G  A  G
gctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggt
 A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G
gcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggc
 A  G  P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G
ccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaa
 P  E  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E
ggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgca
 G  A  G  A  G  A  G  P  E  G  A  G  A  G  A  G  P  E  G  A
ggcgctggtgcgggcccggaaggtgcccgcatgccgaaaaaagttagcgatgttccacgt
 G  A  G  A  G  P  E  G  A  R  M  P  K  K  V  S  D  V  P  R
gatctggaagttgtggcggccaccccgacgtctctgttaatcagctgggatgccccggcg
```

D L E V V A A T P T S L L I S W D A P A
gtgacggttcgctattatcgtatcacctatggcgaaacgggcggcaactctccggtgcaa
V T V R Y Y R I T Y G E T G G N S P V Q
gaatttacggtgccgggttctaagagcaccgcgaccatttctggcttaaaaccgggcgtt
E F T V P G S K S T A T I S G L K P G V
gattacacgattaccgtgtatgcggttaccccgcgtggcgattggaatgagggctctaaa
D Y T I T V Y A V T P R G D W N E G S K
ccaatcagcatcaactatcgcacgactagctatcgcgatccgatgggtgccggcgctggt
P I S I N Y R T T S Y R D P M G A G A G
gcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggc
A G P E G A G A G A G P E G A G A G A G
ccggaaggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaa
P E G A G A G A G P E G A G A G A G P E
ggtgcaggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgca
G A G A G A G P E G A G A G A G P E G A
ggcgctggtgcgggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgct
G A G A G P E G A G A G A G P E G A G A
ggtgcgggcccggaaggtgcccgcatgccgactagtggatccctggaaatccgtgcggcg
G A G P E G A R M P T S G S L E I R A A
gcgctgcgtcgtgaaaacaccgcgctggaaaccgaagttgcggaactggaacagcgtgtt
A L R R E N T A L E T E V A E L E Q R V
cagcgtctgcgtaacgaagtttctcagtacgaaacccgttacggtccgctgggtggtggt
Q R L R N E V S Q Y E T R Y G P L G G G
agatctcatcaccatcaccatcactaagcttaattagctgagcttggactcctgttgata
R S H H H H H H –

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACtATGaaAGGATCCCTGGAAATCCGTGCGGCGGCGCTGCG
TCGTGAAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGCGT
GTTCAGCGTCTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCC
GCTGTATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTG
CAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGA
AGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGC
CCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTG
CGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGC
TGGTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCC
CGCATGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGGC
GGCCACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACGG
TTCGCTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGTGC
AAGAATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCTTA
AAACCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGTGG

CGATTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGACTA
GCTATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCA
GGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAG
GTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCC
GGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCG
GGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTG
GTGCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCG
CATGCCGACTAGTGGATCCCTGGAAATCCGTGCGGCGGCGCTGCGTCGTG
AAAACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGCGTGTTCA
GCGTCTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCCGCTGT
ATCGCGAGCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGG
CGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGT
GCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGG
AAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGG
CCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGT
GCGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCGCA
TGCCGAAAAAAGTTAGCGATGTTCCACGTGATCTGGAAGTTGTGGCGGCC
ACCCCGACGTCTCTGTTAATCAGCTGGGATGCCCCGGCGGTGACGGTTCG
CTATTATCGTATCACCTATGGCGAAACGGGCGGCAACTCTCCGGTGCAAG
AATTTACGGTGCCGGGTTCTAAGAGCACCGCGACCATTTCTGGCTTAAAA
CCGGGCGTTGATTACACGATTACCGTGTATGCGGTTACCCCGCGTGGCGA
TTGGAATGAGGGCTCTAAACCAATCAGCATCAACTATCGCACGACTAGCT
ATCGCGATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGGC
GCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTG
CAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGA
AGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGC
CCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTG
CGGGCCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCGCAT
GCCGACTAGTGGATCCCTGGAAATCCGTGCGGCGGCGCTGCGTCGTGAAA
ACACCGCGCTGGAAACCGAAGTTGCGGAACTGGAACAGCGTGTTCAGCGT
CTGCGTAACGAAGTTTCTCAGTACGAAACCCGTTACGGTCCGCTGGGTGG
TGGTAGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGAGCTTGG
ACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTCCATCTGGATTTGTT
CAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCT
AGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAA
TCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACAT
TTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAG
CTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTT
TTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGA
ATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTC
ACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCT
GGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAA
GATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTAAAGGGTTTATT

GAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGTTTT
GATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATG
CATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGA
TTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTA
ATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTT
TAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCTCTAGCTT
GAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTT
TTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGC
TCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACAC
ATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAG
CAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC
GCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAAC
TATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGA
AATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGG
TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACC
GTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG
AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTA
CCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCT
GGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAA
AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTG
GAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGA
TCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA
GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC
CCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAG
CAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAAC
TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAG
TAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT
CGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTA
GCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTAT
CACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG

TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAT
ACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTC
TTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCA
GCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA
ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTAT
AAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC

A.10 pQE9_ $PC_{10}A$

Submitted by Kechun Zhang **Date** 8 January 2007

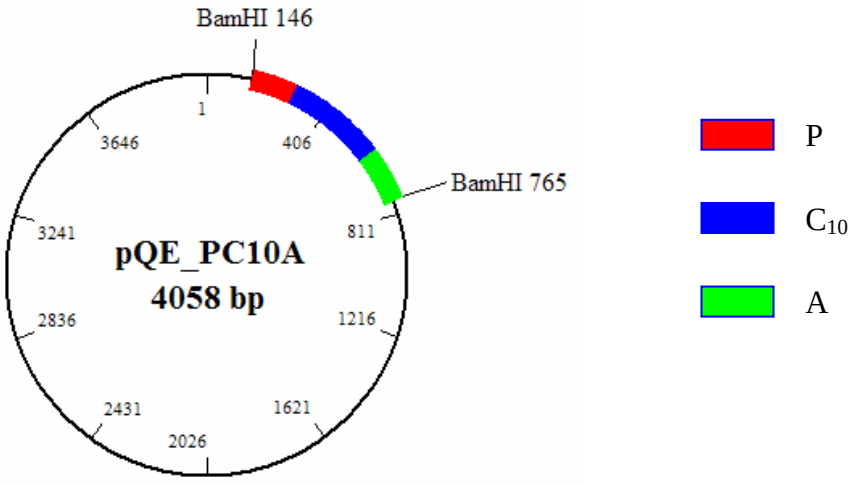
Strain name AF-IQ

Vector pQE9_ $PC_{10}A$

Available Sources:

1. 20% glycerol culture in Kechun Zhang's -80 °C freezer box
   Ampicilin and chloramphenicol resistant.
2. Plasmid DNA in Kechun Zhang's -20 °C freezer box

Plamsid map



Gene sequence

```
atgagaggatcgcatcaccatcaccatcacggatccggtgacctggcgccgcagatgctg
 M  R  G  S  H  H  H  H  H  H  G  S  G  D  L  A  P  Q  M  L
cgtgaactgcaggaaaccaacgcggcgctgcaggacgttcgtgaactgctgcgtcagcag
 R  E  L  Q  E  T  N  A  A  L  Q  D  V  R  E  L  L  R  Q  Q
gttaaagaaatcaccttcctgaaaaacaccgttatggaatctgacgcgtctggtactagc
 V  K  E  I  T  F  L  K  N  T  V  M  E  S  D  A  S  G  T  S
tatcgcgatccgatgggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcg
 Y  R  D  P  M  G  A  G  A  G  A  G  P  E  G  A  G  A  G  A
```

ggcccggaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccg
G P E G A G A G A G P E G A G A G A G P
gaaggtgccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaaggt
E G A G A G A G P E G A G A G A G P E G
gccggcgctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgccggc
A G A G A G P E G A G A G A G P E G A G
gctggtgcgggcccggaaggtgcaggcgctggtgcgggcccggaaggtgcccgcatgccg
A G A G P E G A G A G A G P E G A R M P
actagcggtgacctggaaaacgaagtggcccagctggaaagggaagttagatctctggaa
T S G D L E N E V A Q L E R E V R S L E
gatgaagcggctgaactggaacaaaaagtctcgagactgaaaaatgaaatcgaagacctg
D E A A E L E Q K V S R L K N E I E D L
aaagccgaaattggtgaccatgtggcgcctcgagacactagttggtaaggatccgtcgac
K A E I G D H V A P R D T S W -

Full sequence

CTCGAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATA
ATAGATTCAATTGTGAGCGGATAACAATTTCACACAGAATTCATTAAAGA
GGAGAAATTAACTATGAGAGGATCGCATCACCATCACCATCACGGATCCG
GTGACCTGGCGCCGCAGATGCTGCGTGAACTGCAGGAAACCAACGCGGC
GCTGCAGGACGTTCGTGAACTGCTGCGTCAGCAGGTTAAAGAAATCACCT
TCCTGAAAAACACCGTTATGGAATCTGACGCGTCTGGTACTAGCTATCGC
GATCCGATGGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGGCGCTG
GTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGTGCAGG
CGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGGAAGGT
GCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGGCCCGG
AAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCGGCGCTGGTGCGGG
CCCGGAAGGTGCAGGCGCTGGTGCGGGCCCGGAAGGTGCCCGCATGCCG
ACTAGCGGTGACCTGGAAAACGAAGTGGCCCAGCTGGAAAGGGAAGTTA
GATCTCTGGAAGATGAAGCGGCTGAACTGGAACAAAAAGTCTCGAGACT
GAAAAATGAAATCGAAGACCTGAAAGCCGAAATTGGTGACCATGTGGCG
CCTCGAGACACTAGTTGGTAAGgGATCCGTCGACCTGCAGCCAAGCTTAA
TTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGAACTC
CATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGT
GAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGAAGCTAAAA
TGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCAT
CGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAAC
CAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAA
TAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAA
TGCTCATCCGGAATTTCGTATGGCAATGAAAGACGGTGAGCTGGTGATAT
GGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGT
TTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACACA
TATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCTA

AAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTT
TCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCG
TTTTCACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCG
CTGGCGATTCAGGTTCATCATGCCGTTTGTGATGGCTTCCATGTCGGCAGA
ATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTA
ATTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGGGTAATGACTCT
CTAGCTTGAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC
CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAA
TCCGCCCTCTAGAGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTC
TGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGT
CGGGGCGCAGCCATGACCCAGTCACGTAGCGATAGCGGAGTGTATACTGG
CTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCG
GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGC
TCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGG
CGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC
AGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCT
AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCA
GAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATC
AAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAAT
CAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC
CTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTG
GCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTC
CTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA
GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCG
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAA
GCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCA
GTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATG

CCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTC
TGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACG
GGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAA
AACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCC
AGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTT
TTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACAT
TAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAC