Recombinant production of periodic polypeptides

Introducing structural order
in polymeric materials

Jurgen M. Smeenk

Recombinant production of periodic polypeptides

Introducing structural order in polymeric materials

Een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen,
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen

in het openbaar te verdedigen op dinsdag 21 november 2006,
des namiddags om 3.30 uur precies

door

Jurgen Marinus Smeenk

geboren op 4 april 1976
te Winterswijk

Table of contents

Chapter 1

General introduction

1.1 Introduction

For many years polymer chemists have been developing new synthetic routes for the preparation of better defined polymers. Although considerable progress has been made, especially with the development of controlled polymerization techniques[1], the level of control present in biological polymers remains unsurpassed. In nature, macromolecules, like nucleic acids and proteins fulfill a large variety of functions, including information storage and transfer, catalysis and tissue support. The fidelity of the synthesis of these macromolecules is therefore of utmost importance and receives special attention of the cellular machinery. From a polymer chemistry point of view three aspects of biological macromolecules are remarkable: Firstly, they are monodisperse, i.e. they are of a specific molecular weight in contrast to synthetic polymers, which have a molecular weight distribution. Secondly, biological polymers are synthesized with absolute control over monomer sequence. Thirdly, natural polymers can attain well-defined three-dimensional structures via multiple levels of organization as a result of the interplay of a variety of non-covalent interactions, like steric, hydrophobic, electrostatic and hydrogen bonding interactions.

The importance of sequence information and non-covalent interactions is evident for proteins, where a multitude of structurally and functionally different types of proteins are made with a set of 20 amino acid building blocks. Some proteins evolved as enzymes catalyzing a wide variety of reactions, whereas others acquired regulatory or structural functions. The amino acid sequence affects the possible spatial arrangement of the polypeptide chain and can result in the regular hydrogen bond stabilized α-helical and β-sheet secondary structures. In the next level of organization, the three-dimensional arrangement of the secondary structural elements, stabilization primarily occurs via hydrophobic and electrostatic interactions. Furthermore, proteins can associate into macromolecular assemblies.

These organizational principles have been an inspiration source for materials scientists interested in designing precise nano- and mesoscopic structures using chemically synthesized compounds. In this area of research, termed "supramolecular chemistry"[2] components are organized into patterns or structures with the use of multiple non-covalent interactions. This process where components organize without human intervention is often termed "self-assembly". The interest in self-assembly arises among other reasons from the wish to make materials with ever-smaller feature sizes for use in nanotechnology.

The interest of material scientists in proteinaceous materials results from the wish to understand how assembly at different scales contributes to the macroscopic, mechanical properties of materials. In this respect, fibrous structural proteins, such as collagens, silks and elastins have drawn much attention. Although composed of the same amino acid building blocks, their distinct organization gives these materials a variety of properties, like rigidity, toughness or elasticity. Better understanding of how their organization translates into final material properties may lead to the development of new kinds of high performance materials.

The synthesis of high molecular weight polypeptides with control over amino acid sequence via chemical routes is still a troublesome procedure (see Chapter 2). A valuable alternative for their synthesis is therefore the use of the biosynthetic pathway of (micro-)organisms. This approach has been followed for many years for the production or modification of globular proteins, resulting in the development of a whole range of techniques for gene construction, protein expression and purification. However, only since the last decade, protein engineering has been used extensively as a tool for the preparation of structural proteins. One type of structural protein that has received much attention is silk. Since this material is the source of inspiration of the research described in this thesis, the characteristics of this class of structural proteins, as well as their application in materials science will be discussed in more detail in the next paragraphs.

1.2 Silk worm cocoon silk and spider dragline silk

Silks are fibrous structural proteins that are produced by more than 30.000 known species of spiders and by over 100.000 known species of insects.[3] Of all those silks, two types of silk have been studied most extensively, namely the cocoon silk of the domesticated silk worm *Bombyx mori* and spider dragline silk from the golden orb weaver *Nephila clavipes* and the garden cross spider *Araneus diadematus*. The interest in silk worm silk is, of course, the result of its large-scale production for use as high-quality textile fiber. Dragline silk, produced in the major ampullate gland and forming the web frame, has attracted attention because it is stronger than other biomaterials and tougher (energy to break) than engineered materials, like Kevlar and high-tensile steel.[4]

Silkworm silk has been best characterized of all silks. The fiber consists of a heavy (H) and a light (L) chain polypeptide (molecular weights of ~ 390 kDa and 26 kDa, respectively) which are linked via their C-terminus by a disulfide bridge.[5-8] In addition, another 25 kDa polypeptide associates with the H-L complex primarily by hydrophobic interactions.[5] The structurally important H-chain has been shown to consist of a repetitive motif of alternating amorphous and crystalline domains.[9-13] Only recently, a complete amino acid sequence was reported as deduced from genomic DNA.[14, 15] However, some variability in molecular weight and composition of the H-chain exists between *Bombyx mori* stocks.[11, 12] The 12 crystalline domains contain multiple repeats of the hexapeptide GAGAGS (G = glycine, A = alanine and S = serine), separated by a more variable repeating motif. The strict G-X alternation (X = A in 65%, S in 23% and Y (tyrosine) in 9%) is occasionally interrupted by the GAAS tetrapeptide sequence. A typical sequence of a crystalline domain is depicted in Figure 1.1a. The number of tandem repeats (n) within these domains is variable. The crystalline domains are connected by 11 conserved, non-repetitive linker sequences of approximately 30 residues, containing a proline, a tryptophan, and charged residues (these residues are completely absent from the crystalline domains).

a) **Bombyx mori heavy chain fibroin**

- $\{(GAGAGS)_{n=5-8} - [(GA)_{n=1-3} - GY]_{n=4-7} - GAGAGS - GAAS\}_{n=2-8}$ -
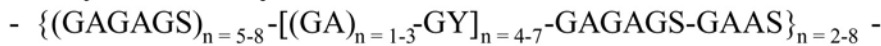
b) **Nephila clavipes major ampullate spidroins**

MaSp-1: - AGRGGLGGQGAG-AAAAAAA-GGAGQGGYGGLGSQG -
MaSp-2: - GPGQQGPGGYGPGQQGPSGPGS-AAAAAAAAA-GPGGYGPGQQGPGGY -

Figure 1.1 *Amino acid sequence motifs for (a) a crystalline domain of Bombyx mori heavy chain fibroin and (b) Nephila clavipes dragline silk fibroins (G = glycine, A = alanine, S = serine, Y = tyrosine, R = arginine, L = leucine, Q = glutamine and P = proline).*

The crystallinity of silk was already known in 1913[16] via X-ray diffraction and has later been assigned to the regular G-X repeats in the fibroin.[17, 18] The diffraction pattern is characteristic of an antiparallel pleated β-sheet structure.[19-21] Silkworm silk can thus be considered as a semicrystalline material made of amorphous flexible chains reinforced by strong and stiff β-sheet crystals (Figure 1.2). The β-sheet crystals (with a size of $6 \times 2 \times 21$ nm) can be considered as non-covalent crosslinks and are oriented with the polypeptide chains lying parallel to the fiber axis.[22]
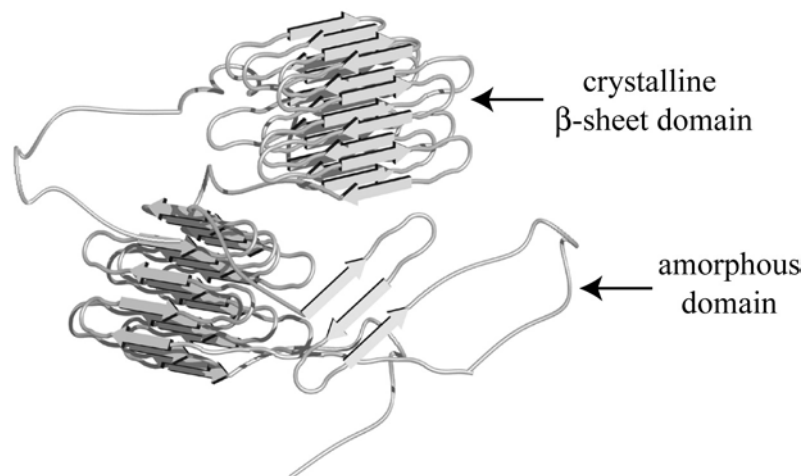
Figure 1.2 *Schematic representation of the silk structure. The crystalline β-sheet domains are embedded in an amorphous matrix, which allows the crystalline domains to orient themselves under strain, and makes the fibers flexible.*

In comparison to silkworm silk, limited information on the composition of spider dragline silk is available. Dragline silks are considered to be composed of two major proteins.[23] Lewis et al. reported the presence of two proteins, termed major ampullate spidroin 1 and 2 (MaSp-1 and MaSp-2), in the dragline silk from *Nephila clavipes*.[24, 25] Based on cloned cDNA sequences containing the C-terminal part of these silk fibroins their repetitive character became apparent. However, no complete sequence data are available due to the repetitiveness and high molecular

masses of these fibroins. Apparent molecular masses between 180 and 720 kDa have been reported depending on the applied analytical technique.[26-28] Repeating sequence motifs of approximately 34 and 45 amino acid residues are found for MaSp-1 and MaSp-2, respectively. Within these repeat motifs, short poly(alanine) blocks of 7-10 residues in length are present (Figure 1.1b). These poly(alanine) stretches have been found to form crystalline β-sheet stacks, giving the material its strength.[29-33] The β-sheet crystals (2 × 5 × 6 nm) are dispersed in a less well-defined glycine-rich matrix, which provides elasticity to the fiber.[34] Two motifs can be recognized in the glycine-rich sequence, namely GPGXX in MaSp-2 (with X often representing glutamine (Q)) and GGX in MaSp-1. The first motif has been suggested to form a β-turn spiral, and most likely provides elasticity, based on the predominance of this motif in the more elastic, flagelliform silk and the presence of comparable proline-containing motifs in other elastic proteins, such as elastin and gluten.[35-38] The GGX motif has been proposed to form a 3_1-helix.[39] Similar repetitive motifs were found in major ampullate fibroins of *Araneus diadematus* and other spiders.[23, 40] Furthermore, the conserved, non-repetitive carboxyl-terminal sequence[40] contains a cysteine residue that may participate in interfibroin disulfide cross-linking.[28]

1.3 Recombinant synthesis of silk proteins

The production of silk by genetically modified organisms has been focused on spider dragline silk, which is tougher than the brittle cocoon silk of *Bombyx mori*. The isolation of large amounts of spider silk from natural sources is not feasible, due to the territorial nature of spiders. Spinning of spider silk occurs at ambient temperatures and pressures using water as solvent and yields a biodegradable fiber. Therefore it is considered as a "green" alternative to traditional synthetic fibers.[41] Furthermore, recombinant fibroin synthesis allows variation of silk primary sequences and thus enables the production of fibers with a wide range of properties.[42] An anticipated application of silk fibers is, for example, as fine monofilament sutures in microsurgery.[43]

Factors that complicate the recombinant production of spider silk fibroins are the length and repetitiveness of the genes and their high *gc*-content. Only partial cDNA clones of dragline silk genes have been obtained and have been used for the production of recombinant silk in a number of host organisms. Arcidiacono et al. reported the low level expression (4 mg L^{-1}) in *E. coli* of a partial cDNA clone from the 3'-end of the *Nephila clavipes* spidroin 1 gene resulting in a 43 kDa recombinant silk protein.[44] More recently, mammalian cells (bovine and hamster)[43], and insect cell lines (derived from the fall armyworm *Spodoptera frugiperda*) using the baculovirus expression system[45], were used for the production of recombinant silk fibroins. The company Nexia Biotechnologies even investigated the use of transgenic goats that secrete silk protein in their milk.[46]

Synthetic genes have been engineered to optimize codon usage for the applied expression host and to minimize repetitiveness of the silk genes. Building up genes from small

oligonucleotide repeats allows control over gene and protein sequence and the final protein size. Synthetic genes have been used mainly in *E. coli*[47-52], but also in the yeast *Pichia pastoris*[53] and plants like tobacco and potato[54, 55] and silk fibroins with a molecular weight of up to 150 kDa have been produced. Fahnestock and coworkers were the first to express synthetic analogues of both *Nephila clavipes* dragline silk fibroins in *E. coli*.[47] To minimize repetitiveness they used long DNA monomers (303 bp), encoding four variant repeats of the natural protein. The genes consisting of 8 or 16 monomer repeats were quite stable (less than a few percent of deletion observed after 400 doublings) and expressed well (300 mg L^{-1}). However, truncated protein products were formed predominantly due to premature termination of translation. The fraction of full length protein was determined to be approximately 50% for the 8-mer gene and decreased to 9% for the 16-mer gene. This limitation was overcome by using the methylotrophic yeast *Pichia pastoris* as a production host. By expressing the genes under control of the methanol-inducible AOX1 promoter, yields of 1g L^{-1} were obtained and longer products up to 2000 amino acid residues could be synthesized without evidence of truncated synthesis.[53]

1.4 Silk assembly

The production of recombinant silk proteins similar to the natural counterparts is, by itself, not sufficient to produce fibers with good mechanical properties. Another critical factor is the artificial spinning of these silks. Several physiological and mechanical parameters play an important role in natural silk spinning. For example, the spinneret involved in the production of dragline silk fibers by *Nephila clavipes* has been well studied.[56] The silk fibroins are stored at high concentrations (up to 50% w/v) in the silk gland without apparent aggregation.[57] This "spinning dope" is liquid crystalline, meaning that neighbouring molecules are aligned approximately parallel to one another. This allows extrusion of the fiber using minimal forces.[58] Thread assembly occurs during passage of the silk dope through the spinning duct and is accompanied by extraction of water, sodium and chloride ions from the lumen.[59, 60] At the same time potassium ions and protons are secreted into the lumen which results in lowering of the pH from 6.9 to 6.3. It is thought that this triggers partial unfolding of the proteins and the transition of the polyalanine stretches from an α-helical to a β-sheet conformation. This results in interchain hydrogen bonds, and thus in physical cross-linking.[61-63] The presence of these cross-linked domains results in the remarkable mechanical properties of spider dragline silk. Mechanical tests show that it is among the strongest biomaterials known and is tougher (energy absorption before breakage) than engineering materials, like Kevlar (Figure 1.3).

Recombinant spider silk proteins have been spun from the strongly denaturing solvent hexafluoroisopropanol or from dilute solutions in concentrated formic acid using methanol as coagulant [64-66], but the properties of the natural silks have not been reproduced. A procedure more similar to the natural spinning process is the one described by Lazaris and coworkers.[43] Silk fibroin from *Araneus diadematus* expressed in mammalian cells was spun into fibers from

an aqueous spinning dope (>23% w/v of protein) by coagulation in 70-80% methanol. The fibers were subjected to postspinning draw, where higher draw ratios resulted in higher strength, toughness and stiffness values. However, the strength values were typically 5-fold lower than for dragline silk, but similar to those measured for fibers spun from regenerated spider silk.[67] The toughness and stiffness values were comparable to native dragline silk.
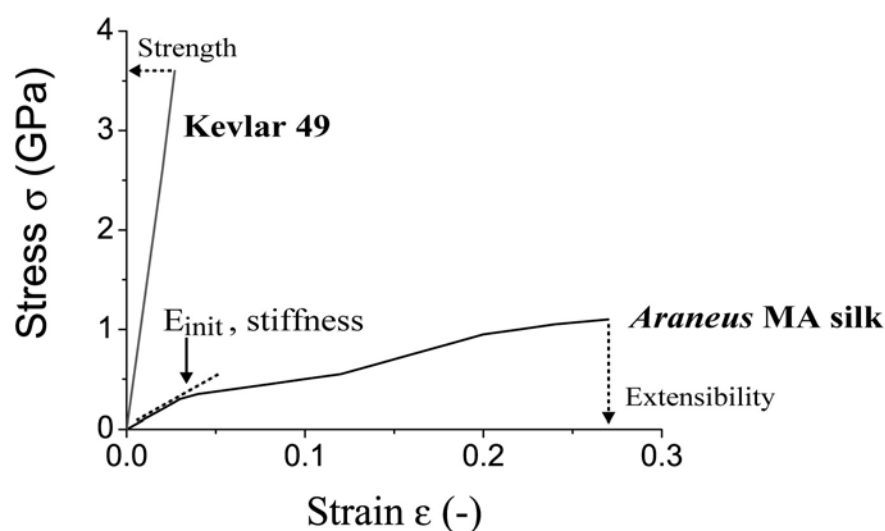
Material	Stiffness, E_{init} (GPa)	Strength, σ_{max} (GPa)	Extensibility, ε_{max} (-)	Toughness (MJ m^{-3})
Araneus MA silk	10	1.1	0.27	160
Bone	20	0.16	0.03	4
Elastin	0.001	0.002	1.5	2
Kevlar 49 fibre	130	3.6	0.027	50
High tensile steel	200	1.5	0.008	6

Figure 1.3 *Schematic stress-strain curves for major ampullate gland (MA) silk and Kevlar 49 showing the characteristic parameters for tensile tests. Some values for biological and synthetic materials are given. The stress is the normalized force (force divided by cross-sectional fiber area) and the strain is the normalized deformation (change in fibril length divided by initial length). The toughness represents the area under the stress-strain curve. The values are from reference [4].*

1.5 Artificial protein polymers

The resemblance of structural proteins to segmented multiblock copolymers has instigated researchers to combine blocks from different structural proteins to design new materials, and furthermore get insight in the individual contributions of the blocks on the final material properties. Designed repetitive block copolypeptides which combine motifs like β-sheets in silk, β-spirals in elastin, animal cell adhesion sequences in fibronectin, and leucine zipper motifs in DNA-binding proteins, have been reported.[68-72]

Alternatively, low-complexity sequences were designed to control polypeptide assembly. An example is the repeating sequence "(EAEAKAK)$_n$", which forms extended β-strand

structures as a result of the alternating polar and non-polar residues.[73] These polypeptides self-assemble into fibrils and form self-sustaining gels. Tirrell and coworkers have explored the possibility to control the solid-state structure and in particular the crystal structure of artificial polypeptides.[74] They reported the biosynthetic production of a series of polypeptides, with the generic sequence -[(AG)$_x$EG]$_n$- (x = 3, 4, 5, and 6), with the aim to have control over polypeptide folding after crystallization. For conventional polymers, including the aliphatic polyamides (the nylons), a commonly occurring crystalline element is the adjacent re-entry chain-folded lamella (Figure 1.4a). Keller and others showed already in the 1950s for polyethylene single crystals that the polymers must bend back on themselves forming a regularly folded configuration.[75, 76] It is a generally accepted configuration for linear polymers that are sufficiently flexible to form hairpin-like folds.
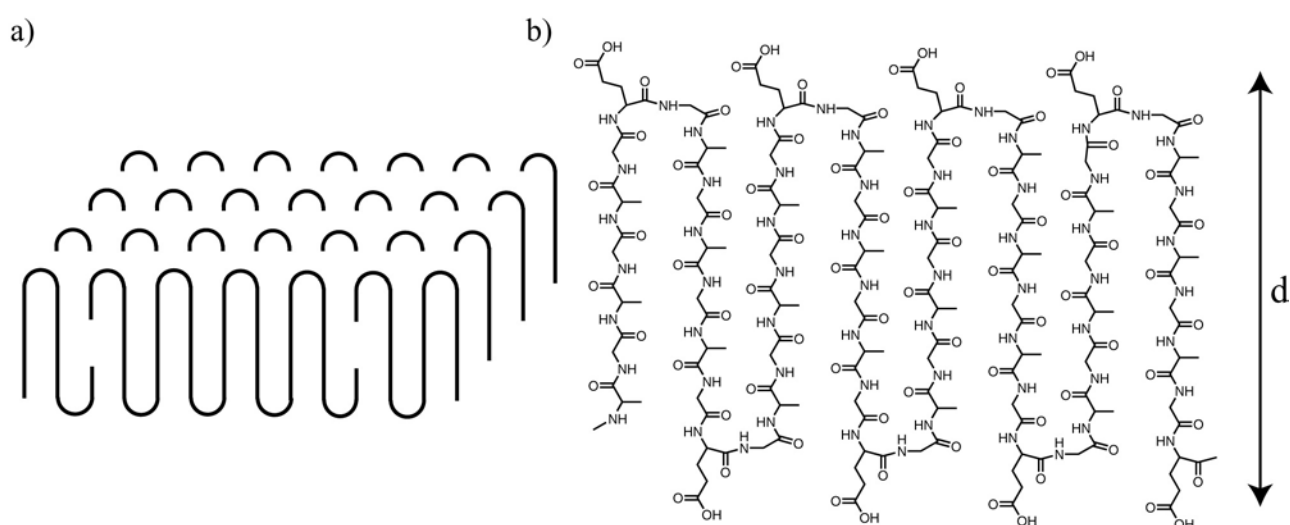
Figure 1.4 (a) A single polymeric crystal in the form of a folded-chain lamella. The polymer chains fold regularly and sharply with a uniform fold-period: the adjacent-reentry model. (b) Lamella formation for poly-[(AG)$_x$EG] (x = 3). The alanylglycine (AG) repeats form antiparallel β-strands in the lamellar interior, whereas the glutamic acid (E) residues are located at the lamellar surface. The number of AG-repeats (x) determines the lamellar thickness, d.

By introducing sequence information in polypeptide-based materials the objective was to obtain control over lamellar crystals with respect to (1) chain conformation (2) lamellar thickness and (3) lamellar surface structure. The design of the –[(AG)$_x$EG]$_n$- sequence was based on the observation that poly(alanylglycine) (AG) and AG-rich polypeptides, including *Bombyx mori* silk fibroin have an antiparallel β-sheet arrangement in the solid state.[77, 78] Secondly, glutamic acid (E) is a poor β-sheet former and its size and polarity makes inclusion in the lamellar interior sterically unfavorable.[79] Indeed, crystallization of these polypeptides from formic acid produced antiparallel β-sheet structures with the glutamic acid residues at the lamellar surface (Figure 1.4b). This crystal structure was supported by X-ray diffraction, vibrational spectroscopy and cross-polarization magic angle spinning ^{13}C nuclear magnetic resonance.[80-82]

In particular, the diffraction results showed that the lamellar thickness, d, is dependent on the number of AG-repeats, ranging from 3.6 nm for -[(AG)$_3$EG]- to 5.5 – 6.2 nm for -[(AG)$_5$EG]-.[74] In addition, artificial polypeptides with other amino acids at the turn positions were reported. The crystallization of periodic polypeptides into β-sheet crystals was reported for the sequences –[(AG)$_3$XG], where X is serine (S), asparagine (N), valine (V), phenylalanine (F), tyrosine (Y) and alternating glutamic acid (E) and lysine (K).[83] The β-sheet stacking periodicity was shown to increase with an increasing bulkiness of the amino acid at the turn position. These examples show that protein engineering is a powerful tool for the synthesis of artificial polypeptides which have a higher level of organization than traditional polymers.

1.6 Aim and outline of this thesis

In this thesis, the preparation of "hybrid" block copolymers, consisting of artificial β-sheet polypeptide blocks and synthetic, non-peptidic blocks is described. The polypeptide blocks are based on the β-sheet sequence of *Bombyx mori* silkworm silk and have the generic sequence -[(AG)$_3$XG]- (where X = glutamic acid or lysine). They adopt a regular antiparallel β-sheet structure in the solid state (see Section 1.5). The aim of this thesis is to combine these β-sheet elements with synthetic polymers. The resulting silk-like model structures could give insight in the relationship between the structure of silk and its unique physical properties. The amorphous synthetic polymer part should be beneficial for the processability of the final composite material. In addition, the presence of glutamic acid or lysine residues at the surface of crystallized poly-[AG)$_3$XG] could be used for functionalization with e.g. mesogenic groups, resulting in new liquid crystalline materials. A literature review on the various synthetic methods for the preparation of "hybrid" block copolymers is given in *Chapter 2*.

Chapter 3 describes the construction of artificial genes coding for the repetitive sequence -[(AG)$_3$EG]$_n$-. The design considerations and the expression of this sequence in *E. coli* is discussed. Furthermore, the expression of a variant with lysine residues at the turn position, -[(AG)$_3$KG]$_n$-, is described.

Chapter 4 deals with the conjugation of synthetic polymer blocks to recombinantly produced poly-[(AG)$_3$EG]. This polypeptide was equipped with telechelic cysteine residues, which could be used for the selective attachment of maleimide-functionalized polymers. The emphasis is on triblock copolymers, consisting of a central poly-[(AG)$_3$EG] β-sheet block and poly(ethylene glycol) end blocks. Furthermore, conjugation with a more hydrophobic synthetic polymer, poly(methyl methacrylate), is described.

Chapter 5 describes the secondary structure characterization of these triblock copolymers using infra-red and circular dichroism spectroscopy. Furthermore, the effect of the attachment of poly(ethylene glycol) on the assembly of the β-sheet block has been investigated by transmission electron microscopy and atomic force microscopy.

Chapter 6 deals with the design and synthesis of artificial α-helical polypeptides. These polypeptides could be useful for the preparation of monodisperse glycopolymers, where spacing between carbohydrate epitopes can be controlled. Multivalent glycopolymers can be potent inhibitors of cell surface carbohydrate-binding receptors and are considered as attractive drug candidates for certain disease states, such as chronic and acute inflammation. An initial polypeptide design with the repetitive sequence $[(MAKA)_2MAA]_n$ is presented. In this design lysine and methionine residues are located on opposing sides of the helix, and could be useful for the attachment of carbohydrate epitopes at regular intervals along the helical axis. This could give more insight in multivalent interactions between glycopolymers and carbohydrate receptors.

1.7 References

[1]   K. Matyjaszewski, in *Advances in Controlled/Living Radical Polymerization, Vol. 854*, **2003**, pp. 2.

[2]   J.-M. Lehn, *Angew. Chem. Int. Edit.* **1988**, *27*, 89.

[3]   D. Kaplan, in *Silk Polymers: Materials Science and Biotechnology, Vol. 544* (Eds.: D. Kaplan, W. W. Adams, B. Farmer, C. Viney), American Chemical Society, Washington, **1994**, pp. 2.

[4]   J. M. Gosline, P. A. Guerette, C. S. Ortlepp, K. N. Savage, *J. Exp. Biol.* **1999**, *202*, 3295.

[5]   K. Tanaka, N. Kajiyama, K. Ishikura, S. Waga, A. Kikuchi, K. Ohtomo, T. Takagi, S. Mizuno, *Biochim. Biophys. Acta* **1999**, *1432*, 92.

[6]   K. Yamaguchi, Y. Kikuchi, T. Takagi, A. Kikuchi, F. Oyama, K. Shimura, S. Mizuno, *J. Mol. Biol.* **1989**, *210*, 127.

[7]   F. Takei, Y. Kikuchi, A. Kikuchi, S. Mizuno, K. Shimura, *J. Cell Biol.* **1987**, *105*, 175.

[8]   T. Sasaki, H. Noda, *Biochim. Biophys. Acta* **1973**, *310*, 76.

[9]   Y. Tsujimoto, Y. Suzuki, *Cell* **1979**, *18*, 591.

[10]   Y. Tsujimoto, Y. Suzuki, *Cell* **1979**, *16*, 425.

[11]   L. P. Gage, R. F. Manning, *J. Biol. Chem.* **1980**, *255*, 9444.

[12]   R. F. Manning, L. P. Gage, *J. Biol. Chem.* **1980**, *255*, 9451.

[13]   K. Mita, S. Ichimura, T. C. James, *J. Mol. Evol.* **1994**, *38*, 583.

[14]   C. Z. Zhou, F. Confalonieri, M. Jacquet, R. Perasso, Z. G. Li, J. Janin, *Proteins-Structure Function and Genetics* **2001**, *44*, 119.

[15]   C. Z. Zhou, F. Confalonieri, N. Medina, Y. Zivanovic, C. Esnault, T. Yang, M. Jacquet, J. Janin, M. Duguet, R. Perasso, Z. G. Li, *Nucleic Acids Res.* **2000**, *28*, 2413.

[16]   S. Nishikawa, S. Ono, *Proc. Tokyo Math.-Phys. Soc.* **1913**, *7*, 131.

[17]   J. O. Warwicker, *Acta Cryst.* **1954**, *7*, 565.

[18]   F. Lucas, J. T. B. Shaw, S. G. Smith, *Biochem. J.* **1957**, *66*, 468.

[19]   L. Pauling, R. B. Corey, *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **1953**, *141*, 21.

[20]   R. E. Marsh, R. B. Corey, L. Pauling, *Biochim. Biophys. Acta* **1955**, *16*, 1.

[21]   Y. Takahashi, M. Gehoh, K. Yuzuriha, *Int. J. Biol. Macromol.* **1999**, *24*, 127.

[22]   R. D. B. Fraser, T. P. MacRae, *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*, Academic Press, New York, **1973**.

[23]   J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, R. Lewis, *Science* **2001**, *291*, 2603.

[24]   M. B. Hinman, Z. Dong, R. V. Lewis, *Mol. Biol. Cell* **1992**, *3*, A113.

[25]   M. Xu, R. V. Lewis, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120.

[26]   C. M. Mello, K. Senecal, B. Yeung, P. Vouros, D. Kaplan, in *Silk Polymers, Vol. 544*, **1994**, pp. 67.

[27]   G. C. Candelas, J. Cintron, *J. Exp. Zoo.* **1981**, 1.

[28]   C. Jackson, J. P. O'Brien, *Macromolecules* **1995**, *28*, 5975.

[29]   A. Simmons, E. Ray, L. W. Jelinski, *Macromolecules* **1994**, *27*, 5235.

[30]   A. H. Simmons, C. A. Michal, L. W. Jelinski, *Science* **1996**, *271*, 84.

[31]   D. T. Grubb, L. W. Jelinski, *Macromolecules* **1997**, *30*, 2860.

[32]   M. B. Hinman, J. A. Jones, R. V. Lewis, *Trends Biotechnol.* **2000**, *18*, 374.

[33]   C. Y. Hayashi, N. H. Shipley, R. V. Lewis, *Int. J. Biol. Macromol.* **1999**, *24*, 271.

[34]   Z. Yang, D. T. Grubb, L. W. Jelinski, *Macromolecules* **1997**, *30*, 8254.

[35]   C. Y. Hayashi, R. V. Lewis, *Science* **2000**, *287*, 1477.

[36]   C. Y. Hayashi, R. V. Lewis, *Bioessays* **2001**, *23*, 750.

[37]   D. W. Urry, *J. Phys. Chem. B* **1997**, *101*, 11007.

[38]   A. A. Van Dijk, L. L. Van Wijk, A. Van Vliet, P. Haris, E. Van Swieten, G. I. Tesser, G. T. Robillard, *Protein Sci.* **1997**, *6*, 637.
[39]   J. Kummerlen, J. D. van Beek, F. Vollrath, B. H. Meier, *Macromolecules* **1996**, *29*, 2920.
[40]   P. A. Guerette, D. G. Ginzinger, B. H. F. Weber, J. M. Gosline, *Science* **1996**, *272*, 112.
[41]   Y. Termonia, in *Structural Biological Materials: Design and Structure-Property Relationships, Vol. 10* (Ed.: M. Elices), American Chemical Society, Washington, DC, **2000**, pp. 271.
[42]   C. Viney, in *Structural Biological Materials: Design and Structure-Property Relationships, Vol. 10* (Ed.: M. Elices), American Chemical Society, Washington DC, **2000**, pp. 295.
[43]   A. Lazaris, S. Arcidiacono, Y. Huang, J. F. Zhou, F. Duguay, N. Chretien, E. A. Welsh, J. W. Soares, C. N. Karatzas, *Science* **2002**, *295*, 472.
[44]   S. Arcidiacono, C. Mello, D. Kaplan, S. Cheley, H. Bayley, *Appl. Microbiol. Biotechnol.* **1998**, *49*, 31.
[45]   D. Huemmerich, T. Scheibel, F. Vollrath, S. Cohen, U. Gat, S. Ittah, *Curr. Biol.* **2004**, *14*, 2070.
[46]   www.nexiabiotech.com.
[47]   S. R. Fahnestock, S. L. Irwin, *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23.
[48]   R. V. Lewis, M. Hinman, S. Kothakota, M. J. Fournier, *Protein Expr. Purif.* **1996**, *7*, 400.
[49]   D. Huemmerich, C. W. Helsen, S. Quedzuweit, J. Oschmann, R. Rudolph, T. Scheibel, *Biochemistry US* **2004**, *43*, 13604.
[50]   J. T. Prince, K. P. McGrath, C. M. Digirolamo, D. L. Kaplan, *Biochemistry US* **1995**, *34*, 10879.
[51]   S. Winkler, S. Szela, P. Avtges, R. Valluzzi, D. A. Kirschner, D. Kaplan, *Int. J. Biol. Macromol.* **1999**, *24*, 265.
[52]   S. Szela, P. Avtges, R. Valluzzi, S. Winkler, D. Wilson, D. Kirschner, D. L. Kaplan, *Biomacromolecules* **2000**, *1*, 534.
[53]   S. R. Fahnestock, L. A. Bedzyk, *Appl. Microbiol. Biotechnol.* **1997**, *47*, 33.
[54]   J. Scheller, K. H. Guhrs, F. Grosse, U. Conrad, *Nat. Biotechnol.* **2001**, *19*, 573.
[55]   J. Scheller, D. Henggeler, A. Viviani, U. Conrad, *Transgenic Res.* **2004**, *13*, 51.
[56]   F. Vollrath, D. P. Knight, *Nature* **2001**, *410*, 541.
[57]   D. H. Hijirida, K. G. Do, C. Michal, S. Wong, D. Zax, L. W. Jelinski, *Biophys. J.* **1996**, *71*, 3442.
[58]   D. P. Knight, F. Vollrath, *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **1999**, *266*, 519.
[59]   F. Vollrath, D. P. Knight, X. W. Hu, *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **1998**, *265*, 817.
[60]   E. K. Tillinghast, S. F. Chase, M. A. Townley, *J. Insect Physiol.* **1984**, *30*, 591.
[61]   D. P. Knight, F. Vollrath, *Naturwissenschaften* **2001**, *88*, 179.
[62]   D. P. Knight, M. M. Knight, F. Vollrath, *Int. J. Biol. Macromol.* **2000**, *27*, 205.
[63]   X. Chen, D. P. Knight, Z. Shao, F. Vollrath, *Biochemistry US* **2001**, *41*, 14944.
[64]   S. R. Fahnestock, Z. Yao, L. A. Bedzyk, *Rev. Mol. Biotechnol.* **2000**, *74*, 105.
[65]   R. L. Lock, in *PCT Int. Appl.*, (du Pont de Nemours, E. I., and Co., USA). Wo, **1993**, p. 16.
[66]   J. P. O'Brien, S. R. Fahnestock, Y. Termonia, K. C. H. Gardner, *Adv. Mater.* **1998**, *10*, 1185.
[67]   A. Seidel, O. Liivak, S. Calve, J. Adaska, G. D. Ji, Z. T. Yang, D. Grubb, D. B. Zax, L. W. Jelinski, *Macromolecules* **2000**, *33*, 775.
[68]   F. Ferrari, J. Cappello, in *Protein-based materials* (Eds.: K. P. McGrath, D. L. Kaplan), Birkhauser, Boston, MA, **1997**, pp. 37.
[69]   J. Cappello, J. W. Crissman, M. Crissman, F. A. Ferrari, G. Textor, O. Wallis, J. R. Whitledge, X. Zhou, D. Burman, L. Aukerman, E. R. Stedronsky, *J. Control. Release* **1998**, *53*, 105.

[70]   A. Nicol, D. C. Gowda, T. M. Parker, D. W. Urry, *Biotechnology of Bioactive Polymers*, Plenum Press, New York, **1994**.

[71]   A. Panitch, T. Yamaoka, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1999**, *32*, 1701.

[72]   W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389.

[73]   N. L. Goeden-Wood, J. D. Keasling, S. J. Muller, *Macromolecules* **2003**, *36*, 2932.

[74]   M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. J. Mason, D. A. Tirrell, *Science* **1994**, *265*, 1427.

[75]   A. Keller, *Philos. Mag.* **1957**, *2*, 1171.

[76]   E. W. Fischer, *Z. Naturforsch.* **1957**, *12a*, 753.

[77]   J. M. Anderson, H. H. Chen, W. B. Rippon, A. G. Walton, *J. Mol. Biol.* **1972**, *67*, 459.

[78]   R. D. B. Fraser, T. P. MacRae, F. H. C. Steward, E. Suzuki, *J. Mol. Biol.* **1965**, *11*, 706.

[79]   P. Chou, G. Fasman, *Biochemistry US* **1974**, *13*, 222.

[80]   C. C. Chen, M. T. Krejchi, D. A. Tirrell, S. L. Hsu, *Macromolecules* **1995**, *28*, 1464.

[81]   J. X. Wang, A. D. Parkhe, D. A. Tirrell, L. K. Thompson, *Macromolecules* **1996**, *29*, 1548.

[82]   M. T. Krejchi, S. J. Cooper, Y. Deguchi, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1997**, *30*, 5012.

[83]   E. J. Cantor, E. D. T. Atkins, S. J. Cooper, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Biochem.* **1997**, *122*, 217.

Chapter 2

Peptide-containing block copolymers

Abstract

The progress in both peptide and polymer chemistry has led to the preparation of hybrid molecular architectures with properties not achievable with the separate components. The current synthetic methodologies for the preparation of peptide-containing block copolymers, either being completely peptidic or as hybrids with synthetic polymer blocks, are described. For the hybrid block copolymers two categories are distinguished, i.e. block copolymers with the peptide segments either in the main chain or in the side chain. The different techniques to prepare these polymers are ordered by level of control over peptide sequence and hence their molecular structure.

2.1 Introduction

The development of synthetic approaches for the preparation of well-defined polypeptide-based materials has attracted increasing attention in the polymer chemistry field. The many levels of supramolecular organization present in proteinaceous materials in nature show that only a limited number of monomer building blocks is sufficient for the construction of complex structures, with a wide range of properties. The chemical information stored in the amino acid side chains does not only result in the folding of the peptide main-chain into a specific secondary structure, but is also responsible for tertiary folding as a result of the delicate interplay of a variety of non-covalent interactions, like hydrophobic, hydrogen-bonding and ionic interactions. Despite the large variety of monomers that can be polymerized chemically and the increasing level of control over polymer size distribution using controlled polymerization techniques[1], a fundamental limitation is the lack of absolute control over monomer sequence.

In many cases, only small peptide sequences are already sufficient to introduce specific functionality such as folding, recognition, biodegradability and mechanical properties into molecular structures. This realization has led to an increased use in polymer science of polypeptide preparation methods of organic synthetic and biosynthetic origin. Especially, the combination of amino acid sequences and synthetic, non-peptidic polymers has drawn much attention. The synthetic polymer part can either act simply as a carrier or contribute to the overall molecular features with its specific solubility or mechanical properties. These block copolymer systems are expected to be useful for compatibilization of synthetic and biological systems.

In this chapter the current (bio)synthetic methodologies are described for the preparation of peptide-containing block copolymers, which can be either completely peptidic or hybrids with synthetic polymer blocks. The different techniques are ordered by level of control over peptide sequence and molecular structure. Furthermore, both main and side chain peptide containing polymers will be discussed.

2.2 Homopolypeptide block copolymers

The most frequently used synthetic methodology for the preparation of block copolymers containing homopolypeptides is the ring-opening polymerization of protected α-amino acid-N-carboxyanhydrides (α-NCAs) initiated by a primary amino-end functional polymer (Figure 2.1). This technique allows multigram-scale synthesis, but has long been plagued by chain-breaking transfer and termination reactions, resulting in products with a wide range of molecular weights (polydispersity index, PDI > 1.4).[2] Furthermore, an often observed phenomenon of NCA-polymerizations is precipitation of the growing polypeptide chain at a certain molecular weight and/or the formation of secondary structure, which may result in

physical termination. Fractionation has therefore commonly been applied before physical characterization of these polymers.
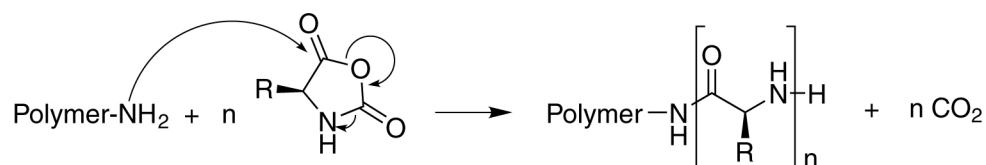
Figure 2.1 *Preparation of a polypeptide containing block copolymer by ring-opening polymerization of α-amino acid N-carboxyanhydrides initiated by an amino-functional polymer.*

A large number of different block copolymers have been prepared using ω-amino or α,ω-diamino end-functionalized synthetic polymer blocks as initiator. This subject has been reviewed by Gallot[3], Klok et al.[4] and Schlaad et al.[5] Commonly used NCAs are γ-benzyl L-glutamate and N-ε-benzyloxycarbonyl L-lysine, since their polymerization is known to be the best controlled.[2] The synthetic polymer blocks have been varied extensively from hydrophobic blocks, like polystyrene and polybutadiene to hydrophilic blocks like poly(ethylene glycol) and poly(vinyl alcohol).[3] Polypeptide blocks based on poly(glutamic acid) or poly(lysine) are known to form α-helical secondary structures in the solid-state and in solution, resulting in rod-like molecules which behave as mesogens. By combining them with coil-like synthetic polymers these rod-coil block copolymers show a strong preference in the solid state for the formation of a lamellar structure of alternating polyvinyl and polypeptide sheets. In addition to these lamellae, the α-helical polypeptides are arranged in a hexagonal array and for longer polypeptides the chain backfolds. The final result is a hexagonal-in-lamellar morphology as depicted in Figure 2.2.[5]
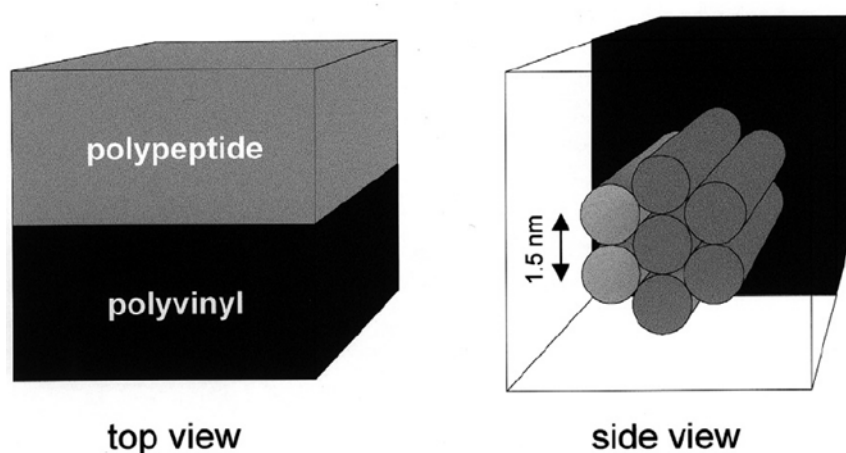
Figure 2.2 *Schematic model of the hexagonal-in-lamellar morphology of polyvinyl-polypeptide block copolymers (α-helical polypeptides are represented as cylinders).[5]*

Oligomeric rod-coil block copolymers have been described by Klok et al. who studied the supramolecular organization of diblock copolymers with an α-helical oligopeptide rod based on either γ-benzyl-L-glutamate (10 – 80 units) or ε-benzyloxycarbonyl-L-lysine (20 – 80 units) and a short oligostyrene block (10 units).[6, 7] The supramolecular organization of these block copolymers was shown to be sensitive to temperature. Depending on the amino acid and the polypeptide length, a partial transition from an α-helical to a β-sheet secondary structure was observed upon increasing the temperature. A small fraction of peptide segments with a β-sheet conformation was found to be sufficient to prevent the regular organization of the α-helical blocks, resulting in a change in observed supramolecular organization from a hexagonal to a lamellar β-sheet morphology.

An interesting example of structure formation in dilute solution is the work on polybutadiene-block-poly(L-glutamate) (PB-b-PGA). Kukula et al. reported on the formation of spherical micelles and large vesicles ("peptosomes", diameter ~ 150 nm) in aqueous solution, depending on the chemical composition of the copolymer.[8] Chécot et al. described the formation of vesicular aggregates by PB_{40}-b-PGA_{100}. The secondary structure of the polypeptide block changed from α-helix at pH 4.5 to random coil at a higher pH.[9] This conformational transition was reflected in the hydrodynamic radius of the vesicles, which increased from approximately 100 nm at pH 4.5 to 150 nm above pH 7. The proposed model for the build-up of these vesicles is depicted in Figure 2.3. The 1,2-vinyl bonds of the polybutadiene block could be used for UV-induced cross-linking and resulted in stable vesicles which might be suitable for drug encapsulation or sensor nanodevices.[10]
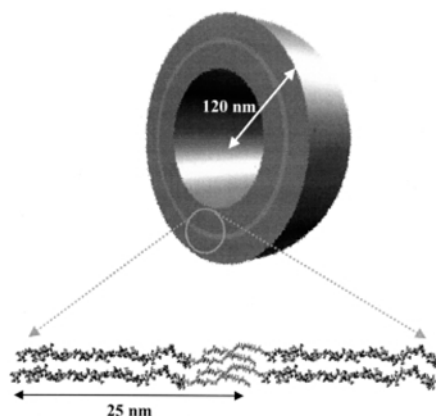
Figure 2.3 *Proposed model for the self-assembly of polybutadiene-block-poly(L-glutamate) (PB_{40}-b-PGA_{100}) into vesicular aggregates.*[9]

The synthesis of well-defined block copolypeptides by NCA polymerization has become accessible since the use of transition metal-amine initiator complexes like bipyNi(COD) (bipy = 2,2'-bipyridyl, COD = 1,5-cyclooctadiene) instead of primary amine initiators. Block copolypeptides could be prepared via this method with better control over chain length and with

low polydispersity (PDI < 1.2).[11, 12] Recently described alternatives for this "coordination" polymerization process include the use of hydrochloride salts of primary amines as initiator and the use of high vacuum techniques.[13, 14] However, only for the block copolypeptides prepared via "coordination" polymerization, physical characterization has been reported. For example, amphiphilic diblock copolymers have been synthesized with a poly(L-lysine bromide salt) or poly(L-glutamate sodium salt) hydrophilic block and a poly(L-leucine) or poly(L-valine) hydrophobic block. These diblock amphiphiles were found to form rigid hydrogels (instead of micelles as would be expected) that retained their mechanical strength up to temperatures of about 90 °C and recovered rapidly after stress.[15-17] The gelation properties of these copolymers were not solely dependent on the amphiphilic nature of the polypeptide, instead it appeared that gelation was tied to the conformations of the hydrophobic domains, where α-helical segments formed by the poly(L-leucine) blocks were slightly better gelators than β-strands formed by poly(L-valine).

Another block copolypeptide, consisting of L-cysteine and L-lysine, acted as a morphology-directing component in the formation of ordered silica structures.[18] This block copolymer, which mimics the protein silicatein, self-assembles as a result of its amphiphilic character, whereas the L-cysteine residues enable hydrolysis of the tetraethoxysilane precursor at neutral pH. These examples show that although NCA polymerization does not provide absolute sequence control, the assembly properties of these materials can make them useful as tissue engineering scaffolds, drug carriers or templates for bio-mimetic composite formation.[17, 19-25]

2.3 Solution phase synthesis of peptide-containing block copolymers

Solution phase synthesis of peptides is useful for the preparation of short sequences of up to 10 residues in length. Difficulties that limit this technique are non-quantitative coupling reactions and the often problematic solubility of oligopeptides. Only few examples have been reported in which small peptidic elements have been used in the preparation of "hybrid" multiblock copolymers. An interesting example in which the peptide segment was used to contribute to the mechanical properties of the final material is the work of Sogah et al.[26-28] They prepared *Bombyx mori* silk worm silk and *Nephila clavipes* spider dragline silk mimetic block copolymers. Silk worm silk is of interest because of its use in textile fibers, whereas spider dragline silk is one of the types of spider silk known to have a remarkable combination of strength and toughness.[29] Both silks contain repetitive amino acid sequences which form crystalline and amorphous domains in the silk fiber. The crystalline domains give the material strength, whereas the amorphous protein matrix allows the crystalline domains to orient under strain to increase the strength of the material and introduces flexibility to increase the energy to break. The crystalline β-sheet sequences are poly(alanylglycine) in silkworm silk and poly(alanine) in dragline silk. Sogah et al. used either the tetrapeptide GlyAlaGlyAla (Gly =

glycine; Ala = alanine) or poly(Ala) and replaced the amorphous silk sequences with poly(ethylene glycol). Multiblock copolymers were prepared via polycondensation using two approaches. The first one made use of an aromatic hairpin residue to force formation of parallel β-sheets as is schematically depicted in Figure 2.4a. A second approach is depicted in Figure 2.4b and is a fully linear system in which the β-sheet segments were free to form intra- and intermolecular parallel or antiparallel β-sheets.
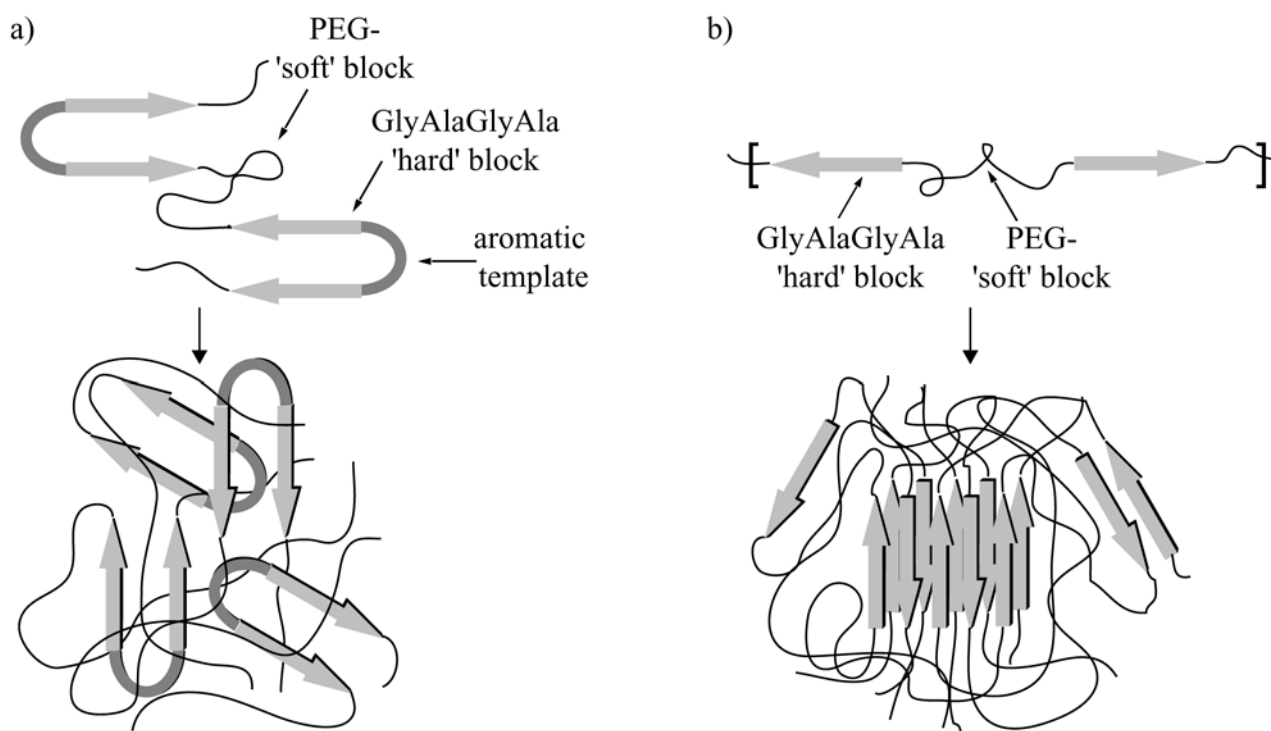
Figure 2.4 *Schematic drawing of multiblock copolymers consisting of a poly(ethylene glycol) "soft" block and a tetrapeptide GlyAlaGlyAla, crystalline "hard" block in two variants:* ***(a)*** *Templated system in which an aromatic hairpin turn is used to force parallel β-sheet formation.* ***(b)*** *Non-templated system in which peptide segments are free to form parallel and/or antiparallel β-sheets. Adapted from ref. [26].*

For both designs a microphase separated morphology was observed with 20 – 50 nm peptide domains dispersed in a continuous poly(ethylene glycol) phase. Furthermore, a 100 – 150 nm superstructure was observed in cast films which were explained to result from the polydispersity and multiblock character of the polymers. The mechanical properties of fibers and films made from these block copolymers could be modulated by manipulating the length and nature of the constituent blocks. Similar work was reported by Shao et al.[30]

In an alternative approach atom transfer radical polymerization (ATRP) was used for the introduction of the tetrapeptide AlaGlyAlaGly in the side chain of a polymer.[31] For this purpose the peptide sequence was functionalized at the C-terminus with a methacrylate moiety, which was polymerized in a controlled fashion. This side chain polymer was subsequently used

as a macroinitiator for ATRP of methyl methacrylate to yield the corresponding amorphous poly (methyl methacrylate). Infra-red spectroscopy clearly showed that the resulting well-defined triblock copolymer possessed a β-sheet secondary structure. The same ATRP methodology has also been applied for the construction of side chain polymers containing pentapeptide repeats commonly found in tropoelastin (ValProGlyValGly).[32] In this case the pentapeptide was synthesized either by solution or solid phase peptide chemistry. The N-terminus was modified with a methacrylate handle. Subsequent ATRP from a bifunctional poly(ethylene glycol) macroinitiator resulted in a triblock copolymer with a narrow molecular weight distribution (PDI = 1.2). This block copolymer showed a temperature and pH dependent aggregation behavior which was similar to elastin.

A unique way of introducing small peptides in the side chain of polymers has been reported by Cornelissen et al.[33] They prepared amphiphilic block copolymers containing a polystyrene tail and a charged poly(isocyanide) headgroup derived from isocyano-L-alanine-L-alanine and isocyano-L-alanine-histidine. This type of rod-coil block copolymer formed micelles, vesicles and bilayer aggregates in aqueous solution as is depicted in Figure 2.5. Furthermore, the chirality of the poly(isocyanodipeptide) block was shown to be dependent on the dipeptide used. Moreover, helical superstructures were of opposite chirality to that of the constituent block copolymer.
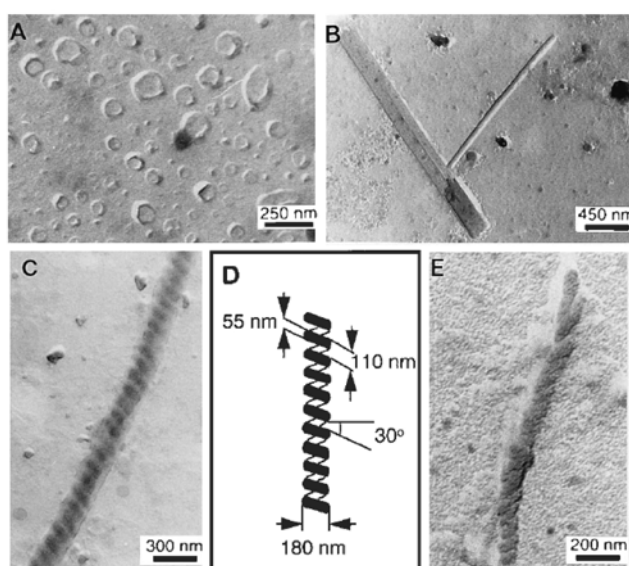
Figure 2.5 *TEM micrographs of the morphologies formed by polystyrene-block-poly(isocyanodipeptide)s in aqueous solution: (a) vesicles (b) bilayer filaments (c) left-handed superhelix, (d) schematic representation and (e) right-handed helical aggregate.[33]*

2.4 Solid phase synthesis of peptide-containing block copolymers

For the synthesis of small to medium-sized peptides solid phase peptide synthesis is the method of choice. This technique involves the stepwise addition of N-protected amino acids to a peptide

chain anchored with its C-terminus to a polymeric support.[34] With this method a peptide is constructed via sequential coupling and deprotection steps and thus sequence-specific peptides can be prepared. The immobilization of the peptide allows the use of excess of coupling and deprotection reagents, resulting in high yields per step. The immobilization of the desired product obviates the need to purify intermediates. Deletions and truncations are however unavoidable, making this method less trivial for the preparation of larger peptides (> 50 residues).

Several studies of "hybrids" prepared by a fully solid phase based synthetic procedure have been reported. An interesting example was reported by Burkoth et al.[35-37], who investigated fibril formation of β-amyloid peptide, the primary component of amyloid plaques in Alzheimer's disease. They derivatised the C-terminus of the central domain of β-amyloid peptide (Aβ(10-35)) with poly(ethylene glycol)-3000 using standard Fmoc solid phase peptide synthesis. Although some models for β-amyloid structure (with the C-terminal hydrophobic domain in the fibril interior) suggested the attachment of PEG would disturb fibril formation, the conjugate did self-associate into fibrils. The fibril formation was, unlike the native peptide, completely reversible and the solubility of the formed fibrils was greatly improved. The lateral association of the fibrils into bundles was prevented, as can be seen in the electron micrographs of Figure 2.6.

YEVHHQK**LVFFA**EDVGSNKGAIIGLM

YEVHHQK**LVFFA**EDVGSNKGAIIGLM
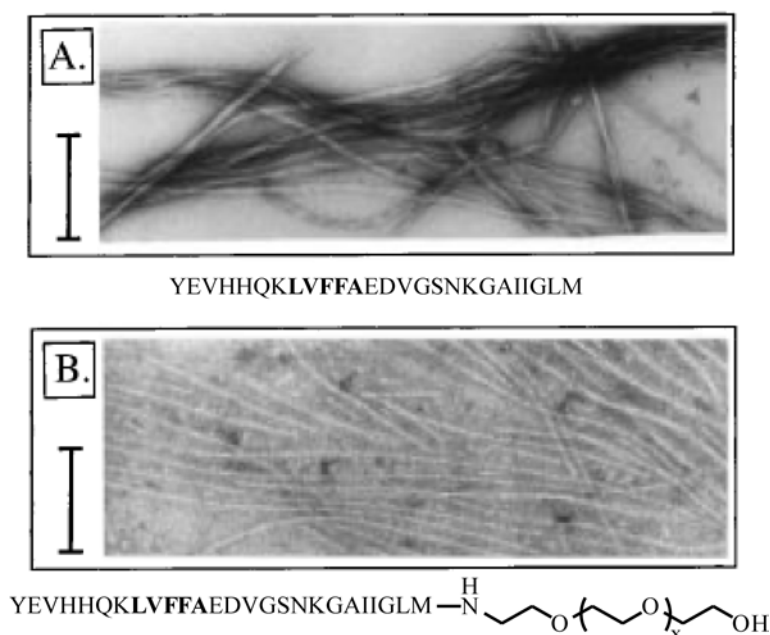
*Figure 2.6 Electron microscopy images of fibrils formed by **(a)** β-amyloid peptide (Aβ (10-35)) and **(b)** poly(ethylene glycol)-3000 conjugate of Aβ (10-35). Scale 200 nm. Adapted from ref. [35].*

Using the same synthetic procedure, Rosler et al. investigated the solid state and melt structures of poly(ethylene glycol) di- and triblock copolymers, containing amphiphilic β-strand peptide

sequences.[38] These hybrids formed superstructures consisting of alternating PEG layers and peptide domains in an antiparallel β-sheet conformation.
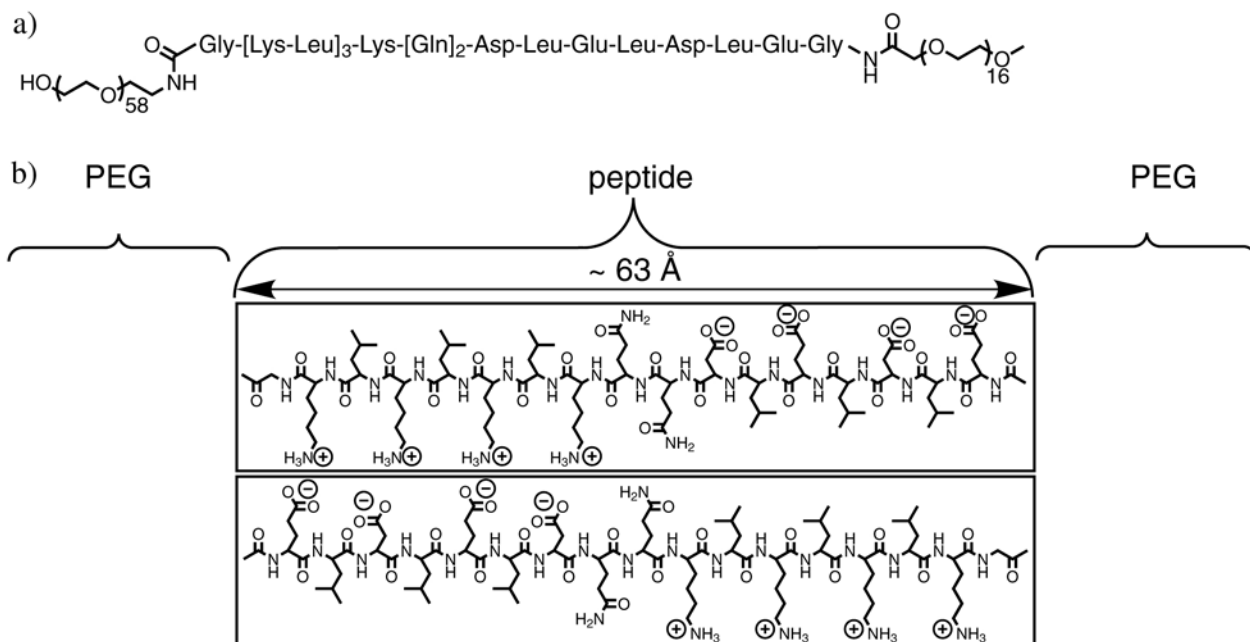
Figure 2.7 (a) *Triblock copolymer consisting of a central amphiphilic β-strand peptide and flanking poly(ethylene glycol) blocks.* ***(b)*** *Model for the nanoscale organization of the block copolymer. The plane of amino acid side-chain interactions is shown (Coulombic interaction between lysines (Lys) and glutamic acids (Glu)/aspartic acids (Asp), hydrophobic interaction between leucines (Leu) and hydrogen bonding between glutamines (Gln)). Hydrogen bonding takes place perpendicular to the plane of the paper.*[38]

A report that emphasizes the synthetic procedure for the preparation of a hybrid triblock copolymer in the solid phase was reported by Reynhout et al.[39] The triblock copolymer consisted of a central β-turn peptide (from the circumsporozoite (CS) protein of the malaria parasite *Plasmodium falciparum*) and polystyrene end blocks. In this approach amine-functionalized polystyrene was coupled to an aldehyde modified resin by reductive amination (Figure 2.8). The resulting secondary amine was then used for the build-up of the peptide segment, followed by coupling of a carboxylic acid functionalized polystyrene at the N-terminus. Non-quantitative coupling of the first amino acid to the secondary amine and the second polymer to the N-terminus of the peptide resulted in a mixture of polystyrene, PS-b-peptide and PS-b-peptide-b-PS. The triblock copolymer could however be purified by column chromatography resulting in a final yield of 15% based on initial loading. Alternative total solid phase based strategies for the preparation of polymer-peptide block copolymers were based on the controlled polymerization of the synthetic polymer block from the supported peptide segment using either nitroxide-mediated radical polymerization (NMRP) or ATRP (Figure 2.9).[40, 41]
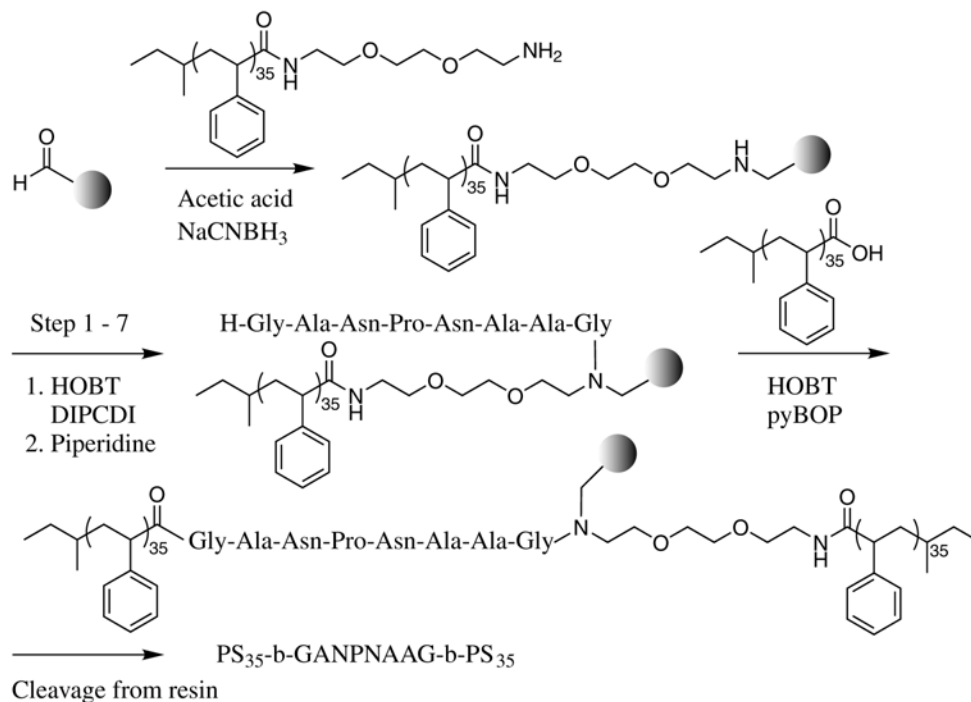
Figure 2.8 *Solid phase synthetic route towards triblock copolymer consisting of a β-turn peptide middle block and polystyrene end blocks.*[39]
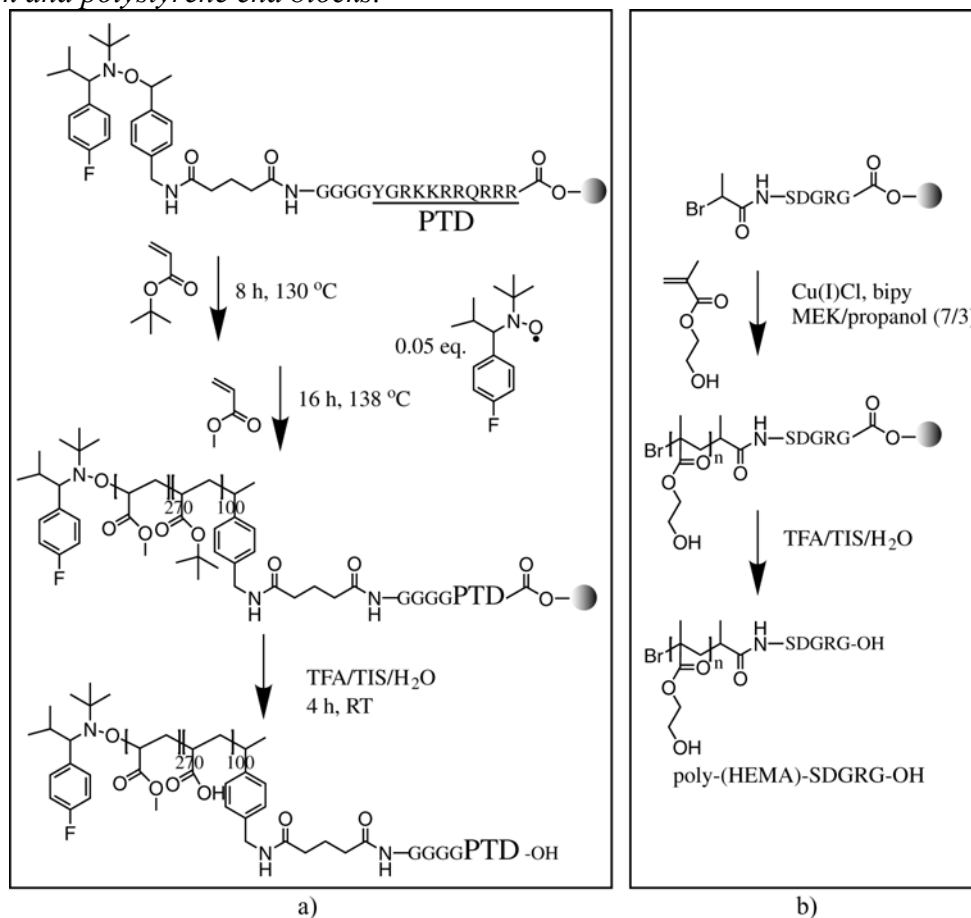
Figure 2.9 *Two solid phase strategies towards peptide containing block copolymers by polymerization from the supported peptide via **(a)** nitroxide-mediated radical polymerization and **(b)** atom transfer radical polymerization. Adapted from ref. [40, 41].*

For NMRP a small peptide, containing a peptide transduction domain (PTD) sequence was prepared via standard Fmoc solid phase synthesis. The N-terminus was subsequently functionalized with a fluorine-labeled alkoxyamine initiator. This could be used for the sequential solvent-free polymerization of tert-butyl acrylate and methyl acrylate, resulting in an ABC triblock copolymer. Traditional characterization of this triblock copolymer by gel permeation chromatography and MALDI-TOF mass spectroscopy was however complicated partly due to solubility problems. Therefore characterization of this block copolymer was limited to ^{1}H and ^{19}F NMR and no conclusive evidence on molecular weight distribution and homopolymer contaminants was obtained. The same principle was shown for ATRP of hydroxyethyl methacrylate from a resin-supported fibronectin-based RGD peptide sequence. Polymerization was carried out in methyl ethyl ketone/propanol (7:3) and resulted in a bio-hybrid with a polydispersity of 1.5.

Attachment of a polymerisable handle to peptides synthesized by solid phase peptide synthesis has been used for generating high molecular weight polymers with peptide moieties pending from the backbone. These side chain peptide polymers are especially of interest in the development of synthetic vaccines. The preparation of defined, chemically synthesized peptide-antigens could exclude the use of infectious material and thus result in safer vaccines. However, short peptides generally do not elicit good immune responses, therefore several synthetic approaches have been used to generate higher molecular weight polymers from them.[42] One approach is the so called "multiple antigen peptide" (MAP) system.[43] These molecules consist of a branched core resulting from the use of both amino groups of lysine (N-α and N-ε, Figure 2.10). Antigenic peptides are either chemically synthesized onto this core or coupled after synthesis. An alternative approach has been described using acryloyl modified peptides which are polymerized by free radical polymerization, allowing the production of copolymers of different epitopes.[44]
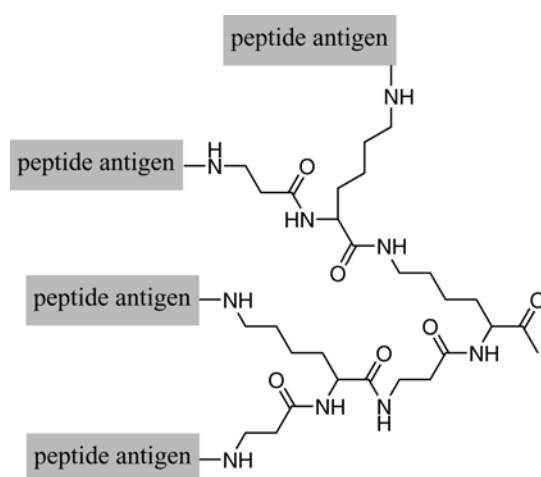
Figure 2.10 *Example of a branched "multiple antigen peptide" consisting of a lysine-b-alanine core. Adapted from ref. [43].*

The limitations of solid phase peptide synthesis with respect to length can be overcome by "chemoselective ligation" of smaller unprotected peptide fragments. This strategy has become especially attractive with the introduction of "native chemical ligation" , in which the coupling results in the formation of the native, amide bond (in contrast to earlier chemical ligation strategies which resulted in e.g. thioester, thioether or oxime bonds).[45] Native chemical ligation is based on the reaction of a peptide-α-thioester with another peptide segment containing an amino-terminal cysteine residue (Figure 2.11a). The thioester-linked intermediate that is formed spontaneously rearranges, resulting in the formation of an amide bond and the regeneration of the free sulfhydryl group. An interesting practical extension of this concept is the fully solid-phase based chemical ligation which allowed the build-up of longer peptides (up to 118 amino acids) in either the N→C or the C→N direction.[46] Furthermore, the development of the 1-phenyl-2-sulfanylethyl auxiliary group has allowed the ligation of peptides without the requirement of a cysteine group (Figure 2.11b).[47]
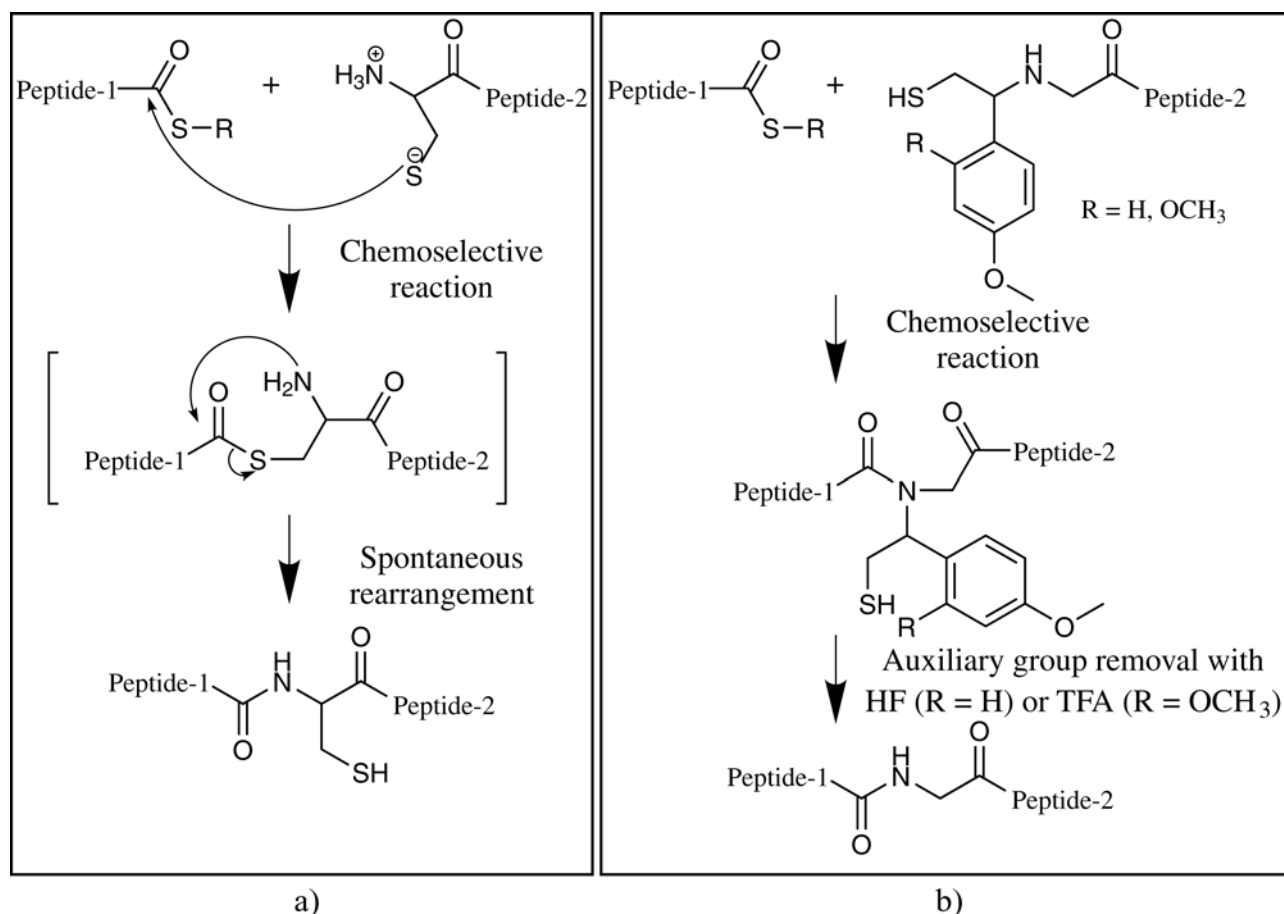
Figure 2.11 (a) *The principle of native chemical ligation. Peptide 1, containing a C-terminal thioester, undergoes a nucleophilic attack by the cysteine residue at the N-terminus of peptide 2. The intermediate rearranges spontaneously to form the native peptide bond. **(b)** Native chemical ligation without the requirement of a cysteine residue using the 1-phenyl-2-sulfanylethyl auxiliary group. Adapted from ref. [45, 47].*

This cleavable thiol-containing auxiliary is added to the alpha-amino group of one peptide segment to facilitate amide bond-forming ligation. The amine-linked 1-phenyl-2-mercaptoethyl moiety is stable under the conditions used to cleave and deprotect peptides after solid-phase peptide synthesis. However, after ligation the auxiliary group can be removed by treatment with anhydrous hydrogen fluoride or trifluoroacetic acid. This methodology has been demonstrated in the synthesis of cytochrome b562.[48] Furthermore, also the Staudinger ligation has been proposed as a peptide ligation method which does not depend on the amino acids present at the junction site. The potential of this reaction for the ligation of unprotected peptide segments is, however, still not established.[49] An extensive review on the chemical synthesis of proteins by ligation has been reported by Borgia et al.[50]

   Elegant work in which the chemical ligation methodology was used for the preparation of a polymer-modified protein was reported by Kochendoerfer et al.[51] They prepared a monodisperse, polymer-modified, synthetic erythropoiesis protein (SEP) (Figure 2.12). SEP was designed to be analogous to human erythropoietin (Epo), a glycoprotein hormone that regulates the proliferation, differentiation and maturation of erythroid cells. SEP was assembled by native chemical ligation of four individual peptide segments (from 27 to 55 amino acids) resulting in a 166-amino acids polypeptide chain.
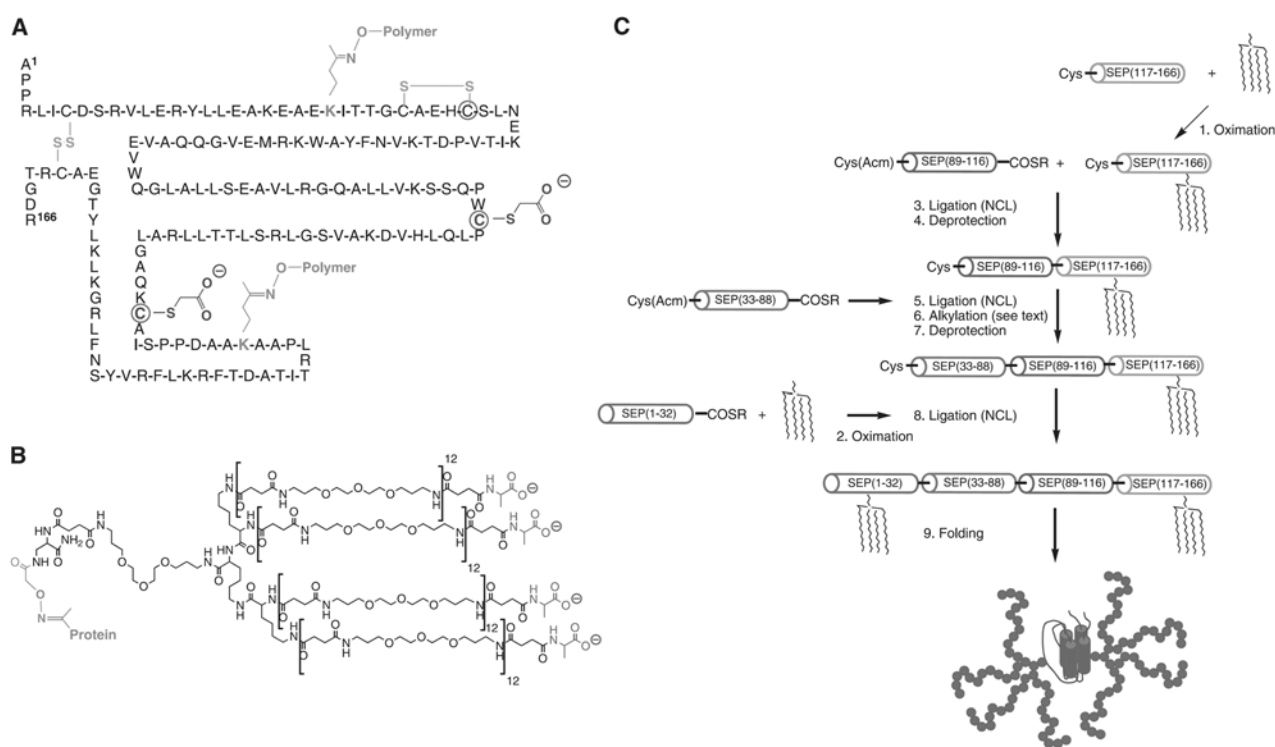
Figure 2.12 Molecular structure of synthetic erythropoietin protein (SEP). (a) Primary amino acid sequence with three ligation sites circled in red. (b) Structure of the branched, negatively charged poly(ethylene glycol) based polymer. (c) Scheme for the synthesis of SEP by chemical ligation. Branched polymers were first attached to the individual peptide segments by oxime-forming ligation followed by native chemical ligation.[51]

Before this ligation, two of the peptide segments were functionalized via oxime forming ligation[52] with a monodisperse, negatively charged and branched poly(ethylene glycol) derivative. These polymer moieties were attached to sites corresponding to two glycosylation sites of Epo and their negative charge enabled prolonged duration of action in vivo. This synthetic methodology resulted in a precisely defined 51 kDa protein-polymer conjugate that was produced in quantities over 100 mg, which is a considerable synthetic improvement in comparison to common protein-polymer conjugates that are often heterogeneous with respect to the polymers attached and the attachment sites.

2.5 Preparation of artificial polypeptides by protein engineering

The term "protein engineering" encompasses a whole range of techniques, which are focused on altering the expression or properties of proteins in a specific manner. It is applied in a variety of areas such as biochemistry, microbiology, molecular biology and genetics and is used for example to change the cellular localization of enzymes or to modify the properties of enzymes with respect to catalysis and stability. The cloning and expression of genes has become a standard tool in biochemistry and molecular biology[53] and has been recognized only since the last decade as a useful synthetic tool for the preparation of polypeptide-based materials. This technique allows the preparation of monodisperse, high molecular weight polypeptides with complete sequence control. Protein engineering has been successfully used for the preparation of recombinant, structural proteins, such as silks, collagen and elastin.[54, 55] The common feature of these fibrous protein polymers is the presence of repetitive sequence motifs which form defined secondary structures.

The repetitive amino acid sequences present in structural proteins offer the possibility to construct artificial genes by multimerization of small synthetic oligonucleotide sequences and thus build up high molecular weight proteins. The constructed artificial genes can be incorporated into an expression plasmid, which can subsequently be transferred to a bacterial host for production of the desired polypeptide (Figure 2.13). The most commonly used host is *E. coli*.

The resemblance of structural proteins to segmented multiblock copolymers has instigated researchers to combine blocks from different structural proteins to design new materials, and furthermore get insight in the individual contributions of the blocks on the final material properties. Some examples are given on designed repetitive block copolypeptides which combine motifs like β-sheets in silk, β-spirals in elastin, leucine zipper motifs in DNA-binding proteins and animal cell adhesion sequences in fibronectin.

Multiblock copolymers consisting of silk-like β-sheet blocks and elastin-like blocks were reported by Capello and Ferrari.[56] The silk-like blocks consisted primarily of GAGAGS, derived from silkworm silk fibroin and the elastin-like blocks had the general amino acid sequence VPGVG, derived from mammalian elastin. By varying the ratio of crystallizing β-sheet blocks and solubilizing elastin blocks, the rate of irreversible hydrogel formation,

bioresorption and the release of solutes from these gels could be influenced. These characteristics make these block copolymers interesting for controlled drug delivery.[57]
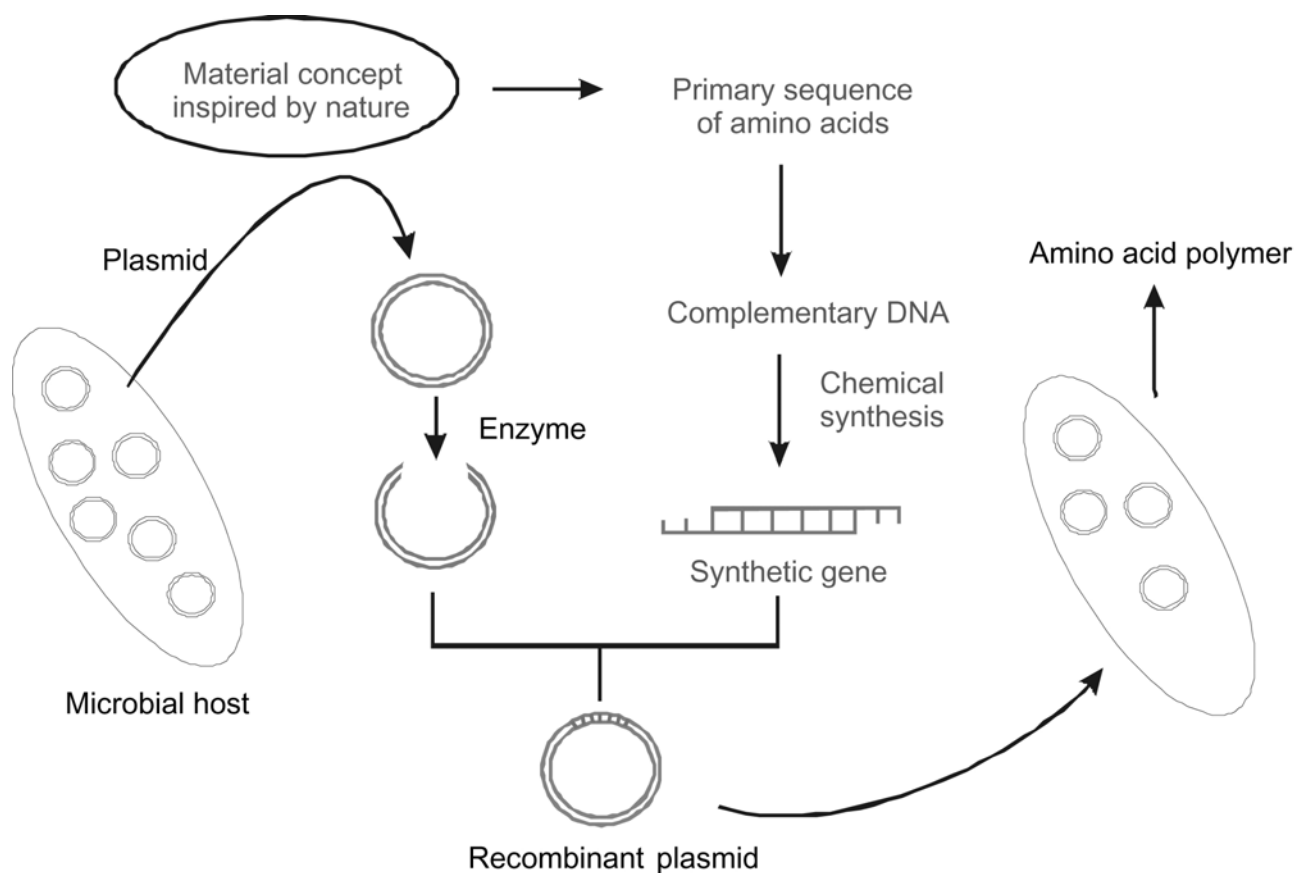
Figure 2.13 *Protein biosynthesis by bacterial expression of artificial genes. Based on the desired amino acid sequence an artificial gene is built up from chemically synthesized oligonucleotides and introduced via a plasmid in the microbial host, which in turn can produce the desired protein polymer.*

Conticello et al. and Chilkoti et al. investigated the self-assembly of elastin-like block copolymers.[58-60] The temperature at which aggregation of elastin-like polypeptides occurs, i.e. the lower critical solution temperature (LCST), can be tuned by variation of the fourth amino acid residue of the consensus pentapeptide repeat VPGVG. With an increasing polarity of the amino acid residue at this position, the LCST is increased.[61] Thermoreversible micelles and hydrogels could be prepared using AB diblock copolypeptides and BAB triblock copolypeptides, respectively. In these block copolypeptides the B block is more hydrophobic and aggregates selectively, whereas the more hydrophilic A block stays solvated. The possibility for entrapment of hydrophobic drugs in the micellar core and the triggered disassembly of the micelles, suggests their potential application in targeted drug delivery.

Block copolypeptides with envisioned applications in tissue engineering were reported by several groups. These polypeptides combined structural domains from silk or elastin with cell-binding domains from the natural extracellular matrix protein fibronectin. Cappello and

coworkers prepared block copolypeptides consisting of the silkworm silk β-sheet sequences and the tripeptide sequence arginine-glycine-aspartic acid (RGD), which is the cell adhesion motif of fibronectin.[56] The crystalline β-sheet blocks have been described to adsorb to hydrophobic plastic surfaces, while exposing the cell adhesion sequences. Urry et al. and Panitch et al. used elastin-like sequences which they combined with RGD or an alternative fibronectin sequence REDV.[62, 63] These materials had mechanical properties similar to those of the arterial wall and supported the adhesion of vascular endothelial cells.

Coiled-coil motifs are present in a large variety of proteins, like DNA binding proteins, keratins and muscle proteins. An interesting example in which a coiled-coil motif has been used to prepare pH dependent reversible hydrogels, is the work on triblock copolymers comprising a central random coil, polyelectrolyte domain $[(AG)_3PEG]_{10}$ flanked by leucine zipper domains based on the Jun oncogene product (Figure 2.14).[64]

a)

MRGSHHHHHHGSDDDDKWA- Helix -IGKHVAPRDTSYRDPMG- $[(AG)_3PEG]_{10}$

ARMPTSGD- Helix -IGDHVAPRDTSMGGC

Helix = SGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE

b)

pH, temperature

Gel   Viscous Liquid
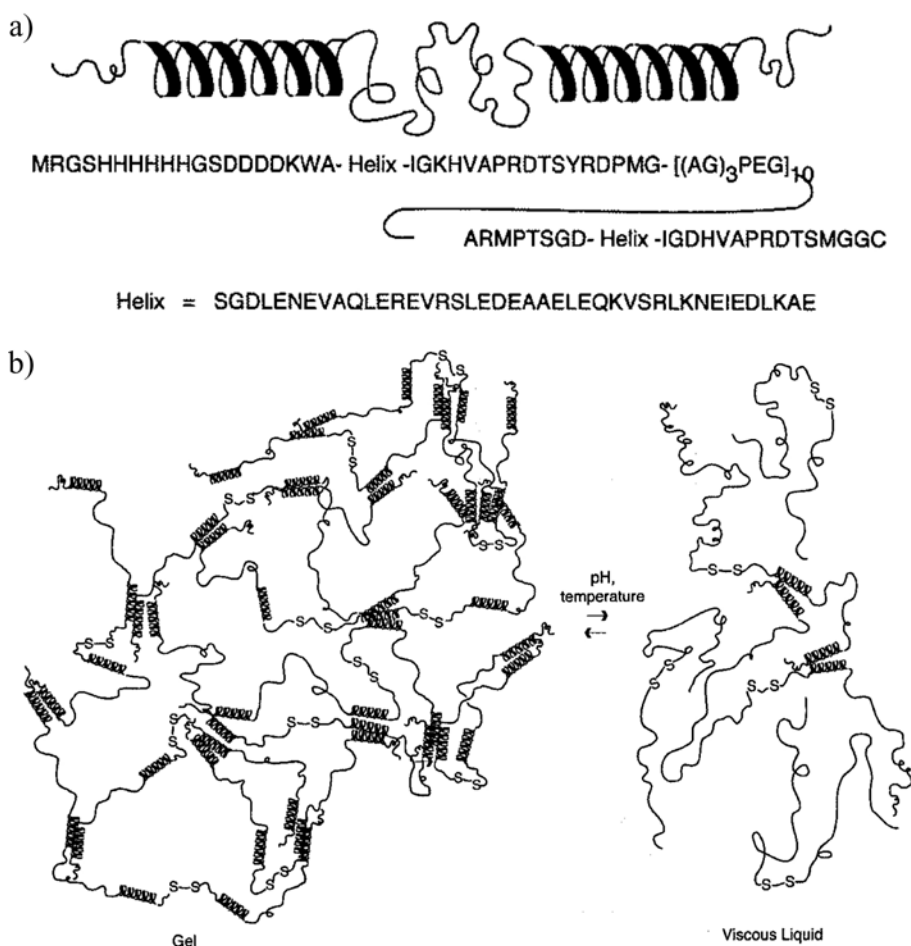
Figure 2.14 (a) *Triblock protein polymer consisting of a polyelectrolyte middle block with the repetitive sequence [(AG)₃PEG]₁₀ and two leucine zipper end blocks (helix). (b) Physical gelation of this triblock copolymer is the result of leucine zipper dimerization and can be modulated by pH and temperature.[64]*

The leucine zipper motif is characterized by a consensus heptad repeat (*abcdefg*), with *a* and *d* the hydrophobic amino acids, whereas *e* and *g* are usually charged. The hydrophobic amino acids are located on one side of the helix and cause the formation of a dimeric coiled-coil structure. The charged amino acids modulate the stability of the dimer. The result is a switchable hydrogel: At low pH and low temperature the material formed elastic gels, while gelation was lost upon increase in pH or temperature.

2.6 Conclusion

The improvement of synthetic methodologies has allowed the synthesis of increasingly better defined "hybrid" block copolymers. The elucidation of the new structural organizations found in these block copolymer systems and the design of materials with multiple levels of structural hierarchy will be a challenge for material scientists. Direct applications of these materials will most likely be in the biomedical area, where peptide elements can serve as recognition elements in targeted drug delivery. Furthermore, the development of truly chemoselective ligation reactions, like the "Staudinger ligation"[49] and "click chemistry"[65], together with the ability to incorporate non-natural amino acids as functional handles in proteins in vivo, will allow the preparation of bioconjugates in a complex biological environment.

This review has shown that the preparation of peptide-containing block copolymers is mostly performed via chemical synthetic methodologies in which the polypeptide blocks are often limited in length or sequence information. The complexity of the materials that can be made is therefore also limited. The preparation of polypeptide blocks by protein engineering allows the build-up of more complex polypeptide sequences. Combining these polypeptides with traditional synthetic polymers is still a relatively unexplored research area.

2.7 References

[1]   K. Matyjaszewski, in *Advances in Controlled/Living Radical Polymerization, Vol. 854*, **2003**, pp. 2.

[2]   H. R. Kricheldorf, *α-Amino acid N-carboxyanhydrides and related heterocycles*, Springer, Berlin, **1987**.

[3]   B. Gallot, *Prog. Polym. Sci.* **1996**, *21*, 1035.

[4]   H. A. Klok, S. Lecommandoux, *Adv. Mater.* **2001**, *13*, 1217.

[5]   H. Schlaad, M. Antonietti, *Eur. Phys. J. E* **2003**, *10*, 17.

[6]   H. A. Klok, J. F. Langenwalter, S. Lecommandoux, *Macromolecules* **2000**, *33*, 7819.

[7]   S. Lecommandoux, M. F. Achard, J. F. Langenwalter, H. A. Klok, *Macromolecules* **2001**, *34*, 9100.

[8]   H. Kukula, H. Schlaad, M. Antonietti, S. Forster, *J. Am. Chem. Soc.* **2002**, *124*, 1658.

[9]   F. Checot, S. Lecommandoux, Y. Gnanou, H. A. Klok, *Angew. Chem. Int. Edit.* **2002**, *41*, 1339.

[10]   F. Checot, S. Lecommandoux, H. A. Klok, Y. Gnanou, *Eur. Phys. J. E* **2003**, *10*, 25.

[11]   T. J. Deming, *Adv. Mater.* **1997**, *9*, 299.

[12]   T. J. Deming, *Nature* **1997**, *390*, 386.

[13]   I. Dimitrov, H. Schlaad, *Chem. Commun.* **2003**, 2944.

[14]   T. Aliferis, H. Iatrou, N. Hadjichristidis, *Biomacromolecules* **2004**, *5*, 1653.

[15]   A. P. Nowak, V. Breedveld, L. Pakstis, B. Ozbas, D. J. Pine, D. Pochan, T. J. Deming, *Nature* **2002**, *417*, 424.

[16]   A. P. Nowak, V. Breedveld, D. J. Pine, T. J. Deming, *J. Am. Chem. Soc.* **2003**, *125*, 15666.

[17]   L. M. Pakstis, B. Ozbas, K. D. Hales, A. P. Nowak, T. J. Deming, D. Pochan, *Biomacromolecules* **2004**, *5*, 312.

[18]   J. N. Cha, G. D. Stucky, D. E. Morse, T. J. Deming, *Nature* **2000**, *403*, 289.

[19]   M. Yu, T. J. Deming, *Macromolecules* **1998**, *31*, 4739.

[20]   M. D. Wyrsta, A. L. Cogen, T. J. Deming, *J. Am. Chem. Soc.* **2001**, *123*, 12919.

[21]   E. G. Bellomo, M. D. Wyrsta, L. Pakstis, D. J. Pochan, T. J. Deming, *Nature Materials* **2004**, *3*, 244.

[22]   L. E. Euliss, S. G. Grancharov, S. O'Brien, T. J. Deming, G. D. Stucky, C. B. Murray, G. A. Held, *Nano Lett.* **2003**, *3*, 1489.

[23]   E. A. Minich, A. P. Nowak, T. J. Deming, D. J. Pochan, *Polymer* **2004**, *45*, 1951.

[24]   T. J. Deming, *Adv. Drug Deliv. Rev.* **2002**, *54*, 1145.

[25]   J. N. Cha, H. Birkedal, L. E. Euliss, M. H. Bartl, M. S. Wong, T. J. Deming, G. D. Stucky, *J. Am. Chem. Soc.* **2003**, *125*, 8285.

[26]   O. Rathore, D. Y. Sogah, *Macromolecules* **2001**, *34*, 1477.

[27]   O. Rathore, D. Y. Sogah, *J. Am. Chem. Soc.* **2001**, *123*, 5231.

[28]   O. Rathore, M. J. Winningham, D. Y. Sogah, *J. Polym. Sci. Pol. Chem.* **2000**, *38*, 352.

[29]   J. M. Gosline, M. E. Demont, M. W. Denny, *Endeavour* **1986**, *10*, 37.

[30]   J. M. Yao, D. Xiao, X. Chen, P. Zhou, T. Yu, Z. Shao, *Macromolecules* **2003**, *36*, 7508.

[31]   L. Ayres, H. H. M. Adams, D. W. P. M. Lowik, J. C. M. van Hest, *Biomacromolecules* **2005**, *6*, 825.

[32]   L. Ayres, M. R. J. Vos, P. Adams, I. O. Shklyarevskiy, J. C. M. van Hest, *Macromolecules* **2003**, *36*, 5967.

[33]   J. Cornelissen, M. Fischer, N. Sommerdijk, R. J. M. Nolte, *Science* **1998**, *280*, 1427.

[34]   M. Bodanszky, A. Bodanszky, *The practise of peptide synthesis*, 2 ed., Springer Verlag, Berlin, **1994**.

[35] T. S. Burkoth, T. L. S. Benzinger, D. N. M. Jones, K. Hallenga, S. C. Meredith, D. G. Lynn, *J. Am. Chem. Soc.* **1998**, *120*, 7655.

[36] T. S. Burkoth, T. L. S. Benzinger, V. Urban, D. G. Lynn, S. C. Meredith, P. Thiyagarajan, *J. Am. Chem. Soc.* **1999**, *121*, 7429.

[37] T. S. Burkoth, T. L. S. Benzinger, V. Urban, D. M. Morgan, D. M. Gregory, P. Thiyagarajan, R. E. Botto, S. C. Meredith, D. G. Lynn, *J. Am. Chem. Soc.* **2000**, *122*, 7883.

[38] A. Rosler, H. A. Klok, I. W. Hamley, V. Castelletto, O. O. Mykhaylyk, *Biomacromolecules* **2003**, *4*, 859.

[39] I. C. Reynhout, D. Lowik, J. C. M. van Hest, J. Cornelissen, R. J. M. Nolte, *Chem. Commun.* **2005**, *5*, 602.

[40] Y. Mei, K. L. Beers, H. C. M. Byrd, D. L. Vanderhart, N. R. Washburn, *J. Am. Chem. Soc.* **2004**, *126*, 3472.

[41] M. L. Becker, J. Q. Liu, K. L. Wooley, *Chem. Commun.* **2003**, 180.

[42] W. Zauner, K. Lingnau, F. Mattner, A. von Gabain, M. Buschle, *Biol. Chem.* **2001**, *382*, 581.

[43] J. P. Tam, *J. Immun. Methods* **1996**, *196*, 17.

[44] N. M. O'Brien-Simpson, N. J. Ede, L. E. Brown, J. Swan, D. C. Jackson, *J. Am. Chem. Soc.* **1997**, *119*, 1183.

[45] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776.

[46] L. E. Canne, P. Botti, R. J. Simon, Y. J. Chen, E. A. Dennis, S. B. H. Kent, *J. Am. Chem. Soc.* **1999**, *121*, 8720.

[47] P. Botti, M. R. Carrasco, S. B. H. Kent, *Tetrahedron Lett.* **2001**, *42*, 1831.

[48] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6554.

[49] M. Kohn, R. Breinbauer, *Angew. Chem. Int. Edit.* **2004**, *43*, 3106.

[50] J. A. Borgia, G. B. Fields, *Trends Biotechnol.* **2000**, *18*, 243.

[51] G. G. Kochendoerfer, S. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884.

[52] K. Rose, J. Vizzanova, *J. Am. Chem. Soc.* **1994**, *121*, 30.

[53] J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*, 2 ed., Cold Spring Laboratory Press, Cold Spring, **1989**.

[54] S. R. Fahnestock, *Biopolymers: Polyamides and Complex Proteinaceous Materials II, Vol. 8*, Wiley-VCH Gmbh & Co. KGaA, Weinheim, **2003**.

[55] J. C. M. van Hest, D. A. Tirrell, *Chem. Commun.* **2001**, 1897.

[56] F. Ferrari, J. Cappello, in *Protein-based materials* (Eds.: K. P. McGrath, D. L. Kaplan), Birkhauser, Boston, MA, **1997**, pp. 37.

[57] J. Cappello, J. W. Crissman, M. Crissman, F. A. Ferrari, G. Textor, O. Wallis, J. R. Whitledge, X. Zhou, D. Burman, L. Aukerman, E. R. Stedronsky, *J. Control. Release* **1998**, *53*, 105.

[58] T. A. T. Lee, A. Cooper, R. P. Apkarian, V. P. Conticello, *Adv. Mater.* **2000**, *12*, 1105.

[59] A. Chilkoti, M. R. Dreher, D. E. Meyer, *Adv. Drug Deliv. Rev.* **2002**, *54*, 1093.

[60] E. R. Wright, V. P. Conticello, *Adv. Drug Deliv. Rev.* **2002**, *54*, 1057.

[61] D. W. Urry, *J. Phys. Chem. B* **1997**, *101*, 11007.

[62] A. Nicol, D. C. Gowda, T. M. Parker, D. W. Urry, *Biotechnology of Bioactive Polymers*, Plenum Press, New York, **1994**.

[63]   A. Panitch, T. Yamaoka, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1999**, *32*, 1701.

[64]   W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389.

[65]   Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192.

Chapter 3

Construction and expression of artificial genes encoding repetitive β-sheet polypeptides

Abstract

The construction of artificial genes coding for the β-sheet polypeptides with the repetitive sequence $[(AG)_3EG]_n$ (n = 10 – 50; A = alanine, G = glycine and E = glutamic acid) and their expression in E. coli is described.

Whereas expression of poly-$[(AG)_3EG]$ with a minimal number of additional amino acids was unsuccessful, the polypeptides were effectively produced as a fusion with glutathione-S-transferase (GST, 26 kDa). The shake flask expression of the fusion protein GST-$[(AG)_3EG]_{20}$ resulted in a yield of approximately 80 mg per liter of culture after GST affinity purification. After proteolytic cleavage of the fusion protein with thrombin, the polypeptide $[(AG)_3EG]_{20}$ was separated from GST by reversed phase chromatography, resulting in a final yield of approximately 20 mg per liter of culture. In an alternative approach $[(AG)_3EG]_{20}$ was produced with a small (11 amino acids) N-terminal "T7-tag", which gave a yield of approximately 20 mg per liter of culture after Ni-NTA chromatography. Expression of this construct resulted in the formation of truncated polypeptide, which could be removed by reversed phase chromatography.

In addition, the expression of a polypeptide with the repetitive sequence $[(AG)_3KG]_{24}$ (K = lysine) is described. The production of this polypeptide was, however, complicated by the sensitivity of this sequence to proteolysis. The polypeptide had to be purified directly after expression by Ni-NTA chromatography. Polypeptides with the complete repetitive part were separated from the truncated polypeptides by gel filtration chromatography, resulting in a final yield of approximately 2 mg per liter of culture.

3.1 Introduction

The design and synthesis of artificial, repetitive polypeptides is an area of research which is important for materials science. In nature, structural proteins such as silks, elastin and collagen consist largely of highly repetitive sequences. Recombinant DNA technology and bacterial protein expression have been used for the biosynthesis of repetitive polypeptides based on sequences found in these structural proteins[1-7] as well as for the production of artificial polypeptides without a direct parallel in nature.[8-18]

However, the first large, repetitive polypeptides that were produced in *E. coli*, were not intended for application in materials. Doel et al. constructed a synthetic gene coding for multiple repeats of the dipeptide L-aspartyl-L-phenylalanine, which is the basis of an artificial sweetener.[19] The dipeptide could be released by proteolytic digestion. A number of other, larger peptides have been produced as tandem repeats in *E. coli* with the aim to release the monomeric form afterwards by chemical or proteolytic cleavage.[20-23] For example, tandemly linked proinsulin could be produced in a stable form, whereas the monomeric form was subject to proteolytic degradation.[20]

With an increased interest in the recombinant production of fibrous proteins, the number of reports describing the construction and expression of repetitive, synthetic genes increased. The natural genes for fibrous proteins derived from cDNA are long and repetitive and their corresponding protein products have a restricted amino acid composition. The propagation and expression of such genes in *E. coli* can present typical problems.[3, 24-26] Firstly, maintenance of repetitive DNA sequences on plasmid vectors can result in deletions or to a lesser degree elongation of the genes. The second difficulty results from premature translation termination as a consequence of a limiting stock of appropriate aminoacyl-tRNAs. These problems can be alleviated by the construction of synthetic genes with a codon usage adapted to *E. coli* and a reduced repetitiveness at the genetic level, as was, for example, demonstrated in the production of *Nephila clavipes* silk fibroins.[3, 24]

The construction of synthetic genes has, however, not only been carried out with the purpose to faithfully reproduce the natural gene products. In addition, low-complexity protein polymers have been prepared which consist of tandem repeats of small peptide sequences found in structural proteins as well as "de novo" designed peptide sequences. A few examples are provided in Table 3.1. The construction of synthetic genes gives the opportunity to synthesize new protein structures or improve the functionality or processability of existing structural proteins beyond what is offered by the natural gene product.

The biosynthesis of these protein polymers requires the construction of large synthetic genes by enzymatic multimerization of double-stranded synthetic oligonucleotide sequences. A variety of strategies for an efficient generation of such multimeric genes have been reported.[13, 23, 27-30] Expression of these genes has been carried out most effectively using the "pET expression system".[31] In this system the genes are placed under the control of the bacteriophage T7 promoter and transcription is carried out by the highly processive T7 RNA

polymerase.[32] High yields have been obtained for some protein polymers, such as poly(L-alanylglycine), elastin-like polypeptides and sea mussel bioadhesive precursor protein (see Table 3.1), despite the incongruity of the proteins to the endogenous proteins of *E. coli* in terms of amino acid composition.

Table 3.1 *Some examples of protein polymers prepared via E. coli expression.*

Protein polymer type	Amino acid sequence	Yield[a] (mg L^{-1})	Properties	Ref.
Spider dragline silk-like polymer	(PGGYGPGQQG PGGYGPGQQG PSGPGSAAAAA AAAG)$_{16}$	4	β-sheet crystalline fibers	[2]
Elastin-like polypeptide (ELP)	(VPGVG)$_{251}$	60, 200-800[b]	Elastic chains assemble above phase transition temperature	[4, 5]
Sea mussel bioadhesive precursor protein	(KAKPSYPPTY)$_{20}$	300	Precursor for water-compatible adhesive	[7]
Poly[(AG)$_3$PEG]	[(AG)$_x$PEG]$_n$ (x = 3 - 4; n = 10 − 54)	10	Amorphous glass	[8, 33]
Poly(L-alanylglycine)	(AG)$_{240}$	50, 1000[b]	β-sheet silk-like	[9]
Poly[(AG)$_3$XG]	[(AG)$_x$XG]$_n$ (X = E, F, N, S, V, Y; x = 3 – 6; n = 13 - 59)	10 - 40	β-sheet silk-like forming needle-shaped lamellar crystals	[10-12]
Poly(EAK)	(AEAEAKAK)$_{18}$	5	Amyloid-like stable fibrils	[13, 14]
Alanine-rich helices	[(AAAQ)$_x$(AAAE)(AAA)$_x$]$_n$ (n = 2 – 6)	5 - 10	Water-soluble helices for controlled placement of carbohydrate ligands	[15, 16]
Poly(L-glutamic acid)	[(E)$_{17}$D]$_n$ (n = 3 – 6)	5 – 10[c]	Rod-like polymer with smectic ordering after side-chain benzylation	[17, 18]

[a] Typical expression yield in shake flask experiments. [b] Yield for high cell density fermentation.
[c] Yield of fusion protein (21.5 kDa N-terminal mouse dihydrofolate reductase).

The expression of some protein polymers has been more troublesome. For example, the biosynthesis of poly(L-glutamic acid) (PLGA) was complicated by the resemblance of the coding sequence to the consensus Shine-Dalgarno (SD) sequence found at translation initiation sites in bacterial mRNA.[34] Induction of PLGA mRNA, containing multiple noninitiator SD-like sequences, resulted in inhibition of protein synthesis by complex formation with 16S and 23S rRNA. For the protein polymer with the sequence [(AG)$_4$PEG]$_{14}$, the formation of truncated products was observed and explained by the action of peptidases.[33]

In this chapter the construction of artificial, repetitive genes coding for poly-[(AG)$_3$EG] and approaches to produce this polypeptide in *E. coli* are described. Poly-[(AG)$_3$EG] is known to crystallize into antiparallel β-sheets with glutamic acid residues located at the turn positions (see Chapter 1). The glutamic acid residues could be useful for functionalization with e.g. mesogenic groups resulting in a material with liquid crystalline properties. In addition, the expression of poly-[(AG)$_3$KG] is described. Chemical modification of the lysine residues of this polypeptide is expected to be more efficient.

3.2 Results and discussion

3.2.1 Construction of artificial genes coding for poly-[(AG)$_3$EG]

The cloning strategy for the construction of repetitive genes encoding poly-[(AG)$_3$EG] is depicted in Figure 3.1. A 48 basepair synthetic oligonucleotide was used which encoded two copies of the octapeptide repeating unit (Figure 3.1a). The sequence of this oligo was designed with the following considerations: (1) avoidance of codons with a low frequency of occurrence in *E. coli*[35], (2) minimization of sequence repetitiveness and (3) use of non-palindromic overhangs for unidirectional multimerization by T4 DNA ligase. A cloning vector was used containing two inversely oriented recognition sites for the restriction enzyme *Bsp*MI (Figure 3.1b).[27] Digestion with this enzyme resulted in cleavage at a single position (indicated with the arrows) and overhangs were generated which were compatible with the introduction of the synthetic oligonucleotide. After insertion of this oligonucleotide and verification of its sequence, the plasmid was digested with *Bsp*MI, followed by isolation of the monomer. After multimerization of the monomer by T4 DNA ligase, the multimer was ligated with the linearized cloning vector. The same procedure was repeated until multimers of sufficient length were obtained (Figure 3.1c). Multimers consisting of up to 5 repeating units were visible on DNA gel (Figure 3.2a) after the first multimerization step. The 5-mer was extracted from gel and ligated with the linearized vector. After plasmid amplification and *Bsp*MI digestion, a larger amount of 5-mer was extracted from gel and a second multimerization was carried out. The linearized plasmid was added to this multimerization reaction and resulted in a series of plasmids, carrying synthetic genes coding for 10 to 50 repeating units of -(AG)$_3$EG- (240 to 1200 basepairs in size) (Figure 3.2b). These multimeric genes were transferred to the pET-3b expression vector using the *Nde*I and *Bam*HI restriction sites.
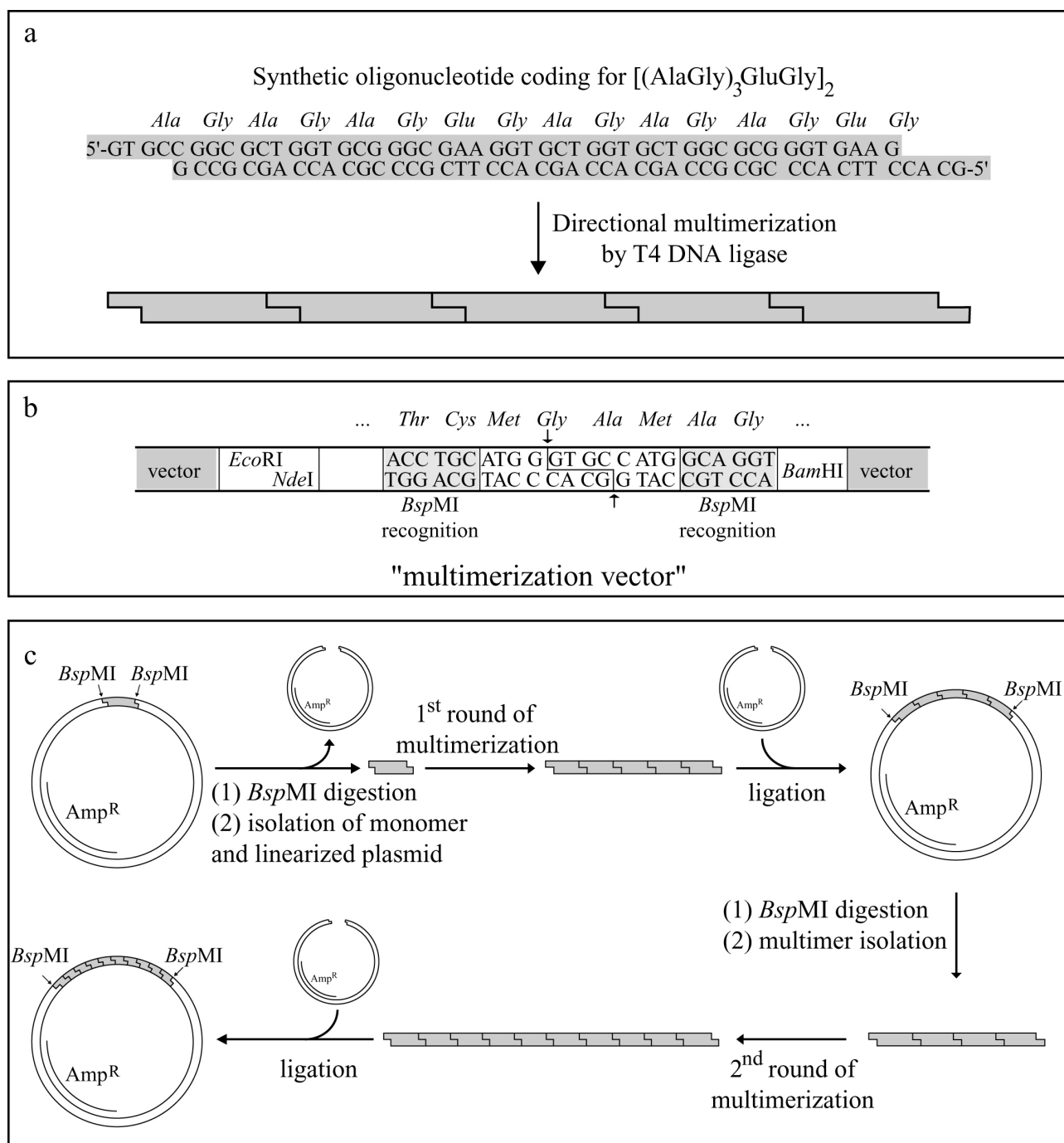
Figure 3.1 *Cloning strategy for the construction of repetitive genes encoding poly-[(AG)₃EG]. **(a)** Multimerization of synthetic oligonucleotides coding for 2 octapeptide repeating units. **(b)** Cloning cassette for the introduction of multimeric genes. Two inversely oriented BspMI recognition sites are positioned to allow cleavage to occur at a single position (indicated with the arrows). NdeI and BamHI sites are used for transfer to the pET-3b expression vector. **(c)** Ligation of multimers with BspMI digested cloning vector. After amplification of the resulting plasmid, the multimers can be isolated by BspMI digestion and a second round of multimerization can be carried out.*
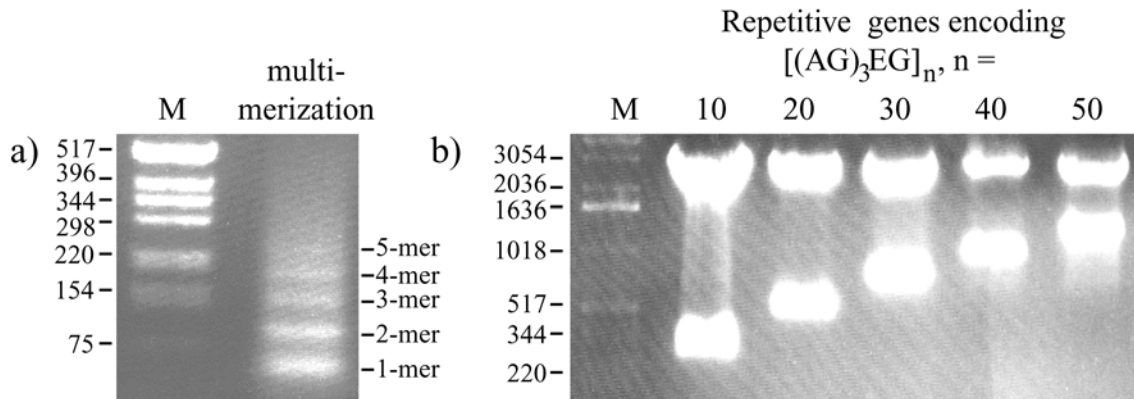
Figure 3.2 (a) *Multimerization of 48 basepair synthetic oligonucleotide analyzed on 2.5% agarose gel.* ***(b)*** *BspMI restriction analysis of plasmids carrying synthetic genes coding for 10 to 50 repeating units of -(AG)₃EG- (240 to 1200 basepairs).*

3.2.2 Expression and purification of His-tagged and GST-fused poly-[(AG)₃EG]

Approach 1: Poly-[(AG)₃EG] with a minimal number of additional amino acids. In this first approach the cloning strategy was designed with the aim to obtain a polypeptide in which the desired β-sheet sequence was flanked with a minimal number of amino acids (Figure 3.3). The constructed expression vector coded for a target polypeptide with an N-terminal $6 \times$ histidine tag for purification by immobilized metal affinity chromatography. Furthermore, the repetitive polypeptide part was flanked on both sides by two methionine residues, which should enable the removal of the $6 \times$ histidine tag and the C terminus by cyanogen bromide cleavage.[35] Cleavage at the carboxyl-terminal side of methionine should leave only one and two extra amino acids at the N- and C-terminal side of the repetitive polypeptide part, respectively.
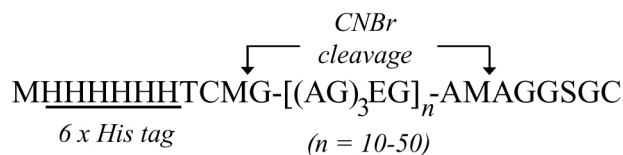
Figure 3.3 *Target sequence of poly-[(AG)₃EG] containing a minimal number of non-repetitive amino acids. An N-terminal $6 \times$ histidine tag was included for purification by immobilized metal affinity chromatography. Methionine residues flank the β-sheet element for removal of the $6 \times$ histidine tag and the residual amino acids at the C terminus by cyanogen bromide (CNBr) cleavage.*

Expression experiments were carried out in *E. coli* strains BL21(DE3) and BL21(DE3)pLysS cells and conditions were varied with respect to temperature (30 and 37 °C), inducer concentration (0.1 – 1 mM IPTG; IPTG = isopropyl-β-D-thiogalactopyranoside) and induction times (1 – 24 hours). However, no expression of target polypeptides with 10 to 50 repeats of the β-sheet sequence, and ranging in calculated molecular weight from 7.7 to 30.5 kDa, was

observed in whole cell lysates analyzed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining. Also after native and denaturing purification by nickel-nitrilotriacetate (Ni-NTA) affinity chromatography no bands corresponding to expressed polypeptide were detected using either Coomassie or silver staining (data not shown).

Approach 2: Poly-[(AG)₃EG] as GST-fusion protein. As a second attempt the expression of poly-[(AG)₃EG] as a fusion with glutathione-S-transferase (GST) was undertaken. This 26 kDa protein can be highly expressed in *E. coli*[36] and is therefore a good way to assess the effect of fusing the poly-[(AG)₃EG] sequence to this protein on the expression level. Therefore, the repetitive genes developed for approach 1 were transferred to the pGEX-2T expression vector (described in the experimental section 3.4.3), which directs the expression of poly-[(AG)₃EG] with GST at the N terminus (Figure 3.4a). The presence of a thrombin cleavage site allowed removal of the GST-protein from the poly-[(AG)₃EG] sequence. Furthermore, cyanogen bromide could be used to cleave off the 6 × His tag and the C terminus. Purification of GST-fusion proteins was performed using a sepharose matrix functionalized with glutathione, which is a tripeptide with high affinity for the binding site of GST. Expression of the repetitive polypeptides from the pGEX-2T plasmids in *E. coli* strain BL21 was successful for all the poly-[(AG)₃EG] variants, from 10 to 50 repeats in length (Figure 3.4b).

Figure 3.4 (a) *Target protein consisting of poly-[(AG)₃EG] fused with the amino-terminal glutathione-S-transferase (GST-protein). The GST-protein can be removed by cleavage with the site-specific protease thrombin. Cyanogen bromide (CNBr) can cleave off the residual amino acids.* ***(b)*** *SDS-PAGE analysis of expression of GST-fused poly-[(AG)₃EG] in E. coli BL21 cells, before (0) and 3 hours (3h) after induction with IPTG. Cultures were grown at 30 °C, induced by the addition of IPTG to a final concentration of 0.4 mM and incubated for an additional 3 hours. Proteins were visualized by Coomassie staining.*

Expressed protein bands appeared 3 hours after induction with increased molecular weights for longer poly-[(AG)$_3$EG] sequences. All fusion proteins appeared at the expected molecular weight (for n = 10 - 50 molecular weights are 35 - 58 kDa) and the attachment of longer poly-[(AG)$_3$EG] sequences did not seem to have a major effect on expression yield.

The GST-fusion proteins were purified with GST-affinity chromatography followed by thrombin cleavage to remove the GST protein. This is shown for GST-[(AG)$_3$EG]$_{20}$ and GST-[(AG)$_3$EG]$_{40}$ in Figure 3.5a. After thrombin cleavage a band corresponding to GST at the 28 kDa molecular weight marker appeared, together with an additional protein band which corresponded with poly-[(AG)$_3$EG]. The expected molecular weights of poly-[(AG)$_3$EG] for n = 20 and 40 were 14.3 and 25.7 kDa, respectively. It is clear from Figure 3.5a, however, that these proteins were migrating slower on an SDS-polyacrylamide gel than would be expected from their molecular weights. This has been reported before for this sequence and for other highly acidic proteins.[8, 10]
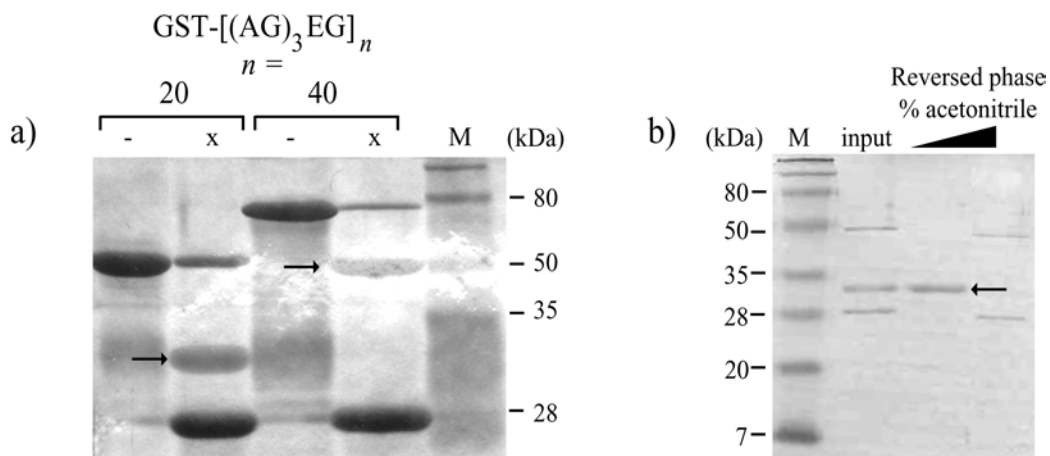
Figure 3.5 *SDS-PAGE analysis of **(a)** purified GST-[(AG)₃EG]ₙ with n = 20 and 40 before (-) and after (×) thrombin cleavage and **(b)** reversed phase chromatography of thrombin cleaved GST-[(AG)₃EG]₂₀. The [(AG)₃EG]₂₀ and [(AG)₃EG]₄₀ products after thrombin cleavage are indicated with an arrow. Proteins were visualized by Coomassie staining.*

Expression of GST-[(AG)$_3$EG]$_{20}$ on a 2.5 liter scale resulted in approximately 200 mg polypeptide after purification with glutathione-sepharose column chromatography. The β-sheet polypeptide was cleaved from the GST-fusion protein by proteolytic cleavage using thrombin. Because this cleavage was incomplete, the β-sheet polypeptide had to be purified from both GST as well as GST-fusion protein. This could be performed by reversed phase chromatography, since the β-sheet polypeptide eluted at a lower concentration acetonitrile (~ 5%) than the GST and GST-fusion proteins (which eluted simultaneously at ~ 25% acetonitrile). The purification procedure resulted in approximately 50 mg of the desired polypeptide after freeze-drying. Amino acid compositional analysis showed peaks for the major amino acids alanine, glycine and glutamic acid and their ratios agreed well with the expected

values (calculated amino acid analysis for [(AG)$_3$EG]$_{20}$ (mol percent): Ala, 37.1; Gly, 50.9; Glu, 12.0. Values found: Ala, 37.3; Gly, 49.4; Glu, 13.3). The chemical composition of the polypeptide was furthermore confirmed by ^{1}H-NMR spectroscopy (Figure 3.6).
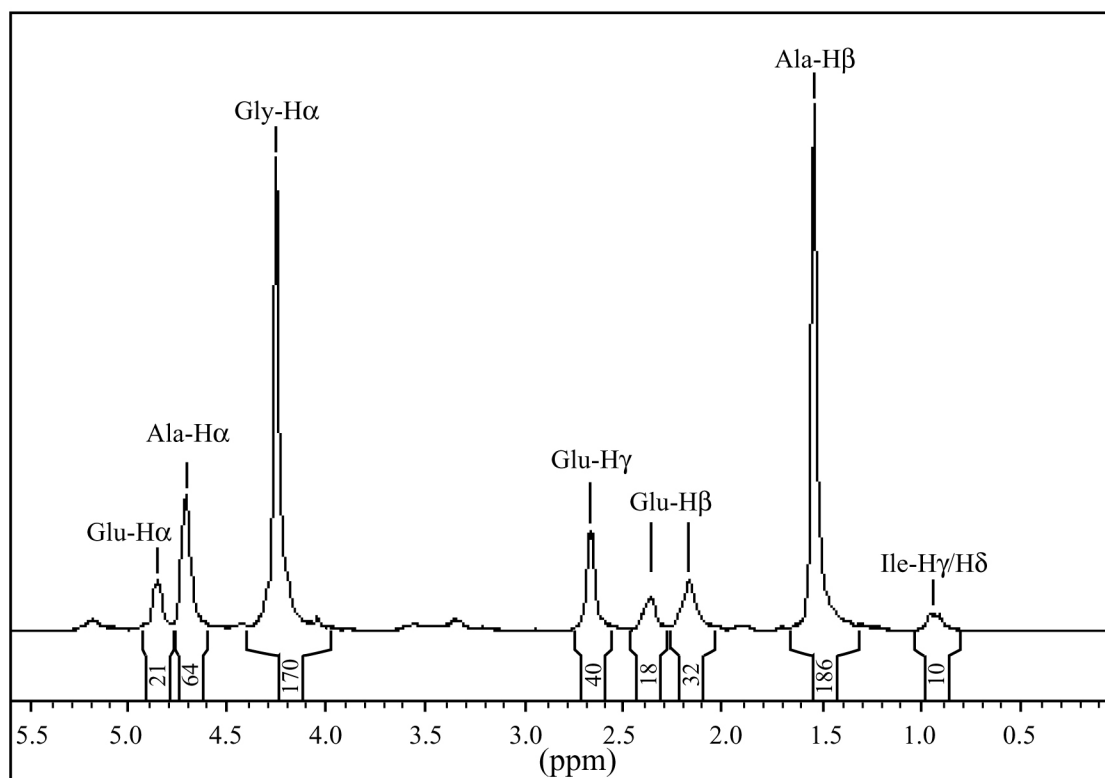
Figure 3.6 *400 MHz ^{1}H-NMR spectrum of [(AG)$_3$EG]$_{20}$ polypeptide in TFA-d obtained after thrombin cleavage of the corresponding GST-fusion.*

Cyanogen bromide cleavage of the polypeptide was carried out in 70% formic acid. After cyanogen bromide cleavage, the ability to stain the polypeptide by Coomassie or silver staining was lost.

The MALDI-TOF spectra of the polypeptide before and after cyanogen bromide cleavage are depicted in Figure 3.7a and b, respectively. Before CNBr cleavage, the major peak at 14.6 kDa is somewhat higher than the expected molecular mass of 14.3 kDa and could indicate the formation of an adduct of the polypeptide with glutathione (which was used for elution of the polypeptide from the GST column) via a disulfide bridge with the single cysteine residue. The calculated mass of this adduct is 14625 Da. Ionization of the polypeptide after cyanogen bromide cleavage seemed to be more difficult, but the molecular mass of the major peak at 11706 was slightly larger than the expected mass of 11659 Da (calculated average mass for H-G-[(AG)$_3$EG]$_{20}$-A-homoserine-OH).
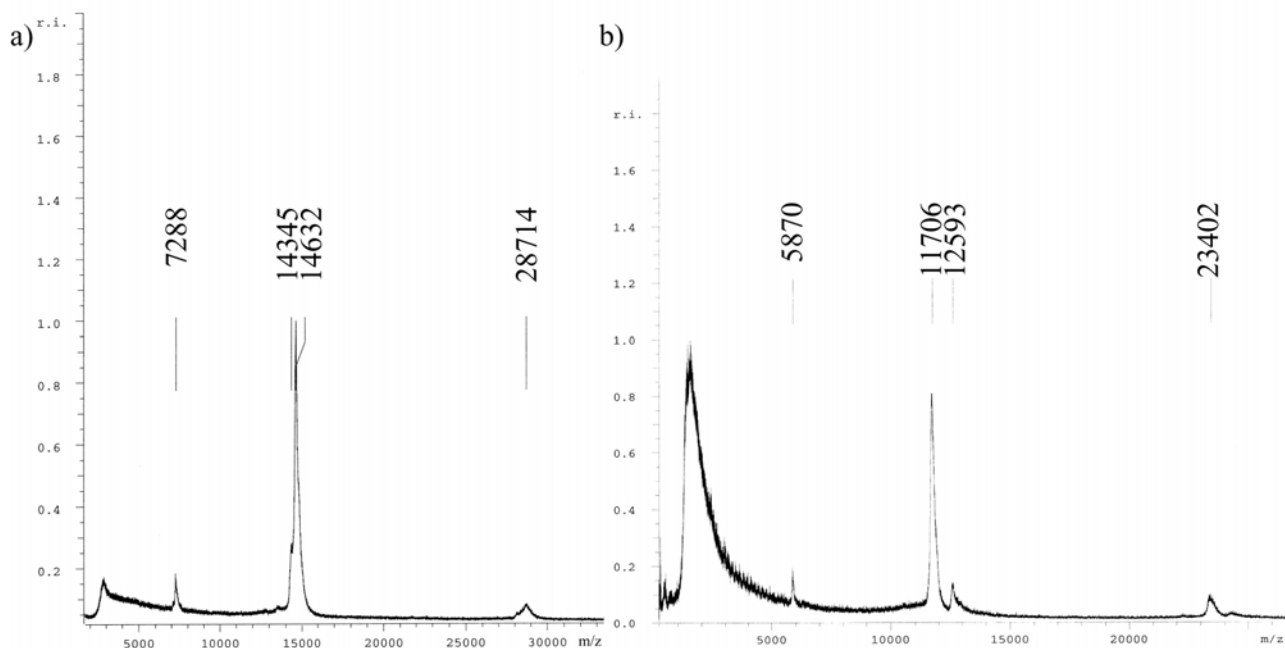
Figure 3.7 *MALDI-TOF spectra of [(AG)₃EG]₂₀ polypeptide after **(a)** thrombin cleavage and **(b)** after cyanogen bromide cleavage.*

In conclusion, we can state that we successfully expressed poly(AG)₃EG up to 50 repeats as a fusion with glutathione-S-transferase. Purification of [(AG)₃EG]₂₀ by GST affinity purification, thrombin cleavage and reversed phase chromatography was demonstrated. The secondary structure characterization of this polypeptide is described in Chapter 5.

Although the expression of poly-[(AG)₃EG] was successful as GST-fusion, the use of the rather expensive thrombin was not very compatible with the desire to produce this polypeptide in larger quantities. Omitting the thrombin cleavage step and instead directly cleaving the GST-fusion protein with cyanogen bromide was attempted, but purification of poly-[(AG)₃EG] from the multiple peptide fragments was unsuccessful. Therefore it was decided to return to the pET-expression system and remake the construct with the smaller T7 tag at the N terminus and additional amino acids at the C terminus. Since it was shown for cyanogen bromide cleaved [(AG)₃EG]₂₀ that Coomassie or silver staining was not possible, the additional amino acids could be beneficial for staining and the T7 tag would furthermore be useful for visualization of expression by Western blotting, since T7 antibodies were available.

Approach 3: Poly-[(AG)₃EG] with an N-terminal "T7 tag". For the expression of poly-[(AG)₃EG] with an N-terminal T7 tag, the same pET-3b vector as in approach 1 was used, but the repetitive genes were introduced in an alternative way (described in experimental section 3.4.3), which resulted in a target polypeptide with a sequence as is depicted in Figure 3.8a. The expression of these polypeptides was successful, as is depicted in Figure 3.8b. The theoretical molecular weights for [(AG)₃EG]ₙ for n is 10, 20 and 30 are 10.9, 16.6 and 22.3 kDa,

respectively. The apparent molecular weights on SDS-polyacrylamide gel were much higher and were comparable with the migration behaviour of the β-sheets which were prepared after thrombin cleavage of the GST-fusions (Figure 3.5), as would be expected. Purification of $[(AG)_3EG]_{20}$ was performed with Ni-NTA chromatography under denaturing conditions using guanidine hydrochloride. Three bands were detected in the elution fraction (Figure 3.8c).
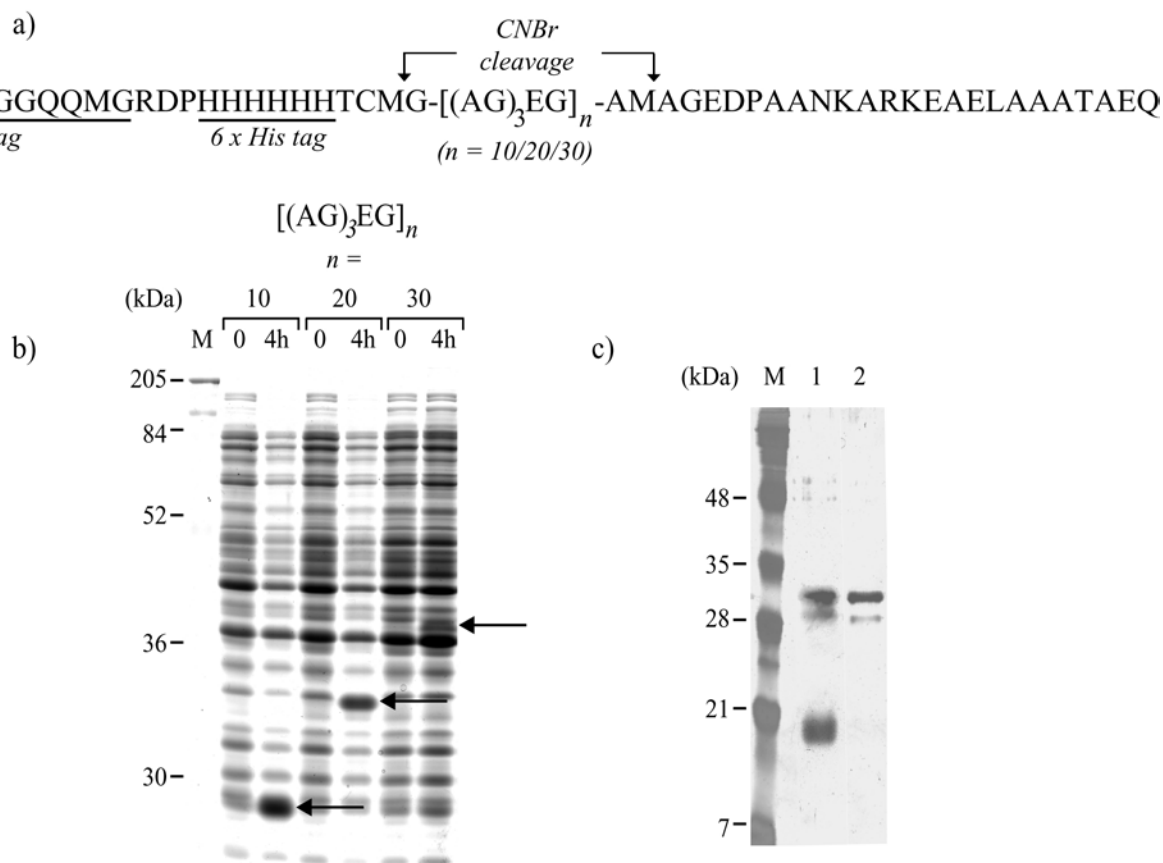
Figure 3.8 (a) Amino acid sequence of poly-[(AG)₃EG] β-sheets. Cyanogen bromide (CNBr) cleavage allows removal of the residual amino acids introduced in the cloning process. (b) Expression of [(AG)₃EG]ₙ (n = 10, 20 and 30) in E. coli BL21(DE3)pLysS, before (0) and 4 hours (4h) after induction with IPTG. The arrows indicate the formation of the β-sheet proteins. (c) [(AG)₃EG]₂₀ after denaturing Ni-NTA purification (lane 1) and reversed phase chromatography (lane 2).

MALDI-TOF mass spectrometry showed that the upper band corresponded to full-length polypeptide (Figure 3.9a). The observed mass of 16748 Da was somewhat higher than the calculated average mass of 16459 Da. This could indicate the formation of a glutathione adduct via the one cysteine residue (calculated average mass is 16764 Da). The polypeptide migrating at the height of the 20 kDa marker was removed by reversed phase chromatography. This band most likely corresponded to truncated polypeptide. The truncated nature of the band at 20 kDa was confirmed for a variant of the $[(AG)_3EG]_{20}$ sequence carrying N- and C-terminal cysteine residues and is described in Chapter 4. The final yield of the purified product was

approximately 20 mg per liter of cell culture. Cyanogen bromide cleavage was carried out to remove the non-repetitive N and C termini. MALDI-TOF analysis showed that the major product was completely cleaved polypeptide (calculated average mass is 11659 Da for H-G-[(AG)$_3$EG]$_{20}$-A-homoserine-OH). However, in addition, partially cleaved product was present (Figure 3.9b).
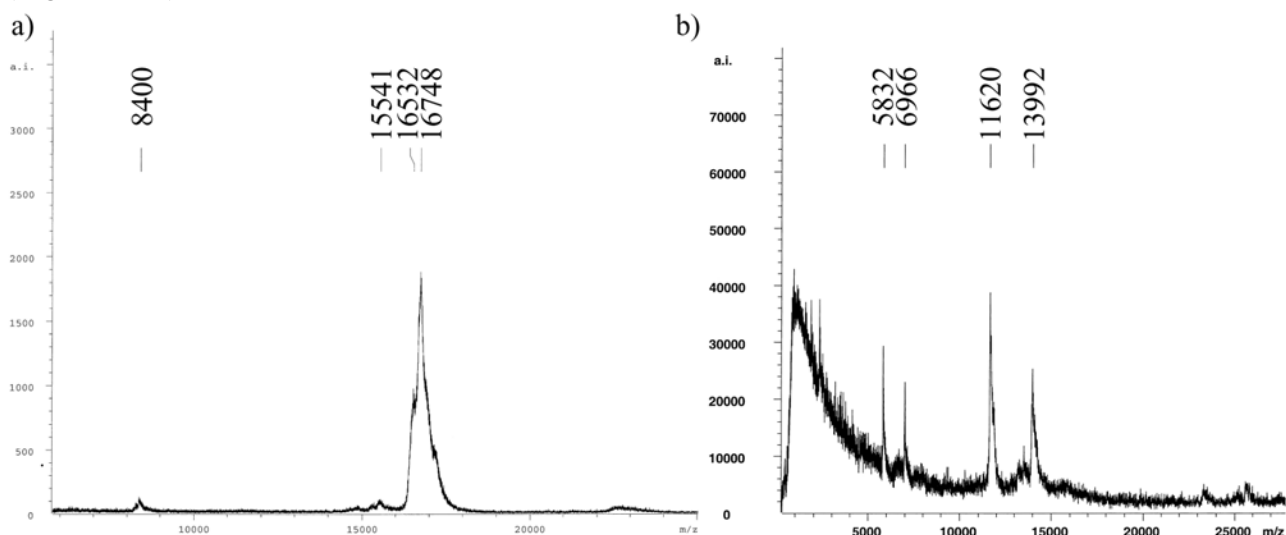
Figure 3.9 *MALDI-TOF spectra of [(AG)$_3$EG]$_{20}$ polypeptide **(a)** before and **(b)** after cyanogen bromide cleavage.*

The difference in expression results reported for the two pET expression vectors (approach 1 and 3) used in this research is likely to be the result of the presence of an N-terminal T7 tag in the successful construct. This 11 amino acids leader sequence is derived from the bacteriophage T7 gene 10 which codes for the highly expressed major capsid protein. It has been proposed by Sprengart and coworkers that in messenger RNA from genes 0.3 and 10 of bacteriophage T7, a specific region located downstream (+15 to +26) of the initiation codon serves as an independent translational signal, besides the upstream Shine-Dalgarno sequence.[37, 38] This region of the mRNA sequence has been termed "downstream box" and it has been shown that an increasing complementarity of this sequence to the penultimate stem of the 16S ribosomal RNA resulted in a significant increase (10 to 20-fold) in protein accumulation.[39, 40] For the expression vector without the T7 tag, 6 consecutive histidine codons are positioned directly downstream of the initiation codon, which might result in problems with translation initiation. However, no experiments were performed to confirm this. Although the T7 tag is not an absolute necessity for expression, it has been reported that, at least, it promotes protein expression.[41]

In conclusion, poly-[(AG)$_3$EG] can be expressed as a GST-fusion as well as with the small T7 gene 10 leader sequence. The final yields for [(AG)$_3$EG]$_{20}$ were comparable (20 mg L^{-1}) for both systems. The expression system based on T7 RNA polymerase (approach 3) was considered to be more straightforward for large-scale expression and purification, since no

enzymatic cleavage step is necessary. A complication in this system was, however, the appearance of a truncated polypeptide fraction, which was not observed for the GST-fusions.

3.2.3 Expression and purification of poly-[(AG)₃KG]

Repetitive polypeptides in which the alanylglycine repeats are interrupted by lysine (K) residues were expressed analogous to approach 3 for poly-[(AG)$_3$EG]. An expression vector was used which contained 24 repeats of (AG)$_3$KG (kindly provided by J. Thies and D. Tirrell). The target sequence of this polypeptide is depicted in Figure 3.10a. SDS-PAGE analysis of the induction of protein expression for [(AG)$_3$KG]$_{24}$ in *E. coli* BL21 (DE3) cells is depicted in Figure 3.10b. A new band appeared after IPTG induction with an apparent molecular weight of 26 kDa, which is higher than the calculated molecular weight of 19.0 kDa. In addition, lower molecular weight bands were observed on gel, which could be distinguished from *E. coli* proteins due to their specific purple color after Coomassie staining. After 2 hours of protein expression the amount of presumed full-length product no longer increased and the laddering at lower molecular weights became more apparent, suggesting that degradation already took place during expression. Because product degradation was observed during storage of the cell pellet at -20 °C, Ni-NTA purification had to be carried out directly after cell harvesting. The purity and extent of degradation after native Ni-NTA affinity chromatography in the presence of the protease inhibitor phenylmethanesulphonyl fluoride (PMSF) is shown in Figure 3.10c. Alternatively, denaturing purification using 6M guanidine hydrochloride was carried out (not shown), which however did not diminish the amount of degradation. Figure 3.10d shows that the majority of the truncated products could be removed by gel filtration chromatography, using a Superdex-75 column. The final product yield after dialysis and freeze-drying was however low with approximately 2 mg per liter of cell culture.

MALDI-TOF analysis of the final product indicated two peaks at a molecular weight of 17.7 and 17.9 kDa (Figure 3.11a), which is lower than the calculated molecular weight of 19.0 kDa for the full-length polypeptide. Since the N terminus was present as determined by immunoblot analysis (data not shown), apparently part of the C terminus was not. Based on these findings, the observed molecular weight with MALDI-TOF seems to correspond to cleavage after the last lysine residue. For removal of the non-repetitive residues cyanogen bromide cleavage was carried out (Figure 3.10e). The presence of formic acid resulted already in partial cleavage of the polypeptide, which occurred presumably at the acid-sensitive aspartyl-prolyl (DP) bond.[42] In the presence of cyanogen bromide, two bands were formed on SDS-PAGE gel with the major product located just below the 20 kDa marker.

a)

cyanogen bromide
cleavage sites

MASMTGGQQMGRDPGHHHHHHMG-[(AG)$_3$KG]$_{24}$ -AMSRPELDPAANKARKEAELAAATAEQ

T7 tag   6 x His tag
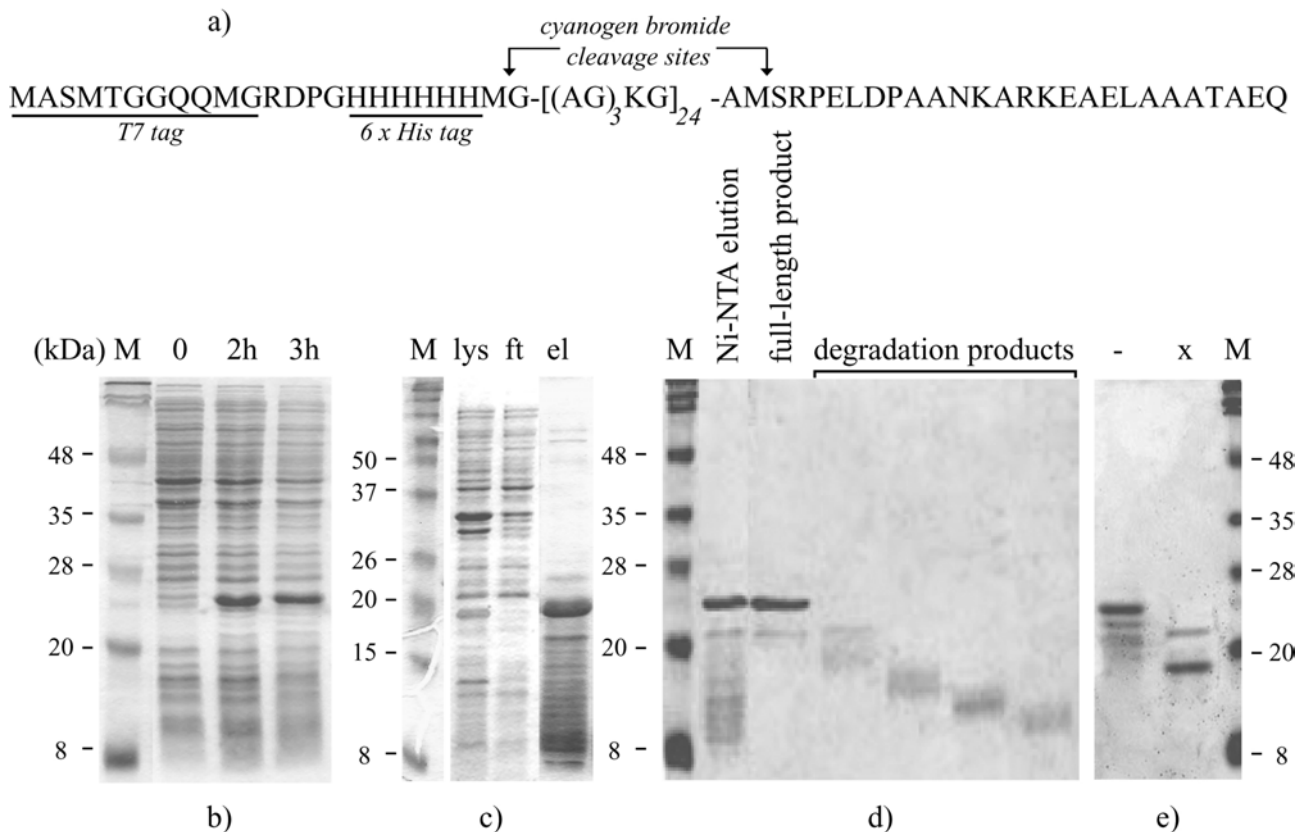
Figure 3.10 (a) *Target sequence of [(AG)$_3$KG]$_{24}$. SDS-polyacrylamide analysis of **(b)** Expression of [(AG)$_3$KG]$_{24}$, before and 2, respectively 3 hours after induction with IPTG. **(c)** Native Ni-NTA purification (lys = soluble lysate, ft = column flow-through, el = column elution fraction). **(d)** Separation of full-length product and degradation products by Superdex-75 gel filtration chromatography. **(e)** Cyanogen bromide cleavage of [(AG)$_3$KG]$_{24}$. (-) and (×) correspond to polypeptide after incubation for 2 days in 70% formic acid without and with cyanogen bromide, respectively. Proteins were visualized by Coomassie staining.*

MALDI-TOF analysis showed a peak with a molecular mass of 15.4 kDa (Figure 3.11b), which is higher than the calculated mass of 13.9 kDa for completely cleaved product. The observed molecular mass indicates that CNBr cleavage at the C-terminal methionine residue was not successful. This is likely to be the result of the presence of a serine residue at the C-terminal side of this methionine, which is known to prevent cleavage of the peptide backbone and instead results in conversion of methionine to homoserine.[43] The susceptibility of poly-[(AG)$_3$KG] for proteolytic degradation, prohibited a large-scale purification of this polypeptide. This difference in comparison with poly-[(AG)$_3$EG], combined with the regular ladder of cleaved products observed on SDS-PAGE gel suggests cleavage at the lysine residues.
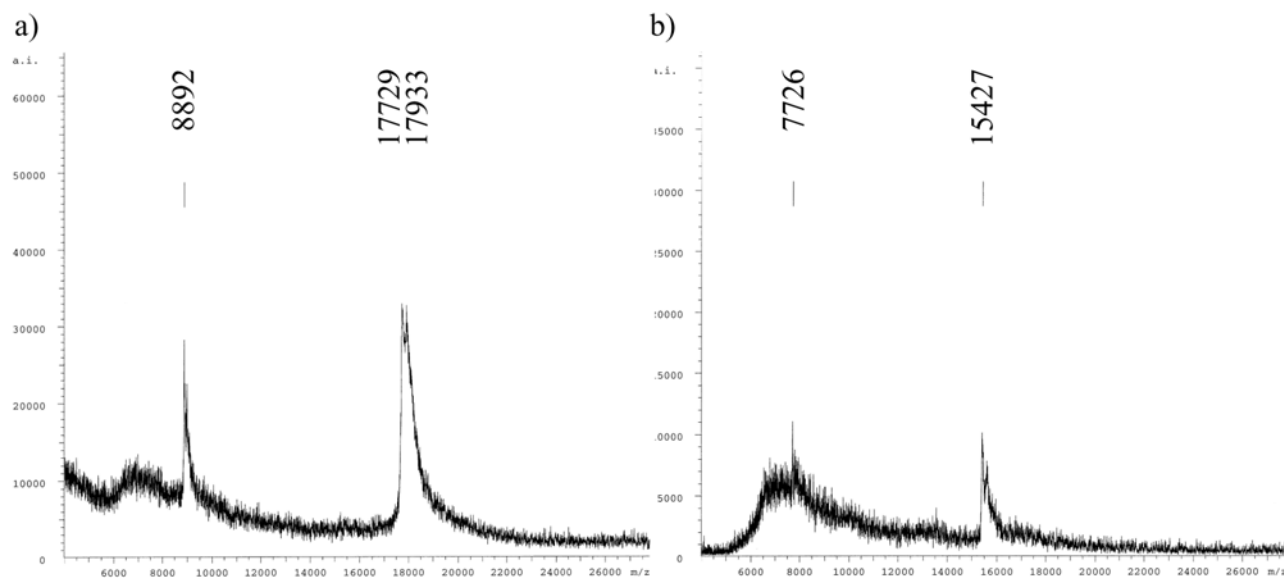
Figure 3.11 *MALDI-TOF spectra of [(AG)₃KG]₂₄ polypeptide (a) before and (b) after cyanogen bromide cleavage.*

A possible candidate responsible for proteolytic degradation of poly-[(AG)₃KG] is oligopeptidase B, a cytoplasmic protease which cleaves at the C-terminal side of lysine and arginine residues.[44] Alternatively, degradation may be caused by energy-dependent heat shock proteins (with the exception of the cytoplasmic Lon protease, which is not present in the BL21(DE3)pLysS strain), which are responsible for 90% of the overall protein degradation in the cytoplasm.[45] Several approaches to decrease protein degradation may be considered, such as the use of protease and heat shock deficient strains[46-48] or the addition of protease inhibitors to the culture medium.[49] In addition, "capping" with fusion partners on both termini ("dual fusion strategy") may protect against proteolytic cleavage and allow an easier purification of full-length product.[50]

3.3 Conclusions

The expression of artificial genes coding for poly-[(AG)₃EG] was successful via two approaches: they could be expressed either as a fusion protein with glutathione-S-transferase (GST) or as a polypeptide containing the T7 gene 10 leader sequence. Expression of GST-[(AG)₃EG]₂₀ resulted in a high yield of approximately 80 mg L⁻¹ of cell culture. Since the part of interest only makes up approximately 35% of the fusion protein, the final isolated yield for [(AG)₃EG]₂₀ after thrombin cleavage and reversed phase chromatography was 20 mg L⁻¹ of cell culture. The use of the bacteriophage T7 expression system was only successful when the T7 gene 10 sequence was included at the N terminus. In contrast to the GST-system, truncated polypeptide product was produced. After removal of this truncated product by reversed phase chromatography an isolated yield of 20 mg L⁻¹ of cell culture was obtained.

The expression of poly-[(AG)$_3$KG] was complicated by the sensitivity of this sequence to proteolytic degradation. Despite direct Ni-NTA purification a substantial amount of truncated product was formed. Using gel filtration chromatography most of it could be removed, but the final isolated yield was low (2 mg L^{-1} of cell culture). This product contained the complete repetitive part, but was still truncated at the C terminus. An efficient strategy for the purification of this polypeptide could be the use of two different affinity tags on both termini.

3.4 Experimental section

3.4.1 Materials

E. coli strain XL1-Blue used for cloning was obtained from Stratagene. The plasmids pBluescript® II SK(-) and pET-3b were from Stratagene. The plasmid pGEX-2T was obtained from Amersham Biosciences. Expression strains BL21, BL21(DE3) and BL21(DE3)pLysS were from Novagen. Synthetic oligonucleotides were purchased from Biolegio (Malden, the Netherlands). Restriction endonucleases (Gibco/New England Biolabs), T4 polynucleotide kinase (Gibco), T4 DNA ligase (Promega), calf intestine phosphatase (New England Biolabs) and Klenow fragment of DNA polymerase I (New England Biolabs) were used as recommended by the manufacturers. Plasmid midiprep kits and Qiaquick® gel extraction kits were from Qiagen. Lysozyme was purchased from Serva, RNAse A was from Sigma and DNAse I was from Boehringer Mannheim. Isopropyl-β-D-thiogalactopyranoside (IPTG) and phenylmethanesulphonyl fluoride (PMSF) were purchased from Sigma. Ni-NTA agarose beads were obtained from Qiagen. Glutathione sepharose 4B beads, reduced glutathione and thrombin (from bovine plasma, T3399, 95 units mg^{-1} protein) were obtained from Sigma. Dithiotreitol (DTT) was purchased from Sigma. The reversed phase column (RPC-15μm C2/C8 column HR16/10) and the gel filtration column (Superdex-75 Hi-Load™) were from Amersham Biosciences. Purifications were performed on a Biologic FPLC instrument from Biorad. Dialysis membranes were purchased from Spectra. Cyanogen bromide was purchased from Aldrich.

3.4.2 General methods

Cloning and expression. DNA manipulations, bacterial growth media and transformation conditions were carried out as described by Maniatis.[52]

NMR spectroscopy. ^{1}H-NMR spectra were recorded on a Varian Inova-400 instrument at 298 K.

MALDI-TOF mass spectrometry. Spectra were measured on a Bruker Biflex III spectrometer. Freeze-dried products were dissolved in 1:1 (v/v) water:acetonitrile with 0.1% trifluoroacetic acid to a final concentration of 10 mg mL^{-1} and mixed in a 1:1 ratio with a solution of 20 mg mL^{-1} of sinapinic acid (Sigma) in the same solvent and spotted on a MALDI-plate.

Amino acid analysis. 1 mg of polypeptide was placed in a hydrolysis tube and dissolved in 1 mL 5.7 N HCl. After purging with nitrogen for 10 min, the solution was frozen with dry ice/acetone and evacuated with high vacuum. The tube was closed by melting and heated for 48 hours at 110 °C. The solution was then evaporated and the residual solid was evaporated three times with water at 40 °C. The amino acids in the hydrolyzate were derivatized with 9-fluorenylmethylchloroformate, separated by high performance liquid chromatography (Varian HPLC 9000 series) on a reversed phase C18 silica column (TSK-RP18) and detected at a wavelength of 269 nm.[51]

SDS-polyacrylamide gel electrophoresis (SDS-PAGE). SDS-PAGE was performed using 15% resolving gels on a mini-Protean II slab cell apparatus (BioRad) according to standard procedures.[35]

Cyanogen bromide cleavage. Polypeptides were dissolved in 70% formic acid (2 mg mL^{-1}) and an equal volume of CNBr (100 mg mL^{-1}) in 70% formic acid was added, followed by incubation on a rotary arm for 2 days at room temperature in the dark. The samples were dried under vacuum in a centrifugal dryer at room temperature. The pellets were redissolved in demi-water and dialyzed for 2 days against demi-water using a dialysis membrane (3.5 kDa molecular weight cut-off).

3.4.3 Synthetic gene construction

Approach 1: Poly-[(AG)$_3$EG] with a minimal number of additional amino acids. The vector pSK-BspMI, which was used for creating DNA multimers encoding poly-[(AG)$_3$EG], was prepared with the use of the synthetic oligonucleotide 1 (depicted below).
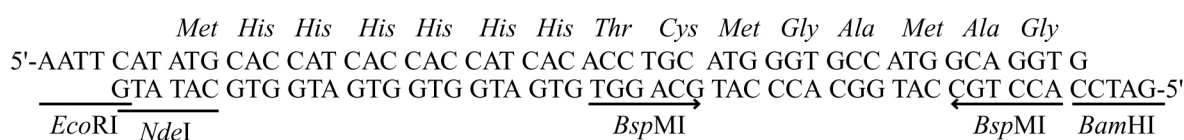
oligonucleotide 1

Both single-stranded oligos were dissolved in deionized water (80 ng μL^{-1}). 2.5 μL of each oligonucleotide solution was mixed with 2 μL 680 mM KCl and 1 μL water, and annealing was carried out by boiling for 4 min followed by cooling slowly to room temperature. 0.5 ng of annealed product was ligated with 25 ng *Eco*RI/*Bam*HI double digested pBluescriptII® SK(-) using 1 unit of T4 DNA ligase. The ligation was performed in 10 μL in a buffer composed of 50 mM Tris.HCl (pH 7.6), 10 mM MgCl$_2$, 10 mM dithiotreitol (DTT) and 1 mM ATP. After ligation for 3 hours at room temperature, 5 μL was transformed to 50 μL XL1-Blue cells using a heat shock procedure.[52] The transformed cells were plated on LB plates containing 100 μg mL^{-1} ampicillin and grown for 16 hours at 37 °C. Positive clones were identified by restriction analysis and the sequence was confirmed.

The resulting vector (pSK-BspMI) was successively digested with *Bsp*MI, dephosphorylated using calf intestine phosphatase (CIP) and extracted from agarose gel.

The two single-stranded oligonucleotides used for the construction of the gene coding for poly-[(AG)$_3$EG] (Figure 3.1a) were phosphorylated using 1 μg of oligonucleotide and 5 U of T4 polynucleotide kinase for 10 min at 37 °C. After heat inactivation at 65 °C for 10 min, the complementary oligonucleotides were annealed as described above. Multimerization was carried out overnight at room temperature in ligation buffer using 1U of T4 DNA ligase. Multimerization products were separated on a 2.5% TAE agarose gel and a 5-mer product was extracted from gel. The 5-mer was ligated with the *Bsp*MI-digested and dephosphorylated vector pSK-BspMI. Clones containing the multimeric insert were analyzed by restriction analysis and the sequence was verified. After amplification of the correct plasmid and *Bsp*MI digestion, the 240 bp 5-mer was extracted from gel. After multimerization of this fragment for 3 hours by T4 DNA ligase at room temperature, the *Bsp*MI-digested and dephosphorylated vector pSK-BspMI was added and ligation was continued overnight. Transformation of this ligation mixture resulted in clones containing multimers of up to 25 oligonucleotide repeats (designated pSK-BspMI-[(AG)$_3$EG]$_n$ (n = 10, 20, 30, 40 and 50)). The multimeric genes were transferred to the pET-3b expression vector using the *Nde*I and *Bam*HI restriction sites.

Approach 2: Poly-[(AG)$_3$EG] as GST-fusion protein. Expression vectors for the expression of GST-fused poly-[(AG)$_3$EG] were constructed using the pGEX-2T vector. This vector was linearized with *Sma*I resulting in blunt ends and subsequently dephosphorylated by calf intestine phosphatase. The multimeric genes were obtained from the vectors described in approach 1 by digestion with *Eco*RI and *Bam*HI. The 5' overhangs of the isolated multimeric genes were filled in using Klenow fragment of DNA polymerase I. The blunt-end multimeric genes were ligated overnight by T4 DNA ligase. Plasmids carrying the multimeric genes in the correct orientation were used for expression experiments.

Approach 3: Poly-[(AG)$_3$EG] with an N-terminal "T7 tag". For the expression of poly-[(AG)$_3$EG] with an N-terminal T7-tag, a new cloning vector, pSK-JS1, was constructed by ligation of the double stranded synthetic oligonucleotide 2 with *Eco*RI/*Bam*HI-digested plasmid pBluescriptII® SK(-).

```
              Asp  Pro  His  His  His  His  His  His  Thr  Cys   Met  Gly  Ala   Met  Ala  Gly
5'-AAT TCG GAT CCC CAC CAT CAC CAC CAT  CAC ACC TGC  ATG GGT GCC ATG GCA GGT G
     GC CTA GGG GTGGTA GTG GTG GTA GTG TGG ACG  TAC CCA CGG TAC CGT CCA CCTAG-5'
   EcoRI  BamHI                           BspMI              BspMI     BamHI
```

oligonucleotide 2
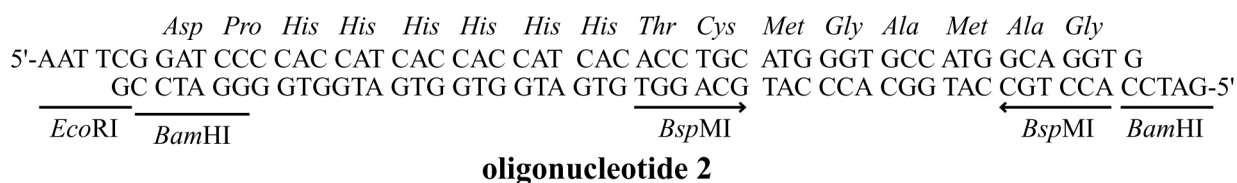
This vector was comparable to pSK-BspMI (Approach 1), but the *Nde*I site was replaced by a *Bam*HI site. The multimeric genes were obtained from approach 1 by digestion of pSK-BspMI-[(AG)$_3$EG]$_n$ (n = 10, 20 and 30) with *Bsp*MI, followed by gel extraction and ligation with

*Bsp*MI-digested and dephosphorylated pSK-JS1. The resulting vectors were digested with *Bam*HI and the multimeric genes were subsequently introduced into the *Bam*HI-digested and dephosphorylated expression vector pET-3b.

3.4.4 Expression of poly-[(AG)₃EG] using pET plasmids

Competent cells of *E. coli* strain BL21(DE3) or BL21(DE3)pLysS were transformed with 100 ng of the respective pET-3b plasmids and the transformation mixture was plated on LB agar plates with ampicillin (100 μg mL^{-1}) for BL21(DE3) and a mixture of ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) for the LysS strain. The plates were incubated overnight at 30 °C. A positive colony was selected and streaked out on a new LB-plate containing the appropriate antibiotic(s). A preculture of $1 \times$ LB medium with the appropriate antibiotic(s) and 1% glucose was inoculated with a single colony and grown overnight at 30 °C. The cells were harvested by centrifugation (5000 g, 5 min) and resuspended in fresh medium which was used to inoculate the main culture to an OD of 0.1. This culture was incubated at 30 °C or 37 °C. When an OD$_{600}$ of 0.6 was reached, IPTG (0.1 - 1 mM) was added to the culture to initiate protein expression. Cells were cultured for an additional 4 to 24 hours and then harvested by centrifugation (5000 g, 5 min). The supernatant was decanted, and the remaining cell pellet was stored at –70 °C.

3.4.5 Ni-NTA purification of poly-[(AG)₃EG]

Native conditions. 5 mL of lysis buffer (50 mM NaH$_2$PO$_4$ (pH 8.0); 300 mM NaCl; 1 mM PMSF; 1 mg mL^{-1} lysozyme) was added per gram (wet weight) of cell pellet. The pellet was resuspended and incubated for 30 min at 4 °C. The suspension was sonicated for 15 min on ice (Branson sonicator, 50% duty cycle, 5 units power). RNAse A (10 μg mL^{-1}) en DNAse I (5 μg mL^{-1}) were added, followed by incubation on ice for 15 min. The lysate was centrifuged for 30 min at 10000 g at 4 °C. Ni-NTA agarose was added to the lysate (1 mL per liter culture) and the mixture was incubated for 1 hour at 4 °C. A column was loaded, which was subsequently washed with 5 bed volumes of lysis buffer and wash buffer (with the same composition as the lysis buffer, with the addition of 20 mM imidazole). Elution was performed with 5 bed volumes of elution buffer (as lysis buffer with the addition of 250 mM imidazole).

Denaturing conditions. 5 mL of 6M guanidine hydrochloride (100 mM NaH$_2$PO$_4$, 10 mM Tris.HCl and 6 M guanidine.HCl, pH 8.0) was added per gram (wet weight) of cell pellet, followed by stirring for 1 hour at room temperature. After 30 min of centrifugation at 10000 g, Ni-NTA agarose was added to the supernatant (1 mL per liter culture) and the mixture incubated for 1 hour at room temperature. A column was loaded which was washed subsequently with 5 bed volumes of 6M guanidine.HCl with pH 8.0, 6.3 and 5.9 respectively. Elution was performed with 5 bed volumes of 6M guanidine.HCl (pH 4.0).

For the purification of $[(AG)_3EG]_{20}$ produced via approach 3 the elution fractions were subjected to reversed phase chromatography using a RPC-15μm C2/C8 column HR16/10. The composition of the eluents was as follows: Eluent A: water + 0.1% TFA; eluent B: 80% MeCN + 0.1% TFA. A 100 mL gradient was applied from 0 to 100% eluent B with a flow of 1 mL min^{-1}. Sample volumes of 5 mL were loaded. The samples were dialyzed against demi-water (3.5 kDa molecular weight cut-off) and freeze-dried. Isolated yield: 20 mg L^{-1} cell culture.

Characterization of $[(AG)_3EG]_{20}$ (obtained via approach 3): MS (MALDI-TOF): m/z = 16532 and 16748 (calcd. mass: 16459 Da). After cyanogen bromide cleavage: m/z = 11620 (calcd. mass for H-G-$[(AG)_3EG]_{20}$-A-homoserine-OH: 11659 Da), 13992 (incomplete cleavage), 5832 (+2) and 6966 Da (+2, incomplete cleavage).

3.4.6 Expression and purification of GST-fusion proteins

50 μL of competent *E. coli* BL21 cells were transformed with 100 ng pGEX-2T plasmids carrying the multimeric genes coding for poly-$[(AG)_3EG]$, and plated on LB agar plates with ampicillin (100 μg mL^{-1}). After growth overnight at 30 °C, single colonies were used to inoculate 100 mL LB medium, containing 100 μg mL^{-1} ampicillin. The cultures were incubated overnight at room temperature and placed at 30 °C in the morning. The culture was induced at an optical density (OD$_{600}$) of 0.6 by adding IPTG to a final concentration of 0.4 mM, followed by incubation for 3 hours. The cells were sedimented by centrifugation at 6000 g for 10 min at 4 °C and the cell pellet was stored at -70 °C. The cells were resuspended in 5 mL ice-cold 1 × PBS (phosphate buffered saline) and disrupted by sonication on ice (250 W Branson sonicator, four 45-seconds bursts with a 45 second cooling period between each burst, 50% duty cycle, 5 units power). The lysate was centrifuged at 12000 g for 10 min at 4 °C and the supernatant was used for purification. Purification was performed with 0.1 g glutathione sepharose 4B beads which were equilibrated by swelling in 10 mL 1 × PBS overnight at 4 °C, resulting in a 1 mL bed volume. The cleared lysate was added to the beads, the suspension was incubated for 2 hours at 4 °C on a rotating wheel and subsequently transferred to a disposable column. The flow-through was collected and the column was washed 3 times with 10 bed volumes of 1 × PBS. The fusion protein was eluted with glutathione elution buffer (10 mM reduced glutathione in 50 mM Tris.HCl, pH 8.0).

Large-scale expression and purification of GST-$[(AG)_3EG]_{20}$ was carried out analogous as described above. To 125 mL lysate obtained from 2.5 L culture, 20% Triton X-100 was added to a final concentration of 1%, followed by incubation for 30 min at 4 °C. Purification was performed using a 20 mL glutathione sepharose 4B column. The eluted polypeptide was dialyzed against demi-water for 2 days (3.5 kDa molecular weight cut-off) and freeze-dried. For thrombin cleavage, the freeze-dried GST-fusion proteins were redissolved in 1 × PBS at a concentration of 5 mg mL^{-1}. The thrombin protease used was dissolved in 1 × PBS to a

concentration of 1 unit per μL. Per milligram of fusion protein 5 μL of thrombin solution was added and the solution was incubated for 16 hours at room temperature.

Purification of [(AG)$_3$EG]$_{20}$ from GST and GST-fusion protein was performed by reversed phase chromatography (RPC-15μm C2/C8 column HR16/10) at room temperature. The composition of the eluents was as follows: Eluent A: water + 0.1% TFA; eluent B: 80% MeCN + 0.1% TFA. 5 mL samples were loaded for each run and after washing with 20 mL eluent A, a 100 mL gradient was applied from 0 to 100% eluent B.

^{1}H-NMR (400 MHz, TFA-d): δ = 1.53 (d, 186H, Ala-Hβ), 2.17 and 2.53 (s, 40H, Glu-Hβ), 2.66 (s, 40H, Glu-Hγ), 4.26 (s, 170H, Gly-Hα), 4.72 (s, 62H, Ala-Hα), 4.86 (s, 20H, Glu-Hα). MS (MALDI-TOF): m/z = 14631 Da (calcd. mass: 14320 Da), m/z = 7288 (+2) and 28714 Da (dimer). Amino acid analysis (mol percent): Ala, 37.3; Gly, 49.4; Glu, 13.3. Calculated amino acid percentages for [(AG)$_3$EG]$_{20}$ (mol percent): Ala, 37.1; Gly, 50.9; Glu, 12.0. After cyanogen bromide cleavage: MS (MALDI-TOF): m/z = 11706 Da (calcd. mass for H-G-[(AG)$_3$EG]$_{20}$-A-homoserine-OH: 11659 Da), m/z = 5870 (+2) and 23402 Da (dimer).

3.4.7 Expression and purification of poly-[(AG)$_3$KG]

E. coli BL21(DE3)pLysS strain was transformed with pET-624 (coding for 24 repeats of (AG)$_3$KG), plated on LB agar plates supplemented with ampicillin (100 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) and grown overnight at 30 °C. One colony was used to inoculate a 200 mL LB medium containing the appropriate antibiotics, followed by growing at 30 °C overnight. This preculture was used to inoculate 4 L of main culture, which was grown at 37 °C and induced with 0.4 mM IPTG at an OD of 0.6. After 3 hours of protein production the cells were centrifuged for 15 min at 6000 g at 4 °C. The pellet was immediately resuspended in 100 mL lysisbuffer (50 mM NaH$_2$PO$_4$ (pH 8.0), 300 mM NaCl and 1 mM PMSF) and sonicated for 10 min on ice (250 W Branson sonicator, 50% duty cycle, 5 units power). The lysate was clarified by centrifugation for 15 min at 4000 g at 4 °C. The lysate was incubated with 5 mL Ni-NTA beads for 1 hour at 4 °C, and the beads were subsequently poured into a column. The column was washed with 100 mL lysis buffer, followed by a 100 mL gradient from 0 to 200 mM imidazole (50 mM NaH$_2$PO$_4$ (pH 8.0), 300 mM NaCl). The elution fractions were stored at -20 °C. Preparative gel filtration chromatography was performed using a Superdex-75 Hi-LoadTM 26/60 column at 4 °C. The eluent composition was as follows: 50 mM NaH$_2$PO$_4$ (pH 8.0), 150 mM NaCl, 1 mM PMSF. A flow of 2 mL min^{-1} was used. Sample volumes up to 10 mL were used per run.

Characterization of [(AG)$_3$KG]$_{24}$: MS (MALDI-TOF): m/z = 17932, 17767 (calcd. mass of full-length polypeptide: 18897 Da) and 8890 Da(+2). ^{1}H-NMR (400 MHz, D$_2$O): δ = 1.35 (d, 231H, Ala-Hβ), 1.65 (m, 52H, Lys-Hγ), 1.85 (m, 104H, Lys-Hβ/γ/δ), 2.85 (t, 52, Lys-Hε), 3.92 (m, 202H, Gly-Hα), 4.30 (s, 103H, Ala/Lys-Hα). After CNBr cleavage: MS (MALDI-TOF):

m/z = 15427 (calcd. mass for H-G-[(AG)$_3$KG]$_{24}$-A-homoserine-OH: 13918 Da) and 7726 Da (+2).

3.5 References

[1] J. T. Prince, K. P. McGrath, C. M. Digirolamo, D. L. Kaplan, *Biochemistry US* **1995**, *34*, 10879.

[2] R. V. Lewis, M. Hinman, S. Kothakota, M. J. Fournier, *Protein Expr. Purif.* **1996**, *7*, 400.

[3] S. R. Fahnestock, Z. Yao, L. A. Bedzyk, *Rev. Mol. Biotechnol.* **2000**, *74*, 105.

[4] D. T. McPherson, J. Xu, D. W. Urry, *Protein Expr. Purif.* **1996**, *7*, 51.

[5] H. Daniell, C. Guda, D. T. McPherson, X. Zhang, J. Xu, D. W. Urry, *Methods Mol. Biol.* **1997**, *63*, 359.

[6] I. Goldberg, A. J. Salerno, T. Patterson, J. I. Williams, *Gene* **1989**, *80*, 305.

[7] A. J. Salerno, I. Goldberg, *Appl. Microbiol. Biotechnol.* **1993**, *39*, 221.

[8] K. P. McGrath, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Am. Chem. Soc.* **1992**, *114*, 727.

[9] A. Panitch, K. Matsuki, E. J. Cantor, S. J. Cooper, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1997**, *30*, 42.

[10] M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. J. Mason, D. A. Tirrell, *Science* **1994**, *265*, 1427.

[11] E. J. Cantor, E. D. T. Atkins, S. J. Cooper, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Biochem.* **1997**, *122*, 217.

[12] M. T. Krejchi, S. J. Cooper, Y. Deguchi, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1997**, *30*, 5012.

[13] N. L. Goeden-Wood, V. P. Conticello, S. J. Muller, J. D. Keasling, *Biomacromolecules* **2002**, *3*, 874.

[14] N. L. Goeden-Wood, J. D. Keasling, S. J. Muller, *Macromolecules* **2003**, *36*, 2932.

[15] R. S. Farmer, K. L. Kiick, *Biomacromolecules* **2005**, *6*, 1531.

[16] Y. Wang, K. L. Kiick, *J. Am. Chem. Soc.* **2005**, *127*, 16392.

[17] S. M. Yu, C. H. Soto, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 6552.

[18] S. J. M. Yu, V. P. Conticello, G. H. Zhang, C. Kayser, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Nature* **1997**, *389*, 167.

[19] M. T. Doel, M. Eaton, E. A. Cook, H. Lewis, T. Patel, N. H. Carey, *Nucleic Acids Res.* **1980**, *8*, 4575.

[20] S. H. Shen, *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 4627.

[21] M. Lennick, J. R. Haynes, *Gene* **1987**, *61*, 103.

[22] A. Kuliopulos, C. T. Walsh, *J. Am. Chem. Soc.* **1994**, *116*, 4599.

[23] T. Kempe, S. B. H. Kent, F. Chow, S. M. Peterson, W. I. Sundquist, J. J. Litalien, D. Harbrecht, D. Plunkett, W. J. Delorbe, *Gene* **1985**, *39*, 239.

[24] S. R. Fahnestock, S. L. Irwin, *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23.

[25] M. B. Hinman, Z. Dong, R. V. Lewis, *Mol. Biol. Cell* **1992**, *3*, A113.

[26] S. Arcidiacono, C. Mello, D. Kaplan, S. Cheley, H. Bayley, *Appl. Microbiol. Biotechnol.* **1998**, *49*, 31.

[27] J. H. Lee, P. M. Skowron, S. M. Rutkowska, S. S. Hong, S. C. Kim, *Genetic Anal. Biomol. Engin.* **1996**, *13*, 139.

[28] R. A. McMillan, T. A. T. Lee, V. P. Conticello, *Macromolecules* **1999**, *32*, 3643.

[29] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2002**, *3*, 357.

[30] E. R. Wright, V. P. Conticello, *Adv. Drug Deliv. Rev.* **2002**, *54*, 1057.

[31] F. Ferrari, J. Cappello, in *Protein-based materials* (Eds.: K. P. McGrath, D. L. Kaplan), Birkhauser, Boston, MA, **1997**, pp. 37.

[32] F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, *Methods in Enzymology* **1989**, *185*, 60.

[33]   R. C. Beavis, B. T. Chait, H. S. Creel, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Am. Chem. Soc.* **1992**, *114*, 7584.
[34]   M.-V. Mawn, M.-J. Fournier, D.-A. Tirrell, T.-L. Mason, *J. Bacteriol.* **2002**, *184*, 494.
[35]   F. Ausubel, R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, K. Struhl, *Current protocols in molecular biology, Vol. 3*, John Wiley & sons, New York, **1999**.
[36]   D. B. Smith, K. S. Johnson, *Gene* **1988**, *67*, 31.
[37]   M. L. Sprengart, H. P. Fatscher, E. Fuchs, *Nucleic Acids Res.* **1990**, *18*, 1719.
[38]   M. L. Sprengart, E. Fuchs, A. G. Porter, *EMBO Journal* **1996**, *15*, 665.
[39]   M. Faxén, J. Plumbridge, L. A. Isaksson, *Nucleic Acids Res.* **1991**, *19*, 5247.
[40]   J. P. Etchegaray, M. Inoue, *J. Biol. Chem.* **1999**, *274*, 10079.
[41]   C.-H. Luan, et al. (2004), *Genome Res.* **2004**, *14*, 2102.
[42]   M. Landon, *Methods in Enzymology* **1977**, *47*, 145.
[43]   W. A. Schroeder, J. B. Shelton, J. R. Shelton, *Arch. Biochem. Biophys.* **1969**, *130*, 551.
[44]   M. Pacaud, C. Richaud, *J. Biol. Chem.* **1975**, *250*, 7771.
[45]   M. R. Maurizi, *Experientia* **1992**, *48*, 178.
[46]   S. O. Enfors, *Trends Biotechnol.* **1992**, *10*, 310.
[47]   S. Gottesman, *Methods in Enzymology* **1990**, *185*, 119.
[48]   M. Murby, M. Uhlen, S. Ståhl, *Protein Expr. Purif.* **1996**, *7*, 129.
[49]   W. F. Prouty, A. L. Goldberg, *J. Biol. Chem.* **1972**, *247*, 3341.
[50]   M. Murby, L. Cedergren, J. Nilsson, P. A. Nygren, B. Hammarberg, B. Nilsson, S. O. Enfors, M. Uhlen, *Biotechnol. Appl. Biochem.* **1991**, *14*, 336.
[51]   R. Cunico, A. G. Mayer, C. T. Wehr, T. L. Sheenan, *Biochromatography* **1986**, *1*.
[52]   J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*, 2 ed., Cold Spring Laboratory Press, Cold Spring, **1989**.

Chapter 4

Preparation of hybrid block copolymers comprising β-sheet polypeptides

Abstract

The synthesis of silk-like triblock copolymers consisting of a central β-sheet polypeptide block and poly(ethylene glycol) (PEG) end blocks is described. Polypeptides consisting of 10 and 20 repeats of the β-sheet forming sequence -(AG)$_3$EG- (A = alanine, G = glycine and E = glutamic acid), flanked by cysteine residues, were expressed in E. coli and purified by nickel-nitrilotriacetate (Ni-NTA) chromatography. Further purification by gel filtration or reversed phase chromatography was necessary for both polypeptides to remove truncated product, which was formed during expression. The isolated yield of [(AG)$_3$EG]$_{10}$ produced in shake flask cultures was approximately 6 mg per liter. The polypeptide [(AG)$_3$EG]$_{20}$ was produced in high cell density culture, resulting in isolated yields of approximately 10 mg per liter. Unfortunately, high cell density fermentation was unsuccessful for the production of larger amounts of polypeptide, since the amount of produced polypeptide per cell decreased dramatically in comparison with shake flask expressions.

The thiol functionality of the cysteine residues was subsequently used for the conjugation of maleimide-functionalized PEGs. The polypeptide [(AG)$_3$EG]$_{10}$ was conjugated with PEGs with molecular weights of 750, 2000 and 5000 g mol^{-1}. The polypeptide [(AG)$_3$EG]$_{20}$ was conjugated only with PEG-750. The excess PEG which was used to obtain complete functionalization, could be efficiently removed by Ni-NTA chromatography. Cyanogen bromide cleavage was then used to remove the non-repetitive N- and C-terminal amino acids. MALDI-TOF mass spectrometry indicated that the main products that were formed, were polypeptides functionalized with two PEG chains.

In addition, the synthesis of triblock copolymers in which the central polypeptide block was flanked by poly(methyl methacrylate) (pMMA) end blocks was performed. PMMA was prepared by atom transfer radical polymerization and subsequently functionalized using ε-maleimidocaproic acid. Although SDS-PAGE analysis indicated successful coupling, the chemical characterization of this conjugate proved to be troublesome.

4.1 Introduction

Synthetic polymer-protein conjugates are of considerable interest because of their application potential in a variety of fields. The usefulness of these biological-synthetic hybrids was first established in the area of biomedicine in the 1970s.[1-3] The conjugation of poly(ethylene glycol) (PEG) to peptide- or protein-based drugs was found to have beneficial effects on the therapeutic properties of these drugs, in terms of reduced immunogenicity, higher stability and increased plasma half-life. Several protein-PEG conjugates have been commercialized.[4] More recently, the use of synthetic polymer-protein conjugates for bioanalytical application, bioseparations or for modulation of recognition properties has been investigated.[5, 6] For these applications thermosensitive polymers (e.g. poly(N-isopropylacrylamide) are used, whose reversible, temperature-induced precipitation behaviour and solubilization is an interesting tool for enzyme recovery and immuno-assays.

Relatively new is the use of hybrid block copolymers for the build-up of supramolecular architectures. The folding and recognition properties of peptides may allow the construction of more complex, hierarchically organized materials with properties unequalled by traditional synthetic polymers. Most research in this area has focused on the combination of small peptide motifs (β-strands and coiled-coils) and PEG and the majority of these conjugates were prepared by solid-phase synthesis. One of the first examples was reported by Meredith et al., who investigated the aggregation properties of a diblock copolymer of the central domain of β-amyloid peptide (A$_\beta$(10–35) and PEG.[7] Fibril formation by the conjugate was, unlike for the native peptide, completely reversible and the solubility of the formed fibrils was improved. It was suggested this could facilitate the characterization of the different stages of fibril formation. The self-assembly properties of di- and triblock copolymer based on PEG and amphiphilic β-strand sequences have been reported more recently.[8] These hybrids formed lamellar structures consisting of alternating PEG layers and peptide domains in an antiparallel β-sheet conformation in the solid-state.

An interesting example in which the peptide segment was used to contribute to the mechanical properties of the final material is the work of Sogah and coworkers.[9, 10] They prepared multiblock copolymers of oligopeptide β-sheet silk domains (poly(alanine) or GAGA sequence) and PEG via polycondensation. Another example worth mentioning is the work by Nolte and coworkers., who studied the self-assembly properties of a synthetic polymer-protein conjugate composed of a single polystyrene chain (degree of polymerization of approximately 40) and lipase B from *Candida antarctica*.[11] This hybrid block copolymer can be considered as a giant amphiphile, consisting of a polar enzyme head group and an apolar polymer tail. Indeed, this conjugate formed well-defined micrometer long fibers, consisting of micellar rods and seemed to behave similar to its low molecular weight counterparts.

In this chapter, the preparation of triblock copolymers, consisting of a β-sheet forming central poly-[(AG)$_3$EG] sequence and flanking synthetic polymer blocks, is described. The

rationale behind the attachment of synthetic polymer blocks at the N and C termini was to restrict macroscopic crystallization and to allow translation of the β-sheet design characteristics of width, height and surface functionality into the self-assembled structures. With the aim to investigate the effect of the nature of the synthetic polymer part on the assembly behaviour of the β-sheet polypeptides, both hydrophilic PEG and the more hydrophobic poly(methyl methacrylate) were conjugated to the β-sheet polypeptides. The strategy for the preparation of these hybrid block copolymers is depicted in Figure 4.1.
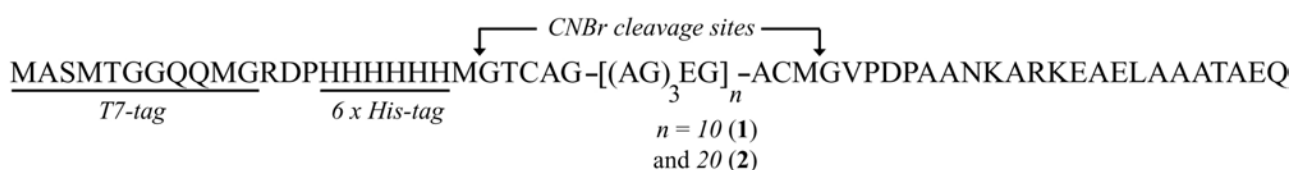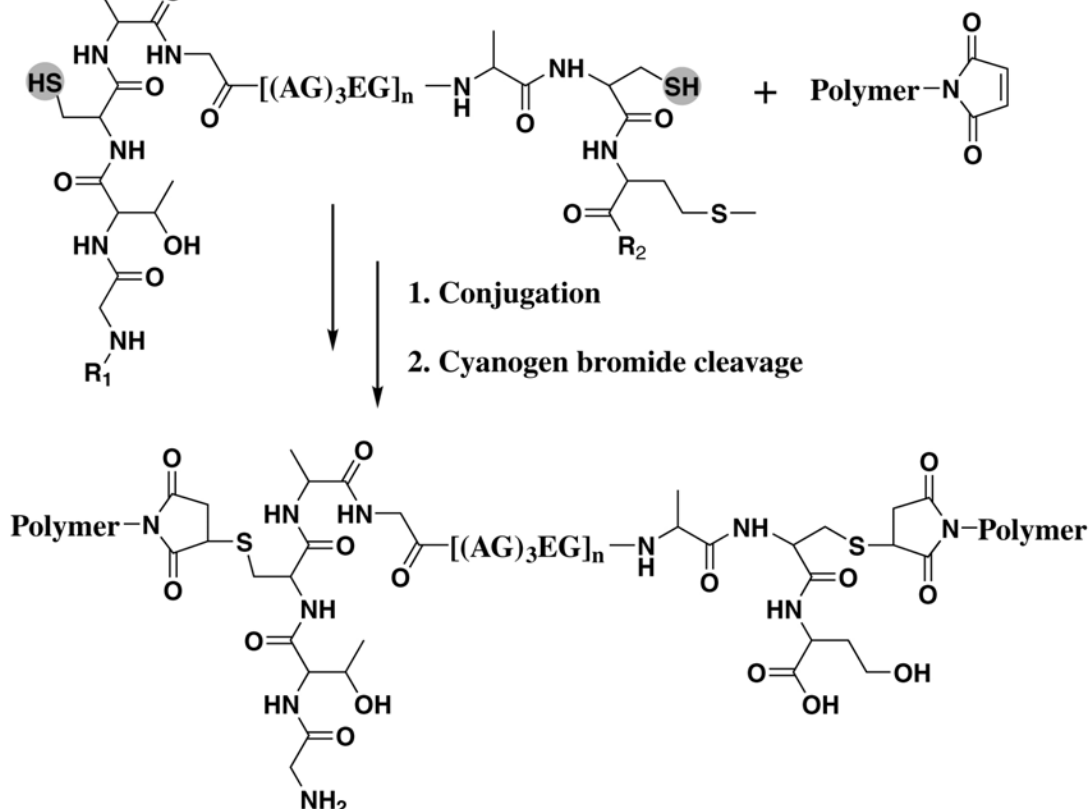
Figure 4.1 (a) *Amino acid sequence of cysteine flanked poly-[(AG)₃EG]. Cyanogen bromide cleavage allows removal of the residual amino acids introduced in the cloning process.* *(b)* *Conjugation of maleimide functionalized polymers to poly-[(AG)₃EG] using selective reaction of maleimide functionality and thiol group of the cysteine residues.*

For conjugation of synthetic polymers, the poly-[(AG)₃EG] sequence is outfitted with N- and C-terminal cysteine residues. The thiol functionality of the cysteine residues can subsequently be used for selective alkylation with maleimide functionalized polymers. The 6 × histidine-tag can

be used for purification of the conjugates. The use of methionine residues allows removal of the residual amino acids by cyanogen bromide cleavage. This chapter describes subsequently the expression and purification of these polypeptides, the synthesis of the maleimide functionalized polymers and their subsequent conjugation and characterization.

4.2 Results and discussion

4.2.1 Expression and purification of cysteine-flanked poly-[(AG)₃EG]

For the preparation of cysteine-flanked poly-[(AG)$_3$EG] the multimeric genes encoding 10 and 20 repeats of the -(AG)$_3$EG- sequence (described in Chapter 3), were transferred to a new plasmid containing the genetic information for the T7-tag, the His-tag and the N- and C-terminal cysteines. The expression of cysteine-flanked [(AG)$_3$EG]$_{10}$ (**1**) in *E. coli* BL21(DE3) cells was followed by immunoblot analysis using T7 tag mouse monoclonal antibodies (Figure 4.2a), since expression was not or barely visible by Coomassie staining. Whereas already before induction a band at the 28 kDa marker was present, the intensity of this band increased after induction with IPTG. In addition, a second band appeared at the height of the 21 kDa marker. Both products were present in the soluble fraction after cell disruption and were purified by native Ni-NTA chromatography (Figure 4.2b). Gel filtration chromatography with a Superdex-75 column was performed under reducing conditions (10 mM β-mercaptoethanol) and resulted in separation of both products. The band located at the 28 kDa marker was expected to correspond to full-length product based on the expression results for the poly-[(AG)$_3$EG] variants without cysteine residues (Chapter 3, Paragraph 2.2). This was confirmed by MALDI-TOF mass spectrometry (Figure 4.2c). The observed average mass of 11134 Da is higher than the calculated average mass for the reduced polypeptide (10980 Da). However, it corresponds well with the calculated average mass for an adduct of the polypeptide with β-mercaptoethanol (11132 Da). The band located at the height of the 21 kDa marker was shown to consist of a mixture of truncated products from 5.9 kDa to 6.9 kDa with the dominant peak at 6.2 kDa. This peak corresponds to a protein product from N-terminus to alanine-75, which is the first alanine after six octapeptide repeats. This result was confirmed by amino acid analysis by comparison of amino acid content for amino acids located at the N-terminus and the three amino acids located in the repeats. The final yield of full-length polypeptide after dialysis and freeze-drying was approximately 6 mg per liter of culture.

    The polypeptide [(AG)$_3$EG]$_{20}$ (**2**) was produced by high cell density fermentation. The fermentation was carried out according to a protocol described by Panitch and coworkers[12], who used it for the production of poly(L-alanylglycine). Medium with a chemically defined nutrient composition was used to obtain a high cell density.[13] Batch growth of *E. coli* BL21(DE3)pLysS cells was carried out using glucose as carbon source. After depletion of glucose (occurring at an OD$_{600}$ of approximately 25), determined by a rise in pH and a reduced oxygen consumption, fed-batch growth was established using a pH-stat feeding strategy.[14]
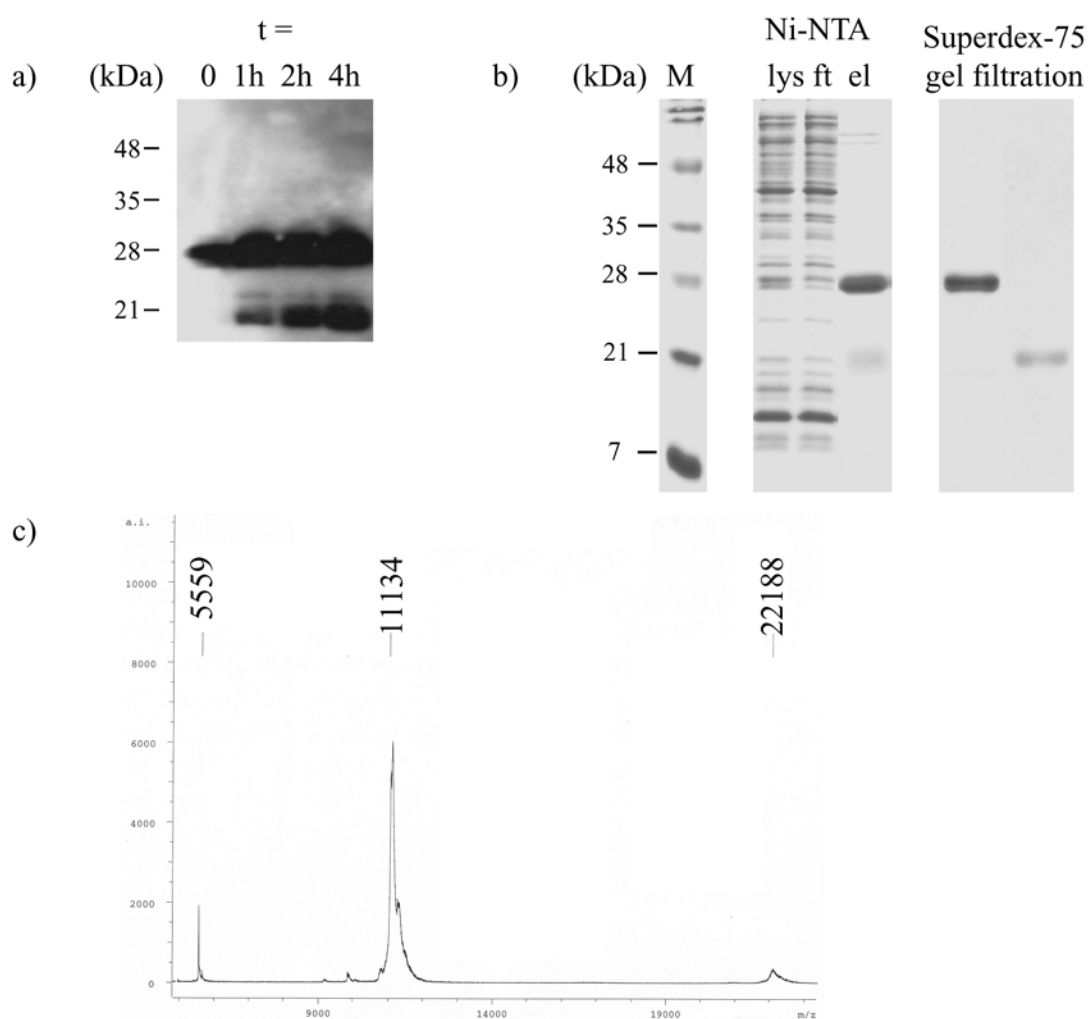
Figure 4.2 (a) *Expression of cysteine-flanked [(AG)₃EG]₁₀ (**1**) in E. coli BL21(DE3) cells analyzed by immunoblotting using T7-tag antibodies. Expression was induced by addition of IPTG to a final concentration of 1 mM at t = 0.* ***(b)*** *Purification of cysteine-flanked [(AG)₃EG]₁₀ analyzed by SDS-PAGE. The polypeptide was purified by native Ni-NTA chromatography (lys = soluble lysate, ft = column flow-through; el = elution). Gel filtration chromatography (Superdex-75) was used to separate full-length product (at the height of the 28 kDa marker) from truncated product. Proteins were visualized by Coomassie staining.* ***(c)*** *MALDI-TOF spectrum of cysteine-flanked [(AG)₃EG]₁₀.*

During this stage glucose was fed to the reactor through the acid pump in response to increased pH. The expression was induced at the end of the batch phase by the addition of IPTG to a final concentration of 1 mM. After 4 hours of expression an optical density of 25 – 30 was reached. The expression was analyzed by immunoblot analysis (Figure 4.3a). Two main protein products were formed. The lower band was located at the same height as the byproduct observed during [(AG)₃EG]₁₀ expression and corresponded to the same truncated product. The full-length polypeptide had an apparent molecular weight of 32 kDa and was located in the soluble fraction of the lysate. Since the column size used for native Ni-NTA purification was not adequately adapted to the amount of expressed polypeptide, still a large amount of contaminating *E. coli*

proteins were present in the elution fraction (Figure 4.3b). However, by a simple heating step (10 min at 70 °C), most *E. coli* proteins precipitated and could be removed by centrifugation. Finally, reversed phase chromatography was used to isolate full-length product, since the truncated product eluted at a lower acetonitrile concentration (15%) than the full length product (20% acetonitrile).
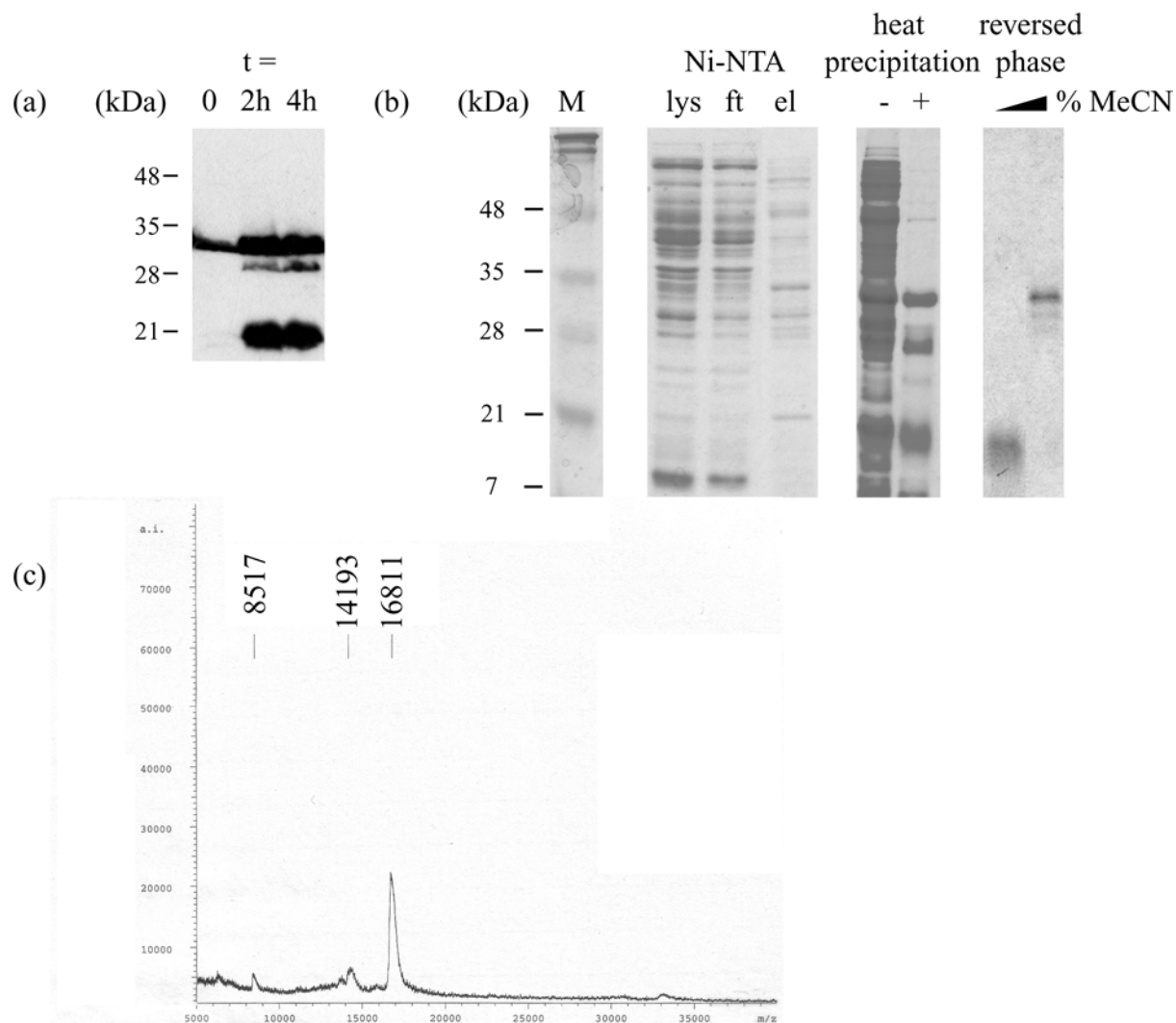
Figure 4.3 (a) *Expression of cysteine-flanked [(AG)₃EG]₂₀ (2) during high cell density fermentation of E. coli BL21(DE3)pLysS cells analyzed by immunoblotting using T7-tag antibodies. Expression was induced by addition of IPTG to a final concentration of 1 mM at t = 0. **(b)** Purification of cysteine-flanked [(AG)₃EG]₂₀ analyzed by SDS-PAGE. The polypeptide was purified by native Ni-NTA chromatography (lys = soluble lysate, ft = column flow-through; el = elution). Heat treatment (10 min at 70 °C) was used to precipitate E. coli proteins. Results before (-) and after (+) heat precipitation are depicted. Reversed phase chromatography (RPC-15µm C2/C8) was used to separate full-length product (between the 28 and the 35 kDa marker) from truncated polypeptides. **(c)** MALDI-TOF spectrum of cysteine-flanked [(AG)₃EG]₂₀.*

MALDI-TOF mass spectrometry gave an average molecular mass of 16811 Da, which is higher than the calculated average mass for the reduced polypeptide (16686 Da), and could correspond to an adduct of the polypeptide with β-mercaptoethanol (calculated average mass is 16836 Da) (Figure 4.3c). The isolated yield after freeze-drying was approximately 10 mg per liter of culture. Alternatively, gel filtration chromatography (Superdex-75 column) could be used for the separation of both polypeptides (not shown).

The expression level of $[(AG)_3EG]_{20}$, in terms of produced protein per cell, was assessed to be approximately 5 times lower in high cell density fermentation in comparison to shake flask experiments (based on immunoblot analysis of whole cell lysates). However, since approximately 20-fold higher cell densities were reached in comparison to shake flask experiments, the amount of polypeptide produced per liter of culture was higher. However, this high cell density fermentation could not be called satisfactory, since more effort was needed in the purification procedure. In later high cell density fermentations, carried out for the production of $[(AG)_3EG]_{10}$, the amount of produced polypeptide per cell was more than 50-fold lower than in shake flask expressions, making this method with its current protocol unsuitable for the production of this polypeptide.

Although high cell density fermentation of recombinant *E. coli* is a well-established technique for the large-scale production of proteins[15, 16], and also repetitive proteins such as poly(L-alanylglycine) (240 repeats)[12] or the sea mussel adhesive protein (75 repeats of a decapeptide sequence)[17] have been produced in inclusion bodies with yields of 1 to 5 gram per liter of culture, a successful large-scale production of poly-$[(AG)_3EG]$ has not been established, unfortunately.

One of the possible causes for low expression in high cell density culture is plasmid instability. Because of the higher number of generations in comparison to shake flask experiments, the plasmid might have been lost from a large fraction of the *E. coli* cells. Incompletely repressed expression and loss of selective pressure as a consequence of ampicillin degradation by secreted β-lactamase or acidic conditions can cause overgrowth of the culture with plasmid-free cells. The latter effect may be alleviated by using the ampicillin analog carbenicillin, which is less susceptible for degradation.[18] Furthermore, the production of acetate in the presence of excess glucose has been shown to reduce recombinant protein synthesis and growth rate.[19-23] Overfeeding of glucose should not have taken place with a pH-stat feeding strategy, but actual measurement of glucose and acetate concentrations in the fermentor during the fed-batch phase was not carried out. Feeding strategies different from pH-stat, such as DO-stat (constant level of dissolved oxygen)[24], constant[25] or exponential feeding[26] were not performed, but could also have an effect on protein production.[17]

Furthermore, the lack of certain metabolites might hinder higher levels of expression. Since the expressed polypeptide deviates considerably from the average *E. coli* protein, amino acid starvation might occur. The response includes the down regulation of genes involved in transcription, translation and amino acid biosynthesis.[27] Furthermore, enhanced proteolysis

might occur to release amino acids from nonessential proteins for incorporation into proteins essential for cell survival.[28, 29] The addition of the appropriate amino acids might alleviate this stress response and can result in higher levels of protein expression.[30, 31]

Besides the difficulties with high cell density fermentation for large-scale polypeptide production, another complicating factor was the presence of trunctated product. Remarkably, the truncated products for $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ were the same, corresponding to a mixture of polypeptides from N-terminus to around 6 -$(AG)_3EG$- repeats. Besides the negative effect on yield of full-length product (up to 50 weight percent of truncated polypeptide was formed for $[(AG)_3EG]_{20}$), it also was a complicating factor for purification. For $[(AG)_3EG]_{36}$ reported by Tirrell and coworkers, also truncated product seemed to have formed, but no further comment on the nature of this product was given.[32] However, for the similar repetitive sequence $[(AG)_4PEG]_{14}$, the appearance of a mixture of truncated product was reported.[33] The fragments consisted of the complete N-terminus followed by four to six copies of the repeating undecapeptide sequence. It was explained by the action of exo- and endopeptidases. The formation of truncated products was also reported for the expression of silk fibroins in *E. coli* and was determined to be caused by premature translation termination.[34] It is not clear why during the expression of poly-$[(AG)_3EG]$ a relatively discrete truncated product was formed. This would not be expected in the case of proteolysis or premature termination of translation.

The purified full-length polypeptides produced were used for the attachment of the synthetic polymer poly(ethylene glycol).

4.2.2 Preparation of maleimide functionalized poly(ethylene glycols)

For the preparation of poly(ethylene glycol) (PEG) conjugates of poly-$[(AG)_3EG]$ β-sheets, maleimide functionalized PEGs were prepared, which are capable of reacting with the thiol groups of the cysteine residues located on both sides of the poly-$[(AG)_3EG]$ segment. These were prepared by reaction of monoamine-functionalized PEGs with ε-maleimidocaproic acid using BOP (benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate) as a coupling agent (Figure 4.4). This reaction was performed for PEGs with molecular weights of 750, 2000 and 5000 g mol^{-1}. The reaction was carried out for 24 hours at room temperature using an equimolar amount of ε-maleimidocaproic acid. Gel filtration chromatography with a Sephadex LH-20 column was performed to remove the formed hexamethylphosphoramide (HMPA) and hydroxybenzotriazole (HOBt) and the isolated yield was 68 to 89%. MALDI-TOF analysis showed that the PEGs were completely functionalized with the maleimide group.
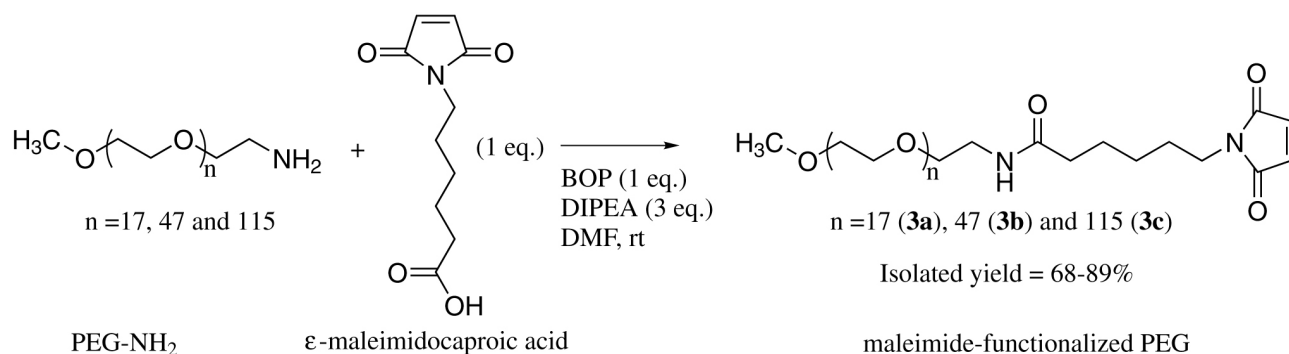
Figure 4.4 *Coupling reaction of monoamine-functionalized poly(ethylene glycols) with ε-maleimidocaproic acid using BOP (benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate) as a coupling agent.*

4.2.3 Conjugation of maleimide functionalized poly(ethylene glycols) and poly-[(AG)₃EG]

Initial conjugation experiments were carried out with $[(AG)_3EG]_{20}$ and maleimide functionalized poly(ethylene glycol) with a number average molecular weight of 750 g mol^{-1} (PEG-750). The free thiol content after dissolution of the freeze-dried polypeptide in sodium phosphate buffer was determined with Ellman's essay[35], which indicated that only 5% of the cysteines was in the reduced form. Therefore the oxidized cysteines were reduced with an excess of dithiotreitol (DTT). Initially dialysis was used for removal of DTT, but turned out to be impractical since only 30% of the cysteines were in the reduced form after dialysis. Non-reducing SDS-PAGE analysis showed bands at higher molecular weights indicating intermolecular disulfide bridge formation (data not shown).

A more practical alternative was to precipitate the polypeptide with the use of trichloroacetic acid (TCA), which is a commonly used protein precipitation method[36], and turned out to be suitable for the precipitation of this acidic polypeptide. Subsequent washing with 20% TCA and demi-water was carried out to remove DTT. The resulting pellet could be redissolved in sodium phosphate buffer (pH 6.8) for coupling with maleimide functionalized PEG-750. Determination of free thiol content showed that the number of free thiols typically was 1.5- to 2-fold higher than the theoretical number of cysteines, indicating the presence of some residual DTT. However, this procedure was chosen to avoid re-oxidation of cysteines and the presence of DTT could be compensated by the addition of excess of PEG-750.

Coupling experiments were performed with increasing equivalents (0.4 - 40) of PEG relative to cysteine residues to determine the amount needed for complete conversion, i.e. bifunctionalization of $[(AG)_3EG]_{20}$ with PEG-750. SDS-PAGE showed that only by using 20 equivalents of PEG-750 a complete shift of the polypeptide band to a higher molecular weight was realized (Figure 4.5a).
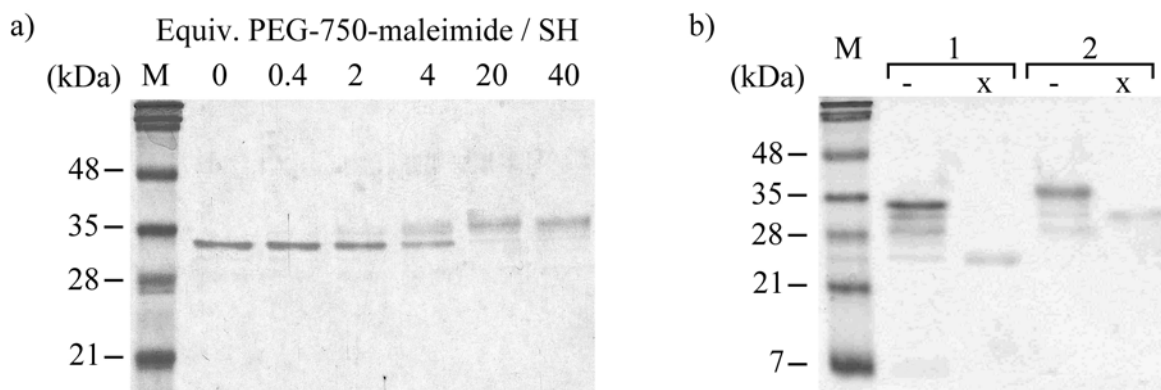
Figure 4.5 (a) *SDS-PAGE analysis of conjugation reaction of [(AG)₃EG]₂₀ and maleimide functionalized PEG-750. Reactions were carried out in 100 mM NaH₂PO₄ buffer (pH 6.8) containing 150 mM NaCl at room temperature for 1 day.* ***(b)*** *Cyanogen bromide cleavage of (1) [(AG)₃EG]₂₀ and (2) [(AG)₃EG]₂₀-PEG-750 conjugate. (-) sample incubated in 70% formic acid without CNBr and (×) sample incubated in 70% formic acid containing CNBr. Visualization of bands for CNBr cleaved samples required contrast enhancement.*

After conjugation, the 6 × histidine tag was used to remove the excess of PEG by Ni-NTA chromatography. After loading the conjugation mixture, the column was washed until the absorbance at 214 nm no longer decreased, indicating complete removal of PEG. Subsequently the conjugate was eluted from the column with the use of imidazole. Then, the fractions were dialyzed and freeze-dried. In Figure 4.5b the CNBr cleavage of [(AG)₃EG]₂₀ and its PEG-750 conjugate is depicted and shows a shift of both bands to a lower molecular weight. The removal of hydrophobic and basic residues by cyanogen bromide cleavage resulted in a product which was almost not stainable with the Coomassie Brilliant Blue R-250 dye, as was observed for the analogous proteins in Chapter 3. In this experiment the non-cleaved control was incubated in the presence of 70% formic acid without cyanogen bromide. Under these conditions additional bands are present below the full-length product. This is most likely the result of cleavage at the acid-labile aspartyl-prolyl bond, which is present twice in the expressed polypeptide. It has to be noted here that when methionine reacts with cyanogen bromide, it is converted to homoserine lactone, which is even under acidic conditions in equilibrium with homoserine.[37] However, the exact chemistry at the C-terminal side of the polypeptide has not been determined.

Although MALDI-TOF mass spectrometry gave relatively weak signals the presence of [(AG)₃EG]₂₀ bi-functionalized with PEG-750 as the predominant product was established. However, also smaller peaks were present for monofunctionalized and trifunctionalized polypeptide. The latter is most likely the result of functionalization of the C-terminal lysines or the N-terminus, since the maleimide functionality eventually also reacts with amines (although ~ 1000 times slower than with thiol groups).[38] A ¹H NMR spectrum of [(AG)₃EG]₂₀-PEG-750 conjugate dissolved in D₂O is depicted in Figure 4.6. Because of the large water peak at 4.8 ppm, only the spectrum from 0 to 4.5 ppm is shown. Comparison of the major signals from the

polypeptide (Ala-Hβ, 1.40 ppm) and PEG (ethylene oxide, 3.71 ppm) showed that the ethylene oxide signal was approximately 1.2 times the value expected for bifunctionalized product. Furthermore, it has to be noted that for unknown reasons the signal for glycine-Hα is too small relative to alanine-Hβ (only a factor 0.6 of the expected value), whereas this ratio was still as expected for the unfunctionalized polypeptide.
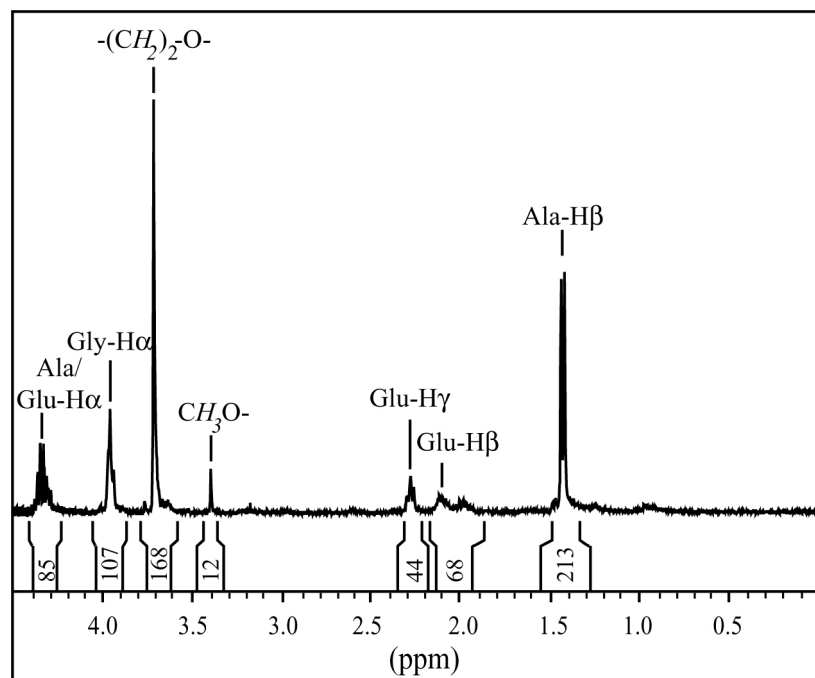
Figure 4.6 *400 MHz ^{1}H NMR spectrum of conjugate of [(AG)$_3$EG]$_{20}$ and poly(ethylene glycol)-750 in D$_2$O (NaOH was added for solubilization, 40 mM final concentration).*

For [(AG)$_3$EG]$_{10}$ a series of conjugates were made using ε-maleimidocaproic acid and PEGs with number average molecular weights of 750, 2000 and 5000 g mol^{-1}. The conjugation procedure was similar as used for the preparation of conjugates of [(AG)$_3$EG]$_{20}$ and PEG-750, with the only difference that in the trichloroacetic acid precipitation step the second wash of the pellet was performed with 1% TCA instead of demi-water, since the [(AG)$_3$EG]$_{10}$ polypeptide partly redissolved by washing with demi-water. After resuspension in coupling buffer the cysteine content was determined to be 75% of the theoretical value (based on weighed in product). In contrast to the conjugation experiments described for [(AG)$_3$EG]$_{20}$, one equivalent of maleimide functionalized PEG relative to cysteine residues was sufficient for complete functionalization, since SDS-PAGE analysis showed no further shift of bands on use of more equivalents of PEG (data not shown). However, for subsequent experiments 5 equivalents of maleimide functionalized PEGs were used. The SDS-PAGE analysis of these conjugation reactions is depicted in Figure 4.7a. Lane 1 shows the product after conjugation with ε-maleimidocaproic acid and a slight downshift was observed in comparison to unreacted polypeptide (lane 5). Some unreacted polypeptide still seemed to be present. Surprisingly, also

the attachment of PEG-750 resulted in a slightly increased electrophoretic mobility (lane 2), whereas the attachment of longer PEG-chains (PEG-2000 and PEG-5000; lane 3 and 4) again resulted in a shift of bands to higher molecular weight. The strange migration behavior on SDS-PAGE gel cannot be explained easily. One has to take into account, however, that the unreacted polypeptide already migrates anomalously slow on gel. Furthermore, particularly for the PEG-5000 conjugate, higher molecular weight bands were detected. These bands are most likely physically aggregated conjugates, since SDS-PAGE was carried out under reducing conditions (loading buffer contained 2% β-mercaptoethanol).

The expected molecular weights for $[(AG)_3EG]_{10}$ conjugated with two PEG-chains of PEG-750, PEG-2000 and PEG-5000 are 13.1, 15.6 and 21.7 kDa, respectively. MALDI-TOF mass spectrometry showed that the predominant PEG-conjugates formed were in agreement with these values. Some impurities could be observed corresponding to monofunctionalized and trifunctionalized product. Trifunctionalization is likely to be the result of reaction with lysine residues. Comparison of the major ^{1}H NMR signals from the polypeptide (Ala-Hβ, 1.40 ppm) and PEG (ethylene oxide, 3.71 ppm) showed that the ethylene oxide signal was approximately 1.1 times the value expected for bifunctionalized product for the PEG-750 conjugate, whereas this factor was 1.2 for the PEG-2000 and PEG-5000 conjugates.
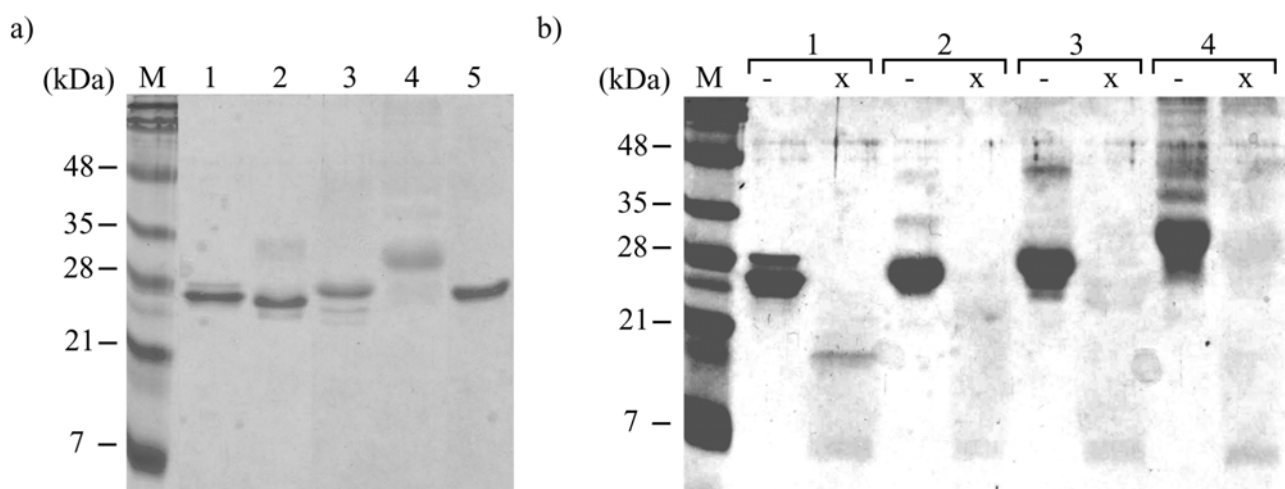
Figure 4.7 *SDS-PAGE analysis of **(a)** conjugation of $[(AG)_3EG]_{10}$ and (1) ε-maleimidocaproic acid, (2) poly(ethylene glycol) -750 (PEG-750), (3) PEG-2000 and (4) PEG-5000. (5) unreacted $[(AG)_3EG]_{10}$. **(b)** cyanogen bromide (CNBr) cleavage of (1) $[(AG)_3EG]_{10}$-ε-maleimidocaproic acid, (2) $[(AG)_3EG]_{10}$-PEG-750, (3) $[(AG)_3EG]_{10}$-PEG-2000 and (4) $[(AG)_3EG]_{10}$-PEG-5000. Samples were loaded in the presence of 1% β-mercaptoethanol. For both the non-cleaved samples (-) and the CNBr-cleaved (×) 10 μg was loaded. Visualization of bands for CNBr cleaved conjugates required contrast enhancement.*

After removal of PEG by Ni-NTA chromatography and subsequent dialysis, cyanogen bromide cleavage was carried out. After this reaction the capability of staining the conjugates with Coomassie was almost completely lost (Figure 4.7b). Only after contrast enhancement, vague

bands could be detected at slightly lower molecular weight than the original conjugates (compare lanes – and ×). With contrast enhancement and at these high loadings of 10 μg the aggregated, high-molecular weights can be detected clearly. Furthermore, for the conjugate with ε-maleimidocaproic acid two bands are present before cyanogen bromide cleavage. The upper band was located at the height of unreacted polypeptide (visible in Figure 4.7a), but could also corresponds to polypeptide monofunctionalized with ε-maleimidocaproic acid.

Although the quality of the spectra obtained by MALDI-TOF mass spectrometry was poor, the main signals for the conjugates of $[(AG)_3EG]_{10}$ with PEG-750, PEG-2000 and PEG-5000 (Figure 4.8a, b and c, respectively) were in reasonable agreement with the calculated average masses (Figure 4.8d). For the $[(AG)_3EG]_{10}$-PEG-5000 conjugate an additional peak at 9215 Da was present, which could only correspond to partially cleaved $[(AG)_3EG]_{10}$ without conjugated PEG.

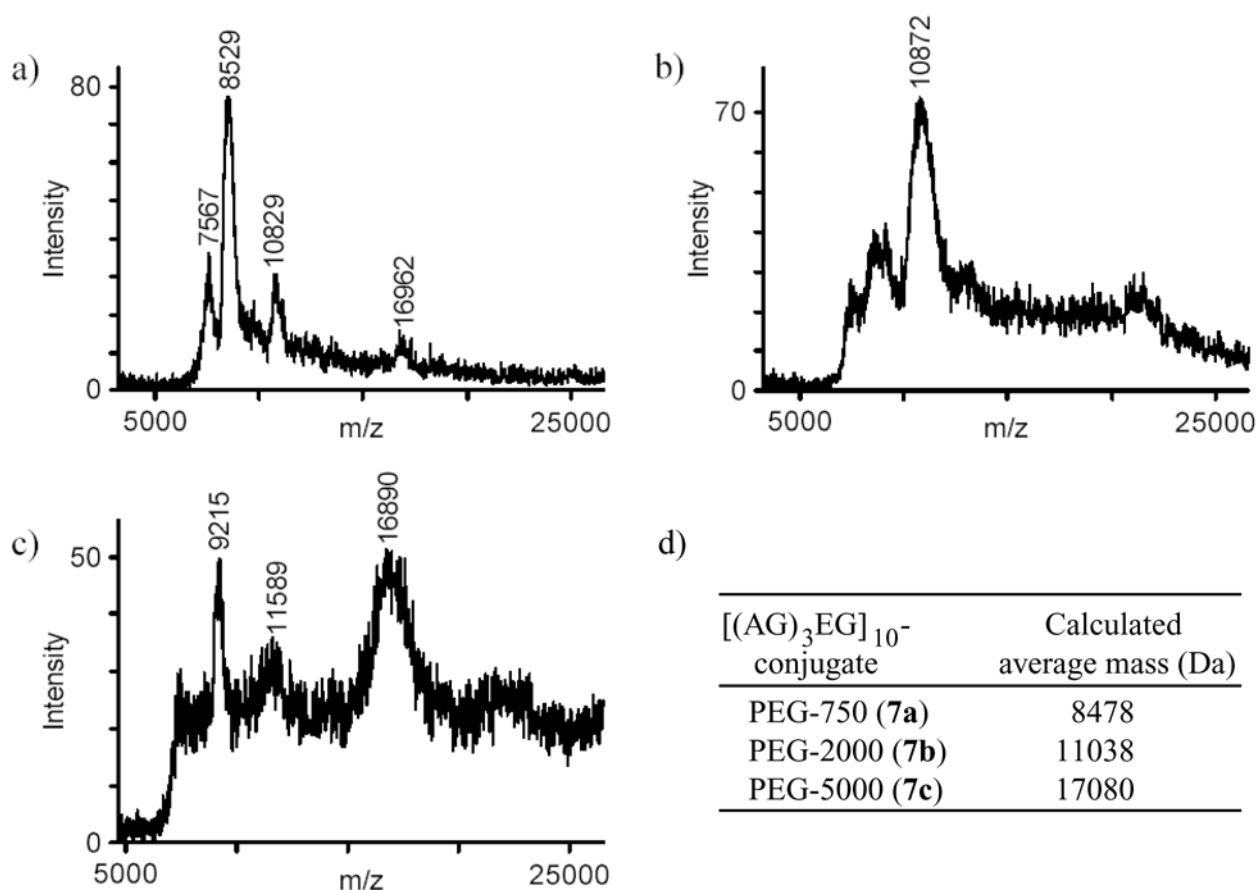***Figure 4.8*** *MALDI-TOF mass spectra of cyanogen bromide cleaved conjugates of $[(AG)_3EG]_{10}$ and **(a)** PEG-750, **(b)** PEG-2000, **(c)** PEG-5000. **(d)** Calculated average masses of $[(AG)_3EG]_{10}$ bi-functionalized with PEG (polypeptide part: H-GTCAG-$[(AG)_3EG]_{10}$-AC-homoserine-OH).*

4.2.4 Conjugation of maleimide functionalized poly(methyl methacrylate) and poly-[(AG)₃EG]

Besides the preparation of poly(ethylene glycol) (PEG) conjugates, the conjugation of the more apolar, synthetic polymer poly(methyl methacrylate) (pMMA) was attempted. The effect of the synthetic polymer block on the aggregation behaviour of the β-sheet polypeptide can thus be investigated. Atom transfer radical polymerization (ATRP) was used to prepare pMMA of which one end could be functionalized with a maleimide group (Figure 4.9). An amine-functional α-bromoester, of which the amine moiety was protected with a tert-butyloxycarbonyl (Boc) group, was used to initiate polymerization of MMA using Cu(I)Cl as a catalyst. The polymerization was carried out in deuterated DMSO, which enabled determination of polymerization kinetics. A polymer was obtained with a number average molecular weight (M_n) of 3.4 kg mol^{-1} and a polydispersity (PDI) of 1.22 (**10a**). After polymerization, the Boc-group was deprotected and the resulting free amine group was used for reaction with ε-maleimidocaproic acid, using a similar procedure as was described for the functionalization of PEG (Paragraph 4.2.3). This polymer (**10b**) was used for coupling to [(AG)₃EG]₂₀.
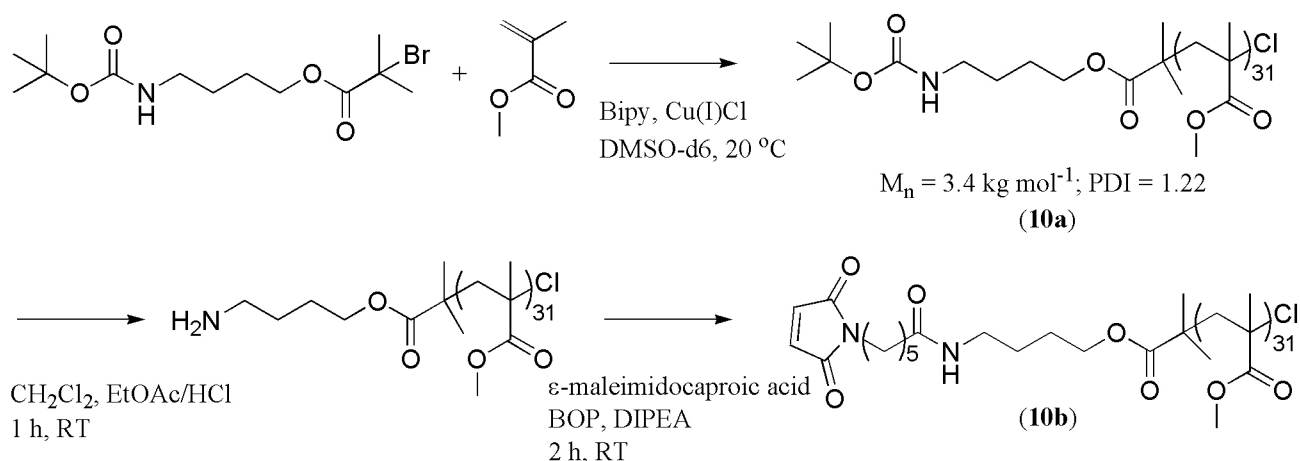
Figure 4.9 *Preparation of maleimide functionalized poly(methyl methacrylate) by atom transfer radical polymerization.*

For conjugation to [(AG)₃EG]₁₀, pMMA was used with a M_n of 3.9 kg mol^{-1} (**11**; kindly provided by Joost Opsteen). Before conjugation the cystine disulfide bridges of the polypeptides were reduced with dithiothreitol (DTT) using the same procedure as described in Paragraph 4.2.3. Maleimide functionalized pMMA was dissolved in tetrahydrofuran (THF) and slowly added to the reduced polypeptide, which was dissolved in sodium phosphate buffer (pH 6.8) (final ratio THF-buffer 3:2).

For [(AG)₃EG]₁₀ the effect of conjugation of pMMA-maleimide was analyzed by SDS-PAGE (Figure 4.10a). Whereas for Boc-protected pMMA (lane 1) no band shift was observed, the use of increasing equivalents of pMMA-maleimide (lane 3 – 6), resulted finally for 20 equivalents of pMMA-maleimide in a shift of the polypeptide band to lower molecular weight.

A shift to lower apparent molecular weight was observed earlier after the conjugation of PEG-750 (Figure 4.7). Because of the anomalous migration behaviour of the polypeptide alone, the effect of conjugation on electrophoretic mobility seems to be difficult to predict. For unclear reasons, with the use of less equivalents of pMMA a band at high molecular weight was also present.
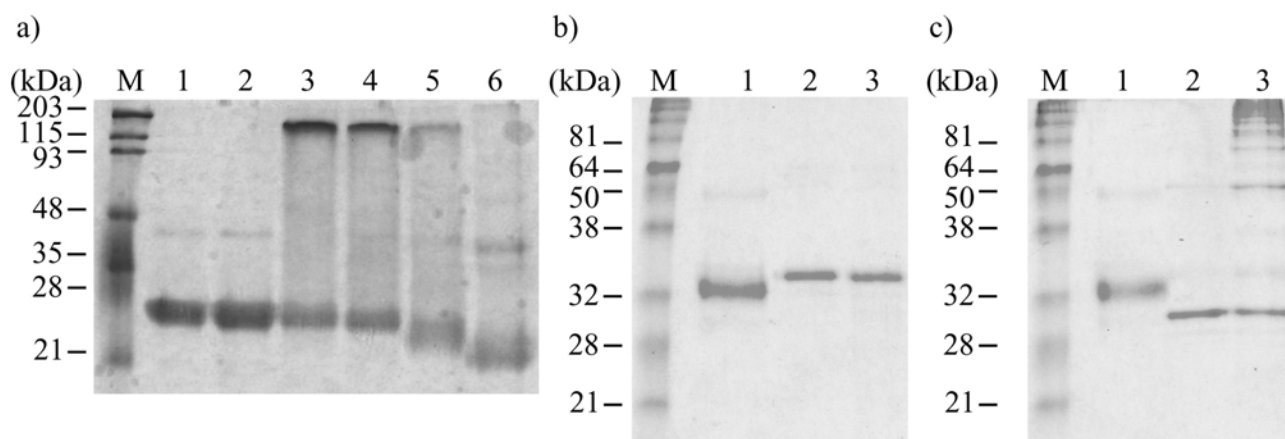
Figure 4.10 SDS-PAGE analysis of **(a)** conjugation of [(AG)$_3$EG]$_{10}$ and pMMA (M_n = 3.9 kg mol^{-1}). Lane 1: without polymer (only THF); lane 2: 5 eq. of Boc-protected (unreactive) pMMA; ; lane 3 – 6: 0.5, 1, 5 and 20 eq. (relative to cysteine residues) of maleimide functionalized pMMA. Electrophoresis was carried out under reducing conditions, using 2% β-mercaptoethanol in the loading buffer. **(b)** conjugation of [(AG)$_3$EG]$_{20}$ and pMMA (M_n = 3.4 kg mol^{-1}). Lane 1: 20 eq. pMMA-maleimide; lane 2: Boc-protected (unreactive) pMMA; lane 3: no polymer (only THF), respectively. Electrophoresis was carried out under reducing conditions. **(c)** like b, using non-reducing SDS-PAGE.

For [(AG)$_3$EG]$_{20}$, the conjugation of pMMA was analyzed by SDS-PAGE under both reducing (Figure 4.10b) and non-reducing conditions (Figure 4.10c). Also for this polypeptide conjugation with pMMA resulted in a shift of the polypeptide band to lower molecular weight under reducing conditions (compare lane 1 and 2). For Boc-protected (unreactive) pMMA (lane 3) no shift was observed. The product formed after reaction with pMMA-maleimide was not able to form disulfide bridges as was observed by SDS-PAGE under non-reducing conditions (Figure 4.10c). The incubation in the presence of Boc-protected pMMA in THF (lane 2) and THF without polymer (lane 3) resulted in disulfide bridge formation. The band below the 32 kDa marker is likely the result of intramolecular disulfide bridge formation. Furthermore, for incubation with THF (lane 3) multimers were visualized as a result of the formation of intermolecular disulfide bridges. The extent of intramolecular disulfide bridge formation seemed to be more extensive after incubation with Boc-protected pMMA (lane 2), since no laddering was observed and presumably most of the product was insoluble and remained in the wells.

Attempts were made to isolate the conjugates from the excess of pMMA used. Thereto, the solvents were evaporated, followed by extraction with water. Although SDS-PAGE analysis

seemed to indicate that the product could be extracted, ^{1}H-NMR of the product in DMSO only showed signals corresponding to the polypeptide. Also MALDI-TOF analysis was unsuccessful. No final proof for conjugation of pMMA was obtained, since no chemical characterization of the conjugates was realized.

4.3 Conclusions

The expression and purification of poly-[(AG)$_3$EG] sequences with flanking cysteine residues has been successfully carried out, although the production was complicated by low yields and the formation of truncated polypeptide product. The feasibility of the preparation of block copolymers of recombinantly prepared polypeptide sequences and synthetic polymer blocks has been demonstrated. Triblock copolymers consisting of a central poly-[(AG)$_3$EG] block and two PEG end blocks could be prepared by using the selective reaction between the sulfhydryl group of cysteines flanking the poly-[(AG)$_3$EG] sequence and the maleimide functionality of PEG. The structural characterization of these block copolymers is described in Chapter 5.

The preparation of block copolymers of poly[(AG)$_3$EG] and poly(methyl methacrylate) was more complicated. Although the successful coupling was indicated by SDS-PAGE, chemical characterization of these conjugates was not accomplished.

4.4 Experimental section

4.4.1 General methods

Protein purification. Ni-NTA, gel filtration and reversed phase chromatography were performed using a Biologic FPLC instrument (BioRad).

NMR spectroscopy. ^{1}H-NMR spectra were recorded on a Varian Inova-400 instrument at 298 K. Chemical shifts are reported in ppm relative to tetramethylsilane ($\delta = 0.00$ ppm) or the appropriate solvent signal. ^{13}C-NMR spectra were recorded on a Bruker AC-300 instrument at 298 K.

MALDI-TOF mass spectrometry. Spectra were measured on a Bruker Biflex III spectrometer. Freeze-dried products were dissolved in 1:1 (v/v) water:acetonitrile with 0.1% trifluoroacetic acid to a final concentration of 10 mg mL^{-1} and mixed in a 1:1 ratio with a solution of 20 mg mL^{-1} of sinapinic acid (Sigma) in the same solvent and spotted on a MALDI-plate.

Amino acid analysis. 1 mg of polypeptide was placed in a hydrolysis tube and dissolved in 1 mL 5.7 N HCl. After purging with nitrogen for 10 min, the solution was frozen with dry ice/acetone and evacuated with high vacuum. The tube was closed by melting and heated for 48 hours at 110 °C. The solution was then evaporated and the residual solid was evaporated three times with water at 40 °C. The amino acids in the hydrolyzate were derivatized with 9-fluorenylmethylchloroformate, separated by high performance liquid chromatography (Varian HPLC 9000 series) on a reverse phase C18 silica column (TSK-RP18) and detected at a wavelength of 269 nm.[39]

Gel permeation chromatography. Molecular weight distributions were measured using GPC, on a Shimadzu system equipped with a guard column and a PL gel 5 μm mixed D column (Polymer Laboratories) with differential refractive index and UV (254 nm) detection, using THF as an eluent at 1 mL min^{-1} at 35 °C. Poly(styrene) standards were used to calibrate the SEC with THF as an eluent.
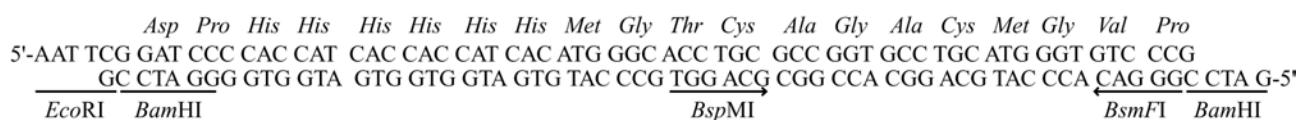
SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot analysis. SDS-PAGE was performed using 15% resolving gels on a mini-Protean II slab cell apparatus (BioRad) according to standard procedures.[40] For immunoblot analysis the separated whole cell lysate samples were transferred to nitrocellulose membranes (Protran, Schleicher & Schuell BioScience) using the Mini Trans-Blot cell (BioRad). The composition of the used transfer buffer was as follows: 25 mM Tris.HCl, 192 mM glycine (pH 8.3), 20% methanol. Subsequently the blots were incubated in TBST (10 mM Tris.HCl (pH 8.0), 150 mM NaCl, 0.1% Tween 20) containing 5% milk powder for 1 hour at room temperature, followed by incubation with T7 tag mouse monoclonal antibody (Novagen; 1 : 10000) in 5% milk powder in TBST overnight at 4 ºC. The blots were washed 3 times for 10 min with TBST and then incubated with rabbit-anti-mouse peroxidase (Dako Diagnostics; 1 : 2000) for 1 hour at room temperature. After washing 3 times for 5 min with TBST at room temperature and drying with Whatmann paper, the blots were developed using the ECL^{TM} Chemiluminescent Detection Reagents (Amersham Biosciences). The detection solution was prepared by mixing equal volumes of solution 1 and 2 on a glass plate and the blots were placed with the protein-containing side down on top of this solution for 30 seconds. After drying with Whatmann paper the blots were enclosed in plastic wrap and exposed to film (KODAK® chemiluminescence film) for 30 seconds before development.

Ellman's assay for free thiol determination.[35] To 50 μl of polypeptide solution of suitable concentration 950 μl of 100 mM NaH$_2$PO$_4$ buffer (pH 6.8) containing 150 mM NaCl and 1mM EDTA was added. To this solution 50 μl of 3 mM 5,5'-dithio-bis-2-nitrobenzoic acid (DTNB; Sigma, dissolved in the same buffer) was added. The absorbance difference between the polypeptide sample and a reference without protein was measured at 412 nm (ε_{412} = 10900 M^{-1} cm^{-1}).

Cyanogen bromide cleavage.[40] Conjugates were dissolved in 70% formic acid (2 mg mL^{-1}), and an equal volume of CNBr (100 mg mL^{-1}) (Sigma) in 70% formic acid (Merck) was added followed by incubation on a rotary arm for 2 days at room temperature in the dark. The samples were dried under vacuum. The pellets were redissolved in demi-water and dialyzed for 2 days against demi-water using a dialysis membrane (Spectra Por dialysis tubing, 3.5 kDa molecular weight cut-off).

4.4.2 Construction of expression vectors for cysteine flanked poly-[(AG)$_3$EG]

For the expression of poly-[(AG)$_3$EG] with flanking cysteine residues, a new cloning vector, pSK-JS2, was constructed by ligation of the double stranded synthetic oligonucleotide 1 (Sigma Genosys) with *Eco*RI/*Bam*HI-digested pBluescriptII®SK(-).

```
            Asp Pro His His  His His  His His Met Gly Thr Cys  Ala Gly Ala Cys Met Gly Val  Pro
5'-AAT TCG GAT CCC CAC CAT  CAC CAC CAT CAC ATG GGC ACC TGC  GCC GGT GCC TGC ATG GGT GTC  CCG
       GC CTA GGG GTG GTA  GTG GTG GTA GTG TAC CCG TGG ACG CGG CCA CGG ACG TAC CCA CAG GGC CTA G-5'
   EcoRI  BamHI                                 BspMI                          BsmFI  BamHI
```

oligonucleotide 1

The multimeric genes coding for [(AG)$_3$EG]$_{10}$ and [(AG)$_3$EG]$_{20}$, prepared as described in Chapter 3, were cloned into *Bsp*MI-digested pSK-JS2. The resulting vector was digested with *Bam*HI and the multimeric gene fragment was cloned into the *Bam*HI digested pET-3b expression vector.

4.4.3 Expression and purification of cysteine-flanked [(AG)$_3$EG]$_{10}$ (1)

Expression. The pET-3b expression vectors carrying the artificial genes coding for 10 repeats of the octapeptide sequence -(AG)$_3$EG- were transformed to *E. coli* BL21(DE3) cells (Novagen) and grown overnight at 30 °C. A single colony was used to inoculate 250 mL 2 × YT medium containing 100 μg mL^{-1} ampicillin, 34 μg mL^{-1} chloramphenicol and 1% glucose. After growth overnight at 30 °C this preculture was used to inoculate 4.5 L of 2 × YT medium to OD = 0.1 and cells were grown at 37 °C. Protein expression was induced during logarithmic growth by addition of IPTG to a final concentration of 1 mM.

Purification by Ni-NTA and gel filtration chromatography. Cells were harvested after 4 hours of expression by centrifugation at 6000 g for 15 min at 4 °C. Cells were resuspended in 50 mL lysis buffer (50 mM NaH$_2$PO$_4$ (pH 8.0), 300 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol and 1 mM PMSF) and disrupted by sonication on ice for 5 min using a 250 W Branson sonicator (50% duty cycle, 5 units power). RNase A (10 μg mL^{-1}) and DNase I (5 μg mL^{-1}) were added, followed by incubation on ice for 15 min. The lysate was centrifuged at 10000 g for 30 min to pellet the cellular debris. The supernatant was incubated with 5 mL Ni-NTA agarose beads for 1 hour at 4 °C. The suspension was then loaded onto the column followed by washing with 40 mL wash buffer (as lysis buffer, except with the addition of 20 mM imidazole). The protein was eluted with 10 mL elution buffer (as lysis buffer, but 200 mM imidazole). The elution fractions were further purified by gel filtration chromatography using a Superdex-75 Hi-Load™ 26/60 column (Amersham Biosciences) operated at room temperature. The eluent composition was as follows: 50 mM NaH$_2$PO$_4$ (pH 8.0), 150 mM NaCl, 10 mM β-mercaptoethanol. A flow of 2 mL min^{-1} was used. Typically 10 mL of sample was loaded on the column. The isolated yield after dialysis for 2 days against demi-water (Spectra Por dialysis

tubing, 3.5 kDa molecular weight cut-off) and freeze-drying was 6 mg per liter of culture. Full-length $[(AG)_3EG]_{10}$ (**1**): MS (MALDI-TOF): m/z = 11134 Da (calcd. mass: 10980 Da).

4.4.4 Expression and purification of cysteine-flanked $[(AG)_3EG]_{20}$ (2)

High cell density fermentation. Competent cells of *E. coli* strain BL21(DE3)pLysS were transformed with the pET-3b vector carrying the artificial gene coding for cysteine-flanked $[(AG)_3EG]_{20}$. The transformation mixture was plated on LB agar medium under ampicillin (200 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) selection for 16 h at 30 °C. A single colony of a positive transformant was used to inoculate 200 mL of 2 × YT medium, supplemented with 1% glucose, ampicillin (200 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}). This preculture was incubated at 30 °C for 16 h. The cells were isolated by centrifugation at 4000 g and 4 °C for 15 min and resuspended in 100 mL minimal medium. The 5 L fermentor (Bioflo 110, New Brunswick Scientific Co.), containing 4 L of minimal medium was inoculated to an optical density of 0.1. The minimal medium was made according to a recipe described by Panitch and coworkers.[12] MgSO$_4$, thiamine hydrochloride and glucose were separately filter sterilized in ¼ of the final medium volume. The other medium components were autoclaved in the fermentor in ¾ of the final medium volume. Before start of the fermentation, the filter sterilized components were added to the fermentor and the pH was adjusted to 6.8 with 4N NH$_4$OH. Antifoam (Antifoam A concentrate, Sigma), ampicillin (200 μg mL^{-1}) and chloramphenicol (34 μg mL^{-1}) were added just before inoculation. The culture was grown at 37 °C. During the fermentation the pH was kept constant (6.8) by the addition of 4 N NH$_4$OH (also used as nitrogen source). The dissolved oxygen was set at 20% air saturation. The agitation speed was 500 rpm and the aeration rate was 2 L min^{-1}. Dissolved oxygen was kept constant by cascading to pure O$_2$. The doubling time of the cells was approximately 1 hour under these conditions. The depletion of glucose (after approximately 10 hours and an optical density of approximately 20) was observed by an increase of the dissolved oxygen concentration and a rise in pH above the setpoint. A glucose feed (50% w/v) was established at that point. Glucose was fed through the acid pump of the pH controller, thus maintaining the pH at 6.8. Expression was induced by addition of IPTG to a final concentration of 1 mM and the temperature was lowered to 30 °C. The cells were harvested after 4 hours by centrifugation for 15 min at 6000 g and 4 °C and stored at -20 °C.

Purification by Ni-NTA chromatography, heat precipitation and reversed phase chromatography. 70 grams of cells (wet weight) were resuspended in 350 mL lysisbuffer (50 mM NaH$_2$PO$_4$, pH 8.0; 300 mM NaCl; 1 mM PMSF). Lysozyme was added to 1 mg mL^{-1} and the suspension was incubated on ice for 30 min. Sonication was carried out for 15 min on ice using a Branson sonicator (50% duty cycle, 5 units power). RNase A (10 μg mL^{-1}) and DNase I (5 μg mL^{-1}) were added, followed by incubation on ice for 15 min. The lysate was centrifuged

for 30 min at 10000 g. To the lysate 110 mL of 50% pre-equilibrated Ni-NTA slurry was added and mixed by shaking on a rotary shaker at 4 °C for 1 hour. The Ni-NTA agarose beads were spun down by centrifugation at 1500 g for 5 min and the supernatant was removed. The beads were loaded into a column using wash buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 1 mM PMSF; 10 mM imidazole). The beads were washed with in total 300 mL wash buffer. Then a 30 mL imidazole gradient was applied from 10 mM to 250 mM, followed by applying 90 mL elution buffer (50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 1 mM PMSF; 250 mM imidazole). Beta-mercaptoethanol was added to a final concentration of 10 mM. The eluate was then heated for 10 min at 70 °C, followed by centrifugation at 6000 g for 15 min. Reversed phase chromatography was performed with a RPC-15μm C2/C8 column (HR16/10; Amersham Biosciences) operated at room temperature. The composition of the eluents was as follows: Eluent A: water + 0.1% TFA; eluent B: 80% MeCN + 0.1% TFA. A 250 mL gradient was applied from 0 to 50% eluent B with a flow rate of 2 mL min^{-1}. The sample volume was 5 mL. The final yield of full length product was 10 mg L^{-1} of culture.

Full-length $[(AG)_3EG]_{20}$ (**2**): MS (MALDI-TOF): m/z = 16811 Da (calcd. average mass for unmodified polypeptide: 16686). ^{1}H-NMR (400 MHz, D_2O): δ = 4.34 (m, 93H, Ala/Glu-Hα), 3.96 (s, 172H, Gly-Hα), 2.26 (m, 44H, Glu-Hγ), 2.10 (m, 44H, Glu-Hβ), 1.40 (d, 210H, Ala-Hβ). Truncated polypeptide: Amino acid analysis: (mol percent) Ala, 28.6; Gly, 37.3; Glx, 12.5; Arg, 1.5; Ser, 1.3; Asx, 1.8; Thr, 2.5; Met, 2.1; Val, 0.3; His, 9.1; Lys, 0.2. MS (MALDI-TOF): m/z (major peak) = 6237 Da, m/z (additional peaks) = 6109, 6166, 6294, 6365 and 6422 Da.

4.4.5 Synthesis of maleimide functionalized poly(ethylene glycols)

Materials. Poly(ethylene glycols) monofunctionalized with an amine group with molecular weights of 750, 2000 and 5000 g mol^{-1} were purchased from Rapp Polymere GmbH (Tübingen, Germany). Epsilon-maleimidocaproic acid (Sigma), benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP) (Advanced Chemtech), diisopropylethylamine (Fluka), 5,5'-dithio-bis-2-nitrobenzoic acid (Sigma) and sodium 2-mercaptoethanesulfonate (Sigma) were used as received. Thin layer chromatography analyses were performed on Merck precoated silica gel 60 F254 plates (layer thickness 0.25 mm) using the solvent mixtures indicated. Product purification was performed with a Sephadex LH-20 gel filtration resin obtained from Amersham Biosciences.

Maleimide detection for thin-layer chromatography.[41] Maleimide derivatives were detected by spraying the thin layer plates with a 0.1% solution of 5,5'-dithio-bis-2-nitrobenzoic acid in 1:1 ethanol : 100 mM Tris.HCl buffer (pH 8.2) and then with a 2% solution of sodium 2-mercaptoethanesulfonate in 80% ethanol until the background was bright yellow. Maleimide derivatives appeared as white spots.

Maleimide functionalized poly(ethylene glycol)-750 (3a). 391 mg PEG-750-NH$_2$ (0.52 mmol) was dried by azeotropic evaporation with benzene. The dried product was dissolved in 5 mL DMF together with 110 mg ε-maleimidocaproic acid (0.52 mmol), 230 mg BOP coupling reagent (0.52 mmol) and 192 mg diisopropylethylamine (1.60 mmol). After 24 hours stirring at room temperature DMF was evaporated and the resulting solid was redissolved in dichloromethane. This solution was subsequently extracted with 1N HCl (twice), water, 5% NaHCO$_3$ (twice), water and saturated NaCl. After evaporation the product was further purified by gel filtration chromatography using methanol/dichloromethane 1 : 1 v/v as eluent. The final yield was 217 mg pure product (0.35 mmol, 68%). R$_f$ = 0.60 – 0.75 (methanol/chloroform = 1:4 v/v, maleimide detection); ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 6.69 (s, 2H, C*H*=C*H*), 6.27 (br s, 1H, N*H*CO), 3.61 - 3.68 (br m, 68H, O(C*H$_2$*)$_2$O), 3.55 (t, 2H, C*H$_2$*CH$_2$NHCO), 3.51 (t, 2H, NC*H$_2$*), 3.44 (m, 2H, C*H$_2$*NHCO), 3.38 (s, 3H, C*H$_3$*O), 2.17 (t, 2H, C*H$_2$*CONH), 1.66 (m, 2H, NHCOCH$_2$C*H$_2$*), 1.60 (m, 2H, C*H$_2$*CH$_2$N), 1.31 (m, 2H, CH$_2$C*H$_2$*CH$_2$CH$_2$N); ^{13}C-NMR (75 MHz, CDCl$_3$): δ = 172.1 (1C, *C*ONH), 163.8 (2C, N*C*O), 133.5 (2C, *C*=*C*), 71.6 (1C, *C*H$_2$CH$_2$NH), 70.2 (30C, O(*C*H$_2$)$_2$O), 58.7 (1C, *C*H$_3$O), 38.9 (1C, *C*H$_2$NH), 37.5 (1C, *C*H$_2$N), 36.1 (1C, *C*H$_2$CONH), 28.1 (1C, *C*H$_2$CH$_2$N), 26.2 (1C, *C*H$_2$CH$_2$CH$_2$), 24.9 (1C, *C*H$_2$CH$_2$CONH); MS (MALDI-TOF): Calcd. for n = 15, C$_{43}$H$_{80}$N$_2$O$_{19}$ [M + Na]$^+$: 952.05, found m/z = 951.70, M$_n$ (MALDI-TOF) = 1045 g mol^{-1}, M$_w$/M$_n$ = 1.01.

Maleimide functionalized poly(ethylene glycol)-2000 (3b). The maleimide functionalization and subsequent purification was carried out analogous to the route described for PEG-750. The reaction was performed with 800 mg PEG-2000-NH$_2$ (0.4 mmol), 85 mg ε-maleimidocaproic acid (0.4 mmol), 177 mg BOP (0.4 mmol) and 155 mg DIPEA (1.2 mmol). The final yield after gel filtration chromatography (Sephadex LH-20; methanol/dichloromethane 1:1 v/v as eluent) was 784 mg (0.36 mmol, 89%). R$_f$ = 0.50 – 0.65 (methanol/chloroform = 1:5 v/v, maleimide detection); ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 6.69 (s, 2H, C*H*=C*H*), 6.20 (br s, 1H, N*H*CO), 3.61 - 3.68 (br m, 188H, O(C*H$_2$*)$_2$O), 3.55 (t, 2H, C*H$_2$*CH$_2$NHCO), 3.51 (t, 2H, NC*H$_2$*), 3.44 (m, 2H, C*H$_2$*NHCO), 3.38 (s, 3H, C*H$_3$*O), 2.17 (t, 2H, C*H$_2$*CONH), 1.66 (m, 2H, NHCOCH$_2$C*H$_2$*), 1.60 (m, 2H, C*H$_2$*CH$_2$N), 1.31 (m, 2H, CH$_2$C*H$_2$*CH$_2$CH$_2$N); MS (MALDI-TOF): Calcd. for n = 48, C$_{109}$H$_{212}$N$_2$O$_{52}$ [M + H]$^+$: 2383.72, found m/z = 2383.127, M$_n$ (MALDI-TOF) = 2325 g mol^{-1}, M$_w$/M$_n$ = 1.02.

Maleimide functionalized poly(ethylene glycol)-5000 (3c). The maleimide functionalization and subsequent purification was carried out analogous to the route described for PEG-750. The reaction was performed with 850 mg PEG-5000-NH$_2$ (0.17 mmol), 36 mg ε-maleimidocaproic acid (0.17 mmol), 75 mg BOP (0.17 mmol) and 66 mg DIPEA (0.51 mmol). The final yield after gel filtration chromatography (Sephadex LH-20; methanol/dichloromethane 1:1 v/v as eluent) was 617 mg (0.12 mmol, 70%). R$_f$ = 0.50 – 0.65 (methanol/chloroform = 1:5 v/v, maleimide detection); ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 6.69 (s, 2H, C*H*=C*H*), 6.20 (br s, 1H, N*H*CO), 3.60 - 3.67 (br m, 460H, O(C*H$_2$*)$_2$O), 3.55 (t, 2H, C*H$_2$*CH$_2$NHCO), 3.51 (t, 2H, NC*H$_2$*), 3.44 (m, 2H, C*H$_2$*NHCO), 3.38 (s, 3H, C*H$_3$*O), 2.17 (t, 2H, C*H$_2$*CONH), 1.66 (m, 2H,

NHCOCH$_2$C*H$_2$*), 1.60 (m, 2H, C*H$_2$*CH$_2$N), 1.31 (m, 2H, CH$_2$C*H$_2$*CH$_2$CH$_2$N); MS (MALDI-TOF): M$_n$ = 5346 g mol^{-1}, M$_w$/M$_n$ = 1.01.

4.4.6 Conjugation of poly-[(AG)$_3$EG] and poly(ethylene glycols)

10 mg polypeptide was dissolved in 5 mL 20 mM NaH$_2$PO$_4$ buffer (pH 8.0) containing 150 mM NaCl, 1 mM EDTA and 200 mM dithiotreitol (DTT; Sigma) and incubated for 1 hour at room temperature. To remove the excess DTT the polypeptide was precipitated with trichloroacetic acid (TCA; Sigma) using a protocol adapted from Maniatis.[36] Precipitation by addition of 0.25 volumes of ice-cold 100% TCA (w/v) was followed by incubation at -20 °C for 30 min. After centrifugation at 13000 rpm for 10 min at 4 °C the pellet was washed with 2.5 mL ice-cold 20% TCA followed by a second wash with 2.5 mL 1% TCA for [(AG)$_3$EG]$_{10}$ and 2.5 mL milli-Q for [(AG)$_3$EG]$_{20}$. After each wash the solution was centrifuged for 5 min at 4 °C. The pellet was redissolved in 5 mL 100 mM NaH$_2$PO$_4$ buffer (pH 6.8) containing 150 mM NaCl. For complete PEGylation of the cysteine residues an excess of maleimide functionalized PEG was immediately added as a 5 mL solution in the same buffer. For [(AG)$_3$EG]$_{10}$ a 5-fold excess was used whereas a 20-fold excess was used for [(AG)$_3$EG]$_{20}$. For [(AG)$_3$EG]$_{10}$ a conjugation reaction with ε-maleimidocaproic acid (5 equivalents) was carried out. The reaction mixture was incubated overnight on a rotating arm.

Ni-NTA chromatography was used to remove the excess PEG or ε-maleimidocaproic acid. The Ni-NTA beads were pre-equilibrated in 100 mM NaH$_2$PO$_4$ buffer (pH 8.0) containing 150 mM NaCl. The PEG-conjugation solution was added to 10 mL equilibrated 50% Ni-NTA suspension and incubated for 1 hour at room temperature. The suspension was centrifuged for 5 min at 1500 rpm and the supernatant was removed. Wash buffer (10 mL of 100 mM NaH$_2$PO$_4$ (pH 8.0), 150 mM NaCl) was added and a column was loaded. The agarose beads were washed until all the PEG was removed (monitored by measuring absorption profile at 214 nm). Subsequently the polypeptide-PEG conjugate was eluted by increasing the imidazole concentration to 200 mM. Finally, the product was dialysed against demi-water using a dialysis membrane (Spectra Por dialysis tubing, 3.5 kDa molecular weight cut-off) for 2 days and lyophilized.

[(AG)$_3$EG]$_{20}$-PEG-750 conjugate (4a). ^{1}H-NMR (400 MHz, D$_2$O): The following major signals were assigned: δ = 4.34 (m, 94H, Ala/Glu-Hα), 3.96 (s, 172H, Gly-Hα), 3.71 (s, 136H, O(C*H$_2$*)$_2$O), 3.39 (s, 6H, C*H$_3$*O), 2.26 (m, 46H, Glu-Hγ), 2.10 (m, 46H, Glu-Hβ), 1.40 (d, 213H, Ala-Hβ). MS (MALDI-TOF): m/z = 18679 (calcd. mass for bifunctionalized product: 18907 Da) and 19807 Da (calcd. mass for trifunctionalized product: 19952 Da). After cyanogen bromide cleavage (**5a**): MALDI-TOF analysis gave no signal. Cleavage was confirmed by SDS-PAGE analysis (Figure 4.5).

[(AG)$_3$EG]$_{10}$-PEG-750 conjugate (6a). ^{1}H-NMR (400 MHz, D$_2$O): The following major signals were assigned: δ = 4.33 (m, 54H, Ala/Glu-Hα), 3.96 (s, 92H, Gly-Hα), 3.71 (s, 136H,

O(CH_2)$_2$O), 3.39 (s, 6H, CH_3O), 2.26 (m, 26H, Glu-Hγ), 2.10 (m, 26H, Glu-Hβ), 1.40 (d, 123H, Ala-Hβ). (MALDI-TOF): m/z = 13280 (calcd. mass for bifunctionalized product: 13070 Da), 12317, 14315 Da (calcd. mass for mono- and trifunctionalized product: 12025 and 14115 Da, respectively) and 26139 Da (dimeric products). After cyanogen bromide cleavage (**7a**): MS (MALDI-TOF): m/z = 8529 (calcd. mass for bifunctionalized polypeptide (polypeptide part: H-GTCAG-[(AG)$_3$EG]$_{10}$-AC-homoserine-OH): 8478 Da) and 10829 Da (incomplete cleavage).

[(AG)$_3$EG]$_{10}$-PEG-2000 conjugate (6b). ^{1}H-NMR (400 MHz, H$_2$O/D$_2$O 9:1 v/v): The following major signals were assigned: δ = 4.28 (m, 54H, Ala/Glu-Hα), 3.91 (s, 92H, Gly-Hα), 3.71 (s, 376H, O(CH_2)$_2$O), 3.39 (s, 6H, CH_3O), 2.26 (m, 26H, Glu-Hγ), 2.10 (m, 26H, Glu-Hβ), 1.40 (d, 123H, Ala-Hβ). MS (MALDI-TOF): m/z = 15417 (calcd. mass for bifunctionalized product: 15630 Da), 13332 and 17462 Da (calcd. mass for mono- and trifunctionalized product: 13305 and 17955 Da, respectively). After cyanogen bromide cleavage (**7b**): MS (MALDI-TOF): m/z = 10872 Da (calcd. mass for bifunctionalized polypeptide (polypeptide part: H-GTCAG-[(AG)$_3$EG]$_{10}$-AC-homoserine-OH): 11038 Da).

[(AG)$_3$EG]$_{10}$-PEG-5000 conjugate (6c). ^{1}H-NMR (400 MHz, H$_2$O/D$_2$O 9:1 v/v): The following major signals were assigned: δ = 1.40 (d, 123H, Ala-Hβ), 2.10 (m, 26H, Glu-Hβ), 2.26 (m, 26H, Glu-Hγ), 3.39 (s, 6H, CH_3O), 3.71 (s, 920H, O(CH_2)$_2$O), 3.91 (s, 92H, Gly-Hα), 4.28 (m, 54H, Ala/Glu-Hα). MS (MALDI-TOF): m/z = 21552 (calcd. mass for bifunctionalized product: 21672 Da), 10926, 16219 and 26323 Da (calcd. mass for non-, mono- and trifunctionalized product: 10980, 16326 and 27018 Da, respectively). After cyanogen bromide cleavage (**7c**): MS (MALDI-TOF): m/z = 16890 (calcd. mass for bifunctionalized polypeptide (polypeptide part: H-GTCAG-[(AG)$_3$EG]$_{10}$-AC-homoserine-OH): 17080 Da), 11589 (calcd. mass for monofunctionalized product: 11734 Da) and 9215 Da (not assigned).

[(AG)$_3$EG]$_{10}$-ε-maleimidocaproic acid conjugate (8). ^{1}H-NMR (400 MHz, H$_2$O/D$_2$O 9 : 1 v/v): The following major signals were assigned: δ = 1.40 (d, 123H, Ala-Hβ), 2.10 (m, 26H, Glu-Hβ), 2.26 (m, 26H, Glu-Hγ), 3.91 (s, 92H, Gly-Hα), 4.28 (m, 54H, Ala/Glu-Hα). MS (MALDI-TOF): m/z = 11588 (calcd. mass for bifunctionalized product: 11402 Da), 11356 (calcd. mass for monofunctionalized product: 11191 Da) and 22215 Da (dimeric product). After cyanogen bromide cleavage (**9**): MS (MALDI-TOF): 13600, 20233, 26842 (di-, tri- and tetrameric product), 9154 Da.

4.4.7 Synthesis of maleimide functionalized poly(methyl methacrylate)

PMMA used for conjugation to [(AG)$_3$EG]$_{20}$ (10b). PMMA was prepared via atom transfer radical polymerization. 46.4 mg Cu(I)Cl (0.47 mmol), 127.6 mg 2,2-bipyridyl (0.82 mmol) were added to a Schlenk tube and the tube was placed under nitrogen. 67.6 mg tert-butyloxycarbonyl protected 2-bromoisobutyryloxy(4-amino-butane) (0.39 mmol) (kindly provided by Joost Opsteen) and 1.61 g methyl methacrylate (16 mmol) were dissolved in 8 mL DMSO-d6 and 1 mL toluene and purged with nitrogen for 5 min. This purged solution was

added to the Schlenk tube and the mixture was purged with nitrogen for another 15 min. Polymerization was carried out for 5 hours at 20 °C and a conversion of 57% was reached. After polymerization a small amount of dichloromethane was added and the solution was extracted twice with 2.5% EDTA in water. The dichloromethane layer was dried with Na_2SO_4 and filtered. The solvent was removed in vacuo. Yield (**10a**): 986 mg. ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 4.03 (m, 2H, C*H$_2$*CO$_2$), 3.60 (s, 48H, CH$_2$C(CH$_3$)CO$_2$C*H$_3$*, 3.16 (t, 2H, C*H$_2$*NH), 2.1 − 1.7 (br m, 32H, C*H$_2$*C(CH$_3$)CO$_2$CH$_3$, 1.45 (s, 9H, (C*H$_3$*)$_3$CO), 1.13 (s, 6H, (C*H$_3$*)$_2$CO$_2$), 1.02 (s, 16H, CH$_2$C(C*H$_3$*)CO$_2$CH$_3$, 0.85 (s, 29H, CH$_2$C(C*H$_3$*)CO$_2$CH$_3$). Not assigned: 4H, (C*H$_2$*)$_2$CH$_2$O. GPC: M_n = 3.4 kg mol^{-1}, PDI = 1.22. Subsequently the maleimide functionality was introduced according to the following procedure. 386 mg (0.11 mmol) **10a** was dissolved in dichloromethane. Ethyl acetate/HCl was added, resulting in a white precipitate. 1M NaHCO$_3$ was added until the pH was 7 − 8. The mixture was extracted twice with dichloromethane. The dichloromethane layer was washed with water, dried with MgSO$_4$ and filtered. The solvent was removed in vacuo. The deprotected polymer was dissolved in 2 mL DMF. 28.3 mg ε-maleimidocaproic acid (0.11 mmol), 48.5 mg benzotriazole-1-yl-oxy-tris-(dimethylamino)-phosphonium hexafluorophosphate (BOP; 0.11 mmol) and diisopropylethylamine (to raise the pH to 7 - 8) were added. The coupling was carried out for 2 hours at room temperature. The polymer was precipitated by addition of water. The crude product was filtered and washed successively with water, 1M HCl, 1M NaHCO$_3$, water and hexane. The polymer was dissolved in dichloromethane and dried with MgSO$_4$. After filtration, the solvent was removed in vacuo. Yield (**10b**): 308 mg (75%). ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 6.70 (s, 2H, C*H*=C*H*), 4.04 (m, 2H, C*H$_2$*CO$_2$), 3.60 (s, 48H, CH$_2$C(CH$_3$)CO$_2$C*H$_3$*), 3.52 (t, 2H, NC*H$_2$*), 3.28 (m, 2H, NHC*H$_2$*), 2.17 (t, 2H, C*H$_2$*CONH), 2.1 − 1.7 (br m, 32H, C*H$_2$*C(CH$_3$)CO$_2$CH$_3$, 1.13 (s, 6H, (C*H$_3$*)$_2$CO$_2$), 1.03 (s, 16H, CH$_2$-C(C*H$_3$*)CO$_2$CH$_3$), 0.85 (s, 29H, CH$_2$-C(C*H$_3$*)CO$_2$CH$_3$). Not assigned: 2H, NHCOCH$_2$C*H$_2$*; 2H, C*H$_2$*CH$_2$N; 4H, (C*H$_2$*)$_2$CH$_2$O. GPC: M_n = 3.6 kg mol^{-1}, PDI = 1.32.

PMMA used for conjugation to [(AG)$_3$EG]$_{10}$ (11). Tert-butyloxycarbonyl (Boc) protected poly(methyl methacrylate) was kindly provided by Joost Opsteen. 316 mg pMMA-Boc was deprotected by stirring in 10 mL ethyl acetate/HCl overnight. The solvent was evaporated and the excess HCl was removed by the addition of 20 mL 1 : 1 tert-butanol : dichloromethane. After 1 hour the solvent was evaporated. 313 mg of dried product was dissolved in 10 mL dichloromethane together with 33 mg ε-maleimidocaproic acid (0.16 mmol), 69 mg BOP coupling reagent (0.16 mmol) and 61 mg diisopropylethylamine (0.47 mmol). The reaction was allowed to proceed for 24 hours at room temperature. The reaction mixture was successively extracted with an equal volume of 1N HCl (twice), water, 5% NaHCO$_3$ (twice), water and a saturated NaCl solution. Subsequently the polymer was precipitated in an excess heptane. After filtration, the polymer was vacuum-dried. Yield: 254 mg (76 %). ^{1}H-NMR (400 MHz, CDCl$_3$): δ = 6.69 (s, 2H, C*H*=C*H*), 4.04 (m, 2H, C*H$_2$*CO$_2$), 3.60 (s, 119H, CH$_2$C(CH$_3$)CO$_2$C*H$_3$*), 3.52 (t, 2H, NC*H$_2$*), 3.28 (m, 2H, NHC*H$_2$*), 2.17 (t, 2H, C*H$_2$*CONH), 1.81 (s, 33H, C*H$_2$*C(CH$_3$)CO$_2$CH$_3$), 1.13 (s, 6H, (C*H$_3$*)$_2$CO$_2$), 1.02 (s, 46H, CH$_2$-C(C*H$_3$*)CO$_2$CH$_3$), 0.85 (s,

79H, CH_2-C(CH_3)CO$_2$CH$_3$). Not assigned: 2H, NHCOCH$_2$CH_2; 2H, CH_2CH$_2$N; 4H, (CH_2)$_2$CH$_2$O. GPC: M_n = 3.9 g mol^{-1}, PDI = 1.19.

4.4.8 Conjugation of poly-[(AG)$_3$EG] and poly(methyl methacrylate)

The polypeptides [(AG)$_3$EG]$_{10}$ and [(AG)$_3$EG]$_{20}$ (5.7 and 8.5 mg, respectively) were dissolved to a concentration of 2 mg mL^{-1} in 20 mM NaH$_2$PO$_4$ buffer (pH 8.0) containing 150 mM NaCl and 1 mM EDTA. A 1 M solution of dithiothreitol (DTT) in the same buffer was added to a final concentration of 200 mM. After incubation at room temperature for 1 hour, the reduced polypeptide was precipitated by the addition of 0.25 volumes of ice-cold 100% (w/v) trichloroacetic acid (TCA), followed by incubation at -20 °C for 30 min. After centrifugation at 13000 rpm for 10 min at 4 °C the pellet was washed with 1 volume of ice-cold 20% TCA followed by a second wash with 1 volume of 1% TCA for [(AG)$_3$EG]$_{10}$ and 1 volume of demi-water for [(AG)$_3$EG]$_{20}$. After each wash the solution was centrifuged for 5 min at 4 °C. The pellet was redissolved in 1 volume 100 mM NaH$_2$PO$_4$ buffer (pH 6.8) containing 150 mM NaCl. To this reduced polypeptide solution a pMMA solution (0.45 - 18 mg mL^{-1}; corresponding to 0.5 - 20 equivalents per thiol group) in tetrahydrofuran (THF) was slowly added, resulting in a THF : buffer ratio of 3 : 2. This solution was incubated overnight on a rotating arm for [(AG)$_3$EG]$_{10}$ to 3 days for [(AG)$_3$EG]$_{20}$. For SDS-PAGE analysis, 10 µl samples were dried by vacuum evaporation and redissolved in 10 µl 1 × Laemmli buffer and loaded on gel according to standard procedures.

4.5 References

[1]   A. Abuchowski, J. R. McCoy, N. C. Palczuk, T. Vanes, F. F. Davis, *J. Biol. Chem.* **1977**, *252*, 3582.

[2]   A. Abuchowski, T. Vanes, N. C. Palczuk, F. F. Davis, *J. Biol. Chem.* **1977**, *252*, 3578.

[3]   A complete issue on protein/peptide PEGylation, in *Adv. Drug Deliv. Rev., Vol. 54*, **2002**.

[4]   R. Duncan, *Nat. Rev. Drug Discovery* **2003**, *2*, 347.

[5]   A. S. Hoffman, P. S. Stayton, *Macromol. Symp.* **2004**, *207*, 139.

[6]   A. S. Hoffman, P. S. Stayton, V. Bulmus, G. H. Chen, J. P. Chen, C. Cheung, A. Chilkoti, Z. L. Ding, L. C. Dong, R. Fong, C. A. Lackey, C. J. Long, M. Miura, J. E. Morris, N. Murthy, Y. Nabeshima, T. G. Park, O. W. Press, T. Shimoboji, S. Shoemaker, H. J. Yang, N. Monji, R. C. Nowinski, C. A. Cole, J. H. Priest, J. M. Harris, K. Nakamae, T. Nishino, T. Miyata, *J. Biomed. Mater. Res.* **2000**, *52*, 577.

[7]   T. S. Burkoth, T. L. S. Benzinger, D. N. M. Jones, K. Hallenga, S. C. Meredith, D. G. Lynn, *J. Am. Chem. Soc.* **1998**, *120*, 7655.

[8]   A. Rosler, H. A. Klok, I. W. Hamley, V. Castelletto, O. O. Mykhaylyk, *Biomacromolecules* **2003**, *4*, 859.

[9]   O. Rathore, D. Y. Sogah, *Macromolecules* **2001**, *34*, 1477.

[10]   O. Rathore, D. Y. Sogah, *J. Am. Chem. Soc.* **2001**, *123*, 5231.

[11]   K. Velonia, A. E. Rowan, R. J. M. Nolte, *J. Am. Chem. Soc.* **2002**, *124*, 4224.

[12]   A. Panitch, K. Matsuki, E. J. Cantor, S. J. Cooper, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1997**, *30*, 42.

[13]   D. Riesenberg, V. Schulz, W. A. Knorre, H. D. Pohl, D. Korz, E. A. Sanders, A. Ross, W. D. Deckwer, *J. Biotechnol.* **1991**, *20*, 17.

[14]   S. Y. Lee, H. N. Chang, *Biotechnol. Lett.* **1993**, *15*, 971.

[15]   S. Y. Lee, *Trends Biotechnol.* **1996**, *14*, 98.

[16]   D. Riesenberg, R. Guthke, *Appl. Microbiol. Biotechnol.* **1999**, *51*, 422.

[17]   H. H. Wong, Y. C. Kim, S. Y. Lee, H. N. Chang, *Biotechnol. Bioeng.* **1998**, *60*, 271.

[18]   H. P. Sorensen, K. K. Mortensen, *J. Biotechnol.* **2005**, *115*, 113.

[19]   H. L. MacDonald, J. O. Neway, *Appl. Environ. Microbiol.* **1990**, *56*, 640.

[20]   G. L. Kleman, W. R. Strohl, *Appl. Environ. Microbiol.* **1994**, *60*, 3952.

[21]   Y. F. Ko, W. E. Bentley, W. A. Weigand, *Appl. Biochem. Biotechnol.* **1995**, *50*, 145.

[22]   N. Shimizu, S. Fukusono, K. Fujimori, K. Gotoh, Y. Yamazaki, *J. Ferment. Bioeng.* **1992**, *74*, 196.

[23]   C. Turner, M. Gregory, M. Turner, *Biotechnol. Lett.* **1994**, *16*, 891.

[24]   D. J. Seo, B. H. Chung, Y. B. Hwang, Y. H. Park, *J. Ferment. Bioeng.* **1992**, *74*, 196.

[25]   K. Tomson, T. Paalme, P. S. Laakso, R. Vilu, *Biotechnol. Tech.* **1995**, *9*, 793.

[26]   K. Hellmuth, D. J. Korz, E. A. Sanders, W.-D. Deckwer, *J. Biotechnol.* **1994**, *32*, 289.

[27]   D. E. Chang, D. J. Smalley, T. Conway, *Mol. Microbiol.* **2002**, *45*, 289–306.

[28]   A. D. Grossman, W. E. Taylor, Z. F. Burton, R. R. Burgess, C. A. Gross, *J. Mol. Biol.* **1985**, *186*, 357.

[29]   S. W. Harcum, W. E. Bentley, *Appl. Biochem. Biotechnol.* **1999**, *80*, 1.

[30]   D. M. Ramirez, W. E. Bentley, *Biotechnol. Bioeng.* **1993**, *41*, 557.

[31]   S. O. Ramchuran, E. N. Karlsson, S. Velut, L. de Mare, P. Hagander, O. Holst, *Appl. Microbiol. Biotechnol.* **2002**, *60*, 408.

[32]   M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. J. Mason, D. A. Tirrell, *Science* **1994**, *265*, 1427.

[33]   R. C. Beavis, B. T. Chait, H. S. Creel, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Am. Chem. Soc.* **1992**, *114*, 7584.

[34]   S. R. Fahnestock, *Biopolymers: Polyamides and Complex Proteinaceous Materials II, Vol. 8*, Wiley-VCH Gmbh & Co. KGaA, Weinheim, **2003**.

[35]   G. L. Ellman, *Arch. Bioch. Biophys.* **1959**, *82*, 70.

[36]   J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*, 2 ed., Cold Spring Laboratory Press, Cold Spring, **1989**.

[37]   E. Gross, *Methods in Enzymology, Vol. 11*, Academic Press, New York, **1967**.

[38]   G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, San Diego, **1996**.

[39]   R. Cunico, A. G. Mayer, C. T. Wehr, T. L. Sheenan, *Biochromatography* **1986**, *1*.

[40]   F. Ausubel, R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, K. Struhl, *Current protocols in molecular biology, Vol. 3*, John Wiley & sons, New York, **1999**.

[41]   M. Bodanszky, A. Bodanszky, *The practise of peptide synthesis*, 2 ed., Springer Verlag, Berlin, **1994**.

Chapter 5

Structural characterization of block copolymers of β-sheet polypeptides and poly(ethylene glycol)

Abstract

The assembly behavior of silk-like triblock copolymers consisting of a central β-sheet polypeptide block and poly(ethylene glycol) (PEG) end blocks is described. Crystallization of the poly-[(AG)$_3$EG] polypeptide block (A = alanine, G = glycine and E = glutamic acid) was induced by vapour diffusion of methanol into a solution of the triblock copolymers in 70% formic acid. After crystallization the secondary structure of the polypeptide block was characterized with infra-red spectroscopy. The predominant absorption bands in the amide I and amide II region had frequencies of 1625 cm^{-1} and 1520 cm^{-1}, respectively. These frequencies are typical for an antiparallel β-sheet conformation of the polypeptide block. These were the predominant absorption bands for both the polypeptide alone as well as for the PEG-conjugates. For the polypeptide [(AG)$_3$EG]$_{10}$ conjugated with various PEG chains (M$_n$ = 750, 2000 and 5000 g mol^{-1}) the absorption band around 1655 cm^{-1} became stronger with increasing PEG chain length, indicating that the fraction of the polypeptide chain in the β-sheet conformation decreased.

The microstructure of the block copolymers was analyzed by transmission electron microscopy (TEM). It was shown that the block copolymers formed fibrils, whereas this fibrillar structure was not present for the polypeptides without conjugated PEG. To obtain information about the organization of the block copolymers within the fibrils, the length of the β-sheet and PEG block was varied. The fibrils for the block copolymer [(AG)$_3$EG]$_{20}$-PEG-750 had a width of approximately 12 nm. For the block copolymer [(AG)$_3$EG]$_{10}$-PEG-750 (where the polypeptide block has half of the length in comparison to [(AG)$_3$EG]$_{20}$-PEG-750), no fibrils were visible with TEM. Atomic force microscopy (AFM) was carried out on samples applied to a mica surface. From this it was clear that crystallization of the block copolymer [(AG)$_3$EG]$_{10}$-PEG-750 also resulted in fibril formation. Furthermore, AFM analysis showed that the fibrillar height (~ 2 nm) was independent of the polypeptide block length. Variation of the PEG block length in conjugates of [(AG)$_3$EG]$_{10}$ showed that the PEG-750 and PEG-2000 conjugates still readily formed fibrils, whereas for PEG-5000 only small fibrillar fragments were present. No height differences could be established, however.

With the present information an assembly model is proposed in which fibril formation occurs in the β-sheet stacking direction and the hydrogen-bond direction is perpendicular to the length-axis of the fibrils, whereas the PEG-chains prevent side-to-side aggregation of the fibrils. Diffraction data are, however, required to provide a direct evidence of this model. This class of polypeptide-polymer hybrids has the potential to allow control over height, width and surface functionality of the fibrils, and could therefore be useful as well-defined building blocks for the preparation of materials with organization at the nanometer scale.

5.1 Introduction

Peptides and proteins are actively being used as building blocks for the fabrication of higher order nanostructures via self-assembly. Knowledge of the relationship between primary amino acid sequence and folding behavior, as well as an increasing insight in protein-protein interactions make them attractive candidates for the creation of functional nanostructured materials. Naturally occurring proteins with an inherent property to self-assemble have been used to construct new materials. A variety of such systems have been investigated, such as bacterial S-layer proteins[1] and the yeast prion protein Sup35p[2]. S-layer proteins are crystalline two-dimensional protein arrays at the cell surface of certain bacteria. Because of the presence of defined pores they have been used e.g. for ultrafiltration and as template for the deposition of regularly arranged metal nanoparticles.[1] Yeast prion protein fibrils have been used for the preparation of conducting nanowires by selective metal deposition.[2]

Alternatively, new assemblies have been created by "polyvalent design"[3], using oligomeric proteins in combination with multivalent ligands. An example is the formation of a 'diamond-like' lattice resulting from the binding of the tetrameric protein concanavalin A and a two-headed carbohydrate molecule.[4] In another approach, a two-dimensional network with controllable mesh-size could be created by using a tetrameric aldolase conjugated with biotin in combination with streptavidin as a rigid linker.[5]

Besides building blocks of known assembly behavior, "de novo designed" peptide or protein building blocks have become increasingly important for the construction of new self-assembling systems. Especially small peptides forming β-strands and β-hairpins, which can be synthesized easily by solid-phase synthesis, have been investigated for their self-assembly properties. Simple design principles can be applied, such as amphiphilicity in β-strands by the alternation of polar and non-polar residues, charge arrangement in β-strands leading to formation of salt-bridges, and introduction of turns and interruption of hydrogen-bonding by the use of proline. These factors together determine the orientation of β-strands and the shape of the resulting aggregates and offer in addition the opportunity to introduce stimulus-responsiveness. A variety of architectures such as ribbons[6, 7], nanotubes[8, 9], monolayers with nanoscale order[10-12], gels[13, 14] and membranes[15] have been reported for β-sheet peptides. These systems may be used e.g. as scaffolds for tissue engineering or in targeted drug delivery.[16]

Higher molecular weight designed β-sheet polypeptides have been prepared via protein engineering by expression of a combinatorial library of synthetic genes coding for 7-residue long amphiphilic β-strands separated by turn sequences.[17] This entire class of well-defined polypeptides assembled into fibrils, formed monolayers at the air-water interface[18] or flat sheets on highly oriented pyrolytic graphite (HOPG)[19]. The gelation properties of the high molecular weight repetitive β-sheet fibril forming sequence poly-(AEAEAKAK) has been reported by Muller and coworkers.[20] Interestingly, they showed that the fibrils were of similar

diameter as the low-molecular weight analogues[15], but that the high molecular weight material was more elastic.

Despite the large amount of research that is performed on peptide and protein assembly, relatively few reports have described the self-assembly of hybrid block copolymers, i.e. copolymers which combine traditional synthetic polymers with peptide sequences. This kind of block copolymers have the potential to combine properties of the peptide block, such as assembly, recognition or bioactivity with the specific chemical and physical properties of synthetic polymers, such as reactivity, solubility and mechanical properties. Most reported hybrid block copolymers are composed of a synthetic polymer combined with either a homopolypeptide sequence prepared via ring-opening polymerization of an N-carboxyanhydride[21, 22] or a small peptide prepared via solid-phase synthesis. In particular, the latter approach is of interest for self-assembly. Meredith and coworkers reported the formation of fibrils of a diblock copolymer consisting of the amyloid-β(10-35) peptide and poly(ethylene glycol) (PEG).[23, 24] Side-to-side aggregation of fibrils was hindered by the PEG-coated surface of the fibrils. The assembly behavior of hybrid block copolymers consisting of PEG and amphiphilic β-strand peptide sequences was investigated by Klok and coworkers.[25, 26] These hybrids formed superstructures consisting of alternating PEG layers and peptide domains in an antiparallel β-sheet conformation in the solid-state and fibrillar structures after assembly in water. Silk-like multi-block copolymers of PEG and the β-sheet forming sequences AGAG (A = alanine, G = glycine) and poly(alanine) based on silkworm and spider silk, respectively, were reported by Sogah and coworkers.[27, 28] A microphase separated architecture of peptide domains (20 – 200 nm) dispersed in a continuous PEG phase was observed for these polymers.

In this chapter the secondary structure and assembly properties of block copolymers of poly-[(AG)$_3$EG] (E = glutamic acid) and PEG is investigated. The preparation of these block copolymers was described in Chapter 4. Tirrell and coworkers showed that crystallization of poly-[(AG)$_3$EG] from formic acid resulted in the formation of lamellar crystals (Figure 5.1).[29] The polar and bulky glutamic acid residues were confined to the lamellar surface.[30, 31] Although control over polypeptide folding was demonstrated with this sequence, crystallization resulted in extended plate-like structures. Therefore some of the information present in the polypeptide design, such as height and width of the β-sheet elements, was lost in the aggregation process.

With the attachment of N- and C-terminal PEG chains, we attempt to restrict macroscopic aggregation of the β-sheet polypeptides and allow translation of the β-sheet width and height in the assembled structures. This chapter describes the effect of attachment of PEG with varying chain length on the secondary structure of the polypeptide block, as determined by infra-red (IR) spectroscopy. The microstructure of the various triblock copolymers was investigated by transmission electron microscopy (TEM) and atomic force microscopy (AFM).
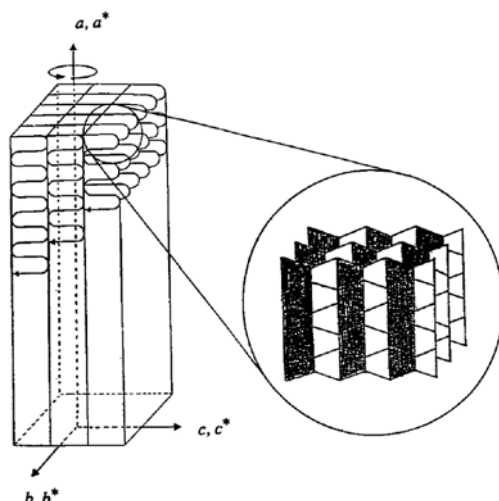
Figure 5.1 *Schematic illustration of the basic crystalline entity for poly-[(AG)₃EG]. The a-axis corresponds to the hydrogen-bond direction. The chain-folded lamellae stack in the b direction. From reference [32].*

5.2 Results and discussion

The conjugates of poly-[(AG)$_3$EG] and poly(ethylene glycol), of which the preparation has been described in Chapter 4, are summarized with their compound number in Figure 5.2.
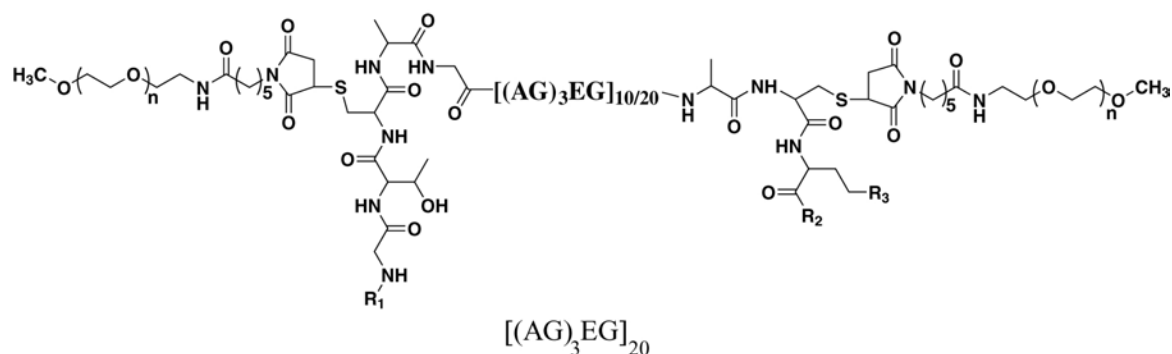
$[(AG)_3EG]_{20}$

DP of PEG block, n =	uncleaved	CNBr cleaved
17	4a	5a

$[(AG)_3EG]_{10}$

DP of PEG block, n =	uncleaved	CNBr cleaved
17	6a	7a
47	6b	7b
115	6c	7c
ε-maleimidocaproic acid	8	9

Figure 5.2 *Prepared conjugates of [(AG)$_3$EG]$_{10}$ and [(AG)$_3$EG]$_{20}$ with poly(ethyleneglycol) and ε-maleimidocaproic acid. Uncleaved conjugates still contain the N- and C-terminal non-repetitive amino acids (R$_1$ = H-ASMTGGQQMGRDPHHHHHHM, R$_2$ = GVPDPAANKARKEAELAAATAEQ-OH and R$_3$ = S-CH$_3$). For cyanogen bromide (CNBr) cleaved conjugates these amino acids have been removed (R$_1$ = H, R$_2$ and R$_3$ = OH).*

For [(AG)$_3$EG]$_{20}$ a conjugate was prepared with PEG-750. The effect of PEG conjugation on secondary structure was assessed by IR spectroscopy. Furthermore, the effect of the presence of the N- and C-terminal non-repetitive amino acids, on the assembly behaviour of the conjugate was investigated by TEM and AFM. Whereas the uncleaved conjugate (**4a**) still contains approximately 20 amino acids (including the 6 × His tag) at both the N terminus and the C terminus, these have been removed in case of the cyanogen bromide cleaved conjugate (**5a**).

For [(AG)$_3$EG]$_{10}$, the PEG-chain length was varied from PEG-750 to PEG-5000 and the effect on secondary structure and assembly was investigated for the cyanogen bromide cleaved variants (**7a**, **7b** and **7c**). In addition, as a control, the effect of conjugation of ε-maleimidocaproic acid (**8** and **9**) on assembly behaviour was investigated. The effect of the polypeptide block length on structure formation was assessed by comparison between [(AG)$_3$EG]$_{20}$-PEG-750 (**5a**) and [(AG)$_3$EG]$_{10}$-PEG-750 (**7a**).

5.2.1 Infra-red and circular dichroism spectroscopy

The crystallization behaviour of poly-[(AG)$_3$EG] is known to be dependent on crystallization conditions.[30] Crystallization from an aqueous lithium bromide solution by dialysis against water resulted in a crystal structure analogous to the structure of silk I (the fibroin of *Bombyx mori* in form I)[33]. The antiparallel secondary structure (as in silk II) resulted from crystallization from aqueous formic acid. In this chapter poly-[(AG)$_3$EG] and its various PEG-conjugates (Figure 5.2) were crystallized via the latter method. Crystallization was induced by vapour diffusion of methanol into a solution of the polypeptide in 70% formic acid. Formic acid serves as a chaotropic agent for the dissolution of the polypeptide. Furthermore, in this solvent the glutamic acid side chains will be protonated (pK$_a$ of formic acid and the glutamic acid side chain are 3.8 and 4.6[34], respectively), which prevents electrostatic repulsion between the turn amino acid residues and facilitates folding.

IR spectroscopy was carried out for the determination of the secondary structure of gelated samples of poly-[(AG)$_3$EG]. Furthermore, the effect of conjugation of PEGs of varying chain length on the secondary structure of poly-[(AG)$_3$EG] was investigated. The infra-red spectra were recorded after drying of the gelated samples. The frequency of the amide I (80% C=O stretch vibration) and amide II (60% N-H bend vibration) were used for indication of the secondary structure of the polypeptides.[35-37] The effect of methanol-induced gelation of [(AG)$_3$EG]$_{20}$ (without flanking cysteine residues, obtained from its GST fusion as described in Chapter 3, Paragraph 2.2) on the secondary structure of this polypeptide was clear after comparison of the IR-spectra for freeze-dried polypeptide and gelated polypeptide. The freeze-dried polypeptide [(AG)$_3$EG]$_{20}$ showed amide I and II bands at 1656 and 1547 cm^{-1}, respectively, which are indicative of a random coil/α-helical conformation (specific assignment to one of these conformations could not be made). For gelated [(AG)$_3$EG]$_{20}$ amide I and II bands at 1626 and 1519 cm^{-1} were observed, which are characteristic for the β-sheet conformation (characteristic frequencies for the β-sheet conformation are 1610 – 1640 cm^{-1} for

amide I and 1510 – 1550 cm^{-1} for amide II). The weak band at 1702 cm^{-1} and the amide II signal below 1630 cm^{-1} were furthermore indicative for the antiparallel orientation of the β-strands.

The effect of conjugation of PEGs with varying chain length on the secondary structure of poly-[(AG)$_3$EG] was investigated for the polypeptide [(AG)$_3$EG]$_{10}$. Figure 5.3a shows the IR-spectra for gelated conjugates of [(AG)$_3$EG]$_{10}$ with ε-maleimidocaproic acid, PEG-750, PEG-2000 and PEG-5000 after cyanogen bromide cleavage (**9**, **7a**, **7b** and **7c**, respectively). These IR-spectra are very similar, showing three bands around 1697, 1654 and 1623 cm^{-1} in the amide I region, as well as a band around 1522 cm^{-1} in the amide II region. For the conjugate of [(AG)$_3$EG]$_{20}$ with PEG-750 (**5a**) a similar spectrum was observed (Figure 5.3b).
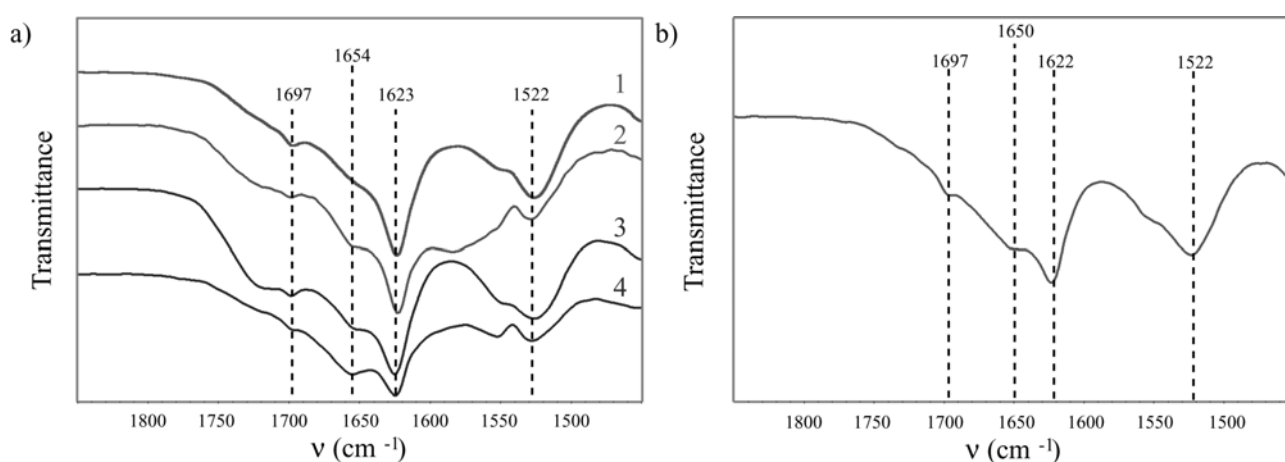

Figure 5.3 *Infrared spectra of cyanogen bromide cleaved conjugates of **(a)** [(AG)$_3$EG]$_{10}$ and (1) ε-maleimidocaproic acid, (2) PEG-750, (3) PEG-2000 and (4) PEG-5000. **(b)** [(AG)$_3$EG]$_{20}$ and PEG-750. Amide I and II regions are shown. Gelated samples (gelation from 10 mg mL^{-1} solution in 70% formic acid) were dried prior to measurement.*

The strong bands at 1623 cm^{-1} and 1522 cm^{-1} and the weak band at 1697 cm^{-1} are again indicative for the antiparallel β-sheet conformation. The amide I band around 1654 cm^{-1} (and also the amide II bands around 1550 cm^{-1}) indicates that some fraction of the polypeptide chain has adopted a secondary structure different from the antiparallel β-sheet conformation. This amide I component has been assigned previously to reverse turns of the β- or γ-type[29], but may also (partly) result from a random coil/α-helical conformation. It has also been assigned to the silk I structure.[31] This component becomes relatively stronger with longer PEG chains, although the main secondary structure is not significantly affected.

PEG-conjugates which were not treated with cyanogen bromide and therefore still contain the N- and C-terminal amino acids flanking the poly-[(AG)$_3$EG] sequence, were also subjected to gelation. The IR spectra for these uncleaved PEG-conjugates of [(AG)$_3$EG]$_{10}$ and [(AG)$_3$EG]$_{20}$ are depicted in Figure 5.4. The spectra are very similar to the cleaved variants (Figure 5.3), showing a large β-sheet content. For the PEG-750 and PEG-2000 conjugates the

band at 1655 cm^{-1} was relatively more intense. However, no contribution of the N- and C-terminal amino acids to specific bands in the final spectrum could be established.

Overall, we can conclude that the antiparallel β-sheet structure is retained upon attachment of PEG and that the spectra do not change significantly for the varying PEG chain lengths. The values obtained for amide I and II vibrations of poly-[(AG)$_3$EG] and its conjugates were identical to the values reported by Krejchi et al. for [AG)$_3$EG]$_{36}$.[29]
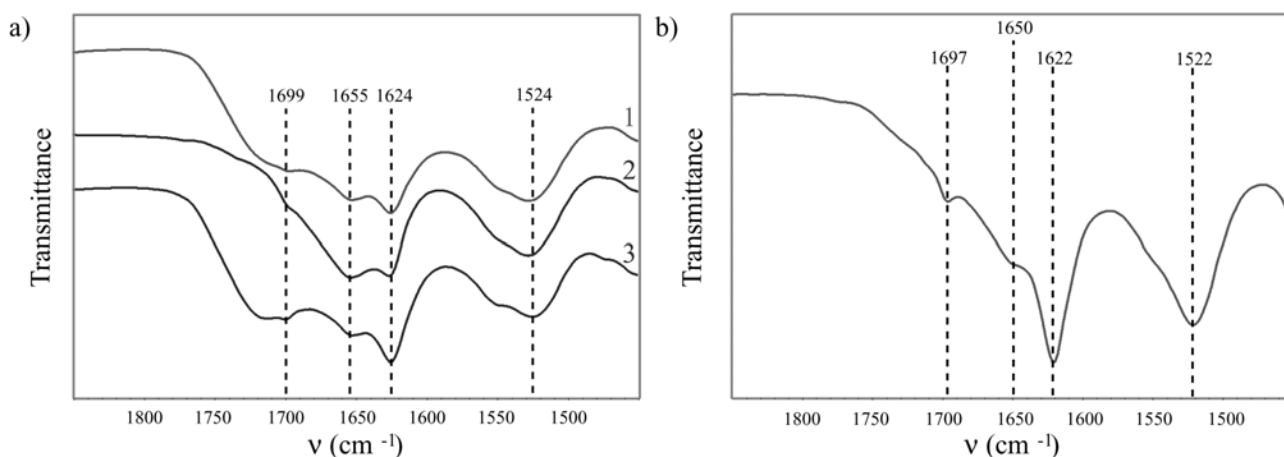

Figure 5.4 *Infrared spectra of uncleaved (i.e. containing N- and C-terminal amino acids flanking poly-[(AG)$_3$EG]) conjugates of (a) [(AG)$_3$EG]$_{10}$ and (1) PEG-750, (2) PEG-2000 and (3) PEG-5000. (b) [(AG)$_3$EG]$_{20}$ and PEG-750. Amide I and II regions are shown. Gelated samples (gelation from 10 mg mL^{-1} solution in 70% formic acid) were dried prior to measurement.*

In addition, circular dichroism (CD) spectroscopy was carried out for secondary structure determination. The gelated polypeptide [(AG)$_3$EG]$_{20}$ and its PEG-750 conjugate (both the uncleaved and the CNBr cleaved structures) were transferred to the surface of a quartz substrate and allowed to dry. Only uncleaved [(AG)$_3$EG]$_{20}$-PEG-750 conjugate gave a measurable CD-effect and confirmed the β-sheet conformation with a negative ellipticity at 215 nm and a positive ellipticity at 195 nm (Figure 5.5).[38] However, since the preparation of films of gelated polypeptide was irreproducible, with the result that absorption was sometimes too high to measure CD, secondary structure characterization of the other conjugates was limited to infrared spectroscopy.
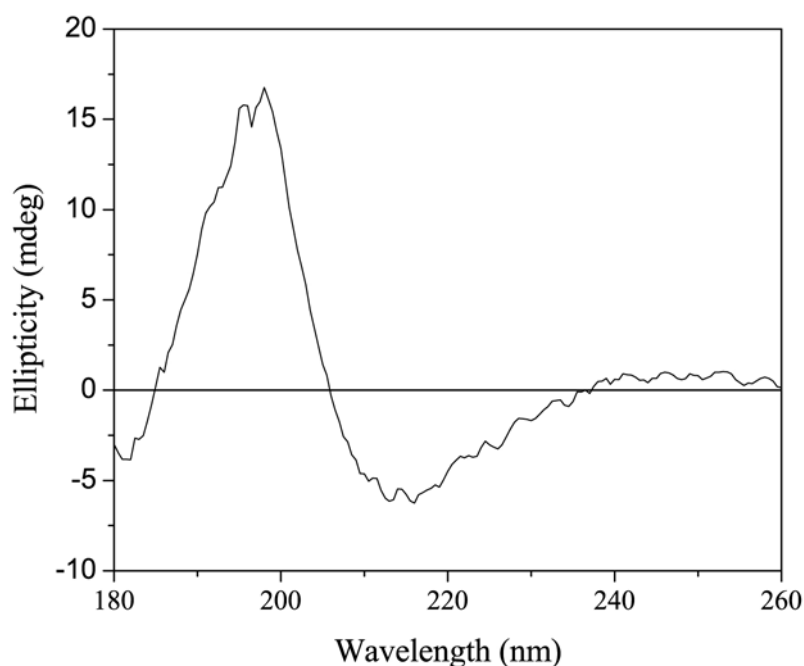
Figure 5.5 *Circular dichroism spectrum of uncleaved [(AG)₃EG]₂₀-PEG conjugate (T = 20 °C). 50 μL gelated sample (1 mg mL⁻¹ in methanol) was applied to a quartz surface and allowed to dry.*

5.2.2 Transmission electron microscopy analysis

To study the microstructure of gelated poly-[(AG)$_3$EG] and its conjugates, transmission electron microscopy (TEM) analysis was performed. Samples were deposited on carbon-coated grids by incubating the grid with the carbon side down on top of a gel. After drying, visualization of the microstructure was achieved by platinum shadowing. Initial TEM measurements were performed with a gelated sample of [(AG)$_3$EG]$_{20}$-PEG-750 conjugate, which still contained the N- and C-terminal non-repetitive amino acids (**4a**). Figure 5.6 illustrates that this sample exhibited a clear fibrillar microstructure. The fibrils were micrometer long, had a uniform width of approximately 12 nm (measured on fibrils lying side-by-side) and were aligned. For visualization of single fibrils, the gelated sample was diluted with methanol and applied to the TEM grid. It was clear, however, that most the fibrils aggregate into larger bundles (Figure 5.6, inset). Comparison of this conjugate with the cyanogen bromide cleaved variant (**5a**), showed that the latter did not have this alignment but instead consisted of a network of fibrils (compare Figure 5.7a and b). For [(AG)$_3$EG]$_{20}$ without conjugated PEG only few (for the uncleaved variant) or no fibrils (for the CNBr cleaved variant) were found on the grid surface (Figure 5.7c and d).
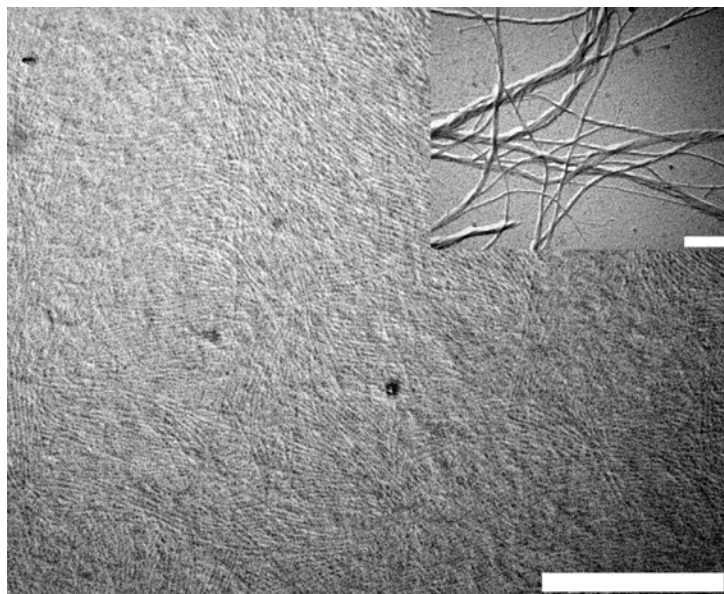
Figure 5.6 TEM micrograph showing the fibrillar microstructure of uncleaved $[(AG)_3EG]_{20}$-PEG-750 conjugate *(4a)*. The scale bar represents 500 nm. Inset: TEM micrograph of diluted sample (~ 0.1 mg mL^{-1}) showing fibril bundles of various thicknesses. The scale bar represents 200 nm.
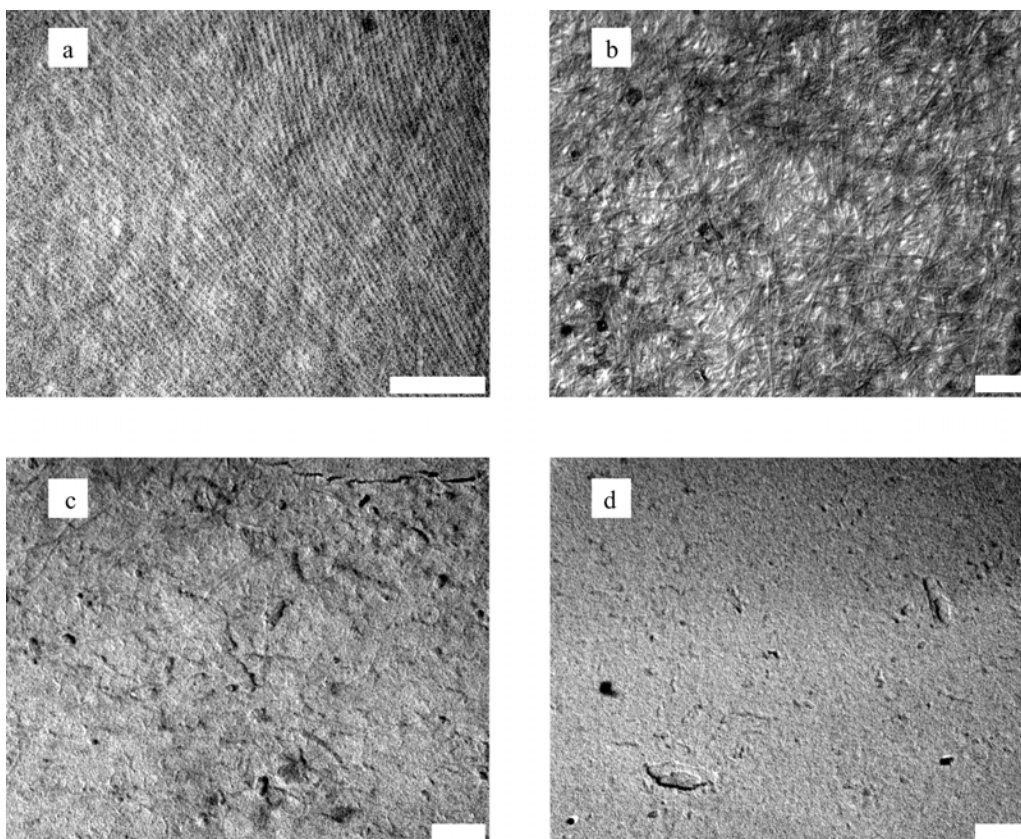

Figure 5.7 TEM micrographs of crystallized $[(AG)_3EG]_{20}$ polypeptides with and without conjugated poly(ethylene glycol)-750. Platinum shadowed samples (1 mg mL^{-1}) of *(a)* uncleaved $[(AG)_3EG]_{20}$-PEG-750 conjugate *(4a)*; *(b)* CNBr cleaved $[(AG)_3EG]_{20}$-PEG-750 conjugate *(5a)*; *(c)* uncleaved $[(AG)_3EG]_{20}$; *(d)* CNBr cleaved $[(AG)_3EG]_{20}$. The scale bars represent 200 nm.

For $[(AG)_3EG]_{10}$, TEM measurements were performed on conjugates of this polypeptide with PEGs of different molecular weight. For the $[(AG)_3EG]_{10}$-PEG-750 conjugates a clear fibrillar microstructure was observed only for the uncleaved variant (**6a**) (Figure 5.8a). The fibrillar width was clearly smaller than for the $[(AG)_3EG]_{20}$-PEG-750 conjugate, since it could not be measured. The platinum particles used for visualization seemed to be too large relative to the fibril width. For the CNBr-cleaved $[(AG)_3EG]_{10}$-PEG-750 conjugate (**7a**) single fibrils were difficult to distinguish, because of dense coverage of the grid (Figure 5.8b). Attempts to improve visualization of these fibrils by diluting this sample and by applying negative staining (uranyl acetate and phosphotungstic acid) were not successful.
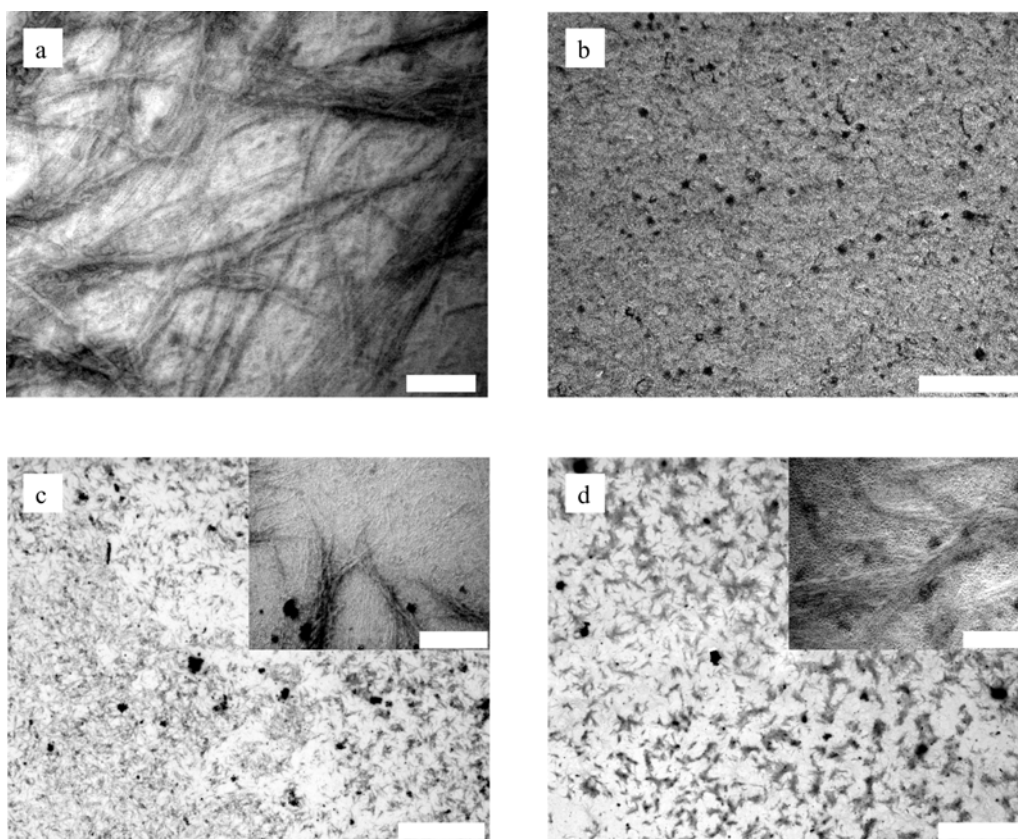

Figure 5.8 *TEM micrographs of crystallized $[(AG)_3EG]_{10}$ polypeptides with conjugated poly(ethylene glycols) of 750, 2000 and 5000 g mol^{-1}* ***(a)*** *uncleaved $[(AG)_3EG]_{10}$ conjugated to PEG-750 (**6a**); the scale bar represents 200 nm.* ***(b)*** *CNBr cleaved $[(AG)_3EG]_{10}$ conjugated to PEG-750 (**7a**); the scale bar represents 200 nm.* ***(c)*** *uncleaved $[(AG)_3EG]_{10}$ conjugated to PEG-2000 (**6b**); the scale bars represent 10 μm and 400 nm for the inset.* ***(d)*** *uncleaved $[(AG)_3EG]_{10}$ conjugated to PEG-5000 (**6c**); the scale bars represent 10 μm and 200 nm for the inset.*

For conjugates of $[(AG)_3EG]_{10}$ with PEG-2000 (**6b** and **7b**) and PEG-5000 (**6c** and **7c**) we were not able to visualize any structure for samples crystallized from a 10 mg mL^{-1} solution. Only for the uncleaved samples crystallized from a more dilute, 1 mg mL^{-1} solution, irregular, μm-sized aggregates were visible (Figure 5.8c and d). This morphology is alike the spherulitic texture of

a recombinant silk-like protein polymer reported by Anderson and coworkers[39] and is also commonly seen in thin films of semi-crystalline polymers.[40] A zoomed-in image of these aggregates (inset) showed the presence of a fibrillar morphology. For $[(AG)_3EG]_{10}$ conjugated to ε-maleimidocaproic acid (**8** and **9**) no fibrillar structure or other morphology could be observed. It seemed that no material could be transferred to the grid. This is likely to reflect the larger extent of physical cross-linking for this conjugate as a result of the increased interaction between the β-sheet polypeptides.

The presence of PEG end-blocks is apparently a prerequisite for the formation of a fibrillar microstructure. Without conjugated PEG, crystallization most likely results in the formation of larger aggregated structures, because of the less directed interaction between the β-sheet polypeptides. The PEG chains contribute a directional force to the β-sheet assembly, analogous to the orienting forces in block copolymer phase separation. The observed fibril alignment, especially visible for uncleaved $[(AG)_3EG]_{20}$-PEG-750 (**4a**) could be the combined result of the presence of the PEG chains and the N- and C-terminal amino acids. Since these flanking peptide elements have a net positive charge under the conditions used for crystallization their presence might lead to an additional repulsive force between the unorganized domains of the triblock copolymers. This could further limit aggregation of fibrils and might facilitate alignment by shear. After attachment of longer PEG-chains (PEG-2000 and PEG-5000) to $[(AG)_3EG]_{10}$, no fibrillar microstructure could be detected for the CNBr cleaved variants. Since it was clear that the absence of a microstructure could also be the result of the fact that the limits of resolution of TEM were reached, atomic force microscopy analysis was carried out to obtain more information about the assembly of the various conjugates.

5.2.3 Atomic force microscopy analysis

The fibrillar microstructure for the various PEG-conjugates of poly-$[(AG)_3EG]$ was analyzed more extensively using atomic force microscopy (AFM). Initial measurements were performed on conjugates of $[(AG)_3EG]_{20}$ with PEG-750 still containing the N- and C-terminal non-repetitive amino acids (**4a**). The gelated conjugate (10 mg mL^{-1}) was diluted with methanol (~ 0.01 - 0.1 mg mL^{-1}) and applied on a freshly cleaved mica surface by spin-coating. Figure 5.9 depicts representative images obtained for this sample. The fibrillar microstructure of this conjugate, as observed by TEM (Figure 5.6) was confirmed by AFM. A similar alignment and side-by-side aggregation of fibrils can be seen in Figure 5.9a. Upon further dilution a mixture of single and aggregated fibrils could be seen with varying height from approximately 1.5 to 10 nm (Figure 5.9b). A high resolution phase and height image of a branching point of fibrils is depicted in Figure 5.9c and d. The side-by-side aggregation for this block copolymer was especially clear from the phase images.
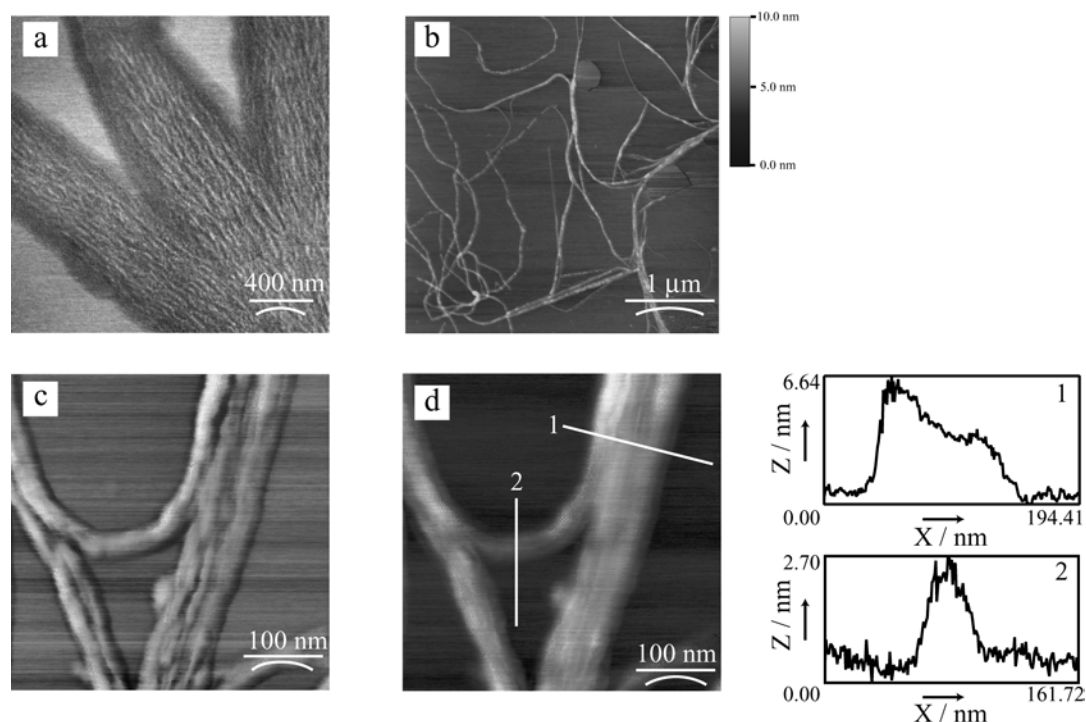
Figure 5.9 *Representative atomic force microscopy images (tapping mode) of β-sheet polypeptide [(AG)₃EG]₂₀ conjugated with poly(ethylene glycol)-750 (uncleaved; **4a**) showing a fibrillar microstructure. Samples were diluted with methanol from a 10 mg mL⁻¹ gelated sample and spincoated onto a freshly cleaved mica surface. **(a)** Phase image of aggregated fibrils (100-fold diluted sample). **(b)** Height image showing single and aggregated fibrils (1000-fold diluted sample) **(c)** High resolution phase image of a branching point of fibrils and **(d)** its corresponding height image and height profiles.*

For the $[(AG)_3EG]_{10}$-PEG-750 conjugate, AFM images were recorded for the cyanogen bromide cleaved variant (**7a**), which proved the presence of a fibrillar microstructure (Figure 5.10a). These fibrils could not be clearly visualized previously using TEM (Figure 5.8b). Individual fibrils were also visualized for CNBr cleaved $[(AG)_3EG]_{20}$-PEG-750 (**5a**) (Figure 5.10b), but relatively more large aggregates were observed for this sample in comparison to the uncleaved variant.

To obtain information about the molecular organization of the β-sheet conjugates within the fibrils, a series of height and width measurements were made for single fibrils of PEG-750 conjugates of $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ (**7a** and **5a**). Within a measurement session, the same height values were found for both conjugates, albeit with a broad distribution. Values found for $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ conjugates were 1.9 ± 0.7 nm and 1.6 ± 0.4 nm, respectively. Furthermore, in later measurement sessions lower values were found (typically 0.8 nm). It can therefore be concluded that no height difference was observed, but that quantification of the height proved to be difficult. Possible factors affecting the measured height values that have been mentioned in literature are e.g. deformation caused by the tip's pressure and the influence of air humidity.[41]

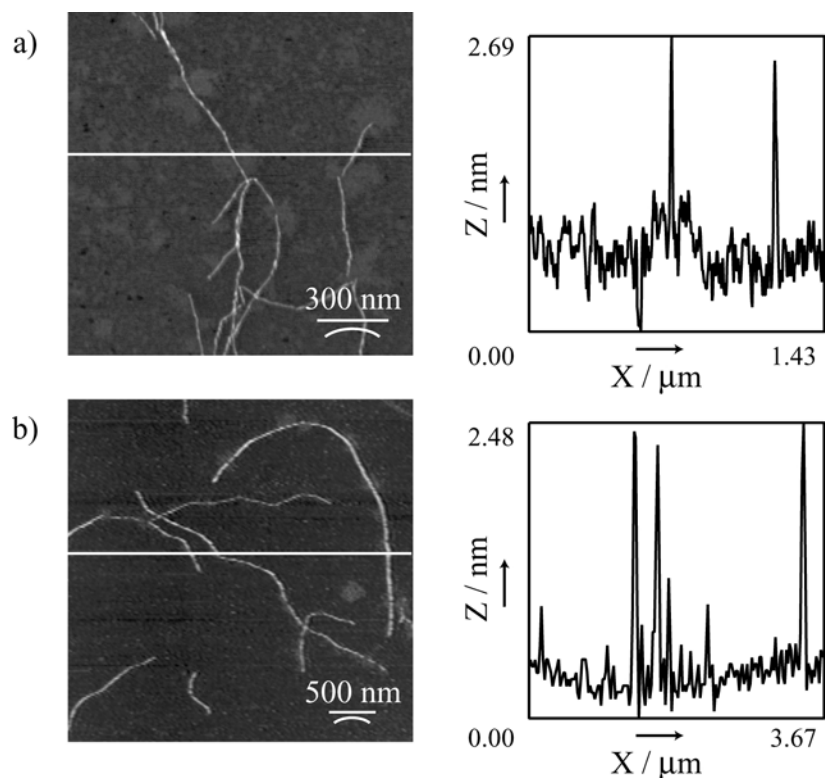**Figure 5.10** *AFM height images of cyanogen bromide cleaved PEG-750-conjugates of (a) [(AG)₃EG]₁₀ (7a) and (b) [(AG)₃EG]₂₀ (5a) on mica supports. Height profiles are depicted next to the images.*

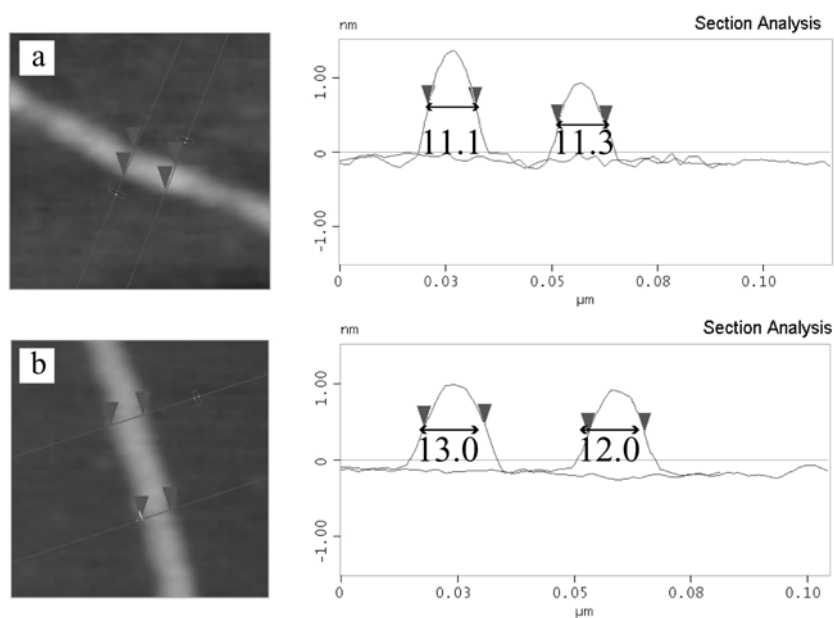**Figure 5.11** *Width-at-half-height measurements of (a) [(AG)₃EG]₁₀ and (b) [(AG)₃EG]₂₀ fibrils measured with an ultrasharp AFM-tip.*

Although it is well established that the measurement of lateral dimensions of objects by AFM is complicated by the finite dimensions of the AFM tip (see Paragraph 5.4.4), width measurements were made using a high-resolution AFM tip (carbon, typical radius of 1 nm). These measurements showed a similar width for both conjugates. Width-at-half-height values of 11.7 ± 1.7 nm and 12.6 ± 1.8 nm were found for cleaved PEG-750 conjugates of $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ fibrils, respectively (Figure 5.11).

The effect of the surface on structure formation was assessed by using instead of mica more hydrophobic graphite surfaces (highly oriented pyrolytic graphite, HOPG). For the uncleaved $[(AG)_3EG]_{20}$-PEG-750 conjugate (**4a**) (Figure 5.12a), the interaction between the fibrils and the surface was weak, since the fibrils moved over the surface during the AFM measurements. For the cleaved $[(AG)_3EG]_{10}$-PEG-750 conjugate (Figure 5.12b and c) the fibrils had a twisted appearance, especially visible in the phase image. However, no major effect of the surface on the observed structures was observed.
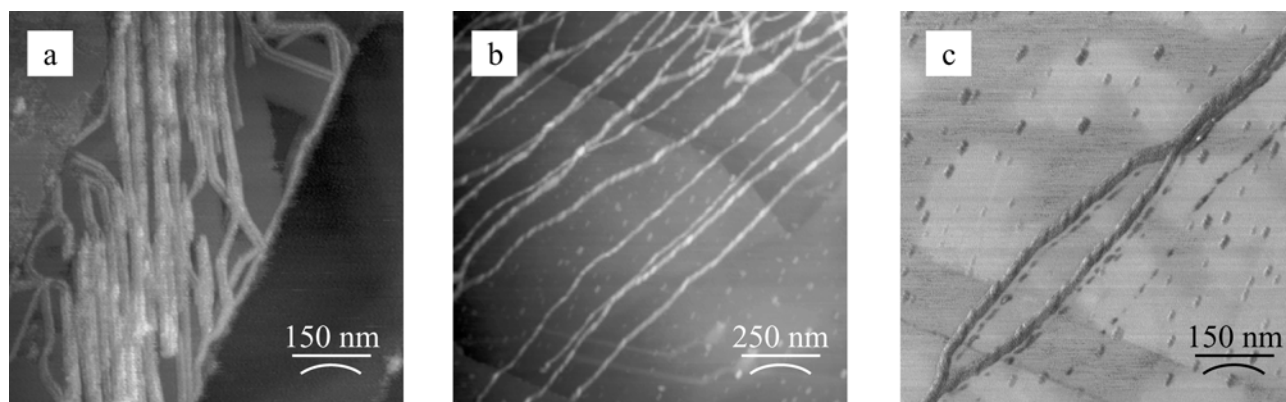

Figure 5.12 *AFM measurement of fibrils on graphite showing height image of* ***(a)*** *uncleaved $[(AG)_3EG]_{20}$-PEG-750 conjugate and* ***(b)*** *cleaved $[(AG)_3EG]_{10}$-PEG-750 conjugate.* ***(c)*** *Phase image of cleaved $[(AG)_3EG]_{10}$-PEG-750 conjugate showing branching and an apparent twisted shape of the fibrils.*

AFM analysis of cyanogen bromide cleaved $[(AG)_3EG]_{10}$-PEG-2000 (**7b**) and $[(AG)_3EG]_{10}$-PEG-5000 (**7c**) showed fibrillar microstructures for both conjugates (Figure 5.13a and b), in contrast to TEM measurements. Whereas for $[(AG)_3EG]_{10}$-PEG-2000 (**7b**), micrometer-long fibrils were observed, the fibrillar structure seemed to be less defined for $[(AG)_3EG]_{10}$-PEG-5000 (**7c**). Furthermore, no fibrils were observed for $[(AG)_3EG]_{10}$-ε-maleimidocaproic acid (**9**) (Figure 5.13c), again confirming that the formation of fibrils is a result of the presence of PEG.

For the determination of fibrillar heights of the PEG-conjugates more diluted samples were applied to the mica surface. Long fibrils were still present for $[(AG)_3EG]_{10}$-PEG-2000 (**7b**), but only short fibril fragments were observed for $[(AG)_3EG]_{10}$-PEG-5000 (**7c**). Presumably, conjugation of longer PEG-chains hinders the formation of long fibrils. A relatively broad distribution in the measured heights was obtained with values of 2.2 ± 0.3 and 2.4 ± 0.9 nm for the PEG-2000 and PEG-5000 conjugates, respectively. This is in the same

range as measured for the PEG-750 conjugates. The width-at-half height values for both conjugates were 26.5 ± 1.2 nm, higher than for the PEG-750 conjugates. However, direct comparison of these results is not possible, since these measurements were not performed with the same tip.

The difference in TEM and AFM observations for the PEG-2000 and PEG-5000 conjugates can not be explained easily. As was already observed for the PEG-750 conjugate, clear visualization of the fibrillar structure was only obtained for the non-cleaved sample, where fibrils were aligned. Since the samples were applied to the surface for TEM and AFM analysis (carbon and mica, respectively) after crystallization (and presumably fibril formation) in a similar procedure, no major differences would be expected. It seems more likely that the resolution of TEM analysis was the limiting factor as a result of the grain size of the platinum particles used for shadowing.
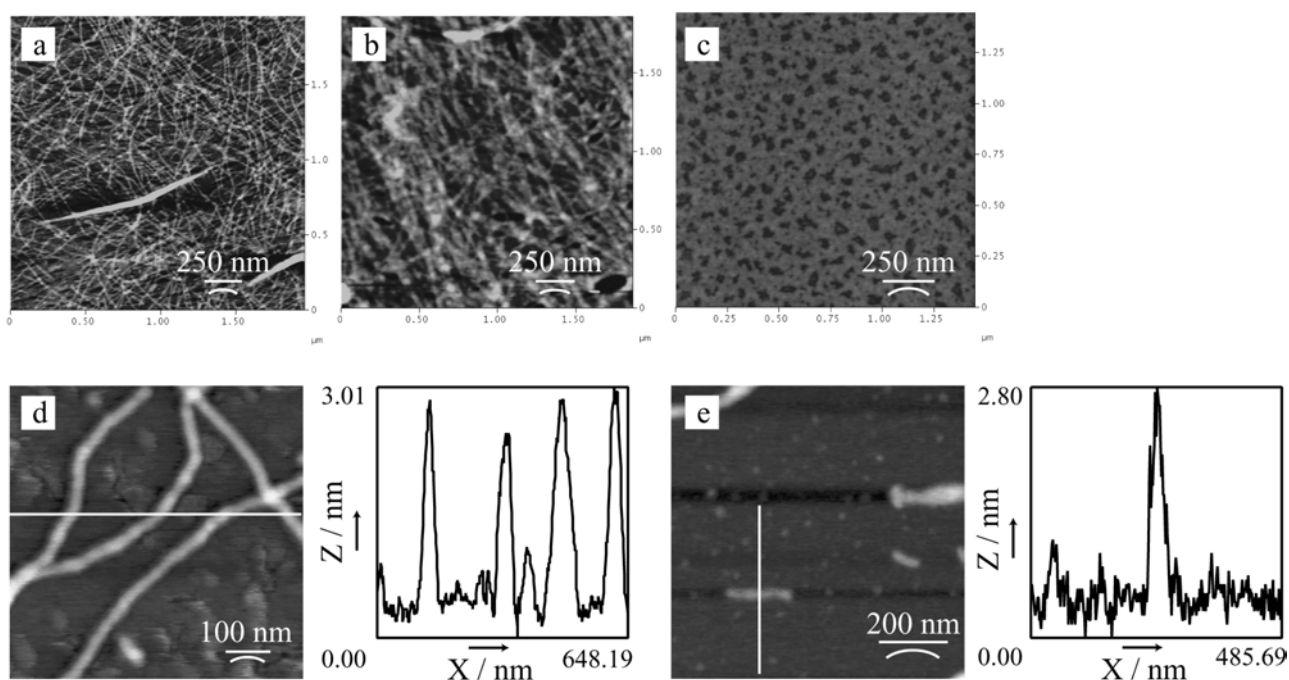

Figure 5.13 *AFM analysis of microstructures of [(AG)₃EG]₁₀ conjugated with **(a)** PEG-2000, **(b)** PEG-5000 and **(c)** ε-maleimidocaproic acid. Height profiles are shown for **(d)** PEG-2000 and **(e)** PEG-5000 conjugates. All samples were cleaved with cyanogen bromide. Crystallized samples were diluted with methanol and dropcasted on mica (concentrations in the range 0.01 – 0.1 mg mL⁻¹).*

5.2.4 Proposed model for fibril formation

Based on the combined results from AFM and TEM measurements it can be stated that the formation of a fibrillar microstructure for these conjugates was shown to depend on the presence of PEG end-blocks. With the chosen crystallization method, the PEG blocks are expected to stay dissolved, since methanol is a good solvent for this polymer.[42] A possible arrangement of the triblock copolymers within the fibrils is depicted for PEG-conjugates of

[(AG)₃EG]₁₀ in Figure 5.14. Fibril formation occurs in the β-sheet stacking direction resulting in an assembly in which the hydrophobic β-sheet surfaces are shielded from methanol, whereas the PEG-chains are in contact with the solvent. Based on this model the expected values for the width of the $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ polypeptide blocks are 4.3 and 8.9 nm, respectively.[29] The radius of gyration of the PEG-chains can be estimated to be 0.8, 1.5 and 2.6 nm for PEG-750, PEG-2000 and PEG-5000, respectively.[42] The total fibril width for the $[(AG)_3EG]_{10}$-PEG-750 and $[(AG)_3EG]_{20}$-PEG-750 conjugates can then be estimated to be ∼ 7.5 and 12.1 nm, respectively, when it is assumed that the PEG-chains can adopt a random coil structure. The fibrillar width or more accurately expressed, a repeating distance of ∼ 12 nm could be measured for the $[(AG)_3EG]_{20}$-PEG-750 conjugate with TEM. These measurements together with the TEM data showing the smaller fibrillar width of $[(AG)_3EG]_{10}$-PEG-750 are supportive for the suggested assembly model.
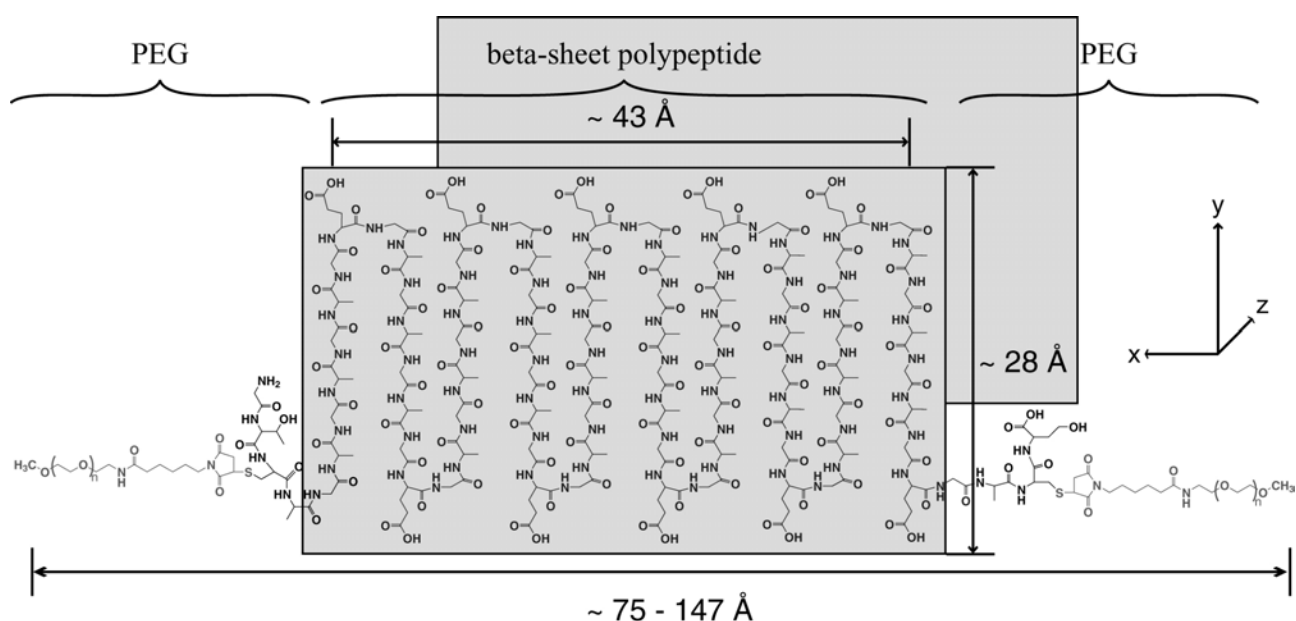

Figure 5.14 *Proposed model for the organization of triblock copolymers of a central $[(AG)_3EG]_{10}$ peptide block and PEG end blocks within fibrils. The model comprises folded β-sheets stacked along the z-axis, leading to fibril formation (perpendicular to the plane of the page). Antiparallel β-strands are formed (x = hydrogen bonding direction, y = chain axis direction) with glutamic acid residues at the fibril surface. PEG chains hinder further lateral aggregation by preventing further hydrogen-bonding along the x-axis. Depending on the molecular weight of the PEG-chain used (with radius of gyration from ∼ 8Å for PEG-750 to 26 Å for PEG-5000), widths of ∼ 75 to 147 Å are expected.*

AFM analysis showed similar height values of ∼ 2 nm for all conjugates of $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$, which are in reasonable agreement with the calculated height of 2.8 nm[29], particularly when one takes into account that AFM measurements do not give absolute height values as a result of possible deformation caused by the tip's pressure and the influence of air humidity.[41] For an alternative assembly arrangement, in which cylindrical instead of

rectangular fibrils are formed, as reported for PEG-conjugates of small β-sheet peptides by a number of research groups[24, 25], changing the PEG-chain length is expected to be reflected in the height values as measured by AFM. In our case we seem to be able to exclude such an arrangement, since no difference in height could be established for the varying PEG-conjugates.

However, the length of the PEG-chains seems to have some effect on fibril formation. With a calculated width of the $[(AG)_3EG]_{10}$ polypeptide block of ~ 4.3 nm, it is clear that especially with PEG-5000 (radius of gyration of ~ 2.6 nm), the size of the PEG-chain is larger than the polypeptide structure. Although the model in Figure 5.14 suggests that the PEG-chains are only responsible for lining up the β-sheet polypeptides and prevent further aggregation in the hydrogen-bonding direction, it seems likely that blocking of aggregation also occurs in the β-sheet stacking direction. Fibril formation by β-sheet stacking will become more difficult with increasing PEG-content, resulting in shorter or less stable fibrils.

Although for poly-$[(AG)_3EG]$ the confinement of glutamic acid residues at the lamellar surface was supported particularly by solid-state NMR spectroscopy data[43], for the PEG-conjugates the decoration of the fibrillar surface with glutamic acid residues still has to be proven. The most straightforward way to test the model for fibril formation (Figure 5.14) is the preparation of polypeptides with an increasing number of consecutive alanylglycine repeats. The effect of this on fibril height can then be estimated by AFM. For example, fibril formed by PEG-conjugates of poly-$[(AG)_5EG]$ should have a calculated height of approximately 4.2 nm.

In addition, the fibrillar width measured by TEM should increase with increasing number of -$(AG)_3EG$- repeats. Diffraction experiments can furthermore give more information on the regular arrangement of the β-sheet polypeptide in the fibrils as well as possible crystallinity of the PEG-blocks after drying. X-ray diffraction measurements were carried out for $[(AG)_3EG]_{20}$ and its PEG-conjugate. However, no diffraction pattern was obtained, which could be the result of the use of insufficient material and/or lack of alignment within the sample. Electron diffraction on a selected area containing aligned fibrils will be the most suitable method for an unambiguous assignment of the direction of β-sheet hydrogen bonding relative to the fibril axis.

5.3 Conclusions

Methanol-induced crystallization of conjugates of poly-$[(AG)_3EG]$ and PEG resulted in folding of the polypeptide chains in an antiparallel β-sheet conformation, as was observed by infra-red spectroscopy. Conjugation of PEG did not have a significant effect on the polypeptide backbone conformation, since the values for the amide I and II vibrations were similar as for the polypeptide alone.[29] Furthermore, it was shown that the attachment of PEG end blocks to a central poly-$[(AG)_3EG]$ β-sheet block prevents the macroscopic aggregation of the β-sheet blocks into needle-shaped lamellar crystals. Instead, crystallization resulted in well-defined fibrils. Microscopy data suggest an organization in which fibrils are formed in the β-sheet

stacking direction. The absolute control over the amino acid sequence might offer the possibility of introducing specific amino acid residues at the turns of the β-sheets, creating a regular array of functional moieties at the fibril surface. Patterning of the surface with different amino acid combinations might provide a method for positioning different chemical functionalities at the nanometer scale.

5.4 Experimental section

5.4.1 Crystallization experiments

Crystallization was induced by vapour diffusion of methanol (p.a. Merck) into a solution of 1 or 10 mg mL^{-1} of polypeptide or conjugate in 70% formic acid (Merck). This was achieved by placing an eppendorf tube with polypeptide or conjugate solution (100 μL) in a Duran bottle filled with 20 mL methanol. The bottle was closed and incubation was carried out for 2 days.

5.4.2 Fourier transform infrared and circular dichroism spectroscopy

Transmission-FTIR spectra were measured with a Thermo Mattson Genesis IR-300 spectrometer using a deuterated triglycine sulphate detector (resolution, 4 cm^{-1}; number of scans, 64). Samples were allowed to dry on the surface of the ATR-crystal prior to measurement. Circular dichroism spectroscopy was performed on a Jasco J-810 spectropolarimeter (band width: 1nm, response: 1 sec., sensitivity: standard, λ range: 260-180 nm, data pitch: 0.5 nm, scanning speed: 100 nm min.$^{-1}$, accumulation: 5).

5.4.3 Transmission electron microscopy studies

All samples were prepared on copper specimen grids covered with a carbon support film (CF200-Cu; Electron Microscopy Sciences, Washington, USA). The grids were placed with the carbon film down on top of a 5 μl droplet of protein suspension (1 mg mL^{-1}) in methanol or on a gel (10 mg mL^{-1}) for 30 seconds. The grids were allowed to dry and were subsequently shadowed with platinum at an elevation angle of 45 degrees. Images were recorded using a JEOL JEM 1010 electron microscope (60 kV) equipped with a CCD camera.

5.4.4 Atomic force microscopy studies

Polypeptide or conjugate suspensions diluted with methanol to concentrations of 10 - 100 μg mL^{-1} were spincoated onto freshly cleaved mica surface. For Figure 5.12 and 5.13 dropcasting was used for applying the protein suspensions to graphite and mica, respectively. The morphologies of the samples were analyzed by means of tapping mode AFM using a Nanoscope IIIa instrument operating in air at room temperature. Height and phase images were recorded with microfabricated silicon cantilevers (length 100 mm and width 35 mm) having a

spring constant between 5.5 and 22.5 N m^{-1}, using scan rates of 1 – 2 lines s^{-1} and a resolution of 512 × 512 pixels.

It is well established that the width measured by AFM is a convolution of the actual size of the objects and that of the AFM tip. Obviously, the distortion is especially large for objects with a similar or smaller size as the AFM tip (with a typical radius of 10 nm). For rectangular objects (Figure 5.15a) the true width W* can be calculated from the observed width W with the equation W* = W – 2X, where X = {H(2R - H)}$^{1/2}$.[44] Alternatively, the true width W* can be calculated from the measured width at half height W$_{1/2}$ with the equation W* = W$_{1/2}$ – 2X, where X = (RH – ¼H^2)$^{1/2}$. For spherical objects (Figure 5.15b) the observed width W is equal to 4(Rr)$^{1/2}$. R is the radius of curvature of the AFM tip and r is the radius of the measured object.[45] Therefore, the true radius of the sphere (r) can be calculated from the observed width W with the equation r = W^2/(16R). Alternatively, the true radius can be calculated from the measured width at half height W$_{1/2}$ with the equation r = (R^2 + ¼W$_{1/2}$^{2})$^{1/2}$ – R.[41]
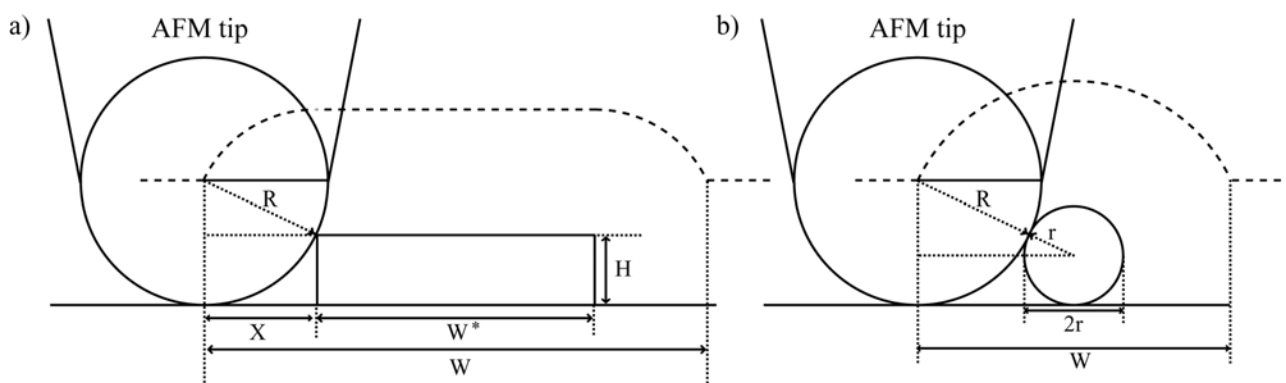

Figure 5.15 *Widths measured by atomic force microscopy are a convolution of AFM tip size and the shape of the object. A schematic representation for **(a)** spherical and **(b)** rectangular objects is depicted.*

5.5 References

[1]   U. B. Sleytr, B. Schuster, D. Pum, *IEEE Eng. Med. Biol. Mag.* **2003**, *22*, 140.
[2]   T. Scheibel, R. Parthasarathy, G. Sawicki, X. M. Lin, H. Jaeger, S. L. Lindquist, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4527.
[3]   T. O. Yeates, J. E. Padilla, *Curr. Opin. Struct. Biol.* **2002**, *12*, 464.
[4]   N. Dotan, D. Arad, F. Frolow, A. Freeman, *Angew. Chem. Int. Edit.* **1999**, *38*, 2363.
[5]   P. Ringler, G. E. Schulz, *Science* **2003**, *302*, 106.
[6]   A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong, N. Boden, *J. Am. Chem. Soc.* **2003**, *125*, 9619.
[7]   D. M. Marini, W. Hwang, D. A. Lauffenburger, S. G. Zhang, R. D. Kamm, *Nano Lett.* **2002**, *2*, 295.
[8]   K. Lu, J. Jacob, P. Thiyagarajan, V. P. Conticello, D. G. Lynn, *J. Am. Chem. Soc.* **2003**, *125*, 6391.
[9]   S. Santoso, W. Hwang, H. Hartman, S. G. Zhang, *Nano Lett.* **2002**, *2*, 687.
[10]   S. Colfer, J. W. Kelly, E. T. Powers, *Langmuir* **2003**, *19*, 1312.
[11]   H. Rapaport, G. Moller, C. M. Knobler, T. R. Jensen, K. Kjaer, L. Leiserowitz, D. A. Tirrell, *J. Am. Chem. Soc.* **2002**, *124*, 9342.
[12]   E. T. Powers, S. I. Yang, C. M. Lieber, J. W. Kelly, *Angew. Chem. Int. Edit.* **2002**, *41*, 127.
[13]   A. Aggeli, M. Bell, N. Boden, L. M. Carrick, A. E. Strong, *Angew. Chem. Int. Edit.* **2003**, *42*, 5603.
[14]   J. P. Schneider, D. J. Pochan, B. Ozbas, K. Rajagopal, L. Pakstis, J. Kretsinger, *J. Am. Chem. Soc.* **2002**, *124*, 15030.
[15]   S. Zhang, T. C. Holmes, C. Lockshin, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3334.
[16]   S. G. Zhang, X. J. Zhao, *J. Mater. Chem.* **2004**, *14*, 2082.
[17]   M. W. West, W. X. Wang, J. Patterson, J. D. Mancias, J. R. Beasley, M. H. Hecht, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11211.
[18]   G. Xu, W. Wang, J. T. Groves, M. H. Hecht, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3652.
[19]   C. L. Brown, I. A. Aksay, D. A. Saville, M. H. Hecht, *J. Am. Chem. Soc.* **2002**, *124*, 6846.
[20]   N. L. Goeden-Wood, J. D. Keasling, S. J. Muller, *Macromolecules* **2003**, *36*, 2932.
[21]   H. Schlaad, M. Antonietti, *Eur. Phys. J. E* **2003**, *10*, 17.
[22]   B. Gallot, *Prog. Polym. Sci.* **1996**, *21*, 1035.
[23]   T. S. Burkoth, T. L. S. Benzinger, D. N. M. Jones, K. Hallenga, S. C. Meredith, D. G. Lynn, *J. Am. Chem. Soc.* **1998**, *120*, 7655.
[24]   T. S. Burkoth, T. L. S. Benzinger, V. Urban, D. G. Lynn, S. C. Meredith, P. Thiyagarajan, *J. Am. Chem. Soc.* **1999**, *121*, 7429.
[25]   I. W. Hamley, A. Ansari, V. Castelletto, H. Nuhn, A. Rosler, H. A. Klok, *Biomacromolecules* **2005**, *6*, 1310.
[26]   A. Rosler, H. A. Klok, I. W. Hamley, V. Castelletto, O. O. Mykhaylyk, *Biomacromolecules* **2003**, *4*, 859.
[27]   O. Rathore, D. Y. Sogah, *Macromolecules* **2001**, *34*, 1477.
[28]   O. Rathore, D. Y. Sogah, *J. Am. Chem. Soc.* **2001**, *123*, 5231.
[29]   M. T. Krejchi, E. D. T. Atkins, A. J. Waddon, M. J. Fournier, T. J. Mason, D. A. Tirrell, *Science* **1994**, *265*, 1427.
[30]   M. T. Krejchi, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *J. Macromol. Sci.-Pure Appl. Chem.* **1996**, *A33*, 1389.
[31]   C. C. Chen, M. T. Krejchi, D. A. Tirrell, S. L. Hsu, *Macromolecules* **1995**, *28*, 1464.
[32]   M. T. Krejchi, S. J. Cooper, Y. Deguchi, E. D. T. Atkins, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Macromolecules* **1997**, *30*, 5012.

[33] B. Lotz, H. D. Keith, *J. Mol. Biol.* **1971**, *61*, 201.
[34] C. R. Cantor, P. R. Schimmel, in *Biophysical Chemistry Part I, Vol. 49*, W. H. Freeman and Company, **1980**.
[35] S. Krimm, J. Bandekar, *Adv. Protein Chem.* **1986**, *38*, 181.
[36] M. Miyazawa, E. R. Blout, *J. Am. Chem. Soc.* **1960**, *83*, 712.
[37] J. T. Pelton, L. R. McLean, *Anal. Biochem.* **2000**, *277*, 167.
[38] N. Greenfield, G. D. Fasman, *Biochemistry US* **1969**, *8*, 4108.
[39] J. P. Anderson, J. Cappello, D. C. Martin, *Biopolymers* **1994**, *34*, 1049.
[40] P. H. Geil, *Polymer Single Crystals*, R.E. Krieger Pub. Co., Huntington, NY, **1973**.
[41] *Atomic force microscopy: Biomedical methods and applications, Vol. 242*, Totowa, NJ: Humana Press, **2004**.
[42] S. Kinugasa, H. Nakahara, J. I. Kawahara, Y. Koga, H. Takaya, *J. Polym. Sci. Pt. B-Polym. Phys.* **1996**, *34*, 583.
[43] J. X. Wang, A. D. Parkhe, D. A. Tirrell, L. K. Thompson, *Macromolecules* **1996**, *29*, 1548.
[44] S. Y. Fung, C. Keyes, J. Duhamel, P. Chen, *Biophys. J.* **2003**, *85*, 537.
[45] J. Vesenka, M. Guthold, C. L. Tang, D. Keller, E. Delaine, C. Bustamante, *Ultramicroscopy* **1992**, *42*, 1243.

Chapter 6

Alpha-helical polypeptides: scaffolds for the preparation of well-defined glycopolymers

Abstract

The interaction between cell-surface carbohydrate ligands and protein receptors is important in many biological recognition events. Multivalent glycopolymers that mimic carbohydrate ligands can be potent inhibitors of carbohydrate-binding receptors. They are therefore considered as attractive drug candidates for certain disease states such as bacterial infections and chronic inflammation. Chemical syntheses of glycopolymers result in heterogeneous products with a limited control over the positioning of the saccharide epitopes. The protein engineering method can therefore be useful in the generation of monodisperse polypeptide scaffolds which can be functionalized with saccharide moieties at defined positions. Through the use of a polypeptide backbone with an α-helical conformation, it will be possible to place reactive handles at various spatial intervals along the axis of the helix and at defined sides of the helix. A design of such an α-helical polypeptide scaffold is described as well as attempts to produce this artificial polypeptide recombinantly in E. coli.

For the design of an α-helical polypeptide scaffold amino acid residues were chosen with a high intrinsic propensity for the α-helical structure as well as amino acids with reactive handles for functionalization. An initial design with the repetitive amino acid sequence $[(MAKA)_2MAA]_n$ (M = methionine, A = alanine and K = lysine) was made. In an α-helical conformation the methionine and lysine residues are located at opposing sides of the helical axis. The lysine residues can be directly used for the conjugation with saccharide residues, whereas replacement of methionine with the non-proteinogenic analogue 2-amino-5-hexynoic acid via the auxotroph methodology provides a handle for the highly efficient and chemoselective [3 + 2]-cycloaddition with azide-functionalized saccharides.

Circular dichroism spectroscopy was carried out on small $[(MAKA)_2MAA]_n$ polypeptides, with n = 1, 2 and 3 to determine whether the polypeptides adopted an α-helical conformation. Indeed, it was demonstrated that the sequence is prone to adopt this conformation. Although the percentage α-helix was low, it increased with increasing polypeptide length and pH, being 32% for $[(MAKA)_2MAA]_3$ at pH 12. The predicted higher α-helical content of longer polypeptide scaffolds encouraged us to investigate the recombinant production of longer $[(MAKA)_2MAA]_n$ polypeptides in E. coli. Unfortunately, bacterial expression of repetitive polymers of this sequence was unsuccessful. Although immunoblot analysis of expression of T7-tagged $[(MAKA)_2MAA]_n$ (n = 6 and 12) suggested polypeptide expression, Ni-NTA purification was unsuccessful. Expression as a fusion with NusA, which is highly expressed in E. coli, was attempted. Unfortunately, after fusion of the repetitive $[(MAKA)_2MAA]_n$ gene, no expression could be observed.

6.1 Introduction

The creation of materials with structural control at the nanometer level has received increasing attention in the last decades. With modern lithography construction methods well-defined two-dimensional structures in the order of 50 - 100 nm can nowadays efficiently be generated.[1, 2] Besides this 'top-down' procedure, a complementary approach is the build-up of materials from the 'bottom-up' through molecular self-assembly. Molecular self-assembly is characterized by the spontaneous and specific association of molecules as a result of multiple non-covalent interactions. The design of new nano-structured materials has been carried out, inspired by the design principles found in nature using a variety of building blocks, such as small molecules, lipids, DNA, proteins and peptides. Envisioned applications are, e.g. smaller computational devices, smart drug delivery systems and new tissue engineering scaffolds.[3, 4]

Peptides and proteins are particularly attractive building blocks for the construction of new materials, since much is known about their folding behaviour and their interactions. In the field of materials design via protein and peptide assembly we can distinguish between two types of hydrogen-bonded structures, namely the α-helix and the β-sheet. Whereas in β-sheets hydrogen-bonding occurs between different β-strands, hydrogen-bonding in α-helices occurs internally. Because of this capacity to form intermolecular hydrogen bonds the field of materials science is more advanced for β-sheet-based and amyloid-like structures, since it enables the formation of all kinds of regular structures, like tapes, ribbons, tubes, fibrils and monolayers (see previous chapters).

The study of assemblies based on the α-helix lags behind. Research in this field mainly focuses on the well-studied folding motif, the coiled coil. It was first described in 1953 by Pauling, Corey[5] and Crick[6] and is a common motif identified in a large variety of proteins such as keratins, myosin, DNA transcription factors and smaller peptides such as hormones and venoms.[7] A coiled coil is a bundle of α-helices that are wound into a left-handed superhelix. The coiled coil is characterized by the amino acid heptad repeating units designated as '*abcdefg*' with hydrophobic residues at position '*a*' and '*d*', and polar residues at the other positions. This arrangement of amino acid residues results in amphipathic α-helices with opposing polar and nonpolar faces oriented along the helix axis. This drives the association by excluding the hydrophobic surface from the aqueous environment and leads to the formation of coiled coils in general consisting of two to four helices with either parallel or antiparallel orientation.

The self-assembly properties of the coiled coil motif have been used to prepare new fibrous materials. For example, mixing of two peptides designed to form "sticky ended" heterodimeric coiled coils resulted in micrometer-long fibres which are tens of nanometers thick (Figure 6.1a and b).[8, 9] The thickness of these fibres is larger than would be expected for a dimeric coiled coil (~ 2 nm), indicating that lateral assembly takes place.
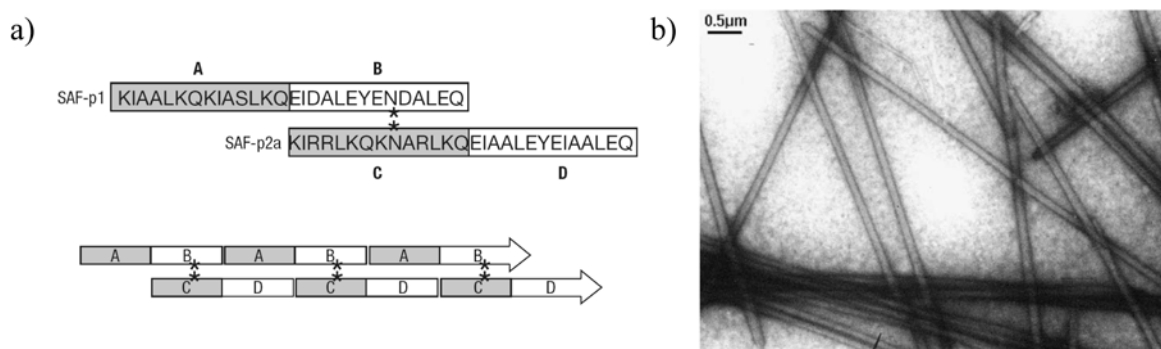
Figure 6.1 (a) *The sequences of two peptides forming sticky ended coiled coil dimers, where A complements D and B complements C, resulting in fibre formation.[8]* ***(b)*** *Transmission electron micrograph of coiled coil self-assembling fibres.[10]*

This concept has been worked out extensively using e.g. T-shaped or dendritic peptides, resulting in branched fibres and fibre networks.[11-15] Furthermore, hybrid diblock copolymers of coiled coil peptides and poly(ethylene glycol) have been prepared and their self-assembly in solution and in the solid state has been investigated.[16-18]

Alternatively, coiled coils have been used for the preparation of hydrogels, which are interesting materials for their application in drug delivery and tissue engineering. Triblock copolymers consisting of a central polyelectrolyte domain with the repetitive sequence -(AG)$_3$PEG- and terminal coiled coil domains, were prepared via bacterial protein expression. Reversible gelation of this artificial polypeptide could be carried out by switching between the aggregated and dissociated state of the coiled coil domains via change in pH or temperature.[19, 20] In a hybrid system, which exhibited a similar stimulus responsive gelation behaviour, coiled coil domains were bound via their His-tag to an iminodiacetate functionalized polyacrylamide polymer.[21]

Besides coiled coil proteins, an example of another α-helical protein which has been prepared via protein engineering is poly(γ-benzyl α,L-glutamate) (PBLG).[22, 23] Monodisperse poly(α-L-glutamic acid) was prepared via bacterial expression and subsequently chemically modified to yield PBLG. Polydisperse PBLG prepared by N-carboxyanhydride (NCA) polymerization was known to form nematic liquid crystalline phases, characterized by orientational, but not positional molecular order.[24] However, solutions of the recombinantly produced polymer exhibited smectic liquid crystalline order. The equal chain length of this polymer resulted in ordering in layers, which is not possible with the polydisperse polymers.

The sequence control of the protein engineering methodology was used for the preparation of artificial, helical glycopolymers with controlled placement of saccharide moieties.[25] Alanine-rich, repetitive protein polymers were prepared with glutamic acid residues placed at defined positions along the helical backbone.[26] The carboxylic acid group of the glutamic acid residues was then modified with the sugar β-D-galactosylamine. The multivalent presentation of galactose was shown to result in an increased binding to the galactose-binding

sites of the cholera toxin in comparison to monovalent galactose. This methodology allows a detailed investigation of saccharide number, saccharide spacing and polymer conformation on the binding properties of glycopolymers.[25, 26]

These artificial glycopolymers could also be of interest as anti-inflammatory drugs.[27-29] The inflammatory response after injury, infection, or disease involves the recruitment of leukocytes from the bloodstream to damaged tissues or lymphoid organs.[30-32] Important mediators in this process are the "selectins", which are transmembrane proteins present at the cell surface of leukocytes, endothelial cells and platelets. L-selectin, for example, is constitutively expressed on leukocytes and binds to cell adhesion molecules on activated endothelial cells. This is the first step of leukocyte migration into inflamed tissue.[33] Inhibition of L-selectin-mediated leukocyte recruitment is considered as a promising approach in the development of anti-inflammatory drugs for the treatment of chronic and acute inflammation diseases, where aggressive leukocyte migration is detrimental.[30, 34] The design of inhibitors can be inspired on the natural ligands for L-selectin. Several ligands that bind L-selectin have been identified, including glycosylated, extended polypeptides. One of these ligands, GlyCAM-1, is a heavily O-glycosylated protein that presents sulfated derivatives of the tetrasaccharide sialyl Lewis x (sLe^x).[35, 36] Inhibitors of L-selectin have therefore been designed that present multiple copies of the sulfated saccharide epitopes. Multivalent selectin inhibitors have demonstrated enhanced binding relative to their monomeric derivatives in cellular and cell free binding inhibition essays.[28]

The design of the appropriate scaffold used for the display of saccharide units is important. Multivalent saccharide displays have been mimicked by the conjugation of saccharides to bovine serum albumin, to short synthetic polypeptides or to lipids.[37, 38] Other approaches are the use of small molecule templates that allow the attachment of a few carbohydrates[39-41] and spherical carbohydrate displays in the form of dendrimers (highly branched oligomers)[42] or liposomes[43]. However, the most popular method for generating multivalent ligands is through the free radical polymerization of modified acrylamides, resulting in linear carbohydrate displays.[44-47] Lack of control over polymer chain length is a disadvantage of this method. An improvement has been the use of ring-opening metathesis polymerization (ROMP).[48-51] ROMP can be carried out in polar solvents and protection of the carbohydrate residues is not required. The synthesized polymers have narrower polydispersities and can be end-capped with reporter groups for detection.[50] Although these investigations have given insights in specificity and avidity of multivalent interactions, a more quantitative understanding is complicated by the heterogeneity of the polymers and the lack of control over saccharide spacing.

The protein engineering method can therefore be useful in the generation of monodisperse polypeptide scaffolds which can be functionalized with saccharide moieties at defined positions. Through the use of a polypeptide backbone with an α-helical conformation, it will be possible to place reactive handles at various spatial intervals along the axis of the helix

and at defined sides of the helix. In this chapter, a design of such an α-helical polypeptide scaffold is described as well as attempts to produce this artificial polypeptide recombinantly in *E. coli*.

6.2 Results and discussion

6.2.1 Helical peptide design

For the design of an α-helical polypeptide scaffold amino acid residues had to be chosen with a high intrinsic propensity for the α-helical structure as well as amino acids with reactive handles for functionalization. Alanine (A) residues were used in positions where a reactive handle was not needed, since this amino acid is a good helix-former.[52, 53] Two amino acid residues, lysine (K) and methionine (M), were chosen for functionalization of the helix. Both of them have high helix-forming propensities. The primary amine-group of lysine can be readily used for a range of chemical modification reactions.[54] Methionine residues can be replaced by non-proteinogenic analogues via the protein engineering approach.[55-58] Some of the methionine analogues that have been incorporated in proteins through the use of methionine auxotroph host strains are depicted in Figure 6.2a. One of the non-natural amino acids that can replace methionine in good efficiency is 2-amino-5-hexynoic acid. Its alkyne functionality can be used for a highly efficient and chemoselective reaction with azide-functionalized molecules via [3 + 2]-cycloaddition (Figure 6.2b).[59, 60]

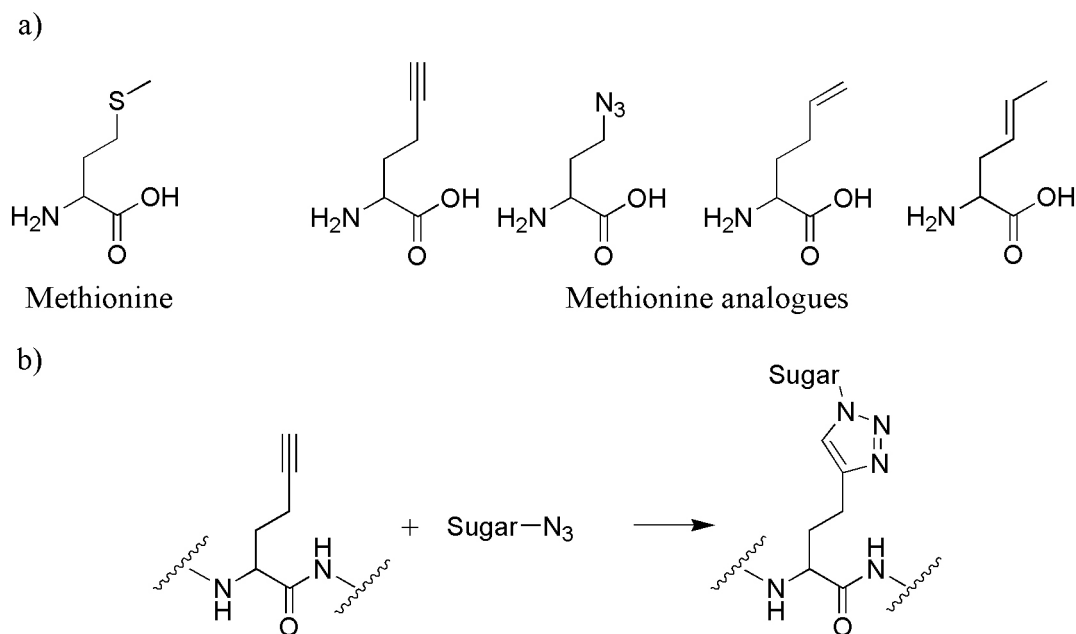*Figure 6.2 (a) Some amino acid analogues that can replace methionine during in vivo protein synthesis. [55-58] (b) After incorporation of 2-amino-5-hexynoic acid into the polypeptide the alkyne functionality can be used for the highly efficient and chemoselective reaction with azide-functionalized sugars via [3 + 2]-cycloaddition.[59, 60]*

A design was made with the aim to incorporate lysine and methionine residues at opposing sides of the helical axis. Several amino acid sequences were tested with different orders and ratios of alanine, methionine and lysine by fitting them in perfect α-helix geometry using the "Macromodel" program. One of them, having the repetitive amino acid sequence [(MAKA)$_2$MAA]$_n$, fitted well with the criterium of aligned methionine and lysine residues which are located at opposing sides of the helical axis (Figure 6.3). With this initial design one alkyne functionality can, in principal, be incorporated at every helical turn (i.e. with spacings of ~ 5.4 Å). At the other side of the helical axis lysine residues occupy approximately 2 of the 3 turns.
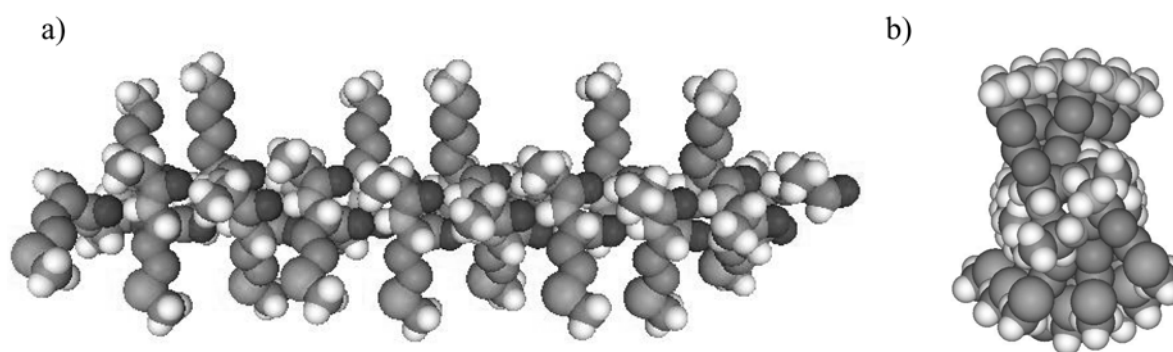

Figure 6.3 *Model of the [(MAKA)$_2$MAA]$_n$ polypeptide (n =3). **(a)** side view **(b)** view along the helix axis. Methylene hydrogens have been omitted for clarity.*

Besides replacing the methionine residues with its 2-amino-5-hexynoic acid analogue, this residue can be used for immobilization of the polypeptide scaffold on a gold surface. Although its thioether group does not bind to gold as strongly as the more frequently used thiol group, this should be compensated by the presence of multiple interactions.[61, 62] This kind of immobilization could be used in surface plasmon resonance studies (SPR, Biacore).[63] With sugar moieties conjugated via the lysine residues, the resulting glycopolymer can be directly coated on the gold surface of the biosensor chip, without the requirement of performing immobilization chemistry. The interaction between the artificial glycopolymer and sugar-binding proteins (lectins) can then be studied.[64]

Based on these considerations the repetitive -(MAKA)$_2$MAA- sequence was chosen as our model peptide for the construction of α-helical scaffolds. [(MAKA)$_2$MAA]$_n$ oligopeptides were prepared via solid phase peptide synthesis and their secondary structure was investigated by circular dichroism spectroscopy (Paragraph 6.2.2). Furthermore, the feasibility of recombinant production of this sequence in *E. coli* was investigated (Paragraph 6.2.3).

6.2.2 Secondary structure determination of [(MAKA)₂MAA]ₙ polypeptides

To determine whether the designed $[(MAKA)_2MAA]_n$ polypeptide scaffold adopted an α-helical structure, polypeptides consisting of one, two and three repeats (n = 1, 2 and 3) of this sequence were prepared by solid phase peptide synthesis. MALDI-TOF analysis of the synthesized polypeptides showed in addition to full-length polypeptide, some signal corresponding to polypeptides having an alanine or lysine deletion. The crude polypeptide product after cleavage from the resin was used for an initial indication of the secondary structure of this sequence. Circular dichroism (CD) spectroscopy was used for the determination of the conformation of $[(MAKA)_2MAA]_n$ in solution. Figure 6.4a shows the effect of polypeptide chain length (n = 1, 2 and 3) on the secondary structure. The spectra were recorded at pH 12, which ensured a deprotonated state of the amine groups of the lysine residues. For n = 1 a spectrum typical of a random coil, with a minimum at 197 nm was observed. However, with increasing polypeptide chain length (n = 2 and 3) the CD spectra with two minima located at 222 nm and around 208 nm, became typical for the α-helical structure.[65]
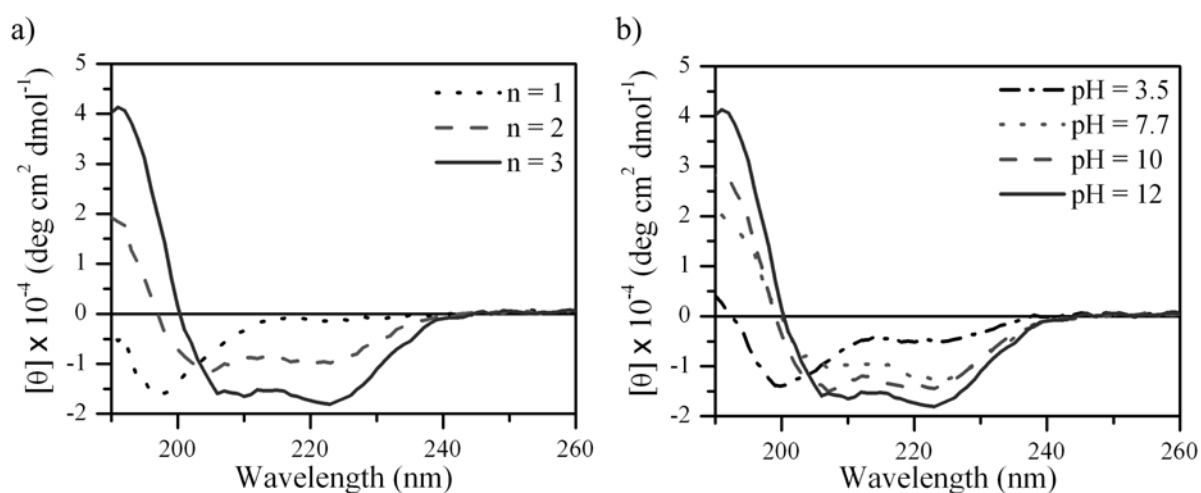

Figure 6.4 *CD spectra of (a) [(MAKA)₂MAA]ₙ polypeptide for n = 1, 2 and 3 at pH 12 and (b) [(MAKA)₂MAA]₃ at varying pH. Spectra were recorded at 25 ºC at polypeptide concentrations of 0.1 mg mL⁻¹.*

The increased optical activity with increasing polypeptide chain length could be expected based on theoretical calculations; the mean residue ellipticity at 222 nm, $[\theta]_{222}$, was shown to vary linearly with the inverse of the number of amino acids.[66, 67] The helical content for the peptides can be estimated from the mean residue ellipticity at 222 nm with the equation, %Helix = $[\theta]_{222}$/{-61000(1 – 2.5/n)}, where n is the number of amino acids.[68] The use of this equation results in a helical content of 3, 18 and 32% for peptides containing one, two and three repeats of the -(MAKA)₂MAA- sequence, respectively. The dependence of the secondary structure on pH for $[(MAKA)_2MAA]_3$ is depicted in Figure 6.4b. With a decreasing pH, the

random coil character increased at the cost of the α-helical secondary structure, as would be expected when the lysine residues become protonated. These results clearly indicate that the designed sequence is prone to adopt an α-helical conformation. The low helical contents can be expected for these relatively short polypeptides, since the stabilizing interactions do not compensate the entropic cost associated with the folding of the polypeptide chain. Furthermore, since the synthesized polypeptides are not protected with uncharged groups (e.g. by N-terminal acetylation and C-terminal amidation), electrostatic repulsion between the charged ends and the helix macrodipole takes place. This results in helix destabilization.[69]

However, the helical content is expected to increase with increasing peptide length.[68, 70] Using the algorithm "Agadir" which predicts the helical behaviour of monomeric peptides, the effect of polypeptide length on α-helix percentage was calculated.[71, 72] For [(MAKA)$_2$MAA]$_n$ polypeptides, where n = 3, 6 and 12, an average helix content of respectively 34, 74 and 87% was calculated. The helix content at the residue level for these polypeptides is depicted in Figure 6.5. This indicated that higher plateau values are reached for longer polypeptides and a value of up to 100% is reached for n = 12. The predicted higher α-helical content of longer polypeptide scaffolds encouraged us to investigate the recombinant production of longer [(MAKA)$_2$MAA]$_n$ polypeptides in *E. coli*.

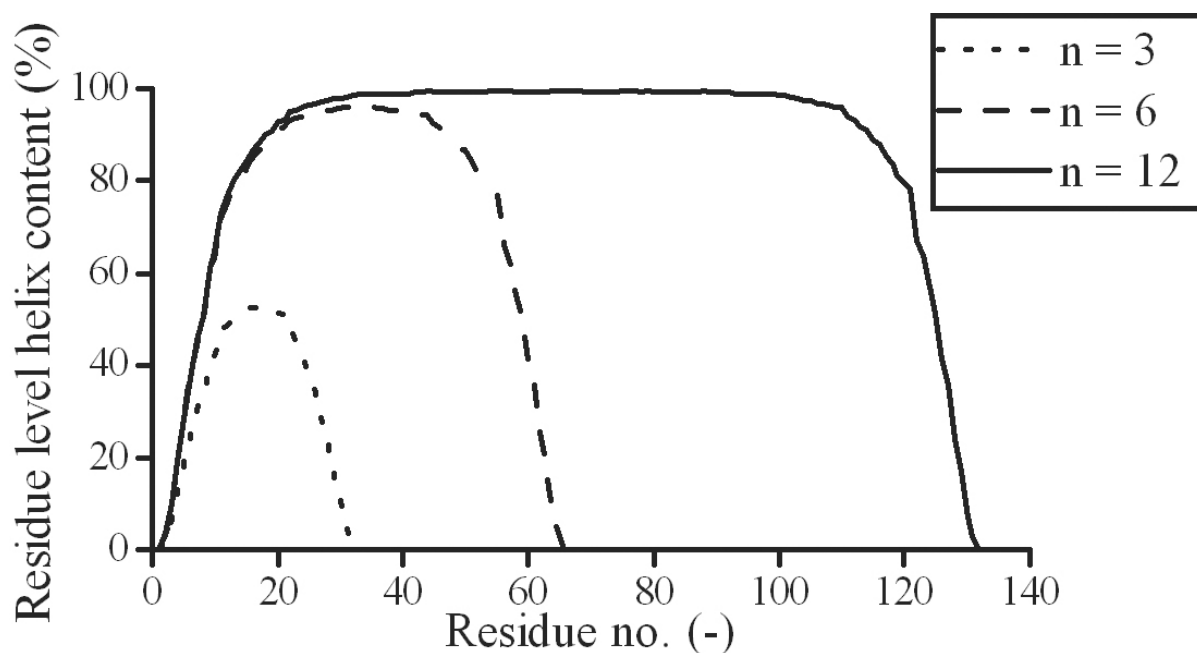**Figure 6.5** *Prediction of α-helix content at the residue level for polypeptides with the repetitive sequence [(MAKA)$_2$MAA]$_n$ of varying length (n = 3, 6, 12). The prediction algorithm "Agadir"[71, 72] was used with the following settings: free N- and C-termini, T = 298 K, pH 7.0, [NaCl] = 0.1 M.*

6.2.3 *E. coli* expression of [(MAKA)$_2$MAA]$_n$ polypeptides

To determine whether bacterial expression would be a suitable method for production of protein polymers consisting of repeats of the α-helical -(MAKA)$_2$MAA- sequence, synthetic genes coding for this sequence were constructed. The cloning strategy used was analogous as described for poly-[(AG)$_3$EG] (Chapter 3, paragraph 3.1) and is described in the experimental section. Artificial genes were built up encoding 6 and 12 repeats of the -(MAKA)$_2$MAA- sequence. The target polypeptide sequence is depicted in Figure 6.6. It contains a T7- and a 6 × His-tag for immunoblot analysis and purification, respectively. Furthermore, the acid-labile aspartyl-prolyl bond allows removal of the tags to result in a polypeptide consisting predominantly of the (MAKA)$_2$MAA sequence.[73]
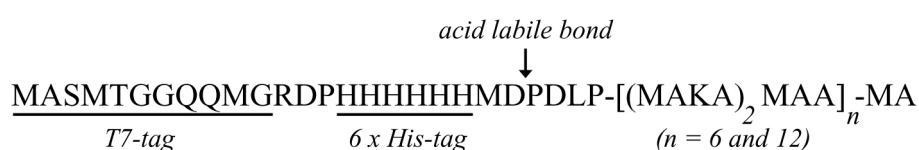
acid labile bond

MASMTGGQQMGRDPHHHHHHMDPDLP-[(MAKA)$_2$MAA]$_n$-MA

T7-tag   6 x His-tag   (n = 6 and 12)

Figure 6.6 *Target sequence of [(MAKA)$_2$MAKA]$_n$ polypeptide. The acid-labile aspartyl-prolyl peptide bond allows removal of the N-terminal T7 and 6 × His tag.*

The expression of the polypeptides in *E. coli* BL21(DE3)pLysS was followed by SDS-PAGE. Coomassie staining did not reveal the presence of overexpressed bands. Immunoblot analysis with T7-tag antibodies showed the presence of multiple bands for both 6 and 12 repeats of the -(MAKA)$_2$MAA- sequence after induction with IPTG (Figure 6.7). Whereas for [(MAKA)$_2$MAA]$_6$, the lowest band could correspond to the desired polypeptide with an expected molecular weight of 9.5 kDa, an additional band at 14 kDa was visible. Furthermore, signal was observed at high molecular weight, in the form of product which did not migrate into the 'separating gel'. Also for [(MAKA)$_2$MAA]$_{12}$, bands were observed at 14 kDa and at high molecular weight. The expected molecular weight of this polypeptide is 15.9 kDa. Purification of the expressed polypeptides was attempted using Ni-NTA chromatography under denaturing conditions. Unfortunately, dot blot analysis with T7-tag antibodies showed that only signal was present in the flow-through (8 M urea, pH 8.0) and the wash fraction (8 M urea, pH 6.3), but not in the elution fractions (8 M urea, pH 5.9 and 4.5) (data not shown). A possible explanation is that, due to its cationic nature, the polypeptide only weakly binds the Ni-NTA beads. However, also in an alternative purification procedure using agarose beads functionalized with T7-tag monoclonal antibodies, no binding seemed to occur.
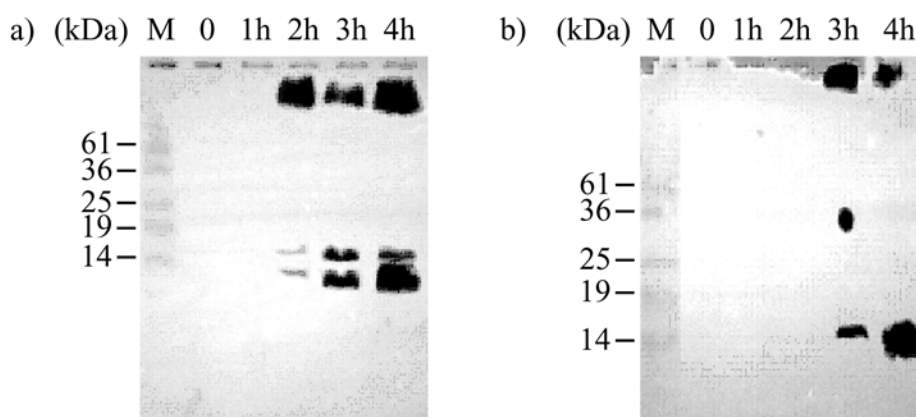
Figure 6.7 *T7-tag immunoblot analysis of expression of polypeptides with the sequence* *[(MAKA)$_2$MAA]$_n$ for **(a)** n = 6 and **(b)** n = 12. Expression was carried out in E. coli BL21(DE3)pLysS cells at 37 °C. Expression was induced by the addition of IPTG to a final concentration of 0.4 mM.*

In an alternative approach, it was attempted to express the [(MAKA)$_2$MAA]$_n$ sequence as a fusion protein. The protein NusA (N utilization substance A; 66 kDa) was chosen a fusion partner, since it can be highly expressed in *E. coli* in the soluble form.[74] Although relatively large in comparison to the desired polypeptide, it was a good way to determine the effect of the [(MAKA)$_2$MAA]$_n$ sequence on the expression level. In addition, experiments with the synthetic [(MAKA)$_2$MAA]$_3$ peptide suggested that this peptide could be toxic to *E. coli* cells when added to the culture medium (data not shown). Similar amphipathic helical peptides have been shown to have antimicrobial activity ascribed to their ability to form pores in the lipid bilayer.[75-77] Fusion with a large protein might prevent this. Furthermore, the host strain Rosetta(DE3)pLysS was used which contains a plasmid carrying among others the methionyl tRNA gene. Insufficient tRNA is known to result in translation problems, which could play a role in the expression of [(MAKA)$_2$MAA]$_n$ because of the relatively high methionine content of this sequence. The production of additional methionyl tRNA might be a solution for expression problems.[78]

The fusion protein consisted of an N-terminal NusA protein, followed by a 6 × His-tag and the target polypeptide (Figure 6.8a). Figure 6.8b shows the SDS-PAGE analysis of the expression and Ni-NTA purification of NusA and its fusion with [(MAKA)$_2$MAA]$_n$ (n = 6 and 12). Whereas for NusA alone (66 kDa) clear expression was observed, no overexpressed bands could be observed for both fusions (69 and 75 kDa). Using Ni-NTA chromatography, substantially purified NusA protein was obtained, whereas for the [(MAKA)$_2$MAA] fusion proteins only weak bands were observed around the expected molecular weight. These were however not shifted upwards relative to NusA alone, which would be expected to be visible at least for [(MAKA)$_2$MAA]$_{12}$. Thus, fusion of the [(MAKA)$_2$MAA] sequence to the highly expressed NusA protein resulted in loss of overexpression. No further experiments have been

performed to determine whether low level expression of these fusion proteins had occurred. The possibility that the fusion proteins were produced in an insoluble form and did not migrate into the gel should be verified as well.
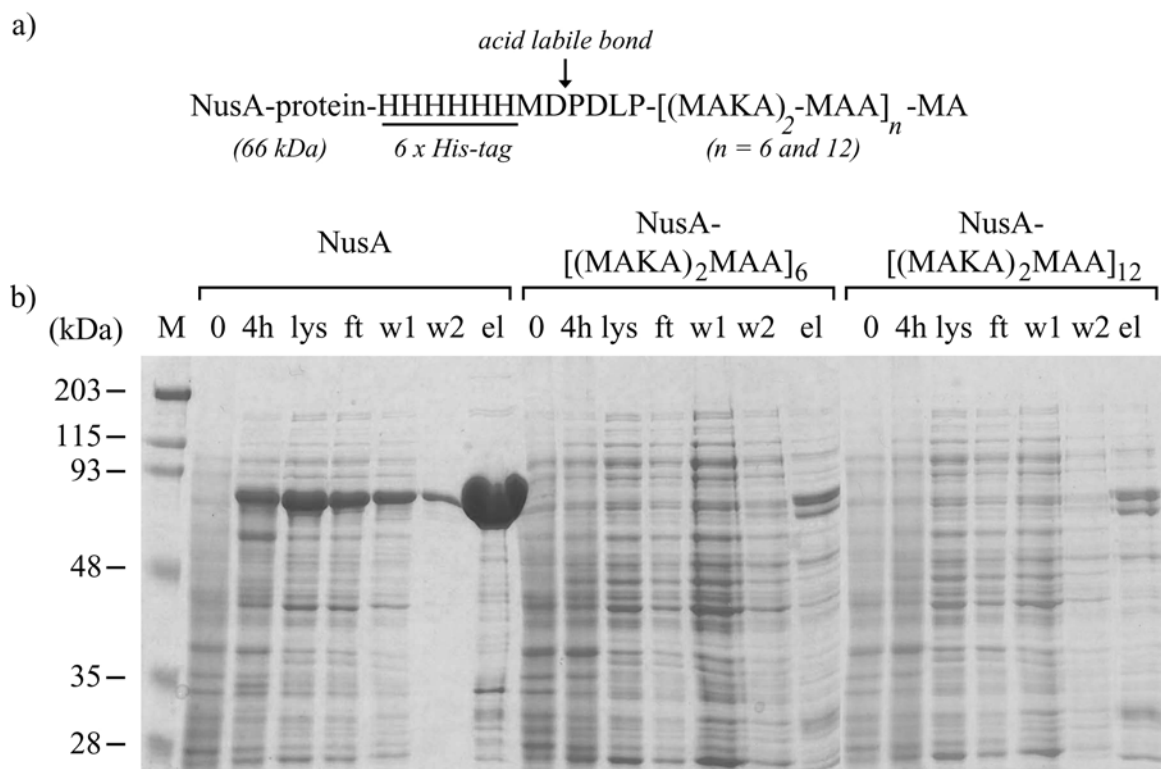
a)

acid labile bond
↓

NusA-protein-HHHHHHMDPDLP-[(MAKA)$_2$-MAA]$_n$-MA
(66 kDa)   *6 x His-tag*   *(n = 6 and 12)*

Figure 6.8 (a) *Target NusA fusion proteins.* **(b)** *SDS-PAGE analysis of expression and native Ni-NTA purification of NusA and its fusions with the repetitive polypeptides [(MAKA)$_2$MAA]$_6$ and (MAKA)$_2$MAA]$_{12}$ (lys = soluble lysate, ft = column flow-through, w1/w2 = column wash and el = elution).*

6.3 Conclusions

The short polypeptides with the sequence [(MAKA)$_2$MAA]$_n$ (n = 1, 2 and 3), which were prepared by solid phase peptide synthesis, were shown to be prone to adopt an α-helical conformation. The α-helical content was shown to increase with increasing peptide length and pH. Although an α-helical content of 32% was still low for [(MAKA)$_2$MAA]$_3$, it was expected to increase significantly for longer polypeptides.

However, the recombinant production of longer polypeptides in *E. coli* was unsuccessful. Although immunoblot analysis seemed to suggest the expression of the T7-tagged [(MAKA)$_2$MAA]$_n$ polypeptide (n = 6 and 12), no binding to Ni-NTA agarose beads was observed. Fusion of the repetitive genes to the 3'-end of the highly expressed NusA gene, resulted in loss of overexpression. Although methionine-rich repetitive proteins have been expressed in *E. coli*[79], expression and purification of this specific sequence has not been successful.

6.4 Experimental section

6.4.1 Materials

The synthetic oligonucleotides were obtained from Isogen. The vectors pBluescriptII® SK(-) and pET-3b and *E. coli* strain XL1-Blue were obtained from Stratagene. The vector pET-43.1b and the *E. coli* strains BL21(DE3)pLysS and Rosetta(DE3)pLysS were from Novagen. Restriction endonucleases (Gibco/New England Biolabs), T4 polynucleotide kinase (Gibco), T4 DNA ligase (Promega), calf intestine phosphatase (New England Biolabs) were used as recommended by the manufacturers. Plasmid midiprep kits and Qiaquick® gel extraction kits were from Qiagen. Isopropyl-β-D-thiogalactopyranoside (IPTG) and phenylmethanesulphonyl fluoride (PMSF) were purchased from Sigma. Ni-NTA agarose beads were obtained from Qiagen. The T7-tag Lumiblot™ Kit was from Novagen. Sinapinic acid was from Sigma.

6.4.2 General methods

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot analysis. SDS-PAGE was performed using 15% resolving gels according to standard procedures.[80] Typically, cell pellets obtained from 0.5 mL cell culture were resuspended in ($OD_{600} \times 100$) μl and heated for 5 min at 95 °C. 10 μL of this sample was loaded on gel. For immunoblot analysis transfer to a nitrocellulose membrane was carried out, followed by T7-tag immunoblot analysis using the 'T7-tag® Lumiblot™ Kit', according to the manufacturer's instructions.

MALDI-TOF mass spectrometry. Spectra were measured on a Bruker Biflex III spectrometer. Freeze-dried products were dissolved in 1:1 (v/v) water:acetonitrile with 0.1% trifluoroacetic acid to a final concentration of 10 mg mL^{-1} and mixed in a 1:1 ratio with a solution of 20 mg mL^{-1} of sinapinic acid in the same solvent and spotted on a MALDI-plate.

6.4.3 Synthetic [(MAKA)$_2$MAA]$_n$ peptides

The synthetic peptides with the sequence [(MAKA)$_2$MAA]$_n$ (n = 1, 2 and 3) were synthesized by Hans Adams by solid phase peptide synthesis on a trityl resin using an Fmoc strategy.

MS (MALDI-TOF) H-[(MAKA)$_2$MAA]$_n$-OH:

n = 1: $C_{45}H_{84}N_{13}O_{12}S_3$ [M + H]$^+$: calc. mass 1094.5 Da, found m/z = 1093.8 Da, [M + Na]$^+$: calc. mass 1116.5 Da, found m/z = 1115.8 Da. In addition to full-length product, peptides with the following deletions were found: Alanine deletion [M + H]$^+$: calc. mass 1023.5 Da, found m/z = 1022.8 Da. Lysine deletion [M + H]$^+$: calc. mass 966.5 Da, found m/z = 965.7 Da.

n = 2: $C_{90}H_{166}N_{26}O_{23}S_6$ [M + H]$^+$: calc. mass 2170.1 Da, found m/z = 2169.3 Additional signals: m/z = 2098.3 (- alanine) and 2060.3 Da (not assigned).

n = 3: $C_{135}H_{245}N_{39}O_{34}S_9$ [M + H]$^+$ calc. mass 3245.6 Da, found m/z = 3245.6 Da. Additional weak signals: m/z = 3117.3 (- lysine) and 2991.6 Da (not assigned).

6.4.4 Circular dichroism spectroscopy

Circular dichroism spectroscopy was performed on a Jasco J-810 spectropolarimeter (band width: 1nm, response: 1 s, sensitivity: standard, λ range: 260 - 180 nm, data pitch: 1 nm, scanning speed: 100 nm min^{-1}, accumulation: 1) in a 1 mm quartz cuvette. Measurements were carried out at a concentration of 0.1 mg mL^{-1}. Dissolution of the peptides in demi-water (milli-Q) resulted in a pH of 3.5. The pH was adjusted to pH 10 and 12, respectively, by the addition of a NaOH solution. The peptides were dissolved in 20 mM NaH$_2$PO$_4$ buffer, pH 7.7 for measurement at this pH. Measurements were performed at 25 ºC. The mean residue ellipticity, [θ] in deg cm^2 dmol^{-1}, was calculated from

$$[\theta] = (\theta \times MW) / (10 \times l \times c \times n) \quad\quad (1)$$

where θ is the ellipticity measured in mdeg, MW is the molecular weight in g mol^{-1}, l is the optical path length of the cell in cm, c is the concentration of the sample in mg mL^{-1} and n is the number of amino acid residues.

6.4.5 Construction of expression vectors with genes coding for [(MAKA)$_2$MAA]$_n$

The double-stranded oligonucleotide 1 coding for 2 repeats of the sequence (n = 2) is depicted below. Annealing was carried out as described in Chapter 3, Paragraph 4.3.

```
       Ala  Lys  Ala  Met  Ala  Lys  Ala  Met  Ala  Ala  Met  Ala  Lys  Ala  Met  Ala  Lys  Ala  Met  Ala  Ala  Met
5'-GGCGAAGGCAATGGCGAAAGCAATGGCGGCGATGGCCAAAGCTATGGCCAAAGCTATGGCGGCCAT
        TTCCGTTACCGCTTTCGTTACCGCCGCTACCGGTTTCGATACCGGTTTCGATACCGCCGGTACCGC-5'
```

oligonucleotide 1

A cloning vector was constructed for insertion of this oligonucleotide. Thereto, *Eco*RI/*Bam*HI digested pBluescriptII$^®$ SK(-) was ligated with the double-stranded oligonucleotide 2, depicted below, resulting in the vector pSK-RT.

```
        Arg  Asp  Pro  His  His  His  His  His  His  Asp  Pro  Asp  Leu  Pro  Met  Ala   *
5'-AATTCGGATCCGCACCATCACCACCATCACGATCCGGACCTGCCGATGGCGTAAGGCAGGTG
        GCCTAGGCGTGGTAGTGGTGGTAGTGCTAGGCCTGGACGGCTACCGCATTCCGTCCACCTAG-5'
  EcoRI   BamHI                                        BspMI          BspMI   BamHI
```

oligonucleotide 2

This vector was digested with *Bsp*MI and dephosphorylated with calf intestine phosphatase. The linearized vector was ligated with oligonucleotide 1 and the sequence of this construct, designated as pSK-RT-MAKA2, was confirmed. For further multimerization, this vector was digested with *Bsp*MI. The 66 bp oligonucleotide was isolated from agarose gel and multimerized by T4 DNA ligase. After 1.5 hours of ligation, *Bsp*MI digested and dephosphorylated pSK-RT was added and ligation was continued overnight. Via this procedure

cloning vectors were obtained carrying 6 and 12 repeats of the -(MAKA)$_2$MAA- sequence (designated as pSK-RT-MAKA6 and pSK-RT-MAKA12). These vectors were digested with *Bam*HI, followed by ligation of the genes with the *Bam*HI digested and dephosphorylated expression vector pET-3b or pET-43.1b.

6.4.6 Expression and purification of T7-tagged [(MAKA)$_2$MAA]$_n$ polypeptides

The expression vectors pET-3b-RT-MAKA6 and pET-3b-RT-MAKA12 (coding for 6 and 12 repeats of the -(MAKA)$_2$MAA- sequence, respectively) were transformed into the *E. coli* strain BL21(DE3)pLysS. A typical starter culture of 50 mL 2 × YT medium, containing 200 μg mL^{-1} ampicillin and 30 μg mL^{-1} chloramphenicol, was inoculated with a single colony from a fresh agar plate and incubated overnight at 37 °C with shaking (300 rpm). After centrifugation at 3000 g for 15 min at 4 °C, the pellet was resuspended in 5 mL 2 × YT medium. This was used for inoculation of 500 mL 2 × YT medium (200 μg mL^{-1} ampicillin and 30 μg mL^{-1} chloramphenicol) to give an OD$_{600}$ of 0.1. Incubation was carried out at 37 °C with shaking (300 rpm) and expression was induced with IPTG (0.4 mM), when an OD$_{600}$ of ~ 0.8 was reached. The cells were harvested 4 hours after induction by centrifugation at 3000 g for 15 min at 4 °C and the pellet was stored at -80 °C.

The pellets were resuspended in lysis buffer (8 M urea, 100 mM NaH$_2$PO$_4$, 10 mM Tris.HCl, pH 8.0) at 5 mL per gram wet weight. The cells were stirred for 30 min at room temperature. The lysate was centrifuged at 10000 g for 30 min at room temperature and the supernatant was loaded on a 1 mL Ni-NTA column. The column was washed subsequently with 10 mL wash buffer (8 M urea, 100 mM NaH$_2$PO$_4$, 10 mM Tris.HCl, pH 6.3). Elution was carried out using subsequently 5 mL elution buffer, pH 5.9 and 5 mL elution buffer, pH 4.5 (8 M urea, 100 mM NaH$_2$PO$_4$, 10 mM Tris.HCl).

6.4.7 Expression and purification of NusA fusion proteins

The plasmids pET-43.1b, pET-43.1b-RT-MAKA6 and pET-43.1b-RT-MAKA12 were used for transformation of *E. coli* strain Rosetta(DE3)pLysS. A single colony was used to inoculate 50 mL 2 × YT medium containing 200 μg mL^{-1} ampicillin, 30 μg mL^{-1} chloramphenicol and 1% glucose. After growth overnight at 30 °C, the cells were centrifuged at 4000 rpm for 15 min at 4 °C. The pellet was resuspended in 5 mL 2 × YT and used to inoculate 50 mL 2 × YT medium (200 μg mL^{-1} ampicillin and 30 μg mL^{-1} chloramphenicol) to give an OD$_{600}$ of 0.1. This culture was grown at 37 °C (230 rpm) and expression was induced with IPTG (0.4 mM), when an OD$_{600}$ of ~ 0.8 was reached. Every hour a 1 mL sample was taken and centrifuged at 8000 rpm for 5 min. The pellet was stored at -20 °C for analysis by SDS-PAGE. The cells were harvested 4 hours after induction by centrifugation at 3000 g for 15 min at 4 °C and the pellet was stored at -80 °C.

Pellets were resuspended in 5 mL lysisbuffer (50 mM NaH_2PO_4 (pH8.0), 300 mM NaCl, 1 mM PMSF, 10 mM imidazole) and sonicated on ice for 1 min (250 W Branson sonicator, 50% duty cycle). The lysate was centrifuged at 10000 g for 30 min. at 4 °C. The soluble fraction was used for native Ni-NTA purification at 4 °C. This was carried out using disposable columns (BioRad) filled with 100 µL equilibrated Ni-NTA agarose beads. The lysate was loaded on the column and allowed to empty by gravity. The beads were then subsequently washed with 1 mL lysisbuffer and 1mL washbuffer (50 mM NaH_2PO_4 (pH8.0), 300 mM NaCl, 1 mM PMSF, 20 mM imidazole). Elution was carried out by the addition of 300 µL elution buffer (50 mM NaH_2PO_4 (pH8.0), 300 mM NaCl, 1 mM PMSF, 200 mM imidazole).

6.5 References

[1]   Y. Xia, G. M. Whitesides, *Angew. Chem. Int. Edit.* **1998**, *37*, 550.
[2]   S. Kramer, R. R. Fuierer, C. B. Gorman, *Chem. Rev.* **2003**, *103*, 4367.
[3]   S. G. Zhang, X. J. Zhao, *J. Mater. Chem.* **2004**, *14*, 2082.
[4]   S. G. Zhang, *Nat. Biotechnol.* **2003**, *21*, 1171.
[5]   L. Pauling, R. B. Corey, *Nature* **1953**, *171*, 59.
[6]   F. H. C. Crick, *Acta Crystal.* **1953**, *6*, 689.
[7]   A. Lupas, *Trends Biochem. Sci.* **1996**, *21*, 375.
[8]   M. G. Ryadnov, D. N. Woolfson, *Nature Materials* **2003**, *2*, 329.
[9]   M. J. Pandya, G. M. Spooner, M. Sunde, J. R. Thorpe, A. Rodger, D. N. Woolfson, *Biochemistry US* **2000**, *39*, 8728.
[10]   C. E. MacPhee, D. N. Woolfson, *Curr. Opin. Solid State Mat. Sci.* **2004**, *8*, 141.
[11]   M. G. Ryadnov, D. N. Woolfson, *J. Am. Chem. Soc.* **2004**, *126*, 7454.
[12]   M. G. Ryadnov, D. N. Woolfson, *J. Am. Chem. Soc.* **2005**, *127*, 12407.
[13]   M. G. Ryadnov, D. N. Woolfson, *Angew. Chem. Int. Edit.* **2003**, *42*, 3021.
[14]   M. G. Ryadnov, B. Ceyhan, C. M. Niemeyer, D. N. Woolfson, *J. Am. Chem. Soc.* **2003**, *125*, 9388.
[15]   A. M. Smith, S. F. A. Acquah, N. Bone, H. W. Kroto, M. G. Ryadnov, M. S. P. Stevens, D. R. M. Walton, D. N. Woolfson, *Angew. Chem. Int. Edit.* **2005**, *44*, 325.
[16]   G. W. M. Vandermeulen, D. Hinderberger, H. Xu, S. S. Sheiko, G. Jeschke, H. A. Klok, *Chemphyschem* **2004**, *5*, 488.
[17]   G. W. M. Vandermeulen, C. Tziatzios, R. Duncan, H. A. Klok, *Macromolecules* **2005**, *38*, 761.
[18]   H. A. Klok, G. W. M. Vandermeulen, H. Nuhn, A. Rosler, I. W. Hamley, V. Castelletto, H. Xu, S. S. Sheiko, *Faraday Discussions* **2005**, *128*, 29.
[19]   W. A. Petka, J. L. Harden, K. P. McGrath, D. Wirtz, D. A. Tirrell, *Science* **1998**, *281*, 389.
[20]   S. B. Kennedy, E. R. deAzevedo, W. A. Petka, T. P. Russell, D. A. Tirrell, M. Hong, *Macromolecules* **2001**, *34*, 8675.
[21]   C. Wang, R. J. Stewart, J. Kopecek, *Nature* **1999**, *397*, 417.
[22]   S. J. M. Yu, V. P. Conticello, G. H. Zhang, C. Kayser, M. J. Fournier, T. L. Mason, D. A. Tirrell, *Nature* **1997**, *389*, 167.
[23]   S. M. Yu, C. H. Soto, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 6552.
[24]   C. Robinson, J. C. Ward, *Nature* **1957**, *180*, 1183.
[25]   Y. Wang, K. L. Kiick, *J. Am. Chem. Soc.* **2005**, *127*, 16392.
[26]   R. S. Farmer, L. M. Argust, J. D. Sharp, K. L. Kiick, *Macromolecules* **2006**, *39*, 162.
[27]   R. Roy, *Curr. Opin. Struct. Biol.* **1996**, *6*, 692.
[28]   E. E. Simanek, G. J. McGarvey, J. A. Jablonowski, C.-H. Wong, *Chem. Rev.* **1998**, *98*, 833.
[29]   M. Mammen, S.-K. Choi, G. M. Whitesides, *Angew. Chem. Int. Edit.* **1998**, *37*, 2754.
[30]   R. Gonzalez-Amaro, F. Sanchez-Madrid, *Crit. Rev. Immunol.* **1999**, *19*, 389.
[31]   D. Vestweber, J. E. Blank, *Physiol. Rev.* **1999**, *79*, 181.
[32]   R. A. Worthylake, K. Burridge, *Curr. Opin. Cell Biol.* **2001**, *13*, 569.
[33]   S. D. Rosen, C. R. Bertozzi, *Curr. Opin. Cell Biol.* **1994**, *6*, 663.
[34]   K. Ley, *Trends Mol. Med.* **2003**, *9*, 263.
[35]   L. A. Lasky, M. S. Singer, D. Dowbenko, Y. Imai, W. J. Henzel, C. Grimley, C. Fennie, N. Gillett, S. R. Watson, S. D. Rosen, *Cell* **1992**, *69*, 927.
[36]   S. Hemmerich, H. Leffler, S. D. Rosen, *J. Biol. Chem.* **1995**, *270*, 12035.
[37]   S. Y. C. Wong, *Curr. Opin. Struct. Biol.* **1995**, *5*, 599.
[38]   L. L. Kiessling, N. L. Pohl, *Chem. Biol.* **1996**, *3*, 71.

[39] G. Kretzschmar, D. Seiffge, *Tetrahedron* **1995**, *51*, 13015.
[40] O. Renkonen, S. Toppila, L. Penttila, H. Salminen, J. Helin, H. Maaheimo, C. E. Costello, J. P. Turunen, R. Renkonen, *Glycobiology* **1997**, *7*, 453.
[41] R. Stahn, H. Schaefer, F. Kernchen, J. Schreiber, *Glycobiology* **1998**, *8*, 311.
[42] S. M. Rele, W. Cui, L. Wang, S. Hou, G. Barr-Zarse, D. Tatton, Y. Gnanou, J. D. Esko, E. L. Chaikof, *J. Am. Chem. Soc.* **2005**, *127*, 10132.
[43] R. E. Bruehl, F. Dasgupta, T. R. Katsumoto, J. H. Tan, C. R. Bertozzi, W. Spevak, D. J. Ahn, S. D. Rosen, J. O. Nagy, *Biochemistry US* **2001**, *40*, 5964.
[44] F. D. Tropper, A. Romanowska, R. Roy, *Methods in Enzymology* **1994**, *242*, 257.
[45] R. Roy, W. K. C. Park, O. P. Srivastava, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1399.
[46] Y. Nishida, H. Uzawa, T. Toba, K. Sasaki, H. Kondo, K. Kobayashi, *Biomacromolecules* **2000**, *1*, 68.
[47] K. Sasaki, Y. N., T. Tsurumi, H. Uzawa, H. Kondo, K. Kobayashi, *Angew. Chem. Int. Edit.* **2002**, *41*, 4463.
[48] C. Fraser, R. H. Grubbs, *Macromolecules* **1995**, *28*, 7248.
[49] K. H. Mortell, M. Gingras, L. L. Kiessling, *J. Am. Chem. Soc.* **1994**, *116*, 12053.
[50] E. J. Gordon, J. E. Gestwicki, L. E. Strong, L. L. Kiessling, *Chem. Biol.* **2000**, *7*, 9.
[51] P. Mowery, Z. Q. Yang, E. J. Gordon, O. Dwir, A. G. Spencer, R. Alon, L. L. Kiessling, *Chem. Biol.* **2004**, *11*, 725.
[52] P. Chou, G. Fasman, *Biochemistry US* **1974**, *13*, 222.
[53] A. Chakrabartty, R. L. Baldwin, *Adv. Protein Chem.* **1995**, *46*, 141.
[54] G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, San Diego, **1996**.
[55] K. L. Kiick, D. A. Tirrell, *Tetrahedron* **2000**, *56*, 9487.
[56] K. L. Kiick, J. C. M. van Hest, D. A. Tirrell, *Angew. Chem. Int. Edit.* **2000**, *39*, 2148.
[57] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282.
[58] J. C. M. van Hest, D. A. Tirrell, *Chem. Commun.* **2001**, 1897.
[59] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Edit.* **2001**, *40*, 2004.
[60] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192.
[61] A. E. Strong, B. D. Moore, *J. Mater. Chem.* **1999**, *9*, 1097.
[62] H. Schönherr, F. J. B. Kremer, S. Kumar, J. A. Rego, H. Wolf, H. Ringsdorf, M. Jaschke, H. J. Butt, E. Bamberg, *J. Am. Chem. Soc.* **1996**, *118*, 13051.
[63] K. Nagata, H. Handa, *Real-time analysis of biomolecular interactions: applications of Biacore, Vol. 1*, Springer, Tokyo.
[64] E. Duverger, N. Frison, R. A. C., M. Monsigny, *Biochimie* **2003**, *85*, 167.
[65] J. T. Pelton, L. R. McLean, *Anal. Biochem.* **2000**, *277*, 167.
[66] R. W. Woody, I. Tinoco, *J. Chem. Phys.* **1967**, *46*, 4927.
[67] Y. H. Chen, J. T. Yang, H. Martinez, *Biochemistry US* **1972**, 4120.
[68] J. S. Miller, R. J. Kennedy, D. S. Kemp, *J. Am. Chem. Soc.* **2002**, *124*, 945.
[69] A. J. Doig, R. L. Baldwin, *Protein Sci.* **1995**, *4*, 1325.
[70] L. Williams, K. Kather, D. S. Kemp, *J. Am. Chem. Soc.* **1998**, *120*, 11033.
[71] V. Munoz, L. Serrano, *Nature: Struct. Biol.* **1994**, *1*, 399.
[72] http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html.
[73] M. Landon, *Methods in Enzymology* **1977**, *47*, 145.
[74] R. Harrison, *InNovations* **2000**, *11*, 4.
[75] T. B. Wyman, F. Nicol, O. Zelphati, P. V. Scaria, C. Plank, F. C. Szoka, *Biochemistry US* **1997**, *36*, 3008.
[76] F. Nicol, S. Nir, F. C. Szoka, *Biophys. J.* **1999**, *76*, 2121.
[77] D. R. Frohlich, M. A. Wells, *Int. J. Pept. Protein Res.* **1991**, *37*, 2.

[78] C. Kurland, J. Gallant, *Curr. Opin. Biotechnol.* **1996**, *7*, 489.

[79] M. J. Dougherty, S. Kothakota, T. L. Mason, D. A. Tirrell, M. J. Fournier, *Macromolecules* **1993**, *26*, 1779.

[80] F. Ausubel, R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, K. Struhl, *Current protocols in molecular biology, Vol. 3*, John Wiley & sons, New York, **1999**.

Summary

The synthesis of macromolecules of well-defined length, composition and end-group functionality is of utmost importance in polymer chemistry. Although substantial progress has been made in this area, the biological synthesis of macromolecules, such as proteins and DNA, is unsurpassed. For proteins, the amino acid sequence determines polypeptide folding and assembly into supramolecular structures. The level of complexity in the assembly of proteins has not been equaled by synthetic systems, such as block copolymers. Protein-like self-assembly has therefore become a popular strategy for the preparation of materials with ordering at the nanometer scale.

This thesis describes the recombinant production of artificial, periodic polypeptides that form regular β-sheet structures. These polypeptides have the repetitive amino acid sequence $[(AG)_3EG]_n$ (A = alanine, G = glycine and E = glutamic acid). The repeating alanylglycine (AG) sequence is also found in silk of the domesticated silk worm *Bombyx mori* and is part of the crystalline β-sheet domains, which are responsible for the strength of the silk fiber. By the presence of bulky, polar glutamic acid residues (E) at regular distances the β-strand structure is disturbed. Crystallization of these artificial polypeptides results in a regular folding with polypeptide chains folding back on themselves at the position of the glutamic acid residues. This results in an antiparallel β-sheet structure with alanylglycine repeats in the extended β-strands and glutamic acid residues in the turns (Figure 1).
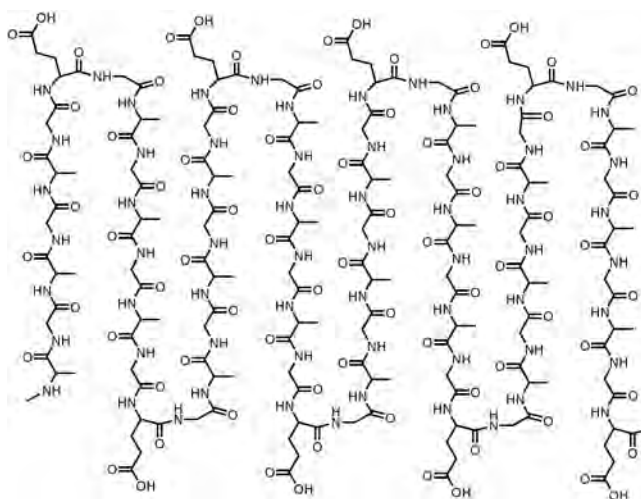
Figure 1. *The formation of an antiparallel β-sheet structure by a polypeptide with the repetitive amino acid sequence $[(AG)_3EG]_n$ in the solid phase. The alanylglycine (AG) repeats form antiparallel β-strands. The glutamic acid residues are located at the turns.*

Although folding of these polypeptides is controlled, crystallization results in large aggregates. The structure information in the polypeptide sequence, such as height and width of the β-sheet structures is lost in the aggregation process. To avoid this "hybrid" block copolymers were

synthesized, in which the aggregating β-sheet polypeptide block is flanked on both sides by a non-aggregating poly(ethylene glycol) (PEG) block. This thesis describes subsequently the recombinant production of repetitive β-sheet polypeptides (Chapter 3), the synthesis of hybrid block copolymers (Chapter 4), and the physical characterization of these block copolymers (Chapter 5). These chapters are preceded by a literature overview of synthetic methods for the preparation of hybrid block copolymers and their potential applications (Chapter 2). Moreover, Chapter 6 describes efforts made for the recombinant production of repetitive α-helical polypeptides, which could be useful for the preparation of well-defined glycopolymers.

The construction of artificial genes coding for polypeptides with the repetitive sequence $[(AG)_3EG]_n$ is described in Chapter 3. The expression of these polypeptides in *E. coli* was carried out via a number of strategies. Polypeptides consisting of 10 to 50 repeats of the -$(AG)_3EG$- sequence were effectively produced as fusion proteins with glutathione-S-transferase (GST, 26 kDa). The expression of the fusion protein GST-$[(AG)_3EG]_{20}$ resulted in a yield of approximately 80 mg per liter of culture. After proteolytic cleavage of the fusion-protein with thrombin, the polypeptide $[(AG)_3EG]_{20}$ was separated from GST by reversed phase chromatography, resulting in a yield of approximately 20 mg $[(AG)_3EG]_{20}$ per liter of culture. In an alternative approach $[(AG)_3EG]_n$ (n = 10, 20, 30) was produced with the T7 bacteriophage expression system (pET). The polypeptide $[(AG)_3EG]_{20}$ was produced without the presence of a fusion partner. Besides complete polypeptide truncated polypeptide was formed, which could be removed with reversed phase chromatography. The yield was also approximately 20 mg per liter of culture. The advantage of this method was that no proteolytic cleavage was necessary and that removal of the N and C termini could be carried out directly by cyanogen bromide cleavage.

Also the expression of a polypeptide with the repetitive sequence $[(AG)_3KG]_n$ is described. The expression of this polypeptide was carried out with the aim to use the lysine residues for functionalization with e.g. mesogenic groups to obtain liquid crystalline materials. However, the production was complicated by the sensitivity of this sequence to proteolysis. The polypeptide $[(AG)_3KG]_{24}$ was purified directly after expression by Ni-NTA chromatography. Polypeptides with the complete repetitive part were separated from the truncated polypeptides by gel filtration chromatography and a yield of approximately 2 mg per liter of culture was obtained.

The synthesis and chemical characterization of block copolymers consisting of the polypeptide block $[(AG)_3EG]_n$ flanked on both sides by a PEG block is described in Chapter 4. The polypeptides $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ were expressed in *E. coli*. Besides complete polypeptide also truncated polypeptides were formed (from N terminus to approximately 6 -$(AG)_3EG$- repeats). The polypeptide $[(AG)_3EG]_{10}$ was produced in erlenmeyer cultures and purified with Ni-NTA chromatography. The complete polypeptide was separated from the truncated product using gel filtration chromatography. The yield was approximately 6 mg per liter of culture. The polypeptide $[(AG)_3EG]_{20}$ was produced with high cell density fermentation

(HCDF) and purified with subsequently Ni-NTA chromatography, heat precipitation and reversed phase chromatography. The yield was approximately 10 mg polypeptide per liter of culture. HCDF was unfortunately not successful in the production of larger quantities of polypeptide, since the amount of expressed polypeptide per cell was clearly lower (5 – 50 times) in HCDF than in erlenmeyer cultures.

The cysteine residues on both sides of the polypeptides $[(AG)_3EG]_{10}$ and $[(AG)_3EG]_{20}$ were used for the coupling of maleimide functionalized PEG (Figure 2). The polypeptide $[(AG)_3EG]_{20}$ was coupled to PEG-750 (M_n = 750 g mol^{-1}). For the polypeptide $[(AG)_3EG]_{10}$ conjugation was carried out to PEGs with various lengths (M_n = 750, 2000 and 5000 g mol^{-1}). The coupling was carried out subsequently by (1) reduction of the cysteine residues with dithiothreitol (DTT), (2) removal of DTT by precipitation of the polypeptide with trichloroacetic acid, (3) coupling of an excess maleimide functionalized PEG, (4) purification of the block copolymer with Ni-NTA chromatography and (5) removal of the non-repetitive ends of the polypeptide by cyanogen bromide cleavage. Mass spectrometry showed that the main product was polypeptide functionalized with two PEG chains.
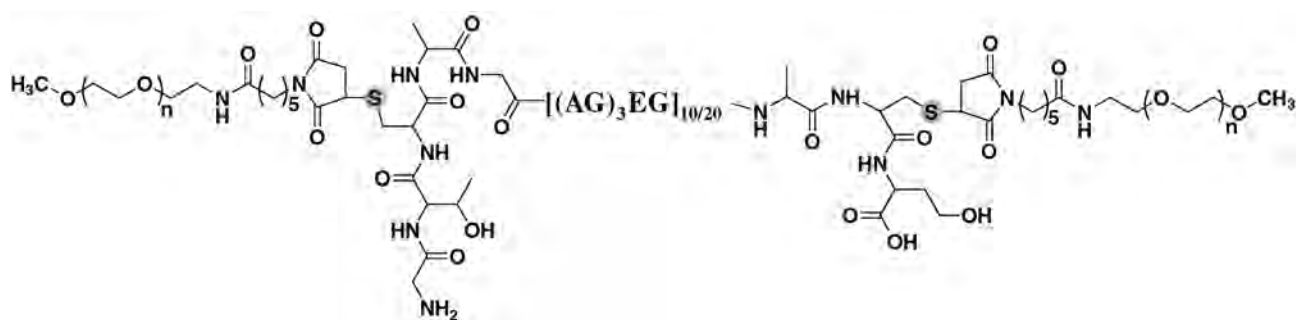

Figure 2. *Block copolymer consisting of the β-sheet polypeptide poly-[(AG)₃EG] flanked by PEG end blocks. Cysteine residues on both sides of the polypeptide block were used for conjugation of maleimide functionalized PEG (n = 17, 47, and 115).*

Chapter 5 describes the physical characterization of the block copolymers described in Chapter 4. After crystallization the secondary structure of the polypeptide block was characterized with infra-red spectroscopy. Thereto the frequencies in the amide I (predominantly C=O stretch vibration) and the amide II (predominantly NH-bend vibration) area were used. The predominant absorption bands had frequencies of ~ 1625 cm^{-1} for amide I and ~ 1520 cm^{-1} for amide II. These frequencies are typical for an antiparallel β-sheet conformation of the polypeptide block. These were the predominant absorption bands for both the polypeptide alone as well as for the PEG-conjugates. With longer PEG-chains the absorption band around 1655 cm^{-1} became relatively stronger, indicating that the fraction of the polypeptide chain in the β-sheet conformation decreased.

The microstructure of the block copolymers was analyzed by transmission electron microscopy (TEM). It was shown that the block copolymers formed fibrils, whereas this fibrillar structure was not present for the polypeptides without conjugated PEG. The fibrils for

the block copolymer [(AG)$_3$EG]$_{20}$-PEG-750 had a width of approximately 12 nm (Figure 3). For the block copolymer [(AG)$_3$EG]$_{10}$-PEG-750 (where the polypeptide block has half of the length) no fibrils were visible with TEM. Atomic force microscopy (AFM) was carried out on samples applied to a mica surface. From this it was clear that crystallization of the block copolymer [(AG)$_3$EG]$_{10}$-PEG-750 also resulted in fibril formation. The height of the fibrils was approximately 2 nm for both [(AG)$_3$EG]$_{10}$-PEG-750 as well as [(AG)$_3$EG]$_{20}$-PEG-750. For the polypeptide [(AG)$_3$EG]$_{10}$ conjugated to PEG-2000 a clear fibrillar structure was present, whereas for the conjugate with PEG-5000 the fibrils were less defined and only short fibril fragments were present.
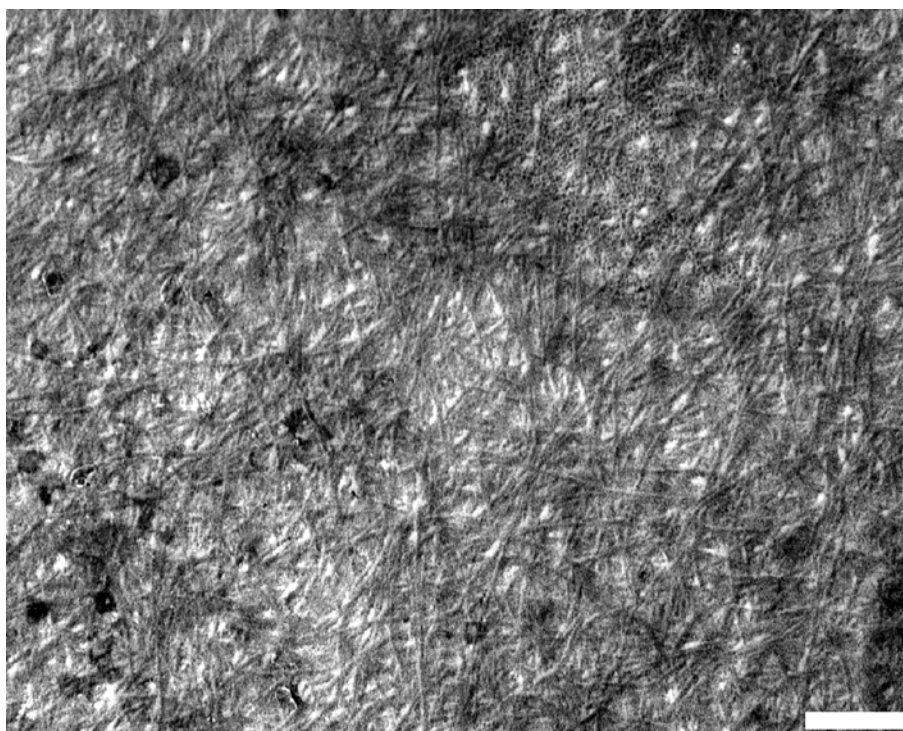

Figure 3. Electron microscopy images of the crystallized block copolymer [(AG)$_3$EG]$_{20}$-PEG-750. The scale bar represents 200 nm.

A possible ordening of the block copolymers within the fibrils was proposed based on TEM and AFM results. The model consists of a stack of β-sheets along the length-axis of the fibrils. The hydrogen-bond direction is perpendicular to the length-axis of the fibrils. The PEG-chains prevent a side-to-side aggregation of the fibrils. Diffraction data are, however, required for a direct evidence of this model.

Chapter 6 describes the design of an α-helical polypeptide for the preparation of well-defined glycopolymers. The interaction between cell-surface carbohydrate ligands and protein receptors is important in many biological recognition events. Multivalent glycopolymers that mimic carbohydrate ligands can be potent inhibitors of carbohydrate-binding receptors. They are therefore considered as attractive drug candidates for certain disease states such as chronic

and acute inflammations. Chemical syntheses of glycopolymers result in heterogeneous products with a limited control over the positioning of the saccharide epitopes. The recombinant production method can therefore be useful in the generation of monodisperse polypeptides which can be functionalized with saccharide moieties at defined positions. Through the use of a polypeptide backbone with an α-helical conformation, it will be possible to place reactive handles at various spatial intervals along the axis of the helix and at defined sides of the helix. A design of such an α-helical polypeptide is described as well as attempts to produce this artificial polypeptide recombinantly in *E. coli*.

For the design of an α-helical polypeptide amino acid residues were chosen with a high propensity for the α-helical conformation as well as amino acids with reactive handles for functionalization. An initial design with the repetitive amino acid sequence $[(MAKA)_2MAA]_n$ (M = methionine, A = alanine and K = lysine) was made. In an α-helical conformation the methionine and lysine residues are located at opposing sides of the helical axis. The lysine residues can be directly used for the conjugation with saccharide residues, whereas replacement of methionine with the non-proteinogenic analogue 2-amino-5-hexynoic acid via the auxotroph methodology provides a handle for the highly efficient and chemoselective [3 + 2]-cycloaddition with azide-functionalized saccharides.

Circular dichroism spectroscopy was carried out on small $[(MAKA)_2MAA]_n$ polypeptides, with n = 1, 2 and 3 to determine whether the polypeptides adopted an α-helical conformation. Indeed, it was demonstrated that the sequence is prone to adopt this conformation. Although the percentage α-helix was low, it increased with increasing polypeptide length and pH, being 32% for $[(MAKA)_2MAA]_3$ at pH 12. The predicted higher α-helical content of longer polypeptides encouraged us to investigate the recombinant production of longer $[(MAKA)_2MAA]_n$ polypeptides in *E. coli*. Unfortunately, bacterial expression of repetitive polymers of this sequence was unsuccessful. Although immunoblot analysis of expression of $[(MAKA)_2MAA]_n$ (n = 6 and 12) with an N-terminal T7-tag suggested polypeptide expression, Ni-NTA purification was unsuccessful. Expression as a fusion with NusA, a protein that can be highly expressed in *E. coli*, was attempted. Unfortunately, after fusion of the repetitive $[(MAKA)_2MAA]_n$ gene, no expression could be observed.

Samenvatting

In de polymeerchemie is de synthese van macromoleculen met een goed gedefinieerde lengte, compositie en eindgroepfunctionaliteit uiterst belangrijk. Hoewel grote vooruitgang op dit gebied is geboekt, is de biologische synthese van macromoleculen, zoals eiwitten en DNA, onovertroffen. Voor eiwitten bepaalt de aminozuurvolgorde de vouwing van de polypeptideketen en de assemblage van eiwitten tot supramoleculaire structuren. De mate van complexiteit in de assemblage van eiwitten is niet geëvenaard door synthetische systemen, zoals blok copolymeren. Eiwit-gebaseerde assemblage is daarom een belangrijke strategie in de synthese van materialen met structuur op de nanometer schaal.

Dit proefschrift beschrijft de recombinante productie van kunstmatige, periodieke polypeptiden die regelmatige β-plaat structuren vormen. Deze polypeptiden hebben de repetitieve aminozuurvolgorde $[(AG)_3EG]_n$ (A = alanine, G = glycine en E = glutaminezuur). De zich herhalende alanylglycine (AG) sequentie komt ook voor in zijde van de gedomesticeerde zijderups *Bombyx mori* en maakt daar onderdeel uit van de kristallijne β-plaat domeinen, die verantwoordelijk zijn voor de sterkte van de zijdevezel. Door nu op regelmatige afstand grote, polaire glutaminezuur residuen (E) in te bouwen wordt de β-streng structuur verstoord. Kristallisatie van deze kunstmatige polypeptiden resulteert in een regelmatige vouwing waarbij de polypeptideketen op zichzelf terugvouwt ter plaatse van de glutaminezuur residuen. Hierdoor ontstaat een antiparallele β-plaat structuur met alanylglycine herhalingen in de uitgestrekte β-strengen en glutaminezuur residuen in de bochten (Figuur 1).
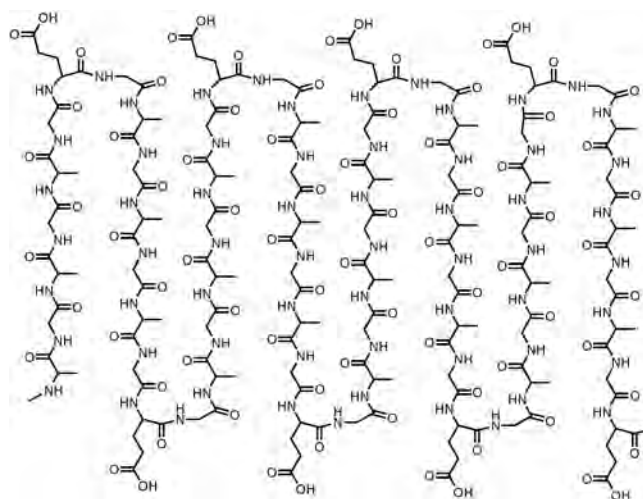
Figuur 1. *De vorming van een antiparallelle β-plaat structuur door een polypeptide met de repetitieve aminozuurvolgorde $[(AG)_3EG]_n$ in de vaste fase. The alanylglycine (AG) herhalingen vormen antiparallelle β-strengen. De glutaminezuur residuen zijn gelocaliseerd in de bochten.*

Hoewel voor deze polypeptiden de vouwing gecontroleerd is, leidt kristallisatie tot grote aggregaten. De structuurinformatie in de polypeptide sequentie, zoals de hoogte en de breedte

van de β-plaat structuren gaat verloren in het aggregatieproces. Om dit te voorkomen werden "hybride" blok-copolymeren gesynthetiseerd, waarbij het aggregerende β-plaat polypeptide blok aan beide zijden wordt geflankeerd door een niet-aggregerend poly(ethyleenglycol) (PEG) blok. In dit proefschrift wordt achtereenvolgens de recombinante productie van repetitieve β-plaat polypeptiden (Hoofdstuk 3), de synthese van de hybride blok-copolymeren (Hoofdstuk 4), en de fysische karakterisatie van deze blok-copolymeren beschreven (Hoofdstuk 5). Deze hoofdstukken worden voorafgegaan door een literatuuroverzicht van synthetische methodes voor het maken van hybride blok-copolymeren en hun mogelijke toepassingen (Hoofdstuk 2). Daarnaast beschrijft hoofdstuk 6 het ontwerp en de synthese van repetitieve α-helix polypeptiden, die gebruikt kunnen worden voor het maken van goed gedefinieerde glycopolymeren.

In hoofdstuk 3 wordt de constructie van kunstmatige genen coderend voor een polypeptide met de repetitieve sequentie $[(AG)_3EG]_n$ beschreven. De expressie van dit polypeptide in *E. coli* werd uitgevoerd via een aantal strategieën. Polypeptides met 10 tot 50 herhalingen van de -$(AG)_3EG$- sequentie konden effectief worden geproduceerd als fusie-eiwit met glutathion-S-transferase (GST, 26 kDa). De expressie van het fusie-eiwit GST-$[(AG)_3EG]_{20}$ resulteerde na GST-affiniteitschromatografie in een opbrengst van ca. 80 mg per liter culture. Na proteolytische klieving van het fusie-eiwit met thrombine werd het polypeptide $[(AG)_3EG]_{20}$ gescheiden van GST d.m.v. omgekeerde fase chromatografie, resulterend in een opbrengst van ca. 20 mg $[(AG)_3EG]_{20}$ per liter culture. Als alternatief werd $[(AG)_3EG]_n$ (n = 10, 20, 30) geproduceerd met het T7 bacteriofaag expressiesysteem (pET). Het polypeptide $[(AG)_3EG]_{20}$ werd geproduceerd zonder de aanwezigheid van een grote fusiepartner. Hierbij werd naast compleet polypeptide verkort produkt gevormd, dat verwijderd kon worden met omgekeerde fase chromatografie. De opbrengst was ook ca. 20 mg per liter culture. Het voordeel van deze methode was dat geen proteolytische klieving nodig was en dat het verwijderen van de aminozuren aan de N- en C-uiteinden direct met cyanogeen bromide klieving uitgevoerd kon worden.

Ook de expressie van een polypeptide met de repetitieve sequentie $[(AG)_3KG]_n$ is beschreven. De expressie van dit polypeptide was uitgevoerd met het doel om de lysine residuen te gebruiken voor de functionalisatie met bijv. mesogene groepen om vloeibaar-kristallijne materialen te verkrijgen. De productie werd bemoeilijkt door de gevoeligheid van deze sequentie voor proteolyse. Het polypeptide $[(AG)_3KG]_{24}$ werd na expressie direct gezuiverd met Ni-NTA chromatografie. Polypeptide met het complete repetitieve deel werd gescheiden van de verkorte producten d.m.v. gelfiltratie chromatografie en een opbrengst van ca. 2 mg per liter culture werd verkregen.

In hoofdstuk 4 wordt de synthese en chemische karakterisatie beschreven van blok-copolymeren waarin het polypeptide blok $[(AG)_3EG]_n$ aan beide zijden wordt geflankeerd door een PEG blok. De polypeptiden $[(AG)_3EG]_{10}$ en $[(AG)_3EG]_{20}$ werden tot expressie gebracht in *E. coli*. Gedurende de expressie werd naast het complete polypeptide ook verkorte polypeptiden

gevormd (van N-terminus tot ongeveer 6 -(AG)$_3$EG- herhalingen). Het polypeptide [(AG)$_3$EG]$_{10}$ werd geproduceerd in erlenmeyer cultures en gezuiverd door middel van Ni-NTA chromatografie. Het complete polypeptide werd gescheiden van de verkorte produkten middels gelfiltratie chromatografie. Een opbrengst van 6 mg per liter culture werd verkregen. Het polypeptide [(AG)$_3$EG]$_{20}$ werd geproduceerd met behulp van hoge celdichtheid fermentatie (HCDF) en gezuiverd met achtereenvolgens Ni-NTA chromatografie, hitte precipitatie en omgekeerde fase chromatografie. De opbrengst was 10 mg polypeptide per liter culture. HCDF was helaas niet succesvol in de productie van grotere hoeveelheden polypeptide, aangezien de hoeveelheid tot expressie gebracht polypeptide per cel duidelijk lager was (5 – 50 maal) in HCDF dan in erlenmeyer cultures.

De cysteine residuen aan beide zijden van de polypeptiden [(AG)$_3$EG]$_{10}$ en [(AG)$_3$EG]$_{20}$ werden gebruikt voor de koppeling van maleimide gefunctionaliseerd PEG (Figuur 2). Het polypeptide [(AG)$_3$EG]$_{20}$ werd gekoppeld met PEG-750 (M_n = 750 g mol^{-1}). Voor het polypeptide [(AG)$_3$EG]$_{10}$ werd conjugatie uitgevoerd met PEGs van verschillende lengtes (M_n = 750, 2000 en 5000 g mol^{-1}). De koppeling werd uitgevoerd door achtereenvolgens (1) reductie van de cysteine residuen met dithiothreitol (DTT), (2) verwijdering van DTT d.m.v. precipitatie van het polypeptide met trichloorazijnzuur, (3) koppeling met een overmaat maleimide gefunctionaliseerd PEG, (4) zuivering van het blok-copolymeer m.b.v. Ni-NTA chromatografie en (5) verwijdering van de niet-repetitieve uiteinden van het polypeptide d.m.v. klieving met cyanogeen bromide. Massa spectrometrie liet zien dat het hoofdproduct overeenkwam met polypeptide gefunctionaliseerd met twee PEG ketens.
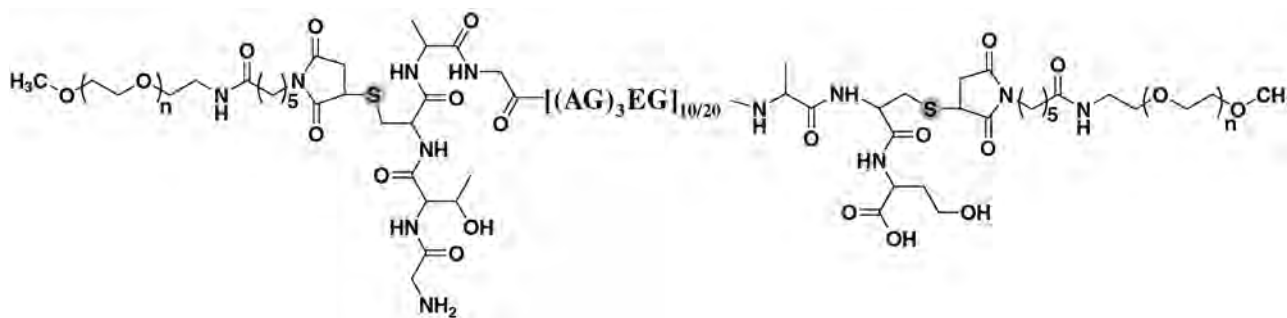

Figuur 2. *Blok-copolymeer bestaande uit het β-plaat polypeptide poly-[(AG)$_3$EG] geflankeerd door poly(ethyleenglycol) (PEG) eindblokken. De cysteine residuen aan weerszijden van het polypeptide blok werden gebruikt voor conjugatie van maleimide gefunctionaliseerd PEG (n = 17, 47 en 115).*

Hoofdstuk 5 beschrijft de fysische karakterisatie van de in hoofdstuk 4 beschreven blok-copolymeren. Na kristallisatie werd de secundaire structuur van het polypeptide blok gekarakteriseerd met behulp van infra-rood spectroscopie. Hiervoor werden de frequenties in het amide I (vnl. C=O strekvibratie) en het amide II (vnl. NH-buigvibratie) gebied gebruikt. De overheersende absorptiebanden hadden frequenties van ~ 1625 cm^{-1} voor amide I en ~ 1520 cm^{-1} voor amide II. Deze frequenties zijn typisch voor een antiparallelle β-plaat conformatie van het polypeptide blok. Dit waren de overheersende absorptiebanden zowel voor het

polypeptide alleen als voor de PEG-conjugaten. Voor langere PEG-ketens werd de absorptieband rond 1655 cm^{-1} relatief sterker, wat erop duidt dat de fractie van de polypeptideketen in de β-plaat conformatie afneemt.

De microstructuur van de blok-copolymeren werd geanalyseerd met electronenmicroscopie (TEM). Hierbij werd aangetoond dat de blok-copolymeren fibrillen vormden, maar dat deze fibrillaire structuur niet aanwezig was voor de polypeptiden zonder geconjugeerd PEG. De fibrillen voor het blok-copolymeer [(AG)$_3$EG]$_{20}$-PEG-750 hadden afmetingen van ongeveer 12 nm (Figuur 3). Voor het blok-copolymeer [(AG)$_3$EG]$_{10}$-PEG-750 (waarbij het polypeptide blok de helft van de lengte heeft) waren de fibrillen niet zichtbaar met electronenmicroscopie. Atoomkracht microscopie (AFM) werd uitgevoerd op monsters aangebracht op een mica oppervlak. Hieruit bleek dat ook kristallisatie van het blok-copolymeer [(AG)$_3$EG]$_{10}$-PEG-750 leidde tot de vorming van fibrillen. De hoogte van de fibrillen was ongeveer 2 nm voor zowel [(AG)$_3$EG]$_{10}$-PEG-750 als [(AG)$_3$EG]$_{20}$-PEG-750. Voor het polypeptide [(AG)$_3$EG]$_{10}$ geconjugeerd met PEG-2000 was er nog steeds een duidelijke fibrillaire structuur aanwezig, terwijl voor het conjugaat met PEG-5000 de fibrillen minder goed gedefinieerd waren en slecht korte fragmenten te zien waren.
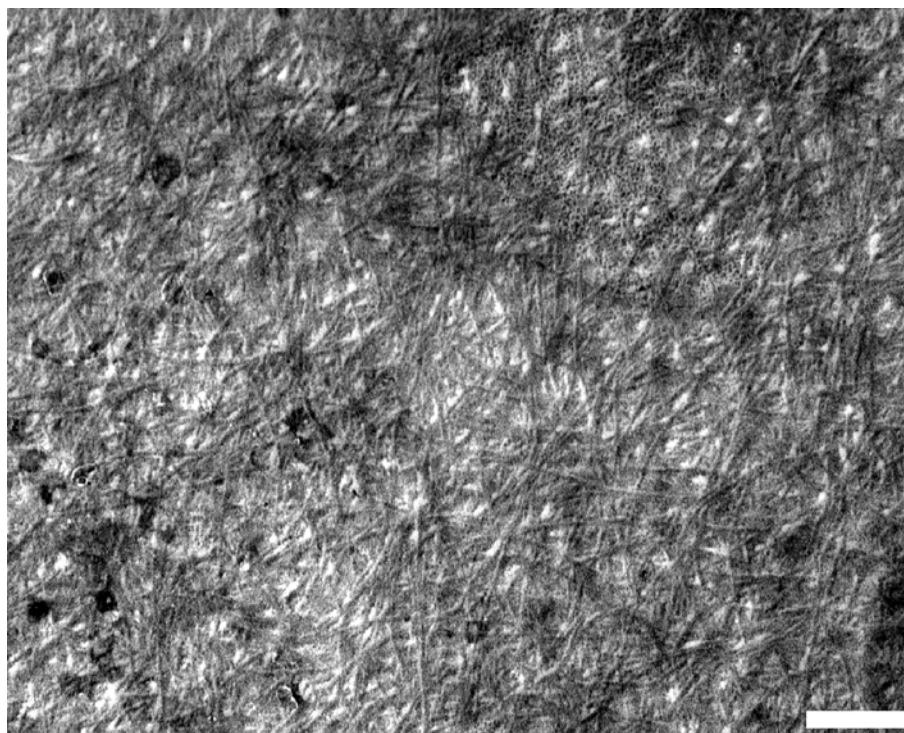
Figuur 3. *Electronenmicroscopie opname van het gekristalliseerde blok-copolymeer [(AG)$_3$EG]$_{20}$-PEG-750. De schaalstreep komt overeen met 200 nm.*

Gebaseerd op de TEM en AFM resultaten is een mogelijke ordening van de blok-copolymeren in de fibrillen voorgesteld. Het model bestaat uit een stapeling van β-platen in de lengte-as van de fibrillen. Hierbij staat de richting van waterstofbruggen loodrecht op de lengte-as van de

fibrillen. De PEG-ketens voorkomen daarbij een zijdelingse aggregatie van de fibrillen. Voor een direct bewijs van dit model zijn diffractie gegevens echter noodzakelijk.

Hoofdstuk 6 beschrijft het ontwerp van een α-helicaal polypeptide welke gebruikt kan worden voor het maken van goed gedefinieerde glycopolymeren. De interactie tussen suiker liganden op het celoppervlak en eiwit receptoren is belangrijk in veel biologische herkenningsprocessen. Multivalente glycopolymeren welke suiker liganden nabootsen kunnen sterke inhibitoren zijn van suiker-bindende receptoren. Ze worden daarom beschouwd als interessante kandidaten om als medicijn ingezet te worden in ziektes, zoals chronische en acute ontstekingen. De chemische synthese van glycopolymeren resulteert in heterogene producten met een beperkte controle over de plaatsing van de suiker epitopen. Recombinante productie methoden kunnen daarom nuttig zijn om monodisperse polypeptides te maken die gefunctionaliseerd kunnen worden met suikereenheden op gedefinieerde posities. Door het gebruik van een polypeptide met een α-helicale conformatie is het mogelijk om reactieve groepen op verschillende afstanden van elkaar langs de helix-as te positioneren. Het ontwerp van zo'n α-helix polypeptide en pogingen om dit kunstmatige polypeptide recombinant te produceren in *E. coli* zijn beschreven.

Voor het ontwerp van een α-helicaal polypeptide werden aminozuur residuen gekozen die een sterke neiging hebben om een α-helix conformatie aan te nemen en die reactieve groepen hebben voor functionalisatie. Een eerste ontwerp is gemaakt met de repetitieve aminozuurvolgorde $[(MAKA)_2MAA]_n$ (M = methionine, A = alanine en K = lysine). In een α-helix conformatie zijn de methionine en lysine residuen aan tegenovergestelde zijden van de helix as geplaatst. De lysine residuen kunnen direct gebruikt worden voor de conjugatie met suiker residuen. De methionine residuen kunnen middels het gebruik van methionine auxotrofe *E. coli* stammen vervangen worden door de niet-natuurlijke aminozuur analoog 2-amino-5-hexynzuur. Daarmee wordt een functionele groep aangebracht die gebruikt kan worden voor de zeer efficiënte en chemoselectieve [3 + 2]-cycloadditie met azide-gefunctionaliseerde suikergroepen.

Circulair dichroisme spectroscopie werd uitgevoerd aan kleine $[(MAKA)_2MAA]_n$ polypeptiden met n = 1, 2 en 3 om te bepalen of de polypeptides een α-helix conformatie aannemen. Er werd aangetoond dat de aminozuurvolgorde geneigd is tot het aannemen van deze conformatie. Hoewel het α-helix percentage laag was, nam het toe met langere polypeptide lengte en hogere pH. Voor $[(MAKA)_2MAA]_3$ was het α-helix percentage 32% bij pH 12. Het voorspelde hogere α-helix percentage voor langere polypeptiden was bemoedigend om de recombinante productie van deze polypeptiden in *E. coli* te onderzoeken. Helaas was de bacteriële expressie van deze repetitieve polypeptidende niet succesvol. Hoewel immunoblot analyse een succesvolle expressie van $[(MAKA)_2MAA]_n$ (n = 6 en 12) met een N-terminale T7-tag suggereerde, was de zuivering met Ni-NTA chromatografie niet succesvol. Expressie als

fusie met NusA, een eiwit dat zeer goed tot expressie komt in *E. coli,* werd uitgeprobeerd. Bacteriële expressie van deze fusie-eiwitten werd echter niet aangetoond.

Dankwoord

Op deze laatste pagina's wil ik de ruimte nemen om de mensen te bedanken zonder wiens hulp en steun dit proefschrift nooit tot stand gekomen was. Allereerst wil ik mijn promotor, prof. Jan van Hest, bedanken voor het geven van de mogelijkheid om dit onderzoek uit te voeren. Je hebt me de vrijheid gegeven om het onderzoek op mijn eigen manier aan te pakken. Zeker toen het onderzoek aanvankelijk moeizaam verliep, heb ik veel steun gehad aan je enthousiaste en motiverende ideeën en je geduld. Ik heb veel van je geleerd, zowel op het professionele als op het persoonlijke vlak.

Ik wil ook mijn copromotor, prof. Henk Stunnenberg, bedanken. Niet alleen heb je me de mogelijkheid gegeven om (langer dan oorspronkelijk gepland) in je lab te werken, maar ik heb ook veel steun gehad aan je hulp bij allerlei praktische zaken (zoals het uitvoeren van eiwitzuiveringen) en je kritische commentaren.

Gelukkig was ik met het feit dat ik in het begin van mijn AiO-schap het kantoortje heb mogen delen met Colin, Paul en Torill. De sfeer in dit kantoortje heb ik als zeer bijzonder ervaren. Colin wil ik daarnaast speciaal bedanken voor de tijd die hij voor me heeft vrijgemaakt om me te helpen met het opzetten van het DNA-werk. Verder wil ik Anita bedanken wiens praktische hulp onmisbaar was o.a. bij het op een degelijke manier uitvoeren van kloneringen en eiwitzuiveringen. Daarnaast wil ik Josephine en Maria bedanken voor al hun hulp.

Theo Sonke en Wilco Peeters van DSM Research wil ik bedanken voor het feit dat ze me de mogelijkheid hebben gegeven om fermentaties uit te voeren in het lab in Geleen. Jens Thies van DSM Research ben ik erkentelijk voor zijn aanzet tot het maken van zijde-analoge hybride blok copolymeren en het verschaffen van plasmiden.

Het belang van degelijke chemische analyses is gedurende mijn promotieonderzoek steeds duidelijker geworden. Mijn dank gaat daarbij uit naar Rien van der Gaag (aminozuuranalyse en HPLC) en Ad Swolfs (NMR), die uitgebreid de tijd hebben genomen om mij te helpen. Ook belangrijk voor mijn proefschrift waren de microscopie analyses. Huub Geurts wil ik danken voor zijn begeleiding bij het uitvoeren van de electronenmicroscopie metingen. Matthijs Otten en Peter Schön wil ik bedanken voor het uitvoeren van de atoomkracht microscopie metingen en de enthousiaste samenwerking die we gehad hebben.

Verder wil ik natuurlijk iedereen van de Bio-organische Chemie groep bedanken. Hans Adams ben ik niet alleen erkentelijk voor de peptiden die hij voor me gesynthetiseerd heeft, maar ook voor al zijn praktische hulp, zoals de uitvoering van eiwithydrolyses en de tijd die hij voor me genomen heeft om mijn chemische vragen te beantwoorden. Ook Dennis Löwik en Joris Meijer wil ik bedanken voor al hun hulp. Lee Ayres, Henri Spijker en Joost Opsteen wil ik bedanken voor o.a. hun hulp bij het uitvoeren van polymerisaties en de GPC-karakterisaties. Peter van Dijk wil ik daarnaast bedanken voor zijn hulp bij het verkrijgen van

fermentoronderdelen. De secretaresses Maria Versteeg en Désirée van der Weij ben ik erkentelijk voor hun hulp bij alle administratieve zaken.

Gedurende mijn promotieonderzoek heb ik het geluk gehad om drie studenten geheel of gedeeltelijk te mogen begeleiden. Als eerste Joost Clerx, die gedurende zijn bijvakstage de aanzet heeft gegeven voor het ontwerp van α-helix polypeptiden die in het laatste hoofdstuk van dit proefschrift zijn beschreven. Door de ervaring die je al had opgedaan in je hoofdvak had je een hoge mate van zelfstandigheid en het was daarom ook leerzaam om met je samen te werken. Ik hoop dat de opgedane technieken je geholpen hebben bij het doen van je eigen promotieonderzoek. Tjaart-Jan Huizing, polymeerchemie student aan de Hogeschool van Arnhem en Nijmegen, wil ik bedanken voor het uitvoeren van gecontroleerde polymerisaties van methyl methacrylaat en de moeilijke experimenten die hij ondernomen heeft om deze relatief eenvoudig te synthetiseren polymeren te koppelen aan de wat minder voorradige β-plaat polypeptiden. Rosalie Teeuwen, ik wil je bedanken voor de inzet die je getoond hebt bij je pogingen om de α-helix polypeptiden recombinant te produceren, iets was geenszins eenvoudig bleek te zijn. Ik hoop dat de ervaringen die je daarbij hebt opgedaan, nuttig zijn geweest voor de uitvoering van je eigen promotieonderzoek.

Coen, Torill en Xavier, jullie aanwezigheid in het lab en daarbuiten heeft ervoor gezorgd, dat ik een geweldig leuke en interessante tijd heb gehad. Zowel de vele spontane avondjes eten en drinken als onze actieve vakanties waren een goede afleiding van het labwerk. Goede herinneringen heb ik dan ook aan de rustige avondjes praten, de wat meer fanatieke avondjes "Kolonisten", het langlaufen naar de top van de Synnfjell (en de ongecontroleerde afdaling ervan), de koude, maar ontspannen sleetochtjes in Røros, het door de modder kruipen in de grotten van de Ardennen en het paardrijden aan de Franse kust.

Belangrijk was ook de ontspanning op sportief gebied. Ik wil daarom ook iedereen van het voetbalteam Uni vv1 en in het bijzonder trainer Wil de Haard bedanken voor deze leuke tijd. Tot slot wil ik mijn vrienden en familie bedanken voor hun steun, vertrouwen en geduld.

List of publications

J. M. Smeenk, T. J. Huizing, H. G. Stunnenberg, J. C. M. van Hest "Well-defined polymer architectures for controlled assembly: Beta-sheet containing hybrid block copolymers" *Polym. Prep.* **226,** U360 (2003).

J. M. Smeenk, L. Ayres, H. G. Stunnenberg, J. C. M. van Hest "Polymer protein hybrids" *Macromol. Symp.* **225,** 1 (2005).

J. M. Smeenk, D. Lowik, J. C. M. van Hest "Peptide-containing block copolymers: Synthesis and potential applications of bio-mimetic materials" *Curr. Org. Chem.* **9,** 1115 (2005).

J. M. Smeenk, M. B. J. Otten, J. Thies, D. A. Tirrell, H. G. Stunnenberg, J. C. M. van Hest "Controlled assembly of macromolecular beta-sheet fibrils" *Angew. Chem. Int. Edit.* **44,** 1968 (2005).

J. M. Smeenk, P. Schön, M. B. J. Otten, S. Speller, H. G. Stunnenberg, J. C. M. van Hest "Fibril formation by triblock copolymers of silk-like beta-sheet polypeptides and poly(ethylene glycol)" *Macromolecules* **39,** 2989 (2006).

Curriculum Vitae

Jurgen Smeenk werd geboren op 4 april 1976 te Winterswijk. Na het behalen van het VWO diploma aan de R.K. Scholengemeenschap Marianum in Groenlo, begon hij in 1994 met de studie Bioprocestechnologie aan de Wageningen Universiteit. Met de specialisatie "Toegepaste biokatalyse" werd afstudeerwerk verricht in de vakgroep Proceskunde (prof. K. van 't Riet) met als titel "Integratie van oligosacharidenverwijdering d.m.v. adsorptie aan actief kool en enzymatische oligosacharidensynthese". Daarna werd in de vakgroep Biotechnologie (prof. P. Adlercreutz, Universiteit van Lund, Zweden) een afstudeervak gedaan met als titel "Porcine liver esterase in organic solvents", waarin het gebruik van enzymen in niet-waterig milieu werd bestudeerd. In maart 2000 behaalde hij het doctoraal examen. Van maart 2000 tot mei 2004 was hij aangesteld als AiO in de vakgroep Organische Chemie van de Radboud Universiteit Nijmegen. Onder leiding van Prof. J.C.M. van Hest en Prof. H.G. Stunnenberg (Moleculaire Biologie) werd gewerkt aan het in dit proefschrift beschreven onderzoek. Daarna werkte hij als Post-doc in het polymeerchemie laboratorium van Prof. O. Nuyken aan de Technische Universiteit München.