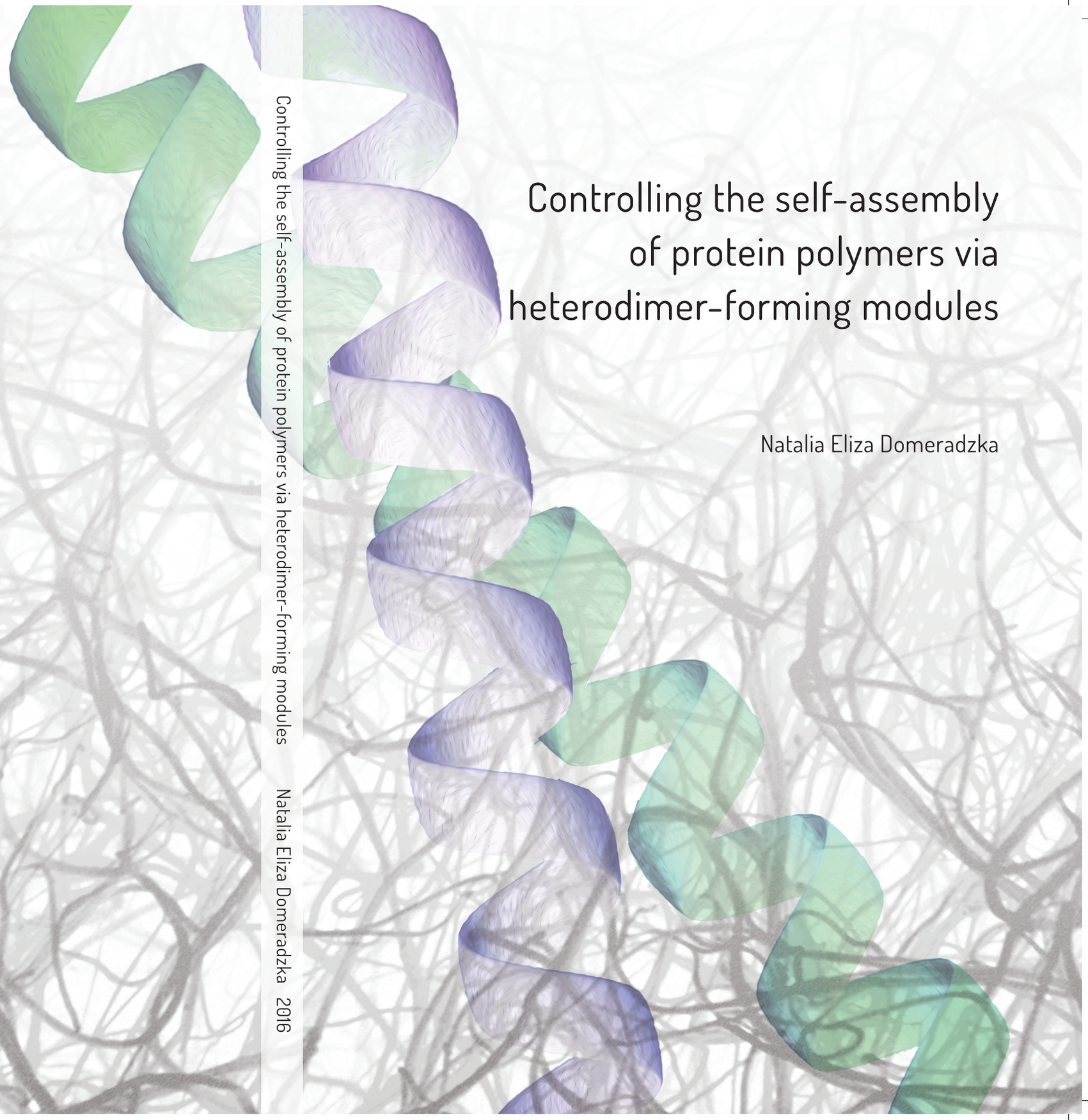
Controlling the self-assembly of protein polymers via heterodimer-forming modules

Natalia Eliza Domeradzka



Controlling the self-assembly of protein polymers via heterodimer-forming modules Natalia Eliza Domeradzka 2016

Controlling the self-assembly of protein polymers via heterodimer-forming modules

Natalia Eliza Domeradzka

Thesis committee

Promotor

Prof. Dr F. A. M. Leermakers
Personal chair at Physical Chemistry and Soft Matter
Wageningen University

Co-Promotors

Dr R. J. de Vries
Associate professor, Physical Chemistry and Soft Matter
Wageningen University

Dr F. A. de Wolf
Senior scientist
Wageningen Food and Biobased Research

Other members

Prof. Dr M. H. M. Eppink, Wageningen University
Prof. Dr G. Eggink, Wageningen University
Prof. Dr T. Scheibel, University of Bayreuth, Germany
Prof. Dr J. C. M. van Hest, Radboud University, Nijmegen

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Controlling the self-assembly of protein polymers via heterodimer-forming modules

Natalia Eliza Domeradzka

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A. P. J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 12 October 2016
at 4 p.m. in the Aula.

Natalia Eliza Domeradzka
Controlling the self-assembly of protein polymers via heterodimer-forming modules
174 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)
with references, with summary in English

ISBN: 978-94-6257-866-1
DOI: 10.18174/386360

Contents

1 General Introduction

1.1. Molecular self-assembly as a fabrication tool for nanomaterials

Nanotechnology is a field of science that deals with the engineering and manufacturing of structures with sizes in the 1-100 nm range in at least one dimension [1]. It is believed that nanotechnology will play a key role in the construction of future generations of materials required for solving problems in fields such as water, energy, and health [2]. Currently, high-tech nanostructured materials can already be found in many applications: in high performance electronics materials, as biomaterials, in sensors, water filters, fuel cells and batteries, medicine, and many consumer products [3,4].

Important concepts of material fabrication used in nanotechnology are the "top-down" and "bottom-up" approaches. The top-down fabrication is a subtractive process in which material undergoes a particular treatment involving use of, for instance, electron beams, radiation, masking or etching, to produce features of a controlled shape [5,6]. The "bottom-up" fabrication is an additive process in which components (particles or bigger molecules) are used to build up desired nanostructures [7]. Nowadays, the "bottom-up" approach becomes more and more employed in the construction of modern materials [5,7]. These approach largely relies on molecular self-assembly (Fig. 1.1), i.e. the spontaneous organisation of molecules (or particles) into structurally well-defined arrangements through numerous weak noncovalent bonds such as hydrogen bonds, ionic bonds, π-π stacking, hydrophobic interactions and van der Waals' interactions [6].

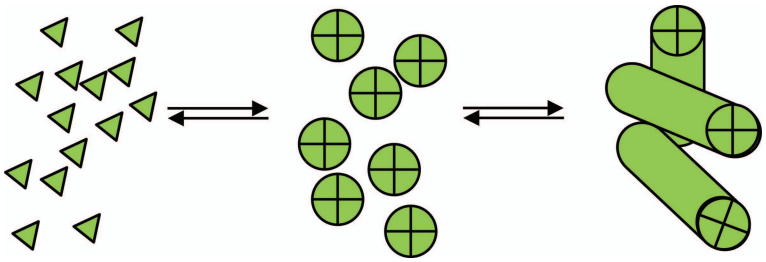


Figure 1.1: A *schematic representation of molecular self-assembly, where green complementary units form self-assemblies that organise into more complex macrostructures.*

Although association constants for single noncovalent bonds are rather low, collectively, many of these weak bonds can guarantee well-defined and very stable macroscopic structures, called supramolecular structures [8]. Many inspiring examples of self-assembled supramolecular structures formed by

noncovalent bonds can be found in nature: the DNA double helix, phospholipid cell membranes, large protein machines such as ribosomes, and complex networks of structural proteins such as collagen and elastin, to mention just a few. Indeed, nature exclusively uses "bottom-up" approach, building up molecular assemblies bit by bit [9]. It is important to mention that the structural complexity of these assemblies is strongly associated with their biological function. The sophistication and success of the natural "bottom-up" nanofabrication process is a great inspiration for material scientists, and although not yet fully understood, it serves as source of ideas for the fabrication of novel supramolecular nanostructured materials with complex functionalities [10-12].

The formation of supramolecular materials is usually studied in liquid phases, since the components must be able to move with respect to one another in order to assemble [10]. The characteristics of individual components such as shape, surface properties, charge, polarizability and mass are very important, since these encode the weak interactions that lead to the formation of supramolecular structures. Any component has to be chemically complementary and structurally compatible with the rest to assure a proper formation of physical bonds. The type of self-assembly and the final properties of the self-assembled materials can be controlled by chemically tailoring the different components [13]. Additionally, the interactions between these components can be manipulated by controlling the environment in which they interact. By testing and mixing different self-assembling components, scientists have discovered many novel self-assembled structures with various functionalities [9].

In this thesis we are concerned with a particular class of self-assembling molecules, viz. proteins. Over the last decade, significant progress has been made in the development of self-assembled, nanostructured protein materials [14-16]. Building on this progress, in this thesis we explore the biotechnological production and use of specific organizing protein elements: **small heterodimer-forming protein modules that can be used to organise supramolecular structures**. To introduce this topic, we first briefly review the structure of peptides and proteins, and then review the basis of their self-assembly into nanomaterials.

1.2. Self-assembling protein polymers

Proteins consist of amino acid units connected by peptide bonds into long chains. There are twenty DNA-encoded amino acids, which means that, for instance, for a 100-residue amino acid chain there are 20^{100} different possible

sequences. In addition, there is a wide range of possible post-translational modifications. The variation of amino acid side-chains is very important for the protein folding, since interactions are partly determined by the nature of the side-chains. Aliphatic groups give rise to a locally hydrophobic environment, and aromatic groups allow for π-π stacking. Hydrophilic groups allow for hydrogen bond formation, and charged groups can be involved in ionic interactions. The spontaneous process of protein folding and protein assembly is best described as occurring at different levels. A first step is the formation of secondary structure elements (α-helix, β-sheet, random coil). Next, secondary structure elements combine into a three-dimensional tertiary structure. Finally, the quaternary structure is the arrangement of multiple proteins into complexes, as determined by a range of weak interactions such as salt bridges, hydrogen bonding, and disulfide bonds. This hierarchical process of the protein folding can be considered as a reflection of the evolution process. Nature evolved dozens of small protein motifs with specific functions that subsequently, stimulated by environment, started to self-assemble into bigger and bigger structures with combined functionalities. Indeed, next to the amino acid sequence, nature assigned an important role in the protein folding also to the cell environment, which consists of many elements [17]. Due to that, it is extremely difficult to predict the *in vitro* folding behaviour, however by observing nature we can try to mimic its successful solutions.

Various natural protein motifs have been analysed in great detail and knowledge generated from these studies is very useful in the fields of the *de novo* protein design [15,18]. An especially fruitful area has been that of self-assembling peptides, where thanks to recent advances in peptide synthesis, many different sequences have been explored with respect to their capabilities to self-assemble into functional nanostructured materials. Self-assembled peptide- and protein materials are now raising the interest of material scientists in a very broad range of application areas, including for example, food, medicine or electronics [19-22].

Rather than with the self-assembly of chemically synthesized peptides, in this thesis we are concerned with the self-assembly of much larger polypeptides. Specifically, we are concerned with so-called protein-based block copolymers [22], that we will also refer to as protein polymers. In brief the idea is to use simple amino acid motifs as building blocks for repetitive polypeptides, or protein-based polymers. Two or more types of protein-based polymers can be combined into fusions, thus obtaining protein-based block-copolymers (Fig. 1.2). By incorporating blocks that can guide self-assembly process between

these copolymers, several supramolecular structures can be formed. Examples of such structures formed by protein polymers in aqueous solution are micelles, fibrils, ribbons, sheets, and even more complex networks, called hydrogels [23]. Supramolecular structures can be very useful especially as biological materials for 2D and 3D tissue cultures, regenerative and reparative medicine, tissue engineering, as well as therapeutic delivery and medical imaging [20,24-26]. Protein-based hydrogels are of particular interest in the medical industry because of their great resemblance to the extracellular matrix (ECM) [27,28]. Some prototypical examples of protein blocks developed to mimic ECM are silk-like, resilin-like, collagen-like, elastin-like, coiled coils and various peptide amphiphiles [22]. Hydrogels made of different protein polymers, obviously, have different rheological, structural and biological properties [29]. A great advantage is that by varying the length or composition of the different blocks in the protein polymer these properties can be tuned very precisely (at the amino acid level) to optimize them for a given application. Since protein blocks are derived from natural sources (i.e. amino acids), such hydrogels are typically biodegradable and biocompatible, which can be also considered as very important advantages [14].

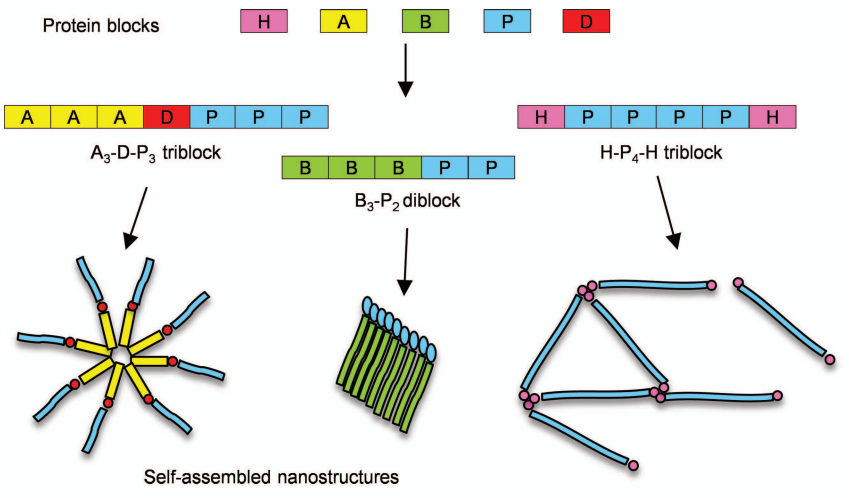


Figure 1.2: *Different protein blocks can be combined in an arbitrary fashion into protein block copolymers. Each block has a specific function, which influences the self-assembly of the protein block copolymers: (H) triple helix formation, (A) nonpolar, (B) β-sheet formation, (P) polar, (D) degradation.*

1.3. Biosynthesis of protein polymers

In the past, many protein polymer sequences have been designed by using motifs obtained from natural protein materials [30,31]. Now that relations between amino acid sequence, secondary structure and self-assembly are becoming more and more clear, especially at the level of short peptides, *de novo* design of self-assembling protein polymers becomes an increasingly realistic possibility [15]. In order to actually obtain protein polymers, one has to use genetic engineering since the chemical synthesis of very long polypeptides is still technically challenging and very expensive [22]. The genetic engineering approach allows the production of arbitrary sequences and additionally, in principle, offers absolute control over their molecular weight and composition [24]. In this strategy, the protein polymer sequence to be produced is first reversely translated into a DNA sequence. Chemically synthesised DNA sequences encoding different pieces of the protein polymer can be combined into single artificial protein polymer genes. With help of the biotechnology toolbox such genes can be transferred into any chosen expression system. Popular expression systems include for example, the bacterium *Escherichia coli* and the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* [32]. Each of these systems has its own advantages and disadvantages. We focus on heterologous expression in *P. pastoris*. *P. Pastoris* based expression system is able to produce proteins at high yields, can grow to very high cell densities and does not need an expensive carbon sources for growth. Additionally, *P. pastoris* as a protein production host is able to secrete high titres of properly folded, post-translationally processed and active recombinant proteins into the culture media.

The process of protein production in *P. pastoris* is rather straightforward [33]. The gene encoding protein of interest is placed in to an expression vector that allows integration of this gene with the chromosomal DNA of *Pichia* cell via homologous recombination. The expression vector chosen for this study is pPIC9. This vector is easy to handle and guarantees a high percentage of correct transformants. For the secretion of the protein polymers to the medium, α-MF signal peptide from *S. cerevisiae* was chosen. The α-MF signal sequence has proven to be most effective in directing proteins through the secretory pathway in *P. pastoris*. The ability to secret protein outside the host cell is a big advantage especially from the downstream processing point of view. To control protein synthesis, we use the alcohol oxidase (AOX1) promoter, which is tightly regulated by presence of methanol in the medium. This allows partial uncoupling of the yeast growth from the production phase of heterologous

proteins. Thanks to AOX1, a considerable amount of biomass is accumulated prior to protein overexpression.

1.4. Blocks and sequence motifs

In our group, we have developed a diverse collection of protein polymer blocks that can be produced in *P. pastoris* at high yield, to be discussed in more detail below. Each of these blocks presents a different functionality and most of them are stimulus-responsive. As described above, these protein polymer blocks can be freely combined in diblocks, triblocks, and etc. Furthermore, they can also be easily extended with biofunctional blocks such as comprising RGD and other cell-recognition motifs, enzymatic degradation sides, blocks that selectively bind certain biological compounds or protein motifs, etc. [34]. In this thesis we extend our list of protein block copolymer designs by incorporation of terminal blocks that specifically interact with each other to form (hetero)-dimers, with the purpose to obtain additional control over self-assembly.

1.4.1. Collagen-inspired blocks

The structural proteins that belong to collagen family are the most abundant components of the extracellular matrix. Many distinct collagen types have been identified to date on the basis of protein and/or DNA sequence [35]. The characteristic feature of collagens is sequence that contains domains with repetitions of the proline-rich tripeptide Gly-X-Y involved in the formation of trimeric collagen triplehelices [36]. In this tripeptide the position X is often occupied by proline and Y is frequently hydroxyproline (which is a product of post-translational modification performed by the enzyme prolyl 4-hydroxylase) [37]. Collagen provides structural support and tensile strength to various tissues of the human body; therefore it has proven itself to be a very versatile material for many medical applications [38]. The common form of collagen in use is gelatine. Gelatine is a denatured collagen typically isolated from bovine or porcine skin or bone by acid or base extraction [35]. There are increasing concerns with the continued use of animal-derived collagens and gelatines such as biocompatibility issues, the ability to transmit pathogenic vectors or lack of product homogeneity, to name a few [39]. In order to avoid issues related to animal-derived material and to be able to produce collagens in large quantities, the collagen production is increasingly shifting towards recombinant technologies [38]. Genetic engineering has made great progress in the areas of recombinant collagen and gelatine expression, and now these materials can be even custom-designed to enhance product performance [38].

In our group, *P. pastoris* has been utilized to produce several custom-designed collagen-inspired blocks [40-42]. Two of these blocks, namely **C** block and $\mathbf{T_9}$ block, are used in this thesis (Fig. 1.3). As first described by Werten et al [41], the **C** block is a 99 amino acids long sequence consisting of Gly-X-Y repeats, where proline is the second after glycine most abundant amino acid. The **C** block was developed in two variants, $\mathbf{C^P}$ and $\mathbf{C^R}$, which have the same molecular weight and amino acid composition, but variant $\mathbf{C^R}$ is a randomized version of $\mathbf{C^P}$ [42]. Both variants are highly hydrophilic and assume a random coil conformation, irrespective of pH and temperature. In this thesis, we used only $\mathbf{C^P}$ variant and to simplify the nomenclature, in some chapters, the superscript **P** was omitted. The **C** block has been successfully used to construct many different protein polymers [42-44]. The number in subscript indicates how many blocks of the same type are repeated in tandem within the entire protein polymer. In most cases, the **C** block serves mainly as a stabilizing spacer, however its nature also allows easy protein purification from the fermentation broth by ammonium sulfate precipitation [45]. The $\mathbf{T_9}$ block consists of nine amino acid triplets Pro-Gly-Pro and, incorporated as a terminal block, is able to form helical homotrimers. The process of $\mathbf{T_9}$ block self-assembly is dependent on temperature and concentration [46]. In addition, the type of neighbouring blocks influences the concentration at which $\mathbf{T_9}$ triple helixes can be formed [47].

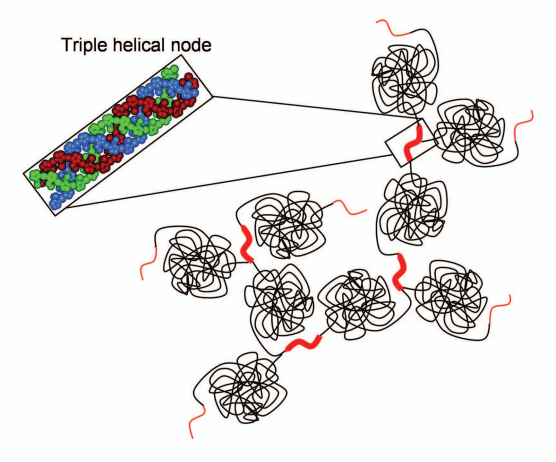


Figure 1.3: *Schematic representation of the hydrogel formed by* $\boldsymbol{T_9\text{-}C_4\text{-}T_9}$ *protein polymer. Triple helixes formed by* $\boldsymbol{T_9}$ *block facilitate self-assembly, while* $\boldsymbol{C_4}$ *block acts as hydrophilic spacer.*

1.4.2. Silk-inspired blocks

Silks are natural protein fibers produced by spiders of the Arachnida class and several worms of the order Lepidoptera [48]. The most popular silk, which is widely used for millennia in the textile industry, is obtained from the cocoons of *Bombyx mori*. Typically, silks fibers are composites of two main proteins, sericin and fibroin, and other associated macromolecules, i.e. polysaccharides and lipids [49]. Relatively conserved fibroin sequences consist of big hydrophobic domains with highly repetitive motifs of Ala, Ala-Gly or Ala-Gly-Ser, which are interrupted by small hydrophilic domains [50]. The hydrophobic domains can form tightly packed anti-parallel β-sheets and are responsible for the remarkable strength and resiliency of silk fibers [30,51] (Fig 1.4A). These impressive mechanical properties of silks as well as their biocompatibility and biodegradability have inspired the preparation of silk-mimetic proteins [52-54].

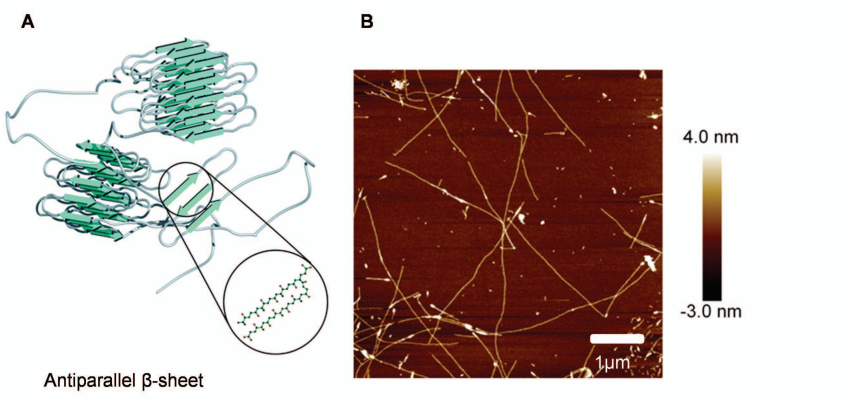


Figure 1.4: *(A) Schematic representation of fibrils formation via β-sheets stacking by silk-inspired protein polymers. (B) An image of fibrils formed by* C_2-S^H_{48}-C_2 *protein polymer obtained by atomic force microscopy.*

A sequence consisting of repeating $(Gly\text{-}Ala)_3Gly\text{-}Glu$ octapeptides [52], denoted as $\mathbf{S^E}$ block, was a first silk-inspired product obtained in *P. pastoris* in our laboratory [55]. In the $\mathbf{S^E}$ block, the hydrophobic fragments consisting of Gly-Ala repeats facilitate the process of self-assembly, as it is in the natural silks, while the presence of glutamic acid imparts self-assembly at low pH values [55] (Fig 1.4B). The other variant of **S** block frequently used in our protein polymers is $\mathbf{S^H}$ block, where glutamic acid was replaced by histidine. In the $\mathbf{S^H}$ block, histidine at neutral and high pH values becomes uncharged, which allows the formation of anti-parallel β-sheets and protein self-assembly [56]. Usually, the **S**

block is fused to other blocks in form of multimers, for instance, $\mathbf{C_2\text{-}S^H{}_8\text{-}C_2}$, $\mathbf{C_2\text{-}S^H{}_{24}\text{-}C_2}$, $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ [57]. In this thesis we used one of the protein polymers that consist $\mathbf{S^H{}_{48}}$, namely $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$, which can form fibrillar hydrogels at neutral pH values. It has been reported that $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ hydrogels can reassemble after rupture, which is called self-healing [43]. Moreover, as shown by Biegun-Wlodarczyk et al., they constitute an attractive material for medical applications [34].

1.4.3. Coiled coils motifs

Predictions based on analyses of primary sequences indicate that approximately 10% of all protein amino acid sequences in eukaryotic cells form coiled coils [58]. There are many functions assigned to different coiled coils, for instance, oligomerization, regulatory function, DNA binding, membrane fusion, or energy transfer [59]. The arrangements of coiled coils are also very diverse. They can consist of two to seven parallel or antiparallel amphipathic α-helices that are wrapped around each other in superhelical fashion [60] (Fig. 1.5A).

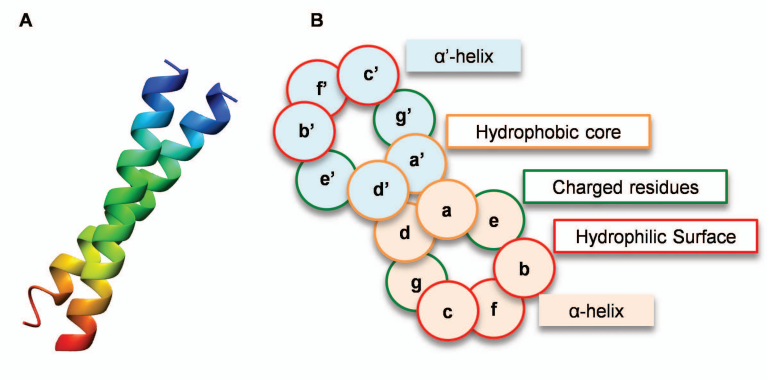


Figure 1.5: *(A) 3-D structure of antiparallel coiled-coil determined by X-ray crystallography. (B) Helical wheel diagram showing interacting residues in a coiled coil.*

Moreover, they can be either left-handed or right-handed [61]. Here, we discuss only small, left-handed and parallel coiled coils, called leucine zippers [62]. This name is derived from leucine that occurs periodically every seven residues in the sequence. The sequence consists of a heptad repeat pattern $(\mathbf{abcdefg})_n$, where n is the number of repeats (Fig. 1.5B). Positions **a** and **d** are typically occupied by hydrophobic amino acids that form the hydrophobic core of the coiled coil. The amino acids located at positions **e** and **g** often contain complementary charged side chain functional groups, which stabilize the

strands through salt bridges. Recently, many *de novo* designed coiled coils have been reported that show selective and specific binding properties [63]. Due to that coiled coils have stimulated the interest in the use of these motifs for directed self-assembly [61].

In this thesis we aimed to test use of two pairs of heterodimer-forming leucine zippers to direct self-assembly of chosen protein polymers. The first pair was developed by Moll et al. to target a radionuclide to tumor-specific antibodies [64]. The complementary units are 47 amino acids long and form heterodimers with the melting temperature of 74 °C, dissociation constant of 1.3×10^{-11} M and molecular weight of 11 kDa (See Table 3.1, Chapter 3 for amino acid sequence). The second pair was developed as a universal peptide capture and delivery system [65]. The complementary units are 21 amino acids long and form heterodimers with the free energy of unfolding equal 9.6 $kcal\times mol^{-1}$, dissociation constant of 7×10^{-8} M and molecular weight of 4 kDa (See Table 4.1, Chapter 4 for amino acid sequence). Both couples have been shown to form heterodimers in very selective and specific manner [64,65]. Although Dooling et al. reported use of coiled coils as midblocks [66], here we tested only terminal orientation.

1.4.4. WW domain motifs

WW domains can be found in various natural proteins. With 35–40 amino acid residues, WW domains are among the smallest independently folding domains [67]. Their sequence has a high content of aromatic, hydrophobic and proline residues, and two highly conserved tryptophans (hence WW domain). In aqueous solutions, the domain assumes a slightly bent three-stranded antiparallel β-sheet, the concave side of which forms a binding pocket for the ligand [68] (Fig. 1.6A). The ligands that WW domains recognise and bind in a specific manner are small proline-rich peptides with the core motif PPXY [69,70]. It has been shown that the interaction between WW domain and a proline-rich peptide can be successfully employed to form supramolecular structures [71] (Fig. 1.6B).

In this thesis, we explore two types of WW domains and two proline-rich motifs as an alternative to coiled coils described above. The WW domains that we have chosen are WWP1-1 and CC43. The first one is derived from the human ubiquitin ligase homolog WWP1 [73] and the second one was designed with the aid of computer modelling, by Russ et al. [74]. As proline-rich ligands we analysed a PPxY peptide derived from the p53-binding protein-2 [73] and PPxY-1, derived from p45 murine activation domain [75]. According to Russ et

al., the dissociation constant for CC43 and PPxY heterodimer is approximately 1.7×10^{-5} M [74]. The combination of PPxY and WWP1-1 was among the best performing pairs tested by Pirozzi et al. [73], but to our knowledge no dissociation constant values have been published. The possibility to use WW domains and their PPxY ligands as parts of longer constructs (two component hydrogels) has been described previously [71]. Both the WW and PPxY modules can be used in the middle of a sequence or at its end, in this thesis we have explored using them as terminal blocks in protein-polymer designs.

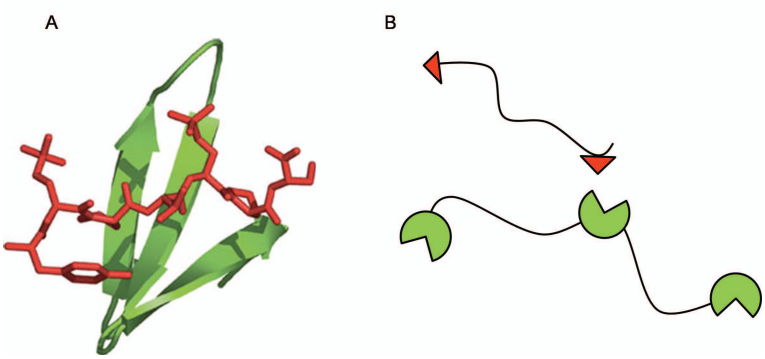


Figure 1.6: *(A) Schematic representation of WW domain [72] bound to its proline-rich ligand (red) . (B) An example of WW domain and its ligand incorporated in the protein polymer chain (black) .*

1.5. Aim and organisation of the thesis

We use the incorporation of heterodimer-forming modules to obtain additional control over protein polymer self-assembly. Specifically, this thesis is aimed at the design, biosynthetic production and characterization of protein polymers with complementary terminal blocks A and B, where A and B are small, A-B heterodimer forming domains. The idea is that the A and B modules act as linkers, physically coupling the two different protein polymers to which the A and B domains are attached (Fig. 1.7). The ability to precisely control the process in which protein polymers are assembled into supramolecular structures is very important for their final applications.

In the next chapter, **Chapter 2,** we review various protein-engineering approaches that have been used to obtain control over the assembly of protein polymers. We discuss techniques that allow for the design of complex, hierarchical polypeptide self-assemblies and hybrid materials, and techniques to cross-link and stabilize weak supramolecular structures, or to incorporate into these structures, specific bio-functional groups via self-assembly.

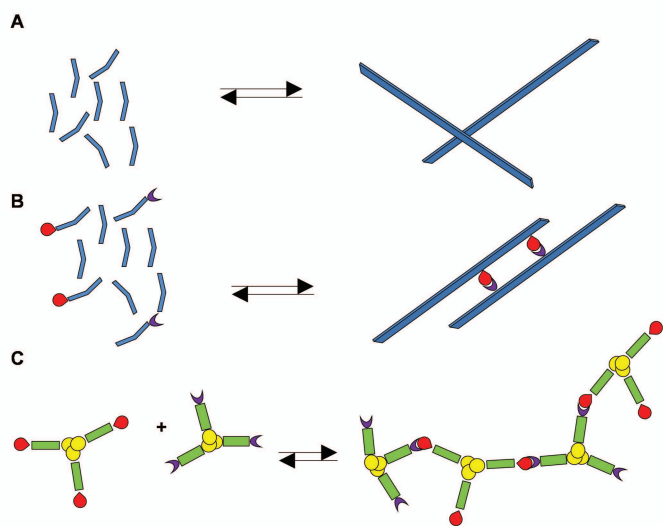


Figure 1.7: *The exemplary application of A/B heterodimer-forming modules (red/purple) for directed self-assembly of protein polymers. A/B can be incorporated in to the fibril-forming protein polymer (blue) (A) in order to organise fibrils into bundles (B), or they can be used for mixing-induced network formation, where block in yellow can form homotrimers and blocks in green serve as spacers (C).*

Next, in **Chapter 3**, **4** and **5** we study the capability of three different heterodimer-forming A-B peptide couples to act as linkers. These peptides were attached either to previously developed $\mathbf{C^P_4}$ protein polymer or to its modified variant called $\mathbf{T_9\text{-}C^P_4}$. Obviously, a mixture of $\mathbf{T_9\text{-}C^P_4\text{-}A}$ with $\mathbf{T_9\text{-}C^P_4\text{-}B}$ should also form a network, if A and B together form A-B heterodimers, and thus, we chose to construct $\mathbf{T_9\text{-}C^P_4\text{-}A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}B}$. For comparison, $\mathbf{A\text{-}C^P_4\text{-}A}$ and $\mathbf{B\text{-}C^P_4\text{-}B}$, which should form long chains upon mixing, and the even simpler $\mathbf{C^P_4\text{-}A}$ and $\mathbf{C^P_4\text{-}B}$, which should form dimers, were also constructed. All the protein polymers discussed in this thesis were designed via genetic engineering and produced via fermentation process in yeast *P. pastoris*.

In **Chapter 3** we study production of protein polymers with incorporated heterodimer-forming coiled coils. We present the strategy allowing significant reduction of product proteolysis during fermentation. Furthermore, we investigate self-assembling properties of di- and triblock protein polymers. We show that coiled coils attached to protein polymers, are functional and can couple, otherwise non-interacting protein polymers. Here we also highlight that this type of linkers can form homo-multimers, which can be avoided by introduction of certain concentration regime.

In **Chapter 4** another type of coiled coils incorporated into di- and triblocks is investigated. Here, we report the case of proteolytic product degradation and several approaches in order to obtain intact product. We show that these coiled coils can be used to form protein polymer network, however we could not exclude that homodimers also participate in the network formation.

In **Chapter 5** we discuss production and characterization of diblocks containing heterodimer-forming WW and proline-rich domains. We also analyse a case of *Picha*-derived glycosylation of polymer variant with proline-rich domain and we show the steps undertaken to escape glycosylation. In this study, we confirm that WW and proline-rich domains incorporated into protein polymers are functional and do not form homo-assemblies. Additionally, in the appendix to this chapter, we discuss the study on production and primary characterization of triblocks.

In **Chapter 6**, we study the heterodimer-forming couple, described in **Chapter 3,** built into silk-like protein polymer that can, already without the A/B linkers self-organize into long, pH-responsive nanofibrils. First we analyse the impact of pH on the linkers incorporated into the fibrils. Furthermore, we investigate whether, by mixing A/B polymer variants with linker-free polymers, we can control the level of interaction between fibrils. Since at the higher concentrations silk-like protein polymers can form hydrogel, we also explore the impact of introducing these linkers on the mechanical properties of such hydrogels.

Finally, in **Chapter 7,** we provide a general discussion of the results obtained in this thesis. We summarize our results, recommend further steps, place our findings into the broader context of possible applications, and analyse possible risk that such application may entail.

References

1. National Research, C.; *The National Academies Press: Washington, DC*, **2013**.
2. Ulijn, R. V. *Nat Nano* **2015**, 10, 295-296.
3. Hirst, A.; Escuder, B.; Miravet, J.; Smith, D. *Angew Chem Int* **2008**, 47, 8002-8018.
4. Lakshmanan, A.; Zhang, S.; Hauser, C. A. E. *Trends Biotechnol* **2012**, 30, 155-165.
5. Silva, G. A. *Surg Neurol* **2004**, 61, 216-220.
6. Whitesides, G. M.; Boncheva, M. *PNAS* **2002**, 99, 4769-4774.
7. Shimomura, M.; Sawadaishi, T. *Curr Opin Colloid Interface Sci* **2001**, 6, 11-16.
8. Zhang, S. *Nat Biotech* **2003**, 21, 1171-1178.
9. Philp, D.; Stoddart, J. F. *Angew Chem Int* **1996**, 35, 1154-1196.
10. Whitesides, G. M.; Grzybowski, B. *Science* **2002**, 295, 2418-2421.
11. Boal, A. K.; Ilhan, F.; DeRouchey, J. E.; Thurn-Albrecht, T.; Russell, T. P.; Rotello, V. M. *Nature* **2000**, 404, 746-748.
12. Zhang, S.; Marini, D. M.; Hwang, W.; Santoso, S. *Curr Opin Chem Biol* **2002**, 6, 865-871.
13. Ward, M. D.; Raithby, P. R. *Chem Soc Rev* **2013**, 42, 1619-1636.
14. Ulijn, R. V.; Smith, A. M. *Chem Soc Rev* **2008**, 37, 664-675.
15. Yanlian, Y.; Ulung, K.; Xiumei, W.; Horii, A.; Yokoi, H.; Zang, S. *Nano Today* **2009**, 4, 193-210.
16. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.*; et al. Science* **2013**, 340, 595-599.
17. Dobson, C. M. *Nature* **2003**, 426, 884-890.
18. Tiwari, M. K.; Singh, R.; Singh, R. K.; Kim, I.-W.; Lee, J.-K. *Comput Struct Biotechnol J* **2012**, 2, e201209002.
19. Rajagopal, K.; Schneider, J. P. *Curr Opin Struct Biol* **2004**, 14, 480-486.
20. Kyle, S.; Aggeli, A.; Ingham, E.; McPherson, M. J. *Biomaterials* **2010**, 31, 9395-9405.
21. Ardoña, H. A. M.; Tovar, J. D. *Bioconjugate Chem* **2015**, 26, 2290-2302.
22. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2011**, 12, 269-289.

23. Loo, Y.; Zhang, S.; Hauser, C. A. E. *Biotechnol Adv* **2012**, 30, 593-603.
24. Ghandehari, H.; Cappello, J. *Pharm Res* **1998**, 15, 813-815.
25. Liu, X.; Pi, B.; Wang, H.; Wang, X.-M. *Front Mater Sci* **2014**, 9, 1-13.
26. Ferrer-Miralles, N.; Rodríguez-Carmona, E.; Corchero, J. L.; García-Fruitós, E.; Vázquez, E.; Villaverde, A. *Crit Rev Biotechnol* **2015**, 35, 209-221.
27. Woolfson, D. N.; Mahmoud, Z. N. *Chem Soc Rev* **2010**, 39, 3464-3479.
28. Romano, N. H.; Sengupta, D.; Chung, C.; Heilshorn, S. C. *Biochim Biophys Acta, Gen Subj* **2011**, 1810, 339-349.
29. Sathaye, S.; Mbi, A.; Sonmez, C.; Chen, Y.; Blair, D. L.; Schneider, J. P.; Pochan, D. J. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2015**, 7, 34-68.
30. Wang, Y.; Kim, H.-J.; Vunjak-Novakovic, G.; Kaplan, D. L. *Biomaterials* **2006**, 27, 6064-6082.
31. Maude, S.; Ingham, E.; Aggeli, A. *Nanomedicine* **2013**, 8, 823-847.
32. Yokoyama, S. *Curr Opin Chem Biol* **2003**, 7, 39-43.
33. Macauley-Patrick, S.; Fazenda, M. L.; McNeil, B.; Harvey, L. M. *Yeast* **2005**, 22, 249-270.
34. Włodarczyk-Biegun, M. K.; Werten, M. W. T.; de Wolf, F. A.; van den Beucken, J. J. J. P.; Leeuwenburgh, S. C. G.; Kamperman, M.; Cohen Stuart, M. A. *Acta Biomater* **2014**, 10, 3620-3629.
35. Ricard-Blum, S.; Ruggiero, F. *Pathol Biol* **2005**, 53, 430-442.
36. Gelse, K.; Pöschl, E.; Aigner, T. *Adv Drug Delivery Rev* **2003**, 55, 1531-1546.
37. de Bruin, E. C.; Werten, M. W. T.; Laane, C.; de Wolf, F. A. *FEMS Yeast Res* **2002**, 1, 291-298.
38. Browne, S.; Zeugolis, D. I.; Pandit, A. *Tissue Eng Part A* **2012**, 19, 1491-1494.
39. Olsen, D.; Yang, C.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perälä, M.; Hämäläinen, E.-R.; Jarvinen, M.*; et al. Adv Drug Delivery Rev* **2003**, 55, 1547-1567.
40. Werten, M. W T.; de Wolf, F. A. *Yeast* **1999**, 15, 1087-1096.
41. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng* **2001**, 14, 447-454.
42. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009**, 10, 1106-1113.

43. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W. T.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 699-706.
44. Beun, L. H. PhD thesis, Wageningen University, Wageningen, 2015.
45. Moers, A. P. H. A.; Wolbert, E. J. H.; de Wolf, F. A.; Werten, M. W. T. *J Biotechnol* **2010**, 146, 66-73.
46. Skrzeszewska, P. J.; de Wolf, F. A.; Werten, M. W. T.; Moers, A. P. H. A.; Cohen Stuart, M. A.; van der Gucht, J. *Soft Matter* **2009**, 5, 2057-2062.
47. Teles, H.; Skrzeszewska, P. J.; Werten, M. W. T.; van der Gucht, J.; Eggink, G.; de Wolf, F. A. *Soft Matter* **2010**, 6, 4681-4687.
48. Vepari, C.; Kaplan, D. L. *Prog Polym Sci* **2007**, 32, 991-1007.
49. Hardy, J. G.; Scheibel, T. R. *Biochem Soc Trans* **2009**, 37, 677-681.
50. Kundu, B.; Rajkhowa, R.; Kundu, S. C.; Wang, X. *Adv Drug Delivery Rev* **2013**, 65, 457-470.
51. Wang, X.; Kim, H. J.; Wong, C.; Vepari, C.; Matsumoto, A.; Kaplan, D. L. *Mater Today* **2006**, 9, 44-53.
52. Krejchi, M.; Atkins, E.; Waddon, A.; Fournier, M.; Mason, T.; Tirrell, D. *Science* **1994**, 265, 1427-1432.
53. Kundu, B.; Kurland, N. E.; Bano, S.; Patra, C.; Engel, F. B.; Yadavalli, V. K.; Kundu, S. C. *Prog Polym Sci* **2014**, 39, 251-267.
54. Dinjaski, N.; Kaplan, D. L. *Curr Opin Biotech* **2016**, 39, 1-7.
55. Werten, M. W. T.; Moers, A. P. H. A.; Vong, T.; Zuilhof, H.; van Hest, J. C. M.; de Wolf, F. A. *Biomacromolecules* **2008**, 9, 1705-1711.
56. Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Cohen Stuart, M. A.; de Wolf, F. A. *Macromolecules* **2009**, 42, 1002-1009.
57. Beun, L. H.; Storm, I. M.; Werten, M. W. T.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 3349-3357.
58. Grigoryan, G.; Keating, A. E. *Curr Opin Struct Biol* **2008**, 18, 477-483.
59. Lupas, A. *Trends Biochem Sci* **1996**, 21, 375-382.
60. Burkhard, P.; Stetefeld, J.; Strelkov, S. V. *Trends Cell Biol* **2001**, 11, 82-88.
61. Parry, D. A.; Fraser, R. B.; Squire, J. M. *J Struct Biol* **2008**, 163, 258-269.
62. Alber, T. *Curr Opin Genet Dev* **1992**, 2, 205-210.
63. Woolfson, D. N. *Adv Protein Chem* **2005**, 70, 79-112.
64. Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci* **2001**, 10, 649-655.

65. Litowski, J. R.; Hodges, R. S. *J Biol Chem* **2002**, 277, 37272-37279.

66. Dooling, L. J.; Buck, M. E.; Zhang, W. B.; Tirrell, D. A. *Adv Mater* **2016**.

67. Dodson, E. J.; Fishbain-Yoskovitz, V.; Rotem-Bamberger, S.; Schueler-Furman, O. *Exp Biol Med* **2015**, 240, 351-360.

68. Macias, M. J.; Hyvönen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. *Nature* **1996**.

69. Macias, M. J.; Wiesner, S.; Sudol, M. *FEBS letters* **2002**, 513, 30-37.

70. Chen, H. I.; Sudol, M. *PNAS* **1995**, 92, 7819-7823.

71. Foo, C. T. W. P.; Lee, J. S.; Mulyasasmita, W.; Parisi-Amon, A.; Heilshorn, S. C. *PANS* **2009**, 106, 22067-22072.

72. Greenfield, N. J. *Methods Mol Biol* **2004**, 55-77.

73. Pirozzi, G.; McConnell, S. J.; Uveges, A. J.; Carter, J. M.; Sparks, A. B.; Kay, B. K.; Fowlkes, D. M. *J Biol Chem* **1997**, 272, 14611-14616.

74. Russ, W. P.; Lowery, D. M.; Mishra, P.; Yaffe, M. B.; Ranganathan, R. *Nature* **2005**, 437, 579-583.

75. Mosser, E. A.; Kasanov, J. D.; Forsberg, E. C.; Kay, B. K.; Ney, P. A.; Bresnick, E. H. *Biochemistry* **1998**, 37, 13686-13695.

2 Protein cross-linking tools for the construction of nanomaterials

Abstract

Across bioengineering there is a need to couple proteins to other proteins, or to peptides. Although traditional chemical conjugations have dominated in the past, more and more highly specific coupling strategies are becoming available that are based on protein engineering. Here we review the use of protein modification approaches such as enzymatic and autocatalytic protein-protein coupling, as well as the use of hetero-dimerizing (or hetero-oligomerizing) modules, applied to the specific case of linking together *de novo* designed recombinant polypeptides into precisely structured nanomaterials. Such polypeptides are increasingly being investigated for biomedical and other applications. In this chapter, we describe the protein-engineering based cross-linking strategies that dramatically expand the repertoire of possible molecular structures and, hence, the range of materials that can be produced from them.

Published as: Domeradzka, N.E..; Werten, M.W.T.; de Wolf, F.A; de Vries, R. *Curr Opin Biotechnol* **2016**, 39, 61-67.

2.1. Introduction

Self-assembling proteins are intensively being investigated as building blocks for nanostructured soft materials [1,2]. Sequence motifs from natural structural proteins can be used to design novel protein materials consisting of polypeptides with simple repetitive sequences ("protein-based polymers", or simply "protein polymers") [3,4], or one can design entirely novel peptide or polypeptide sequences with predictable self-assembly behaviour [5,6]. Self-assembling (poly)peptides can be either chemically synthesized or genetically engineered [7,8]. The latter route, which is the focus of this chapter, allows the production of long precisely defined amino acid sequences, thus offering a level of control that is highly attractive in a range of biomedical applications [8-14].

These applications depend on precise control over the way in which the polypeptides are assembled into their final supramolecular structure. It is here that novel protein-engineering techniques for cross-linking can play a crucial role. Although chemical cross-linking reactions can, and have been used extensively [15], they typically lack the specificity needed to create precisely defined polypeptide nanomaterials. Various protein-engineering tools such as enzymatic coupling [16-20] and the use of bio-recognition modules [21] do offer such specificity. By bio-recognition modules, we here specifically refer to the use of peptide- or small protein domains that can be incorporated in polypeptide chains in order to mediate self-assembly of polypeptides [22] via hetero-dimerization (or hetero-oligomerization). Although these engineering tools can be used to link specific bioactive groups to proteins [23-25], this chapter mainly focusses on their use in linking designed polypeptides into higher order nanostructures. An overview of the tools discussed is given in Table 2.1.

2.2. Covalent protein cross-linking

For permanently linking different protein subunits into larger assemblies, a range of cross-linking enzymes and autocatalytic protein modules are available. Protein cross-linking enzymes differ substantially in their degree of substrate specificity. First we consider enzymes with a rather broad substrate specificity. They are especially useful for the creation of irreversibly cross-linked (hydrogel) networks. Key examples are: Transglutaminase [26], Horse-radish Peroxidase (*HRP*), Lysyl Oxidase (*LOX*) and Plasma Amine Oxidase (*PAO*), which couple one specific type of amino acid to another, as illustrated in Fig. 2.1 (e.g. *TG* couples lysines to glutamines).

Table 2.1: *Cross-linking tools and their application to designer polypeptides*

Strategy	Cross-linking mechanism	Polypeptide	Application	Ref.
Transglutaminase	Formation of an isopeptide bond between Lys and Gln	· Q11 peptide · Collagen-like dendrimers · Elastin-like proteins · Designer protein polymers K_n and $(BQ)_n$	· Network functionalization · Two-component network · Cross-linking of fibers	32, 34-37
Lysyl oxidase and plasma amine oxidase	Spontaneous aldol condensation or formation of a Schiff base between two Lys residues	· $K_2(SL)_6K_2$	· Cross-linking of self-assemblies	40
Horse radish peroxidase	Spontaneous-radical couplings between two Tyr residues	· Silk proteins · Resilin-like polypeptides	· Cross-linking of self-assemblies	45, 46
Sortase A	Formation of threonine-glycine peptide bond between C-terminal LPETGG and N-terminal poly-G motifs	· P1 peptide · Micelles made of elastin-like proteins	· Network functionalization	51, 52
Split-inteins	Formation of a covalent bond by (1) non-covalent association of the halves of a split-intein pair, and (2) joining of the flanking protein fragments and concomitant self-excision by the reconstituted intein	· CutA trimer-forming peptide	· Two-component network	57
SpyTag-SpyCatcher	SpyTag and SpyCatcher undergo autocatalytic isopeptide bond formation between Asp_{117} on SpyTag and Lys_{31} on SpyCatcher	· Elastin-like proteins · CsgA protein	· New self-assembling topologies, two-component network · Enzyme immobilization	60-62
Heterodimer-forming coiled coils	Hydrophobic forces and electrostatic interactions between complementary units	· Hydrophilic random coli · Self-assembling polypeptides	· New self-assembling topologies · Two-component network; network functionalization	72-74
WW domain and its proline rich ligand	Selective binding of WW domain to specific tyrosine-containing sequences, based on a combination of interaction types	· Hydrophilic random coli	· New self-assembling topologies; two-component network; network functionalization	80-83
PDZ domain and its ligand	Strong protein-protein interaction based on hydrophobicity, reinforced by an engineered disulphide linkage	· CutA trimer-forming peptide	· Two-component network; network functionalization	86

Their substrate specificity is broad in the sense that these enzymes typically provide the engineer with considerable freedom to choose the residues flanking those to be cross-linked. A second group of enzymes and autocatalytic modules have a much more narrow substrate specificity: Sortase A (*SrtA*), Inteins, SpyTag/SpyCatcher, see Fig. 2.2. These tools are particularly useful for creating molecular conjugates with precisely defined architectures.

2.2.1. Transglutaminase

*TG*s catalyse the formation of isopeptide bonds between glutamine and lysine residues [27] and are found in both microbial [28] and mammalian cells [27]. They are quite promiscuous with respect to the residues flanking the lysine substrate in the polypeptide chain [29,30], but less so with respect to the preferred context of the glutamine [31,32]. For *de novo* designed polypeptides one can often exploit some freedom of design, such that *TG* can be a very useful tool to create well-defined polypeptide architectures. Collier and Messersmith showed that *TG* can be used to couple the glutamines in the self-assembling peptide Q11 (QQKFQFQFEQQ) to lysines in other peptides and proteins [33]. The same group proposed more *TG* substrate peptides, including an amine donor sequence K (EDGFFKI) that can be targeted by *TG* when placed at the C-terminus [34], and which can be coupled to the amine acceptor substrate peptide Q (APQQEA) [35]. The Chilkoti group studied elastin-like polypeptide (ELP) sequences, tandem repeats of the motif VPGXG(VPAGVG)$_6$, and showed that variants for which X is either lysine (K) or glutamine (Q), were efficiently cross-linked by tissue-derived *TG*s [36]. Bozzini et al. designed protein polymers with an (A-B)$_8$ structure, where A is an ELP sequence and B is the motif AAAAAAKAAAKAAQGFL. These protein polymers were efficiently cross-linked by microbial *TG* [37]. Finally, Davis et al. developed a protein hydrogel system that was cross-linkable by *TG* [38]. This system consists of two *de novo* designed protein polymers K_n and $(BQ)_n$, where the K_n block contains lysines, B is a random coil hydrophilic block, and the Q block serves as the glutamine substrate.

2.2.2. Lysyl oxidase and plasma amine oxidase

LOX are a family of oxido-deaminases that cross-link components of the extracellular matrix, in particular collagen and elastin [39]. *LOX* cross-linking is based on the oxidation of the primary amine of lysine to an aldehyde. This aldehyde can spontaneously react with another amine and form a Schiff base or undergo an aldol condensation with another aldehyde [40]. In this way a covalent

bond is established between two lysines. *PAO* is a commercially available alternative, which also functions by oxidation of primary amines. Bakota et al. showed that *LOX* and *PAO* can cross-link the self-assembled structures of the *de novo* designed peptide $K_2(SL)_6K_2$ [26]. It appears that the activity of *LOX* is less affected by sequence context than that of *TG* [19], although some preferences for the residues surrounding the lysines have been identified [41]. Because *LOX* is naturally present in serum [42], it may be employed for *in vivo* cross-linking of polypeptide nanomaterials.

2.2.3. Horse radish peroxidase

HRP is used for a range of different biochemical applications [43]. The enzyme catalyses redox reactions between hydrogen peroxide and a reducing substrate. Typical substrates include aromatic phenols, phenolic acids, indoles, amines and sulfonates. It can be used for protein cross-linking when tyrosines are available that are sufficiently accessible. Minamihata et al. showed that *HRP* can be used for site-specific cross-linking of recombinant proteins that contain a tyrosine tag (GGGGY or GGYYY) [44]. The Kaplan laboratory explored the use of *HRP* to generate highly elastic hydrogels by cross-linking tyrosines in natural silk proteins [45] and recombinant resilin-like proteins [46]. A potential problem is that prolonged cross-linking may lead to the formation of tri-tyrosines and higher order cross-links [47,48].

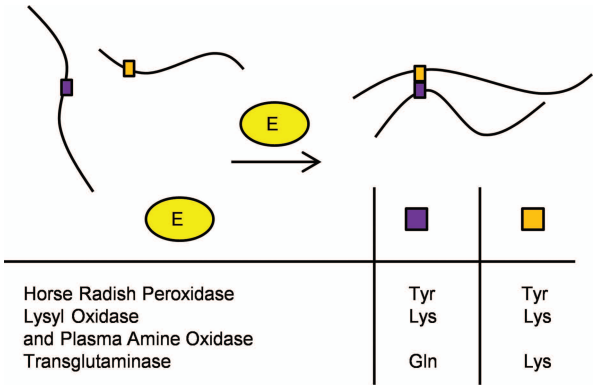


Figure 2.1: *Enzymatic cross-linking with broad specificity. Transglutaminase catalyses protein crosslinking through a transamination reaction between lysine and glutamine. Lysyl oxidase and plasma amine oxidase convert the amino group of lysines into aldehydes, which can form a Schiff base with the amino group of another lysine. Horseradish peroxidase mediates a redox reaction between hydrogen peroxide and tyrosine.*

2.2.4. Sortase A

Sortase A is a highly specific cross-linking enzyme derived from Gram-positive bacteria, and belongs to the group of transpeptidases. *In vitro*, *SrtA* catalyses the transpeptidation by cleaving between threonine and glycine in a C-terminal LPXTG recognition motif and subsequently joining the carboxyl group of threonine to the amino group of an N-terminal glycine of other molecule [49] (Fig. 2.2A). Many *de novo* designed peptides have been screened as potential *SrtA* substrates [50]. Piluso et al. used *SrtA* to functionalize a self-assembling ionic-responsive, peptide hydrogel with a DNA-binding protein (Tus) [51]. A diglycine sequence was incorporated at the N-terminal side of self-assembling peptide P1 (GGFEFEFKFKK) and the Tus protein featured an LPETGG sequence at its C-terminus. The conjugated Tus protein remained functional and accessible in the hydrogel. Similarly, Kim et al. functionalized ELP-based micelles with bioactive protein domains using sortase-mediated bioconjugation [52]. Relatively large proteins (40 kDa) were coupled to the micelles without loss of activity. Not much work has been reported yet where *SrtA* is used specifically to engineer novel self-assembling architectures, which may be due to the fact that *SrtA* conjugations are quite difficult to drive to complete conversion. Also, in some cases, side reactions can occur [53].

2.2.5. Inteins

Inteins are part of a larger class of proteins known as Hedgehog/intein (Hint) domains. An intein carries out a unique auto-processing event known as protein splicing, in which it excises itself from a larger precursor polypeptide through the cleavage of two peptide bonds and ligates the previously flanking protein sequences through the formation of a new peptide bond. Inteins have been exploited in the chemoenzymatic synthesis of N-terminally modified proteins [54], protein purification [55] and in protein biotinylation [56]. The so-called split-inteins [54] contain two separately encoded fragments that can assemble into a functional intein through non-covalent interactions. Such pairs both exist naturally, and have been created artificially by splitting a contiguous intein into an N-terminal and a C-terminal fragment. A protein containing one half of a split-intein pair can thus be coupled to another protein containing the other half, through reconstitution of the functional intein and subsequent self-excision; a process known as *trans*-splicing (Fig. 2.2B). Ramirez et al. used this approach to generate a hydrogel by mixing two soluble protein block copolymers, each containing a trimerizing CutA domain and half of a split-intein pair [57]. Although split-inteins are certainly interesting for linking *de novo* designed

polypeptides into higher order nanostructures, it must be mentioned that flanking residues can considerably influence splicing efficiency [58].

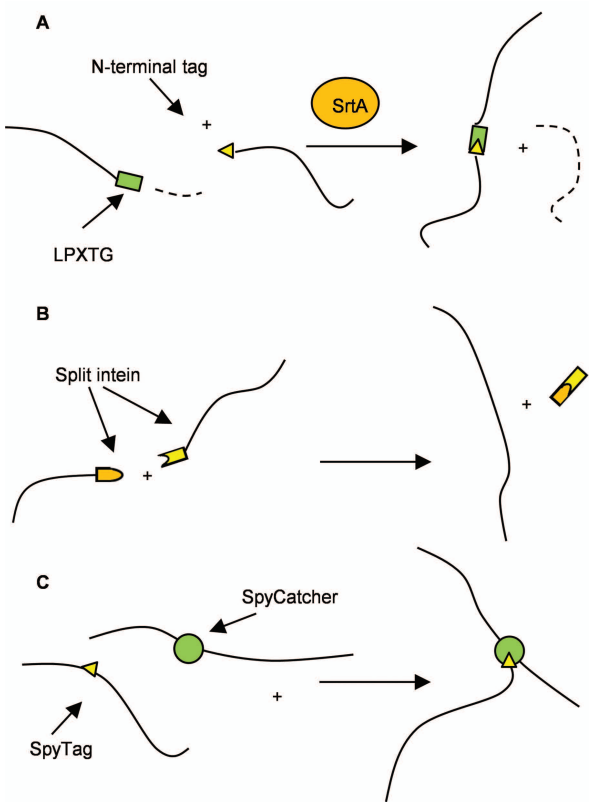


Figure 2.2: *Enzymatic cross-linking with narrow substrate specificity (A) Sortase A cleaves between threonine and glycine at an LPXTG recognition motif and subsequently joining the carboxyl group of threonine to an amino group of glycine. (B) An Intein, after association and refolding of split fragments, regains the ability to auto-catalytically excise itself and form a covalent bond between the previously flanking residues. (C) SpyTag and SpyCatcher undergo autocatalytic isopeptide bond formation between Asp117 on SpyTag and Lys31 on SpyCatcher.*

2.2.6. SpyTag/SpyCatcher

The SpyTag/SpyCatcher system originates from the fibronectin binding protein FbaB of *Streptococcus pyogenes* [59]. This biotechnologically exploited system consists of a small protein (SpyCatcher, 12kDa) and a peptide (SpyTag, 1.1kDa). The Howarth group demonstrated that upon mixing, SpyTag and SpyCatcher undergo autocatalytic isopeptide bond formation between Asp_{117} on SpyTag and Lys_{31} on SpyCatcher [59] (Fig. 2.2C). The Tirrell group used this

system in order to design new ELP polypeptide topologies [60]. Neither SpyTag nor SpyCatcher needs to be at a terminus, which allowed ELP designs that formed 3-arm, 4-arm, and H-shape ELP topologies [60]. The SpyTag-SpyCatcher approach can also be used to biofunctionalize self-assembled protein nanomaterials [61,62]. The SpyTag/SpyCatcher system is highly attractive for creating precisely defined polypeptide nanostructures, since it appears to be quite robust with respect to reaction conditions and is reported to have a high yield. A drawback for some applications, however, may be the relatively large size of the SpyCatcher protein, as this may influence the nanostructure's physicochemical properties and possibly its biocompatibility.

2.3. Reversible cross-linking via biorecognition modules

In some cases, links between polypeptides in nanomaterials should be reversible rather than irreversible, for example, in self-healing materials, injectable gels, etc. Such reversible physical bonds can be engineered via "bio-recognition" modules: small proteins/peptides known to have specific interactions with each other. Much work in the past has dealt with homo-oligomerizing domains that can be used to direct self-assembly of larger polypeptide constructs. Examples are: silk-inspired blocks, collagen-like triple helices [63], and condensed phases of elastin-like polypeptides [64]. Here we will restrict ourselves to more recent work on domains that can be used to physically couple different polypeptides via hetero-oligomerizing domains, which offer more control over macromolecular topology and allow assembly upon mixing different components.

2.3.1.Coiled coils and triple helices

Alpha-helical coiled coils are structural motifs present in many different proteins [65]. The fundamentals of coiled-coil sequence design have been very well studied and as a consequence, hetero-oligomeric coiled coils are increasingly being used in the fabrication of protein nanoarchitectures [66-71]. The power of using sets of orthogonal coiled coils was beautifully illustrated by the "protein origami" work of the Jerala group, in which a sequence of orthogonal coiled-coils was expressed as a single self-assembling recombinant polypeptide [72]. Zhang et al. have developed a hetero-hexameric coiled coil system that was used to physically cross-link self-assembling nanofibers, leading to formation of injectable hydrogels [73]. We have shown that heterodimer-forming coiled coils can be fused to hydrophilic random coil and self-assembling polypeptides, and can be produced efficiently by secretory

expression in the yeast *Pichia pastoris* [74]. Although coiled coils are important structural motifs with great potential in directing the assembly of *de novo* designed polypeptides into higher order structures, a potential problem is that homodimerization may compete with heterodimerization, especially at high concentrations [74].

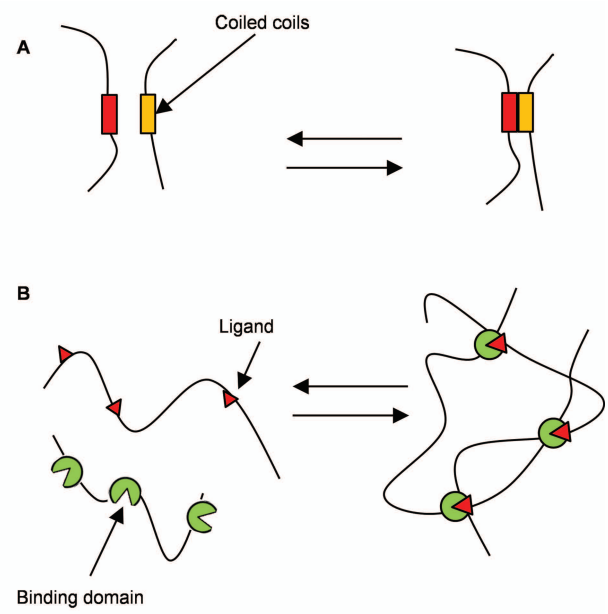


Figure 2.3: *Reversible cross-linking via biorecognition modules. (A) Coiled coils, assembling into heterodimers. (B) Biorecognition modules such as WW domains or PDZ domains specifically (but typically weakly) interact with their peptide ligands. The example shows protein polymers with multiple WW domains interacting with multiple PPxY ligands to form injectable hydrogels [80].*

2.3.2. Protein modules that bind peptide ligands

Many protein-protein or protein-peptide interactions in nature are elicited by small and well-delimited domains of the interacting polypeptide chains. Such small domains can be exploited as selective binding modules in *de novo* designed recombinant polypeptides, for the construction of supramolecular nanostructures and materials. So far, only a limited number of them has been exploited to direct supramolecular assembly. The so-called WW domains are small protein modules with two highly conserved tryptophans that have been identified in many signalling and regulatory proteins [75]. These modules, and their engineered variants [76] have a specific binding affinity for proline-rich

peptide motifs *in vitro* [77-79] (Fig. 2.3B). They have been used for engineering self-assembly in particular by the group of Heilshorn, which has developed two-component hydrogels based on the WW domain and proline-rich peptide [80-83]. Since the protein–peptide ligand interaction is relatively weak, the authors incorporated multiple WW and multiple proline-rich motifs. Other well-known protein interaction domains are the so-called PDZ domains that typically interact with C-terminal parts of target proteins through beta-strand addition [84,85]. Guan et al. used a single-domain PDZ protein called Tip-1 for constructing a two component self-assembling hydrogel system, where the interaction was reinforced by an engineered disulphide linkage [86]. A similar set of specifically interacting domains with disulphide locking is the Cys-tag/Ad-C pair derived from human RNase I [87], but it has so far not been employed for the construction of protein nanomaterials.

2.4. Conclusion and outlook

The examples discussed here illustrate the potential of using cross-linking tools to assemble designed polypeptides into precisely structured nanomaterials. Given the rapid progress in biomolecular engineering, it can be expected that in the near future these tools will become still more precise and versatile: new cross-linking enzymes, autocatalytic cross-linking modules and hetero-oligomerizing modules may become available via genomic and data mining studies, whereas existing ones may be improved for engineering purposes, e.g. via combinations of targeted engineering and directed evolution. For example, directed evolution was already used to identify *SrtA* mutants with improved catalytic activity [88]. Such improved tools will allow for the creation of novel protein materials with still better controlled structures at the nanoscale, and hence, for biomedical materials with more precisely defined interactions with living cells and organisms.

Reference

1. Liu, L.; Busuttil, K.; Zhang, S.; Yang, Y.; Wang, C.; Besenbacher, F.; Dong, M. *Phys. Chem. Chem. Phys.* **2011**, 13, 17435-17444.
2. Luo, Q.; Dong, Z.; Hou, C.; Liu, J. *Chem. Commun.* **2014**, 50, 9997-10007.
3. Heslot, H. *Biochimie* **1998**, 80, 19-31.
4. Annabi, N.; Mithieux, S. M.; Camci-Unal, G.; Dokmeci, M. R.; Weiss, A. S.; Khademhosseini, A. *Biochem. Eng. J.* **2013**, 77, 110-118.
5. Butterfoss, G. L.; Kuhlman, B. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35, 49-65.
6. Khoury, G. A.; Smadbeck, J.; Kieslich, C. A.; Floudas, C. A. *Trends Biotechnol.* **2014**, 32, 99-109.
7. Deming, T. J. *Adv Mate* **1997**, 9, 299-311.
8. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2011**, 12, 269-289.
9. Ghandehari, H.; Cappello, J. *Pharm. Res.* **1998**, 15, 813-815.
10. Shi, P.; Gustafson, J. A.; MacKay, J. A. *Int J Nanomedicine* **2014**, 9, 1617-1626.
11. Price, R.; Poursaid, A.; Ghandehari, H. *J. Control. Release* **2014**, 190, 304-313.
12. Costa, A. M. S.; Mano, J. F. *Eur. Polym. J.* **2015**, 72, 344-364.
13. Matson, J. B.; Stupp, S. I. *Chem. Commun.* **2012**, 48, 26-33.
14. Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater.* **2014**, 10, 1542-1557.
15. Fichman, G.; Gazit, E. *Acta Biomater.* **2014**, 10, 1671-1682.
16. Ahadian, S.; Sadeghian, R. B.; Salehi, S.; Ostrovidov, S.; Bae, H.; Ramalingam, M.; Khademhosseini, A. *Bioconjug. Chem.* **2015**, 26, 1984-2001.
17. Walper, S. A.; Turner, K. B.; Medintz, I. L. *Curr. Opin. Biotechnol.* **2015**, 34, 232-241.
18. Ta, H. T.; Peter, K.; Hagemeyer, C. E. *Trends Cardiovasc. Med.* **2012**, 22, 105-111.
19. Teixeira, L. S.; Feijen, J.; van Blitterswijk, C. A.; Dijkstra, P. J.; Karperien, M. *Biomaterials* **2012**, 33, 1281-1290.
20. Heck, T.; Faccio, G.; Richter, M.; Thony-Meyer, L. *Appl. Microbiol. Biotechnol.* **2013**, 97, 461-475.
21. Kopecek, J.; Yang, J. *Angew. Chem. Int. Ed. Engl.* **2012**, 51, 7396-7417.
22. Kopecek, J.; Yang, J. *Acta Biomater.* **2009**, 5, 805-816.

23. Wang, A. Y.; Leong, S.; Liang, Y. C.; Huang, R. C.; Chen, C. S.; Yu, S. M. *Biomacromolecules* **2008**, 9, 2929-2936.

24. Onaizi, S. A.; Leong, S. S. *Biotechnol. Adv.* **2011**, 29, 67-74.

25. Tessmar, J. K.; Gopferich, A. M. *Adv Drug Deliv Rev* **2007**, 59, 274-291.

26. Bakota, E. L.; Aulisa, L.; Galler, K. M.; Hartgerink, J. D. *Biomacromolecules* **2011**, 12, 82-7.

27. Greenberg, C. S.; Birckbichler, P. J.; Rice, R. H. *FASEB J.* **1991**, 5, 3071-3077.

28. Ando, H.; Adachi, M.; Umeda, K.; Matsuura, A.; Nonaka, M.; Uchio, R.; Tanaka, H.; Motoki, M. *Agr Biol Chem Tokyo* **1989**, 53, 2613-2617.

29. Dennler, P.; Bailey, L. K.; Spycher, P. R.; Schibli, R.; Fischer, E. *ChemBioChem* **2015**, 16, 861-861.

30. Tanaka, T.; Kamiya, N.; Nagamune, T. *FEBS Lett.* **2005**, 579, 2092-2096.

31. Pastor, M. T.; Diez, A.; Pérez-Payá, E.; Abad, C. *FEBS Lett.* **1999**, 451, 231-234.

32. Siegmund, V.; Schmelz, S.; Dickgiesser, S.; Beck, J.; Ebenig, A.; Fittler, H.; Frauendorf, H.; Piater, B.; Betz, U. A. K.; Avrutina, O.*; et al. Angew. Chem. Int. Ed.* **2015**, 54, 13420-13424.

33. Collier, J. H.; Messersmith, P. B. *Bioconjug. Chem.* **2003**, 14, 748-755.

34. Hu, B. H.; Messersmith, P. B. *J. Am. Chem. Soc.* **2003**, 125, 14298-14299.

35. Khew, S. T.; Yang, Q. J.; Tong, Y. W. *Biomaterials* **2008**, 29, 3034-3045.

36. McHale, M. K.; Setton, L. A.; Chilkoti, A. *Tissue Eng.* **2005**, 11, 1768-1779.

37. Bozzini, S.; Giuliano, L.; Altomare, L.; Petrini, P.; Bandiera, A.; Conconi, M. T.; Fare, S.; Tanzi, M. C. *J. Mater. Sci. Mater. Med.* **2011**, 22, 2641-2650.

38. Davis, N. E.; Ding, S.; Forster, R. E.; Pinkas, D. M.; Barron, A. E. *Biomaterials* **2010**, 31, 7288-7297.

39. Herchenhan, A.; Uhlenbrock, F.; Eliasson, P.; Weis, M.; Eyre, D.; Kadler, K. E.; Magnusson, S. P.; Kjaer, M. *J. Biol. Chem.* **2015**, 290, 16440-16450.

40. Kagan, H. M.; Li, W. *J. Cell. Biochem.* **2003**, 88, 660-672.

41. Kagan, H. M.; Williams, M. A.; Williamson, P. R.; Anderson, J. M. *J. Biol. Chem.* **1984**, 259, 11203-12037.
42. Lucero, H. A.; Kagan, H. M. *Cell. Mol. Life Sci.* **2006**, 63, 2304-2316.
43. Veitch, N. C. *Phytochemistry* **2004**, 65, 249-259.
44. Minamihata, K.; Goto, M.; Kamiya, N. *Bioconjug. Chem.* **2011**, 22, 74-81.
45. Partlow, B. P.; Hanna, C. W.; Rnjak-Kovacina, J.; Moreau, J. E.; Applegate, M. B.; Burke, K. A.; Marelli, B.; Mitropoulos, A. N.; Omenetto, F. G.; Kaplan, D. L. *Adv. Funct. Mater.* **2014**, 24, 4615-4624.
46. Qin, G.; Lapidot, S.; Numata, K.; Hu, X.; Meirovitch, S.; Dekel, M.; Podoler, I.; Shoseyov, O.; Kaplan, D. L. *Biomacromolecules* **2009**, 10, 3227-3234.
47. Heijnis, W. H.; Dekker, H. L.; de Koning, L. J.; Wierenga, P. A.; Westphal, A. H.; de Koster, C. G.; Gruppen, H.; van Berkel, W. J. *J. Agric. Food Chem.* **2011**, 59, 444-449.
48. Dhayal, S. K.; Gruppen, H.; de Vries, R.; Wierenga, P. A. *Food Hydrocolloid* **2014**, 36, 53-59.
49. Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 12424-12429.
50. Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc.* **2004**, 126, 2670-2671.
51. Piluso, S.; Cassell, H. C.; Gibbons, J. L.; Waller, T. E.; Plant, N. J.; Miller, A. F.; Cavalli, G. *Soft Matter* **2013**, 9, 6752-6756.
52. Kim, W.; Haller, C.; Dai, E.; Wang, X.; Hagemeyer, C. E.; Liu, D. R.; Peter, K.; Chaikof, E. L. *Angew. Chem. Int. Ed. Engl.* **2015**, 54, 1461-1465.
53. Mohlmann, S.; Mahlert, C.; Greven, S.; Scholz, P.; Harrenga, A. *ChemBioChem* **2011**, 12, 1774-1780.
54. Li, Y. *Biotechnol. Lett.* **2015**, 37, 2121-2137.
55. Blot, M.; Chong, S.; Perler, F. In *Prokaryotic Genomics*; Birkhäuser Basel: 2003; pp 172-193.
56. Lesaicherre, M. L.; Lue, R. Y.; Chen, G. Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2002**, 124, 8768-8769.
57. Ramirez, M.; Guan, D.; Ugaz, V.; Chen, Z. *J. Am. Chem. Soc.* **2013**, 135, 5290-5293.
58. Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 11005-11010.
59. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, E690-E697.

60. Zhang, W. B.; Sun, F.; Tirrell, D. A.; Arnold, F. H. *J. Am. Chem. Soc.* **2013**, 135, 13988-13997.
61. Sun, F.; Zhang, W. B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, 111, 11269-11274.
62. Botyanszki, Z.; Tay, P. K.; Nguyen, P. Q.; Nussbaumer, M. G.; Joshi, N. S. *Biotechnol. Bioeng.* **2015**, 112, 2016-2024.
63. Huang, J.; Wong Po Foo, C.; Kaplan, D. L. *Polymer Reviews* **2007**, 47, 29-62.
64. Floss, D. M.; Schallau, K.; Rose-John, S.; Conrad, U.; Scheller, J. *Trends Biotechnol.* **2010**, 28, 37-45.
65. Moutevelis, E.; Woolfson, D. N. *J. Mol. Biol.* **2009**, 385, 726-732.
66. Petka, W. A.; Harden, J. L.; McGrath, K. P.; Wirtz, D.; Tirrell, D. A. *Science* **1998**, 281, 389-392.
67. Shen, W.; Zhang, K.; Kornfield, J. A.; Tirrell, D. A. *Nat. Mater* **2006**, 5, 153-158.
68. Apostolovic, B.; Danial, M.; Klok, H. A. *Chem. Soc. Rev.* **2010**, 39, 3541-3475.
69. Doles, T.; Bozic, S.; Gradisar, H.; Jerala, R. *Biochem. Soc. Trans.* **2012**, 40, 629-634.
70. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.*; et al. Science* **2013**, 340, 595-599.
71. Kocar, V.; Bozic Abram, S.; Doles, T.; Basic, N.; Gradisar, H.; Pisanski, T.; Jerala, R. *Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology* **2015**, 7, 218-237.
72. Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. *Nat. Chem. Biol.* **2013**, 9, 362-366.
73. Zhang, X.; Zhou, H.; Xie, Y.; Ren, C.; Ding, D.; Long, J.; Yang, Z. *Adv Healthc Mater* **2014**, 3, 1804-1811.
74. Domeradzka, N. E.; Werten, M. W. T.; de Vries, R.; de Wolf, F. A. *Biotechnol. Bioeng.* **2016**, 39, 61-67.
75. Andre, B.; Springael, J. Y. *Biochem. Biophys. Res. Commun.* **1994**, 205, 1201-1205.
76. Russ, W. P.; Lowery, D. M.; Mishra, P.; Yaffe, M. B.; Ranganathan, R. *Nature* **2005**, 437, 579-583.
77. Chen, H. I.; Sudol, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 7819-7823.

78. Dodson, E. J.; Fishbain-Yoskovitz, V.; Rotem-Bamberger, S.; Schueler-Furman, O. *Exp. Biol. Med.* **2015**, 240, 351-360.
79. Macias, M. J.; Hyvonen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. *Nature* **1996**, 382, 646-649.
80. Wong Po Foo, C. T.; Lee, J. S.; Mulyasasmita, W.; Parisi-Amon, A.; Heilshorn, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 22067-22072.
81. Mulyasasmita, W.; Cai, L.; Hori, Y.; Heilshorn, S. C. *Tissue engineering. Part A* **2014**, 20, 2102-2114.
82. Mulyasasmita, W.; Lee, J. S.; Heilshorn, S. C. *Biomacromolecules* **2011**, 12, 3406-3411.
83. Parisi-Amon, A.; Mulyasasmita, W.; Chung, C.; Heilshorn, S. C. *Adv Healthc Mater* **2013**, 2, 428-432.
84. Lee, H. J.; Zheng, J. J. *Cell Commun Signal* **2010**, 8, 8.
85. Mohanty, S.; Ovee, M.; Banerjee, M. *Biology* **2015**, 4, 88-103.
86. Guan, D.; Ramirez, M.; Shao, L.; Jacobsen, D.; Barrera, I.; Lutkenhaus, J.; Chen, Z. *Biomacromolecules* **2013**, 14, 2909-2916.
87. Backer, M. V.; Backer, J. M. In *Methods Mol. Biol.*; 2009; pp 257-266.
88. Chen, I.; Dorr, B. M.; Liu, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, 108, 11399-11404.

3 Production in *Pichia pastoris* of protein-based polymers with small heterodimer-forming modules

Abstract

Some combinations of leucine zipper peptides are capable of forming α-helical heterodimeric coiled coils with very high affinity. These can be used as physical cross-linkers in the design of protein-based polymers that form supramolecular structures, for example hydrogels, upon mixing solutions containing the complementary modules. Such two-component physical networks are of interest for many applications in biomedicine, pharmaceutics and diagnostics. This chapter describes the efficient secretory production of A and B type leucine zipper peptides fused to protein-based polymers in *Pichia pastoris*. By adjusting the fermentation conditions, we were able to significantly reduce undesirable proteolytic degradation. The formation of A-B heterodimers in mixtures of the purified products was confirmed by size exclusion chromatography. Our results demonstrate that protein-based polymers incorporating functional heterodimer-forming modules can be produced with *P. pastoris* in sufficient quantities for use in future supramolecular self-assembly studies and in various applications.

Published as: Domeradzka, N.E.; Werten, M.W.T.; de Vries, R. and de Wolf, F.A., Production in *Pichia pastoris* of protein-based polymers with small heterodimer-forming blocks *Biotechnol Bioeng* **2016**, 113: 953–960.

3.1. Introduction

Protein materials with precisely defined novel functionalities can now be designed via recombinant DNA technology. Indeed, *de novo* designed functional polypeptide blocks can be fused together into protein block copolymers and be produced as heterologous proteins in recombinant host cells by the expression of synthetic genes [1,2]. Such recombinant protein block copolymers can be tailored to self-organize into various nano- and higher order structures, often in response to external stimuli [3-5].

As compared to chemically synthesized block copolymers, recombinant protein polymers can offer superior purity, monodispersity, biocompatibility, biodegradability, and tunability of *in vivo* half-life [2,6]. This makes protein polymers attractive candidate materials for medical applications such as drug encapsulation, controlled drug release, tissue engineering, tissue augmentation, and biosensors [7-11].

In our laboratory, we have developed a range of self-assembling protein polymers containing silk-, collagen-, and elastin-inspired blocks, and produced these as secreted proteins in the yeast *Pichia pastoris* at g L^{-1} levels [12-15]. Self-assembly of these polymers is mainly controlled by pH and temperature of the environment. For several medical applications, it would be desirable to be able to physically link together different protein polymers or protein domains, at a user-chosen time, by simple mixing of separately produced A and B proteins. For example, a possible application could be injectable gels that quickly self-assemble *in situ* after mixing of protein polymers A and B.

In this chapter we investigate the incorporation, into some of our previously developed protein polymer designs, of short leucine zipper peptides capable of forming high affinity α-helical heterodimeric coiled coils. We chose a well characterized pair of leucine zipper peptides, that was first studied by Moll et al. [16], and later used by Glasgow et al. [17] to physically couple single-chain antibody fragments to adenovirus proteins.

The more acidic peptide of the pair will further be denoted as $\mathbf{D^A}$, the more basic peptide as $\mathbf{D^B}$. For proof-of-concept, we genetically incorporated $\mathbf{D^A}$ and $\mathbf{D^B}$ domains at the C-terminus of a previously-developed $\mathbf{T_9\text{-}C^P{}_4}$ diblock copolymer. The C-terminal domain of this diblock protein polymer, $\mathbf{C^P{}_4}$, is a highly hydrophilic random coil and consists of four identical copies of a 99 amino acid long sequence [18]. The N-terminal domain, $\mathbf{T_9}$, is a collagen-inspired $(Pro\text{-}Gly\text{-}Pro)_9$ peptide capable of forming temperature-sensitive homotrimers [15]. The combination of $\mathbf{T_9}$ and $\mathbf{C^P{}_4}$ blocks has been studied previously, for example in the hydrogel-forming triblock copolymer $\mathbf{T_9\text{-}C^P{}_4\text{-}}$

$\mathbf{T_9}$ [15,19]. As a control, to study $\mathbf{D^A/D^B}$ dimer formation in the absence of any other supramolecular structure formation, the $\mathbf{D^A}$ or $\mathbf{D^B}$ modules were also incorporated into a $\mathbf{C^P_4}$ protein polymer without $\mathbf{T_9}$ blocks.
The secretory production in *P. pastoris* of the four protein polymers $\mathbf{C^P_4\text{-}D^A}$, $\mathbf{C^P_4\text{-}D^B}$, $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$, and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ initially resulted in partial proteolytic degradation. By altering the fermentation conditions, essentially intact protein polymers were recovered from benchtop bioreactors in gram quantities. We show that the heterodimerizing modules are functional and can indeed physically couple two different (otherwise non-interacting) protein polymers in solution.

3.2. Materials and Methods

3.2.1. Construction of Expression Vectors and Strains

The double-stranded gene fragments encoding the more acidic heterodimer-forming leucine zipper peptide $\mathbf{D^A}$ and the more basic peptide $\mathbf{D^B}$ (previously denoted as $RR_{12}EE_{345}L$ and $EE_{12}RR_{345}L$, respectively [16]) were assembled via overlap extension PCR [20] from the oligonucleotides shown in Table 3.S1. For the amino acid sequences of $\mathbf{D^A}$ and $\mathbf{D^B}$ see Table 3.1.

Table 3.1. *Amino acid sequences of $\mathbf{D^A}$ and $\mathbf{D^B}$. In our design, both sequences end C-terminally in the cloning-derived amino acid sequence PAGG (not indicated), which also prevents removal of the C-terminal Lys residue of both leucine zippers by the P. pastoris Kex1 carboxypeptidase.*

Dimer-forming modules	Amino acid sequence
$\mathbf{D^A} = RR_{12}EE_{345}L$	LEIRAAFLRQRNTALRTEVAELEQEVQRLENEVSQYETRYGPLGGGK
$\mathbf{D^B} = EE_{12}RR_{345}L$	LEIEAAFLERENTALETRVAELRQRVQRLRNRVSQYRTRYGPLGGGK

The gene fragments were digested with *Xho*I/*Eco*RI and cloned into *Xho*I/*Eco*RI-digested vector pMTL23Δ*Bsa*I [12], in order to obtain two vectors pMTL23Δ*Bsa*I-$\mathbf{D^A}$ and pMTL23Δ*Bsa*I-$\mathbf{D^B}$. The vector pMTL23-$\mathbf{C^P_4}$ contains the sequence coding for $\mathbf{C^P_4}$ [18]. The vector pMTL23-$\mathbf{T_9\text{-}C^P_4}$, encodes, at the N-terminal side of $\mathbf{C^P_4}$, an additional $\mathbf{T_9}$ sequence [15]. The vectors were opened at the 3' end of the $\mathbf{C^P_4}$ or $\mathbf{T_9\text{-}C^P_4}$ gene with *Eco*RI/*Van*91I. The newly prepared constructs pMTL23Δ*Bsa*I-$\mathbf{D^A}$ and pMTL23Δ*Bsa*I-$\mathbf{D^B}$ were digested with *Eco*RI/*Dra*III to release inserts $\mathbf{D^A}$ and $\mathbf{D^B}$. The inserts were ligated to the opened vectors, resulting in pMTL23-$\mathbf{C^P_4\text{-}D^A}$, pMTL23-$\mathbf{T_9\text{-}C^P_4\text{-}D^A}$, pMTL23-$\mathbf{C^P_4\text{-}D^B}$ and pMTL23-$\mathbf{T_9\text{-}C^P_4\text{-}D^B}$. All four inserts were then released with *Xho*I/*Eco*RI and cloned into the likewise-digested *Pichia pastoris* expression

vector pPIC9 (Invitrogen). This resulted in the construction of the vectors: pPIC9-$\mathbf{C^{P}_{4}\text{-}D^{A}}$, pPIC9-$\mathbf{T_{9}\text{-}C^{P}_{4}\text{-}D^{A}}$, pPIC9-$\mathbf{C^{P}_{4}\text{-}D^{B}}$ and pPIC9-$\mathbf{T_{9}\text{-}C^{P}_{4}\text{-}D^{B}}$, respectively. The vectors were linearized with *Sal*I to target for integration at the *his4* locus. Transformation of *P. pastoris* GS115 by electroporation and selection of Mut^{+} transformants were performed as described previously [21]. Vectors pPIC9-$\mathbf{C^{P}_{4}\text{-}D^{B}}$ and pPIC9-$\mathbf{C^{P}_{4}\text{-}D^{A}}$ were also used for transformation of the protease A deficient strain SMD1168 (Invitrogen).

3.2.2. Fermentation

The fermentation setup consisted of a 2.5-L Bioflo 3000 stirred-tank bioreactor (New Brunswick Scientific) interfaced with BioCommand Software (New Brunswick Scientific) and a homebuilt methanol sensor-controller. The fermentations were performed as described previously [15], as follows. A starting volume of 1.25 L minimal basal salts medium [22] was used. The cultures were always inoculated with precultures grown to similar OD_{600}. Growth temperature was 30 °C and the pH was controlled at 3.0 throughout the entire fermentation. The inlet air was supplemented with 20% oxygen for the duration of the glycerol fed-batch phase and the methanol fed-batch phase, which is achieved by the bioreactor by interrupting the air stream with short pulses of pure oxygen. During the induction phase, methanol levels were kept at a constant level of 0.2% (w/v). Agitation control in this phase was programmed to maintain dissolved oxygen levels at 20%, for as long as the oxygen demand of the methanol non-limited cells did not exceed the maximum oxygen transfer capacity of the bioreactor. Where indicated in Results, some of the fermentation parameters were varied in an attempt to reduce proteolysis. This involved various combinations of (a) a culture pH of 5.0 throughout the entire fermentation, (b) supplementation of the medium with 1% casamino acids, (c) supplementation of the medium with 10 mM ascorbic acid, and not supplementing the inlet air with oxygen during the methanol fed-batch phase, and (d) a growth temperature of 20 °C during the induction phase. In case of medium supplementation, casamino acids or ascorbic acid were added approximately 10 min. before methanol induction. For induction at 20 °C, the temperature was linearly decreased from 30 °C to 20 °C during the last 40 min. prior to methanol induction. The induction phase lasted two or three days, for fermentations at 30 °C or 20 °C, respectively, where dissolved oxygen levels typically reached 0% by one or two days, respectively.

3.2.3. Protein Purification

Purification of all protein polymers was done by ammonium sulfate precipitation essentially as described, which typically results in a purity of ~99% at the protein level [15]. Shortly, the pH of the cell-free broth was raised to 8.0 by addition of sodium hydroxide to allow precipitation of medium salts. After centrifugation for 30 min. at 20,000 × g (RT), the protein was precipitated from the supernatant by addition of ammonium sulfate to 40% of saturation, followed by incubation on ice for 30 min and centrifugation for 30 min. at 20,000 × g (4 °C). The protein pellet was resuspended in Milli-Q water and precipitated once more. The pellet was then resuspended in Milli-Q water, desalted by extensive dialysis against Milli-Q water, and finally lyophilized.

3.2.4. SDS-PAGE, Densitometry, and N-Terminal Protein Sequencing

SDS-PAGE was performed using the NuPAGE Novex System with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 pre-stained molecular mass markers. Prior to electrophoresis, all samples were heated for 10 min. at 70 °C in NuPAGE LDS Sample Buffer with NuPAGE Sample Reducing Agent, as per manufacturer's recommendations for denaturing and reducing PAGE. For gel staining, Coomassie SimplyBlue SafeStain (Invitrogen) was used. Gel images were acquired using a GS-800 calibrated densitometer and Quantity One software (Bio-Rad). Densitometric analyses were done using both lane-based density traces and band-based quantification in Quantity One, and were in good mutual accordance. Blotting of proteins for N-terminal sequencing by Edman degradation was described previously [21]. Protein sequencing was performed by Midwest Analytical (St. Louis, MO).

3.2.5. Mass Spectrometry

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was performed using an ultrafleXtreme mass spectrometer (Bruker). Samples were prepared by the dried droplet method on a 600 µm AnchorChip target (Bruker), using 5 mg mL^{-1} 2,5-dihydroxyacetophenone, 1.5 mg mL^{-1} diammonium hydrogen citrate, 25% (v/v) ethanol and 3% (v/v) trifluoroacetic acid as matrix. Spectra were derived from ten 500-shot (1,000 Hz) acquisitions taken at non-overlapping locations across the sample. Measurements were made in the positive linear mode, with ion source 1, 25.0 kV; ion source 2, 23.3 kV; lens, 6.5 kV; pulsed ion extraction, 680 ns. Protein Calibration Standard II (Bruker) was used for external calibration.

3.2.6. Size Exclusion Chromatography

To determine whether both $\mathbf{D^A}$ and $\mathbf{D^B}$ modules incorporated into protein-based polymers are functional, size exclusion chromatography (SEC) was performed on the $\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{C^P_4}$-$\mathbf{D^B}$ diblocks. The polymers containing a $\mathbf{T_9}$ block are not suitable for this study, owing to the fact that triple-helix formation would complicate the analysis of heterodimer formation. The lyophilized proteins were dissolved in 0.5 M sodium phosphate buffer (pH 7) and incubated overnight at 4 °C to allow dimerization. $\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{C^P_4}$-$\mathbf{D^B}$ were mixed in such a way that the final molar concentration of each diblock was 30 μM. The samples were prepared with different molar fractions of $\mathbf{C^P_4}$-$\mathbf{D^A}$ in the mixture of $\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{C^P_4}$-$\mathbf{D^B}$. The following fractions of $\mathbf{C^P_4}$-$\mathbf{D^A}$ were chosen: 0.75, 0.67, 0.60, 0.55, 0.50, 0.45, 0.40, and 0.33. As controls for homodimerization we tested these diblocks alone at 30 μM concentration. As a molecular size marker we used $\mathbf{C^P_4}$ alone [21] and $\mathbf{T_9}$-$\mathbf{C^P_8}$-$\mathbf{T_9}$ alone [23], both at 30 μM concentration. Void volume was determined using Dextran Blue 2000 (GE Healthcare). Samples were injected in a volume of 50 μL on a Superdex 200 Increase 10/300 GL column (GE Healthcare) at ambient temperature. The column was connected to a Biologic DuoFlow Chromatography system (Bio-Rad). Elution was done using 0.5 M sodium phosphate buffer (pH 7) at a flow rate of 0.5 mL min^{-1}. The eluate was monitored at 214 nm. Prior to injection, the column was equilibrated with the elution buffer.

3.3. Results

3.3.1. Production of $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^B}$

$\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^B}$ strains were grown at 30 °C and pH 3, and proteins were purified from the cell-free broth by precipitation with 40% ammonium sulfate and subsequent dialysis and lyophilisation [15]. The average gravimetrically determined yield obtained after purification for these two proteins was 0.6 ± 0.04 g L^{-1} of culture, with a mean productivity of 14 ± 2 mg L^{-1} per h, and a mean biomass-specific productivity of 25 ± 4 μg h^{-1} per g of wet biomass. Next, the protein polymers were analyzed by SDS-PAGE. The theoretical molecular weights of $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^B}$ are 44,971 and 45,053 Da, respectively. It should be noted, however, that proteins containing the $\mathbf{C^P_4}$ block are known to migrate very slowly in SDS-PAGE owing to their highly hydrophilic character and consequent low binding of SDS [15,18]. For both polymers, two bands were detected in SDS-PAGE, at positions corresponding to marker bands of ~98 kDa and ~188 kDa (Fig. 3.1, lanes 1 and 2). Although

aberrant migration of the proteins was expected, the presence of two bands suggests partial degradation. Because attachment of a short peptide with varied amino acid composition previously increased the migration rate of $\mathbf{C^P_2}$ in SDS-PAGE [24], the lower bands seen here may well correspond to the intact $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ proteins, and the upper bands may represent the proteins with most of the leucine zipper peptides missing owing to degradation. In agreement with this notion, Fig. 3.1.S1 shows that the upper band in $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ migrates at similar speed as the related control protein $\mathbf{T_9\text{-}C^P_4\text{-}T_9}$, known to migrate at similar speed as $\mathbf{C^P_4}$ [15]. We further investigated the molecular weight distribution of purified $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ by MALDI-TOF mass spectrometry (Fig. 3.2A). Also in MALDI-TOF, two peaks were observed for each of these proteins: 44,959 Da and 39,648 Da for $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$, and 45,033 Da and 39,844 Da for $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$. The main peaks correspond, within experimental error, to the expected molecular weights of intact $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ mentioned above. In agreement with the observations in SDS-PAGE, the minor peaks of 39,648 Da and 39,844 Da, respectively, most likely correspond to proteins with almost the entire $\mathbf{D^A}$ and $\mathbf{D^B}$ peptides cleaved off, as the $\mathbf{T_9\text{-}C^P_4}$ part alone has an expected mass of 39,293 Da. The majority of the $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ molecules remained intact, but apparently, in a minor fraction of the molecules, the $\mathbf{D^A}$ and $\mathbf{D^B}$ peptides are degraded by an endogenous protease of *P. pastoris*.

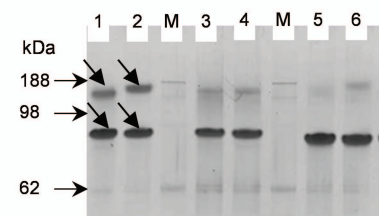


Figure 3.1: *SDS-PAGE. Protein-based polymers produced according to the standard protocol: lane 1, $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$, lane 2, $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$. Protein-based polymers produced after optimization of fermentation conditions: lane 3, $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$, lane 4, $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$, lane 5, $\mathbf{C^P_4\text{-}D^A}$, and lane 6, $\mathbf{C^P_4\text{-}D^B}$. Lanes M: protein molecular weight marker. The lower arrows point to the bands corresponding to the intact proteins and the upper arrows indicate the partly degraded products. 13 μg of protein was applied in each lane.*

3.3.2. Production using the protease A deficient strain SMD1168

We first attempted to remedy the observed protein degradation by using the protease A deficient strain SMD1168 [25], while leaving the fermentation

conditions unchanged. Preliminary indications suggested that the proteins were still degraded, and in view of the relatively poor and unreliable growth of this strain, and a low protein yield (~0.2 g L^{-1} of culture), no further attempts at using this strain were made.

3.3.3. Production at pH 5

One of the most crucial factors influencing proteolysis is known to be the culture pH [21,26,27]. As our standard fermentation conditions for $\mathbf{C^{P}_{4}}$-type polymers involve pH 3, we now attempted pH 5 for the GS115 strain transformed with $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$. The protein yield was comparable to that obtained in the fermentation at pH 3. SDS-PAGE again showed two bands and the upper band was even more present than at pH 3. Accordingly, MALDI-TOF again showed two peaks corresponding to intact and degraded $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$, but this time the two peaks had similar intensity (Fig. 3.1.S2). Apparently, degradation of the protein only increased by changing the pH to 5.

3.3.4. Production under 'best guess' conditions

Next, rather than changing one parameter at a time, we chose a 'best guess' combination of culture conditions known to sometimes remedy problems with proteolysis: (i) the temperature of the induction phase was reduced to 20 °C, (ii) the medium was supplemented with 1% casamino acid, and (iii) 10 mM ascorbic acid was included in the medium. Temperature reduction can reportedly reduce proteolysis of secreted proteins in *P. pastoris* [28,29]. Addition of casamino acids has likewise been helpful for several proteins produced in *P. pastoris* [21,30]. Inclusion of ascorbic acid as an antioxidant can reportedly reduce proteolysis related to cell death resulting from oxidative stress. This stress is caused by intracellular accumulation of reactive oxygen species generated by the yeast's oxygen-dependent methanol metabolism [31]. Because the bioreactor used adds pure oxygen in a pulsed manner to obtain a time-averaged constant supplementation of the inlet air with 20% oxygen, we omitted oxygen supplementation from these fermentations as a precaution, so as to further minimize oxidative stress that might be caused by momentary high oxygen levels to cells near the sparger during a pure oxygen pulse.

Under these combined conditions, fermentation of $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{B}}$ in the GS115 strain resulted in a purified protein yield of 0.5 ± 0.05 g L^{-1} of culture, with a mean productivity of 8 ± 0.6 mg L^{-1} per h, and a mean biomass-specific productivity of 17 ± 0.8 µg h^{-1} per g of wet biomass. For both proteins,

SDS-PAGE again showed the same two bands, although the relative intensity of the upper bands was now much reduced (Fig. 3.1, lanes 3 and 4). MALDI-TOF (Fig. 3.2B) showed major peaks of 44,955 Da and 45,044 Da, respectively, which correspond well to the expected sizes of the intact proteins.

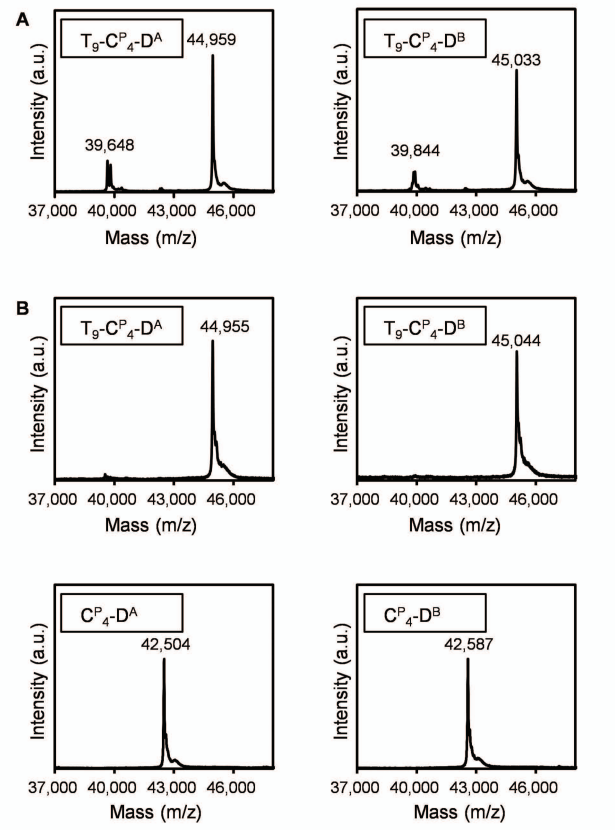


Figure 3.2: *(A)* ***T*$_9$-*C*$^P{}_4$-*D*^{A}** *and* ***T*$_9$-*C*$^P{}_4$-*D*^{B}** *produced according to the standard protocol; peaks at ~40 kDa indicate significant product degradation. (B)* ***T*$_9$-*C*$^P{}_4$-*D*^{A}, *T*$_9$-*C*$^P{}_4$-*D*^{B}, *C*$^P{}_4$-*D*^{A}** *and* ***C*$^P{}_4$-*D*^{B}** *produced after optimization of the fermentation conditions.*

The minor peak of ~39.6 kDa for the degraded species was hardly visible in **T$_9$-C$^P{}_4$-D^A** and completely absent in **T$_9$-C$^P{}_4$-D^B**. Apparently, the altered fermentation conditions indeed led to the production of nearly intact proteins, albeit at somewhat lower yields. Also **C$^P{}_4$-D^A** and **C$^P{}_4$-D^B** were produced under these conditions. SDS-PAGE again showed only minor degradation, as evident from the faint upper band (Fig. 3.1, lanes 5 and 6). MALDI-TOF (Fig. 3.2B) showed only one peak of 42,504 Da for **C$^P{}_4$-D^A** and one peak of 42,587

Da for $\mathbf{C^{P}_{4}\text{-}D^{B}}$, which match their theoretical molecular weights of 42,510 and 42,593 Da, respectively. Therefore, also for these proteins, the 'best guess' fermentation conditions resulted in nearly intact proteins.

As mentioned above, and even though MALDI-TOF suggests the proteins are (nearly) intact, both $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{B}}$ produced under improved conditions still show faint upper bands in SDS-PAGE, although much less intense than before. To verify the nature of both the upper and lower bands in SDS-PAGE, they were subjected to N-terminal sequencing. The sequence of the lower bands matched the expected N-termini of $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{B}}$. Approximately 20% of the signal corresponded to Glu-Ala extended species, which is commonly observed and is due to incomplete processing of the α-factor prepro secretory signal [21]. There was no sign of further secondary reactions, which, in accordance with its migration rate in SDS-PAGE, suggests this band indeed corresponds to the intact protein. Analysis of the faint upper band also revealed sequences matching the expected N-termini of $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{A}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{B}}$ (again with traces of a Glu-Ala extension), but additionally contained ~25% of ambiguous minor components such as Gly, Gln, Ser and Asn. Since these amino acids are abundant in $\mathbf{C^{P}_{4}}$ this could indicate internal degradation, but may also just be background. Given their slow migration rate and ~75% intact N-terminus, the proteins in these bands most likely correspond mainly to C-terminally truncated species (with almost the entire coiled-coil domain removed), as hypothesized above. Nonetheless, it is clear from MALDI-TOF and the faint staining of the upper bands in SDS-PAGE that the preparations contain only negligible amounts of these degraded species. Indeed, densitometry of samples separated by SDS-PAGE shows that proteins produced under standard conditions were degraded for roughly 30-40%, whereas samples obtained under improved conditions contained only around 5-15% of degraded protein. Possibly, the minor fraction of degraded species remaining under 'best guess' conditions is barely visible in MALDI-TOF owing to preferential crystallization/ionization of the intact proteins.

3.3.5. Influence of temperature, casamino acids, and ascorbic acid on product degradation

After establishing that for this new set of fermentation conditions nearly intact products were obtained, we set out to investigate which of the modified conditions were essential for eliminating proteolysis. A number of fermentations were performed, in which each time either one, or two of the

fermentation conditions were changed from the 'best guess' conditions described above, as detailed in Table 3.2.

Table 3.2: *Break-down of fermentation parameters involved in product degradation. Numbers between parentheses refer to numbered fermentation conditions in the text. Qualifications of protein integrity are based on MALDI-TOF (Fig. 3.1.S3). Note that 'degraded' in all cases reflects only moderate degradation compared to standard conditions (30 °C, no ascorbic acid, no casamino acids).*

		Level of product degradation	
Condition	Culture parameters	T_9-C^P_4-D^A	T_9-C^P_4-D^B
'Best guess' (Fig. 3.2B)	20 °C Ascorbic acid Casamino acids	Nearly intact	Nearly intact
No ascorbic acid (1) (Fig. 3.1.S3A)	20 °C No ascorbic acid Casamino acids	Nearly intact	Nearly intact
No casamino acids (2) (Fig. 3.1.S3B)	20 °C Ascorbic acid No casamino acids	Degraded	Nearly intact
Normal temperature, no ascorbic acid (3) (Fig. 3.1.S3C)	30 °C No ascorbic acid Casamino acids	Degraded	Degraded

When the fermentations were repeated in the absence of ascorbic acid (with normal oxygen-supplemented air supply), under otherwise identical conditions (condition 1 in Table 3.2; Fig. 3.1.S3A), **T_9-C^P_4-D^A** and **T_9-C^P_4-D^B** remained as intact as under 'best guess' conditions (Fig. 3.2B). This implies that the use of ascorbic acid, and thereby a putative reduction of oxidative stress [31], was not a critical parameter. When the optimized fermentation at 20 °C was repeated without addition of casamino acids (condition 2 in Table 3.2; Fig. 3.1.S3B), only minor degradation of **T_9-C^P_4-D^A** occurred and negligible degradation of **T_9-C^P_4-D^B**. Also when the fermentation was performed at 30 °C in the presence of casamino acids (condition 3 in Table 3.2; Fig. 3.1.S3C), only minor degradation occurred, albeit slightly more, particularly for **T_9-C^P_4-D^B**, than under condition 2. Apparently, both addition of casamino acids and a growth temperature of 20 °C are effective in reducing proteolysis of the polymers, with the latter parameter having the largest effect on both proteins. The purified protein yield for all fermentations at 20 °C in Table 3.2 was similar (0.4 - 0.5 g L^{-1} of culture). As the final optimized conditions for ongoing work in our

laboratory involving these protein polymers we chose condition 1 (20 °C, supplementation with casamino acids, no ascorbic acid).

3.3.6. Heterodimer formation

To test the functionality of the heterodimer-forming modules incorporated into protein-based polymers, we used size exclusion chromatography (SEC) to analyse $\mathbf{D^A/D^B}$ dimer formation in mixtures of $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$. For this analysis, the $\mathbf{T_9\text{-}C^P{}_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^B}$ triblocks could not be used because the $\mathbf{T_9}$ module can form trimeric $(\mathbf{T_9})_3$ intermolecular cross-links, which would have disturbed the analysis. As a molecular weight marker we used $\mathbf{C^P{}_4}$. This polymer has already previously been characterized by e.g. SEC and MALDI-TOF [18] and does not form any aggregates. Its molecular weight of ~ 37 kDa is fairly close to that of $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$ (~ 42.5 kDa). As a second molecular weight marker, for lack of e.g. a $\mathbf{C^P{}_8}$ polymer, we used $\mathbf{T_9\text{-}C^P{}_8\text{-}T_9}$ [23]. At a low concentration, many of the trimer-forming $\mathbf{T_9}$ blocks will be present in a monomeric state, such that also the monomeric form of the $\mathbf{T_9\text{-}C^P{}_8\text{-}T_9}$ protein will be visible in SEC, in addition to trimers and larger supramolecular forms. The molecular weight of monomeric $\mathbf{T_9\text{-}C^P{}_8\text{-}T_9}$ is ~78 kDa, which is roughly similar to that of a $\mathbf{C^P{}_4\text{-}D^A/C^P{}_4\text{-}D^B}$ dimer (~85 kDa). $\mathbf{C^P{}_4}$ eluted from 16.8 mL to 18.4 mL, and the monomer fraction of $\mathbf{T_9\text{-}C^P{}_8\text{-}T_9}$ eluted from 14 mL to 16 mL (Fig. 3.1.S4). The fraction of $\mathbf{T_9\text{-}C^P{}_8\text{-}T_9}$ eluting in the void volume of the column at up to 14 mL (Fig. 3.1.S4) represents large multimeric forms.

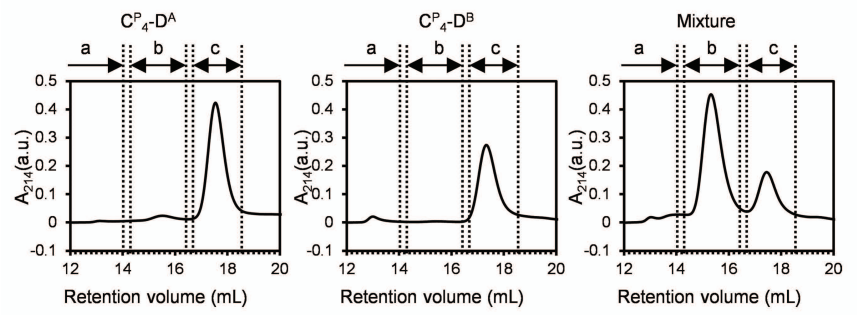


Figure 3.3: *SEC of $\mathbf{C^P{}_4\text{-}D^A}$ alone, $\mathbf{C^P{}_4\text{-}D^B}$ alone, and a mixture of $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$. The retention volume decreases with increasing size of the molecules: (a) void, (b) dimer, and (c) monomer. Each sample was applied at a concentration of 30 μM.*

When individually applied to the column, each at a concentration of 30 μM, $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$ eluted with a similar retention volume as the $\mathbf{C^P{}_4}$ reference (Fig. 3.3), which showed that in isolated form, $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$ occurred as

monomers at this concentration. After overnight incubation at 4 °C, a mixture of 30μM each of $\mathbf{C^P_4\text{-}D^A}$ and $\mathbf{C^P_4\text{-}D^B}$ was applied to the column. The chromatogram showed two peaks (Fig. 3.3). A minor peak was detected at a similar retention volume as monomeric $\mathbf{C^P_4\text{-}D^A}$, $\mathbf{C^P_4\text{-}D^B}$, or $\mathbf{C^P_4}$, and a major peak was detected at a smaller retention volume of 14.4 mL to 16.6 mL, similar to the retention volume of the monomeric form of $\mathbf{T_9\text{-}C^P_8\text{-}T_9}$, which is about twice as large as $\mathbf{C^P_4\text{-}D^A}$ or $\mathbf{C^P_4\text{-}D^B}$. The latter indicates the coexistence of $\mathbf{C^P_4\text{-}D^A}$**::**$\mathbf{C^P_4\text{-}D^B}$ dimers. To further analyze the binding stoichiometry, we analyzed mixtures of $\mathbf{C^P_4\text{-}D^A}$ and $\mathbf{C^P_4\text{-}D^B}$ with different $\mathbf{C^P_4\text{-}D^A}$/$\mathbf{C^P_4\text{-}D^B}$ ratios (f_A) (Fig. 3.4). The fraction of dimer was taken to be the ratio of the area under the dimer peak in the chromatogram divided by the total area under the monomer and dimer peaks. The largest fraction of dimers is obtained at f_A = 0.52, indicating an approximate 1:1 binding stoichiometry.

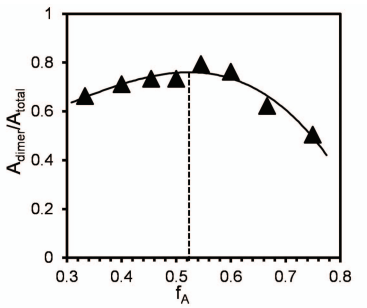


Figure 3.4: *Dimer formation in* $\boldsymbol{C^P_4\text{-}D^A}$ *and* $\boldsymbol{C^P_4\text{-}D^B}$ *mixtures containing different fractions of* $\boldsymbol{C^P_4\text{-}D^A}$ *(f_A). The fraction of dimer was taken to be the ratio (A_{dimer}/A_{total}) of the area under the dimer peak in the chromatogram, divided by the total area under the monomer and dimer peaks. The dashed line indicates the largest fraction of dimers.*

3.4. Discussion

We have shown that small modules forming heterodimers with high specificity and affinity can be incorporated into protein polymers, which can be produced in secretory fashion at g L^{-1} levels using *Pichia pastoris*. The heterodimer forming modules are functional in the context of the protein polymers, thus allowing us to design and produce further two-component systems in the future. With respect to fermentation conditions, we found that a lower induction temperature and, to a slightly lesser degree, also the addition of casamino acids reduced the extent of proteolysis. A low induction temperature has reportedly been effective also in the production of a fusion protein consisting of *Candida antarctica* lipase B and a cellulose-binding domain [32]. Low temperature can

reduce proteolysis simply for thermodynamic reasons, and because of reduced cell death and accompanying release of proteases into the medium [29]. Degradation of herring antifreeze protein in shake flask cultures was also reduced by growth at a low temperature [28]. Addition of casamino acids, mainly acting as competitive protease inhibitors, has been helpful for several proteins produced in *P. pastoris* [21,30]. For our protein polymers, the combination of both low temperature and addition of casamino acids resulted in the lowest level of degradation, indicating that at least for some proteins the effects of both strategies can be additive. The low growth temperature did result in ~20% less protein recovered per culture volume, an ~46% lower mean productivity, and an ~30% lower mean productivity per unit of biomass. Nonetheless, the ~1g of purified protein obtained per benchtop fermentation is more than adequate for laboratory studies.

We did not find a significant effect on proteolysis of supplementing the culture medium with ascorbic acid. This contrasts findings for the production of the peptide hirudin, where addition of ascorbic acid was found to reduce oxidative stress and thereby reduce product degradation [31]. Of course different proteases may be involved in the degradation of hirudin and the protein polymers described here. Furthermore, oxidative stress may play a lesser role in our process because methanol non-limited cultures quickly become oxygen-limited [33,34], whereas Xiao et al. [31] maintained dissolved oxygen levels at >40%.

When studying the interaction between $\mathbf{C^{P}_{4}\text{-}D^{A}}$ and $\mathbf{C^{P}_{4}\text{-}D^{B}}$ using SEC, the maximal extent of heterodimer formation was found at the expected $\mathbf{D^{A}/D^{B}}$ mixing ratio of about 1:1. However, at this mixing ratio and 30 μM of each of the protein polymer components, about 20% of the molecules were still present as monomers (see Fig. 3.3). This is likely caused by the combined effects of association-dissociation kinetics and chromatographic separation of dimers and monomers [35], and by the 5-15% of degraded molecules remaining in the polymers.

From the SEC analysis we also conclude that $\mathbf{D^{B}}$ modules self-interact at higher concentrations; a finding that had not been reported prior to our study. At a concentration of 100 μM of $\mathbf{C^{P}_{4}\text{-}D^{B}}$, in addition to the monomer peak, the chromatogram also shows a peak between 12 and 14 mL (Fig. 3.1.S5), which corresponds to the void volume of the column (Fig. 3.1.S4). The $\mathbf{C^{P}_{4}\text{-}D^{B}}$ diblock probably forms oligomers with sizes exceeding the separation range of the column (10^5 g mol^{-1} for dextrans). According to Moll et al. [16] the melting temperatures of the $\mathbf{D^{A}}$ and $\mathbf{D^{B}}$ homodimers are below 0°C, which implies that homomerization cannot take place at room temperature. Unfortunately, the

authors do not mention the protein concentration at which they did their experiments. Our analysis of their data shows that this concentration must have been around 10 μM, and that for a $\mathbf{D^B}$ concentration of 100 μM the T_m of the homodimer must be around room temperature. This, in retrospect, explains the $\mathbf{D^B}$ self-interaction we report here. We observed that equimolar mixtures of $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^A}$ and $\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^B}$ form hydrogels at high concentrations, in accordance with the possibility of network formation through the occurrence of $\mathbf{D^A/D^B}$ heterodimers and $(\mathbf{T_9})_3$ collagen-like trimers (Fig. 3.1.S6). However, the unexpected interactions between $\mathbf{D^B}$ domains at higher concentrations likely also play a role in these network structures.

It appears, therefore, that the present $\mathbf{D^A/D^B}$ pair of heterodimerizing leucine zippers is best used at low concentrations, where the homodimer melting temperatures are low, and one can safely assume that homodimerization does not occur. For example, we have previously designed and produced silk-like protein polymers that form fibrillar hydrogels [12,13,36]. By mixing in a very small fraction of fibril-forming silk-like protein polymers that additionally have $\mathbf{D^A}$ and $\mathbf{D^B}$ domains, we may be able to bundle and/or cross-link the fibrils in order to modulate the mechanical properties of the gels, as required by many biomedical applications of protein polymer hydrogels [37-39]. The $\mathbf{D^A/D^B}$ pair could also serve as a means of attaching bioactive domains to self-assembled supramolecular structures [41], or could be used in nanocomposite hydrogels. For example, inspired by a study by Appel, et al. [41], we envision using the dimerizing modules to establish connections between self-assembling protein polymers and nanoparticles. This approach is increasingly being explored to develop shear-thinning injectable gels [42,43].

3.1 Supplementary data

Table 3.1.S1: *Oligonucleotides used in gene construction.*

Insert	Oligonucleotide Name	Oligonucleotide sequences
D^A	A-FW1	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTTTAGAAAT TAGAGCTGCCTTTTTGAGACAG-3′
	A-RV1	5′GTCTCTGAACCTCTTGTTCCAACTCTGCAACTTCAGTCCTCAAAGCGGTA TTACGCTGTCTCAAAAAGGCAGCTC-3′
	A-FW2	5′GGAACAAGAGGTTCAGAGACTTGAAAACGAAGTCTCTCAATATGAGACTA GATACGGTCCACTTGGA-3′
	A-RV2	5′GTACGAATTCTATTAGCCACCGGCTGGCTTTCCACCTCCAAGTGGACCGT ATCTAG-3′
D^B	B-FW1	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTTTGGAAAT TGAAGCCGCCTTTCTTGAGC-3′
	B-RV1	5′GAACCCTTTGTCTAAGCTCAGCAACTCTGGTTTCTAAAGCAGTGTTCTCA CGCTCAAGAAAGGCGGCTTC-3′
	B-FW2	5′CTGAGCTTAGACAAAGGGTTCAAAGGTTGAGAAATAGAGTCTCTCAGTAT AGAACTCGTTACGGTCCATTG-3′
	B-RV2	5′GTACGAATTCTATTAGCCACCGGCTGGCTTTCCACCTCCCAATGGACCGT AACGAGTT-3′

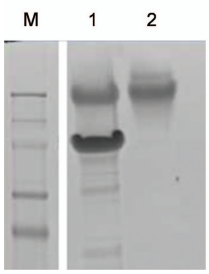


Figure 3.1.S1: *SDS-PAGE migration rate comparison. Lane 1,* T_9-C^P_4-D^B*; lane 2, control protein* T_9-C^P_4-T_9*, known to migrate at similar speed as* C^P_4 [15]*; lane M: protein molecular weight marker. 13 μg of protein was applied in each lane.*

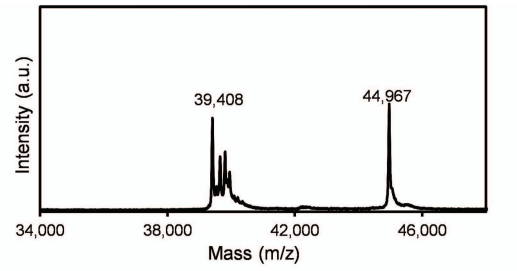


Figure 3.1.S2: *MALDI-TOF spectrum of* T_9-C^P_4-D^A *produced at pH 5.*

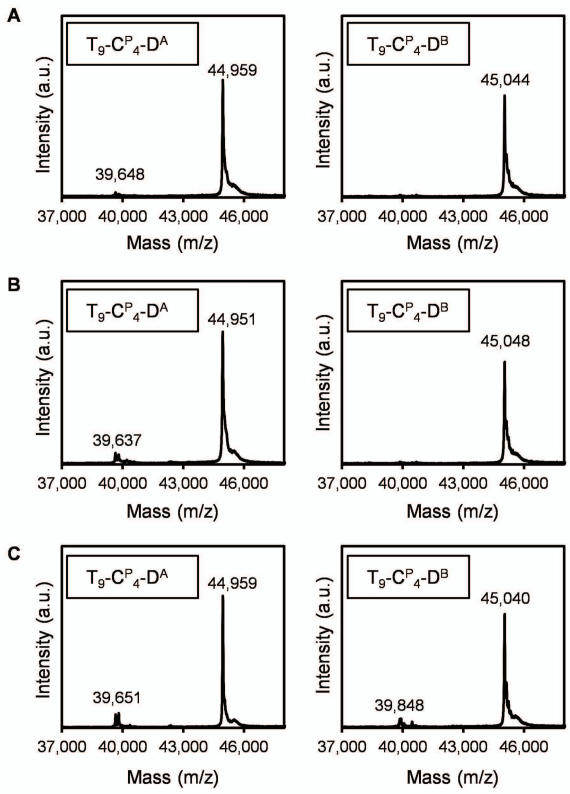


Figure 3.1.S3: *MALDI-TOF of* T_9-C^P_4-D^A *and* T_9-C^P_4-D^B *produced under various conditions. (A) without ascorbic acid (at 20 °C with casamino acids; condition 1), (B) without casamino acids (at 20 °C, with ascorbic acid; condition 2) (C) at 30 °C and without ascorbic acid (with casamino acids; condition 3).*

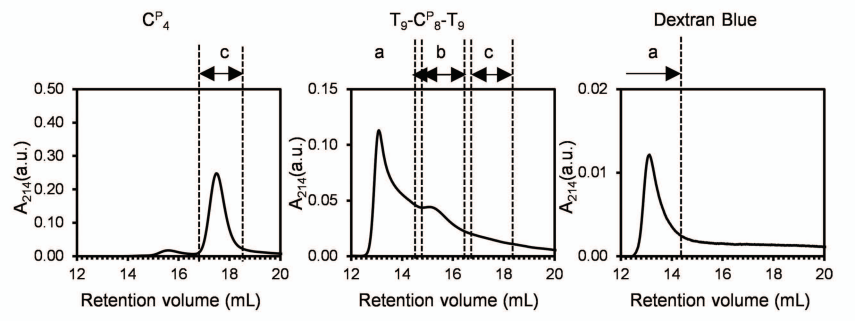


Figure 3.1.S4. *SEC of size markers Dextran Blue 2000,* T_9-C^P_8-T_9*, and* C^P_4*. The retention volume decreases with increasing size of the molecules: (a) void, (b)* T_9-C^P_8-T_9 *monomer, (c)* C^P_4 *monomer. Dextran Blue is used to determine the void volume of the column. Proteins were applied at a concentration of 30 μM. Owning to its collagen-like homotrimer-forming* T_9 *blocks,* T_9-C^P_8-T_9 *at low concentrations is present in solution as monomers, and, eluting in the void, as trimers and multimers.* C^P_4 *does not form multimers.*

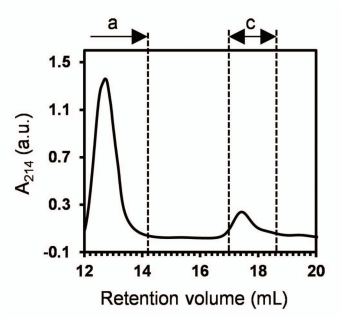


Figure 3.1.S5. *SEC of* C^P_4-D^B *at a concentration of 100 μM, showing (a) multimer and (c) monomer peaks.*

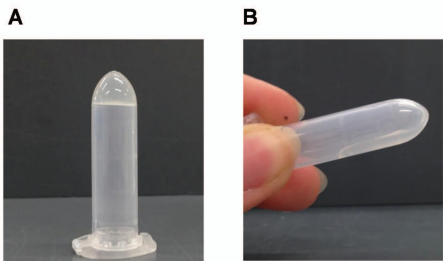


Figure 3.1.S6. *Inverted tube test for an equimolar mixture of* T_9-C^P_4-D^A *and* T_9-C^P_4-D^B*. (A) At 20 °C. (B) Immediately after heating to 45 °C. The final concentration of each protein was 3 mM.*

Reference

1. Luo, Q.; Dong, Z.; Hou, C.; Liu, J. *Chem Commun (Camb).* **2014**, 50, 9997-10007.
2. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2011**, 12, 269-289.
3. Beun, L. H.; Storm, I. M.; Werten, M. W. T.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 3349-3357.
4. MacEwan, S. R.; Chilkoti, A. *Biopolymers.* **2010**, 94, 60-77.
5. Xia, X.-X.; Wang, M.; Lin, Y.; Xu, Q.; Kaplan, D. L. *Biomacromolecules* **2014**, 15, 908-914.
6. Fluegel, S.; Buehler, J.; Fischer, K.; McDaniel, J. R.; Chilkoti, A.; Schmidt, M. *Chemistry.* **2011**, 17, 5503-5506.
7. Ma, P. X. *Adv Drug Deliv Rev.* **2008**, 60, 184-198.
8. Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater.* **2014**, 10, 1542-1557.
9. Ghandehari, H.; Cappello, J. *Pharm Res.* **1998**, 15, 813-815.
10. Malafaya, P. B.; Silva, G. A.; Reis, R. L. *Adv Drug Deliv Rev.* **2007**, 59, 207-233.
11. Nguyen, M. K.; Lee, D. S. *Macromol Biosci.* **2010**, 10, 563-579.
12. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W. T.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 699-706.
13. Martens, A. A.; Portale, G.; Werten, M. W. T.; de Vries, R. J.; Eggink, G.; Cohen Stuart, M. A.; de Wolf, F. A. *Macromolecules* **2009**, 42, 1002-1009.
14. Pham, T. T. H.; Wang, J.; Werten, M. W. T.; Snijkers, F.; de Wolf, F. A.; Cohen Stuart, M. A.; van der Gucht, J. *Soft Matter* **2013**, 9, 8923-8930.
15. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009**, 10, 1106-1113.
16. Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci.* **2001**, 10, 649-655.
17. Glasgow, J. N.; Mikheeva, G.; Krasnykh, V.; Curiel, D. T. *PLoS ONE* **2009**, 4, e8355.
18. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng.* **2001**, 14, 447-454.
19. Skrzeszewska, P. J.; de Wolf, F. A.; Werten, M. W. T.; Moers, A. P. H. A.; Cohen Stuart, M. A.; van der Gucht, J. *Soft Matter* **2009**, 5, 2057-2062.

20. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. *Gene* **1989**, 77, 51-59.
21. Werten, M. W. T.; van den Bosch, T. J.; Wind, R. D.; Mooibroek, H.; de Wolf, F. A. *Yeast* **1999**, 15, 1087-1096.
22. Zhang, W.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M. *Biotechnol Bioeng.* **2000**, 70, 1-8.
23. Teles, H.; Skrzeszewska, P. J.; Werten, M. W. T.; van der Gucht, J.; Eggink, G.; de Wolf, F. A. *Soft Matter* **2010**, 6, 4681-4687.
24. Moers, A. P. H. A.; Wolbert, E. J. H.; de Wolf, F. A.; Werten, M. W. T. *J Biotechnol* **2010**, 146, 66-73.
25. Gleeson, M. G.; White, C.; Meininger, D.; Komives, E. In *Pichia Protocols*; Higgins, D., Cregg, J., Eds.; Humana Press: 1998; pp 81-94.
26. Chang, S. W.; Shieh, C. J.; Lee, G. C.; Akoh, C. C.; Shaw, J. F. *J Mol Microbiol Biotechnol* **2006**, 11, 28-40.
27. Daly, R.; Hearn, M. T. W. *J Mol Recognit.* **2005**, 18, 119-138.
28. Li, Z.; Xiong, F.; Lin, Q.; d'Anjou, M.; Daugulis, A. J.; Yang, D. S. C.; Hew, C. L. *Protein Expr Purif.* **2001**, 21, 438-445.
29. Jahic, M.; Wallberg, F.; Bollok, M.; Garcia, P.; Enfors, S.-O. *Microb Cell Fact.* **2003**, 2, 1-11.
30. Clare, J. J.; Romanes, M. A.; Rayment, F. B.; Rowedder, J. E.; Smith, M. A.; Payne, M. M.; Sreekrishna, K.; Henwood, C. A. *Gene* **1991**, 105, 205-212.
31. Xiao, A.; Zhou, X.; Zhou, L.; Zhang, Y. *Appl Microbiol Biotechnol.* **2006**, 72, 837-844.
32. Jahic, M.; Gustavsson, M.; Jansen, A.-K.; Martinelle, M.; Enfors, S.-O. *J Biotechnol.* **2003**, 102, 45-53.
33. Charoenrat, T.; Ketudat-Cairns, M.; Stendahl-Andersen, H.; Jahic, M.; Enfors, S.-O. *Bioprocess Biosyst Eng* **2005**, 27, 399-406.
34. Khatri, N. K.; Hoffmann, F. *Biotechnol Bioprocess Eng* **2006**, 93, 871-879.
35. Yu, C.-M.; Mun, S.; Wang, N.-H. L. *J Chromatogr A.* **2006**, 1132, 99-108.
36. Włodarczyk-Biegun, M. K.; Werten, M. W. T.; de Wolf, F. A.; van den Beucken, J. J. J. P.; Leeuwenburgh, S. C. G.; Kamperman, M.; Cohen Stuart, M. A. *Acta Biomater.* **2014**, 10, 3620-3629.
37. Ehrick, J. D.; Deo, S. K.; Browning, T. W.; Bachas, L. G.; Madou, M. J.; Daunert, S. *Adv Drug Deliv Rev.* **2005**, 4, 298-302.
38. Seliktar, D. *Science* **2012**, 336, 1124-1128.

39. Tan, H.; Marra, K. G. *Materials* **2010**, 3, 1746-1767.
40. Zhang, X.; Zhou, H.; Xie, Y.; Ren, C.; Ding, D.; Long, J.; Yang, Z. *Adv. Healthcare Mater.* **2014**, 3, 1804-1811.
41. Appel, E. A.; Tibbitt, M. W.; Webber, M. J.; Mattix, B. A.; Veiseh, O.; Langer, R. *Nat Commun* **2015**, 6.
42. Rose, S.; Prevoteau, A.; Elziere, P.; Hourdet, D.; Marcellan, A.; Leibler, L. *Nature* **2014**, 505, 382-385.
43. Schexnailder, P.; Schmidt, G. *Colloid. Polym. Sci.* **2009**, 287, 1-11.

4 Production in *Pichia pastoris* of small coiled coils incorporated into hydrogel-forming protein-based polymers

Abstract

In this chapter we focus on the design, biosynthesis in *Pichia pastoris* and purification of protein polymers $\mathbf{T_9\text{-}C^P{}_4\text{-}D}$ containing small C-terminal heterodimeric coiled coils $\mathbf{D^E}$ = $(EIAALEK)_3$ and $\mathbf{D^K}$ = $(KIAALKE)_3$. The $\mathbf{C^P{}_4}$ is a hydrophilic random coil polypeptide, and the $\mathbf{T_9}$ block is a trimer-forming collagen-like block, such that the $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E/T_9\text{-}C^P{}_4\text{-}D^K}$ couple is expected to form networks upon mixing. We found that protein polymers containing the $\mathbf{D^K}$ module were partially degraded. A different culture pH, and the use of a yapsin 1 protease disruptant did not resolve this problem, but we did show that cation exchange chromatography can be used to purify the intact fraction of $\mathbf{C^P{}_4\text{-}D^K}$ diblocks (that were produced as non-gelling controls). Using dynamic light scattering-based microrheology and oscillation rheology, we show that $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ forms thermosensitive hydrogels, apparently via homotypic interactions of the $\mathbf{D^E}$ modules. At the same concentrations, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ shows a thermosensitive increase in viscosity, but does not gel, indicating the presence of homotypic association of the $\mathbf{D^K}$, but not as strong as in the case of the $\mathbf{D^E}$ module. Mixed systems $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E/T_9\text{-}C^P{}_4\text{-}D^K}$ form hydrogels similarly as the $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ system, indicating that both the heterotypic and homotypic associations may be involved in the hydrogel formation.

4.1. Introduction

As discussed in previous chapters, recombinant protein polymers [1,2] can be used as biocompatible, biodegradable, non-toxic, and non-immunogenic biomaterials, that can additionally contain specific biofunctional domains such as cell attachment sites. These properties make protein polymers a promising class of polymers for a variety of medical applications [3-6]. Several research groups, including ours, successfully demonstrated high-level expression of recombinant protein polymers composed of collagen-like [7-9], elastin-like [10,11] or silk-like [12,13] peptide blocks. Many of these protein polymers can self-assemble into supramolecular nanostructures that often are reversible and stimulus responsive [14-18]. Although for protein polymers we have very good control over the primary sequence, it is still difficult to design sequences that will result in a given type of self-assembly. One way of engineering an additional level of control over supramolecular self-assembly is by the introduction, into the protein polymers, of blocks that can self-assemble with very high specificity [19]. In **Chapter 3**, we have shown that by using heterodimerization of two complementary coiled coils we can bring together two protein polymers, which alone do not show ability to self-assemble. Whereas the coiled-coils used in that chapter were rather long, in this chapter we explore the use of much smaller, *de novo* designed coiled coils. This pair of coiled coils was developed by Litowski et al. in to be used as an expression tag/affinity purification system [20]. According to Litowski et al, the heterodimeric couple can form fully folded coiled-coil with a high conformational stability, despite its small size (21 residues) [20]. The peptide with the highest glutamic acid content $(EIAALEK)_3$ will further be denoted as $\mathbf{D^E}$, and the peptide with the highest lysine content $(KIAALKE)_3$ as $\mathbf{D^K}$. Similar to the dimer-forming modules described in **Chapter 3**, the $\mathbf{D^E}$ and $\mathbf{D^K}$ peptides were fused to previously developed protein polymers $\mathbf{C^P_4}$, and $\mathbf{T_9\text{-}C^P_4}$, at the C-terminus. The $\mathbf{C^P_4}$ block is a highly hydrophilic random coil [21] and $\mathbf{T_9}$ stands for the collagen-inspired sequence $(Pro\text{-}Gly\text{-}Pro)_9$ [8] that forms trimers at low temperatures and high concentrations, such that the $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$/$\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ couple is expected to form networks upon mixing.
In this chapter, we describe the high-yield secretory production of the protein polymers in *Pichia pastoris* and their characterization. The protein polymers $\mathbf{C^P_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ were produced intact and recovered from the cell-free broth in g L^{-1} quantities. However, we found that the $\mathbf{C^P_4\text{-}D^K}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ protein polymers were partially degraded. Neither the fermentation with a yapsin 1 protease disruptant nor the fermentation at

pH 5 resulted the intact product, but we were able to demonstrate that cation exchange chromatography can be used to purify intact $\mathbf{C^P_4\text{-}D^K}$, and to estimate the fraction of intact $\mathbf{C^P_4\text{-}D^K}$ diblocks. Next, we study the rheology of the (partially degraded) protein polymers, both the $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ triblocks separately, and mixtures of the two. We find that $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ forms thermosensitive hydrogels, apparently via homotypic interactions of the $\mathbf{D^E}$ modules. At the same concentrations, $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ shows a thermosensitive increase in viscosity, but does not gel, indicating the presence of homotypic association of the $\mathbf{D^K}$, but not as strong as in the case of the $\mathbf{D^E}$ modules. Mixed systems $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$/$\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ form hydrogels similarly as the $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ system, indicating that both the heterotypic and homotypic associations may be involved in the hydrogel formation.

Although the intent of this work was to control self-assembly via heterodimerization, the $\mathbf{D^E}$ modules might thus still be an interesting homooligomer-forming module for the construction of self-assembling systems.

4.2. Materials and Methods

4.2.1. Construction of expression vectors and strains

The coiled coils used here are short peptides with common amino acid sequence $(XIAALXY)_3$ [20]. The peptide $(EIAALEKE)_3$ is here denoted as $\mathbf{D^E}$; and the complementary peptide $(KIAALKE)_3$ is denoted as $\mathbf{D^K}$. For the amino acid sequences of $\mathbf{D^E}$ and $\mathbf{D^K}$ see Table 4.1. The double-stranded gene fragments encoding the coiled coils were assembled via overlap extension PCR [22] from the oligonucleotides shown in Table 4.2. The gene fragments were digested with *Xho*I/*Eco*RI and cloned into *Xho*I/*Eco*RI-digested vector pMTL23Δ*Bsa*I [23], in order to obtain two vectors pMTL23ΔBsaI-$\mathbf{D^E}$ and pMTL23ΔBsaI-$\mathbf{D^K}$. The vector pMTL23-$\mathbf{C^P_4}$ contains the sequence coding for $\mathbf{C^P_4}$ [21]. The vector pMTL23-$\mathbf{T_9\text{-}C^P_4}$, encodes, at the N-terminal side of $\mathbf{C^P_4}$, an additional $\mathbf{T_9}$ sequence [8]. The vectors were opened at the 3' end of the $\mathbf{C^P_4}$ or $\mathbf{T_9\text{-}C^P_4}$ gene with *Eco*RI/*Van*91I. The newly prepared constructs pMTL23ΔBsaI-$\mathbf{D^E}$ and pMTL23ΔBsaI-$\mathbf{D^K}$ were digested with *Eco*RI/*Dra*III to release inserts $\mathbf{D^E}$ and $\mathbf{D^K}$. The inserts were ligated to the opened vectors, resulting in pMTL23-$\mathbf{C^P_4\text{-}D^E}$, pMTL23-$\mathbf{T_9\text{-}C^P_4\text{-}D^E}$, pMTL23-$\mathbf{C^P_4\text{-}D^K}$ and pMTL23-$\mathbf{T_9\text{-}C^P_4\text{-}D^K}$. All four inserts were then released with *Xho*I/*Eco*RI and cloned into the likewise-digested *Pichia pastoris* expression vector pPIC9 (ThermoFisher, Bleiswijk, The Netherlands). This resulted in the construction

of the vectors: pPIC9-$\mathbf{C^P_4}$-$\mathbf{D^E}$, pPIC9-$\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^E}$, pPIC9-$\mathbf{C^P_4}$-$\mathbf{D^K}$ and pPIC9-$\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^K}$, respectively.

The vectors were linearized with *Sal*I to target for integration at the *his4* locus. Transformation of *P. pastoris* GS115 by electroporation and selection of Mut^+ transformants were performed as described previously [24]. Vector pPIC9-$\mathbf{T_9}$-$\mathbf{C^P_4}$-$\mathbf{D^K}$ was also used for transformation of the *yps*1 single-gene disruptant strain [25].

Table 4.1: *Amino acid sequences of* $\boldsymbol{D^E}$ *and* $\boldsymbol{D^K}$*. In our design, both sequences end C-terminally in the cloning-derived amino acid sequence PAGG (not indicated), which also prevents removal of the C-terminal lysine residues of the* D^E *module by the P. pastoris Kex1 carboxypeptidase.*

Dimer-forming module	Amino acid sequence
$\mathbf{D^E}$ = IAAL E3	EIAALEKEIAALEKEIAALEK
$\mathbf{D^K}$ = IAAL K3	KIAALKEKIAALKEKIAALKE

Table 4.2: *Oligonucleotides used in gene construction*

Insert	Oligonucleotide Name	Oligonucleotide sequences
$\mathbf{D^E}$	E-FW	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTGAAATTGCTGCCTTGGAAAAGGAGATCGCAG-3′
	E-RV	5′GTACGAATTCTATTAGCCACCGGCTGGCTTCTCAAGAGCGGCAATTTCTTTCTCCAAAGCTGCGATCTCCTTTTCCAAG-3′
$\mathbf{D^K}$	K-FW	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTAAGATCGCCGCCTTGAAGGAAAAGATTGCAGCTCT-3′
	K-RV	5′GTACGAATTCTATTAGCCACCGGCTGGTTCTTTCAAAGCAGCAATTTTCTCTTTAAGAGCTGCAATCTTTTCCTTC-3′

4.2.2. Fermentation

The fermentation setup consisted of a 2.5-L Bioflo 3000 stirred-tank bioreactor (New Brunswick Scientific, Nijmegen, The Netherlands) interfaced with BioCommand Software (New Brunswick Scientific, Nijmegen, The Netherlands) and a homebuilt methanol sensor controller. The fermentations were performed as described previously [8], as follows. A starting volume of 1.25

L minimal basal salts medium [26] was used. Growth temperature was 30 °C. The pH was controlled at 3.0, except the one fermentation, where the pH was maintained at 5.0. The air was supplemented with 20% (v/v) oxygen during the glycerol fed-batch phase and the methanol fed-batch phase. During the latter protein production phase, lasting two days, methanol levels were kept at 0.2% (w/v). After harvesting, cells were removed from the broth by centrifugation for 20 min. at 15,000 × g (RT), followed by microfiltration.

4.2.3. Protein purification

Purification of all protein polymers was done by ammonium sulfate precipitation essentially as described, which typically results in a purity of ~99% at the protein level [8]. Shortly, the pH of the cell-free broth was raised to 8.0 by addition of sodium hydroxide to allow precipitation of medium salts. After centrifugation for 30 min. at 20,000 × g (RT), the protein was precipitated from the supernatant by addition of ammonium sulfate to 40% of saturation, followed by incubation on ice for 30 min and centrifugation for 30 min. at 20,000 × g (4 °C). The protein pellet was resuspended in Milli-Q water and precipitated once more. The pellet was then resuspended in Milli-Q water, desalted by extensive dialysis against Milli-Q water, and finally lyophilized. The proteins purified in this manner were used for rheology and Isothermal Titration Calorimetry.

4.2.4. SDS-PAGE

SDS-PAGE was performed using the NuPAGE Novex System (ThermoFisher, Bleiswijk, The Netherlands) with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 prestained molecular mass markers. Prior to electrophoresis, all samples were heated for 10 min. at 70 °C in NuPAGE LDS Sample Buffer with NuPAGE Sample Reducing Agent, as per manufacturer's recommendations for denaturing and reducing PAGE. Gels were stained using Coomassie SimplyBlue SafeStain (ThermoFisher, Bleiswijk, The Netherlands).

4.2.5. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was performed using an ultrafleXtreme mass spectrometer (Bruker, Leiderdorp, The Netherlands). Proteins were desalted using Micro Bio-Spin P-6 columns (Bio-Rad, Veenendaal, The Netherlands), and samples were prepared by the dried droplet method on a 600 μm AnchorChip target (Bruker, Leiderdorp, The Netherlands), using 5 mg mL^{-1} 2,5-

dihydroxyacetophenone, 1.5 mg mL^{-1} diammonium hydrogen citrate, 25% (v/v) ethanol and 3% (v/v) trifluoroacetic acid as matrix. Spectra were derived from ten 500-shot (1,000 Hz) acquisitions taken at non-overlapping locations across the sample. Measurements were made in the positive linear mode, with ion source 1, 25.0 kV; ion source 2, 23.3 kV; lens, 6.5 kV; pulsed ion extraction, 680 ns. Protein Calibration Standard II (Bruker, Leiderdorp, The Netherlands) was used for external calibration.

4.2.6. Cation-exchange chromatography

Both lyophilized proteins $\mathbf{C^{P}_{4}\text{-}D^{K}}$ and $\mathbf{T_{9}\text{-}C^{P}_{4}\text{-}D^{K}}$ were dissolved in 20 mM sodium citrate buffer (pH 4) to concentration of 1g L^{-1} and incubated overnight at RT to allow proper dissolution. After incubation, both samples were centrifuged at 12 000 g for 20 min at RT and finally filtered through 0.45 μm pore Minisart-Plus filters (Sigma-Aldrich, Zwijndrecht, The Netherlands). Samples were injected in a volume of 1 mL on an ENrich S High-Resolution Ion Exchange Chromatography Column (Bio-Rad Laboratories, Veenendaal, The Netherlands) for separation by cation exchange. The column was connected to a Biologic DuoFlow Chromatography system (Bio-Rad Laboratories, Veenendaal, The Netherlands). Prior to injection, the column was equilibrated with the 20 mM sodium citrate buffer at pH 4 (Buffer A), at a flow rate of 1 mL min^{-1}. After injection, the column was washed with 3 ml (3 bed volumes) of Buffer A in order to remove unbound proteins. The elution was performed at RT. Proteins were eluted from the column with a linear gradient from 0 to 2 M NaCl over 30 ml (30 bed volumes) and then an isocratic gradient at 2 M over 1.5 ml (1.5 bed volumes). 1-ml fractions were collected throughout the linear gradient. The eluate was monitored at 214 nm. Fractions eluting between 230 mM and 600 mM NaCl were analysed by SDS-PAGE. The same fractions were desalted on Micro Bio-Spin P6 column (Bio-Rad, Veenendaal, The Netherlands) and analysed by MALDI-TOF.

4.2.7. Dynamic light scattering (DLS)-based microrheology

Dynamic light scattering (DLS)-based microrheology was performed using a Zetasizer Nano ZS (Malvern Instruments, Malvern, UK) equipped with a 4 mW HeNe laser beam with a wavelength of 633 nm, at a scattering angle of 173°. Carboxyl latex microspheres (kindly provided by dr Joris Sprakel) with a diameter of 112 nm were used as tracer particles [27]. Prior to use, the particles were diluted to a final weight to volume fraction of 0.1% and vortexed. Similar latex microspheres had been used for micro-rheology on $\mathbf{T_{9}\text{-}C_{4}\text{-}T_{9}}$ hydrogels,

and showed no undesired interactions with the protein [28]. The lyophilized protein polymers **T_9-C^P_4-D^E** and **T_9-C^P_4-D^K**, purified by ammonium sulfate precipitation, were dissolved in 10 mM sodium phosphate buffer (pH 7), vortexed and incubated overnight at RT. The concentration of both solutions was 80 g L^{-1}. **T_9-C^P_4-D^E** and **T_9-C^P_4-D^K** solutions were mixed in a 1:1 ratio and 10 mM sodium phosphate buffer (pH 7) was added to adjust concentration. Two samples were measured, where the final molar concentration of each triblock was 20 g L^{-1} or 40 g L^{-1} (note that total protein concentration was twice higher). As controls for homodimerization we tested these triblocks alone at concentrations 40 g L^{-1} and 80 g L^{-1}. Prior to the measurement, all samples were incubated at 50 °C for at least 20 min to prevent the triple helix formation by the **T_9** block. The tracer particles were then introduced at a final weight-to-volume fraction of 0.05%, and mixed with the protein solutions by vortexing. The samples with tracer particles were then introduced into a capped cuvette (Malvern Instruments, Malvern, UK) and incubated for 5 min. at 50 °C. The temperature was decreased step-wise from 50 °C with steps of 5 °C until it reached 10 °C. At each step, 40 measurements were performed, 120 sec each. The Dispersion Technology Software (Malvern Instruments, Malvern, UK) was used to calculate the z-averaged diffusion constant D of the tracer particles. The relative viscosity (η/η_0) (where η_0 is the viscosity of the aqueous solvent, and η is the viscosity of the polymer solution) was taken to be equal to the ratio D_0/D, where D_0 is the z-averaged diffusion constant of the tracer particles in aqueous solvent, and D is the z-averaged diffusion constant of the particles in the polymer matrix

4.2.8. Rheology

Freeze-dried proteins purified by AS precipitation were first dissolved in 10 mM sodium phosphate buffer, then vortexed and incubated overnight at RT to allow proper dissolution. **T_9-C^P_4-D^E** and **T_9-C^P_4-D^K** solutions were mixed in a 1:1 ratio. Two samples were prepared at concentration 80 g L^{-1} and 160 g L^{-1}**.** As controls for homodimerization we tested these triblocks alone at concentrations 80 g L^{-1}. Prior to the measurement all samples were heated at 50 °C for at least 20 min to allow any triple helices in the sample to melt completely. Rheological measurements were done with an Anton Paar MCR 501 rheometer equipped with C10/TI Couette geometry, with bob and cup diameter of 9.991 and 10.840 mm, respectively. The temperature was controlled by a Peltier system, which allows quick heating and cooling. A solvent trap containing oil was used to minimize evaporation. The couette was

preheated (50 °C) before adding sample solutions. After inserting the bob into the cup, the temperature was lowered to 20 °C. The gel formation was monitored by applying a sinusoidal deformation to the hydrogel (frequency 1 Hz and amplitude 1%).

4.3. Results and discussion

4.2.1. Protein production and characterisation

The basepair sequences encoding the aminoacid sequence of the coiled coils were introduced at the 3' end of the genes encoding either $\mathbf{C^P_4}$ protein polymer [21] or $\mathbf{T_9\text{-}C^P_4}$ protein polymer. The protein polymers $\mathbf{C^P_4\text{-}D^E}$, $\mathbf{C^P_4\text{-}D^K}$, $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ were secreted by *P. pastoris* into the fermentation medium upon induction with methanol. Proteins were purified from the cell-free medium by differential ammonium sulfate precipitation, and subsequently dialyzed and lyophilized. As previously reported for the protein polymers containing $\mathbf{C^P_4}$ block, this purification technique typically results in a purity of ~99% at the protein level [21]. The gravimetrically determined yields, expressed in grams of purified protein per litre of cell-free broth, are listed in Table 4.3. The yields obtained for all protein polymers are in the same g L^{-1} range as for $\mathbf{C^P_4}$ [21]. In order to assess the product identity, lyophilized samples were characterized by SDS-PAGE and MALDI-TOF. The theoretical molecular weights for the protein polymers were calculated based on their amino acid content and are listed in Table 4.3.

Table 4.3: *Molecular weights (Da) of predesigned protein polymers estimated based on their amino acids composition.*

Protein polymer code	Yield (g L^{-1})*	Theoretical molecular weight (Da)	Molecular weight determined by MALDI-TOF (Da)
$\mathbf{C^P_4\text{-}D^E}$	2.3	39373	39379
$\mathbf{T_9\text{-}C^P_4\text{-}D^E}$	1.6	41834	41833
$\mathbf{C^P_4\text{-}D^K}$	2.5	39370	39372**
$\mathbf{T_9\text{-}C^P_4\text{-}D^K}$	2	41831	41845**

* The gravimetrically determined yield is expressed as g of lyophilized product per L of cell-free broth

**The product additionally contained degraded species

For the protein polymers with a $\mathbf{D^E}$ module, two bands were detected with an apparent molecular weight in SDS-PAGE of ~120 kDa and ~188 kDa (Fig. 4.1, lanes 1 and 2). In both cases, the upper bands were barely visible. The SDS-PAGE of both $\mathbf{D^K}$ variants showed a clear band of ~98 kDa and three less visible upper bands (Fig 4.1, lanes 3 and 4). The occurrence of more than one band in SDS-PAGE suggests that the products may be partly degraded, and this seems most pronounced for protein polymers containing the $\mathbf{D^K}$ module.

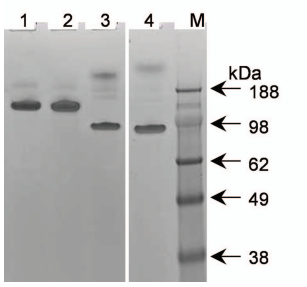


Figure 4.1: *SDS-PAGE of purified protein polymers (12 µg). Lane 1,* $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^E}$*; lane 2,* $\boldsymbol{C^P{}_4\text{-}D^E}$*; lane 3,* $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$*; lane 4,* $\boldsymbol{C^P{}_4\text{-}D^K}$*; lane M, protein molecular weight marker*

It is well known that $\mathbf{C^P{}_4}$ shows aberrant migration behavior in SDS-PAGE [8,10,21,23,29]. The $\mathbf{C^P{}_4}$ block is highly hydrophilic, and therefore its capacity to bind SDS is very low, as has been described before [21]. According to our previous findings, the attachment of peptides that improve SDS binding increased the mobility of $\mathbf{C_2}$– or $\mathbf{C_4}$–containing polymers [29,30]. Because the upper bands in Fig. 4.1 migrate at similar positions as $\mathbf{C^P{}_4}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}T_9}$ [8,21], and because the lower bands show increased mobility, the lower bands most probably correspond to the intact protein polymers, whereas the upper bands likely represent proteins that lack parts of the coiled coils sequences. The difference in the mobility between $\mathbf{D^E}$ and $\mathbf{D^K}$ variants is certainly related to the amino acid composition and the net charge of both modules. The $\mathbf{D^K}$ module is positively charged and probably attracts more anionic SDS molecules than $\mathbf{D^E}$ in accordance with earlier findings on the effect of protein charge on SDS binding [31].

The molecular weight distribution was further investigated by MALDI-TOF. For all analyzed protein polymers, the molecular mass corresponding to main peaks matched the theoretical mass within experimental error, as shown in Table 4.3. For both $\mathbf{C^P{}_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ proteins, only one peak was detected in the MALDI-TOF spectrum (Fig 4.2), which corresponds well to the

expected molecular mass. Because the upper bands in SDS-PAGE are barely visible, the degree of degradation or contamination is probably too low for detection in MALDI-TOF. We did not further investigate the nature of the upper bands at this point and assumed the proteins were sufficiently intact and pure for initial tests. The MALDI-TOF spectra of $\mathbf{C^{P}_{4}\text{-}D^{E}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{E}}$ showed a small shoulder to the right of the main peak. As previously discussed, such species with slightly higher mass frequently occur in *P. pastoris* and originate from incomplete removal of the Glu-Ala spacer located between the α-factor prepro leader and the protein of interest [24].

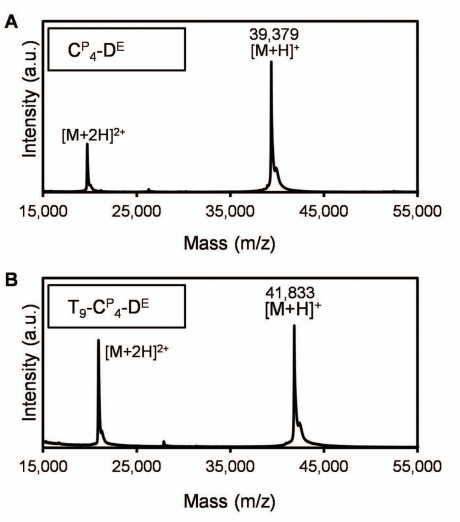


Figure 4.2: *MALDI-TOF of purified protein polymers. (A)* $\boldsymbol{C^{P}_{4}\text{-}D^{E}}$*, (B)* $\boldsymbol{T_9\text{-}C^{P}_{4}\text{-}D^{E}}$.

In contrast, the MALDI-TOF spectrum for both $\mathbf{C^{P}_{4}\text{-}D^{K}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{K}}$ contained three minor peaks corresponding to a lower molecular mass than the main peak (Fig 4.3 A and B). This MALDI-TOF analysis confirms the observations from SDS-PAGE that the $\mathbf{C^{P}_{4}\text{-}D^{K}}$ and $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{K}}$ samples probably constitute a mix of the intact protein and several degraded species from which different fragments of the $\mathbf{D^{K}}$ block are missing.

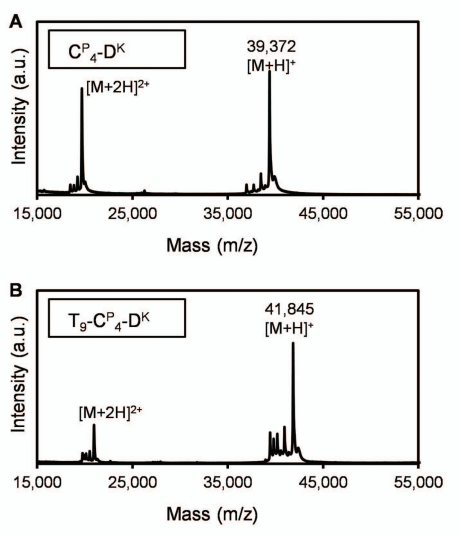


Figure 4.3: *MALDI-TOF of purified protein polymers. (A)* C^P_4-D^K*, (B)* T_9-C^P_4-D^K*.*

4.3.2. Attempts to minimize proteolytic degradation of T_9-C^P_4-D^K

MALDI-TOF of **T_9-C^P_4-D^K** showed a similar spectrum, yet shifted towards higher mass, as **C^P_4-D^K**. This suggests that both protein polymers undergo the same degradation of the **D^K** module. Most probably the amino acid sequence of the **D^K** module is digested in several places by endoproteases. For **T_9-C^P_4-D^K**, we made the two following attempts to remedy the observed protein degradation: (i) fermentation was performed at pH 5 instead of pH 3 [32], and (ii) the yapsin 1 deficient strain *yps1* [25] was used instead of the wild-type GS115 strain, under otherwise unchanged standard fermentation conditions. First, we tested the change of pH, since pH is often an important factor influencing proteolysis [24,33,34]. The protein yield of ~1.9 g L^{-1}, obtained at pH 5, was comparable to that obtained in the fermentation at pH 3. SDS-PAGE again showed a few bands and the top band was even more present than at pH 3 (Fig 4.4, lane 1). Accordingly, MALDI-TOF showed several peaks corresponding to intact and degraded **T_9-C^P_4-D^K,** but this time the intensity of the main peak relative to that of the other peaks decreased (Fig 4.4B). Apparently, degradation of the protein only increased by changing the pH to 5. Yapsin 1 (Yps1) is a well-characterized protease found at the plasma membrane, the extracellular medium and in the late secretory pathway [35]. The involvement of Yps1 in the degradation of secreted heterologous proteins in yeast has been previously reported [35], and was

involved in the degradation of protein polymers produced in *P. pastoris* [36]. The protease specifically cleaves C-terminal to Lys or Arg residues, although little is known about its exact sequence specificity. Because two of the three peaks corresponding to degraded fragments in the mass spectrum of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ seemed to precisely match cleavage after Lys residues present in the $\mathbf{D^K}$ module (not shown), yapsin 1 may well be involved in the observed degradation. To test this, we performed a fermentation with a yapsin 1 deficient strain (*yps1*) previously constructed in our laboratory [25]. This fermentation resulted in a yield of ~2.3 g L^{-1}. A similar yield was obtained for the strain GS115. Unfortunately, SDS-PAGE again showed several bands (Fig 4.4, lane 2). Similarly, MALDI-TOF revealed peaks with the exact same masses as in the wild type (not shown), although the peaks corresponding to the degraded species were perhaps less intense than before (Fig 4.4C). Apparently, other yapsins, or completely different proteases, are (also) involved in the degradation of the $\mathbf{D^K}$ module.

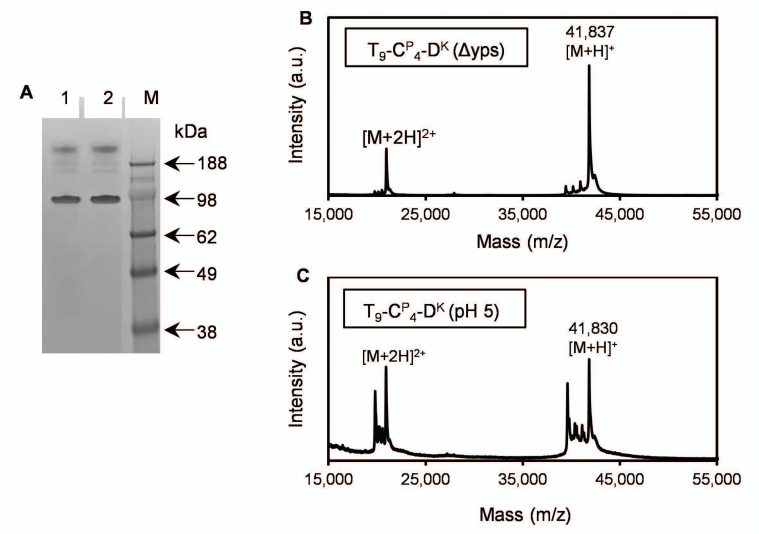


Figure 4.4: *Analysis of purified* $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$ *protein polymers. (A) SDS-PAGE of protein polymer (12 µg): lane 1, protein produced in Δyps strain and lane 2, produced at pH 5. MALDI-TOF of protein polymer (A) produced in Δyps strain and (B) produced at pH 5*

4.3.3. Cation exchange chromatography

For many medical applications monodispersity of protein-based materials is crucial [37]. Unfortunately the steps that we undertook in order to remedy the observed degradation of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ were unsuccessful, and most probably a

similar outcome could be expected for $\mathbf{C^P_4\text{-}D^K}$. Although optimizing culture conditions can sometimes solve proteolysis problems [38], there is no guarantee that this would be effective in this particular case, and further tests could be very time consuming. With a view for future use of protein polymers containing the $\mathbf{D^K}$ module, we therefore tested whether cation-exchange chromatography can be used for the removal of degraded species from both $\mathbf{C^P_4\text{-}D^K}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$. This will furthermore allow us to estimate the fraction of intact material. The $\mathbf{D^K}$ module has a positive net charge at pH 4 and therefore, the intact protein containing the full length $\mathbf{D^K}$ module will probably bind better to the column than degraded molecules without $\mathbf{D^K}$ module. The latter will elute first. The lyophilized protein polymers were dissolved, filtered and then subjected to column chromatography. The chromatogram showed several peaks for both $\mathbf{C^P_4\text{-}D^K}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ (Fig 4.5).

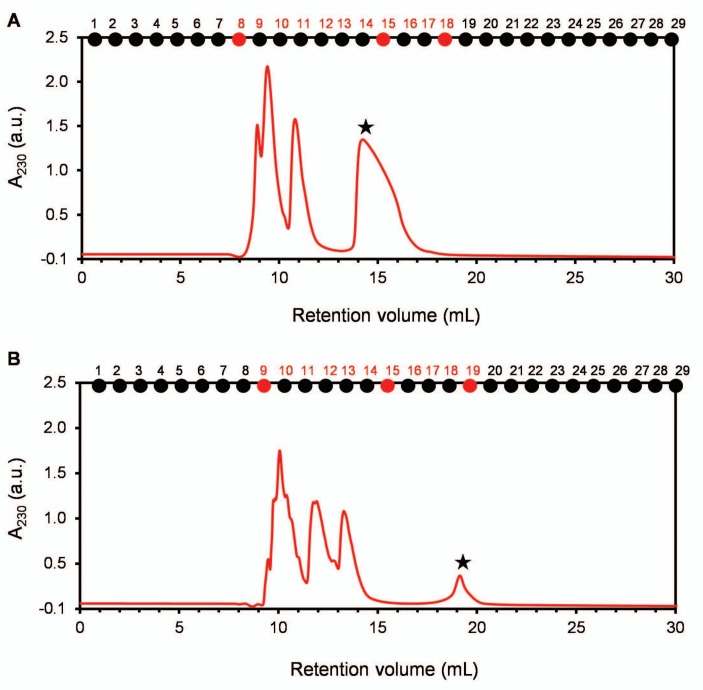


Figure 4.5: *Typical chromatograms for* $\boldsymbol{C^P_4\text{-}D^K}$ *and* $\boldsymbol{T_9\text{-}C^P_4\text{-}D^K}$. *(A)* $\boldsymbol{C^P_4\text{-}D^K}$ *and (B)* $\boldsymbol{T_9\text{-}C^P_4\text{-}D^K}$. *The upper dots indicate collected fractions, the numbers in red indicate the fractions analyzed by SDS-PAGE, and the red upper dots indicate fractions analyzed by MALDI-TOF. The star indicates the fractions containing the intact protein polymer.*

The last peaks (indicated by a star) probably correspond to intact protein, since it binds to the column the strongest and should be retained on the column until the salt concentration will be sufficiently high to allow elution. The peaks detected at early stage of elution represent the degraded species. The more the $\mathbf{D^K}$ module is degraded, the fewer positive charges it will contain, and the less it will bind to the column. Hence, the first peak will probably correspond to the peptide with the most truncated version of $\mathbf{D^K}$ module.

The eluted fractions corresponding to the various peaks were analysed by SDS-PAGE (Fig 4.6A). Elution fraction 8 of $\mathbf{C^P{}_4\text{-}D^K}$ showed only one protein band in SDS-PAGE, migrating to the position of the ~188 kDa marker band. From fraction 9 to fraction 14, more bands were present, and the intensity of the band at ~ 98 kDa was gradually increasing while the upper bands were less and less visible in each subsequent fraction. In fractions 15 to 18, only one band was observed at ~ 98 kDa. The fraction 8, 15 and 17 were analysed with MALDI-TOF (Fig 4.7A). For both fractions 15 and 17, MALDI-TOF showed a major peak at *m/z* 39,382, which is in good agreement with the expected molecular weight of intact $\mathbf{C^P{}_4\text{-}D^K}$. In fraction 8 the major peak was detected at *m/z* 36,970 indicating degraded form of $\mathbf{C^P{}_4\text{-}D^K}$.

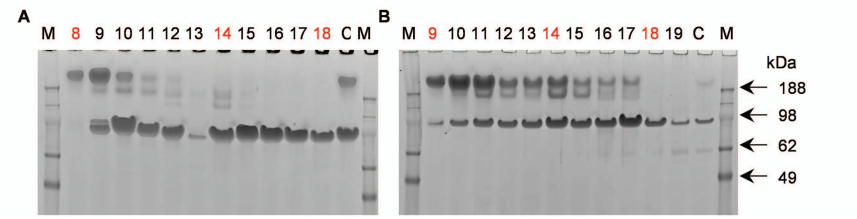


Figure 4.6: *SDS-PAGE of the fractions (A)* $\boldsymbol{C^P{}_4\text{-}D^K}$ *and (B)* $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$ *pulled from cation-exchange chromatography. The lines 9-18 correspond respectively to the fraction number, line C is a bulk sample as control and line M is a protein molecular weight maker. The red numbers indicate fractions chosen for MALDI-TOF analysis.*

From the SDS-PAGE of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$, we can conclude that this separation apparently did not proceed as efficiently as in case of the diblock (Fig 4.6B). The SDS-PAGE for fraction 9 showed two bands at the level corresponding to the marker bands of ~188 kDa and ~98 kDa, and in the fractions from 10 to 17 multiple bands with intermediate migration behaviour were present. Again, the intensity of the band at ~ 98 kDa was increasing whereas the upper bands were less visible in each subsequent fraction, but over a much wider elution range than in the case of $\mathbf{C^P{}_4\text{-}D^K}$. In fractions 18 and 19, only one band was observed at ~ 98 kDa. The fractions 9, 14 and 18 were analysed with MALDI-TOF

(Fig 4.7B). In the fraction 9, MALDI-TOF showed a major peak at *m/z* 39,440 and several minor peaks on both sides of the major peak. All these peaks corresponded to molecular masses lower than that of intact $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$, and thus probably represented degraded forms of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$. A major peak corresponding to the molecular mass of the intact triblock polymer was detected in fraction 14 and 18. Fraction 14, however, showed three additional peaks to the left of the major peak, which indicates that the intact product is contaminated with molecules of lower mass, probably degraded $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ species.

By analysis of chromatograms, we attempted to estimate the amounts of the intact and degraded protein in the original sample. The fraction of the intact protein was taken to be the ratio of the area under the last peak divided by the total area under all peaks. The fractions of the degraded species were derived accordingly. For $\mathbf{C^P{}_4\text{-}D^K}$, the fully intact form of protein polymer constituted 43% of the total sample. The $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ protein polymer could not be properly separated using SEC, therefore we could not obtain correct distribution for the intact and degraded species. Referring to MALDI-TOF spectrum, the fraction 14 contained both intact and degraded species (Fig 4.7B).

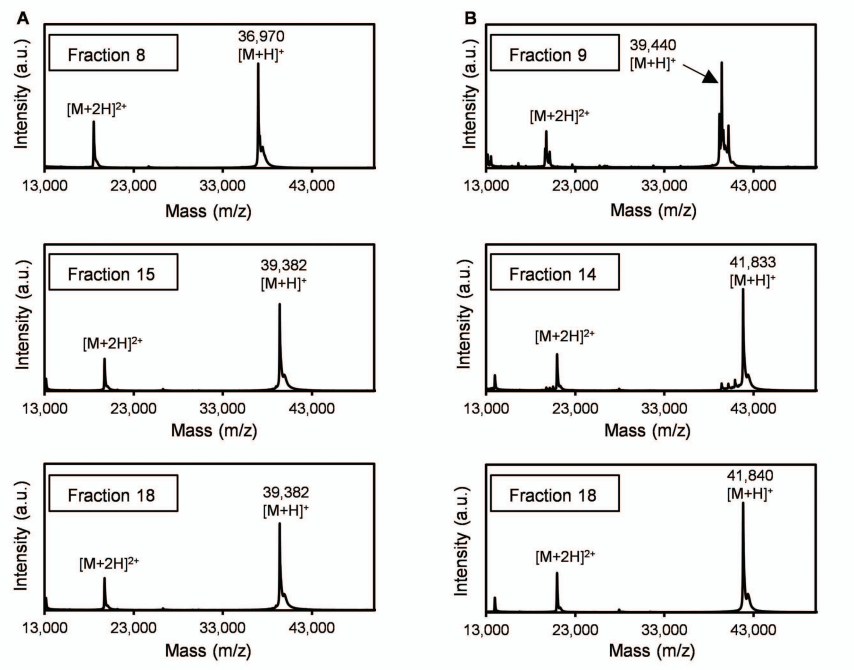


Figure 4.7: *MALDI-TOF of selected HPLC elution fractions of (A)* $\boldsymbol{C^P{}_4\text{-}D^K}$ *and (B)* $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$.

Most likely, the chromatography of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ is influenced by the presence of the $\mathbf{T_9}$ block. This module can form trimeric $\mathbf{(T_9)_3}$ intermolecular cross-links, and therefore the separation of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ is probably not as straightforward as that of $\mathbf{C^P{}_4\text{-}D^K}$. It is likely that part of the intact $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ molecules elutes too early, viz. when they are incorporated in mixed trimers consisting of both intact and degraded molecules.

The SDS-PAGE and MALDI-TOF analysis indicates that the formation of triple helixes by $\mathbf{T_9}$ may indeed interfere with the process of separation, whereas the present cation exchange chromatography procedure is a suitable method to separate intact $\mathbf{C^P{}_4\text{-}D^K}$ from degraded or contaminating molecules. The purification process for $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ needs to be optimized by preventing $\mathbf{T_9}$ self-assembly, for instance by adding urea to the application and elution buffers, [39] or by working at elevated temperature (e.g. 50 °C) [36]. Additional concern is the contamination of the samples with the unattached fragments of the $\mathbf{D^K}$ module. Although the chances are low, we cannot rule out the possibility that the cleaved forms of the $\mathbf{D^K}$ module might co-purify in AS precipitation with the intact and partly intact molecules and be present in the bulk sample. Such small peptides are not in the detection range of SDS-PAGE and MALDI-TOF used here, therefore analysis in this mass range (~2 kDa) should be performed.

4.3.4. Analysis of hydrogel formation

To test the functionality of the heterodimer-forming modules incorporated into protein-based polymers, we used Dynamic Light Scattering-based (DLS-based) microrheology to study gel formation in 1:1 mixtures of $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ at concentrations of 40 g L^{-1} and 80 g L^{-1}. DLS was used to track the diffusion of small amounts of latex tracer particles (diameter 112 nm) added to the protein solutions. The relative viscosity (η/η_0) of the protein polymer solutions was taken to be equal to the ratio D_0/D of (z-averaged) diffusion constants, where D_0 is the diffusion constant of the tracer particles in the aqueous solvent and D is the diffusion constant of the tracer particles in the protein polymer solutions.

Relative viscosities were measured as a function of temperature, and results are shown in Fig 4.8. For both 1:1 mixtures and pure components $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$, no gelation was observed at a total protein polymer concentration of 40 g L^{-1}. A small but noticeable increase in viscosity at low temperatures was found for the 1:1 mixtures. At a concentration of 80 g L^{-1}, both the 1:1 mixed system and the $\mathbf{T_9\text{-}C^P{}_4\text{-}D^E}$ system formed a hydrogel at temperatures

below 35 °C, whereas the **T$_9$-C$^P_4$-D^K** system only showed an increase in viscosity below 35 °C.

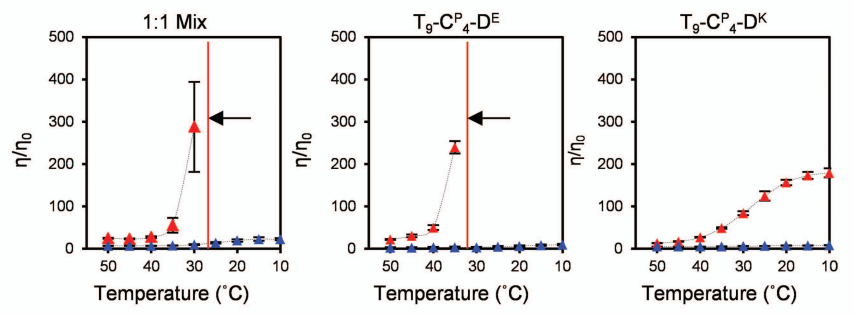


Figure 4.8: *Relative viscosity (η/η_0) versus temperature (˚C) measured for protein solutions (A) 1:1 mixture of **T$_9$-C$^P_4$-D^E** and **T$_9$-C$^P_4$-D^K**, (B) **T$_9$-C$^P_4$-D^E** or (C) **T$_9$-C$^P_4$-D^K**. The relative viscosity (η/η_0) of the protein solutions versus temperature. The blue triangles represent the solutions at total protein concentration 40 g L^{-1} and the red triangles at concentration 80 g L^{-1}.*

The formation of collagen-like triple helices (trimers) by the **T$_9$** block is dependent on both temperature and concentration [40]. For example, a previously-studied polymer, **T$_9$-C$^P_4$-T$_9$**, forms hydrogels with a clearly apparent stiffness only at concentrations higher than 40 g L^{-1} (~1 mM) and temperatures lower than 30 °C [40]. Since **T$_9$-C$^P_4$-T$_9$** contains two **T$_9$** blocks, 1:1 mixture of **T$_9$-C$^P_4$-D^E** and **T$_9$-C$^P_4$-D^K** must contain at least 2 mM of **T$_9$** blocks, equivalent to a total protein polymer concentration of 80 g L^{-1}, in order for the mixture to form a network through the association of both the **T$_9$** blocks and the **D^E/D^K** heterodimers, and this is also what we observe. If the **D^E** and the **D^K** modules would only form heterodimers, the hydrogels should not be formed in the control samples of pure components at 80 g L^{-1}, but only in the 1:1 mixtures of **T$_9$-C$^P_4$-D^E** and **T$_9$-C$^P_4$-D^K**. However, gelation occurred both for the 1:1 mixed system, and for the **T$_9$-C$^P_4$-D^E** system, suggesting that for this case, there must also be strong homotypic **D^E/D^E** associations. Likewise, the strong increase in the viscosity of the 80 g L^{-1} **T$_9$-C$^P_4$-D^K** system points to homotypic **D^K/D^K** associations.

Next, shear moduli were determined using oscillatory shear measurements. Protein polymers were dissolved in 10 mM sodium phosphate buffer (pH 7) and mixed in a 1:1 ratio at total protein concentrations 160 g L^{-1} (~4 mM). Prior to the measurement, the sample was heated at 50 °C for 20 min. and then introduced into the preheated geometry. The system was subsequently quenched to 20 °C. The process of gelation was followed in time by online

measurement of the storage (G′) modulus at 1 Hz at a strain of 0.1%. This measurement was performed at constant temperature (20 °C) for 3.5 h (Fig 4.8A). Subsequently, the temperature was increased by 10 °C every 3.5, up to 50 °C (Fig 4.9B). At 20 °C, the mixture formed a gel with a storage modulus of 905 Pa after 3.5h. An increase of temperature resulted in a decrease of storage modulus and at 40 °C, the storage modulus was zero. Hence, the mixed gels show a temperature dependence that is very similar to that of the gels of the **T_9-C^P_4-T_9** triblocks [40].

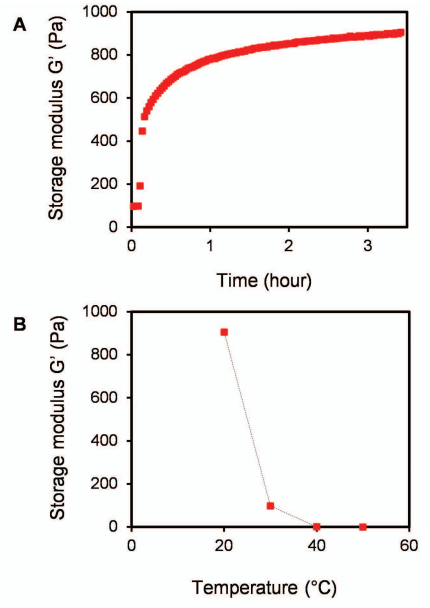


Figure 4.9: *Storage modulus (G') of 160 g L^{-1} gels made of a 1:1 mixture of T_9-C^P_4-D^E and T_9-C^P_4-D^K. (A) Time resolved gel formation at 20 °C; (B) temperature dependence.*

Further, we investigated gel formation of the 1:1 **T_9-C^P_4-D^E/T_9-C^P_4-D^K** mixture at a lower concentration of 80 g L^{-1}. This time we also tested **T_9-C^P_4-D^E** alone and **T_9-C^P_4-D^K** alone at a concentration of 80 g L^{-1}. The sample pre-treatment and measurement settings were the same as for the 1:1 mixed system at 160 g L^{-1}. The increase of the modulus was monitored for 5 h at a constant temperature (20 °C). After 5 h of measurement, the **T_9-C^P_4-D^K** system had not gelled, whereas the **T_9-C^P_4-D^E** sample and the **T_9-C^P_4-D^E/T_9-C^P_4-D^K** mixture reached storage moduli of 104 Pa and 55 Pa, respectively (Fig 4.10). The higher value found for **T_9-C^P_4-D^E**, as compared to the **T_9-C^P_4-D^E/T_9-C^P_4-D^K** mixture, might indicate that only the formation of **D^E** homodimers contributes

to gel formation in both samples. Hence, results for the mixed system do not yet indicate that there is an interaction between $\mathbf{D^E}$ and $\mathbf{D^K}$. However, our results do show that fully intact $\mathbf{D^E}$ itself is an interesting module to be studied for self-assembling systems. Furthermore, it should also be taken into account that for the $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ protein polymers, a certain fraction (which we did not manage to quantify here) has completely or partly truncated $\mathbf{D^K}$ modules. This means that the sample has lower concentration of potentially gel-forming molecules and it also might contain the free $\mathbf{D^K}$ fragments capable of binding to the remaining intact molecules. This causes a seemingly (i.e. apparent or falsely) low tendency of the polymers to take part in the bond formation. Therefore, a more definitive test for the heterodimeric association of the $\mathbf{D^K}$ and $\mathbf{D^E}$ modules, will need to be done with 100% intact $\mathbf{T_9\text{-}C^P{}_4\text{-}D^K}$ protein polymers.

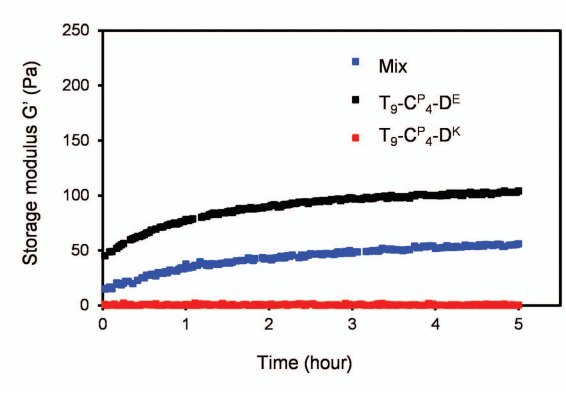


Figure 4.10 *Time course of increase in storage G' modulus for 80 g L^{-1} gels made of $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^E}$ (black), $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$ (red), and a 1:1 mixture of $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^E}$ and $\boldsymbol{T_9\text{-}C^P{}_4\text{-}D^K}$ (blue).*

4.3.5. Detection of heterodimeric association of $\mathbf{D^E}$ and $\mathbf{D^K}$ modules

A first attempt at directly measuring the heterodimeric association of the $\mathbf{D^E}$ and $\mathbf{D^K}$ modules in the context of protein polymers was done using isothermal titration calorimetry. A concentrated stock solution of $\mathbf{C^P{}_4\text{-}D^K}$ of 2.9 mM was used as the titrant. Ligand aliquots of 4 µl were titrated into 1.4 ml of 0.2 mM solution of $\mathbf{C^P{}_4\text{-}D^E}$ protein solution inside the calorimeter cell under continuous stirring. Unfortunately, for these conditions, we found only a very small exothermic signal, with a preliminary analysis of the data giving a very high value for the resulting dissociation constant (K_d on the order of mM). From the work of Litowski et al [20], the dissociation constant of the free peptide pair should be very much lower (K_d value around 70 nM).

There may be a number of reasons why no strong exothermic signal was found in these preliminary ITC measurements, foremost of which the dissociation of homomultimers occurring in parallel to the association of heterodimers might cause the net heat effect to be muted.

4.4. Concluding remarks

We have shown that polymers containing the $\mathbf{D^E}$ and the $\mathbf{D^K}$ coiled coils can be efficiently produced in *Pichia pastoris.* The protein polymers containing the $\mathbf{D^K}$ coiled coil were partly degraded. This degradation could not be resolved by performing the fermentation at pH 5, or by using a yapsin 1 protease disruptant. Using cation exchange chromatography, we did separate the intact $\mathbf{C^P_4\text{-}D^K}$ protein polymer from the degraded protein polymers and we estimated its content in the bulk sample. Although homomerization has not been previously reported for the free $\mathbf{D^E}$ and $\mathbf{D^K}$ peptides at pH 7, our rheology data indicate strong homotypic associations, especially for $\mathbf{D^E}$, when fused to $\mathbf{T_9\text{-}C^P_4}$. Note that homotrimer-formation has been reported previously for the $\mathbf{D^E}$ module, however only at pH 5 (and at low concentrations) [41].

Here we have not yet been able to demonstrate the direct interaction of the $\mathbf{D^E}$ and the $\mathbf{D^K}$ modules in the context of larger protein polymers. This may be in part due to competition by the strong homotypic associations that we have found, and in part by the fact that the protein polymers with the $\mathbf{D^K}$ modules were partly degraded.

Possibly the $\mathbf{D^E}$ module (produced fully intact in *P. pastoris*) might be developed into a homomultimer-forming block for self-assembling protein polymers. In past, several interesting protein-based architectures have been developed using homotypic interaction between coiled coils. The Tirrell group reported on the hydrogel with the tunable erosion rate that was formed by aggregation of coiled coils [42]. A very interesting example of the use of homodimeric coiled-coils is the provided by the self-assembling coiled-coil tetrahedrons described by group of Jerala [43]. Therefore, a further analysis of the $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ and $\mathbf{C^P_4\text{-}D^E}$ protein polymers is certainly interesting, and may reveal new and useful features of the $\mathbf{D^E}$ modules, when used as a self-assembly domain in the context of larger protein polymer constructs.

Reference

1. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2011**, 12, 269-289.
2. Gupta, P.; Nayak, K. K. *Polym Eng Sci* **2015**, 55, 485-498.
3. Lawrence, B. D.; Cronin-Golomb, M.; Georgakoudi, I.; Kaplan, D. L.; Omenetto, F. G. *Biomacromolecules* **2008**, 9, 1214-1220.
4. Qiu, Y.; Park, K. *Adv Drug Delivery Rev* **2012**, 64, 49-60.
5. Gomes, S.; Leonor, I. B.; Mano, J. F.; Reis, R. L.; Kaplan, D. L. *Prog Polym Sci* **2012**, 37, 1-17.
6. Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater* **2014**, 10, 1542-1557.
7. Woolfson, D. N. *Pept Sci* **2010**, 94, 118-127.
8. Werten, M. W.; Teles, H.; Moers, A. P.; Wolbert, E. J.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009**, 10, 1106-1113.
9. Olsen, D.; Yang, C.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perälä, M.; Hämäläinen, E.-R.; Jarvinen, M. *Adv Drug Delivery Rev* **2003**, 55, 1547-1567.
10. Schipperus, R.; Teeuwen, R. L.; Werten, M. W.; Eggink, G.; de Wolf, F. A. *Appl Microbiol Biotechnol* **2009**, 85, 293-301.
11. Rodríguez-Cabello, J. C.; Arias, F. J.; Rodrigo, M. A.; Girotti, A. *Adv Drug Delivery Rev* **2015**.
12. Werten, M. W.; Moers, A. P.; Vong, T.; Zuilhof, H.; van Hest, J. C.; de Wolf, F. A. *Biomacromolecules* **2008**, 9, 1705-1711.
13. Dinjaski, N.; Kaplan, D. L. *Curr Opin Biotechnol* **2016**, 39, 1-7.
14. Hosseinkhani, H.; Hong, P.-D.; Yu, D.-S. *Chem Rev* **2013**, 113, 4837-4861.
15. Kopeček, J. *Eur J Pharm Sci* **2003**, 20, 1-16.
16. Desai, M. S.; Lee, S. W. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2015**, 7, 69-97.
17. Zhang, S.; Marini, D. M.; Hwang, W.; Santoso, S. *Curr Opin Chem Biol* **2002**, 6, 865-871.
18. Beun, L. H.; Storm, I. M.; Werten, M. W.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 3349-3357.
19. Domeradzka, N. E.; Werten, M. W.; de Wolf, F. A.; de Vries, R. *Curr Opin Biotechnol* **2016**, 39, 61-67.
20. Litowski, J. R.; Hodges, R. S. *J Biol Chem* **2002**, 277, 37272-37279.
21. Werten, M. W.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng* **2001**, 14, 447-454.

22. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. *Gene* **1989**, 77, 51-59.
23. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 699-706.
24. Werten, M.; van den Bosch, T.; Wind, R.; Mooibroek, H.; de Wolf, F. *Yeast* **1999**, 15, 1087-1096.
25. Werten, M. W.; de Wolf, F. A. *Appl Environ Microbiol* **2005**, 71, 2310-2317.
26. Zhang, W.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M. *Pap Biotech* **2000**, 18.
27. Appel, J.; Akerboom, S.; Fokkink, R. G.; Sprakel, J. *Macromol. Rapid Commun.* **2013**, 34, 1284-1288.
28. Cingil, H. E.; Rombouts, W. H.; van der Gucht, J.; Cohen Stuart, M. A.; Sprakel, J. *Biomacromolecules* **2014**, 16, 304-310.
29. Moers, A. P.; Wolbert, E. J.; de Wolf, F. A.; Werten, M. W. *T. J Biotechnol* **2010**, 146, 66-73.
30. Domeradzka, N. E.; Werten, M. W.; de Vries, R.; de Wolf, F. A. *Biotechnol Bioeng* **2015**.
31. Dunker, A.; Rueckert, R. R. *J. Biol. Chem.* **1969**, 244, 5074-5080.
32. Zhang, W.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M. *Biotechnol Bioeng* **2000**, 70, 1-8.
33. Chang, S.-W.; Shieh, C.-J.; Lee, G.-C.; Akoh, C. C.; Shaw, J.-F. *J. Mol. Microbiol. Biotechnol.* **2006**, 11, 28-40.
34. Daly, R.; Hearn, M. T. *J. Mol. Recognit.* **2005**, 18, 119-138.
35. Kang, H.; Kim, S.; Choi, E.-S.; Rhee, S.-K.; Chung, B. *Appl. Microbiol. Biotechnol.* **1998**, 50, 187-192.
36. Silva, C. I.; Teles, H.; Moers, A. P.; Eggink, G.; de Wolf, F. A.; Werten, M. W. *Biotechnol Bioeng* **2011**, 108, 2517-2525.
37. Hartmann, L.; Börner, H. *Adv. Mater.* **2009**, 21, 3425-3431.
38. Sinha, J.; Plantz, B. A.; Inan, M.; Meagher, M. M. *Biotechnol. Bioeng.* **2005**, 89, 102-112.
39. Franke, W. W.; Schiller, D. L.; Hatzfeld, M.; Winter, S. *PANS* **1983**, 80, 7113-7117.
40. Skrzeszewska, P. J.; de Wolf, F. A.; Werten, M. W.; Moers, A. P.; Stuart, M. A. C.; van der Gucht, J. *Soft Matter* **2009**, 5, 2057-2062.
41. Apostolovic, B.; Klok, H.-A. *Biomacromolecules* **2008**, 9, 3173-3180.
42. Shen, W.; Zhang, K.; Kornfield, J. A.; Tirrell, D. A. *Nat Mater* **2006**, 5, 153-158.

43. Gradišar, H.; Božič, S.; Doles, T.; Vengust, D.; Hafner-Bratkovič, I.; Mertelj, A.; Webb, B.; Šali, A.; Klavžar, S.; Jerala, R. *Nat Chem Biol* **2013**, 9, 362-366.

5 Production in *Pichia pastoris* of complementary protein-based polymers with heterodimer-forming WW and PPxY domain

Abstract

Specific coupling of *de novo* designed recombinant protein polymers for the construction of precisely structured nanomaterials is of interest for applications in biomedicine, pharmaceutics and diagnostics. An attractive coupling strategy is to incorporate specifically interacting peptides into the genetic design of the protein polymers. An example of such interaction is the binding of particular proline-rich ligands by so-called WW-domains. In this study, we investigated whether these domains can be produced in the yeast *Pichia pastoris* as part of otherwise non-interacting protein polymers, and whether they bring about polymer coupling upon mixing. We constructed two variants of a highly hydrophilic protein-based polymer that differ only in their C-terminal extensions. One carries a C-terminal WW domain, and the other a C-terminal proline-rich ligand (PPxY). Both polymers were produced in *P. pastoris* with a purified protein yield of more than 2 g L^{-1} of cell-free broth. The proline-rich module was found to be O-glycosylated, and uncommonly a large portion of the attached oligosaccharides was phosphorylated. Glycosylation was overcome by introducing a Ser → Ala mutation in the PPxY peptide. Tryptophan fluorescence monitored during titration of the polymer containing the WW domain with either the glycosylated or nonglycosylated PPxY-containing polymer revealed binding. The complementary polymers associated with a K_d of ~3 μM, regardless of glycosylation state of the PPxY domain. Binding was confirmed by isothermal titration calorimetry, with a K_d of ~9 μM. This chapter presents a blueprint for the production in *P. pastoris* of protein polymers that can be coupled using the noncovalent interaction between WW domains and proline-rich ligands. The availability of this highly specific coupling tool will hereafter allow us to construct various supramolecular structures and biomaterials.

Published as: Domeradzka, N.E.; Werten, M.W.T.; de Vries, R. and de Wolf, F.A. *Microb Cell Fact* **2016**, *15(1), 1*.

5.1. Introduction

Protein-based block copolymers (or protein polymers, for short) are *de novo* designed polypeptides that consist of different blocks, where each of the blocks adopts a specific conformation and fulfills a specific function [1]. Self-assembling protein polymers are being intensively explored for use as biomaterials for drug encapsulation, controlled drug release, tissue engineering, tissue augmentation, and biosensors [2-8]. Protein polymers are produced via recombinant DNA technology, and thus, aside from possible problems with the biological production, in principle have a defined sequence, mass, and chemical composition. They are biodegradable, often biocompatible, and easy to functionalize by inclusion of bioactive sequences in the genetic design [1,7].

In early work, the sequences of protein polymers were typically inspired by naturally occurring self-assembling structural proteins such as silk, elastin, collagen, and resilin, with inherent attractive properties as soft materials [9,10]. The repertoire of useful motifs for self-assembly has been expanded over the years with sequences designed or modified using molecular modelling. Several self-assembling modules can be combined into a multifunctional polypeptide block copolymer that can spontaneously self-organize via non-covalent interactions into nanostructures such as micelles, fibrils, or hydrogels [11-13]. However, peptide blocks that can facilitate the assembly of supramolecular structures with particularly high precision are still highly sought after in biomaterial science [14-16].

Our group has successfully produced several protein polymers at high yield using the yeast *Pichia pastoris* [17-23]. Many of these are block copolymers that self-organize into stimulus-responsive supramolecular structures. We currently aim to expand our library of functional protein blocks with modules that allow highly precise heterodimerization, in an effort to gain more control over the self-assembly of supramolecular structures. A pair of block copolymers fitted with complementary interacting modules could then self-assemble into higher order structures upon mere mixing of the two [24]. Besides providing crosslinks in self-assembled structures, heterodimerizing modules can also be used for incorporation of functionalities such as growth factors, antimicrobial peptides, and cell-adhesive peptides [25].

The so-called WW domain is found in various natural proteins and binds particular proline-rich peptides with a high degree of specificity [26]. As its name suggests, the amino acid sequence of the WW domain contains two highly conserved tryptophans. The domain consists of a slightly bent three-stranded antiparallel β-sheet, the concave side of which forms a binding pocket for the

proline-rich ligand [27]. Wong Po Foo et al. successfully used the interaction between two different WW domains (CC43 and Nedd4.3) and a so-called group I proline-rich peptide (PPxY) to generate two-component hydrogels [28]. Here, we investigate the separate incorporation of this same PPxY peptide derived from p53-binding protein-2 [29] and another WW domain (WWP1-1; one of three WW domains present in the human ubiquitin ligase homolog WWP1 [29]) at the C-terminus of the $\mathbf{C^P_4}$ protein polymer previously developed by us [18]. The $\mathbf{C^P_4}$ polymer (formerly referred to as 'P4') consists of four identical copies of a 99 amino acid long, highly hydrophilic random coil polypeptide [18]. We report the high-yield secretory production of these polymers in *Pichia pastoris* and their characterization. Undesired O-glycosylation of the PPxY peptide was overcome, and the polymers interacted as intended. We have thus expanded our toolkit towards the creation of well-defined supramolecular topologies for new biomaterials.

5.2. Materials and methods

5.2.1. Construction of expression vectors and strains

The double-stranded gene fragments encoding $\mathbf{D^{WW}}$ and $\mathbf{D^{PPxY}}$ were assembled via overlap extension PCR [50] from the oligonucleotides shown in Supplementary data 1: Table 5.1.S1. The gene fragments were digested with *Xho*I/*Eco*RI and cloned into *Xho*I/*Eco*RI-digested vector pMTL23Δ*Bsa*I [51], in order to obtain two vectors pMTL23Δ*Bsa*I-$\mathbf{D^{WW}}$ and pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY}}$. The vector pMTL23-$\mathbf{C^P_4}$ contains the sequence encoding $\mathbf{C^P_4}$ (previously referred to as 'P4') [18], and was opened at the 3' end of the $\mathbf{C^P_4}$ gene with *Van*91I/*Eco*RI. The newly prepared constructs pMTL23Δ*Bsa*I-$\mathbf{D^{WW}}$ and pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY}}$ were digested with *Dra*III/*Eco*RI to release inserts $\mathbf{D^{WW}}$ and $\mathbf{D^{PPxY}}$. The inserts were ligated into the opened pMTL23-$\mathbf{C^P_4}$ vector, resulting in pMTL23-$\mathbf{C^P_4}$-$\mathbf{D^{WW}}$ and pMTL23-$\mathbf{C^P_4}$-$\mathbf{D^{PPxY}}$.

A Ser12 → Ala mutant of the $\mathbf{D^{PPxY}}$ module was prepared by annealing of a pair of largely complementary oligos (Supplementary data 1: Table 5.1.S1), and is referred to as $\mathbf{D^{PPxY*}}$.

This double-stranded adapter with *Dra*III/*Eco*RI overhangs was ligated into vector pMTL23-$\mathbf{C^P_4}$ previously digested with *Van*91I/*Eco*RI (at the 3' end of the $\mathbf{C^P_4}$ gene), resulting in pMTL23Δ*Bsa*I-$\mathbf{C^P_4}$-$\mathbf{D^{PPxY*}}$. The $\mathbf{C^P_4}$-$\mathbf{D^{WW}}$, $\mathbf{C^P_4}$-$\mathbf{D^{PPxY}}$, and $\mathbf{C^P_4}$-$\mathbf{D^{PPxY*}}$ inserts were then released with *Xho*I/*Eco*RI and cloned into the likewise-digested *Pichia pastoris* expression vector pPIC9 (ThermoFisher, Bleiswijk, The Netherlands). This resulted in the construction of the vectors: pPIC9-$\mathbf{C^P_4}$-$\mathbf{D^{WW}}$, pPIC9-$\mathbf{C^P_4}$-$\mathbf{D^{PPxY}}$ and pPIC9-$\mathbf{C^P_4}$-$\mathbf{D^{PPxY*}}$,

respectively. The vectors were linearized with *Sal*I to target for integration at the *his4* locus. Transformation of *P. pastoris* GS115 by electroporation and selection of Mut^+ transformants were performed as described previously [17]

5.2.2. Fermentation

The fermentation setup consisted of a 2.5-L Bioflo 3000 stirred-tank bioreactor (New Brunswick Scientific, Nijmegen, The Netherlands) interfaced with BioCommand Software (New Brunswick Scientific, Nijmegen, The Netherlands) and a homebuilt methanol sensor-controller. The fermentations were performed as described previously [19], as follows. A starting volume of 1.25 L minimal basal salts medium [52] was used. The cultures were always inoculated with precultures grown to similar OD_{600}. Growth temperature was 30 °C, and the pH was controlled at 3.0 throughout the entire fermentation. The air was supplemented with 20% (v/v) oxygen during the glycerol fed-batch phase and the methanol fed-batch phase. During the latter protein production phase, lasting two days, methanol levels were kept at 0.2% (w/v). Wet biomass was typically ~150 g L^{-1} at the end of the glycerol fed-batch phase, and ~500 g L^{-1} at the end of the fermentation. After harvesting, cells were removed from the broth by centrifugation for 20 min. at 15,000 × g (RT), followed by microfiltration.

5.2.3. Protein purification

Purification of all protein polymers was done by ammonium sulfate precipitation essentially as described [19], except that heating of the supernatant and acetone precipitation were omitted. Shortly, medium salts were removed by raising the pH of the cell-free broth to 8.0 with sodium hydroxide, followed by 30 min. of centrifugation at 20,000 × g (RT). The protein was precipitated from the supernatant by addition of ammonium sulfate to 40% of saturation, followed by incubation on ice for 30 min. and centrifugation for 30 min. at 20,000 × g (4 °C). The protein pellet was resuspended in Milli-Q water and precipitated using ammonium sulfate at 40% of saturation as before. The pellet was then resuspended in Milli-Q water, desalted by extensive dialysis against Milli-Q water, and finally lyophilized.

5.2.4. SDS-PAGE

SDS-PAGE was performed using the NuPAGE Novex System (ThermoFisher, Bleiswijk, The Netherlands) with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 pre-stained molecular mass markers. Prior to electrophoresis, all samples were heated for 10 min. at 70 °C in NuPAGE LDS Sample Buffer with NuPAGE Sample Reducing Agent, as per manufacturer's recommendations for denaturing and reducing PAGE. Gels were stained using Coomassie SimplyBlue SafeStain (ThermoFisher, Bleiswijk, The Netherlands).
For detection of glycosylated proteins, SDS-PAGE gels were stained using Periodic acid-Schiff staining [33]. The gel was incubated for 1 h in 12.5% TCA, 1 h in 1% periodic acid/3% acetic acid, 1 h in 15% acetic acid (replaced every 10 min.), and 1 h at 4 °C in the dark in Schiff's reagent (Sigma-Aldrich, Zwijndrecht, The Netherlands). The gel was then washed two times for 5 min. in 0.5% sodium bisulfite and destained in 7% acetic acid.

5.2.5. Treatment of proteins with α-mannosidase or phosphatase

For α-mannosidase digestions, 30 µg of glycoprotein was incubated for 24 h at 37 °C under mild agitation with 0.9 U of jack bean α(1-2,1-3,1-6) mannosidase (Sigma-Aldrich, Zwijndrecht, The Netherlands) in 60 µL of 20 mM sodium acetate, 0.4 mM zinc chloride, pH 5. Dephosphorylation involved incubation of 30 µg of glycoprotein for 24 h at 37 °C under mild agitation with 60 U of calf intestinal alkaline phosphatase (NEB, Ipswich, MA) in 60 µL of 50 mM Tris-HCl, 10 mM magnesium chloride, pH 8.5.
For consecutive α-mannosidase/phosphatase digestions, the enzyme after each step was inactivated by heating for 15 min. at 100 °C, followed by desalting using Micro Bio-Spin P-6 columns (Bio-Rad, Veenendaal, The Netherlands). To allow mass spectrometric analysis of the reaction products, it was verified that such analysis of enzyme-only digestions revealed no significant peaks in the relevant ~38-41 kDa range (not shown).

5.2.6. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was performed using an ultrafleXtreme mass spectrometer (Bruker, Leiderdorp, The Netherlands). Proteins were desalted using Micro Bio-Spin P-6 columns (Bio-Rad, Veenendaal, The Netherlands), and samples were prepared by the dried droplet method on a 600 µm AnchorChip target (Bruker, Leiderdorp, The Netherlands), using 5 mg mL^{-1} 2,5-dihydroxyacetophenone, 1.5 mg mL^{-1} diammonium hydrogen citrate, 25%

(v/v) ethanol and 3% (v/v) trifluoroacetic acid as matrix. Spectra were derived from ten 500-shot (1,000 Hz) acquisitions taken at non-overlapping locations across the sample. Wide mass-range measurements were made in the positive linear mode, with ion source 1, 25.0 kV; ion source 2, 23.3 kV; lens, 6.5 kV; pulsed ion extraction, 680 ns. Detailed analyses of glycoproteins in the ~38-41 kDa range were done with ion source 1, 20.0 kV; ion source 2, 18.4 kV; lens, 6.2 kV; pulsed ion extraction, 450 ns, and spectra were derived from ten 1,000-shot (1,000 Hz) acquisitions. Protein Calibration Standard II (Bruker, Leiderdorp, The Netherlands) was used for external calibration.

5.2.7. Steady-state fluorescence spectroscopy

Steady-State Fluorescence Spectroscopy was performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Amstelveen, The Netherlands), monitoring the intrinsic fluorescence emission of the tryptophan residues in the WW domain at 340 nm, with excitation at 295 nm. Proteins were dissolved overnight in 10 mM sodium phosphate buffer (pH 7) at RT. The binding assays for both couples $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ / $\mathbf{C^{P}_{4}\text{-}D^{WW}}$, and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ / $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ were conducted in triplicate at RT. A 500 µL aliquot of 10 µM $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ was pipetted into a quartz fluorescence cuvette (Sigma-Aldrich, Zwijndrecht, The Netherlands). A 500 µM solution of ligand was stepwise added to the cuvette, up to a final ligand concentration of ~24 µM. The time interval between additions was 30 min. The final volume of added ligand solution did not exceed 5% of the starting volume of $\mathbf{C^{P}_{4}\text{-}D^{WW}}$. Curve fitting for K_d determination was done on averaged data of triplicate titrations.

5.2.8. Isothermal titration calorimetry

ITC was conducted on a MicroCal VP-ITC (Malvern Instruments, Malvern, United Kingdom) at 25 °C. All purified protein polymers were dissolved in 10 mM sodium phosphate buffer, pH 7 and filtrated with 0.2 µm Minisart NML Syringe Filters (Sigma-Aldrich, Zwijndrecht, The Netherlands). Prior to titration, each protein polymer solution was degassed under vacuum for 60 min. at RT. The ligand concentration of $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ or $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ in the titration syringe was 2.9 mM. Each titration consisted of 63 injections at an interval of 250 sec. Ligand aliquots of 4 µL were titrated into 1.4 mL of a 200 µM $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ protein solution inside the calorimeter cell under continuous stirring at 329 rpm. Data obtained from the injection of ligand molecules into 1.4 mL of 10 mM PBS buffer (Fig. 5.1.S2) were subtracted as blanks from the experimental data before the data were analyzed using MicroCal Origin

Software (Malvern Instruments, Malvern, United Kingdom). Titrations were performed in triplicate and K_d values were averaged after curve fitting.

5.3. Results and discussion

5.3.1. Protein production and purification

The WW and PPxY domains are referred to hereafter as $\mathbf{D^{WW}}$ and $\mathbf{D^{PPxY}}$, respectively (D for Dimerization). For amino acid sequences see Table 5.1.

Table 5.1: *Amino acid sequences of the C-terminal D^{ww}, D^{PPxY} and D^{PPxY*} modules. All sequences end C-terminally in the cloning-derived amino acid sequence PAGG (not indicated).*

Module	Amino acid sequence
D^{WW}	LPSGWEQRKDPHGRTYYVDHNTRTTTWERPQPLPPGA
D^{PPxY}	EYPPYPPPPYPSG
D^{PPxY*}	EYPPYPPPPYPAG

The domains were cloned at the 3' end of the gene encoding the previously reported $\mathbf{C^P_4}$ protein polymer [18]. The encoded proteins $\mathbf{C^P_4\text{-}D^{WW}}$ and $\mathbf{C^P_4\text{-}D^{PPxY}}$ were produced in secretory fashion using genetically modified *P. pastoris,* grown in methanol fed-batch mode. Proteins were purified from the cell-free broth by differential ammonium sulfate precipitation, and subsequently dialyzed and lyophilized. Previous studies showed that ammonium sulfate precipitation of protein polymers containing the $\mathbf{C^P_4}$ block typically results in a purity of ~99% at the protein level [19,30]. The gravimetrically determined yields, expressed in g per L of cell-free broth, were 2.2 g L^{-1} for $\mathbf{C^P_4\text{-}D^{WW}}$, and 2.3 g L^{-1} for $\mathbf{C^P_4\text{-}D^{PPxY}}$. Although several reports describe the production of proteins with WW domains and proline-rich ligands in *Escherichia coli*, yields have not been reported [28,31,32]. To our knowledge, production of WW and PPxY domains in *P. pastoris* has not been reported before. The non-optimized yields are in the same g L^{-1} range as for $\mathbf{C^P_4}$ [18], showing that the $\mathbf{D^{WW}}$ and $\mathbf{D^{PPxY}}$ modules are not a significant bottle-neck. This offers good prospects towards their further use in the construction of complex supramolecular architectures and biomaterials.

5.3.2. Protein characterization

$\mathbf{C^P_4\text{-}D^{WW}}$ and $\mathbf{C^P_4\text{-}D^{PPxY}}$ were characterized by SDS–PAGE (Fig. 5.1). Based on the well-established aberrant migration behavior of the $\mathbf{C^P_4}$ block in SDS-PAGE, the proteins were expected to migrate much more slowly in SDS-PAGE

than would be expected on the basis of their theoretical molecular weights of ~41 and ~39 kDa, respectively [18,19,23,30]. This is due to the highly hydrophilic character and consequent low SDS-binding capacity of the $\mathbf{C^P_4}$ block [18]. The 13-residue $\mathbf{D^{PPxY}}$ block resembles the proline-rich nature of the $\mathbf{C^P_4}$ block, and as such is not expected to much affect the mobility of the protein in SDS-PAGE relative to that of $\mathbf{C^P_4}$ alone. On the other hand, the 37-residue hydrophobic $\mathbf{D^{WW}}$ block likely will, because attachment to $\mathbf{C^P_4}$ of peptides that improve SDS binding is known to increase the mobility of the polymer [23,30]. Indeed, the proteins migrated accordingly in SDS-PAGE (Fig. 5.1): $\mathbf{C^P_4\text{-}D^{PPxY}}$ migrates as slowly as the control $\mathbf{C^P_4}$ protein, and $\mathbf{C^P_4\text{-}D^{WW}}$ migrates faster than $\mathbf{C^P_4}$. Although SDS-PAGE is not informative about the molecular mass of the polymers, it does show the proteins are relatively pure and intact.

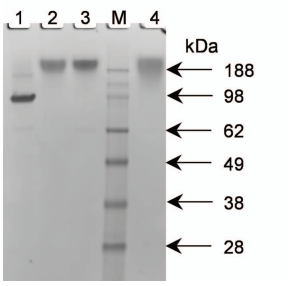


Figure 5.1: *SDS-PAGE analysis of the purified protein polymers. Lane 1,* $\boldsymbol{C^P_4\text{-}D^{WW}}$*; lane 2,* $\boldsymbol{C^P_4\text{-}D^{PPxY}}$*; lane 3,* $\boldsymbol{C^P_4\text{-}D^{PPxY*}}$*; lane M, protein molecular weight marker; lane 4, control protein* $\boldsymbol{C^P_4}$*. Samples were loaded at 13 µg of protein per lane.*

The molecular weight distribution of purified $\mathbf{C^P_4\text{-}D^{WW}}$ and $\mathbf{C^P_4\text{-}D^{PPxY}}$ was further investigated by MALDI-TOF mass spectrometry. A peak at *m/z* 41,440 was observed in the spectrum of $\mathbf{C^P_4\text{-}D^{WW}}$ (Fig. 5.2A). This peak corresponds, within experimental error, to the expected molecular weight of the intact protein (41,437 Da). The small shoulder represents a minor fraction of the molecules with an N-terminal $(Glu\text{-}Ala)_2$ extension. Such extensions commonly occur in *P. pastoris* due to incomplete processing of the α-factor prepro secretory signal [17]. The MALDI-TOF analysis confirms the conclusion from SDS-PAGE that the $\mathbf{C^P_4\text{-}D^{WW}}$ protein is pure and intact. The MALDI-TOF spectrum for $\mathbf{C^P_4\text{-}D^{PPxY}}$ (Fig. 5.2B), however, showed several peaks. The minor low mass peak at *m/z* 38,553 is in accordance with the expected molecular weight of the intact protein (38,552 Da). It seems likely that the other peaks of higher molecular mass represent glycosylated species. Indeed, SDS-PAGE followed by Periodic acid-Schiff staining [33] confirmed that $\mathbf{C^P_4\text{-}D^{PPxY}}$ is

glycosylated (Fig. 5.3). Although the sequence of $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ contains no N-x-[ST] N-glycosylation motifs, *P. pastoris* is also capable of O-glycosylation [34]. Because $\mathbf{C^{P}_{4}}$, as a separate protein, is known to be nonglycosylated [18] (see also Fig. 5.3) most likely the single Ser residue in the added $\mathbf{D^{PPxY}}$ block has been modified with O-glycans.

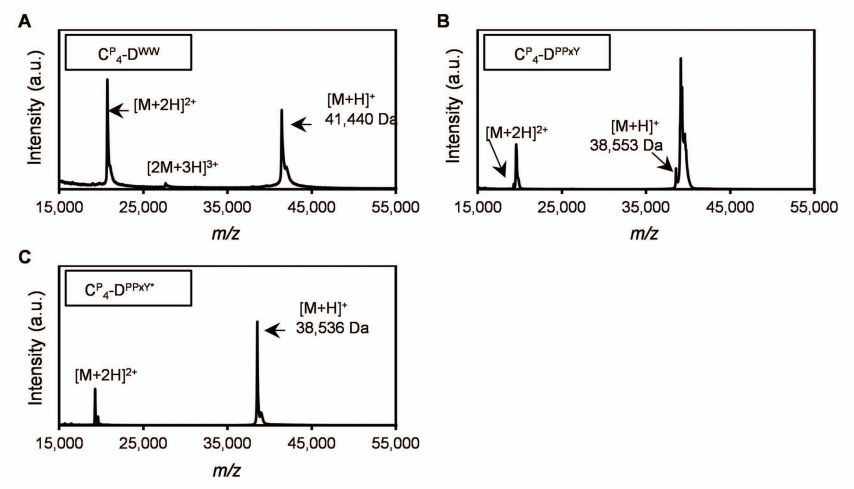


Figure 5.2: *MALDI-TOF analysis of the purified protein polymers. (A)* $\boldsymbol{C^{P}_{4}\text{-}D^{WW}}$*, (B)* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$*, (C)* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY*}}$*.*

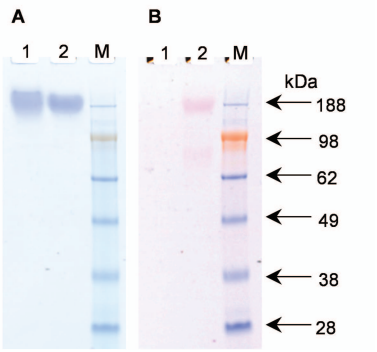


Figure 5.3: *Detection of glycosylation by SDS-PAGE. Lane 1,* $\boldsymbol{C^{P}_{4}}$ *control; lane 2,* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$*; lane M: protein molecular weight marker. (A) Stained with Coomassie Brilliant Blue. (B) Stained using Periodic acid-Schiff staining. Lanes 1 and 2 were loaded with equal amounts of protein.*

5.3.3. Construction of a Ser12 → Ala PPxY mutant

Glyosylation of therapeutic proteins by *P. pastoris* may cause adverse immune responses [35]. Also from a nanomaterials point of view, the polydispersity resulting from glycosylation is undesirable, and desired interactions may well be hindered by the presence of oligosaccharides. To abolish the observed glycosylation of $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$, we constructed a variant with a Ser12 → Ala mutation in the PPxY module (Table 5.1). This variant, denoted $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$, was produced in *P. pastoris* and purified as described above. SDS-PAGE shows a single band (Fig. 5.1), and the yield of purified protein was 2.5 g L^{-1} of cell-free broth. MALDI-TOF of $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ no longer showed the extensive pattern of glycosylated species (Fig. 5.2C). Instead, a major peak at *m/z* 38,536 is seen, matching the expected molecular weight of 38,536 Da. The shoulder represents a minor fraction of the protein not fully processed by *P. pastoris* dipeptidylaminopeptidase, as mentioned above.

The absence of glycosylation in $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ shows that indeed Ser12 of the PPxY module in $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ was glycosylated. Although there is no known consensus sequence for O-glycosylation, serine/threonine rich sequences appear relatively susceptible, particularly when prolines are in the vicinity of the hydroxyl residues [36]. Although both $\mathbf{C^{P}_{4}}$ and $\mathbf{D^{PPxY}}$ are rich in proline and serine, interestingly, all serines in $\mathbf{C^{P}_{4}}$ are followed by proline, while in $\mathbf{D^{PPxY}}$ the single serine is preceded by proline. The observed exclusive glycosylation of serine in the PPxY module seems to agree with the reported enhancement of mannosyl transfer in *S. cerevisiae* for peptide substrates with proline N-terminal to the hydroxyl amino acid [37], and with its inhibition when proline is the C-terminal neighbor [38]. However, this should not be taken to imply that such simple motifs are sufficient to determine O-glycosylation.

5.3.4. Limited glycan characterization

The fact that exclusively Ser12 in the PPxY module was glycosylated allows a direct interpretation of the mass spectrum of the glycosylated protein in terms of the oligosaccharide composition of the population of attached glycans. Although relatively few studies have reported O-glycosylation in *P. pastoris*, it is clear that O-glycans in this host usually consist of up to five mannose units [34], with dimers and trimers being the most abundant species [39,40]. However, longer glycan chains of up to nine mannose residues have been described [41], as well as a phosphorylated Man_6 O-glycan [40]. The glycosidic links between the mannose units are mostly in α1-2 arrangement [39,40], although also α1-3 and α1-6 links have been reported [41,42], as well as ß1-2 links [40].

For closely studying the glycan mass distribution, the MALDI-TOF analysis of $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ was repeated with optimal settings for the *m/z* range of the relevant $[M+H]^+$ ions (Fig. 5.4A). Table 5.2 provides an interpretation of the observed peaks, all of which can be explained in terms of mannose units (+162 Da mass shift) and phosphorylation (+80 Da mass shift). Applying Occam's razor, we assumed maximally one phosphate per glycan structure.

Hypothetically, the $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ protein itself may have been phosphorylated, rather than the oligosaccharides attached to it. Because a nonglycosylated peak shifted by +80 Da is not observed in the MALDI-TOF spectrum of $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$, and because both $\mathbf{C^{P}_{4}}$ [18] and the nonglycosylated $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ are not phosphorylated, such hydroxyl amino acid phosphorylation would need to have occurred specifically in the glycosylated fraction of the $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ molecules. This unlikely notion can be excluded because treatment with alkaline phosphatase did not result in significant changes to the MALDI-TOF spectrum (Fig. 5.1.S1). This finding furthermore indicates that the phosphorylated oligosaccharides do not contain phosphomonoesters, but rather contain phosphorylated Man in diester linkage, in agreement with the findings by Trimble et al. for the above-mentioned phosphorylated Man_6 O-glycan [40].

When $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ was treated with the exoglycosidase jack bean α(1-2,1-3,1-6) mannosidase, a clearly altered mass distribution was obtained in MALDI-TOF (Fig. 5.4B, Table 5.2). Repeated digestions using increasing amounts of enzyme and incubation time all resulted in similar spectra (not shown), suggesting that the glycan structure is partially resistant to the α(1-2,1-3,1-6) mannosidase. The observed phosphorylation provides a likely explanation. Nonphosphorylated species were hardly detectable after digestion with jack bean mannosidase (Fig. 5.4B; Table 5.2). Furthermore, a peak corresponding to Man_2+P is absent in the undigested sample, but appears upon α-mannosidase digestion at the apparent expense of Man_3+P and larger phosphorylated forms. Although jack bean α-mannosidase can trim terminal mannoses linked to phosphate, it cannot thereafter proceed further [43-45]. Thus, the digestion likely halted upon generation of the phosphorylated species observed in Fig. 5.4B (Man_2+P, Man_3+P, and Man_6+P). Treatment of the mannosidase-digested sample with alkaline phosphatase resulted in a shift by -80 Da for these three phosphorylated species (Fig. 5.4C; Table 5.2), confirming that they were present as phosphodiesters prior to α-mannosidase digestion, and that indeed, as assumed above, each glycoform contains only one phosphate. Moreover, dephosphorylation rendered the remaining glycan structures susceptible to further digestion by α-mannosidase (Fig. 5.4D; Table 5.2). Possibly, the minor

amount of remaining Man_6 is capped with a resistant β1-2 Man disaccharide, as has been described for *P. pastoris* [40].

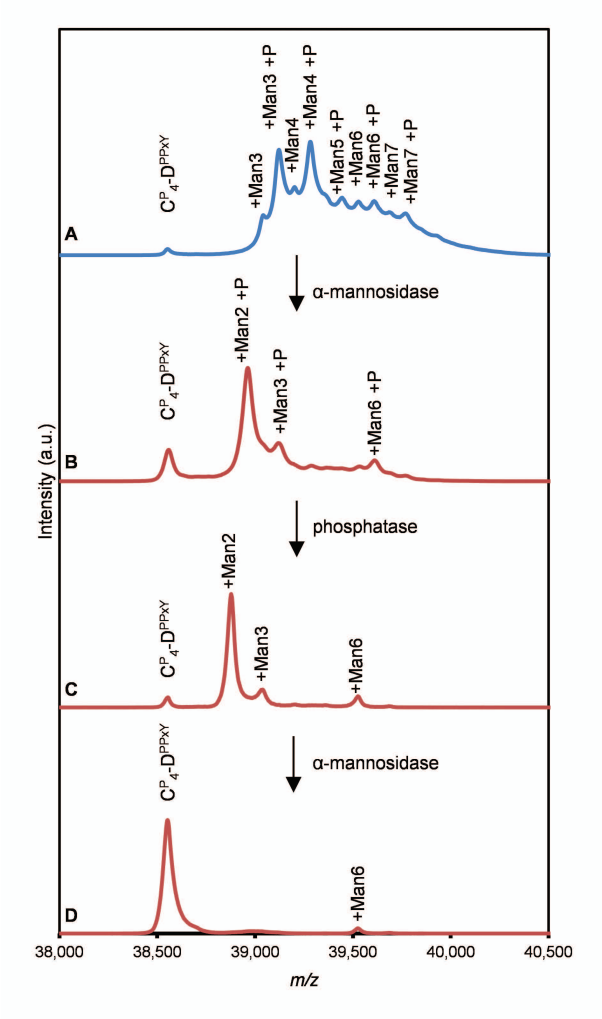


Figure 5.4: *MALDI-TOF of* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$ *treated with α(1-2,1-3,1-6) mannosidase and phosphatase. (A) Untreated* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$*. The protein was digested with (B) α-mannosidase, then with (C) phosphatase, and finally digested again with (D) α-mannosidase. Enzymes were thermally inactivated between consecutive digestions. See Table 5.2 for observed m/z values and theoretical masses of the indicated glycoforms.*

Table 5.2. *Masses observed in MALDI-TOF and tentative glycan structures*

	m/z[a]	Additional mass (Da)[b]	Tentative glycan structure[c]	Theoretical mass of tentative glycan structure (Da)[d]
C^{P}_{4}-D^{PPxY}				
	38,553	-	-	-
	39,042	489	Man_3	486
	39,122	569	Man_3 + P	566
	39,202	649	Man_4	649
	39,283	730	Man_4 + P	729
	39,444	891	Man_5 + P	891
	39,529	976	Man_6	973
	39,609	1056	Man_6 + P	1053
	39,688	1135	Man_7	1135
	39,767	1214	Man_7 + P	1215
C^{P}_{4}-D^{PPxY} *digested with α-mannosidase*				
	38,558	-	-	-
	38,963	405	Man_2 + P	404
	39,121	563	Man_3 + P	566
	39,611	1053	Man_6 + P	1053
C^{P}_{4}-D^{PPxY} *digested with (1) α-mannosidase, and (2) phosphatase*[e]				
	38,553	-	-	-
	38,878	325	Man_2	324
	39,039	486	Man_3	486
	39,526	973	Man_6	973
C^{P}_{4}-D^{PPxY} *digested with (1) α-mannosidase, (2) phosphatase, and (3) α-mannosidase*[e]				
	38,553	-	-	-
	39,526	973	Man_6	973

[a] See Fig. 5.4 (only true peaks are listed; minor inflections in the spectra are ignored)
[b] Relative to the *m/z* value corresponding to the nonglycosylated protein in the same mass spectrum
[c] Assuming mannose (Man) units only, and maximally one phosphate (P) per glycan
[d] Theoretical glycoform masses calculated using 162.14 Da for Man and 79.98 Da for P
[e] Enzymes were thermally inactivated between consecutive digestions

A detailed study of O-glycans chemically released from $\mathbf{C_4\text{-}D^{PPxY}}$ would be needed to determine their precise structure with certainty, but is beyond the scope of this work. Nonetheless, it is clear that the heterogeneous mixture of glycans attached to Ser12 in the PPxY module contains phosphorylated species in diester linkage. Similarly phosphorylated O-glycans have been described for several proteins in *S. cerevisiae* [44-46]. To our knowledge, the present study represents the first confirmation of the occurrence of phosphorylated O-glycans in *P. pastoris* as reported by Trimble et al. [40]. Interestingly, the phosphorylated Man_6 previously reported was described as only a minor component [40], whereas most of the oligosaccharides on $\mathbf{C_4\text{-}D^{PPxY}}$ are phosphorylated.

5.3.5. Binding of proline-rich ligands by the WW domain

Because tryptophans are exclusively present in the WW domain, tryptophan fluorescence can be used to monitor binding of $\mathbf{C^P_4\text{-}D^{WW}}$ to its proline-rich ligands. Upon binding, the local environment of the tryptophans becomes more hydrophobic, causing a blue-shift of the emission maximum and increased fluorescence. Tryptophan fluorescence of a fixed amount of $\mathbf{C^P_4\text{-}D^{WW}}$ (10 µM) was thus followed during titration with a concentrated stock solution of either $\mathbf{C^P_4\text{-}D^{PPxY}}$ or $\mathbf{C^P_4\text{-}D^{PPxY*}}$.

An excitation wavelength of 295 nm was used to prevent excitation of tyrosines [47], which are present in the PPxY domain. With increasing concentration of the ligand, the maximum emission wavelength of $\mathbf{C^P_4\text{-}D^{WW}}$ decreased, indicating the transition of at least one tryptophan of $\mathbf{C^P_4\text{-}D^{WW}}$ from a solvent-exposed to a more hydrophobic environment (Fig. 5.5). As expected, this was accompanied by an increased fluorescence quantum yield. From the titration plots in Fig. 5.5, a K_d of 2.5 µM was calculated by non-linear regression for $\mathbf{C^P_4\text{-}D^{PPxY}}$, and a K_d of 2.7 µM for $\mathbf{C^P_4\text{-}D^{PPxY*}}$. Apparently, the WW domain binds equally well to glycosylated and nonglycosylated $\mathbf{C^P_4\text{-}D^{PPxY}}$.

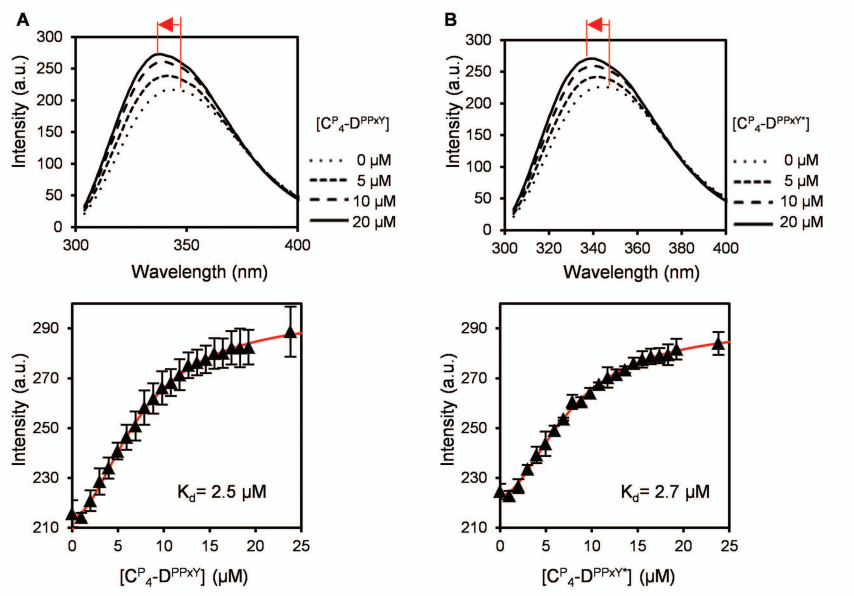


Figure 5.5: *Steady-state tryptophan fluorescence of* $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ *upon titration with (A)* $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ *and (B)* $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$*. The top graphs present tryptophan fluorescence spectra obtained for* $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ *upon addition of different concentrations of* $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ *and* $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$*. The bottom graphs show titration of* $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ *with* $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ *and* $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$*, where fluorescence was monitored at 340 nm. Error bars represent s.d. (n=3).*

Binding affinities were further established using isothermal titration calorimetry. A fixed amount of $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ (200 µM) was titrated with concentrated ligand stock solution (Fig. 5.6). Control experiments where buffer was titrated with ligand, resulted in a relatively small and constant heat of dilution (Fig 5.1.S2A and 5.S2B). Also $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ titrated with control protein $\mathbf{C^{P}_{4}}$ showed only heat of dilution (Fig. 5.1.S2C). The binding isotherms of both $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ showed an immediate decrease in differential power for each consecutive injection (Fig 5.6).

The K_d values derived from the integrated heat plots in Fig. 5.6 are 9.3 and 9.2 µM for $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$, respectively. Both values are similar, and in reasonable agreement with the fluorescence spectroscopy results. In general, the K_d of various WW-domains and their proline-rich ligands are in the high nM to low µM range [48]. According to Russ et al., the PPxY peptide, also used in our $\mathbf{D^{PPxY}}$ block, binds the CC43 WW domain with K_d = 1.7 µM, and the Nedd4.3 WW domain with K_d = 11.2 µM [49]. Wong Po Foo et al. used the same couples in the context of protein polymers and found relatively high K_d values

of 4.6 µM and 62 µM, respectively, for polymers containing *three* PPxY motifs, interacting with polymers containing *three* CC43 or Nedd4.3 domains. The combination of PPxY (p53BP-2) and WWP1-1 used by us was among the best performing pairs tested by Porozzi et al. [29], but to our knowledge no K_d values have been published. In our protein polymer context, the K_d of this combination (~3 to 9 µM) is in a similar range as the above-mentioned literature values for other WW/PPxY combinations. This range is quite sufficient for various supramolecular systems, and multiple **D** blocks could be introduced into the polymer for applications that require even lower working concentrations. We did attempt to produce the CC43 domain as well, but encountered proteolytic degradation in *P. pastoris* that could not be readily resolved. The stoichiometry (N) determined for both $\mathbf{C^P_4\text{-}D^{PPxY}}$ and $\mathbf{C^P_4\text{-}D^{PPxY*}}$ is around 0.9, which, given unavoidable inaccuracies in preparing stock solutions from lyophilized proteins, is in good agreement with the expected 1:1 stoichiometry.

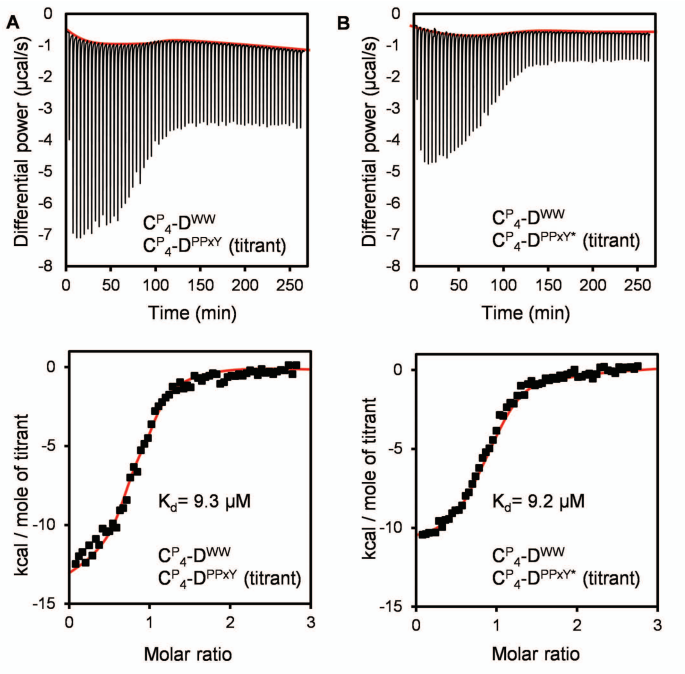


Figure 5.6: *Binding study using isothermal titration calorimetry. (A)* $\boldsymbol{C^P_4\text{-}D^{PPxY}}$ *titrant, (B)* $\boldsymbol{C^P_4\text{-}D^{PPxY*}}$ *titrant. The top graphs show the heat response upon titration of* $\boldsymbol{C^P_4\text{-}D^{WW}}$ *with the proline-rich ligands. Bottom graphs represent the integrated peak areas per mole of ligand.*

5.4. Conclusion

We have shown that polymers containing the WW domain and a proline-rich ligand can be efficiently produced in *Pichia pastoris*. The PPxY module was found to be O-glycosylated, and remarkably a considerable fraction of the oligomannose structures was phosphorylated. O-glycosylation was abolished by changing the serine in the PPxY sequence to alanine. The WW domain effectively bound both the glycosylated and nonglycosylated PPxY modules, with similar binding affinities. This work provides proof-of-concept that otherwise noninteracting protein polymers can be brought together using the specifically interacting WW and PPxY modules. This will allow us to create protein materials with still better controlled structures at the nanoscale, and hence biomedical materials with more precisely defined interactions with living cells.

5.1 Supplementary data

Table 5.1.S1: *Oligonucleotides used in gene construction*

Insert	Oligonucleotide Name	Oligonucleotide sequence
D^{WW}	WW-FW1	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTTTGCCTTCTGGTTGGGAA-3′
	WW-RV1	5′GGTCAACGTAATAAGTACGACCGTGAGGATCCTTTCTTTGTTCCCAACCAGAAGGCAAAG-3′
	WW-FW2	5′GTCGTACTTATTACGTTGACCATAACACTAGGACTACCACATGGGAAAGACCACAGCCATTGCCA-3′
	WW-RV2	5′GTACGAATTCTATTAGCCACCGGCTGGTGCTCCAGGTGGCAATGGCTGTGGTCTT-3′
D^{PPxY}	PPxY-FW	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTGAATACCCTCCATACCCACCAC-3′
	PPxY-RV	5′GTACGAATTCTATTAGCCACCGGCTGGACCAGATGGATAAGGAGGTGGTGGGTATGGAGGGTA-3′
D^{PPxY*}	PPxY*-FW	5′GTGCTGAATACCCTCCATACCCACCACCTCCTTATCCAGCTGGTCCAGCCGGTGGCTAATAG-3′
	PPxY*-RV	5′AATTCTATTAGCCACCGGCTGGACCAGCTGGATAAGGAGGTGGTGGGTATGGAGGGTATTCAGCACCGG-3′

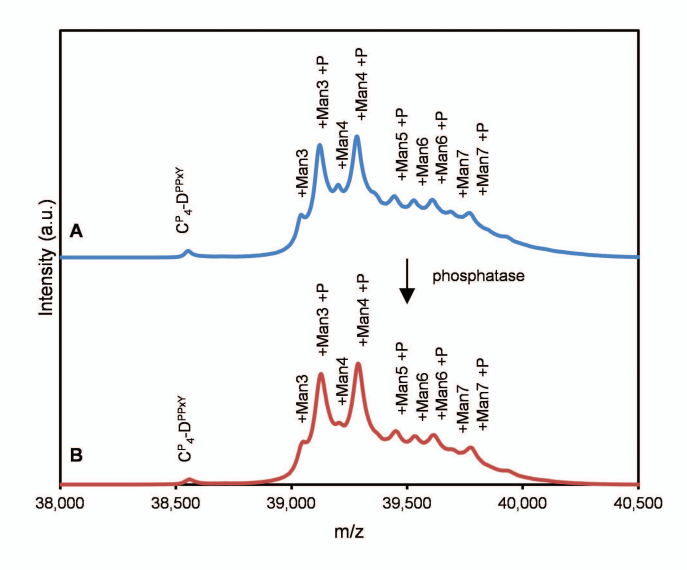


Figure 5.1.S1: *MALDI-TOF of* C^P_4-D^{PPxY} *before (A) and after (B) phosphatase treatment. The same glycoforms are seen in both spectra.*

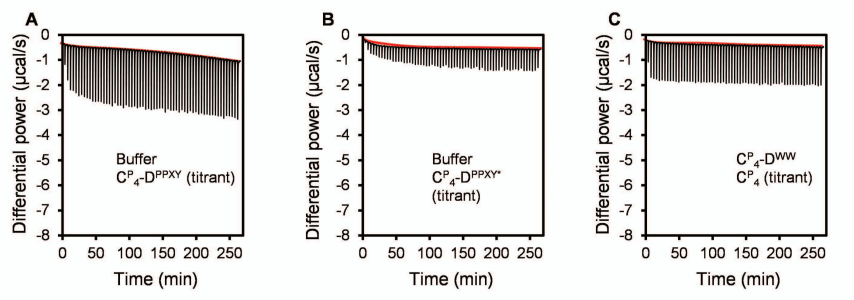


Figure 5.1.S2:*Control ITC measurements. (A) Titration of buffer with* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$*, (B) titration of buffer with* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY*}}$*, (C) titration of* $\boldsymbol{C^{P}_{4}\text{-}D^{WW}}$ *with* $\boldsymbol{C^{P}_{4}}$ *(protein without binding domain).*

5.2 Supplementary data

Material and Methods

Construction of Expression Vectors and Strains

Similarly as in **Chapter 5**, the double-stranded gene fragments encoding $\mathbf{D^{CC43}}$, $\mathbf{D^{PPxY\text{-}1}}$ and $\mathbf{D^{PPxY\text{-}1*}}$ (Table 5.2.S1) were assembled via overlap extension PCR from the oligonucleotides shown in Table 5.2.S2. The gene fragments were digested with *Xho*I/*Eco*RI and cloned into *Xho*I/*Eco*RI-digested vector pMTL23Δ*Bsa*I, in order to obtain vectors pMTL23Δ*Bsa*I-$\mathbf{D^{CC43}}$, pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY\text{-}1}}$ and pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY\text{-}1*}}$. The vector pMTL23-$\mathbf{T_9\text{-}C^P{}_4}$ contains the sequence encoding $\mathbf{T_9\text{-}C^P{}_4}$, and was opened at the 3' end of the $\mathbf{T_9\text{-}C^P{}_4}$ gene with *Van*91I/*Eco*RI. The newly prepared constructs, as well as pMTL23Δ*Bsa*I-$\mathbf{D^{WW}}$, pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY}}$ and pMTL23Δ*Bsa*I-$\mathbf{D^{PPxY*}}$ (discussed in **Chapter 5**) were digested with *Dra*III/*Eco*RI to release inserts $\mathbf{D^{CC43}}$, $\mathbf{D^{WW}}$, $\mathbf{D^{PPxY}}$, $\mathbf{D^{PPxY*}}$, $\mathbf{D^{PPxY\text{-}1}}$ and $\mathbf{D^{PPxY\text{-}1*}}$. The inserts were ligated into the opened pMTL23-$\mathbf{T_9\text{-}C^P{}_4}$ vector, resulting in pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$, pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{WW}}$, pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$, pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$, pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1}}$ and pMTL23-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1*}}$.

Table 5.2.S1: *Amino acid sequences of the C-terminal* $\mathbf{D^{CC43}}$, $\mathbf{D^{PPxY\text{-}1}}$ *and* $\mathbf{D^{PPxY\text{-}1*}}$ *modules. All sequences end C-terminally in the cloning-derived amino acid sequence PAGG (not indicated).*

Module	Amino acid sequence
$\mathbf{D^{CC43}}$	RLPAGWEQRMDVKGRPYFVDHVTKSTTWEDPRPE
$\mathbf{D^{PPxY\text{-}1}}$	YPGPLPPPTYSPSSI
$\mathbf{D^{PPxY\text{-}1*}}$	YPGPLPPPTYSP**ASI**

The vectors pMTL23-$\mathbf{C^P{}_4\text{-}D^{WW}}$ (discussed in **Chapter 5**) was opened at the 5' end of the $\mathbf{C^P{}_4\text{-}D^{WW}}$ gene with *Xho*I/*Van*91*I*. The vector pMTL23Δ*Bsa*I-$\mathbf{D^{WW}}$ was digested with *Xho*I/*Dra*III to release insert $\mathbf{D^{WW}}$. The insert was ligated into the opened vector, resulting in pMTL23-$\mathbf{D^{WW}\text{-}C^P{}_4\text{-}D^{WW}}$.

The $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{WW}}$, $\mathbf{D^{WW}\text{-}C^P{}_4\text{-}D^{WW}}$, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY*}}$, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1}}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1*}}$ inserts were then released with *Xho*I/*Eco*RI and cloned into the likewise-digested *Pichia pastoris* expression vector pPIC9 (ThermoFisher, Bleiswijk, The Netherlands). This resulted in the construction of the vectors: pPIC9-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$, pPIC9-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{WW}}$, pPIC9-$\mathbf{D^{WW}\text{-}C^P{}_4\text{-}D^{WW}}$, pPIC9-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$, pPIC9-$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY*}}$, pPIC9-$\mathbf{T_9\text{-}C^P{}_4}$-

$\mathbf{D^{PPxY\text{-}1}}$ and pPIC9-$\mathbf{T_9}$-$\mathbf{C^{P}_{4}}$-$\mathbf{D^{PPxY\text{-}1*}}$ respectively. The vectors were linearized with *Sal*I to target for integration at the *his4* locus. Transformation of *P. pastoris* GS115 by electroporation and selection of Mut^{+} transformants were performed as described previously.

Table 5. 2.S2: *Oligonucleotides used in gene construction.*

Module	Oligonucleotide Name	Oligonucleotide sequence
$\mathbf{D^{CC43}}$	CC-FW1	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTAGATTGCCTGCTGG-3′
	CC-RV1	5′ATGGCCTACCCTTAACATCCATTCTTTGTTCCCATCCAGCAGGCAATCTAGCACC-3′
	CC-FW2	5′GGATGTTAAGGGTAGGCCATACTTTGTTGATCATGTCACTAAATCTACTACCTGGGAAG-3′
	CC-RV2	5′GTACGAATTCTATTAGCCACCGGCTGGCTCTGGACGAGGGTCTTCCCAGGTAGTAGATTTAG-3′
$\mathbf{D^{PPxY\text{-}1}}$	PPxY-1-FW	5′GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTTATCCAGGTCCATTGCCTCCACCTAC-3′
	PPxY-1-RV	5′GTACGAATTCTATTAGCCACCGGCTGGAATAGAGGAAGGAGAGTAAGTAGGTGGAGGCAATGGAC-3′
$\mathbf{D^{PPxY*}}$	PPxY*-1-1FW	5′GTGCTTATCCAGGTCCATTGCCTCCACCTACTTACTCTCCTGCCTCTATTCCAGCCGGTGGCTAATAG-3′
	PPxY*-1-1RV	5′AATTCTATTAGCCACCGGCTGGAATAGAGGCAGGAGAGTAAGTAGGTGGAGGCAATGGACCTGGATAAGCACCGG-3′

Protein Production and Characterization

All protein polymers were produced, purified and characterized as described in Materials and Methods in **Chapter 5.**

Isothermal Titration Calorimetry

ITC was performed as described in **Chapter 5**. The ligand concentration ($\mathbf{C^{P}_{4}}$-$\mathbf{D^{PPxY}}$ or $\mathbf{C^{P}_{4}}$-$\mathbf{D^{KPPxY*}}$) in the titration syringe was 2.9 mM. Each titration consisted of 63 injections at an interval of 150 sec. Ligand aliquots of 4 µl were titrated into 1.4 ml of a 200 µM $\mathbf{D^{WW}}$-$\mathbf{C^{P}_{4}}$-$\mathbf{D^{WW}}$ protein solution inside the calorimeter cell under continuous stirring at 329 rpm.

Rheology.

Freeze-dried proteins purified by ammonium sulphate precipitation were first dissolved in 10 mM sodium phosphate buffer, then vortexed and incubated overnight at RT to allow proper dissolution. Solutions of protein polymers with the $\mathbf{D^{WW}}$ domain and with the proline-rich domains were mixed in a 1:1 ratio in different combinations. Samples were prepared at concentration 180 g L^{-1}**.** Prior to the measurement all samples were heated at 50 °C for at least 20 min to allow any triple helices in the sample to melt completely. Rheological measurements were done with an Anton Paar MCR 501 rheometer equipped with C10/TI Couette geometry, with bob and cup diameter of 9.991 and 10.840 mm, respectively. The temperature was controlled by a Peltier system, which allows quick heating and cooling. A solvent trap containing oil was used to minimize evaporation. The couette was preheated (50 °C) before adding sample solutions. After inserting the bob into the cup, the temperature was lowered to 20 °C. The gel formation was monitored by applying a sinusoidal deformation to the hydrogel (frequency 1 Hz and amplitude 1%).

Results

Binding of WW domains with proline-rich ligands

The CC43 and PPxY-1 domains are referred to hereafter as $\mathbf{D^{CC43}}$ and $\mathbf{D^{PPxY\text{-}1}}$, respectively. The $\mathbf{D^{PPxY\text{-}1*}}$ domain is a variation of $\mathbf{D^{PPxY\text{-}1}}$, where Ser13→Ala. For amino acid sequences see Table 5.2.S1. The $\mathbf{D^{CC43}}$ belongs to WW domain family [28] and was discussed in **Chapter 5**. The $\mathbf{D^{PPxY\text{-}1}}$ domain is a proline-rich domain [53]. As reported by Wong Po Foo et al. $\mathbf{D^{CC43}}$ can bind $\mathbf{D^{PPxY}}$ [28]. Mosser et al. showed that $\mathbf{D^{WW}}$ can bind $\mathbf{D^{PPxY\text{-}1}}$ [53]. In **Chapter 5**, we showed that $\mathbf{D^{WW}}$ can alternatively bind $\mathbf{D^{PPxY}}$ and its variant with Ser12→Ala, $\mathbf{D^{PPxY*}}$. Building upon that, we attempted to study interaction between WW domains and proline-rich domains in different combinations in the context of $\mathbf{T_9}$-$\mathbf{C^{P}_{4}}$ protein polymer that can form trimers via $\mathbf{T_9}$ tripel helix formation [19].

Protein production and characterization

Following procedure described in **Chapter 5**, all protein polymers were produced in genetically modified *P. pastoris* and purified via ammonium sulfate precipitation. The gravimetrically determined yields, expressed in g per L of cell-free broth, are given in Table 5.2.S3. The non-optimized yields are in the same g L^{-1} range as shown in **Chapter 5**.

Table 5.2.S3: *Theoretical and measured molecular masses of protein polymers produced in this study.*

Protein polymer	Theoretical Molecular Weight (Da)	Measured Molecular Weight (Da)	Product	Yield (g L-1)**
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$	43637	42171	Degraded	0.75
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{WW}}$	43897	43895	Intact	1.8
$\mathbf{D^{WW}\text{-}C^P{}_4\text{-}D^{WW}}$	46060	46051	Intact	2.4
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$	41012	41737*	Glycosylated	1.6
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY*}}$	40996	40994	Intact	1.4
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1}}$	41124	41127*	Glycosylated	1.85
$\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1*}}$	41108	41113*	Glycosylated (to a lesser degree than $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1}}$)	1.6

* The peak of the highest intensity is mentioned, but several peaks of different sizes were observed
* The gravimetrically determined yield is expressed as g of lyophilized product per L of cell-free broth

Obtained protein polymers were characterized by SDS–PAGE (Fig. 5.2.S1). An observed migration is similar as for protein polymers discussed in **Chapter 5**: protein polymers with proline-rich modules migrated as slowly as the control $\mathbf{C^P{}_4}$ protein, and those with WW domains migrated faster than $\mathbf{C^P{}_4}$ due to better binding of SDS. Although SDS-PAGE is not informative about the molecular mass of the polymers, it does show the proteins are relatively pure and intact, except $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$, were multiple bands indicated protein degradation. The molecular weight distribution of purified protein polymers was further investigated by MALDI-TOF mass spectrometry. For all produced protein polymers, the measured molecular weights [17] are compared with theoretical MW in Table 5.2.S3.

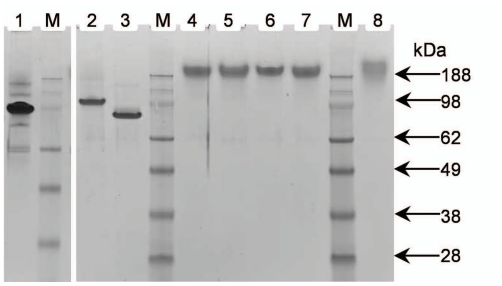


Figure. 5.2.S1: *SDS-PAGE analysis of the purified protein polymers. Lane 1,* T_9-C^P_4-D^{CC43}*; lane 2,* T_9-C^P_4-D^{WW}*; lane 3,* D^{WW}-C^P_4-D^{WW}*; lane 4,* T_9-C^P_4-D^{PPxY}*; lane 5,* T_9-C^P_4-D^{PPxY*}*; lane 6,* T_9-C^P_4-D^{PPxY-1}*; lane 7,* T_9-C^P_4-$D^{PPxY-1*}$*; lane M, protein molecular weight marker, lane 8, control protein* C^P_4.

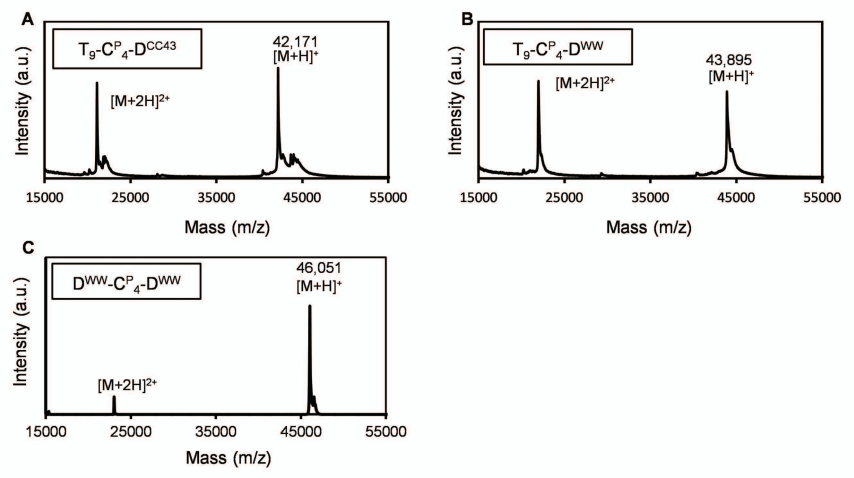


Figure 5.2.S2: *MALDI-TOF analysis of the purified protein polymers. (A)* T_9-C^P_4-D^{CC43}*, (B)* T_9-C^P_4-D^{WW}*, (C)* D^{WW}-C^P_4-D^{WW}.

Multiple peaks were observed in the spectrum of $\mathbf{T_9\text{-}C^P_4\text{-}D^{CC43}}$ (Fig. 5.2.S2A), which is consistent with SDS-PAGE results. Moreover, the weight indicated by the main peak was lower than expected MW, which confirmed that protein is degraded. One peak was observed in spectrum of $\mathbf{T_9\text{-}C^P_4\text{-}D^{WW}}$ (Fig. 5.2.S2B) and spectrum of $\mathbf{D^{WW}\text{-}C^P_4\text{-}D^{WW}}$ (Fig. 5.2.S2C). These peaks correspond, within experimental error, to the expected MW of the intact proteins, which confirms the conclusion from SDS-PAGE that these proteins are pure and intact. The MALDI-TOF spectra for $\mathbf{T_9\text{-}C^P_4\text{-}D^{PPxY}}$, $\mathbf{T_9\text{-}C^P_4\text{-}D^{PPxY\text{-}1}}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^{PPxY\text{-}1*}}$ (Fig. 5.2.S3), showed several peaks of higher MW than expected. Referring to diblock discussed in **Chapter 5**, these peaks most certainly

represent glycosylated species of proteins. Similarly as $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ (**Chapter 5**), MALDI-TOF of $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{PPxY*}}$ no longer showed the extensive pattern of glycosylated species (Fig. 5.2.S3). Although the mutation of the $\mathbf{D^{PPxY\text{-}1}}$ block to $\mathbf{D^{PPxY\text{-}1*}}$ did not remove all glycosylation, it has clearly become less (compare Figures 5.2S3C and D).

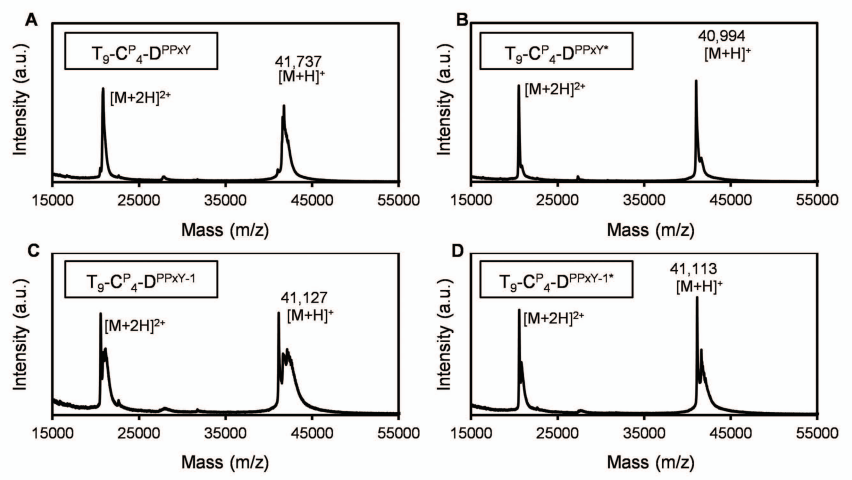


Figure 5.2.S3: *MALDI-TOF analysis of the purified protein polymers. (A)* $\boldsymbol{T_9\text{-}C^{P}_{4}\text{-}D^{PPxY}}$*, (B)* $\boldsymbol{T_9\text{-}C^{P}_{4}\text{-}D^{PPxY*}}$*, (C)* $\boldsymbol{T_9\text{-}C^{P}_{4}\text{-}D^{PPxY\text{-}1}}$*, (D)* $\boldsymbol{T_9\text{-}C^{P}_{4}\text{-}D^{PPxY\text{-}1*}}$*.*

Protein Binding Study

Binding affinities between $\mathbf{D^{WW}\text{-}C^{P}_{4}\text{-}D^{WW}}$ and its glycosylated ligand ($\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$) and nonglycosylated ligand ($\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$) were established using isothermal titration calorimetry. A fixed amount of $\mathbf{D^{WW}\text{-}C^{P}_{4}\text{-}D^{WW}}$ (200 μM) was titrated with concentrated ligand stock solution (Fig. 5.2.S4). Control experiments where buffer was titrated with ligand, resulted in a relatively small and constant heat of dilution (Fig 5.1.S2A and S2B). The binding isotherms of both $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ showed an immediate decrease in differential power for each consecutive injection (Fig. 5.2.S4). The K_d values derived from the integrated heat plots in Fig. 5.2.S4 are 58.8 and 20.8 μM for $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$, respectively. Based on these results, the $\mathbf{D^{WW}\text{-}C^{P}_{4}\text{-}D^{WW}}$ binds nonglycosylated ligand ~3-fold stronger than glycosylated ligand. This is in contrast to results discussed in **Chapter 5**, were $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ binds both ligands with similar affinity. In comparison, $\mathbf{C^{P}_{4}\text{-}D^{WW}}$ binds glycosylated ligand ~6-fold stronger and nonglycosylated ligand ~2-folds stronger than $\mathbf{D^{WW}C^{P}_{4}\text{-}D^{WW}}$.

The stoichiometry (N) determined for both $\mathbf{C^{P}_{4}\text{-}D^{PPxY}}$ and $\mathbf{C^{P}_{4}\text{-}D^{PPxY*}}$ is around 0.9, which, given unavoidable inaccuracies in preparing stock solutions from lyophilized proteins, is in good agreement with the expected 1:1 stoichiometry.

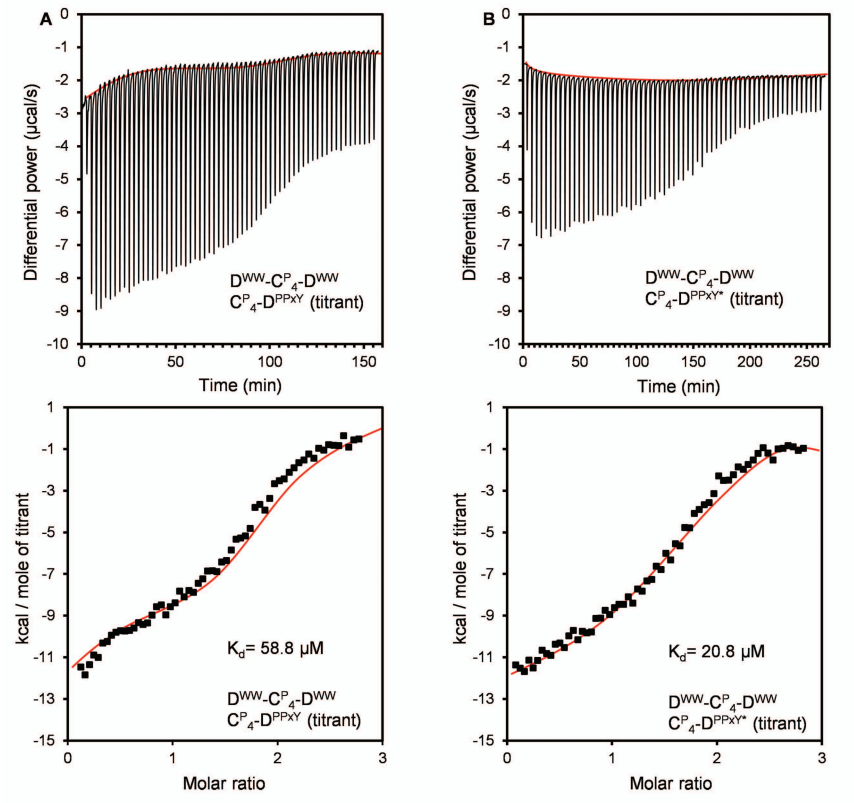


Figure. 5.2.S4. *Binding study using isothermal titration calorimetry. (A)* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY}}$ *titrant, (B)* $\boldsymbol{C^{P}_{4}\text{-}D^{PPxY*}}$ *titrant. The top graphs present the heat response upon titration of* $\boldsymbol{D^{WW}\text{-}C^{P}_{4}\text{-}D^{WW}}$ *with the proline-rich ligands. Bottom graphs represent the integrated peak areas per mole of ligand.*

Hydrogel Formation

Each of four solutions of triblocks with proline-rich domains was mixed with the solution of $\mathbf{T_9\text{-}C^{P}_{4}\text{-}D^{WW}}$ and similarly with the solution of $\mathbf{D^{WW}\text{-}C^{P}_{4}\text{-}D^{WW}}$. Eight obtained samples containing equimolar mixtures were analyzed by oscillatory rheology. The total concentration of samples was 4 mM, which is sufficient concentration allowing formation of $\mathbf{T_9}$ triple helix. For all of analyzed samples, the gel formation was not confirmed by rheology.

Concluding Remarks

Here, we presented an additional information on production of triblocks with WW domains and proline-rich domains. From seven produced protein polymers only $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{CC43}}$ was degraded. The protein polymers $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY}}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1}}$ were glycosylated. The modification of amino acid sequence resulted in nonglycosylated $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY*}}$, but did not completely resolve this issue for $\mathbf{T_9\text{-}C^P{}_4\text{-}D^{PPxY\text{-}1*}}$. In preparing $\mathbf{D^{PPxY\text{-}1*}}$, we did not want to modify more than one residue in the $\mathbf{D^{PPxY\text{-}1}}$ sequence in view of the risk of disturbing binding functionality, but besides Ser13 apparently also Thr9 and/or Ser11 are glycosylated.

Further, we also showed primary analysis of interaction between the heterodimer-forming domains incorporated into triblock. Regarding ITC results, in **Chapter 5** we have envisioned that multiple **D** blocks could be introduced into the protein polymer for applications that require very low working concentrations. Here, however, we found that it is not so straightforward, since $\mathbf{D^{WW}\text{-}C^P{}_4\text{-}D^{WW}}$ showed lower affinity towards its ligands in comparison to the variant with only one binding domain. The decrease in binding affinity for $\mathbf{D^{WW}C^P{}_4\text{-}D^{WW}}$ is not entirely clear. We might assume that very flexible $\mathbf{C^P{}_4}$ linker might allow the $\mathbf{D^{WW}}$ domains to align sterically towards each other in such a way that it affects ligand recognition, however, this is just a speculation and has not been verified. An inability to form hydrogels by any of tested combinations at concentration 4 mM was also very surprising, since in **Chapter 5** we showed that $\mathbf{C^P{}_4\text{-}D^{WW}}$ binds its ligands with affinity within a micromolar range. Since $\mathbf{T_9}$ is also proline-rich, we speculate that $\mathbf{D^{WW}}$ might also bind to $\mathbf{T_9}$, which disturbs network formation. This could be answered by monitoring of tryptophan fluorescence during the titration of protein polymers containing $\mathbf{D^{WW}}$ with $\mathbf{T_9\text{-}C^P{}_4\text{-}T_9}$ [19]. The interaction of WW domain with proline-rich modules should be investigated further in order to find optimal conditions for its application.

References

1. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L. *Biomacromolecules* **2011**, 12, 269-289.
2. Ghandehari, H.; Cappello, J. *Pharm Res* **1998**, 15, 813-815.
3. Tessmar, J. K.; Göpferich, A. M. *Adv Drug Deliv Rev* **2007**, 59, 274-291.
4. Brea, R. J.; Reiriz, C.; Granja, J. R. *Chemical Society Reviews* **2010**, 39, 1448-1456.
5. Matson, J. B.; Stupp, S. I. *Chem Commun* **2012**, 48, 26-33.
6. Price, R.; Poursaid, A.; Ghandehari, H. *J Control Release* **2014**, 190, 304-313.
7. Gagner, J. E.; Kim, W.; Chaikof, E. L. *Acta Biomater* **2014**, 10, 1542-1557.
8. Shi, P.; Gustafson, J. A.; MacKay, J. A. *Int J Nanomedicine* **2014**, 9, 1617-1626.
9. Heslot, H. *Biochimie* **1998**, 80, 19-31.
10. Annabi, N.; Mithieux, S. M.; Camci-Unal, G.; Dokmeci, M. R.; Weiss, A. S.; Khademhosseini, A. *Biochem Eng J* **2013**, 77, 110-118.
11. Liu, L.; Busuttil, K.; Zhang, S.; Yang, Y.; Wang, C.; Besenbacher, F.; Dong, M. *Phys Chem Chem Phys* **2011**, 13, 17435-17444.
12. Kopeček, J.; Yang, J. *Angew Chem Int Ed* **2012**, 51, 7396-7417.
13. Fichman, G.; Gazit, E. *Acta Biomater* **2014**, 10, 1671-1682.
14. Kim, H.; Siu, K.-H.; Raeeszadeh-Sarmazdeh, M.; Sun, Q.; Chen, Q.; Chen, W. *Biotechnol Bioeng* **2015**, 112, 1495-1505.
15. Walper, S. A.; Turner, K. B.; Medintz, I. L. *Curr Opin Biotechnol* **2015**, 34, 232-241.
16. Domeradzka, N. E.; Werten, M. W. T.; de Wolf, F. A.; de Vries, R. *Curr Opin Biotechnol* **2016**, 39, 61-67.
17. Werten, M. W. T.; van den Bosch, T. J.; Wind, R. D.; Mooibroek, H.; de Wolf, F. A. *Yeast* **1999**, 15, 1087-1096.
18. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng* **2001**, 14, 447-454.
19. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009**, 10, 1106-1113.
20. Schipperus, R.; Eggink, G.; de Wolf, F. A. *Biotechnol Progress* **2012**, 28, 242-247.

21. Beun, L. H.; Storm, I. M.; Werten, M. W. T.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 3349-3357.
22. Włodarczyk-Biegun, M. K.; Werten, M. W. T.; de Wolf, F. A.; van den Beucken, J. J. J. P.; Leeuwenburgh, S. C. G.; Kamperman, M.; Cohen Stuart, M. A. *Acta Biomater* **2014**, 10, 3620-3629.
23. Domeradzka, N. E.; Werten, M. W. T.; de Vries, R.; de Wolf, F. A. *Biotechnol Bioeng* **2016,** 113, (5), 953-960.
24. Wang, H.; Heilshorn, S. C. *Adv Mater* **2015**, 27, 3717-3736.
25. Chan, G.; Mooney, D. J. *Trends Biotechnol* **2008**, 26, 382-392.
26. Sudol, M.; Chen, H. I.; Bougeret, C.; Einbond, A.; Bork, P. *FEBS L* **1995**, 369, 67-71.
27. Macias, M. J.; Hyvonen, M.; Baraldi, E.; Schultz, J.; Sudol, M.; Saraste, M.; Oschkinat, H. *Nature* **1996**, 382, 646-649.
28. Wong Po Foo, C. T. S.; Lee, J. S.; Mulyasasmita, W.; Parisi-Amon, A.; Heilshorn, S. C. *PANS* **2009**, 106, 22067-22072.
29. Pirozzi, G.; McConnell, S. J.; Uveges, A. J.; Carter, J. M.; Sparks, A. B.; Kay, B. K.; Fowlkes, D. M. *J Biol Chem* **1997**, 272, 14611-14616.
30. Moers, A. P. H. A.; Wolbert, E. J. H.; de Wolf, F. A.; Werten, M. W. T. *J Biotechnol* **2010**, 146, 66-73.
31. Bao, W.-J.; Gao, Y.-G.; Chang, Y.-G.; Zhang, T.-Y.; Lin, X.-J.; Yan, X.-Z.; Hu, H.-Y. *Protein Express Purif* **2006**, 47, 599-606.
32. Martinez-Rodriguez, S.; Bacarizo, J.; Luque, I.; Camara-Artigas, A. *J Struct Biol* **2015**, 191, 381-387.
33. Zacharius, R. M.; Zell, T. E.; Morrison, J. H.; Woodlock, J. J. *Anal Biochem* **1969**, 30, 148-152.
34. Bretthauer, R. K.; Castellino, F. J. *Biotechnol Appl Biochem* **1999**, 30, 193-200.
35. Daly, R.; Hearn, M. T. W. *J Mol Recognit* **2005**, 18, 119-138.
36. Herscovics, A.; Orlean, P. *FASEB J* **1993**, 7, 540-50.
37. Lehle, L.; Bause, E. *Biochim Biophys Acta* **1984**, 799, 246-251.
38. Strahl-Bolsinger, S.; Tanner, W. *FEBS J* **1991**, 196, 185-190.
39. Duman, J. G.; Miele, R. G.; Liang, H.; Grella, D. K.; Sim, K. L.; Castellino, F. J.; Bretthauer, R. K. *Biotechnol Appl Biochem* **1998**, 28, 39-45.
40. Trimble, R. B.; Lubowski, C.; Hauer, C. R.; Stack, R.; McNaughton, L.; Gemmill, T. R.; Kumar, S. A. *Glycobiology* **2004**, 14, 265-274.
41. Gustafsson, A.; Sjöblom, M.; Strindelius, L.; Johansson, T.; Fleckenstein, T.; Chatzissavidou, N.; Lindberg, L.; Ångström, J.; Rova, U.; Holgersson, J. *Glycobiology* **2011**, 21, 1071-1086.

42. Boraston, A. B.; Sandercock, L. E.; Warren, R. A. J.; Kilburn, D. G. *J Mol Microbiol Biotechnol* **2003**, 5, 29-36.
43. Hernández, L. M.; Ballou, L.; Alvarado, E.; Gillece-Castro, B. L.; Burlingame, A. L.; Ballou, C. E. *J Biol Chem* **1989**, 264, 11849-56.
44. Bergwerff, A. A.; Stark, W.; Fendrich, G.; Knecht, R.; Blommers, M. J. J.; Maerki, W.; Kragten, E. A.; van Oostrum, J. *Eur J Biochem* **1998**, 253, 560-575.
45. Nakayama, K.-i.; Feng, Y.; Tanaka, A.; Jigami, Y. *Biochim Biophys Acta Gen Subj* **1998**, 1425, 255-262.
46. Jars, M. U.; Osborn, S.; Forstrom, J.; MacKay, V. L. *J Biol Chem* **1995**, 270, 24810-24817.
47. Weljie, A. M.; Vogel, H. J. In *Methods in Molecular Biology*; Vogel, H. J., Ed.; Springer New York: Totowa, NJ, 2002; pp 75-87.
48. Kay, B. K.; Williamson, M. P.; Sudol, M. *FASEB J* **2000**, 14, 231-241.
49. Russ, W. P.; Lowery, D. M.; Mishra, P.; Yaffe, M. B.; Ranganathan, R. *Nature* **2005**, 437, 579-583.
50. Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. *Gene* **1989**, 77, 51-59.
51. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W. T.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 699-706.
52. Zhang, W.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M. *Biotechnol Bioeng* **2000**, 70, 1-8.
53. Mosser, E. A.; Kasanov, J. D.; Forsberg, E. C.; Kay, B. K.; Ney, P. A.; Bresnick, E. H. *Biochemistry* **1998**, 37, 13686-13695.

6 Cross-linking and bundling of self-assembled protein-based polymer fibrils via heterodimeric coiled coils

Abstract

Previously, we developed triblock protein polymers that form fibrillar hydrogels at low protein polymer concentrations (denoted $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$). We here demonstrate that the structure of these hydrogels can be tuned via heterodimeric coiled coils that cross-link and bundle the self-assembled protein-polymer fibrils. We fused well-characterized, 47 amino acids-long heterodimeric coiled coil "linkers" ($\mathbf{D^A}$ or $\mathbf{D^B}$) to the C-terminus of the triblock polymer. The resulting $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ polymers, were successfully produced as secreted proteins in *Pichia pastoris*, with titers of purified protein in the order of g L^{-1} of clarified broth. Atomic force microscopy showed that fibrils formed by either $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ alone already displayed extensive bundling, apparently as a result of homotypic ($\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$) interactions. For fibrils prepared from protein polymers having no linkers, plus a small fraction of polymers containing either $\mathbf{D^A}$ or $\mathbf{D^B}$ linkers, no cross-linking and bundling was observed. At these same low concentrations of linkers, fibrils containing both the $\mathbf{D^A}$ and the $\mathbf{D^B}$ linkers did show cross-linking and bundling as a consequence of heterodimer formation. This work shows that coiled coil modules can be employed to control bundling of supramolecular fibrils, which is promising for the further development of materials that mimic the extracellular matrix.

Submitted in modified form as: Domeradzka, N.E.; Werten, M.W.T.; de Vries, R. and de Wolf, F.A. Cross-linking and bundling of self-assembled protein-based polymer fibrils via heterodimeric coiled coils.

6.1. Introduction

The extracellular matrix (ECM) is a network of fibrous proteins [1] that provides structural and biochemical support to the cells [2]. Considerable effort has been devoted to the development of artificial extracellular matrices for medical applications such as tissue engineering and drug delivery [3-7]. Different hydrogels made of synthetic polymers [8-11], natural polymers [12-14] or short synthetic peptides [15-17] have been explored for this purpose. Various recombinant protein polymers (collagen-inspired, elastin-inspired, and silk-inspired) that form various types of physical hydrogels have been developed in our laboratory [18-23]. The triblock protein polymer $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$, which forms fibrillar hydrogels at pH values above 6 and at low protein concentrations, is a promising candidate for use as an ECM-mimicking material [24]. Fibril formation is driven by the silk-like middle block $\mathbf{S^H_{48}}$ = $(GAGAGAGH)_{48}$, and the fibrils are stabilized against aggregation by two end-blocks denoted $\mathbf{C_2}$ [19]. Each end-block consists of two repeats of a 99 amino acid-long hydrophilic random coil "**C**" block [25]. The $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymer self-assembles under physiological conditions and is not cytotoxic [24].

The structure and mechanics of the ECM are determined by a range of complex interactions [26]. For example, there is a wide range of proteins that form different types of physical cross-links between different ECM components. A key role is also played by the precisely controlled bundling of collagen fibrils into fibers [27]. A notable demonstration of the impact of fibril bundling on the properties of hydrogels is provided by Kouwer et al. [28]. The well-defined fibers in their system, consisting of multiple bundled fibrils, were much stiffer than the individual fibrils. Consequently, hydrogel formation already occurred at very low concentrations (1 g L^{-1}) [28].

Our laboratory has similarly sought to influence the cross-linking and bundling of fibrils in $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ hydrogels. For example, shortening the stabilizing end-blocks leads to some degree of fibril cross-linking, fibril bundling, and modulus increase [29]. Also, variants of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymer were developed with N-terminal heparin-binding domains. Fibrils formed by these modified protein polymers could be made to cross-link and bundle via the addition of heparin, again leading to an increased modulus [30].

Here, we explore the use of heterodimerizing peptide modules as an alternative strategy to obtain more control over the physical cross-linking and bundling of fibrils formed by the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ triblock protein polymer. As we will show, it is possible to produce $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ tetrablock protein polymers by secretory expression in *Pichia pastoris*. Using atomic force

microscopy (AFM), we demonstrate that fibrils prepared from mixtures of **C_2-S^H_{48}-C_2**, **C_2-S^H_{48}-C_2-D^A** and **C_2-S^H_{48}-C_2-D^B** have a degree of cross-linking and bundling that can be controlled via the relative ratio of the three polymers.

6.2. Materials and methods

6.2.1. Construction of expression vectors and strain

The construction of vector pMTL23Δ*Bsa*I-**C_2-S^H_{48}-C_2**, which contains the gene encoding the **C_2-S^H_{48}-C_2** triblock copolymer, was previously described by Golinska et al. [19]. The construction of two vectors pMTL23Δ*Bsa*I-**D^A** and pMTL23Δ*BsaI*-**D^B** was previously described by Domeradzka et al [31]. These vectors contain gene fragments encoding, respectively, the relatively acidic heterodimer-forming leucine zipper peptide **D^A** and the relatively basic leucine zipper **D^B**. To add **D^A** and **D^B** modules to the C-terminus of the **C_2-S^H_{48}-C_2** triblock, restriction sites were added to the corresponding **D^A** and **D^B** DNA fragments by PCR, using forward primers that contain a *Ban*I restriction site, and reverse primers that contain a *Not*I site and a downstream *Xho*I site. The **D^A** block was amplified using the primers 5'-TATGGTGCCTTAGAAATTAGAGCTGCCTTTTTGAGAC-3' (forward) and 5'-ATACTCGAGCGGCCGCTTAACCCTTTCCACCTCCAAGTGGACC-3' (reverse). The **D^B** block was amplified using the primers 5'-TATGGTGCCTTGGAAATTGAAGCCGCCTTTCTTG-3' (forward) and 5'-ATACTCGAGCGGCCGCTTAACCCTTTCCACCTCCCAATGGACC-3' (reverse). The vector pMTL23Δ*Bsa*I-**C_2-S^H_{48}-C_2** was opened at the 3' end of the **C_2-S^H_{48}-C_2** gene with *Ban*I/*Xho*I. The **D^A** and **D^B** PCR products were digested with *Ban*I/*Xho*I to obtain sticky ends. The inserts were then ligated into the opened vector, resulting in pMTL23Δ*Bsa*I-**C_2-S^H_{48}-C_2-D^A** and pMTL23Δ*Bsa*I-**C_2-S^H_{48}-C_2-D^B**, respectively. These two inserts were then released with *Eco*RI/*Not*I and cloned into the likewise-digested *Pichia pastoris* expression vector pPIC9 (ThermoFisher, Bleiswijk, The Netherlands). This resulted in the vectors pPIC9-**C_2-S^H_{48}-C_2-D^A** and pPIC9-**C_2-S^H_{48}-C_2-D^B**. The vectors were linearized with *Sal*I to target for integration at the *his4* locus. Transformation of *P. pastoris* GS115 by electroporation and selection of Mut^+ transformants were performed as described previously [23]. For the expected protein sequences upon expression see Supplementary data Table 6.1.S1.

6.2.2. Protein production and purification

Fermentations were done using 2.5-L Bioflo 3000 stirred-tank bioreactors (New Brunswick Scientific, Nijmegen, The Netherlands), using BioCommand Software (New Brunswick Scientific, Nijmegen, The Netherlands) and a homebuilt methanol sensor-controller. The fermentations were performed under conditions optimal for $\mathbf{D^A}$ and $\mathbf{D^B}$, as described previously [31]. A starting volume of 1.25 L minimal basal salts medium [32] was used, and the growth temperature was 30 °C. The pH was controlled at 3.0 throughout the entire fermentation. The inlet air was supplemented with 20% oxygen for the duration of the glycerol fed-batch phase and the methanol fed-batch phase. During the induction phase, methanol levels were kept at a constant level of 0.2% (w/v).
The last 40 min prior to methanol induction the temperature was linearly decreased from 30 °C to 20 °C. Approximately 10 min before methanol induction the medium was supplemented with 1% casamino acids.
The purification of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ was similar to the procedure described previously for $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ [19]. The proteins were precipitated from the supernatant by addition of ammonium sulfate to 45% of saturation, followed by incubation on ice for 30 min and centrifugation for 30 min at 20,000 × g (4 °C). The protein pellet was resuspended in 50 mM formic acid and precipitated once more. The pellet was then resuspended in 50 mM formic acid, and desalted by extensive dialysis using Spectra/Por 7 tubing (Spectrum Laboratories, Breda, The Netherlands) with a 1 kDa molecular weight cut-off against 10 mM formic acid. The desalted protein was freeze-dried for storage until use. The lyophilized $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein was produced as described previously by Golinska et al. [19].

6.2.3. SDS-PAGE

SDS-PAGE was performed using a NuPAGE Novex System (ThermoFisher, Bleiswijk, The Netherlands) with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 pre-stained molecular mass markers. Prior to electrophoresis, all samples were heated for 10 min at 70 °C in NuPAGE LDS Sample Buffer with NuPAGE Sample Reducing Agent, as per manufacturer's recommendations for denaturing and reducing PAGE. For gel staining, Coomassie SimplyBlue SafeStain (ThermoFisher, Bleiswijk, The Netherlands) was used.

6.2.4. Mass spectrometry

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was performed using an ultrafleXtreme mass spectrometer (Bruker, Leiderdorp, The Netherlands). Samples were prepared by the dried droplet method on a 600 µm AnchorChip target (Bruker, Leiderdorp, The Netherlands), using 5 mg mL^{-1} 2,5-dihydroxyacetophenone, 1.5 mg mL^{-1} diammonium hydrogen citrate, 25% (v/v) ethanol and 3% (v/v) trifluoroacetic acid as matrix. Spectra were derived from ten 500-shot (1,000 Hz) acquisitions taken at non-overlapping locations across the sample. Measurements were made in the positive linear mode, with ion source 1, 25.0 kV; ion source 2, 23.3 kV; lens, 6.5 kV; pulsed ion extraction, 680 ns. Protein Calibration Standard II (Bruker, Leiderdorp, The Netherlands) was used for external calibration.

6.2.5. Atomic force microscopy (AFM)

Stock solutions of purified protein polymers for atomic force microscopy (AFM) were prepared by suspending the lyophilized product in 5 mM HCl at a concentration of 10 g L^{-1}, followed by overnight incubation at room temperature to allow complete dissolution.

In a first experiment, aimed at investigating whether the incorporated $\mathbf{D^A/D^B}$ modules disturb self-assembly of the silk-inspired polymers, fibrils were prepared exclusively from either $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$. The influence of pH was investigated as well. To start fibril formation, the pH values of the stock solutions were adjusted to the desired value with NaOH, and the solutions were diluted to 1 g L^{-1} with 10 mM sodium phosphate buffer of the same pH. The pH values evaluated were pH 7, 9, 10, and 12.

In a second experiment, heterodimer formation by $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ was tested in the presence of excess unmodified $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$, such that the final protein concentration was always 10 g L^{-1}. To this end, equimolar mixtures of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ were prepared and immediately added to a solution of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ in four different ratios. In this manner, the final weight fraction of the combined modified polymers constituted 1, 2, 4, or 8% of the total protein. That is, each modified polymer was present at 0.05, 0.1, 0.2, or 0.4 g L^{-1}, respectively. As controls for possible homodimerization under these conditions, analogous mixtures were prepared of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ with either $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$. Thus, always only one modified polymer was present, at concentrations of 0.1, 0.2, 0.4, or 0.8 g L^{-1} , respectively. Fibril formation was induced by adjusting the pH to 7.4 with

NaOH, and dilution to a final protein concentration of 1 g L^{-1} with 10 mM sodium phosphate (pH 7.4).

For AFM imaging, the above polymer solutions were first incubated at room temperature to allow supramolecular assembly to occur. After 24 h, a drop (6 μL) of sample was deposited onto a clean silica wafer and incubated for 20 min. The wafer was then rinsed with 600 μL of Milli-Q water to remove salts, and carefully dried under a stream of nitrogen. Dry samples were imaged using a Nanoscope V (Veeco, NY, U.S.A.) in Scan Asyst imaging mode, using nonconductive silicon nitride probes with a spring constant of the cantilever 0.32 N m^{-1}. Images were recorded at frequency of 0.997 Hz and further processed with NanoScope Analysis 1.20 software (Veeco Instruments Inc. 2010, U.S.A.).

6.2.6. Rheology

Stock solution of purified proteins $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$, $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$, were prepared by dissolving the lyophilized polymers in 5 mM HCl, followed by overnight incubation at room temperature. To start self-assembly into fibrils and hydrogels, the pH was adjusted to pH 7.4 with NaOH, and samples were diluted with Milli-Q to the desired concentrations.

The rheological properties of hydrogels consisting exclusively of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$, $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$, or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ were compared at a concentration of 10 g L^{-1}. We also analyzed hydrogels that consisted of an equimolar mixture of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$, which was subsequently mixed with unmodified $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$. In these hydrogels, the summed weight fraction of polymers containing the $\mathbf{D^A/D^B}$ modules was 2, 4, 8, 16, 32, 64, 80 or 100% of the total protein concentration (10 g L^{-1}). That is, the two modified polymers are each present at final concentrations of 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 4.0, or 5.0 g L^{-1}. Finally, a series of measurements at a range of concentrations was performed for (i) an equimolar mixture of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$, and (ii) $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ alone. In these cases, samples were analyzed at total protein concentrations of 5 g L^{-1}, 7.5 g L^{-1}, 10 g L^{-1}, 15 g L^{-1}, 20 g L^{-1}, and 30 g L^{-1}.

All rheological measurements were performed using an Anton Paar MCR 501 Rheometer (Anton Paar, Oosterhout, The Netherlands) equipped with a CC10/T200 Couette geometry (Anton Paar, Oosterhout, The Netherlands), with bob and cup diameter of 10.002 and 10.845 mm, respectively. A solvent trap was used to minimize evaporation. Frequency sweeps with angular frequency (ω) between 0.01 and 100 rad s^{-1} were performed at a strain of 0.1%. Strain

sweeps were performed between 0.1 and 100% deformation and a frequency of 1 Hz. For measuring the build-up of the modulus in time, oscillatory shear measurements were done at an oscillation frequency of 1 Hz, and a fixed strain of 0.1%. The temperature was kept at 20 °C using a Peltier system.

6.3. Result and discussion

6.3.1. Protein production and characterization

The two protein polymers, $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^B}$, were biosynthesized in genetically modified *Pichia pastoris*. Their full amino acid sequences are listed in Table S1. The yield of purified protein was 2.3 g L^{-1} of cell-free broth for $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^A}$ and 1.4 g L^{-1} for $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^B}$. To assess product identity, purified samples were subjected to SDS–PAGE and MALDI-TOF. For both polymers, a main protein band was detected in SDS-PAGE at a position corresponding to the marker band of ~98 kDa, and it was accompanied by a smear at lower molecular weight (Fig. 6.1).

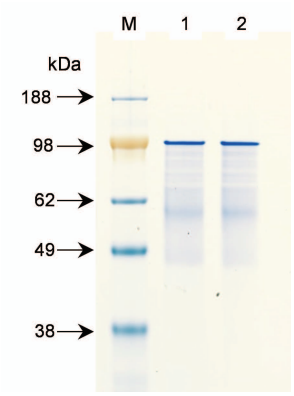


Figure 6.1: *SDS-PAGE analysis of the purified protein polymers. The purified protein polymers (5 µg) were subjected to electrophoresis under denaturing and reducing conditions. Lane 1,* $\boldsymbol{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^A}$*; lane 2,* $\boldsymbol{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^B}$*; lane M, protein molecular weight marker.*

The MALDI-TOF spectrum showed only a single peak for both $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^B}$, at *m/z* 71,569 and *m/z* 71,642, respectively (Fig. 6.2). This corresponded well with the predicted molecular weights of 71,607 Da and 71,689 Da, respectively, within experimental error. The anomalous molecular weights apparent from SDS-PAGE were expected and are related to the hydrophilic nature of the **C** block that poorly binds SDS [20, 21, 31, 34-36]. A similar migration pattern in SDS-PAGE was previously observed for $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ [24].

The protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ migrate slightly faster than $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ because the $\mathbf{D^A}$ and $\mathbf{D^B}$ modules have substantial SDS-binding [31]. The smear visible in SDS-PAGE could indicate that the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ samples contained impurities. MALDI-TOF, however, did not show any evidence of that, and in $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ a similar smear was shown to be an SDS-PAGE artifact [24].

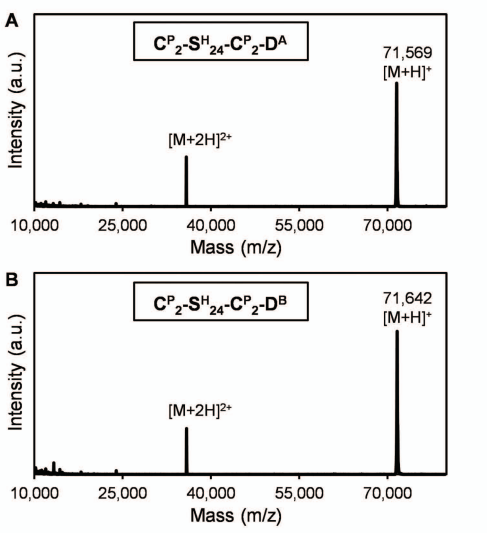


Figure 6.2: *MALDI-TOF analysis of the purified protein polymers. (A)* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^A}$*; (B)* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^B}$.

6.3.2. Pure $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ fibrils

The $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymer forms fibrils at pH ≥6 [19]. The morphology of these fibrils was studied by tapping mode AFM (Fig 6.3). Using AFM we investigated whether the attachment of $\mathbf{D^A/D^B}$ modules at the C-terminus of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymer influenced its ability to form fibrils. Fibril formation of separate $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ solutions was studied, at a concentration of 1 g L^{-1} and at different pH values. As shown in Fig. 6.4A and 6.4B both proteins formed long fibrils. In contrast to $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ (Fig. 6.3), the fibrils appear to be cross-linked and bundled. While fibrils formed by $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ were evenly spread over the surface of the silica wafer, cross-linked and bundled fibrils of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ were observed as patches on an otherwise empty surface (Fig. 6.1.S1). The presence of bundles suggests that both $\mathbf{D^A}$ and $\mathbf{D^B}$ can undergo homotypic ($\mathbf{D^A/D^A}$ and

$\mathbf{D^B/D^B}$) interaction. This is in agreement with our previous study, where we showed that $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ homotypic association may occur at concentrations above ~100 μM [31]. The bundling was found to be strongly pH dependent. $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ fibrils do not appear to bundle when pH ≥ 10, and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ fibrils do not appear to bundle when pH ≥ 12. The disappearance of bundles at high pH is most likely caused by electrostatic repulsion as a consequence of the larger negative charge of the fibrils at these pH values. Also, interactions between coiled-coils are typically at least partly electrostatic in nature [37], and hence the homotypic $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ associations might be pH-dependent.

To selectively suppress homotypic interaction while maintaining heterodimerization, the concentration of $\mathbf{D^A}$ and $\mathbf{D^B}$ must be below the aforementioned ~100 μM. For the samples studied with AFM, the overall protein polymer concentration was 14 μM. However, the local $\mathbf{D^A}$ and $\mathbf{D^B}$ concentration in the fibrils is much larger, and apparently large enough to trigger homotypic associations. Next, we focus on controlling fibril cross-linking and bundling via heterotypic interactions between $\mathbf{D^A}$ and $\mathbf{D^B}$ modules in fibrils.

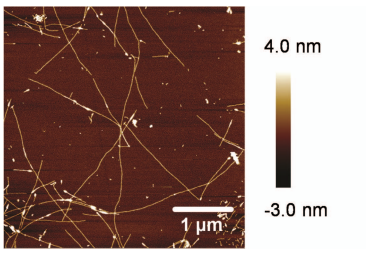


Figure 6.3: *AFM images of 0.1 % (w/v))* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2}$*. Protein solution was incubated for 24 hours at pH 7.4 at room temperature, followed by 20 minute deposition on the silica wafer, rising and drying.*

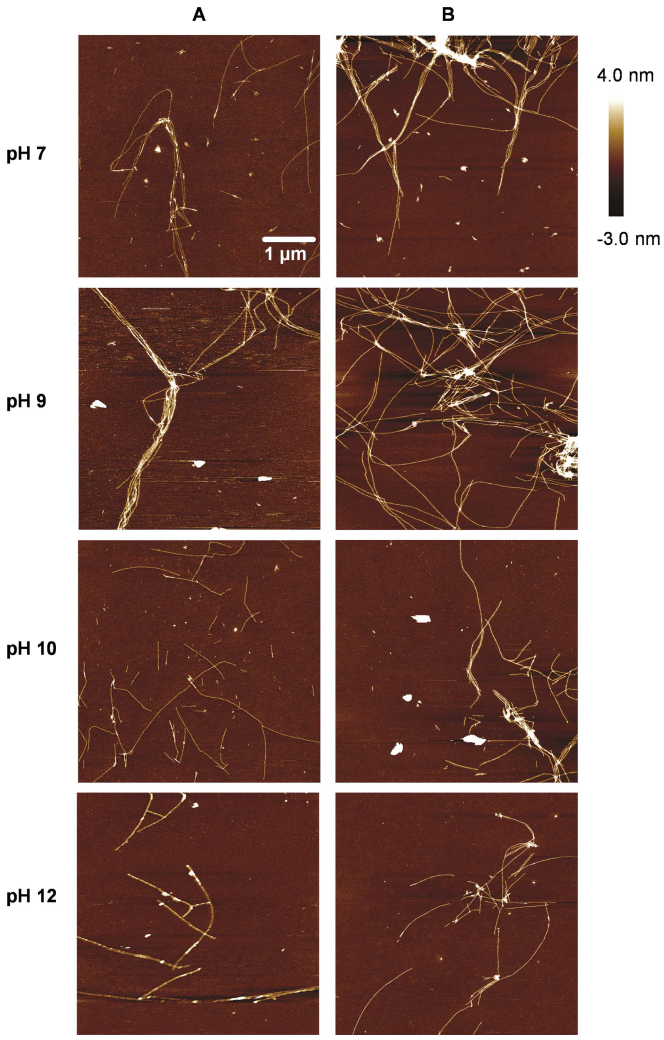


Figure 6.4: *AFM analysis of fibrils of (A)* $\mathbf{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^A}$ *and (B)* $\mathbf{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^B}$ *as a function of pH, 1 g* L^{-1} *protein solutions were incubated for 24 h at different pH values at room temperature, followed by 20 min of deposition on the silica wafer, rinsing, and drying.*

6.3.3. Heterodimer-driven fibrils bundling

It has been shown previously that mixing of $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ with an N-terminally extended variant does not interfere with fibril formation, and that hybrid fibrils are formed with both polymer types incorporated into one fibril [38]. We here similarly mixed an excess of $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ with its modified linker-containing variants with the aim of decreasing the local concentration of $\mathbf{D^A}$ and $\mathbf{D^B}$, while

maintaining a constant total protein concentration. This allows the formation of fibrils (at pH 7.4) that contain only a small fraction of randomly incorporated $\mathbf{D^A}$ and $\mathbf{D^B}$ modules. The protein polymers with heterodimerizing $\mathbf{D^A/D^B}$ modules together constituted 1, 2, 4, or 8% of the total protein polymer concentration (1 g L^{-1}). To determine whether homotypic $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ still play a role, we also studied fibrils formed from solutions of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ mixed with either $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ (Fig. 6.5).

We find that, for the control sample where $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ was mixed with only $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$, fibrils did not form bundles at a $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ content of up to 8%. For the control sample consisting of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$, however, we find that at a relative content of 4% and 8%, fibrils are already spread less evenly over the surface, and show some bundling. A $\mathbf{D^B}$-content of 4 % corresponds to an average distance along the fiber of ~25 nm between the $\mathbf{D^B}$-modules, similar to the average intermolecular distance in a ~100 μM solution. Thus, these fiber results correspond well with our previous findings on concentration dependent $\mathbf{D^B}$ homodimer formation [31]. For the samples containing $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ and different amounts of an equimolar mixture of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$, we observe a clear bundling of fibrils, except at the lowest relative content of linker-containing polymers of 1%. AFM images of samples with 2 and 4% of $\mathbf{D^A/D^B}$-containing proteins appear similar, and show fibrils of a few microns long, linked into straight bundles. For the 8% sample, bundling was much more extensive.

The controls show that the $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ homotypic interactions only occur when the fraction of linkers is high. Also, the fact that the homotypic interaction is strongest for the $\mathbf{D^B}$ linker, is in agreement with our previous findings [31]. This shows that only a small fraction of $\mathbf{D^A/D^B}$ modules need to be incorporated into $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ fibrils to quite drastically change the organization of the fibrils.

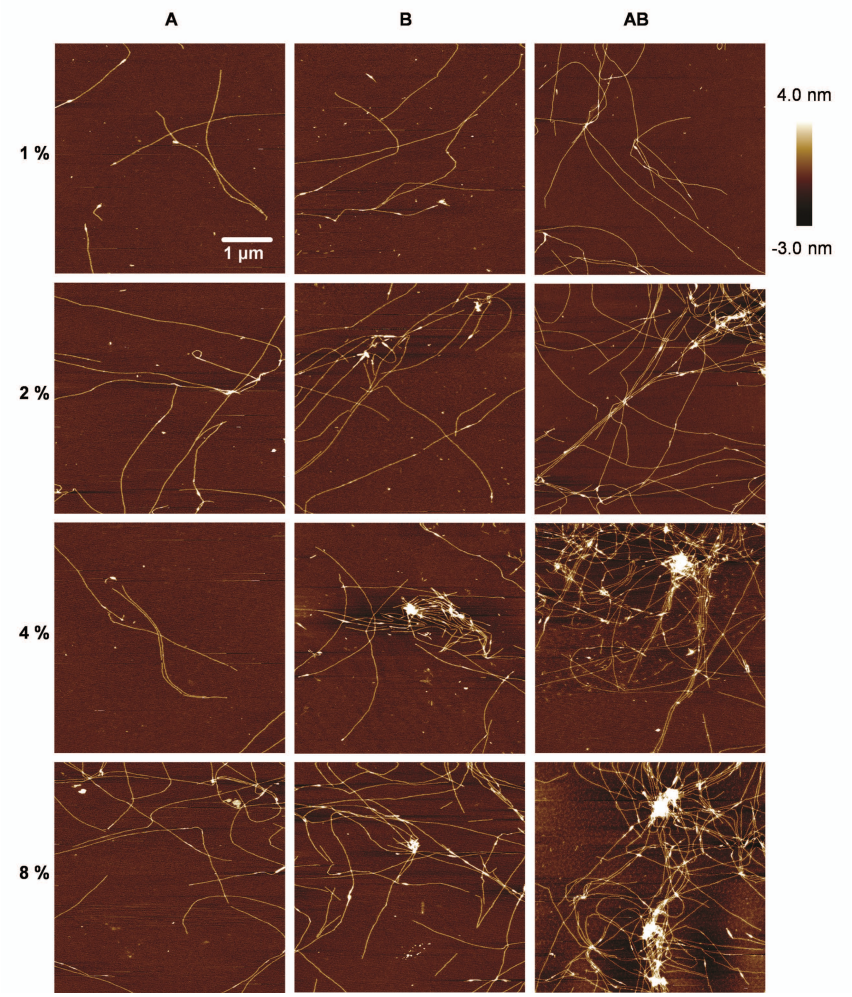


Figure 6.5: *Control over* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$ *cross-linking and bundling using* D^A/D^B *heterodimers. AFM images are shown of* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$ *protein solutions containing different weight fractions of (A)* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$-D^A*, (B)* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$-D^B*, or (AB) an equimolar mixture of* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$-D^A *and* $C^P{}_2$-$S^H{}_{48}$-$C^P{}_2$-D^B*. All proteins solutions were 1 g* L^{-1}*, and were incubated for 24 h at pH 7.4 at room temperature, followed by 20 min of deposition on the silica wafer, rinsing, and drying.*

6.3.4. Mechanical properties of C_2-$S^H{}_{48}$-C_2-D^A and C_2-$S^H{}_{48}$-C_2-D^B hydrogels

6.3.4.1. Gel formation and recovery

The gelation behavior of C_2-$S^H{}_{48}$-C_2-D^A and C_2-$S^H{}_{48}$-C_2-D^B polymers was investigated using oscillatory shear measurements. Prior to the measurement,

the pH was raised to 7.4 in each sample, to trigger gelation. All measurements were performed at a constant temperature of 20 °C. The gelation behavior of the two protein polymers was compared with that of the previously developed $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ [19, 24]. At a concentration of 10 g L^{-1}, all samples formed transparent hydrogels. Visually, the consistency of the protein solutions appeared to change immediately after the pH was increased, and approximately 2 hours later, all samples had formed self-supporting gels.

Gel formation and gel recovery after mechanical failure were followed in time by continuous measurement of the storage (G′) modulus of 10 g L^{-1} protein polymer solutions, at an oscillation frequency of 1 Hz, and at a shear strain of 0.1 % (Fig. 6.6A). The speed of gelation was similar for all three proteins. After 10 h of gelation, the storage modulus of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ gel was 182 Pa, whereas that of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ gel was 276 Pa and 409 Pa, respectively. The storage modulus (G') dominated the loss modulus (G") in all cases. After rupture, network recovery was observed for all gels. The recovery of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ gel was nearly complete after 10 h of healing and reached a storage modulus of 142 Pa, in agreement with previous results [19, 24]. In contrast, $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ gels recovered only partly with respect to their original modulus. After 10 h of healing, the storage modulus for $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ hydrogel reached a final value of 172 Pa and the storage modulus of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ hydrogel reached the final value of 181 Pa. After fracture, all three hydrogels had a similar modulus.

The increased storage moduli of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ hydrogels, in comparison to the storage modulus of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ hydrogel, indicate that both $\mathbf{D^A}$ and $\mathbf{D^B}$ contributed to network formation via homotypic interactions. Moreover, the storage modulus of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ was higher than the storage modulus of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$, which again suggests the $\mathbf{D^B}$ module has stronger homotypic interactions than the $\mathbf{D^A}$ module, as we concluded earlier [31]. The inability of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ hydrogels to fully recover from mechanical failure suggests that fibril-fibril interactions mediated by homotypic $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ association somehow hindered or blocked recovery. Previous findings by our group showed that while full recovery of pure $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ hydrogels was obtained after mechanical failure at lower concentrations, only partial recovery was observed at higher concentrations [19, 24]. The protein polymer variant $\mathbf{C_2\text{-}S^E_{48}\text{-}C_2}$, which contains a glutamic acid instead of a histidine, was previously shown to form fibrillar hydrogels with a higher storage modulus than $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ but with no capacity to recover, after mechanical failure [39]. As suggested by Golinska et al. [19], while

stronger physical cross-links may give rise to higher moduli, they may possibly also impede recovery of the gel after failure.

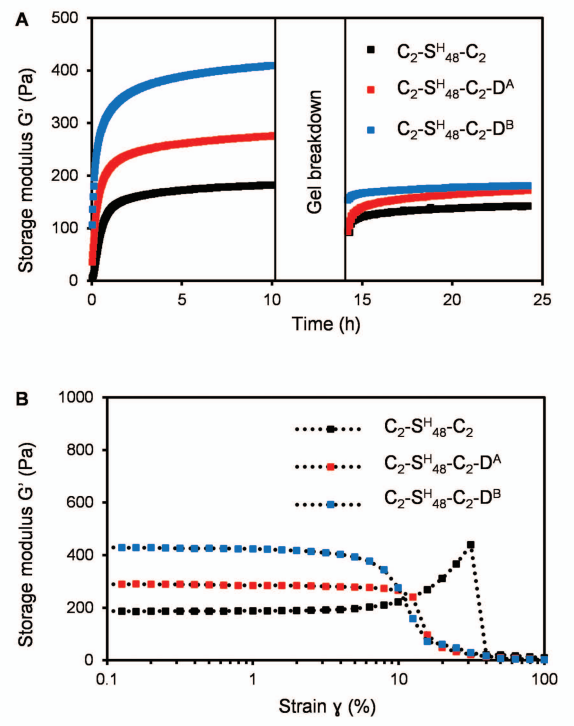


Figure 6.6: *(A) Time course of storage modulus G' for 10 g L^{-1} gels made of* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2}$ *(black),* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^A}$ *(red), or* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^B}$ *(blue). The gels were formed at pH 7.4 and 20 °C. After 10 h the gels were broken by applying a 100% deformation and the recovery was measured as a function of time. Measurements were made in a Couette configuration at 1 Hz and 0.1 % deformation. (B) Strain sweep at 1 Hz for 10 g L^{-1} gels made of* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2}$ *(black),* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^A}$ *(red),* *or* $\mathbf{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^B}$ *(blue). The storage modulus (G') is plotted as a function of deformation (γ).*

6.3.4.2. Stiffening behavior

For the samples discussed in the previous paragraph we also tested whether they exhibited strain stiffening, which is a hallmark feature in the rheology of many fibrillar hydrogel systems such as the ECM. Hydrogels were exposed to a strain sweep up to the point of mechanical failure. While hydrogels composed of the original $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ triblock protein polymer showed significant strain stiffening before failure, this was not the case for hydrogels composed of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ only (Fig. 6.6B). Note that strain stiffening has

not only been reported for hydrogels of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$, but also for hydrogels of the glutamic acid-containing $\mathbf{C_2\text{-}S^E_{48}\text{-}C_2}$ variant [39]. Both $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ and $\mathbf{C_2\text{-}S^E_{48}\text{-}C_2}$ hydrogels are composed of mostly non-crosslinked and non-bundled semi-flexible fibrils that undergo significant thermal deformations [19, 39]. The characteristic strain-hardening of fibrillar hydrogels occurs when stretching forces have pulled out all thermal deformations of the fibrils [40]. The extensive cross-linking and bundling of the fibrils in $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ may have affected the thermal deformations of the fibrils in such a way that shear stiffening was abolished.

6.3.5. Mechanical properties of mixed $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2/C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A/C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ hydrogels as a function of linker concentration

6.3.5.1. Gel formation and recovery

Next, we studied hydrogels composed of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ plus different amounts of an equimolar mixture of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$. Here we can use the fraction of protein polymers with $\mathbf{D^A/D^B}$ linkers as a tuning parameter for the degree of fibril cross-linking and bundling. The total protein polymer concentration was again 10 g L^{-1} and the weight fractions of protein polymers with $\mathbf{D^A/D^B}$ linkers were chosen to be 2, 4, 8, 16, 32, 64, 80, and 100% (so e.g. 2% means 1% $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and 1% $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$). Gelation and recovery after mechanical failure were studied as in the previous paragraph. The storage modulus obtained after 10 h of incubation time was plotted against the fraction of protein polymers with $\mathbf{D^A/D^B}$ in the fibrils (Fig. 6.7). The final modulus of the hydrogels increased with the fraction of protein polymers with $\mathbf{D^A/D^B}$ modules. This shows that the $\mathbf{D^A/D^B}$ modules do indeed strengthen the network of fibrils by cross-linking and bundling the fibrils. Since the rheology experiments were done at concentrations that are a factor 10 higher than those used in the AFM experiments, we expect that especially for the fibrils with a higher content of the $\mathbf{D^A/D^B}$ modules, this strengthening is a consequence of both hetero- and homotypic associations. Consistent with the results obtained with hydrogels made from either $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ alone, we find that the incorporation of more $\mathbf{D^A/D^B}$ modules impedes recovery after mechanical failure. For a few selected samples, Fig. 6.8A shows the development of the storage modulus as a function of time, during gel formation and during recovery. Rheological data for additional hydrogels, including frequency sweeps as the analysis of the behaviour of the samples at different changes of stress, can be found in Supplementary material, Fig. 6.1.S2.

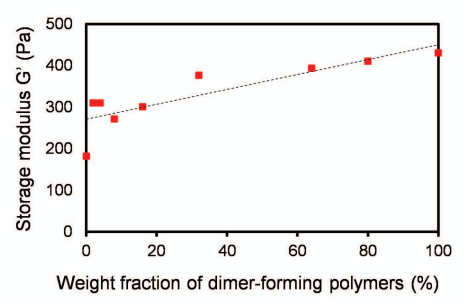


Figure 6.7: *Storage modulus (G') after 10 h of 10 g L^{-1}* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}}$ *gels containing different weight fractions of a 1:1 mixture of* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}\text{-}D^{A}}$ *and* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}\text{-}D^{B}}$.

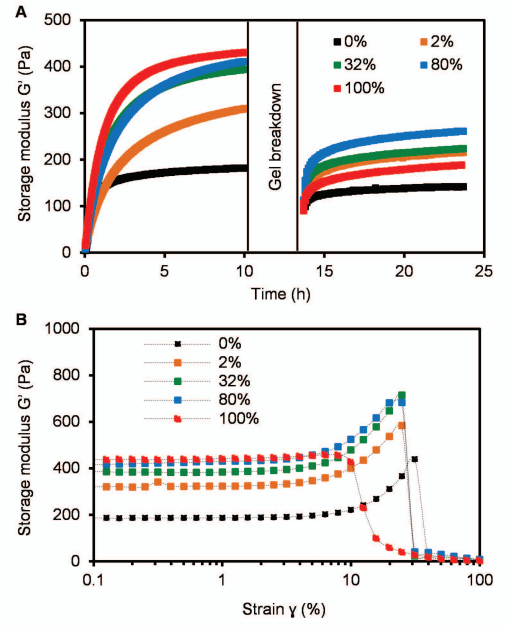


Figure 6.8: *(A) Time course of storage modulus G' for 10 g L^{-1} gels made of* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}}$ *containing different fractions of a 1:1 mixture of* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}\text{-}D^{A}}$ *and* $\mathbf{C^{P}_{2}\text{-}S^{H}_{48}\text{-}C^{P}_{2}\text{-}D^{B}}$. *(B) Strain sweep at 1 Hz for the same samples.*

6.3.5.2. Stiffening behavior

Also for the hydrogels with mixed fibrils, we studied their possible strain hardening. In Fig. 6.8B, the storage modulus is plotted against strain for hydrogels formed by mixed fibrils containing a range of concentrations of the $\mathbf{D^{A}/D^{B}}$ linkers.

We find that strain hardening is lost completely for the case of mixed fibrils with 100% $\mathbf{D^{A}/D^{B}}$ linkers (i.e. fibrils that contain 50% of $\mathbf{C_{2}\text{-}S^{H}_{48}\text{-}C_{2}\text{-}D^{A}}$ and

50% of $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^B}$). Hydrogels composed of fibrils with 2%, 32% and 80% of linkers did show an upturn of the storage modulus, but were broken at lower values of the strain than hydrogels composed of fibrils without linkers.

6.3.6. Concentration dependence

We studied the concentration dependence of the modulus of hydrogels composed exclusively of polymers with linkers, i.e. gels that contain 50% of $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^A}$ and 50% of $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2\text{-}D^B}$. Storage moduli were determined after 10 h of gelation for samples with total protein polymer concentrations of 5, 7.5, 10, 15, 20, and 30 g L^{-1}, both for the mixed fibrils with 100% linkers and for a $\mathbf{C_2\text{-}S^H{}_{48}\text{-}C_2}$ control (without linkers). Results are shown in Fig. 6.9. For all concentrations, the storage moduli obtained were higher for the hydrogels composed of fibrils with linkers, but otherwise the concentration dependence was quite similar for the cases with and without linkers. At a concentration of 5 g L^{-1}, a storage modulus could be detected for the hydrogels with linkers, but not for the hydrogels without the linkers.

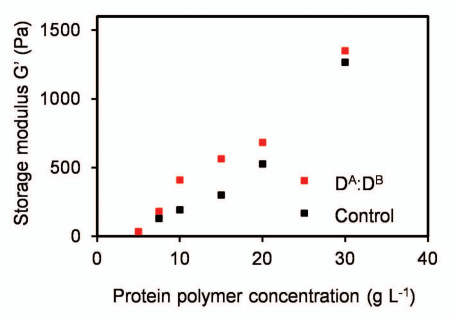


Figure 6.9: *Storage modulus (G') as a function of concentration of a 1:1 mixture of* $\boldsymbol{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^A}$ *and* $\boldsymbol{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2\text{-}D^B}$ *(red), and of a* $\boldsymbol{C^P{}_2\text{-}S^H{}_{48}\text{-}C^P{}_2}$ *control (black).*

6.3.7. Separation of $\mathbf{D^A}$ and $\mathbf{D^B}$ in different fibers

We tried to investigate whether there is a difference between a situation where the $\mathbf{D^A/D^B}$ modules are mixed within each fibril (as a result of mixing solutions prior to inducing self-assembly, as considered here), and a situation where fibrils with only $\mathbf{D^A}$ modules are mixed with fibrils containing only $\mathbf{D^B}$ modules. We expected stronger fibril-fibril interactions in the latter case. Unfortunately it is difficult to realize such a mixture of '$\mathbf{D^A}$-only' and '$\mathbf{D^B}$-only' fibers in practice. One would need to mix samples that have already gelled, or mix dilute samples and then concentrate.

As a preliminary test, we performed the following experiment. At a low concentration of 7.5 g L^{-1}, we induced fibril formation by pH adjustment, and allowed fibril formation to proceed for 2h for separate $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ samples. Next, these solutions were mixed. We also prepared a control sample that was mixed before pH adjustment as normal. We find that delaying the mixing of the protein polymers with $\mathbf{D^A}$ and $\mathbf{D^B}$ linkers by 2h, leads to an increase of the modulus by a factor of about 2 (Fig. 6.1.S3). This is an interesting result, which suggests there is much to be gained in terms of achieving still stronger gels when a convenient procedure could be developed to prepare concentrated hydrogels consisting of fibrils with only $\mathbf{D^A}$ and fibrils with only $\mathbf{D^B}$ linkers.

6.4. Conclusion

We have produced recombinant protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ with $\mathbf{D^A}$ and $\mathbf{D^B}$ linkers that mediate cross-linking and bundling of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ fibrils, via $\mathbf{D^A/D^B}$ heterodimerization. The system allows for a systematic control over the extent of cross-linking and bundling of the fibrils, by varying the concentration of linkers in the fibrils, even though unintended homotypic $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ associations probably do play a role at higher concentrations of linkers in the fibrils. This is an important result, since in many systems developed as artificial ECM, there is some degree of fibril cross-linking and bundling, but it typically cannot be tuned [28]. We found that in mixed hydrogels consisting of only protein polymers with $\mathbf{D^A/D^B}$ linkers, both strain stiffening and full recovery after mechanical failure were lost. This is reminiscent of results for collagen type I hydrogels, where the degree of strain stiffening was found to be smaller for gels at higher concentration and in those displaying thick fibril bundles [41].

The cross-linking and bundling approach we have developed here can be used as an alternative to our previously developed approach, in which heparin was used to cross-link and bundle heparin-binding fibrils[30]. More generally, we have made an important first step to the design of protein polymers for which fibril cross-linking and bundling can be controlled by merely adjusting fibril composition through mixing. Given the exquisite sensitivity of cells for the structure and mechanics of the ECM, the development of approaches that enable control over the nano- and microscale architecture of fibrillar hydrogels is an important challenge for the development of better artificial ECM hydrogels with properties matched to the needs of specific cell types.

Supplementary data 6.1

Table 6.1.S1: *Amino acid sequence of newly constructed proteins. The silk-like* S^{H}_{48} *blocks are underlined, and the heterodimer-forming modules are highlighted in bold red.*

Code	Amino acid Sequence
$C_2\text{-}S^{H}_{48}\text{-}C_2\text{-}D^{A}$	YVEFGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQG NPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPG NPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPG EQGKPGNQGPAGEGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG AGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG HGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGHG AGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAG AGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG AGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAG HGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQP GNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKP GNQGPAGEGA**LEIRAAFLRQRNTALRTEVAELEQEVQRLENEVSQYETRYGPLGGGKG**
$C_2\text{-}S^{H}_{48}\text{-}C_2\text{-}D^{B}$	YVEFGLGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQG NPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPG NPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPG EQGKPGNQGPAGEGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG AGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG HGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGHG AGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAG AGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAG AGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAGHGAGAGAG HGAGAPGEPGNPGSPGNQGQPGNKGSPGNPGQPGNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKN GQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKPGNQGPAGEPGNPGSPGNQGQPGNKGSPGNPGQP GNEGQPGQPGQNGQPGEPGSNGPQGSQGNPGKNGQPGSPGSQGSPGNQGSPGQPGNPGQPGEQGKP GNQGPAGEGA**LEIEAAFLERENTALETRVAELRQRVQRLRNRVSQYRTRYGPLGGGKG**

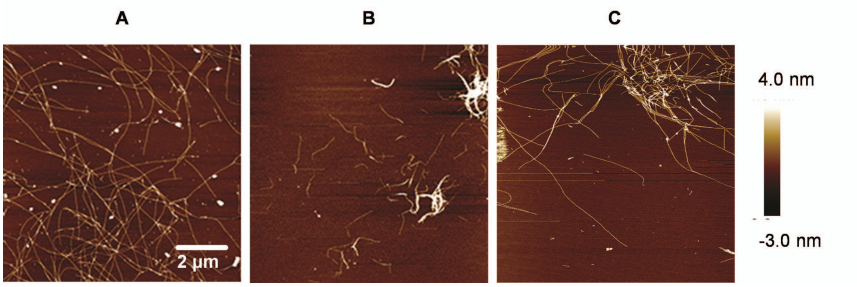


Figure 6.1.S1: *The distribution of fibrils on the surface of the AFM silica wafer. Fibrils formed by (A)* C_2-S^H_{48}-C_2 *are evenly spread, while (B)* C_2-S^H_{48}-C_2-D^A *and (C)* C_2-S^H_{48}-C_2-D^B *are gathered in patches exposing the empty surface.*

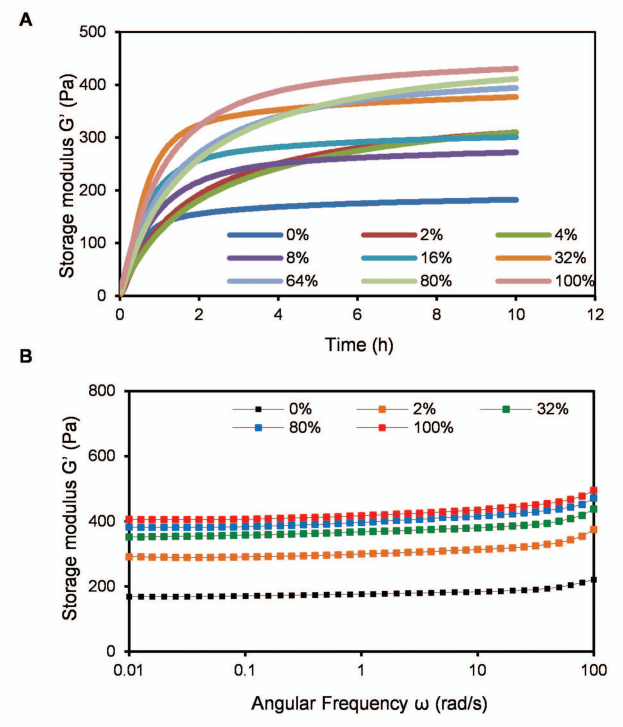


Figure 6.1.S2: *(A) Time course of storage modulus G' for 10 g* L^{-1} *gels made of* C^P_2-S^H_{48}-C^P_2 *containing different fractions of a 1:1 mixture of* C^P_2-S^H_{48}-C^P_2-D^A *and* C^P_2-S^H_{48}-C^P_2-D^B. *(B) Angular frequency (ω) dependence of the storage modulus (G') at 20 ˚C.*

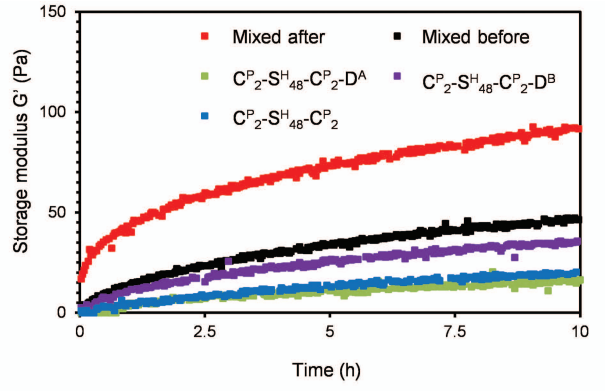


Figure 6.1.S3: *Time course of storage modulus G' for 7.5 g m^{-1} gels made of* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2}$ *(blue),* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^A}$ *(green),* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^B}$ *(purple), and* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^A}$ *and* $\boldsymbol{C^P_2\text{-}S^H_{48}\text{-}C^P_2\text{-}D^B}$ *mixed in a 1:1 ratio before pH was adjusted to 7.4 (black) or 2 h after pH was adjusted to 7.4 (red).*

References

1. Halper, J.; Kjaer, M. In *Progress in Heritable Soft Connective Tissue Diseases*, Springer: 2014; pp 31-47.

2. Watt, F. M.; Huck, W. T. *Nat Rev Mol Cell Biol* **2013,** 14, (8), 467-473.

3. Pellowe, A. S.; Gonzalez, A. L. *WIREs Nanomed Nanobiotechnol* **2016,** 8, (1), 5-22.

4. Annabi, N.; Tamayol, A.; Uquillas, J. A.; Akbari, M.; Bertassoni, L. E.; Cha, C.; Camci‐Unal, G.; Dokmeci, M. R.; Peppas, N. A.; Khademhosseini, A. *Adv Mater* **2014,** 26, (1), 85-124.

5. Swinehart, I. T.; Badylak, S. F. *Dev Dynam* **2016**.

6. Haque, M. A.; Nagaoka, M.; Hexig, B.; Akaike, T. *Sci Technol Adv Mater* **2016**.

7. Lutolf, M.; Hubbell, J. *Nat Biotechnol* **2005,** 23, (1), 47-55.

8. Okamoto, M.; John, B. *Prog Polym Sci* **2013,** 38, (10–11), 1487-1503.

9. Park, I.-K.; Yang, J.; Jeong, H.-J.; Bom, H.-S.; Harada, I.; Akaike, T.; Kim, S.-I.; Cho, C.-S. *Biomaterials* **2003,** 24, (13), 2331-2337.

10. Ohya, S.; Nakayama, Y.; Matsuda, T. *Biomacromolecules* **2001,** 2, (3), 856-863.

11. Song, J.; Malathong, V.; Bertozzi, C. R. *JACS* **2005,** 127, (10), 3366-3372.

12. Hernandez-Gordillo, V.; Chmielewski, J. *Biomaterials* **2014,** 35, (26), 7363-7373.

13. Ponticiello, M. S.; Schinagl, R. M.; Kadiyala, S.; Barry, F. P. *J Biomed Mater Res A* **2000,** 52, (2), 246-255.

14. Sell, S. A.; Wolfe, P. S.; Garg, K.; McCool, J. M.; Rodriguez, I. A.; Bowlin, G. L. *Polymers* **2010,** 2, (4), 522-553.

15. Beniash, E.; Hartgerink, J. D.; Storrie, H.; Stendahl, J. C.; Stupp, S. I. *Acta Biomater* **2005,** 1, (4), 387-397.

16. Anderson, J. M.; Kushwaha, M.; Tambralli, A.; Bellis, S. L.; Camata, R. P.; Jun, H.-W. *Biomacromolecules* **2009,** 10, (10), 2935-2944.

17. Galler, K. M.; Cavender, A.; Yuwono, V.; Dong, H.; Shi, S.; Schmalz, G.; Hartgerink, J. D.; D'Souza, R. N. *Tissue Eng* **2008,** 14, (12), 2051-2058.

18. Silva, C. I. F.; Skrzeszewska, P. J.; Golinska, M. D.; Werten, M. W. T.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2012,** 13, (5), 1250-1258.

19. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W. T.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014,** 15, (3), 699-706.
20. Schipperus, R.; Teeuwen, R. L. M.; Werten, M. W. T.; Eggink, G.; de Wolf, F. A. *Appl Microbiol Biotechnol* **2009,** 85, (2), 293-301.
21. Silva, C. I. F.; Teles, H.; Moers, A. P. H. A.; Eggink, G.; de Wolf, F. A.; Werten, M. W. T. *Biotechnol Bioeng* **2011,** 108, (11), 2517-2525.
22. Beun, L. H.; Storm, I. M.; Werten, M. W. T.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014,** 15, (9), 3349-3357.
23. Werten, M. W. T.; van den Bosch, T. J.; Wind, R. D.; Mooibroek, H.; de Wolf, F. A. *Yeast* **1999,** 15, (11), 1087-1096.
24. Włodarczyk-Biegun, M. K.; Werten, M. W. T.; de Wolf, F. A.; van den Beucken, J. J. J. P.; Leeuwenburgh, S. C. G.; Kamperman, M.; Cohen Stuart, M. A. *Acta Biomater* **2014,** 10, (8), 3620-3629.
25. Werten, M. W. T.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng* **2001,** 14, (6), 447-454.
26. Peyton, S. R.; Ghajar, C. M.; Khatiwala, C. B.; Putnam, A. J. *Cell Biochem Biophys* **2007,** 47, (2), 300-320.
27. Pins, G. D.; Christiansen, D. L.; Patel, R.; Silver, F. H. *Biophys J* **1997,** 73, (4), 2164.
28. Kouwer, P. H.; Koepf, M.; Le Sage, V. A.; Jaspers, M.; van Buul, A. M.; Eksteen-Akeroyd, Z. H.; Woltinge, T.; Schwartz, E.; Kitto, H. J.; Hoogenboom, R. *Nature* **2013,** 493, (7434), 651-655.
21. Beun, L. H. PhD dissertation, Wageningen University, Wageningen University, The Netherlands, 2015, ISBN 978-94-6257-233-1., Wageningen University, Wageningen, 2015.
30. Włodarczyk-Biegun, M. K.; Slingerland, C. J.; Werten, M. W. T.; van Hees, I.; de Wolf, F. A.; de Vries, R.; Cohen Stuart, M. A.; Kamperman, M. *Biomacromolecules* **2016**.
31. Domeradzka, N. E.; Werten, M. W. T.; de Vries, R.; de Wolf, F. A. *Biotechnol Bioeng* **2016,** 113, (5), 953-960.
32. Zhang, W.; Bevins, M. A.; Plantz, B. A.; Smith, L. A.; Meagher, M. M. *Biotechnol Bioeng* **2000,** 70, (1), 1-8.
33. Olsen, D.; Yang, C.; Bodo, M.; Chang, R.; Leigh, S.; Baez, J.; Carmichael, D.; Perälä, M.; Hämäläinen, E.-R.; Jarvinen, M.; Polarek, J. *Adv Drug Delivery Rev* **2003,** 55, (12), 1547-1567.
34. Werten, M. W. T.; Moers, A. P. H. A.; Vong, T.; Zuilhof, H.; van Hest, J. C. M.; de Wolf, F. A. *Biomacromolecules* **2008,** 9, (7), 1705-1711.

35. Moers, A. P. H. A.; Wolbert, E. J. H.; de Wolf, F. A.; Werten, M. W. T. *J Biotechnol* **2010,** 146, (1–2), 66-73.
36. Werten, M. W. T.; Teles, H.; Moers, A. P. H. A.; Wolbert, E. J. H.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009,** 10, (5), 1106-1113.
37. Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Protein Sci* **2001,** 10, (3), 649-655.
38. Włodarczyk-Biegun, M. K.; Farbod, K.; Werten, M. W. T.; Slingerland, C. J.; de Wolf, F. A.; van den Beucken, J. J. J. P.; Leeuwenburgh, S. C. G.; Stuart, M. A. C.; Kamperman, M. *PloS one* **2016,** 11, (5), e0155625.
39. Martens, A. A.; van der Gucht, J.; Eggink, G.; de Wolf, F. A.; Cohen Stuart, M. A. *Soft Matter* **2009,** 5, (21), 4191-4197.
40. Storm, C.; Pastore, J. J.; MacKintosh, F. C.; Lubensky, T. C.; Janmey, P. A. *Nature* **2005,** 435, (7039), 191-194.
41. Motte, S.; Kaufman, L. J. *Biopolymers* **2013,** 99, (1), 35-46.

7 General Discussion

7.1. Heterodimerizing modules as means to better control protein-polymer assembly

Protein self-assemblies present in nature are remarkably complex and diverse. Their assembly and disassembly is a sensitive function of physical-chemical parameters such as pH, ionic strength and temperature [1], and this makes the design of protein-polymers that assemble into desired structures particularly difficult. At the same time, the enormous potential of proteins as building blocks for the "bottom-up" fabrication of novel biomaterials with versatile functionalities is clearly recognized in nanotechnology [2,3].

In this thesis, we explore supramolecular strategies in order to be able to better control protein polymer self-assembly. So far, in our laboratory, we have developed a number of protein polymers that can self-assemble via homotypic associations into simple supramolecular structures such as fibrils or micelles [4-6]. In this thesis we go one step further, by focusing on the construction and physical characterization of protein polymers that additionally have modules with heterotypic interactions, in particular module A and B that form A/B heterodimers but (preferably) do not show A/A or B/B association.

To put the research described in this thesis into a broader context, we first review other general strategies that have been used before to direct protein polymer assembly. We then argue why there is a need for an additional, complementary strategy involving reversible heterodimerizing modules. Next, we summarize the most important outcomes of this research and do some recommendations for further steps that could be taken. Finally, we briefly reflect on the potential applications of our results, present a risk analysis and technology assessment, finally comment on future perspectives for the field of self-assembling protein polymers.

7.2. Strategies to direct protein polymer self-assembly

7.2.1. Strategies to obtain well-defined space-filling networks

Natural protein self-assemblies often exhibit distinct symmetries, that reflect the symmetries of the underlying protein components (e.g. bacterial S-layers, or polyhedral virus particles) [1]. Building on that observation, general strategies for designing large protein assemblies have been recently proposed for building structures from units, where each unit has multiple self-assembling blocks. For example, if a protein block A that forms a tetramer is fused to a protein block B that forms a dimer, a tetrameric unit $(AB)_4$ is formed that in turn can assemble into much larger structures via B-B associations [7,8]. Regular structures can be formed in the A-B linkage is rigid. With flexible A-B linkages, more irregular

assemblies are formed [9]. Several interesting symmetrical peptide self-assemblies designs along these lines have recently been reviewed by Yeates et al. [10]. A variation on this theme is the protein polymer $\mathbf{T_9\text{-}C^P_4\text{-}T_9}$ [11] on which we develop variations in this thesis. In this protein polymer, a long hydrophilic random coil midblock ($\mathbf{C^P_4}$) chemically links two trimer-forming blocks ($\mathbf{T_9}$). Above a critical concentration, and at low enough temperatures, the resulting triblock protein polymer forms space-filling networks.

7.2.2. Strategies to obtain well-defined nanoparticles using templates strategies

Another nature-inspired strategy, templating, can be used to obtain well-defined finite-sized protein assemblies. This holds in particular for linear assemblies. Precise length control for linear protein assemblies is difficult to achieve, but a beautiful example where precise control is obtained is provided by the linear viruses, where the length of the protein assembly is set by the length of a nucleic acid template.

At least two domains are required for templated self-assembly: a domain that allows for the interaction between neighbouring protein polymers, and a domain that interacts with the template [3]. A first example of the implementation of this strategy was described by Grigoryan et al, where a peptide was designed that assembles around a carbon nanotube [12]. Another example is provided by work from our group, where a protein polymer was designed that co-assembles with DNA to form rod-shaped virus-like particles [13].

7.2.3. Biochemical strategies for covalent cross-linking

These strategies have been reviewed in the **Chapter 2** of this thesis. In brief, another powerful approach to assemble protein components is via biochemical reactions between proteins [14]. In our group we have explored use of the enzymes glutaminase [15] and sortase [16] to cross-link different type of protein polymers. In both cases it was found to be difficult to optimize the process of cross-linking. Also the need for an additional component (an enzyme) can be seen as a disadvantage. An example of biochemical cross-linking that does not need an additional component is described by the group of Tirrell *et al.* and spontaneously occurs between the peptide pair called SpyTag/SpyCatcher [17], where the SpyCatcher is in essence a small autocatalytic cross-linking enzyme.

7.2.4. Strategies for reversible cross-linking using blocks with homotypic interactions

While for some applications it is best to use covalently cross-linked protein assemblies, for other applications, reversible, physical cross-links are preferred. For instance, for the construction of smart materials that can self-heal and disassemble in response to particular stimuli such pH or temperature [18], it is crucial to have assemblies be built up via reversible physical bonds, and these are also the types of materials that we aim at in this thesis. As also described in **Chapter 2**, in reversible cross-linking strategies, protein polymer designs include blocks A that have homotypic A/A interactions that drive association of the A blocks into dimers, trimers etc. When multiple A blocks are present on a single protein polymer, space filling networks can be formed [3]. Many natural domains with known sequences can be used that show such self-associations. In addition, computational protein design is now also starting to result in functional *de novo* designed self-associating [19]. Indeed, the toolbox of available self-association domains is still expanding and one can choose between domains with widely differing sizes and association strengths [20,21]. The choice depends on the ultimate application of protein polymer self-assemblies [22]. Indeed, depending on whether we aim to fabricate a nanoreactor, a sensor or a stimuli-responsive material, requirements on the assembly and the final properties will be very different, and this translates into different choices for protein polymer designs and self-association blocks to be used [23].

7.2.5. Reversible cross-linking using blocks with heterotypic interactions

An additional level of control is obtained when not homotypic A/A associations are being used, but heterotypic A/B associations. Ideally of course, in this case, there are no homotypic A/A or B/B associations. The specific case that we focus on is that of A/B heterodimer formation. Employing heterodimerization to form supramolecular structures from protein polymers allows to construct two-component systems, that assemble upon mixing. This can also be used to direct hierarchical assembly, for example a system involving only A first assembles into thin fibrils, and then, by adding another protein polymer that contains B, the fibrils are cross-linked or bundled up into higher-order structures. Heterodimerization could be also used as a tool to combat protein production problems. For example, the low yield of production for $\mathbf{T_{16}\text{-}C^{P}{}_{4}\text{-}T_{16}}$ triblock protein polymers was attributed to network formation during biosynthesis [24]. This problem could possibly be solved by producing the protein

polymer as two halves, that later are made to stick together via A/B heterodimer association.

Heterodimers studied in this thesis have been heterodimeric coiled-coils, and WW domains and their proline-rich peptide ligands. In both cases, the mechanism of interaction is based on hydrophobic forces and/or attraction of the opposite charges [25-28]. The choice for these particular heterodimer couple was partly dictated by the medical applications that are foreseen for protein polymers [29]. Another criterion was the possibility to produce the chosen peptide sequences in the *P. pastoris* expression system that we use for protein polymer production. We chose sequences for heterodimeric couples that were short: less than 50 amino acid. This improves chances that the newly introduced blocks do not interfere too much with the behavior of the template protein polymer in which they are incorporated. Also, heterodimers should still form at physiological conditions (i.e. ~37 ˚C and ~ 7 pH). With these conditions being met, we expect that the risk that the introduction of the heterodimerizing blocks will compromise features such as biocompatibility and biodegradability is rather low. In the next paragraph, we discuss in more detail the functionality of the chosen heterodimers as blocks for controlled assembly of protein polymers.

7.3. Recombinant protein polymers with heterodimer-forming modules

7.3.1. A-B guided protein self-assemblies: from concept to practice

In this thesis, we use the previously developed protein polymers $\mathbf{C^P_4}$ [30], $\mathbf{T_9\text{-}C^P_4}$ [11] and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ [5] as template protein polymers that we extend/modify by the introduction of heterodimerizing blocks, referred hereafter as modules. The self-assembly behaviour of each these (unmodified) protein polymers in aqueous solutions is summarized in Fig. 7.1.

We aimed to constructed A and B variants of each of above mentioned protein polymers (Table 7.1). Here, A and B stand for the two different but complementary peptides that self-assemble with high affinity. By mixing variants A and B, we envisioned that several protein self-assemblies could be achieved, as illustrated in Fig. 7.2. Although the system that we envisioned is still rather simple in comparison to natural protein self-assemblies [1], it is a good starting point in a quest for more and more complex nanostructures [31].

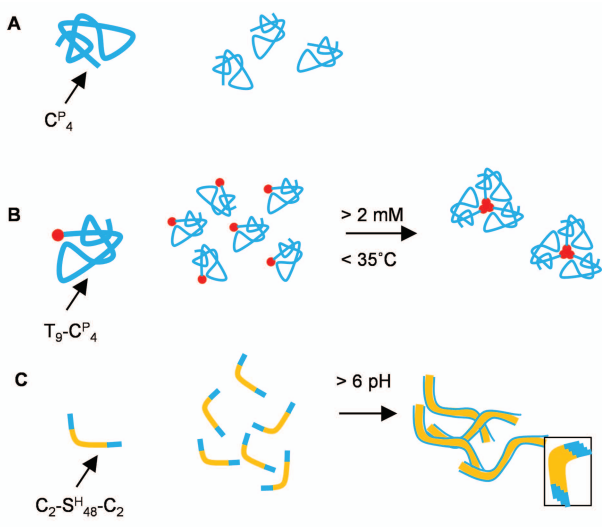


Figure 7.1: *Protein polymers used in this thesis to test in combination with heterodimer-forming modules. (A) The* $\boldsymbol{C^P_4}$ *block (blue) is a hydrophylic random coil block with a high content of glycines and prolines. (B) The* $\boldsymbol{T_9}$ *block of the* $\boldsymbol{T_9\text{-}C^P_4}$ *fusion (*$\boldsymbol{T_9}$ *block in red) forms tripe helixes at protein concentrations higher than 2 mM and temperatures below 35 ˚C. (C) The* $\boldsymbol{C_2\text{-}S^H_{48}\text{-}C_2}$ *triblock forms stiff, long fibrils at pH > 6. It consists of a silk-like midblock* $\boldsymbol{S^H_{48}}$ *(yellow)* $\boldsymbol{C_2}$ *terminal blocks (blue) that act as stabilizers. They are similar in sequence to the* $\boldsymbol{C^P_4}$ *block but shorter (half as long).*

Table 7.1: *Heterodimeric modules studied in this thesis.* ***D*** *stands for dimer-forming. The superscript indicates the identity of the modules: A and E for the more acidic and B and K for the more basic coils of the coiled/coil pairs. WW and CC43 are both WW domains, and the general abbreviation for their proline-rich ligands is PPxY.*

Codes for A/B heterodimer-forming couple		Type of interaction
D^A	D^B	Heterodimeric coiled coil [27]
D^E	D^K	Heterodimeric coiled coil [28]
D^{WW},$D^{CC43\dagger}$	D^{PPxY}, D^{PPxY*}, $D^{PPxY\text{-}1}$ or $D^{PPxY\text{-}1*\dagger}$	WW binding domain and proline-rich ligand [25,26]

† WW domains can recognize different type of proline-rich peptide ligands

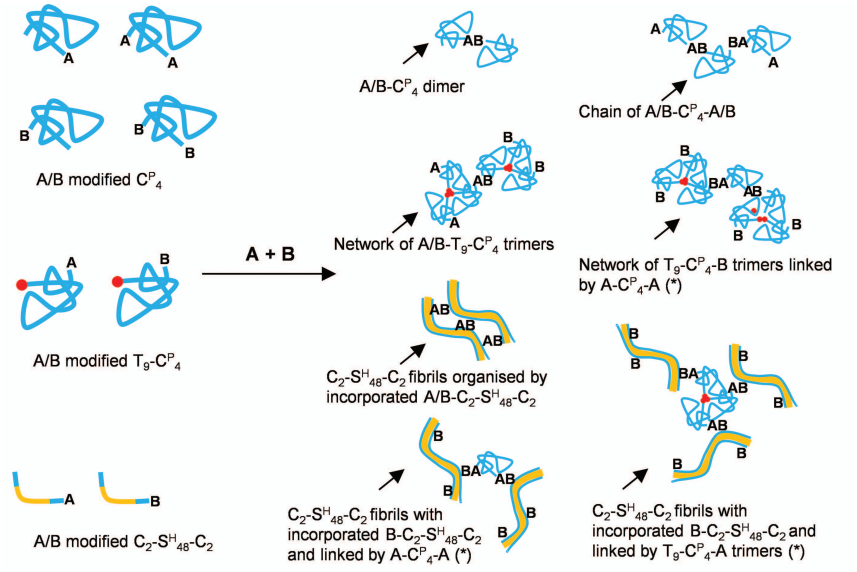


Figure 7.2: *Examples of protein self-assemblies that could be formed in various mixtures of **A** and **B** variants of C^P_4, T_9-C^P_4 and C_2-S^H_{48}-C_2. The reaction labelled as **A+B** symbolizes that variant **A** is mixed with variant **B** in a 1:1 ratio. **(*)** Here also an alternative network is possible, where A and B are reversed.*

7.4. Biosynthesis of protein polymers

For completeness, we give an overview of all new protein polymers produced in this thesis. All protein polymers were produced in *P. pastoris* production strain GS115 in a process established by Werten et al (i.e. standard fermentation) [30]. A list of produced protein polymers with $\mathbf{D^A/D^B}$ modules is given in Table 7.2. For the $\mathbf{D^E/D^K}$ pair, see Table 7.3, and for the WW/PPxY pairs, see Table 7.4. Protein polymers listed in these tables were obtained in high yields. In some cases, the standard fermentation resulted in partly degraded products. This could not be resolved by employing the protease deficient strains [32] **(Chapter 3** and **Chapter 4)**. Conditions tested to reduce degradation are also indicated in the tables. Of the parameters that were varied, we found that a low temperature in the induction phase is best to reduce degradation **(Chapter 3)**. We also showed that *Pichia*-derived glycosylation of protein polymers can be resolved by suitable modifications of the protein sequence (**Chapter 5**). In a few cases no proteins were produced during fermentation, or only protein polymer genes were constructed but no fermentations were done. These cases are listed in Table 7.5.

Table 7.2: *Protein polymers with D^A and D^B coiled coils analyzed in this thesis.*

Code	Fermentation parameters	Yields*	Product	Function
C^P_4-D^A	SMD1168 strain	0.3	Partly degraded	Not tested
	Best-guess	0.6	Nearly intact	Forms both homotypic and heterotypic bonds
T_9-C^P_4-D^A	Standard	1.2	Partly degraded	Not tested
	pH 5	0.7	Degraded	
	C conditions	0.9	Partly degraded	
	B conditions	0.2	Nearly intact	
	A conditions	0.6	Partly degraded	
	Best-guess	0.3	Nearly intact	Forms both homotypic and heterotypic bonds
C^P_4-D^B	SMD1168 strain	0.3	Partly degraded	Not tested
	Best-guess	1	Nearly intact	Forms both homotypic and heterotypic bonds
T_9-C^P_4-D^B	Standard	1.1	Partly degraded	Not tested
	C conditions	0.7	Partly degraded	
	B conditions	0.6	Nearly intact	
	A conditions	0.6	Nearly intact	
	Best-guess	0.4	Nearly intact	Forms both homotypic and heterotypic bonds
C_2-S^H_{48}-C_2-D^A	B conditions	2.3	Intact	Forms both homotypic and heterotypic bonds
C_2-S^H_{48}-C_2-D^B	B conditions	1.4	Intact	

* The gravimetrically determined yield is expressed as g of lyophilized product per L of cell-free broth,

Abbreviations of fermentation conditions: Best-guess: 20 ˚C, no additional oxygen supply, supplementation with ascorbic acid and casamino acids, **A conditions**: 20 ˚C, no additional oxygen supply, supplementation with ascorbic acid, **B conditions**: 20 ˚C, supplementation with casamino acids, **C conditions**: 30 ˚C, supplementation with casamino acids

Table 7.3: *Protein polymers with D^E and D^K coiled coils analysed in this thesis.*

Code	Fermentation parameters	Yields*	Product	Function
C^P_4-D^E	Standard	2.3	Intact	Data unclear
T_9-C^P_4-D^E	Standard	1.6	Intact	Only homotypic interaction confirmed
C^P_4-D^K	Standard	2.5	Partly degraded	Data unclear
T_9-C^P_4-D^K	Standard	2	Partly degraded	Only homotypic interaction confirmed
	pH 5	1.9	Degraded	Not tested
	Yapsin 1 disruptant strain	2.3	Partly degraded	Only homotypic interaction confirmed

* The gravimetrically determined yield is expressed as g of lyophilized product per L of cell-free broth

Table 7.4: *Protein polymers with WW domains and proline-rich domains analysed in this thesis.*

Code	Fermentation parameters	Yields*	Product	Function
C^P_4-D^{WW}	Standard	2.2	Intact	Forms heterotypic bonds
T_9-C^P_4-D^{WW}	Standard	1.8	Intact	Data unclear
D^{WW}-C^P_4-D^{WW}	Standard	2.4	Intact	Forms heterotypic bonds
C^P_4-D^{PPxY}	Standard	2.3	Glycosylated	
T_9-C^P_4-D^{PPxY}	Standard	1.6	Glycosylated	Data unclear
C^P_4-D^{PPxY*}	Standard	2.5	Intact	Forms heterotypic bonds
T_9-C^P_4-D^{PPxY*}	Standard	1.4	Intact	Data unclear
C^P_4-D^{CC43}	Standard	1.64	Degraded	Not tested
T_9-C^P_4-D^{CC43}	Standard	0.73	Degraded	
T_9-C^P_4-D^{PPxY-1}	Standard	1.85	Glycosylated	Not tested
T_9-C^P_4-$D^{PPxY-1*}$	Standard	1.6	Glycosylated	

* The gravimetrically determined yield is expressed as g of lyophilized product per L of cell-free broth

P. pastoris has become a popular host for the expression of recombinant proteins. In our laboratory, we have successfully produced many protein polymers in large quantities in *P. pastoris* [13,30,33-35]. Nonetheless, referring to the results compiled in Tables 7.2-4, when using *P. pastoris* as an expression system, it is not easy to predict how certain products will be processed and, whether they will be secreted or not. The production of recombinant proteins in *P. pastoris* requires luck, patience and empirical knowledge. Parameters influencing product yield in *P. pastoris* have been described in a few recent reviews [36-38] and such information is crucial for planning protein production in *P. pastoris*.

Table 7.5: *The list of genes encoding protein polymers that were constructed but not produced*

Code	Production stage
$\mathbf{D^{E}\text{-}C^{P}_{4}\text{-}D^{E}}$	Transformed GS115 *P. pastoris*
$\mathbf{D^{K}\text{-}C^{P}_{4}\text{-}D^{K}}$	Transformed GS115 *P. pastoris*
$\mathbf{D^{B}\text{-}C^{P}_{4}\text{-}D^{B}}$	Transformed GS115 *P. pastoris*
$\mathbf{C^{P}_{4}\text{-}D^{PPxY\text{-}1}}$	Transformed GS115 *P. pastoris*
$\mathbf{C^{P}_{4}\text{-}D^{PPxY\text{-}1*}}$	pPIC9 in *E. coli*
$\mathbf{D^{PPxY\text{-}1}\text{-}C^{P}_{4}\text{-}D^{PPxY\text{-}1}}$	Transformed GS115 *P. pastoris*
$\mathbf{D^{CC43}\text{-}C^{P}_{4}\text{-}D^{CC43}}$	Product absent in the free-cell broth
$\mathbf{D^{PPxY}\text{-}C^{P}_{4}\text{-}D^{PPxY}}$	Product absent in the free-cell broth
$\mathbf{D^{B}\text{-}C^{P}_{4}\text{-}D^{B}}$	Product absent in the free-cell broth

7.5. Recommendation for follow up research

This study has resulted in many interesting findings, and most of these have already been published in peer reviewed scientific journals. At the same time, many fascinating questions remains unanswered. With respect to the $\mathbf{D^{A}}$ and the $\mathbf{D^{B}}$ modules, fused to different protein polymers, we were able to show that they were functional, and that their dissociation constant was on the order of μM (**Chapter 3**), but we would like to have a more precise value for the K_d. Possibly, such a more precise value can be obtained, for instance, using fluorescence resonance energy transfer (FRET). Furthermore, using circular dichroism (CD), the preferential formation of heterodimers by the $\mathbf{D^{A}}$ and the $\mathbf{D^{B}}$ modules could be analyzed, as well as structures formed through homotypic

interaction of these modules [39]. This would allow defining the concentration regime where we may still find heterodimer-driven self-assembly. In case of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ hydrogels containing heterodimer-forming modules (**Chapter 6**), more information on their structure could be obtained from cryo-transmission electron microscopy. Włodarczyk-Biegun et al. already studied the pore size and erosion of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ hydrogels [29]. A similar analysis could be done for the hydrogels studied in this thesis. Referring to work of Rombouts et al. on composite hydrogels with improved mechanical properties, investigating the effect of mixing two types of protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A/D^B}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^A/D^B}$ on the mechanics of the resulting hydrogels would also be very interesting [16].

Although the $\mathbf{D^E}$ /$\mathbf{D^K}$ pair did not give the anticipated results (**Chapter 4**), further research could still be useful. First, it is important to obtain fully intact $\mathbf{D^K}$ protein polymers variants by testing different protease deficient stain, or, by exploring fermentation conditions as described in **Chapter 3**. Furthermore, also for protein polymers with this heterodimer couple it would be interesting to perform CD at different pH values. Since the intact $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$ can form hydrogels, it would be also interesting to compare its properties to properties of hydrogels formed by $\mathbf{T_9\text{-}C^P_4\text{-}T_9}$. The group of Tirrel et al, for instance, showed that hydrogels made of asymmetric protein polymers with terminal coiled coils had a tunable erosion rate [40].

In the case of the $\mathbf{D^{WW}}$ domain and its proline-rich peptide ligands (**Chapter 5**), the high K_d value obtained from ITC for $\mathbf{D^{WW}\text{-}C^P_4\text{-}D^{WW}}$, should be verified using tryptophan fluorescence studies, as it was done for $\mathbf{C^P_4\text{-}D^{WW}}$. In case of $\mathbf{T_9\text{-}C^P_4\text{-}D^{WW}/T_9\text{-}C^P_4\text{-}D^{PPxY*}}$ hydrogels, the interaction of the $\mathbf{T_9}$ block and the $\mathbf{D^{WW}}$ domain could be analysed by monitoring tryptophan fluorescence during titration of the $\mathbf{D^{WW}}$ domain protein polymer variant with $\mathbf{T_9\text{-}C^P_4\text{-}T_9}$. Since the protein polymer with $\mathbf{D^{CC43}}$ domain was degraded, we did not proceed with its further analysis. Foo et al. showed, however, that this WW domain has the strongest affinity towards $\mathbf{D^{PPxY}}$ [25]. Therefore, it would also be worthwhile to try to remedy the degradation problems for these protein polymers. In Fig. 7.2, we showed many architectures that could be achieved using A/B heterodimer-forming modules. Unfortunately, in this thesis, not all of these possibilities could be tested.

7.6. Protein polymer self-assemblies as functional nanomaterials

Protein polymers are considered as useful building blocks for the “bottom-up” fabrication [41,42] of nanostructured biomaterials. To date, dozens of peptide- and

protein-based nanomaterials have been developed as for very diverse range of applications such as nanoreactors, sensors, electronics, and stimulus-responsive materials [43,44]. Our main interest has been in eventual applications as nanostructured biomaterials for medical use [45], and for that case, certain requirements must be fulfilled simultaneously [46]. First, the protein polymer materials should display appropriate mechanical and chemical properties that trigger a desired cellular response [45]. Next, their toxicity, antigenicity, and inflammation should be lower than established thresholds [46]. The property of protein materials to decompose naturally over time into nontoxic components is a big advantage [46]. General advantages of recombinant protein polymers as biomaterials are that the protein polymers are very monodisperse, and that one can include various bioactive peptide motifs in the designs. Furthermore, they allow for programmable biodegradation and can potentially be produced at low cost and large scale [47].

So far, the only protein-polymer discussed in this thesis that has been investigated as a future biomaterial is the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ triblock, hydrogels of which have been investigated as a scaffold for cell growth, in the context of tissue engineering (TE) [29]. Results obtained were very encouraging, and warrant further efforts to improve performance of these hydrogels as an artificial extracellular matrix (ECM). This is exactly what we have done by investigating the use of heterodimerizing modules for cross-linking and bundling fibrils, thus better mimicking the actual structure of the ECM. In TE, the aim is to repair, maintain or replace tissue function [48]. TE involves growing cell in scaffolds. The scaffold is a highly porous artificial ECM used to accommodate cells and guide their growth [48]. Biomaterials made of recombinant protein polymers are considered as very promising scaffolds [49-51]. This is because of the possibility to combine protein blocks with different functionalities within one protein polymer [49]. The most commonly used protein blocks are collagen-like, silk-like, elastin-like or resilin-like [49]. A common problem is that scaffolds lack the required mechanical strength and porosity [29,49]. Coming back to the example of the $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymer, heterodimers can be used to modulate architecture of scaffold (**Chapter 6**). Heterodimers can be also used to construct hybrid scaffolds made of different types of co-assembling protein polymers. Such hybrid scaffolds have been shown to have promising mechanical properties [52]. Furthermore, heterodimers can be employed to construct two-component mixing-induced hydrogels and shear-thinning hydrogels that are very desired for applications *in situ* [25].

Protein polymers with heterodimer-forming modules can also be used for the targeted delivery and/or controlled release of therapeutic agents [53], in which medicines are delivered to specific locations in the body and released in response to local stimuli [18]. For example, heterodimers can be used to physically immobilize bioactive in a matrix from which it is slowly released.

While recombinant protein polymers have many interesting features as future biomaterials for the medical industry, they still need to improve in order to be useful for actual applications. Using supramolecular strategies such heterodimerization, may be one way to obtain biomaterials that perform better, and therefore should be investigated further. But, we should not forget to also look at the risks that might arise from use of protein polymers as future biomaterials, and this will be done next.

7.7. Risk analysis

The focus of the research described in this thesis was development of a strategy for improving performance of supramolecular and bio-inspired materials from the perspective of the future use in medicine as scaffolds for tissue engineering or controlled drug delivery. This research was designed for the Nanomaterials theme and was funded by the NanoNextNL-consortium. The mission of the Nanomaterial theme is to study how to build materials of well-defined structure via "bottom-up" approach in order to provide to society high quality, cost effective and cutting-edge products of diverse functions. As the theme presumes a transition from academic findings to industrial products, the NanoNextNL program obliges the scientists working within this theme to identify and address potential human, environmental and societal risks which may arise from their inventions.

The "bottom-up" construction based on supramolecular self-assembly has already led to development of many nanomaterials with very attractive physical and chemical properties. While still in their early stages of development, nanomaterials have already made a tremendous impact on the medical industry [23,45]. Concerning the rapid progress in the field of nanomaterials, it might be that answers on the risks posed by nanomaterials are years away and, in any event, are likely to emerge on a case-by-case basis. There is a lot of speculation around the use of nanomaterials and these negatively affect public perception. A critical challenge for the emerging nanomaterials industry is to ensure that the potential health and environmental impacts of nanomaterial fabrication and use are small [54]. Therefore, the assessment of the possible effects, from benefits to risks, and health hazards associated with exposure to

nanomaterials should be considered already at the research and development stage.

The recombinant protein polymers produced at our laboratory are still at a very early stage of development. Ultimately, the aim is to use them for medical purposes, however, specific applications have not been yet defined. The risk that arises from the production and use of protein polymers has not really been investigated yet, therefore we can only speculate here. Concerning the source of the protein polymers, the use of modified *P. pastoris* cells in production may appear hazardous to society. Generally, genetically modified organisms are perceived by the general public in a rather negative way, but this negative perception does not seem to exist for recombinant pharmaceuticals and medical devices. A good example is the recombinant production of insulin [55], which has been approved for human use in 1982. Many recombinant medicines are now available [56] and this does not seem to cause much resistance in society.

When it comes to the general process of using of *P. pastoris* for large-scale production, several approaches for assessment of this technology are available [57]. The effort to generate certified high-purity products is low compared with other expression systems, where media components or by-products of the cells' metabolism can give rise to purification problems [58]. *Pichia*-derived post-translational modification of products that could cause immunogenicity can be also escaped [37]. On the down side, the methanol required for induction, although relatively cheap, presents a fire hazard and a considerable health risk, especially in large-scale operations in which large methanol quantities must be stored on-site [59].

The recombinant protein polymers described in this thesis are made of natural amino acids and are expected to be easily biodegradable and biocompatible, meaning that they are non-toxic, hemocompatible, and histocompatible. Biocompatibility is typically first tested on cell lines selected for relevance to the envisaged application and, in the later stages, *in vivo* in animal studies and clinical studies [60]. Since there is always a risk that some recombinant protein polymers have unintended side-effects or toxicity, it is crucial that such test are performed in order for new recombinant protein polymer biomaterials to be safe and effective.

7.8 Concluding remarks and perspectives

In this study the small heterodimer-forming modules were used to trigger hierarchical self-assembly of recombinant protein polymers. We designed, produced and characterized a number of self-assembling protein polymers that

contain different heterodimer-forming modules. We presented several steps that can be applied in case of unwanted product processing in *Pichia pastoris*. Finally, we showed that coiled coils can be successfully employed to modulate morphology of pH-responsive supramolecular network consisting of silk-like proteins polymers.

What about future protein polymer designs? The enormous diversity of protein structures and functions present in nature continuously inspires scientists to come up with new designs [23]. Employing new computational techniques such as free-energy-based simulations for predicting protein native structures and ligand-binding affinities could be very useful, although it is not yet a guarantee of success when designing new proteins. The detailed analysis of natural and man-made supramolecular systems will help to elucidate the fundamental design considerations that determine the success or failure of self-assembling systems, such as the use of complementary shapes, complementary forces, and appropriate levels of plasticity [31].

From a different angle, the field of genetic engineering is also progressing rapidly and is offering better and better tools to synthesize protein polymers differing in size and complexity, even containing non-natural amino acids [61]. Looking into the future, a better understanding of cell physiology may enable the development of new expression systems with low probabilities of product degradation, proper protein folding, protein post-translational modification and the growth of cells on cheap feedstocks. This may also include systems based on plant- and mammalian cells. Furthermore, ongoing automation technology may mean that less and less human involvement will be needed for producing proteins that are both better and cheaper.

Finally, protein self-assembly is a complex phenomenon that can only be understood through a multidisciplinary effort involving disciplines such as biology, chemistry, physics and materials science. But, given its promise for new biomaterials that address important needs in society, it is also an important topic for which ground-breaking discoveries may certainly be expected in the near future.

Reference

1. Pieters, B. J.; van Eldijk, M. B.; Nolte, R. J.; Mecinović, J. *Chem Soc Rev* **2016**, 45, 24-39.
2. Yeates, T. O.; Padilla, J. E. *Curr Opin Struct Biol* **2002**, 12, 464-470.
3. Bai, Y.; Luo, Q.; Liu, J. *Chem Soc Rev* **2016**, 45, 2756-2767.
4. Golinska, M. D.; Pham, T. T.; Werten, M. W.; de Wolf, F. A.; Cohen Stuart, M. A.; van der Gucht, J. *Biomacromolecules* **2012**, 14, 48-55.
5. Golinska, M. D.; Włodarczyk-Biegun, M. K.; Werten, M. W.; Stuart, M. A. C.; de Wolf, F. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 699-706.
6. Beun, L. H.; Storm, I. M.; Werten, M. W.; de Wolf, F. A.; Cohen Stuart, M. A.; de Vries, R. *Biomacromolecules* **2014**, 15, 3349-3357.
7. Lai, Y.-T.; King, N. P.; Yeates, T. O. *Trends Cell Biol* **2012**, 22, 653-661.
8. King, N. P.; Lai, Y.-T. *Curr Opin Struct Biol* **2013**, 23, 632-638.
9. Lai, Y.-T.; Tsai, K.-L.; Sawaya, M. R.; Asturias, F. J.; Yeates, T. O. *JACS* **2013**, 135, 7738-7743.
10. Yeates, T. O. *The FASEB Journal* **2016**, 30, 250.1.
11. Werten, M. W.; Teles, H.; Moers, A. P.; Wolbert, E. J.; Sprakel, J.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2009**, 10, 1106-1113.
12. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F. *Science* **2011**, 332, 1071-1076.
13. Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F.; Bomans, P. H.; Sommerdijk, N. A.; Thies-Weesie, D. M.; Favretto, M. E.; Brock, R.; de Wolf, F. A.; Werten, M. W. *Nat Nanotechnol* **2014**, 9, 698-702.
14. Hennink, W.; Van Nostrum, C. F. *Adv Drug Deliv Rev* **2012**, 64, 223-236.
15. Goliska, M. D. Proefschrift Wageningen
Met literatuuropgave. - Met samenvatting in het Engels, Nederlands en Pools, Wageningen University, Wageningen, 2014.
16. Rombouts, W. H. Includes bibliographic references. - With summaries in English and Dutch, Wageningen University, Wageningen, 2015.
17. Zhang, W.-B.; Sun, F.; Tirrell, D. A.; Arnold, F. H. *JACS* **2013**, 135, 13988-13997.
18. Kopeček, J. *Eur J Pharm Sci* **2003**, 20, 1-16.
19. Der, B. S.; Kuhlman, B. *Curr Opin Struct Biol* **2013**, 23, 639-646.
20. Cai, L.; Heilshorn, S. C. *Acta Biomater* **2014**, 10, 1751-1760.

21. Domeradzka, N. E.; Werten, M. W.; de Wolf, F. A.; de Vries, R. *Curr Opin Biotech* **2016**, 39, 61-67.
22. Zhao, X.; Zhang, S. *Trends Biotech* **2004**, 22, 470-476.
23. Zhang, S.; Marini, D. M.; Hwang, W.; Santoso, S. *Curr Opin Chem Biol* **2002**, 6, 865-871.
24. Silva, C. I.; Skrzeszewska, P. J.; Golinska, M. D.; Werten, M. W.; Eggink, G.; de Wolf, F. A. *Biomacromolecules* **2012**, 13, 1250-1258.
25. Foo, C. T. W. P.; Lee, J. S.; Mulyasasmita, W.; Parisi-Amon, A.; Heilshorn, S. C. *PNAS* **2009**, 106, 22067-22072.
26. Mosser, E. A.; Kasanov, J. D.; Forsberg, E. C.; Kay, B. K.; Ney, P. A.; Bresnick, E. H. *Biochemistry* **1998**, 37, 13686-13695.
27. Moll, J. R.; Ruvinov, S. B.; Pastan, I.; Vinson, C. *Prot Sci* **2001**, 10, 649-655.
28. Litowski, J.; Hodges, R. *J Pept Res* **2001**, 58, 477-492.
29. Włodarczyk-Biegun, M. K.; Werten, M. W.; de Wolf, F. A.; van den Beucken, J. J.; Leeuwenburgh, S. C.; Kamperman, M.; Stuart, M. A. C. *Acta Biomater* **2014**, 10, 3620-3629.
30. Werten, M. W.; Wisselink, W. H.; Jansen-van den Bosch, T. J.; de Bruin, E. C.; de Wolf, F. A. *Protein Eng* **2001**, 14, 447-454.
31. Whitesides, G. M.; Grzybowski, B. *Science* **2002**, 295, 2418-2421.
32. Werten, M. W.; de Wolf, F. A. *Appl Environ Microbiol* **2005**, 71, 2310-2317.
33. Silva, C. I.; Teles, H.; Moers, A. P.; Eggink, G.; de Wolf, F. A.; Werten, M. W. *Biotechnol Bioeng* **2011**, 108, 2517-2525.
34. Werten, M. W.; Moers, A. P.; Vong, T.; Zuilhof, H.; van Hest, J. C.; de Wolf, F. A. *Biomacromolecules* **2008**, 9, 1705-1711.
35. Schipperus, R.; Eggink, G.; de Wolf, F. A. *Biotechnol Progr* **2012**, 28, 242-247.
36. Daly, R.; Hearn, M. T. *J Mol Recogn* **2005**, 18, 119-138.
37. Garcia-Ortega, X.; Adelantado, N.; Ferrer, P.; Montesinos, J. L.; Valero, F. *Process Biochem* **2016**, 51, 681-691.
38. Tacuri, M.; Flores, D.; Anrango, M.; Naranjo, B.; Koch, A.; Cumbal, L. *Biol Med* **2016**, 2016.
39. Greenfield, N. J. *Protein-Protein Interactions: Methods and Applications* **2004**, 55-77.
40. Shen, W.; Zhang, K.; Kornfield, J. A.; Tirrell, D. A. *Nat Mater* **2006**, 5, 153-158.
41. Drexler, K. E. *PNAS* **1981**, 78, 5275-5278.

42. Zhang, S. *Nat Biotechnol* **2003**, 21, 1171-1178.
43. Rajagopal, K.; Schneider, J. P. *Curr Opin Struct Biol* **2004**, 14, 480-486.
44. De La Rica, R.; Matsui, H. *Chem Soc Rev* **2010**, 39, 3499-3509.
45. Langer, R.; Tirrell, D. A. *Nature* **2004**, 428, 487-492.
46. Peppas, N. A.; Langer, R. *Science* **1994**, 263, 1715-1720.
47. Frandsen, J. L.; Ghandehari, H. *Chem Soc Rev* **2012**, 41, 2696-2706.
48. Yang, S.; Leong, K.-F.; Du, Z.; Chua, C.-K. *Tissue Eng* **2001**, 7, 679-689.
49. Gomes, S.; Leonor, I. B.; Mano, J. F.; Reis, R. L.; Kaplan, D. L. *Prog Polym Sci* **2012**, 37, 1-17.
50. Sengupta, D.; Heilshorn, S. C. *Tissue Eng* **2010**, 16, 285-293.
51. Place, E. S.; Evans, N. D.; Stevens, M. M. *Nat Mater* **2009**, 8, 457-470.
52. Lv, Q.; Hu, K.; Feng, Q.; Cui, F. *J Biomed Mater Res A* **2008**, 84, 198-207.
53. Ghandehari, H.; Cappello, J. *Pharm Res* **1998**, 15, 813-815.
54. Wiesner, M. R.; Lowry, G. V.; Alvarez, P.; Dionysiou, D.; Biswas, P. *Environ Sci Technol* **2006**, 40, 4336-4345.
55. Johnson, I. S. *Science* **1983**, 219, 632-637.
56. Sanchez-Garcia, L.; Martín, L.; Mangues, R.; Ferrer-Miralles, N.; Vázquez, E.; Villaverde, A. *Microb Cell Fac* **2016**, 15, 1.
57. Harms, J.; Wang, X.; Kim, T.; Yang, X.; Rathore, A. S. *Biotechnol Progr* **2008**, 24, 655-662.
58. Fischer, R.; Drossard, J.; Emans, N.; Commandeur, U.; Hellwig, S. *Biotechnol Appl Biochem* **1999**, 30, 117-120.
59. Potvin, G.; Ahmad, A.; Zhang, Z. *Biochem Eng J* **2012**, 64, 91-105.
60. Neffe, A. T.; Lendlein, A. *Adv Healthc Mater* **2015**, 4, 642-645.
61. Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr Opin Biotechnol* **2003**, 14, 603-609.

8 Summary

Supramolecular assemblies formed by protein polymers are attractive candidates for future biomaterials. Ideally, one would like to be able to define the nanostructure, in which the protein polymers should self-assemble, and then design protein polymer sequences that assemble exactly into such nanostructures. Despite progress towards 'programmability' of protein polymer self-assembly, we do not yet have such control. This holds especially for hierarchical structures such as self-assembled fibril bundles, where one would like to have independent control over the structures at the different length-scales. In this thesis we explore the use of heterodimerization as a strategy to control self-assembly of protein polymers at multiple length-scales. We tested a selected set of heterodimer-forming peptide modules. The heterodimer-forming modules are genetically incorporated at the C-terminus of protein polymers with a previously characterized self-assembly behavior. Several newly constructed protein polymers were biosynthesized in the yeast *Pichia pastoris* and, for these new protein polymers we investigated whether the inclusion of the heterodimer-forming blocks improved the control over the assembly of nanostructures.

The incorporation of heterodimer-forming modules into protein polymers is not the only tool that can be used for improving programmability of assembly. In **Chapter 2** we present an overview of several tools that can be use, and we highlighted their advantages and disadvantages.

In **Chapter 3** we test *de novo* designed heterodimerizing coiled coils $\mathbf{D^A}$ = LEIRAAFLRQRNTALRTEVAELEQEVQRLENEVSQYETRYGPLGGGK and $\mathbf{D^B}$ = LEIEAAFLERENTALETRVAELRQRVQRLRNRVSQYRTRYGPLGGGK. These peptides were fused to hydrophilic random coil protein polymer ($\mathbf{C^P{}_4}$) and homotrimer forming protein polymer ($\mathbf{T_9\text{-}C^P{}_4}$). We present data on the production, characterization and functionality for four new protein polymers: $\mathbf{C^P{}_4\text{-}D^A}$, $\mathbf{C^P{}_4\text{-}D^B}$, $\mathbf{T_9\text{-}C^P{}_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P{}_4\text{-}D^B}$. When the new protein polymers were produced using the fermentation process established previously for other protein polymers such as $\mathbf{C^P{}_4}$ (i.e. standard fermentation), we found the new protein polymers to be partly degraded. The use of a protease deficient strain, as well as changes in aeration or pH were found ineffective in preventing degradation, but nearly intact products were obtained from a fermentation in which the induction was done at 20 ˚C and in which the medium was supplemented with casamino acids. With respect to the physical properties of the new protein polymers, size exclusion chromatography (SEC) showed that an equimolar mixture of $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$ contained mostly dimers, whereas unmixed $\mathbf{C^P{}_4\text{-}D^A}$ and $\mathbf{C^P{}_4\text{-}D^B}$ contained only monomers. However, we also

found that $\mathbf{C^P_4\text{-}D^B}$ forms homooligomers at concentrations ≥100 μM. A mixture of $\mathbf{T_9\text{-}C^P_4\text{-}D^A}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^B}$ forms a hydrogel, most probably due to both homotypic and heterotypic $\mathbf{D^A/D^B}$ associations. We conclude that when used at low concentration, this pair of coiled coils seems to be suitable to control self-assembly of protein polymers produced in *Pichia Pastoris*.

Next, in **Chapter 4** we test another pair of *de novo* designed coiled coils. These are much shorter and have lower reported values of the association constant as compared to the $\mathbf{D^A/D^B}$ coiled coils. The systems consist of a peptide $\mathbf{D^E}$ = $(EIAALEK)_3$ and a peptide $\mathbf{D^K}$ = $(KIAALKE)_3$. The two peptides were C-terminally fused to protein polymers $\mathbf{C^P_4}$ and $\mathbf{T_9\text{-}C^P_4}$. The standard fermentations resulted in intact $\mathbf{C^P_4\text{-}D^E}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$, but protein polymers $\mathbf{C^P_4\text{-}D^K}$ and $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$ were found to be partly degraded. The degradation of variants with $\mathbf{D^K}$ module could not be readily resolved by fermentation at higher pH or using proteases deficient strain. For $\mathbf{C^P_4\text{-}D^K}$, ion exchange chromatography showed that about 40% of protein polymer (by mass) was intact. We find that for this pair of coiled-coils, homotypic interactions are so strong that they can drive gel formation in the case of $\mathbf{T_9\text{-}C^P_4\text{-}D^E}$, and a strong increase in viscosity for $\mathbf{T_9\text{-}C^P_4\text{-}D^K}$. Mixtures of the complimentary triblocks also form hydrogels, but it is not yet clear to what extent this is due to homotypic $\mathbf{D^E/D^E}$ and $\mathbf{D^K/D^K}$ associations, and to what extent it is due to $\mathbf{D^E/D^K}$ heterodimer formation.

A very different type of heterodimer-forming block is the so-called WW domain that is found in many natural proteins, and which forms heterodimers with proline-rich peptides **PPxY**. In **Chapter 5** we test the interaction between a naturally occurring WW domain ($\mathbf{D^{WW}}$) and its proline-rich ligand ($\mathbf{D^{PPxY}}$). Both were C-terminally fused to the hydrophilic random coil protein polymer $\mathbf{C^P_4}$. The new protein polymers $\mathbf{C^P_4\text{-}D^{WW}}$ and $\mathbf{C^P_4\text{-}D^{PPxY}}$ were produced intact during standard fermentations, but $\mathbf{C^P_4\text{-}D^{PPxY}}$ was shown to be glycosylated. Using genetic engineering, we mutated the $\mathbf{C^P_4\text{-}D^{PPxY}}$ protein polymer sequence by the substitution Ser12→Ala. A standard fermentation resulted in an intact and non-glycosylated protein polymer $\mathbf{C^P_4\text{-}D^{PPxY*}}$. Interaction studies (ITC and steady state tryptophan fluorescence quenching), showed that both $\mathbf{C^P_4\text{-}D^{PPxY}}$ and $\mathbf{C^P_4\text{-}D^{PPxY*}}$ bind to $\mathbf{C^P_4\text{-}D^{WW}}$ with an equilibrium dissociation constant on the order of μM.

Finally, to demonstrate that heterodimer-forming blocks can be used to independently control protein polymer self-assembly at multiple length-scales, we selected the heterodimer-forming modules $\mathbf{D^A}$ and $\mathbf{D^B}$ to control the lateral interactions of fibrils self-assembled from the previously designed triblock

protein polymer $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$. In **Chapter 6** we construct the protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$. The $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ protein polymers assemble into long and stiff fibrils at neutral pH. The aim of the C-terminal attachment of the $\mathbf{D^A/D^B}$ blocks was to be able to control subsequent physical cross-linking and bundling of the fibrils. Both protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ were produced intact and with high yield during fermentation at optimal conditions as discussed in **Chapter 3**. Using Atomic Force Microscopy (AFM) we show that at neutral pH, fibrils consisting of 100% $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ or $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$ protein polymers bundle up and cross-link via homotypic $\mathbf{D^A/D^A}$ and $\mathbf{D^B/D^B}$ associations. Control over the degree of cross-linking and bundling can be obtained by using mixed fibrils consisting of $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2}$ with controlled amounts of the newly developed protein polymers $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^A}$ and $\mathbf{C_2\text{-}S^H_{48}\text{-}C_2\text{-}D^B}$**.** While the effect of the heterodimers on the structure of the fibril network as judged from AFM is very strong, oscillation rheology shows that the inclusion of the heterodimer forming blocks merely leads to a moderate increase in gel stiffness.

In order to place the research discussed in this thesis into the broader perspective, in **Chapter 7** we provide a General Discussion. We discuss several general strategies that can be used to control protein polymer self-assembly and discuss why and when there is a need for using heterodimer forming blocks. After providing an overview over results obtained in this thesis, we highlight the most urgent questions that need to be answered next. This is followed by a discussion on the benefits that heterodimer-driven self-assembly may bring to possible future applications of protein polymers as biomaterials. We also discuss the possible risks for human health end environment that might arise from the use of protein polymers technology. Finally we present some speculations about the future of the field of self-assembling protein polymers.

List of publications

This thesis:

Domeradzka, N. E., Werten, M. W.T., de Vries, R. and de Wolf, F. A. Production in *Pichia pastoris* of protein-based polymers with WW and PPxY domains. **2016** *Microb Cell Fact* 15(1):105

Domeradzka, N. E., Werten, M. W.T., de Vries, R. and de Wolf, F. A. Production in *Pichia pastoris* of protein-based polymers with small heterodimerforming blocks. **2016** *Biotechnol Bioeng* 113(5):953

Domeradzka, N. E., Werten, M. W.T., de Vries, R. and de Wolf, F. A. Protein cross-linking tools for the construction of nanomaterials. **2016** *Curr Opin Biotechnol*: 39

Domeradzka, N. E., Werten, M. W.T. de Vries, R. and de Wolf, F. A. Cross-linking and bundling of self-assembled protein-based polymer fibrils via heterodimeric coiled coils. *Submitted*

Other work:

Rombouts, W. H., **Domeradzka, N. E.**, Werten, M. W. T., Leermakers, F. A. M., de Vries, R., de Wolf, F.A., van der Gucht, J. Enhanced stiffness of silk-like fibers by loop formation in the corona leads to stronger gels. **2016** *Biopolymers* 105(11): 795

Fros, J.J., **Domeradzka, N.E.**, Baggen, J., Geertsema, C., Flipse, J., Vlak, J.M. and Pijlman, G.P. Chikungunya virus nsP3 blocks stress granule assembly by recruitment of G3BP into cytoplasmic foci. **2012** *J Virol* 86(19): 10873

Fros, J.J., Geertsema, C., Zouache, K., Baggen, J., **Domeradzka, N. E.**, van Leeuwen, D.M., Flipse, J.,Vlak, J.M., Failloux, A.B. and Pijlman, G.P. Mosquito Rasputin interacts with chikungunya virus nsP3 anddetermines the infection rate in *Aedes albopictus*. **2015** *Parasit Vectors* 17(8):464

Acknowledgements

By writing these last pages of my thesis I have to admit that my mind is like a bowl containing mixture of very colourful emotions; happiness, pride, suspense, excitement, nostalgia... When I started to flip the pages, I have realized that in every paragraph and in every result, there are many hidden things. I saw dozens of conversations with my supervisors, their trust, my ups and downs and support of my dear friends. Therefore, here I would like to write, from the bottom of my heart, the words of appreciation and gratefulness.

I would like to thank my supervisors Dr Frits de Wolf, Marc Werten and Dr Renko de Vries. I think that only you truly understand, where this adventure brought me. You warned me that the project would be challenging and indeed it was to me. Despite difficulties, I could always count on your support and guidance. **Frits** thank you for your enthusiasm and patience. I really appreciate your immediate comments and the fact that I could knock your door and you always found time to answer my questions and concerns. **Marc**, I have learnt from you a lot; and not only the laboratory part but also writing. Your keen eye for detail changed the way I work now and I am very grateful for that. **Renko** thank you for the amazing atmosphere you created around my project, for the time you spent with me on brainstorming and for all the words of encouragement. Although I started with a biotechnology background, you made sure that I did not feel stupid and uncomfortable around physicists and chemists. I am very grateful for the time you spent on commenting my thesis and your recommendation that opened me an opportunity to work at Penn State.I would like to thank Prof. **Frans** Leermakers, who agreed to become my promotor and was always interested about my progress. Thank you for the very insightful discussion about my propositions. Frans, your attitude of a cheerful person with huge passion to support students is really inspiring and I hope to meet on my way more of such positive scientists. I also would like to thank Prof. **Martien** Cohen Stuart, who was involved at the design phase of my project. Martien, thank you for your advices and words of encouragement.

I would like to thank all my colleagues from FBR and PCC. I thank **Jan Wichers**, **Marjo** and **Antoine** for creating a nice atmosphere in the office. **Emil** and **Jan Springer**, thank you for all the help I receive from you. **Gosia B**, **Juan** and **Hande**, it was great to share with you 1008, thank you for all the tips that you gave me during writing this thesis and for all the conversations about our “difficult PhD life”. **Marcel** and **Wolf**, you were my fun-fantastic office mates. I have to say that it was very difficult to say

"goodbye" to the office in Helix - you were such cool companion. Wolf thank you for fruitful collaboration in FBR and for all the help with rheology and AFM. **Marcel** and **Jan Maarten**, thank you guys for being so patient when I was terrorizing MCRs. I would like to thank **Gosia W**, **Hanne**, **Wolf** and **Maarten** for organizing together UK PhD trip. Guys, we were so efficient team and it was pleasure to work with you. **Sabine**, **Christian**, **Armando** and **Yunus**, guys, thanks a lot for organizing USA PhD trip. **Monika** and **Thao**, girls thank you for all your advices with cloning. **Inge,** thank you for teaching me how to fix AFM samples and your companion in New York. **Lennart, Johan** and **Helen**, thanks for hang outs during USA PhD trip. **Aljosha,** I adored our lunch conversations - thank you for that. **Junior**, I really enjoyed our beer together - we must repeat it. **Esio** thank you for the philosophic discussions about time and space, and your big help with the fluorescence study. **Amanda** and **Loiane**, thank you for our AMPs collaboration and I really hope that you will make those AMPs work. **Jeroen**, **Maria**, **Rui** and **Vittorio**, thanks for being a great companion at many parties, for fun conversations and your amazing attitude, and Maria thanks for teaching me ITC. **Huanhuan** and **Tingting,** thank you for all your kind words and all the tips that you gave me during writing. **Soumi** and **Surender**, thank you for all the talks and a lovely dinner together. Dear **Kris**, **Jacob**, **Lione**, **Merve**, **Ties**, **Ruben**, **Ilse**, **Joanne**, and **Marco**, we did not have a chance to spend more time together but I would like to thank you for keeping great work atmosphere.

Marleen, thank you for all your advices during protein meetings, **Mieke**, thank you for help with AFM, **Joris**, thanks for organizing these peculiar group meetings with beer, and **Joshua**, you help me a lot not only with rhoemeters, but also you always had a good advice for me, thank you for that.

Josie and **Mara**, thank you for being so positive, amazing and dependable. You both always helped me in any need. I am sure that all of us PhDs would be miserable without you. **Anita** and **Bert**, thank you for helping us with the organization of PhD trip and taking care for all the formalities related to the conferences that I participated. **Rene**, **Peter**, **Dirk**, **Ronald** and **Kamuran**, thank you for being the great colleagues during the practical work with students. **Anton**, **Remco**, **Hannie** and **Diane**, thank you that anytime I had a question, you always had an answer for me. **Hans**, thank you for our little conversations around the coffee machines and your interesting stories about Russia and PCC.

I would like to thank also to all my students for giving me a good lesson.

The big thanks to my friends from VLAG PhD council: **Jonathan**, **Joao**, **Geraldine**, **Anika**, **Andrada**, **Marlies**, **Gudrun**, **Lennart**, **Isabelle**, **Tjerk**,

Ita, **Anne** and **Canan** and our coordinator **Vesna** Prsic. It was pleasure and fun to work with you.

Moja szalona rodzinko! Tak sobie myślę teraz, że chyba zawsze będę dla Was małą Natalką, która uciekała przez bramę do Kany i Justy, gdy „cały dom" oglądał Dynastję. Jesteście najwspanialsi na świecie. Wiem, że zawsze mogę na Was liczyć i kocham Was bezgranicznie. **Mamo** i **Tato**, nie znam słów, żeby Wam podziękować za wszystkie poświecenia, cierpliwość i miłość. Kiedy widzę w Waszych oczach, że jesteście ze mnie dumni, jest to dla mnie największa nagroda. **Ania** i **Magdzia**, jako mała dziewczynka zawsze brałam z Was przykład i jestem bardzo dumna, że mam takie fantastyczne i nietuzinkowe starsze siostry. Dziewczyny, dziękuję za to, że zawsze mogę liczyć na dobrą radę i nie tylko. To samo dotyczy **Wojciecha** i **Michała**, którym równie serdecznie dziękuję. W swoich podziękowaniach, nie mogę oczywiście zapomnieć o najmłodszych i zacznę chronologicznie od **Wojtka "teenegara".** Wojtku, bardzo jestem z Ciebie dumna i cieszę się, że moje wybory inspirują także Ciebie. **Maciuniu**, **Atosiu** i **Piotrusiu**, nasze rezolutne domowe Dusiaki, jesteście naszymi promykami i dziękuję Wam za te wszystkie psotki, żarciki, przytulaski i za to, że jestem dla Was najkochaniszą ciocią Nianią. Na zakończenie, dziękuję też moim **Dziadkom** i **Babciom**, których nie ma już z nami, ale jestem pewna, że czuwają.

Maxim, thank you for everything that you have done for me, for your love and endless support. Thank you for all those lunches, which you brought to the lab, when I forgot that human body does not draw its energy just from determination. This is just one of many proofs how much you care about me. You always know how to bring me up, when I am down and losing my enthusiasm. You are an amazing partner and I want to thank you for the serenades on the guitar, dancing with me when everybody sits, the hitchhiker trips, wine at the middle of the night at the beach in Marseille, travel around the Bali island on the motorcycle, discussing my thesis, even though it is totally not your field and all the other adventures that only you could plan to cheer me up (or accidentally happened to us because of you, like Halloween that I will never forget).

Urbiku i **Olu**, bardzo Was kocham dziewczyny! Dziękuję za wszystkie przygody, te holenderskie i te niemieckie, za te skypy, ploteczki, rozwiązywanie i szukanie problemów, za wzajemne motywowanie się i dyskusje na tematy sercowe. Marta, Tobie dodatkowo muszę podziękować za to, że nigdy nie boisz się powiedzieć mi prosto w oczy, gdy zrobię coś głupiego i umiesz mną potrząsnąć. **Aniu** i **Mariuszu**, nasze drogi się troche rozeszły ostatnio, ale mam nadzieje, że się niebawem spotkamy. Dziękuję za Waszą serdeczność, szaloną zabawę na Waszym weselu, i za szalone lata we Fryczu.

Camila, my love! I am so happy that I met you. Thank you for your spirituality and for being my loyal friend.
Moja rodzinko „zagranico", **Marcin**, **Ilona**, **Michał, Kuba, Ala, Olek** i **Agata**! Nie mogłam sobie wymarzyć lepszej ekipy, kochani. Dziękuję za prezetny urodzinowe, imprezki, posiadówy i wszystkie nasze eskapady. **Kasia** i **Paweł**, nie znamy się długo, ale Wam też dziękuję za pogaduchy. **Meto**, **Francesco**, **Elena** and **Jan** thank you for many awesome moments together, dinners and movie nights.
And last but not least, my beloved Droef 95 and Co, you are deeply in my heart and I would not make it without you. My dear**: Ale, Moni, Alex, Cristian, Ayla, Gigi, Marie, Alexia, Alejandro, Andres, Kostia, Adria, Janis, Giorgos B and Giorgos K, Nikos, Lili, Berta, Maria, Giulio, Tereza, Harmen, Raul, Jacobo, Jordi, Janson, Marco, Emiko, Emmelie, Antonia, Carmen, Leo, Nicolo, Rosemarie and Baastian,** because of you I learnt so much about world and people and no one can take this away from me.

About the author

Natalia Eliza Domeradzka was born on the 15th of April 1987 in Nowe Miasto near the Pilica river, Poland. She completed her secondary education at XVII High School under the auspices of Andreas Fricius Modrevius in Warsaw. In 2006 she started to study at Warsaw University of Life Sciences at the Interfaculty Studies in Biotechnology. During this study she was voluntary working at the Department of Histology and Embryology of Warsaw Medical University on two projects. The first study was focused on analysis of correlation between mutations in keratin genes and disease cases of Meesmann corneal dystrophy. The second study involved the identification of single nucleotide polymorphisms in the genes for proteins involved in the metabolism of an immunosuppressant drugs. In 2009 she participated in the Erasmus exchange program that allowed her completing the internship at the Laboratory of Nematology at Wageningen University. The main focus of the internship was the production and characterization of the venom allergen-like recombinant proteins for immunomodulation of the human cell response. After obtaining her Engineer's Degree, Natalia continued the collaboration with the Laboratory of Nematology as a research assistant. In 2010 she started the Master program in medical biotechnology at Wageningen University. During Master's thesis accomplished at the Laboratory of Virology, she studied self-assembling behavior of non-structural proteins of Chikungunya virus. The results of the thesis were included in the first publication, where Natalia was listed as a co-author. The fluorescent microscopy picture of viral protein self-assembly taken by Natalia was placed at the cover of Journal of Virology (October 2012 vol. 86). In 2012 Natalia pursued a PhD degree at Physical Chemistry and Soft Matter at Wageningen University, where she studied the design, production and characterization of protein-based supramolecular materials. The results of the study are presented in this thesis. During her PhD fellowship Natalia was an active member of the VLAG PhD council. At the moment, she accepted a postdoctoral position in Prof. Demirel`s laboratory at Pennsylvania State University, where she will continue research in the field of material science.

Overview of completed training activities

Discipline specific

Advanced course of Microbial Physiology and Fermentation Technology	Delft	2014
Colloid Science	Wageningen	2013
Gordon Conference: Bioinspired materials*	Sunday River, USA	2014
Integration and combination of microscopy techniques	Amsterdam	2015
Knowledge transfer with Utrecht University	Wageningen	2013
Network meeting, Nanonext NL†	Utrecht	2012
Network meeting, Nanonext NL†	Utrecht	2013
Network meeting, Nanonext NL†	Utrecht	2013
Network meeting, Nanonext NL†	Utrecht	2014
NanoCity 2014*	Utrecht	2014
NanoCity 2015*	Amersfoort	2015

General

VLAG PhD week	Baarlo	2012
Project and time management (WGS)	Wageningen	2013
Communication with the Media and the General Public (Hertz)	Wageningen	2012
Interpersonal communication for PhD students (Hertz)	Wageningen	2013
Mobilising your Scientific network (Dicuore)	Wageningen	2012
Scientific Publishing (WGS)	Wageningen	2013
Dutch for Foreigners	Wageningen	2015

Intellectual property and valorisation awareness (NanonextNL)	Amersfoort	2013
Risk analysis and technology assessment (NanonextNL)	Amersfoort	2013
Techniques for writing and presenting scientific papers (WGS)	Wageningen	2013
Teaching and Supervising Thesis students (ESD&PS)	Wageningen	2014
Entrepreneurship in and outside Science (StartLife Center of Entrepreneurship)	Wageningen	2015
Writing Grant proposals (Wageningen in'to Languages)	Wageningen	2015

Optional

PhD study tour	USA	2013
PhD study tour; Chairwoman of organization committee	UK	2014-2015
Group meetings at Physical Chemistry and Soft Matter	Wageningen	2012-2016
Group meetings at Bioconversion	Wageningen	2012-2016
PhD council member	Wageningen	2012-2016

† *oral presentation*, * *poster presentation*

The research described in this thesis was financially supported by NanoNextNL, a micro and nanotechnology consortium of the government of The Netherlands and 130 partners.

Cover design by Agilecolor.com

Printed by Proefschriftmaken.nl