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.

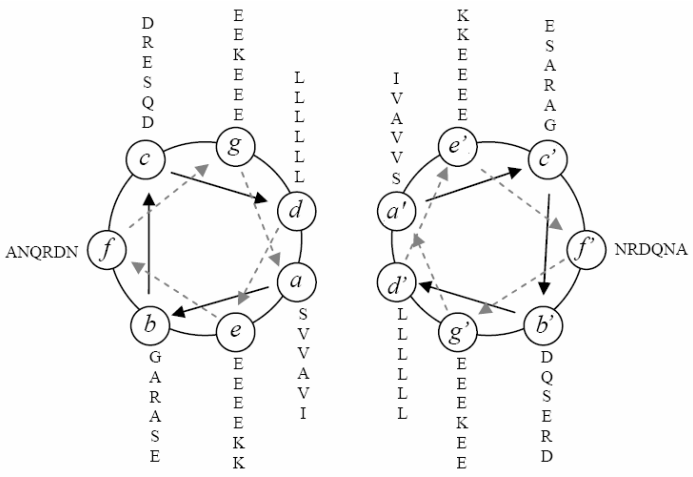

Figure 1.1 Helical wheel representation of the parallel A1 homodimer.

(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]

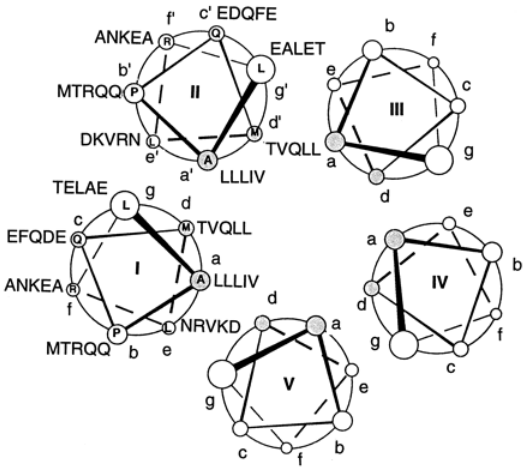

Figure 1.2 Helical wheel representation of the COMPcc pentamer.

(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

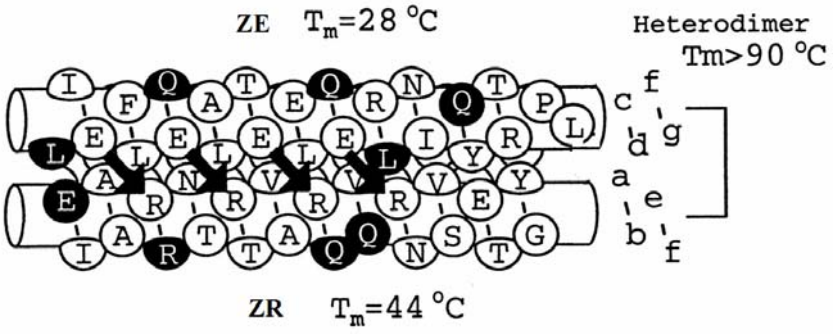

Figure 1.3 Schematic of the ZE/ZR heterodimer.

(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.