

The background of the cover is a deep teal blue. It features several large, semi-transparent blue spheres that resemble cells or droplets. Overlaid on these are numerous thin, tangled, greyish-brown lines that look like protein filaments or polymers. Some of these lines are thicker and more prominent, while others are thin and wispy, creating a complex, organic texture.

Production of protein-based polymers in *Pichia pastoris*

Production of protein-based polymers in *Pichia pastoris* Marc W. T. Werten 2017

Marc W. T. Werten

Production of protein-based polymers in *Pichia pastoris*

Marc W. T. Werten

Thesis committee

Promotors

Prof. Dr M. A. Cohen Stuart

Professor of Physical Chemistry and Colloid Science

Wageningen University & Research

Prof. Dr G. Eggink

Special professor Industrial Biotechnology

Wageningen University & Research

Co-promotor

Dr F. A. de Wolf

Senior Scientist, Biobased Products

Wageningen University & Research

Other members

Prof. Dr N. Callewaert, Ghent University, Belgium

Prof. Dr V. A. P. Martins dos Santos, Wageningen University & Research

Dr I. K. Voets, Eindhoven University of Technology

Prof. Dr H. A. B. Wösten, Utrecht University

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

Production of protein-based polymers in *Pichia pastoris*

Marc W. T. Werten

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 Friday September 15, 2017

at 1.30 p.m. in the Aula.

Marc W. T. Werten

Production of protein-based polymers in *Pichia pastoris*,
250 pages.

PhD thesis, Wageningen University, Wageningen, The Netherlands (2017).
With references, with summary in English.

ISBN: 978-94-6343-606-9

DOI: 10.18174/417761

What I cannot create, I do not understand

Richard P. Feynman

Table of Contents

Chapter 1	1
General introduction	
Chapter 2	11
High-yield secretion of recombinant gelatins by <i>Pichia pastoris</i>	
Chapter 3	29
Reduced proteolysis of secreted gelatin and Yps1-mediated α -factor leader processing in a <i>Pichia pastoris</i> <i>kex2</i> disruptant	
Chapter 4	49
Secreted production of a custom-designed, highly hydrophilic gelatin in <i>Pichia pastoris</i>	
Chapter 5	69
Biosynthesis of an amphiphilic silk-like polymer	
Chapter 6	87
Precision gels from collagen-inspired triblock copolymers	
Chapter 7	113
Production in <i>Pichia pastoris</i> of complementary protein-based polymers with heterodimer-forming WW and PPxY domains	
Chapter 8	135
Nanofibrillar hydrogel scaffolds from recombinant protein-based polymers with integrin- and proteoglycan-binding domains	
Chapter 9	161
General discussion	
Summary	229
Dankwoord	233
About the author	237
Publications	239

Chapter 1

General introduction

Submitted in modified form as part of a review article:
Werten MWT, Eggink G, Cohen Stuart MA, de Wolf FA.

1.1. Protein-based polymers

A long-sought objective in material science is the development of polymers with controlled monomer sequence [1, 2]. Although progress has been made in synthetic chemistry [3], the level of control evident in natural sequential polymers such as DNA and proteins is unparalleled. These biological macromolecules feature a defined molecular size, as well as a controlled sequence of the nucleotide or amino acid monomers. Proteins fold into a three-dimensional structure defined by their primary sequence, resulting in unique properties. From 20 different amino acid monomers, nature has created an awe-inspiring wealth of different proteins, including enzymes, antibodies, and peptide hormones, as well as nonbioactive proteins with a structure-forming, rheological, or colloidal function. The latter category includes proteins such as collagen and elastin that fulfill a major role in the structure of various tissues, and silks used in animal architecture such as silkworm cocoons and spider webs (for more information see Section 9.1 of the General discussion). These proteins typically feature highly repetitive sequences with biased amino acid composition, and can often reversibly self-assemble into supramolecular structures through the formation of noncovalent bonds. The natural materials derived from them display remarkable toughness, elasticity, and other properties that have inspired material scientists to mimic them using modern protein engineering.

These so-called protein-based polymers, or protein polymers for short, are produced as heterologous proteins in a suitable host, just like enzymes and other proteins, although their highly repetitive sequence, biased amino acid composition, and physicochemical properties do present additional difficulty. The genes encoding natural protein polymers are sometimes used, but more often genes are synthesized that encode simplified mimics, or even completely *de novo*-designed protein polymers [4-6]. Multifunctional block copolymers can be prepared by combining different polymer types into one molecule [1, 2, 7-9]. Ever since the pioneering work by Cappello and Ferrari of Protein Polymer Technologies [10, 11], *Escherichia coli* has by far been the most widely used production host for protein polymers. This workhorse of protein engineering is genetically very well accessible, and offers fast growth and efficient low-cost production [12]. Several other hosts have been used for the production of protein polymers, including plants, insect cells, transgenic animals, *Aspergillus nidulans*, *Saccharomyces cerevisiae*,

Hansenula polymorpha (*Pichia angusta*), and *Bacillus brevis* [13-16]. More than two decades ago, our laboratory and others [17, 18] started to explore the use of the methylotrophic yeast *Pichia pastoris* (*Komagataella phaffii*) [19] for the production of various protein polymers. By now, this yeast probably represents the second most-used microbial system for the production of protein polymers after *E. coli*.

1.2. *Pichia pastoris* as a protein production host

As a methylotrophic yeast, *P. pastoris* produces enzymes involved in methanol metabolism at very high levels when grown on methanol [20]. It is this observation that inspired the development of the yeast into a protein expression platform, where typically the strong methanol-inducible alcohol oxidase 1 (*AOX1*) promoter is used to drive transcription of genes encoding heterologous proteins [21]. Many more promoters have been described for use in *P. pastoris* [22, 23], including that of the constitutively expressed glyceraldehyde-3-phosphate dehydrogenase (*GAP*) gene [24]. Nonetheless, the *AOX1* promoter remains the most commonly used promoter. Thus far, all studies on the production of protein polymers in *P. pastoris* have relied on the *AOX1* promoter to drive expression (Section 9.1).

P. pastoris can grow to very high cell densities (>400 g/L wet weight) in cheap chemically defined media, and is highly suited for large-scale cultivation in bioreactors [25]. It can be easily genetically manipulated, similarly to *S. cerevisiae*, providing mitotically stable integration of vectors at desired loci in the genome without the need for selective pressure [26]. For an overview of the fundamentals of the *P. pastoris* expression system and available vectors, the reader is referred to several reviews [25-31]. The availability of the *P. pastoris* genome [32-36] and recent synthetic biology tools for this yeast including CRISPR-Cas9 [37, 38], allow efficient genetic strain improvement.

Although the highest currently published protein yield in *P. pastoris* of 22 g/L for *Hevea brasiliensis* hydroxynitrile lyase was achieved by intracellular expression [39], the highest reported extracellular yield of 18 g/L for *Trichoderma reesei* cellulase [40] does not lag far behind. In general, quite many proteins, including protein polymers, are secreted by *P. pastoris* at g/L levels [28]. Because *P. pastoris* secretes relatively low amounts of endogenous proteins, secretory production constitutes a highly efficient first purification step, and obviates the need for cell

disruption procedures that are cost-prohibitive at an industrial scale. Another advantage, particularly for the production of therapeutic proteins, is that its eukaryotic secretory pathway allows posttranslational modifications such as glycosylation, disulfide bond formation, folding, and proteolytic processing [28]. Overall, for many researchers the capability of *P. pastoris* for efficient secretory production will be the main reason for choosing this host. Accordingly, most protein polymers have been produced in secretory fashion (Section 9.1).

Over the years, *P. pastoris* has been established as an efficient industrial host. In 2006 the FDA conferred GRAS status [41], and a considerable number of therapeutic and other commercial proteins have been produced in the host. For example, Dyax Corp. (now Shire) have obtained FDA approval in 2009 for their kallikrein inhibitor KALBITOR (ecallantide). A recombinant hepatitis B vaccine by Shanta Biotech has been on the market in India since 1999. Mitsubishi Tanabe Pharma has marketed recombinant human serum albumin since 2000. This protein is secreted at more than 10 g/L of broth [42], and is produced at over 40 tons annually [43]. The first protein polymers produced in *P. pastoris* have also already entered the market (Section 9.1).

1.3. Genetic engineering of protein polymers

One of the distinguishing features of protein polymers is their repetitive sequence, and often modular architecture. Because this is directly reflected in the encoding genes, we will shortly introduce the cloning methodology involved.

DNA sequences that encode natural protein polymers such as spider silk or collagen are typically isolated by cDNA library screening or PCR (Chapter 2) [44]. However, in the case of spider silks, only partial cDNAs have been obtained because of the exceptionally large transcript size and high GC-content [45]. Furthermore, PCR of repetitive sequences can be very challenging [46] in view of the high probability for false priming by the primers or by incompletely extended fragments [47]. Similarly, sequencing of long repetitive genes by common primer walking methodology is not well feasible. This is problematic because minor sequence errors can have a detrimental impact on the folding and function of some protein polymers [48]. For these reasons, if natural homologs are not required for the intended application, synthetic biomimetic genes may be more practical than natural sequences. In this approach, protein polymer sequences are oligomerized

from a (preferably sequence-verified) DNA fragment (hereafter referred to as ‘monomer fragment’) that encodes several repeats of a specific amino acid sequence motif. These are often idealized consensus motifs derived from relatively abundant sequence patterns in natural polymer-like proteins. Besides biomimetic sequences, also completely artificial peptide sequences can be used, selected on the basis of combinatorial approaches or by rational design [49]. An important additional benefit of using simple analogs instead of complex natural sequences, is that insight is gained regarding the influence of amino acid sequence on three-dimensional structure and function, potentially offering future control. By combining different sequence types, multifunctional block copolymers can be created.

Although ligation-mediated and PCR-based random concatenation approaches have been widely used for the construction of oligomeric genes for various purposes, including protein polymers [11, 50-57], they offer no control over repeat number and block order when producing block copolymers (Figure 1.1A). Procedures that involve oligonucleotides or PCR furthermore pose a risk of introducing sequence errors, which, in long oligomeric genes, cannot readily be detected. Therefore, techniques have been developed for the concatenation of sequence-verified monomer fragments *via* sequential cloning steps. One such technique was developed by Kempe et al. [58], and it has been successfully used to construct protein polymers [59-61]. Probably the most commonly used sequential oligomerization technique today is an approach independently devised by our laboratory (Chapter 4) [62] and the Chilkoti group [63, 64]. We named the procedure ‘insertional doubling’, focusing on the ability to optionally double insert size with every step. The Chilkoti group aptly termed it ‘Recursive Directional Ligation’ (RDL). See Figure 1.1B for a schematic representation. For details of the method, which involves the use of restriction enzymes of the so-called ‘interrupted palindromes’ category [65], see Chapter 4. Later implementations by both groups use class IIs enzymes, which cut outside of their recognition sequence so that the oligomer can be completely free of cloning scars [66-68].

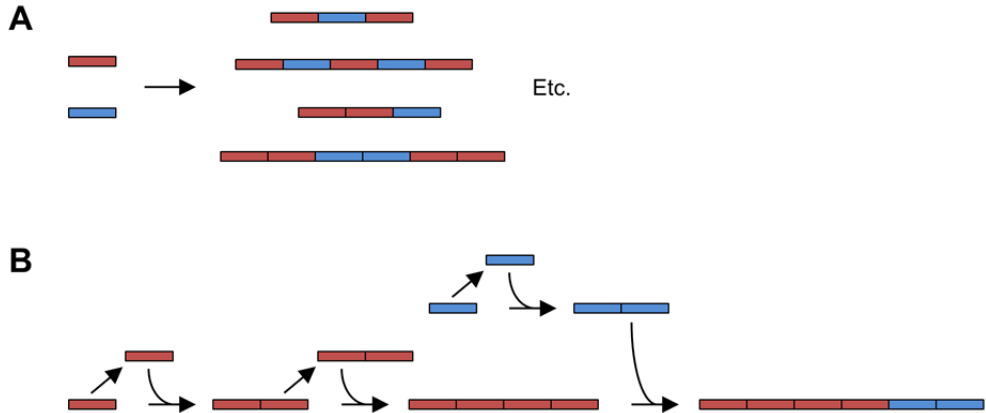


Figure 1.1. (A) Random concatenation results in a heterogeneous gene population. (B) Sequential concatenation allows the construction of block copolymers with defined repeat number and block order.

1.4. Aim and outline of this thesis

The aim of this thesis is to evaluate the suitability of the methylotrophic yeast *Pichia pastoris* for the production of a variety of protein polymers, and the preliminary physico-chemical characterization and functional testing of the products. Because one of the main promises of protein polymers as compared to traditional polymers is their highly defined nature, a major focus will be on protein quality. Protein polymers represent a particularly challenging target in this respect. Their often (partly) unfolded structure renders them highly susceptible to proteolytic attack, or other forms of posttranslational modification. On the other hand, the potential of some protein polymers for supramolecular self-assembly (i.e., the formation of complexes through noncovalent bonds) may cause problems *in vivo*, particularly in secretory production. Thus, owing to their structural features and low-complexity sequence, protein polymers can also provide insight into particular biological bottlenecks relevant to protein expression in general. The protein polymers studied in the experimental chapters of this thesis fall into three categories: (i) single-chain collagen-like proteins with natural or *de novo*-designed synthetic sequences, (ii) *de novo*-designed silk-like proteins, and (iii) block copolymers featuring collagen-inspired, silk-like, random coil, heterodimerizing, or cell-binding modules in various combinations.

In **Chapter 2** we describe the secretory production in *P. pastoris* of single-chain murine collagen fragments, or ‘gelatins’. We optimized fermentation conditions to minimize proteolytic degradation, and increased product levels by using multicopy strains.

In **Chapter 3** we report on the construction of *P. pastoris* strains deficient in Kex2 and yapsin 1 protease, aimed at preventing some of the endoproteolytic degradation of murine gelatins seen in Chapter 2. We evaluated the effect of these gene disruptions on product integrity, and investigated the involvement of both proteases in the processing of the α -factor mating pheromone secretory leader that was used to drive gelatin secretion.

In **Chapter 4** we describe the design and production of a completely artificial highly hydrophilic gelatin. We developed a method for the sequential construction of genes encoding protein (co)polymers, and characterized the protein polymer produced in terms of molecular weight, secondary structure, and surface activity.

In **Chapter 5** we outline the production of *de novo*-designed, amphiphilic and nonamphiphilic, silk-like proteins. Secondary structure, fibril-formation, and surface-active behavior of the proteins were studied.

In **Chapter 6** we present the production and physico-chemical characterization of gel-forming triblock copolymers. These polymers feature collagen-inspired triple-helix-forming end blocks, and a random-coil middle block identical to, or derived from, the hydrophilic gelatin described in Chapter 4. The thermoreversible trimer-forming functionality of the polymers was studied, as well as the rheological, sustained release, and erosion properties of the hydrogels formed.

In **Chapter 7** we explore the possibility of incorporating heterodimerizing so-called WW and PPxY modules into protein polymers, which would allow additional control over the self-assembly of protein polymers. The capacity of the proteins to form heterodimers was studied. O-glycosylation of the PPxY module, featuring extensively phosphorylated glycans, was abolished through site-directed mutagenesis.

In **Chapter 8** we consider the possibility of using protein polymers for cell culture. The proteins studied are functionalized fibril-forming triblock copolymers that consist of random coil end blocks, derived from the hydrophilic gelatin of Chapter 4, and a silk-like middle block, derived from the silk-like proteins of Chapter 5. Functionalization involved the genetic incorporation of two types of

cell-binding motifs. Proteolytic degradation could be prevented by appropriate sequence design. Tuning the density of functional motifs allowed control over cellular response.

The General discussion in **Chapter 9** first provides an overview of protein polymers that have been produced in *P. pastoris* by our group and others. We summarize their physicochemical properties and the details of their biosynthesis, and where applicable place each case in the context of previous work involving *E. coli*. We then proceed to draw conclusions concerning the main challenges faced, with particular emphasis on product quality and secretory production. Sequence design, culture conditions, and protease-deficient strains are discussed in relation to the atypical properties of protein polymers. Finally, we provide our view on the future prospects of the use of *P. pastoris* in the field of protein-based polymers.

References

1. Van Hest JCM, Tirrell DA (2001). *Chem Commun*, 1897-1904.
2. Kumar M, Sanford KJ, Cuevas WA, Du M, Collier KD, Chow N (2006). *Biomacromolecules* 7, 2543-2551.
3. Lutz JF, Ouchi M, Liu DR, Sawamoto M (2013). *Science* 341, 1238149.
4. Heslot H (1998). *Biochimie* 80, 19-31.
5. Grunwald I, Rischka K, Kast SM, Scheibel T, Barge H (2009). *Phil Trans R Soc A* 367, 1727-1747.
6. Kim W (2013). *Front Biosci* 18, 289-304.
7. Rabotyagova OS, Cebe P, Kaplan DL (2011). *Biomacromolecules* 12, 269-289.
8. DiMarco RL, Heilshorn SC (2012). *Adv Mater* 24, 3923-3940.
9. Desai MS, Lee SW (2015). *WIREs Nanomed Nanobiotechnol* 7, 69-97.
10. Cappello J (1990). *Trends Biotechnol* 8, 309-311.
11. Cappello J, Crissman J, Dorman M, Mikolajczak M, Textor G, Marquet M, Ferrari F (1990). *Biotechnol Prog* 6, 198-202.
12. Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MM, Ramadan HA, Saini KS, Baeshen NA, Redwan EM (2015). *J Microbiol Biotechnol* 25, 953-962.
13. Wong Po Foo C, Kaplan DL (2002). *Adv Drug Deliv Rev* 54, 1131-1143.
14. Baez J, Olsen D, Polarek JW (2005). *Appl Microbiol Biotechnol* 69, 245-252.
15. Girotti A, Fernandez-Colino A, Lopez IM, Rodriguez-Cabello JC, Arias FJ (2011). *Biotechnol J* 6, 1174-1186.
16. Heidebrecht A, Scheibel T (2013). *Adv Appl Microbiol* 82, 115-153.
17. Fahnestock SR, Bedzyk LA (1997). *Appl Microbiol Biotechnol* 47, 33-39.
18. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997). *EMBO J* 16, 6702-6712.
19. Kurtzman CP (2009). *J Ind Microbiol Biotechnol* 36, 1435-1438.
20. Couderc R, Baratti J (1980). *Agri Biol Chem* 44, 2279-2289.
21. Wegner GH (1990). *FEMS Microbiol Rev* 7, 279-283.
22. Gasser B, Steiger MG, Mattanovich D (2015). *Microb Cell Fact* 14, 196.
23. Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl AM, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, et al. (2016). *ACS Synth Biol* 5, 172-186.
24. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997). *Gene* 186, 37-44.
25. Cereghino GP, Cereghino JL, Ilgen C, Cregg JM (2002). *Curr Opin Biotechnol* 13, 329-332.
26. Cereghino JL, Cregg JM (2000). *FEMS Microbiol Rev* 24, 45-66.
27. Sreekrishna K, Brankamp RG, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA (1997). *Gene* 190, 55-62.
28. Cregg JM, Cereghino JL, Shi J, Higgins DR (2000). *Mol Biotechnol* 16, 23-52.
29. Daly R, Hearn MT (2005). *J Mol Recognit* 18, 119-138.
30. Lin-Cereghino J, Lin-Cereghino GP (2007). *Methods Mol Biol* 389, 11-25.
31. Ahmad M, Hirz M, Pichler H, Schwab H (2014). *Appl Microbiol Biotechnol* 98, 5301-5317.
32. De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, Rouze P, Van de Peer Y, Callewaert N (2009). *Nat Biotechnol* 27, 561-566.
33. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B (2009). *Microb Cell Fact* 8, 29.
34. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, et al. (2011). *J Biotechnol* 154, 312-320.
35. Sturmberger L, Chappell T, Geier M, Krainer F, Day KJ, Vide U, Trstenjak S, Schiefer A, Richardson T, Soriaga L, et al. (2016). *J Biotechnol* 235, 121-131.
36. Valli M, Tatto NE, Peymann A, Gruber C, Landes N, Ekker H, Thallinger GG, Mattanovich D, Gasser B, Graf AB (2016). *FEMS Yeast Res* 16.

37. Weninger A, Hatzl AM, Schmid C, Vogl T, Glieder A (2016). *J Biotechnol* 235, 139-149.
38. Kang Z, Huang H, Zhang Y, Du G, Chen J (2017). *World J Microbiol Biotechnol* 33, 19.
39. Hasslacher M, Schall M, Hayn M, Bona R, Rumbold K, Luckl J, Griengl H, Kohlwein SD, Schwab H (1997). *Protein Expr Purif* 11, 61-71.
40. Mellitzer A, Ruth C, Gustafsson C, Welch M, Birner-Grunberger R, Weis R, Purkarthofer T, Glieder A (2014). *J Biotechnol* 191, 187-195.
41. Vogl T, Hartner FS, Glieder A (2013). *Curr Opin Biotechnol* 24, 1094-1101.
42. Sumi A, Okuyama K, Kobayashi K, Ohtani W, Ohmura T, Yokoyama K (1999). *Bioseparation* 8, 195-200.
43. Julien C (2006). *Bioprocess Int* 4, 22-31.
44. Arcidiacono S, Mello C, Kaplan D, Cheley S, Bayley H (1998). *Appl Microbiol Biotechnol* 49, 31-38.
45. Hayashi CY, Lewis RV (1998). *J Mol Biol* 275, 773-784.
46. Scheibel T (2004). *Microb Cell Fact* 3, 14.
47. Hommelsheim CM, Frantzeskakis L, Huang M, Ulker B (2014). *Sci Rep* 4, 5052.
48. Persikov AV, Pillitteri RJ, Amin P, Schwarze U, Byers PH, Brodsky B (2004). *Hum Mutat* 24, 330-337.
49. Lakshmanan A, Zhang S, Hauser CA (2012). *Trends Biotechnol* 30, 155-165.
50. Doel MT, Eaton M, Cook EA, Lewis H, Patel T, Carey NH (1980). *Nucleic Acids Res* 8, 4575-4592.
51. Ferrari FA, Richardson C, Chambers J, Causey SC, Pollock TJ (1987). *Pat Appl WO/1988/003533*.
52. Rudert WA, Trucco M (1990). *Nucleic Acids Res* 18, 6460.
53. White MJ, Fristensky BW, Thompson WF (1991). *Anal Biochem* 199, 184-190.
54. Kato T, Suzuki S, Nishiok K, Yamamoto K (1994). *Anal Biochem* 220, 428-429.
55. Kurihara H, Morita T, Shinkai M, Nagamune T (2005). *Biotechnol Lett* 27, 665-670.
56. Sasagawa N, Ishiura S (2006). *Anal Biochem* 357, 308-310.
57. Chu HS, Ryum J, Park SY, Kim BG, Kim DM, Won JI (2011). *Biotechnol Lett* 33, 977-983.
58. Kempe T, Kent SB, Chow F, Peterson SM, Sundquist WI, L'Italien JJ, Harbrecht D, Plunkett D, DeLorbe WJ (1985). *Gene* 39, 239-245.
59. O'Brien JP, Hoess RH, Gardner KH, Lock RL, Wasserman ZR, Weber PC, Salemme FR (1994), in *Silk polymers: materials science and biotechnology*, (Kaplan D, Adams WW, Farmer B, Viney C, Eds.), American Chemical Society, Washington, DC, pp. 104-117.
60. Lewis RV, Hinman M, Kothakota S, Fournier MJ (1996). *Protein Expr Purif* 7, 400-406.
61. Fahnestock SR, Irwin SL (1997). *Appl Microbiol Biotechnol* 47, 23-32.
62. Van Heerde GV, van Rijn AC, Bouwstra JB, de Wolf FA, Mooibroek H, Werten MWT, Wind RD, van den Bosch TJ (1998). *US Pat* 6,150,081.
63. Meyer DE, Chilkoti A (1999). *Nat Biotechnol* 17, 1112-1115.
64. Meyer DE, Chilkoti A (2002). *Biomacromolecules* 3, 357-367.
65. Pitulle C, Hedenstierna KO, Fox GE (1996). *Biotechniques* 21, 619-620, 622.
66. Werten MWT, Moers APHA, Vong T, Zuilhof H, van Hest JCM, de Wolf FA (2008). *Biomacromolecules* 9, 1705-1711.
67. Teeuwen RLM, Zuilhof H, de Wolf FA, van Hest JCM (2009). *Soft Matter* 5, 2261-2268.
68. McDaniel JR, Mackay JA, Quiroz FG, Chilkoti A (2010). *Biomacromolecules* 11, 944-952.

Chapter 2

High-yield secretion of recombinant gelatins by *Pichia pastoris*

Recombinant nonhydroxylated gelatins based on mouse type I and rat type III collagen sequences were secreted from the methylotrophic yeast *Pichia pastoris*, using the *Saccharomyces cerevisiae* α -mating factor prepro signal. Proteolytic degradation could be minimized to a large extent by performing fermentations at pH 3.0 and by adding casamino acids to the medium, even though gelatin is extremely susceptible to proteolysis due to its open, unfolded structure. Proteolytic cleavage at specific mono-arginylic sites, by a putative Kex2-like protease, could be successfully abolished by site-directed mutagenesis of these sites. Production levels as high as 14.8 g/L clarified broth were obtained, using multicopy transformants. To our knowledge, this represents the highest level of heterologous protein secretion reported to date for *P. pastoris*.

2.1. Introduction

Gelatin is a well-known biopolymer and has a long history of use, mainly as a gelling agent in food. It is, in essence, denatured collagen and is prepared by hot acid or alkaline extraction of animal tissues such as bones and hides [1]. The gelling properties of gelatin are due to the remarkable amino acid sequence of its collagen ancestor. The so-called helical domain of this sequence consists of repeating Gly-Xaa-Yaa triplets, where Xaa and Yaa are often proline and hydroxyproline, respectively. *In vivo*, this polymeric structure enables the assembly of three procollagen molecules into a collagen triple helix [2]. This is a highly ordered process by virtue of specific mutual recognition of the collagen propeptides. In contrast, the denatured collagen triple helices in a heated gelatin preparation renature upon cooling to form randomly intertwining triple helices, resulting in the characteristic gel formation.

Apart from its traditional use, gelatin is also employed in sophisticated technical and medical applications, such as intravenous infusions [3], matrix implants [4] and injectable drug delivery microspheres [5]. The suitability of traditional gelatin for certain applications is, however, limited by its inherent characteristics. First, the above-mentioned extraction procedure results in chemically modified gelatin peptides covering a wide range of molecular weights [1]. Second, traditional gelatin has a high gelling temperature due to its high content of helix-stabilizing hydroxyprolines. This can be undesirable for low temperature applications [3].

In principle, gelatins of specific molecular weight can be prepared by biotechnological means. Production of nongelling (i.e., nonhydroxylated) gelatin-like proteins in *Escherichia coli* has been reported. It involves the expression of synthetic genes constructed from repeating (Gly-Xaa-Yaa)_n-encoding units [6-9]. Many problems arise concerning the instability of these highly repetitive genes [7]. Probably, native gelatin sequences derived from collagen genes are more stable than synthetic gelatin-like sequences, because there is greater variation in amino acid usage and less repetitiveness. Production in yeasts rather than bacteria may be preferable, since it has been reported that yeasts are well able to cope with repetitive genes [10-12]. Furthermore, the methylotrophic yeasts especially can produce high levels of recombinant protein [13]. It has recently been shown that proline hydroxylation can be achieved in the yeasts *P. pastoris* and *S. cerevisiae* by

coexpression of heterologous prolyl-4-hydroxylase [14, 15]. Although we currently pursue the production of nongelling gelatins, yeasts may thus also offer the possibility to produce hydroxylated, gelling gelatins. Whichever host organism is chosen, a challenge remains to overcome the extreme susceptibility of non-hydroxylated gelatin to proteolytic degradation, resulting from its random coil conformation. In this paper we report the high yield extracellular production of native nongelling gelatins by the methylotrophic yeast *P. pastoris* and we describe the minimization of their proteolytic degradation.

2.2. Materials and methods

2.2.1. Yeast strains and plasmids

P. pastoris strains GS115 (*his4*) [16] and SMD1168 (*his4*, *pep4*), and expression vector pPIC9 [17] were obtained from Invitrogen. The pPIC9 plasmid contains a *HIS4* selectable marker, an alcohol oxidase 1 (*AOX1*) promoter/terminator cassette and a *S. cerevisiae* α -factor prepro secretory signal. Vector pPIC9K was obtained from M.A. Romanos. It is identical to pPIC9 except that it also contains the *Tn903 kan^r* gene, which enables the selection of multicopy transformants [18].

2.2.2. Media composition

Minimal dextrose plates consisted of 0.4 mg/L biotin, 1.34% Yeast Nitrogen Base without amino acids (Difco), 1% dextrose and 1.5% agar. Minimal glycerol medium contained 0.4 mg/L biotin, 1.34% Yeast Nitrogen Base without amino acids and 1% glycerol. Buffered minimal glycerol medium consisted of minimal glycerol medium supplemented with 100 mM potassium phosphate, pH 6.0. Buffered minimal methanol medium was identical to buffered minimal glycerol medium but contained 0.5% methanol instead of glycerol. Fermentation basal salts medium contained, per liter: 26.7 mL phosphoric acid (85%), 0.93 g calcium sulfate dihydrate, 18.2 g potassium sulfate, 14.9 g magnesium sulfate heptahydrate, 4.13 g potassium hydroxide, 40.0 g glycerol and 4.3 mL trace elements. Trace elements contained, per liter: 4.5 g cupric chloride dihydrate, 0.09 g potassium iodide, 3.5 g manganese chloride tetrahydrate, 0.2 g sodium molybdate dihydrate, 0.02 g boric acid, 1.08 g cobalt sulfate heptahydrate, 42.3 g zinc sulfate heptahydrate, 65.0 g ferrous sulfate heptahydrate, 0.2 g biotin and 5.0 mL sulfuric acid.

2.2.3. Construction of pCOL3A1 and pCOL3A1K expression vectors

Plasmid pRGR5 [19], containing a partial rat pro α 1(III) collagen cDNA (Genbank Accession No. X70369), was digested with *Pst*I, yielding a 0.7 kb fragment encoding part of the helical domain. The resulting fragment was blunt-ended and inserted into the *Sna*BI site of pPIC9, yielding vector pCOL3A1. The vector pCOL3A1K was constructed by inserting a 1.0 kb *Bam*HI/*Eco*RI fragment isolated from pCOL3A1 into the *Bam*HI/*Eco*RI-digested vector pPIC9K.

2.2.4. Construction of pCOL1A1-1 and pCOL1A1-2 expression vectors

Two mouse type I collagen encoding DNA fragments of different length were amplified by PCR, such that both fragments shared the same 5' end. The primers were based on the mouse *COL1A1* sequence (Genbank Accession No. U08020) published by Li et al. [20], and were as follows:

C1A1-FW: 5'-CTTCCCAGATGTCCTATGGCTATGATG-3',

C1A1-RV1: 5'-CCGCTCGAGGCGCTCGCCAGGAGGTCCAGGCAG-3', and

C1A1-RV2: 5'-GCGCTCGAGGGGAGGACCAATGGGACCAGTCAG-3'. PCR was performed on Mouse 17-day Embryo QUICK-Clone cDNA (Clontech) using Advantage KlenTaq Polymerase Mix (Clontech). A 1.0 kb product (COL1A1-1) was obtained using primer combination C1A1-FW / C1A1-RV1 and a 1.8 kb product (COL1A1-2) was obtained using primer combination C1A1-FW / C1A1-RV2. To enable cloning of these fragments, the pPIC9 vector was modified by insertion of an adapter, which was prepared by annealing two complementary oligonucleotides:

5'-TCGAAAAGAGAGAGGCTGAAGCTCCCATGGGATAACTCGAGTAGG-3'

and 5'-AATTCCTACTCGAGTTATCCCATGGGAGCTTCAGCCTCTCTCTTT-3'.

This adapter was inserted into the *Xho*I/*Eco*RI sites of pPIC9, whereby the original *Xho*I site was disrupted. The COL1A1-1 and COL1A1-2 PCR products were digested with *Nco*I/*Xho*I, after which the resulting helical domain encoding DNA fragments were cloned into the *Nco*I/*Xho*I sites within the adapter. The vectors pCOL1A1-1 and pCOL1A1-2 thus obtained were verified by DNA sequencing using an ALF automated sequencer (Pharmacia).

2.2.5. Construction of pCOL1A1-1* and pCOL1A1-2* expression vectors

Using plasmid pCOL1A1-1 as a template and using *Pwo* DNA polymerase (Boehringer-Mannheim), PCR was performed using the commercial vector primers 5'AOX1 and 3'AOX1 (Invitrogen) in combination with the following mutagenic primers (mutagenic positions are underlined):

MUT1FW: 5'-GAGCCTGGCGGTTTCAGGTCCACGAGGTCCAATGGGTCCCCCTGG-3',

MUT1RV: 5'-CCAGGGGGGACCCATTGGACCTCGTGGACCTGAACCGCCAGGCTC-3',

MUT2FW: 5'-GGAGCTCCTGGCCAGCGAGGTCCAATGGGTCTGCCCCGGTGAGAG-3',

MUT2RV: 5'-CTCTACCGGGCAGACCCATTGGACCTCGTGGCCAGGAGCTCC-3'.

The primer combinations used were 5'AOX1/MUT1RV, MUT1FW/MUT2RV and MUT2FW/3'AOX1. The 0.5 kb, 0.3 kb and 0.7 kb PCR products obtained, respectively, were combined by overlap extension PCR [21] to form a doubly mutated 1.5 kb product. Digestion with *Bam*HI/*Apa*I resulted in a 1.0 kb DNA fragment which was subsequently inserted into *Bam*HI/*Apa*I-digested pCOL1A1-1 and pCOL1A1-2. The plasmids pCOL1A1-1* and pCOL1A1-2* thus obtained were verified by sequencing.

2.2.6. Transformation of *P. pastoris*

Plasmids used for transformation were linearized with *Sal*I in order to preferentially obtain Mut⁺ transformants (i.e., by integration at the *his4* locus and thus enabling normal growth on methanol) or with *Bgl*II in order to preferentially obtain Mut^s transformants (i.e., by transplacement of the *AOX1* locus and thus resulting in slow growth on methanol [22]). *P. pastoris* was transformed by electroporation [23] using a GenePulser (Bio-Rad) set at 1,500V, 25μF and 200 Ω and using 0.2 cm cuvettes. After growth on minimal dextrose plates at 30 °C for 3 days, several colonies were selected for PCR confirmation of the Mut genotype according to Linder et al. [24] with the exception that extension at 72 °C was for 2 min and cells were used directly for PCR without any pretreatment.

2.2.7. Selection of multicopy *P. pastoris* transformants

Multicopy Mut⁺ pCOL3A1K transformants were selected on the basis of G418 resistance, essentially as described [18]. Copy numbers were estimated by a quantitative dot blot as described by Clare et al. [22], with the following modifications. The blot was hybridized to a *COL3A1* probe. A *URA3* probe was

used for normalization of the amounts of DNA transferred. The blots were subjected to autoradiography and the signals were quantified by using a densitometric scanner (PDI).

2.2.8. Small-scale production of recombinant gelatins in *P. pastoris*

Transformants were grown overnight in buffered minimal glycerol medium, harvested and resuspended in buffered minimal methanol medium to an OD₆₀₀ of 1.0 (Mut⁺) or 10.0 (Mut^S). Cells were grown in baffled shaking flasks for 4 days at 30 °C and 250 rpm, with methanol being added to 0.5% every day.

2.2.9. Fermentative production of recombinant gelatins in *P. pastoris*

Fermentations were performed in 1, 2, 20, or 140 liter fermentors (Applikon). At the start of the fermentation, the fermentor contained half the working volume of fermentation basal salt medium, which optionally contained 1% casamino acids. Fermentation conditions for Mut⁺ transformants were as follows. The temperature was set at 30 °C, agitation at 500 rpm and aeration rate at 1 vvm. The pH was adjusted to pH 5.0 with 25% ammonium hydroxide. The fermentor was inoculated with 10% of the initial fermentation volume of a culture grown in minimal glycerol medium. A batch culture was grown until the glycerol was completely consumed. Aeration and agitation were increased to 2 vvm and 1,000 rpm, respectively. A glycerol fed-batch phase was initiated by feeding 50% glycerol containing 12 mL/L trace salts, at a rate of 18 mL/h per liter initial fermentation volume. The pH was either maintained at 5.0 by the addition of 25% ammonium hydroxide, or was allowed to drop to 3.0 by the yeast's metabolism. The glycerol feed was discontinued when the cell wet-weight reached approximately 180 g/L. After complete consumption of the glycerol, additional casamino acids were optionally added to 1% of the initial fermentor volume and the culture was induced by initiating a 100% methanol feed containing 12 mL/L trace salts. The feed rate was initially set at 3 mL/h per liter initial fermentor volume and was gradually increased to maximally 9 mL/h per liter initial fermentor volume. Fermentation conditions for Mut^S transformants were similar, but the methanol feed rate was initiated at 1 mL/h per liter initial fermentor volume, and was gradually increased to 3 mL/h per liter. Throughout the fermentation 2 mL culture samples were

collected and spun at 20,000 g for 5 min, after which the wet cell pellets were weighed. The supernatants were filtered using a disposable 0.22 μm filter.

2.2.10. Quantification of produced gelatins

In order to quantify the amounts of gelatin secreted into the medium, the gelatin was separated from extracellular endogenous proteins by acetone fractionation. Endogenous proteins were precipitated at 40% (v/v) acetone. The gelatin in the supernatant was precipitated by further increasing the acetone concentration to 80% (v/v), after which the resulting pellet was washed with 80% (v/v) acetone. The effectiveness of this separation was demonstrated by SDS-PAGE and by the observation that the determined amino acid composition closely matched that of pure gelatin (most markedly the approximate 33% Gly and 22% Pro). Gelatin concentrations were subsequently determined using the enhanced protocol of the BCA Protein Assay (Pierce) with analytical grade gelatin (Merck) as a reference.

2.2.11. SDS-PAGE and N-terminal protein sequencing

SDS-PAGE [25] was performed in a Mini-PROTEAN II system (Bio-Rad) under reducing denaturing conditions. For N-terminal protein sequencing, proteins were blotted onto Immobilon P[®] membrane (Millipore) by applying 100 V for 1 h in a Mini Trans-Blot Cell (Bio-Rad). Transfer buffer was 2.2 g CAPS per liter of 10% methanol, pH 11. Blots were stained with Coomassie Brilliant Blue and selected bands were cut out. N-terminal protein sequencing using Edman degradation was performed by Sequentiecentrum Utrecht, The Netherlands.

2.3. Results

2.3.1. Production of Col3a1 gelatin

A 0.7 kb rat *COL3A1* cDNA fragment, encoding a 21 kDa gelatin, was cloned into the *P. pastoris* expression vector pPIC9. The vector pCOL3A1 thus obtained was used to transform *P. pastoris* GS115. For each construct, several Mut⁺ transformants were randomly chosen and tested for gelatin production in shaking flasks. SDS-PAGE of the culture supernatants showed no significant differences in productivity between different transformants.

A representative COL3A1 transformant was selected for fermentation experiments. A growth and productivity curve of a typical COL3A1 fermentation at pH 3.0 is shown in Figure 2.1. Routine Col3a1 production levels ranged from 4-5 g/L clarified broth at a biomass wet-weight of 300-500 g/L.

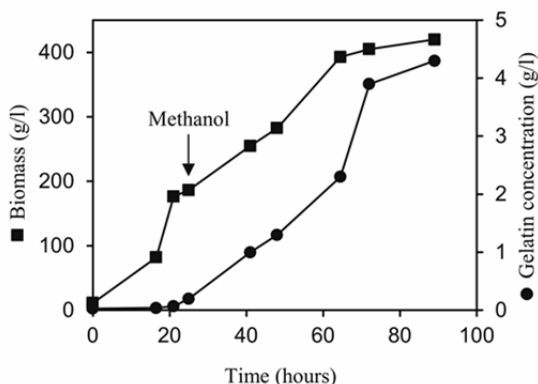


Figure 2.1. Growth and productivity curves of a 20 liter COL3A1 fermentation at pH 3.0. The culture was grown in glycerol batch mode for 0-16.5 h and in glycerol fed-batch mode for 16.5-21 h, after which protein expression was induced by initiating a methanol fed-batch. Biomass is expressed as cell wet-weight (g) per liter of broth. Product concentrations are expressed in gelatin (g) per liter of clarified broth and were determined using the BCA Protein Assay (Pierce) with analytical grade gelatin as a reference, after removal of endogenous secreted proteins by acetone fractionation (see Materials and Methods).

Culture supernatants harvested throughout the fermentation were subjected to SDS-PAGE. Collagenous proteins are known to migrate in SDS-PAGE at an apparent molecular weight approximately 1.4 times higher than the true molecular weight [26]. Accordingly, the apparent molecular weight observed for the 21 kDa Col3a1 gelatin was 29 kDa (Figure 2.2A). Apart from the full-length band, minor degradation products of about 21 kDa were visible. Initially, fermentations were performed at pH 5.0, but during the course of fermentation the full-length band was completely degraded into 21 kDa and smaller bands (Figure 2.2B). This pH effect suggests that the proteases involved were extracellular neutral proteases. It has been reported that supplementation of the fermentation medium with casamino acids can reduce extracellular proteolysis (Clare, 1991a). When applied to Col3a1 fermentations at pH 3.0, no such effect was observed (data not shown).

Circular dichroic spectroscopy performed according to De Wolf and Keller [27] showed that at 5 °C the Col3a1 product remained largely in the random coil conformation and is thus essentially nongelling. This is in accordance with the

absence of helix-stabilizing hydroxyprolines, as confirmed by determination of the amino acid composition (data not shown).

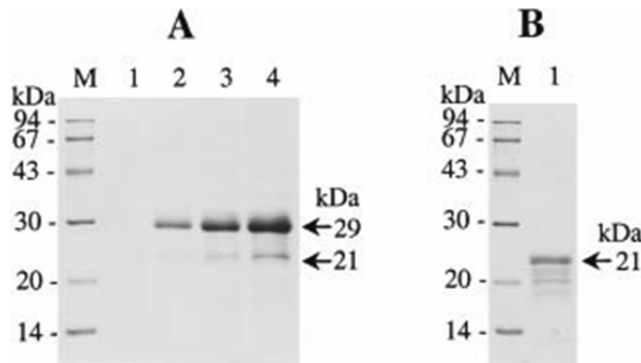


Figure 2.2. SDS-PAGE of COL3A1 fermentations at pH 3.0 and pH 5.0. (A) Lanes 1-4: 2 μ l culture supernatant of a fermentation at pH 3.0, collected at 16.5, 41, 64.5 and 89 h, respectively. Lane M: molecular weight marker (Pharmacia). (B) Lane 1: 2 μ l culture supernatant of a fermentation at pH 5.0, collected at 93 h. Lane M: molecular weight marker.

2.3.2. Production of Col1a1-1 and Col1a1-2 gelatins

A 1.0 kb mouse *COL1A1* cDNA fragment encoding a 28 kDa gelatin (Col1a1-1) and a 1.8 kb *COL1A1* cDNA fragment encoding a 53 kDa gelatin (Col1a1-2), were cloned in pPIC9. The vectors pCOL1A1-1 and pCOL1A1-2 thus obtained were used to transform *P. pastoris*, after which representative transformants were selected for fermentation.

Optimized COL1A1-1 fermentations involved growth at pH 3.0 and supplementation of the medium with casamino acids. Expression levels ranged from 2-3 g/L in the clarified broth. SDS-PAGE results are shown in Figure 2.3A. Apart from minor degradation products, several distinct major bands can be observed. The highest of these major bands has an apparent molecular weight of 38 kDa, which corresponds well to the apparent molecular weight expected for full-length Col1a1-1. Although the results obtained from fermentations at pH 5.0 were not significantly different from those at pH 3.0 (data not shown), the presence of casamino acids in the medium appeared to be important. The full-length band was barely detectable in fermentations that did not contain casamino acids (Figure 2.3B).

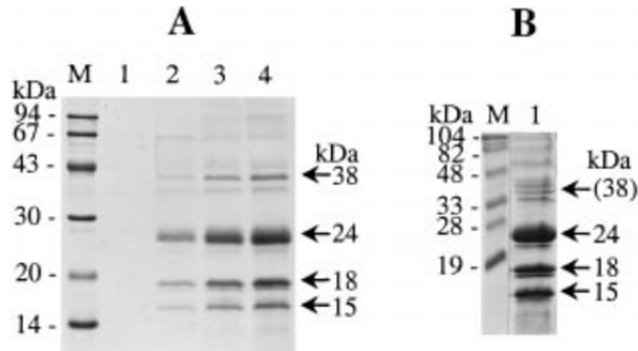


Figure 2.3. SDS-PAGE of COL1A1-1 fermentations at pH 3.0 with or without supplementation of the media with casamino acids. (A) Lanes 1-4: 3 μ l culture supernatant of a fermentation supplemented with casamino acids, collected at 21, 45, 70 and 94 h, respectively. Lane M: molecular weight marker. (B) Lane 1: 10 μ l of culture supernatant of a fermentation without supplementation with casamino acids, harvested at 94 h. Note that, while three times more supernatant was loaded than in lane 4 of (A), the full-length 38 kDa band can barely be discerned from background proteins. Lane M: prestained molecular weight marker (Bio-Rad).

A *P. pastoris* COL1A1-2 transformant was fermented under the same conditions used in the optimized COL1A1-1 fermentations. Expression levels ranged from 2-3 g/L in the clarified broth. SDS-PAGE showed a similar banding pattern to that obtained for COL1A1-1 (Figure 2.4). Interestingly, the lowest two major bands of the Col1a1-2 gelatin have the same apparent molecular weight as the lowest two bands of Col1a1-1. The highest Col1a1-2 band has the expected apparent molecular weight of 74 kDa.

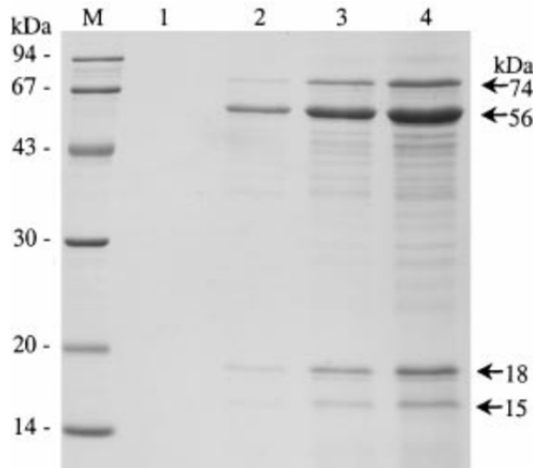


Figure 2.4. SDS-PAGE of a COL1A1-2 fermentation at pH 3.0 supplemented with casamino acids. Lanes 1-4: 3 μ l culture supernatant, collected at 24, 50, 96 and 164 h, respectively. Lane M: molecular weight marker.

2.3.3. Analysis of Col1a1 protein fragments and site-directed mutagenesis

In order to establish their identity, the Col1a1 protein fragments were subjected to N-terminal protein sequencing. Sequences corresponding to both the N-terminus (double underlines in Figure 2.5) and to internal positions were obtained (single underlines in Figure 2.5). The sequences corresponding to the N-termini of full-length Col1a1-1 and Col1a1-2 revealed a Glu-Ala extension. Such an extension of one or more Glu-Ala pairs, derived from the α -factor prepro secretory signal, is known to be present in some cases due to incomplete processing by the *STE13* gene product [28]. In the case of Col1a1, this may well be caused by the presence of a proline residue immediately C-terminal to the Glu-Ala-Glu-Ala spacer [29].

```

1  PMGPSGPRGL PGPPGAPGPQ GFQGGPPGEPG EPGGSGPMGP RGPPGPPGKN GDDGEAGKPG
61  RPPGERGPPGP QGARGLPGTA GLPGMKGHRG FSGLDGAKGD AGPAGPKGEF GSPGENGAPG
121 QMGPPRGLPGE RGRFGPPGTA GARGNDGAVG AAGPPGPTGP TGPPGFPGAV GAKGEAGPQG
181 ARGSEGPPQGV RGEPPGPPGPA GAAGPAGNPG ADGQPGAKGA NGAPGIAGAP GFPGARGPSG
241 PQGPSPGPPGP KNSGEPGAP GNKGDTGAKG EPGATGVQGP PGPAGEEGKR GARGEPGPSG
301 LPGPPGERGG PGSRGFPAGD GVAGPKGPSG ERGAPGPAGP KGSPGEAGRP GEAGLPAGAKG
361 LTGSPGSPGP DGKTGPPGPA GQDGRPGPAG PPGARGQAGV MGFPGPKGTA GEPGKAGERG
421 LPGPPGAVGP AGKDGEAGAQ GAPGPAGPAG ERGEQGPAGS PGFQGLPGPA GPPGEAGKPG
481 EQGVPGDLGA PGPSGARGER GFPGERGVQG PPGPAGPRGN NGAPGNDGAK GDTGAPGAPG
541 SQGAPGLQGM PGERGAAGLP GPKGDRGDAG PKGADGSPGK DGARGLTGPI GPPLLE

```

Figure 2.5. Theoretical Col1a1-2 protein sequence. Col1a1-1 shares residues 1-308 with Col1a1-2 and ends with Leu-Glu, as does Col1a1-2. The N-terminal sequences obtained for the Col1a1 protein fragments are underlined. Double-underlining at the N-terminus indicates that two fragments gave the same N-terminal sequence, both of which had a Glu-Ala extension.

Based on their N-terminal sequences and their molecular weights as determined by SDS-PAGE (corrected for the anomalous migration rate of collagenous proteins), the protein fragments could be assigned to distinct parts of the theoretical Col1a1 sequence (Table 2.1). Interestingly, the N-terminal sequences of the internal Col1a1 peptides were preceded in the theoretical sequence by the sequence Met-Gly-Pro-Arg (Figure 2.5, residues 38-41 and 122-125). This sequence occurs only at these two positions in Col1a1-1 and Col1a1-2, suggesting that it was possibly recognized by a specific endopeptidase. In order to verify this hypothesis, the cleavage sites were altered by site-directed mutagenesis. To preserve the native amino acid composition of Col1a1, the sequences were converted to Arg-Gly-Pro-Met, yielding plasmids pCOL1A1-1* and pCOL1A1-2*, respectively. *P. pastoris* strain GS115 was transformed with plasmids pCOL1A1-1* and pCOL1A1-2* and fermentations were performed as described for COL1A1-1 and COL1A1-2. SDS-

PAGE of the culture supernatant shows that, aside from minor degradation products, one major band of the expected full-length size is formed for COL1A1-1* as well as COL1A1-2* (Figure 2.6). The N-terminal sequences of the full-length bands were determined and were found to be identical to the expected N-terminus of intact Col1a1, although again extended with Glu-Ala.

Table 2.1. Apparent and theoretical molecular weights of Col1a1 fragments

Recombinant gelatin	Apparent molecular weight (kDa)	Corrected molecular weight (kDa)	Assigned residues	Theoretical molecular weight (kDa)
Col1a1-1	38	27	1-310	28
	24	17	126-310	16
	18	13	1-125	12
	15	11	42-125	8
Col1a1-2	74	53	1-595	53
	56	40	126-595	42
	18	13	1-125	12
	15	11	42-125	8

Apparent molecular weights are values derived by comparison with globular molecular weight marker proteins. Corrected molecular weights are the apparent molecular weights divided by 1.4 to account for the aberrant migration rate of collagenous proteins, as compared to globular proteins. Theoretical molecular weights were calculated from the indicated parts of the sequence. Note that both Col1a1 types share corresponding protein fragments, having apparent molecular weights of 15 and 18 kDa.

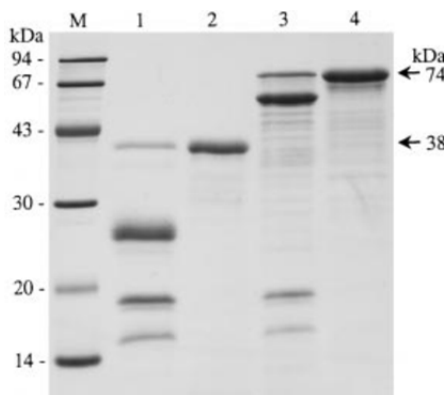


Figure 2.6. SDS-PAGE comparison of Col1a1-1, Col1a1-1*, Col1a1-2 and Col1a1-2*. All fermentations were performed at pH 3.0 and were supplemented with casamino acids. Culture supernatants were collected at the end of each fermentation and purified by acetone fractionation (see Materials and Methods). The apparent molecular weights of the full-length bands are indicated. Lanes 1-4: Col1a1-1, Col1a1-1*, Col1a1-2 and Col1a1-2*, respectively. Lane M: molecular weight marker.

2.3.4. Alternative recombinant gelatin production strategies

In order to investigate whether the Mut phenotype had an effect on productivity or proteolysis, Mut^s COL3A1 and COL1A1-1 transformants of GS115 were generated. Fermentation of these transformants at pH 3.0 did not show any reduction in proteolysis and did not improve yield (data not shown).

It has been reported that the use of the protease A-deficient *P. pastoris* strain SMD1168 can sometimes reduce proteolysis [30]. To verify its applicability for the production of recombinant gelatin, Mut⁺ COL3A1 and COL1A1-1 transformants of this strain were constructed and fermented at pH 3.0. SDS-PAGE of the culture supernatants revealed no decreased proteolysis or improved productivity (data not shown).

To further increase expression levels in *P. pastoris* GS115, the COL3A1 DNA fragment was cloned into the vector pPIC9K, which allows the selection of multicopy transformants [18]. A series of transformants containing approximately 1-15 copies of the vector was obtained, as determined by quantitative dot-blot analysis (data not shown). A shaking flask production expression experiment showed that productivity appeared to increase with increasing gene copy number. To investigate this further, transformants containing 5 or 15 copies were fermented. The 5-copy transformant produced 11.3 g gelatin per liter and the 15-copy transformant produced 14.8 g gelatin per liter of clarified broth. Thus, productivity was considerably improved as compared to the single-copy COL3A1 transformant.

2.4. Discussion

This paper describes the extracellular production of recombinant gelatins in *P. pastoris*. Different strategies were used to reduce the proteolytic degradation of this extremely vulnerable unfolded protein.

Initial fermentations of COL3A1 at pH 5.0 resulted in severe degradation of the full-length protein. This effect could be minimized by performing the fermentation at pH 3.0, suggesting proteolysis by extracellular proteases. In contrast, degradation of Col1a1-1 and Col1a1-2 could not be repressed by fermenting at pH 3.0. These gelatins were degraded to several distinct degradation products from the onset of fermentation. In order to obtain detectable amounts of the full-length Col1a1-1 band, it was necessary to supplement the fermentation medium with

casamino acids. This is in contrast with COL3A1 fermentations, where the use of casamino acids did not have any effect on degradation. Despite the beneficial effect of adding casamino acids to the medium, which has been reported to repress extracellular proteases [17], the distinct major Col1a1 degradation products were still formed. Their formation at the onset of fermentation suggests that some type of proteolytic cleavage may take place intracellularly. This notion is supported by the finding that incubation of the (clarified as well as nonclarified) Col1a1-1 fermentation broth at 30 °C for many days did not lead to degradation of the full-length band to the smaller species (data not shown).

N-terminal sequence analysis of the Col1a1 bands showed that the shorter species were internal fragments. The internal sequences were different from each other, but both were preceded in the theoretical Col1a1 sequence by Met-Gly-Pro-Arg. We hypothesized that Met-Gly-Pro-Arg might represent a motif that is specifically cleaved by an endopeptidase. Indeed, conversion in Col1a1-1 and Col1a1-2 of the Met-Gly-Pro-Arg sequences by site-directed mutagenesis to Arg-Gly-Pro-Met resulted in the disappearance of the major degradation products, in favor of the full-length proteins. Basic residues or pairs of basic residues are a well-known recognition site for proteolytic processing of prohormones in eukaryotes [31]. The *S. cerevisiae* subtilisin-like Kex2 protease was the first known enzyme with a cleavage specificity for paired basic residues [32]. It has been suggested that Kex2 is also able to cleave the mono-arginylic sequence Pro-Arg [33]. If Kex2 were to recognize Pro-Arg in the Met-Gly-Pro-Arg sequence, one might expect cleavage of all occurrences of Pro-Arg, in view of the absence of folded structure in nonhydroxylated gelatin. However, Col1a1 and Col3a1 do contain the sequences Ala-Gly-Pro-Arg, Ser-Gly-Pro-Arg, Gln-Gly-Pro-Arg, Pro-Gly-Pro-Arg and a dibasic Glu-Gly-Lys-Arg which are apparently not cleaved, or at least not to the extent of the Met-Gly-Pro-Arg sites. Furthermore, the conversion of residues 38-41 (Figure 2.5) to Arg-Gly-Pro-Met coincidentally generates the sequence Ser-Gly-Pro-Arg, which is evidently not substantially cleaved since the mutations were effective in enabling the expression of full-length Col1a1-1 and Col1a1-2. Thus, we conclude that neighboring residues are involved in recognition by the protease. As both the +1 and the -3 position (relative to the site of cleavage) are occupied by Gly in all of the above mentioned Pro-Arg sites in the recombinant gelatins, it is likely that the residue at the -4 position is the determining factor for cleavage. Brenner and Fuller

[34] mention that known Kex2-cleaved sites appear to contain aliphatic residues at the -4 position. Elaborating on this, we observe that the -4 position in known cleaved proteins is occupied by the more hydrophobic [Leu-Ile-Val-Met] subset of the aliphatic residues [35-38]. Although probably constituting the preferred Kex2 substrates, a basic or proline residue at the -2 position may not be strictly necessary [34, 38]. We therefore speculate that the mono-arginylic sequence Met-Gly-Pro-Arg is cleaved by Kex2 or a Kex2-like protease, which may recognize sequences conforming to the motif [Leu-Ile-Val-Met]-Xaa-Yaa-Arg. Accordingly, this motif occurs only at the two Met-Gly-Pro-Arg sites in the Col1a1 sequence, in contrast with a mere Pro-Arg motif. It will require the construction of a *P. pastoris* KEX2 gene disruptant to determine whether this protease is indeed involved in the cleavage at Met-Gly-Pro-Arg.

In an attempt to further enhance the yield and integrity of the recombinant gelatins, several strategies were employed. The use of Mut^S transformants or transformants of the protease A-deficient strain SMD1168 did not lead to enhanced productivity or reduced degradation. Productivity was, however, substantially improved by the use of multicopy transformants. A transformant bearing approximately 15 copies of the pCOL3A1 vector produced up to 14.8 g gelatin per liter of extracellular medium. To our knowledge, this represents an unprecedented high yield of secreted heterologous proteins in *P. pastoris*.

In view of the high productivity obtained, *P. pastoris* is apparently well able to cope with the extraordinary sequence of gelatin, which is highly repetitive and codes for some 33% Gly and 22% Pro. The codon usage for these amino acids in the rat COL3A1 sequence fits that of *P. pastoris* highly expressed genes [30] fairly well. The codon usage of the COL1A1-1 and COL1A1-2 sequences, however, differs considerably. In the COL1A1-1 sequence, glycine is encoded by the codon GGC in 28% of occurrences, while this is only 3% in the *P. pastoris* highly expressed genes. In the case of the proline encoding codon CCC this is 31% and 4%, respectively. Possibly, this codon usage may account for the lower expression levels of Col1a1-1 and Col1a1-2 as compared to Col3a1.

It has been suggested that the maximum level of protein secretion is ultimately determined by the protein folding capacity of the endoplasmic reticulum [39, 40]. Exceeding this capacity by the use of multicopy transformants is thought to result in the accumulation of unfolded proteins in the endoplasmic reticulum and a

concomitant vast decrease in the level of expression due to physiological instability. In this respect, the unique characteristics of gelatin may provide an explanation for the exceptionally high secretion levels obtained with our multicopy transformants. As nonhydroxylated gelatin is unfolded, it will probably not drain the folding capacity of the endoplasmic reticulum. Also, in contrast with incorrectly folded globular proteins, it is not likely that the unfolded gelatin will aggregate and accumulate in the endoplasmic reticulum, due to its outstanding solubility. This notion is supported by the observation that lysed cells did not show significant intracellular retention of recombinant gelatin (data not shown).

It can be concluded that *P. pastoris* is a very suitable host for the production of recombinant gelatins. The levels of heterologous protein secretion were exceptionally high. Proteolytic degradation could be minimized despite the unfolded structure of gelatin. Given the ease of upscaling fermentations of *P. pastoris*, these findings open up possibilities for economical large-scale production of recombinant gelatins.

Acknowledgments

We thank J.M.H. Stoop and E.C. de Bruin for critical reading of the manuscript; H.W. Wisselink, C. Kamp, M. Strubel, G. Boswinkel and R.C.A. Keller for technical assistance; E. Vuorio for providing plasmid pRGR5; and M.A. Romanos for providing plasmid pPIC9K.

References

1. Asghar A, Henrickson RL (1982). *Adv Food Res* 28, 231-372.
2. Vuorio E, De Crombrughe B (1990). *Annu Rev Biochem* 59, 837-872.
3. Saddler JM, Horsey PJ (1987). *Anaesthesia* 42, 998-1004.
4. Pollack SV (1990). *J Dermatol Surg Oncol* 16, 957-961.
5. Rao KP (1996). *J Biomater Sci Polymer Edn* 7, 623-645.
6. Goldberg I, Salerno AJ, Patterson T, Williams JI (1989). *Gene* 80, 305-314.
7. Cappello J (1990). *Trends Biotechnol* 8, 309-311.
8. Obrecht G, Lefèvre, J. F. and Meyrueis, P. (1991). *Fr Pat Appl* 91 16215.
9. Gardner K, Lock, R. L., O'Brien, J. P. and Salemme, F. R. (1992). *Pat Appl PCT/US92/09655*.
10. Strausberg R, Link R (1990). *Trends Biotechnol* 8, 53-57.
11. Cappello J, Ferrari F (1994), in *Plastics from microbes*, (Mobley DP, Ed.), Carl Hanser Verlag, Munich, pp. 35-92.
12. Fahnestock SR, Bedzyk LA (1997). *Appl Microbiol Biotechnol* 47, 33-39.
13. Hollenberg CP, Gellissen G (1997). *Curr Opin Biotechnol* 8, 554-560.
14. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997). *EMBO J* 16, 6702-6712.
15. Vaughan PR, Galanis M, Richards KM, Tebb TA, Ramshaw JAM, Werkmeister JA (1998). *DNA Cell Biol* 17, 511-518.
16. Cregg JM, Barringer KJ, Hessler AY, Madden KR (1985). *Mol Cell Biol* 5, 3376-3385.
17. Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA (1991). *Gene* 105, 205-212.
18. Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K (1994). *Nat Biotechnol* 12, 181-184.
19. Glumoff V, Mäkelä JK, Vuorio E (1994). *Biochim Biophys Acta* 1217, 31-40.
20. Li SW, Khillan J, Prockop DJ (1995). *Matrix Biol* 14, 593-595.
21. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989). *Gene* 77, 51-59.
22. Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA (1991). *Nat Biotechnol* 9, 455-460.
23. Becker DM, Guarente L (1991). *Methods Enzymol* 194, 182-187.
24. Linder S, Schliwa M, Kube-Granderath E (1996). *Biotechniques* 20, 980-982.
25. Laemmli UK (1970). *Nature* 227, 680-685.
26. Butkowski RJ, Noelken ME, Hudson BG (1982). *Methods Enzymol* 82, 410-423.
27. De Wolf FA, Keller RCA *Progr Colloid & Polym Sci* 102, 9-14.
28. Vedvick T, Buckholz RG, Engel M, Urcan M, Kinney J, Provow S, Siegel RS, Thill GP (1991). *J Ind Microbiol* 7, 197-201.
29. Van den Bergh CJ, Bekkers ACAPA, Geus P, Verheij HM, Haas GH (1987). *Eur J Biochem* 170, 241-246.
30. Sreekrishna K, Kropp KE (1996), in *Nonconventional Yeasts in Biotechnology*, (Wolf K, Ed.), Springer Verlag, pp. 203-253.
31. Van de Ven WJM, Roebroek AJM, Van Duijnhoven HLP (1993). *Crit Rev Oncog* 4, 115-136.
32. Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984). *Cell* 37, 1075-1089.
33. Mizuno K, Nakamura T, Ohshima T, Tanaka S, Matsuo H (1989). *Biochem Biophys Res Commun* 159, 305-311.
34. Brenner C, Fuller RS (1992). *Proc Natl Acad Sci USA* 89, 922-926.
35. Singh A, Chen EY, Lugovoy JM, Chang CN, Hitzeman RA, Seeburg PH (1983). *Nucleic Acids Res* 11, 4049-4063.
36. Zhu YS, Zhang XY, Cartwright CP, Tipper DJ (1992). *Mol Microbiol* 6, 511-520.
37. Ledgerwood EC, George PM, Peach RJ, Brennan SO (1995). *Biochem J* 308, 321-325.
38. Rourke IJ, Johnsen AH, Din N, Petersen, J. G. L., Rehfeld JF (1997). *J Biol Chem* 272, 9720-9727.
39. Parekh RN, Forrester K, Wittrup KD (1995). *Prot Expr Purif* 6, 537-545.
40. Parekh RN, Wittrup KD (1997). *Biotechnol Prog* 13, 117-122.

Chapter 3

Reduced Proteolysis of Secreted Gelatin and Yps1-Mediated α -factor Leader Processing in a *Pichia pastoris kex2* Disruptant

Heterologous proteins secreted by yeast and fungal expression hosts are occasionally degraded at basic amino acids. We cloned *Pichia pastoris* homologs of the *Saccharomyces cerevisiae* basic residue specific endoproteases Kex2 and Yps1 to evaluate their involvement in the degradation of a secreted mammalian gelatin. Disruption of the *P. pastoris KEX2* gene prevented proteolysis of the foreign protein at specific monoarginylic sites. The *S. cerevisiae* α -factor preproleader used to direct high-level gelatin secretion was correctly processed at its dibasic site in the absence of the prototypical proprotein convertase Kex2. Disruption of the *YPS1* gene had no effect on gelatin degradation or processing of the α -factor propeptide. When both the *KEX2* and *YPS1* genes were disrupted, correct precursor maturation no longer occurred. The different substrate specificities of both proteases and their mutual redundancy for propeptide processing, indicate that *P. pastoris kex2* and *yps1* single gene disruptants can be used for the α -factor leader-directed secretion of heterologous proteins otherwise degraded at basic residues.

3.1. Introduction

Targeting heterologous proteins for secretion can be an advantageous production strategy with eukaryotic microorganisms. Protein secretion allows for folding and posttranslational modification of the foreign protein, and can serve as a means to circumvent toxicity problems associated with the intracellular accumulation of some proteins. Furthermore, downstream processing usually is achieved more easily with secreted proteins than with intracellular proteins.

In *Saccharomyces cerevisiae*, the preproregion of the yeast's α -factor mating pheromone is commonly fused to heterologous proteins to direct their efficient secretion [1]. The α -factor propeptide is removed in a late Golgi compartment through cleavage C-terminal to Lys-Arg by the membranebound protease Kex2 [2]. This prototypical proprotein convertase (also known as kexin; EC 3.4.21.61) is a member of the subtilisin superfamily of serine proteases, and is specific for (mostly paired) basic amino acids.

The preprosequence of the *S. cerevisiae* α -factor also is used widely for the secretion of heterologous proteins by the methylotrophic yeast *Pichia pastoris*, and is the most successful leader sequence employed thus far in this expression system [3]. It generally is assumed that α -factor propeptide processing in *P. pastoris* also is effectuated by a Kex2 homolog. We have used the α -factor preprosequence in *P. pastoris* to obtain efficient secretion of mammalian gelatins with yields as high as ~15 g/L of clarified fermentation broth (Chapter 2). These gelatins are nonhydroxylated segments of collagen α -chains, which are essentially in the random coil conformation and, therefore, nongelling (Chapter 2). Traditional animal gelatin is a versatile biomaterial with both medical and technical applications. Gelatins produced in microbial systems provide the added benefit of being well-defined, free of animal-derived contagious agents, and suitable for modification for specific needs (Chapters 4 and 6) [4].

Although gelatin production in *P. pastoris* was very successful in terms of productivity, a 53 kDa gelatin derived from the murine type I collagen α -chain (referred to as Col1a1-2) was proteolytically cleaved C-terminal to two occurrences of the monobasic sequence Met-Gly-Pro-Arg (Chapter 2). This motif is similar to known nondibasic Kex2 substrates, which led to the hypothesis that the putative *P. pastoris* α -factor leader processing Kex2 homolog was responsible for the observed cleavage. Removal of the Met-Gly-Pro-Arg motifs from the gene construct by site-

directed mutagenesis resulted in the production of predominantly full-length product, and a number of shorter fragments present in lower quantities (Chapter 2). A custom-designed gelatin with no Arg residues was produced completely intact, so we hypothesized that the minor fragments may have originated from limited Kex2-like proteolysis at suboptimal monoarginylic motifs (Chapter 4). In addition to a Kex2 homolog, the aspartic yapsin proteases [5-7] also could be involved, since their substrate specificities overlap those of Kex2 in *S. cerevisiae* [8]. Yps1 (yapsin 1, previously termed Yap3; EC 3.4.23.41) is the best-characterized yapsin and is primarily active at the plasma membrane, but is also found in the extracellular medium and transiently active in the late secretory pathway [9-11]. It has frequently been implicated in the degradation of secreted heterologous proteins at basic amino acids in *S. cerevisiae*. See, for example, Kang et al. [11].

Our objective in this study was to isolate and disrupt the *P. pastoris* genes encoding Kex2 and Yps1, and to evaluate the utility of these gene disruptants for the production of intact Col1a1-2 gelatin. The development of *P. pastoris* strains deficient in proteolysis at basic residues should be valuable for the industrial production of medically or technically important proteins.

3.2. Materials and methods

3.2.1. *P. pastoris* strains and media

The *P. pastoris* GS115-based strain subjected to gene disruptions in this study was the previously described Col1a1-2 gelatin producing strain COL1A1-2 (Chapter 2). The expression vector used in this strain was derived from pPIC9 (Invitrogen, Breda, The Netherlands) and encodes a translational fusion of the *S. cerevisiae* α -factor prepropeptide and 53 kDa of the helical domain of murine type I collagen [12]. Strain COL1A1-2* produces a variant of Col1a1-2 referred to as Col1a1-2*, in which the proteolyzed Met-Gly-Pro-Arg motifs had been removed by site-directed mutagenesis of the encoding gene (Chapter 2).

Yeast extract-peptone-dextrose (YPD) medium (Duchefa, Haarlem, The Netherlands) contained 10 g/L yeast extract, 20 g/L peptone and 20 g/L dextrose. Minimal dextrose medium consisted of 13.4 g/L yeast nitrogen base without amino acids (Sigma, Zwijndrecht, The Netherlands), 10 g/L dextrose, and 0.4 mg/L biotin. Basal salts medium was as described by Zhang et al. [13].

Protein sequence analyses. Pairwise alignment of amino acid sequences and calculation of percent identity and similarity were performed with the EMBOSS Needle program [14]. The PAM250 similarity matrix was used, gap opening penalty was 10 and gap extension penalty was 0.5. The Pfam server [15] was used to search for functional protein domains. Secretory signal prediction was done with the SignalP neural network [16]. Sequences were analyzed for the presence of potential transmembrane domains with SOSUI [17]. Glycosylphosphatidylinositol modification sites were predicted with the fungal big-PI predictor [18].

3.2.2. Cloning of *P. pastoris* KEX2

An ~0.4 kb fragment was amplified from *P. pastoris* GS115 genomic DNA with degenerate primers (forward, 5'-CAYGGNACNMGNTGYGC-3'; reverse, 5'-GTRTANCCRTCSARTTRCA-3'). The PCR product was cloned into pGEM-T Easy (Promega, Leiden, The Netherlands) and sequenced to confirm its similarity to the *S. cerevisiae* KEX2 gene. Inverse PCR [19] was then performed on *Eco*RI-digested and self-ligated genomic DNA with the Expand Long Template PCR System (Roche, Almere, The Netherlands). Subfragments of the ~5.4 kb inverse PCR product were sequenced, and the entire genomic fragment was subsequently amplified with the Expand High Fidelity PCR System (Roche). The PCR product was cloned into pGEM-T Easy (Promega), and both strands were sequenced by primer walking (Eurogentec, Seraing, Belgium).

3.2.3. KEX2 gene disruption

Plasmid pMTL23-KEX2 was constructed by insertion of an ~3.9 kb *Bam*HI KEX2 subfragment into the *Bam*HI site of cloning vector pMTL23 [20]. The plasmid was cut with *Van*91I and *Pin*AI, which removed ~1.7 kb of the KEX2 fragment. An ~1.2 kb Zeocin resistance cassette of vector pPICZ α A (Invitrogen) was amplified with primers that introduce a *Van*91I site at the 5' end and a *Pin*AI site at the 3' end (forward, 5'-CTTCCACAATGTGGCCACACACCATAGCTTC-3'; reverse, 5'-CTGACCGGTAGCTTGCAAATTAAAGCCTTC-3'). The PCR product was cloned into the cut pMTL23-KEX2 plasmid. For gene replacement, the resulting disruption vector pKEX2 Δ was digested with *Bam*HI and the 3.4 kb fragment obtained was used to transform *P. pastoris*. Transformation was done by electroporation [21] in 0.2 cm cuvettes, with a Gene Pulser (Bio-Rad, Veenendaal, The Netherlands) set at

1,500 V, 25 μ F and 200 Ω . The electroporated cells were allowed to recover for 1 h in 1M sorbitol at 30 °C. One volume of YPD medium was added, and the cells were grown for 1 h at 30 °C before being plated on YPD medium supplemented with 1M sorbitol and 100 μ g/mL Zeocin (Invitrogen).

3.2.4. Preparation of membrane extracts for the Kex2 assay

Overnight cultures of *P. pastoris* were grown in 25 mL of minimal dextrose medium. Cells were harvested by centrifugation at 4 °C for 5 min at 1,500 \times g and washed three times with 25 mL of 100 mM Tris-HCl (pH 7.0) by repeated resuspension and centrifugation. The final cell pellet was resuspended in 1 mL of the same buffer. Acid washed 425- to 600- μ m glass beads (Sigma) were added to a final volume of ~1.8 mL and the cells were lysed at 4 °C in a Mixer Mill (Retsch, Haan, Germany) for 5 min at 30 Hz. The glassbeads were allowed to settle and the supernatant was spun twice for 5 min at 1,500 \times g and 4 °C to remove any remaining intact cells. The supernatant was centrifuged at 4 °C for 2 h at 20,800 \times g (rotor *k*-factor, 377), and then the pellet was resuspended in 1 mL of 100 mM Tris-HCl (pH 7.0), 0.5 M NaCl and centrifuged as before. This washing step was repeated with 1 mL of 100 mM Tris-HCl (pH 7.0). The resulting membrane pellet was resuspended in 50 μ L of 100 mM Tris-HCl (pH 7.0), 1% (v/v) Triton X-100. The sample was incubated at 4 °C for 30 min with gentle agitation and then centrifuged for 2 h at 20,800 \times g and 4 °C. The protein content of the supernatant was determined with the bicinchoninic acid protein assay (Pierce, Rockford, IL) with bovine serum albumin as a standard.

3.2.5. Kex2 assay

A modification of the procedure of Achstetter and Wolf [22] was used. An aliquot of 50 μ g of the detergent-solubilized membrane proteins (see above) was incubated in 200 μ L of 100 mM Tris-HCl (pH 7.0), 0.1% (v/v) Triton X-100, 0.5 mM benzyloxycarbonyl-Tyr-Lys-Arg-*p*-nitroanilide (Bachem, Weil am Rhein, Germany), and 5 mM calcium chloride. The release of *p*-nitroaniline at 30 °C was monitored for 16 h by measuring the A_{405} in a Safire microplate spectrophotometer (Tecan, Giessen, The Netherlands). The initial velocity was calculated from the slope of the enzyme progress curve over a period of 2.5 h (16 data points). Enzyme

activity was calculated using a molar absorption coefficient at 405 nm of 9,500 M⁻¹ cm⁻¹ for *p*-nitroaniline [23].

3.2.6. Cloning of *P. pastoris* YPS1

An ~0.4 kb fragment was amplified from *P. pastoris* GS115 genomic DNA with degenerate primers:

Forward, 5'-GCAGGCCACCGTTGACTGCTCGCARTAYGGNACNTT-3';

Reverse, 5'-GGTATACTTGCTTTGGTCCACTGCTCCRAANARDAT-3'. The PCR product was cloned with the TOPO TA Cloning Kit (Invitrogen) and sequenced to verify its similarity to the *S. cerevisiae* *YPS1* gene. The GeneRacer system (Invitrogen) was used according to the manufacturer's recommendations to isolate the unknown 5' and 3' termini of the transcript by RNA ligase-mediated rapid amplification of cDNA ends [24, 25]. Several cloned PCR products were sequenced, and the entire transcript was subsequently amplified by reverse transcription-PCR with the Expand High Fidelity PCR System (Roche). The ~1.9 kb fragment obtained was cloned with the TOPO TA Cloning Kit (Invitrogen), and both strands were sequenced by primer walking (BaseClear, Leiden, The Netherlands).

3.2.7. YPS1 gene disruption

The ~0.4 kb *YPS1* fragment isolated by degenerate PCR was reamplified with primers that introduce a *Bgl*II site at the 5' end and a *Bam*HI site at the 3' end:

Forward, 5'-GCGAGATCTGCAGGCCACCGTTGACTG-3';

Reverse, 5'-GCAGGATCCGGTATACTTGCTTTGGTCCAC-3'. An ~2.0 kb fragment containing a blasticidin resistance cassette and an *Escherichia coli* origin of replication was cut from vector pPIC6 A (Invitrogen) with *Bgl*II/*Bam*HI. The PCR product was digested with *Bgl*II/*Bam*HI and ligated into the vector fragment. Of the two possible orientations, a clone was chosen in which the fragments were oriented such that it had nonfunctional hybrid *Bgl*II/*Bam*HI sites. For insertional gene disruption the vector pYPS1Δ so obtained was linearized within the *YPS1* fragment with *Nde*I prior to transformation. Transformation of *P. pastoris* was as described for *KEX2* disruption, but selection was with 300 μg/mL blasticidin (Invitrogen).

3.2.8. Yapsin assay

The yapsin assay is modified from Azaryan et al. [26]. Membrane extracts were prepared as described for the Kex2 assay, but 100 mM sodium citrate buffer (pH 4.0) was used in all steps instead of 100 mM Tris-HCl (pH 7.0) due to the different pH optima of Kex2 and Yps1. An aliquot of 50 µg of the detergent-solubilized membrane proteins was incubated for 48 h at 30 °C in 200 µl of 100 mM sodium citrate buffer (pH 4.0), 0.1% (v/v) Triton X-100, 0.5 mM benzyloxycarbonyl-Tyr-Lys-Arg-*p*-nitroanilide (Bachem) and 5 mM calcium chloride. Subsequent treatment with aminopeptidase M was required to release *p*-nitroaniline after the initial Yps1 reaction. This treatment is consistent with the reported need for aminopeptidase M treatment with fluorogenic methylcoumarin amide peptides, which are cleaved between the paired basic residues by *S. cerevisiae* Yps1 [26], possibly due to the steric constraints imposed by peptide substrates. The reaction mixtures were heat-inactivated by boiling for 5 min, after which 30 µl of 1 M Tris-HCl (pH 9.0) was added to raise the pH to 7.5 to enable aminopeptidase M activity. A 200 µl portion of each sample was transferred to a microtiter plate, and the A_{405} was read in a microplate spectrophotometer (Tecan Safire). The samples were incubated for ~2 h at 37 °C after addition of 20 mU of aminopeptidase M (Roche) until the absorbance remained constant. The increase in absorbance after addition of aminopeptidase M was used as a measure of endopeptidase cleavage during the incubation of the membrane extracts with the chromogenic substrate. Enzyme activity was calculated using a molar absorption coefficient at 405 nm of 9,500 M⁻¹ cm⁻¹ for *p*-nitroaniline [23].

3.2.9. Fermentation of *P. pastoris*

Fed-batch fermentations were performed in 2.5 L Bioflo 3000 fermentors (New Brunswick Scientific, Nijmegen, The Netherlands), essentially as described by Zhang et al. [13]. Minimal basal salts medium was used, and no protease-inhibiting supplements were added. The pH during the glycerol batch phase was kept at 5.0. The pH was allowed to decrease metabolically to 3.0 during the glycerol fed-batch phase and was maintained at this pH during the 3- to 4-day methanol fed-batch phase. A home-made semiconductor gas sensor-controller, similar to that described by Katakura et al. [27], was used to monitor the methanol level in the off-gas and to maintain a constant level of ~0.2% (w/v) methanol in the broth.

3.2.10. Incubation experiments for extracellular precursor processing

Fresh cells of the *kex2* disruptant were obtained from fermentation broth harvested just prior to methanol induction, such that the pH was 3.0 but gelatin was not yet being produced. The cells were pelleted by centrifugation for 5 min at $1,500 \times g$ and 4 °C, and washed three times with deionized water by repeated resuspension and centrifugation.

Samples of the *kex2 yps1* disruptant were harvested at the end of the fermentation. Cell-free supernatant was prepared by centrifugation of the broth for 5 min at $1,500 \times g$ and 4 °C, followed by microfiltration of the supernatant using disposable 0.2 μm cellulose acetate syringe filters (Schleicher & Schuell, 's Hertogenbosch, The Netherlands).

A 1:1 ratio was used when *kex2 yps1* cell-free supernatant was mixed with *kex2* cell-free supernatant or cell-containing broth. When *kex2* cells were mixed with *kex2 yps1* supernatant, they were resuspended at ~450 mg/mL wet weight to mimic a typical high-cell-density fermentation. The aspartic protease inhibitor pepstatin was optionally added at 10 μM . The mixtures were incubated at 30 °C for 8 h and microfiltered to remove cells where applicable.

3.2.11. SDS-PAGE and N-terminal protein sequencing

The NuPAGE Novex system (Invitrogen) was used for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), with 10% Bis-Tris gels, MES SDS running buffer and SeeBlue Plus2 pre-stained molecular mass markers. Gelatin was purified by differential acetone precipitation (Chapter 2) prior to electrophoresis. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). Blotting of proteins for N-terminal sequencing by Edman degradation was as described previously (Chapter 2). Protein sequencing was performed by Midwest Analytical (St. Louis, MO).

3.2.12. Nucleotide sequence accession number

The nucleotide sequence data for *KEX2* and *YPS1* have been deposited in the GenBank Database under accession no. AY362700 and AY362701, respectively.

3.3. Results

3.3.1. Cloning of *P. pastoris* KEX2 and characterization of the deduced protein sequence

A putative *P. pastoris* KEX2 fragment encoding the catalytic domain was isolated by PCR with degenerate primers based on the amino acid sequences from various yeasts, and the entire gene was isolated by inverse PCR.

Among fungal Kex2 homologs, the deduced *P. pastoris* amino acid sequence is most similar to that of *Kluyveromyces lactis* (EMBL accession no. CAA30088; 45% identity and 66% similarity) and *S. cerevisiae* (GenPept accession no. AAA34718; 44% identity and 66% similarity). The first 21 residues of the *P. pastoris* sequence probably are a signal peptide for translocation into the endoplasmic reticulum. A conserved dibasic site (residues 114 to 115) defines a putative pro-region, which in *S. cerevisiae* is autocatalytically processed in the endoplasmic reticulum [28]. A region downstream of the supposed propeptide probably is a subtilisin domain (residues 152 to 445) and contains the highly conserved Asp, His and Ser subtilisin catalytic triad as well as a conserved Asn residue that stabilizes the oxyanion in the transitional state during hydrolysis of the peptide bond [29]. The region downstream of the catalytic domain probably is a proprotein convertase P-domain (residues 466 to 598), which in *S. cerevisiae* is required for the intramolecular maturation of pro-Kex2 [30]. A Ser/Thr rich domain, probably a target for O-glycosylation, is found C-terminal to this region, followed by a transmembrane domain. The C-terminal cytoplasmic tail contains a conserved Tyr residue (position 716) that, together with contextual sequences, constitutes a *trans*-Golgi network localization signal in *S. cerevisiae* [31].

3.3.2. Cloning of *P. pastoris* YPS1 and characterization of the deduced protein sequence

Degenerate primers were designed on the basis of a yapsin-like gene fragment from *Pichia angusta* (EMBL accession no. AL433341) and conserved amino acid stretches of *S. cerevisiae* Yps1, Yps2, and Yps3 (RefSeq accession no. NP_013221, NP_010428, and NP_013222, respectively). Reverse transcription-PCR with these primers followed by rapid amplification of cDNA ends resulted in the isolation of a putative *P. pastoris* yapsin.

Comparison of the deduced amino acid sequence with the *S. cerevisiae* yapsins showed that it was most similar to Yps1 (41% identity and 65% similarity). The first 23 residues probably are a secretory signal. A dibasic propeptide processing site, which is presumably cleaved autocatalytically in *S. cerevisiae* [10], is conserved in the *P. pastoris* sequence (residues 70 to 71). The location of the two active site Asp residues (residues 104 and 384) and their specific sequence context match those of *S. cerevisiae* Yps1 [32]. Two Cys residues are also conserved (residues 120 and 202). In *S. cerevisiae* Yps1, these residues flank an exposed loop region, which is proteolytically cleaved into α - and β -subunits that are linked by a disulfide bridge [10]. The inferred loop region in the *P. pastoris* sequence contains a stretch of Gly-Ser repeats that may confer conformational flexibility. The C-terminal part of the sequence contains a Ser/Thr-rich region (residues 545-569) that may be O-glycosylated, a hydrophobic tail (residues 579-599), and a putative glycosylphosphatidylinositol (GPI) modification signal [18]. Most GPI-anchored plasma membrane proteins, including *S. cerevisiae* Yps1, have a pair of basic amino acids just prior to the GPI-attachment site [33]. The putative GPI-attachment site of *P. pastoris* Yps1 (residue 574) is immediately preceded by a triplet of acidic residues. The significance of this atypical sequence context is unclear.

3.3.3. Disruption of the *KEX2* and *YPS1* genes

The *KEX2* and *YPS1* genes were disrupted individually and in combination in the present study, using a previously described *P. pastoris* strain that produces Colla1-2 gelatin (Chapter 2) and which is hereafter referred to as the wild type. The part of the *KEX2* locus encoding the catalytic and P domains was replaced with the Zeocin resistance cassette of plasmid pKEX2 Δ (Figure 3.1A). The *YPS1* gene was disrupted by insertion of the blasticidin selectable plasmid pYPS1 Δ in the region that encodes the catalytic β -subunit (Figure 3.1B), N-terminal to the subunit's catalytic Asp residue. Correct disruption of the *KEX2* and *YPS1* genes was verified by PCR (not shown) and Southern blotting (Figures 3.1C and D).

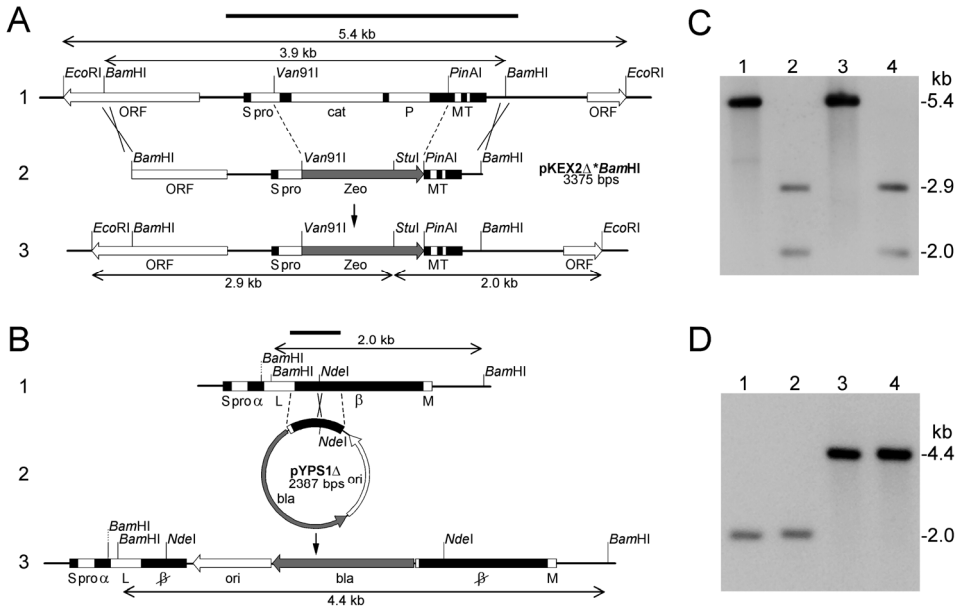


Figure 3.1. *KEX2* and *YPS1* gene disruption. (A) Schematic representation of the *P. pastoris KEX2* locus and its disruption. Abbreviations: ORF, unrelated flanking open reading frame; S, signal peptide; pro, propeptide; cat, catalytic domain; P, P domain; M, membrane spanning domain; T, *trans*-Golgi network localization signal; Zeo, Zeocin resistance cassette of pPICZα A (Invitrogen). The three parts of the scheme represent the *KEX2* locus (1), plasmid pKEX2Δ digested with *Bam*HI (2), and the disrupted *KEX2* locus (3). The black bar indicates the region used as a probe for Southern blotting. (B) Disruption of *P. pastoris YPS1*. Abbreviations: S, signal peptide; pro, propeptide; α, α-subunit; L, loop region; β, β-subunit (striethrough indicates disruption); M, membrane anchoring GPI domain; bla, blasticidin resistance cassette of pPIC6 A (Invitrogen), ori: *E. coli* origin of replication of pPIC6 A. The three parts of the scheme represent the *YPS1* locus (1), plasmid pYPS1Δ digested with *Nde*I (2), and the disrupted *YPS1* locus (3). The black bar represents the probe used for Southern blotting. (C) Verification of *KEX2* disruption by Southern blotting. Genomic DNA digested with *Eco*RI/*Stu*I was hybridized with the *KEX2* probe indicated in panel A. Lanes: 1, wild type; 2, *kex2*; 3, *yps1*; 4, *kex2 yps1*. The hybridization pattern was as expected (5.4 kb for the wild type and the *yps1* disruptant, and 2.9 and 2.0 kb for the *kex2* disrupted strains). (D) Verification of *YPS1* disruption by Southern blotting. Genomic DNA digested with *Bam*HI was hybridized with the *YPS1* probe indicated in panel B. The lanes correspond to the same strains as in panel C. The hybridization pattern was as expected (2.0 kb for the wild type and the *kex2* disruptant, and 4.4 kb for the *yps1* disruptants).

To demonstrate functional disruption of the *KEX2* and *YPS1* genes, the corresponding enzyme activities of all three disruptants and the wild type were assayed (Table 3.1). Kex2 activity was negligible in the *kex2* and *kex2 yps1* disruptants compared to the wild type and the *yps1* disruptant. A control reaction consisting of a wild-type sample and the chelator EGTA instead of the Ca^{2+} cofactor had minimal Kex2 activity, which attests to the specificity of the assay. Yapsin

activity was greatly reduced in the *yps1* and *kex2 yps1* disruptants, demonstrating functional disruption of the *YPS1* gene. A wild-type control reaction performed in the presence of the aspartic protease inhibitor pepstatin appropriately showed little yapsin activity. For unknown reasons, the yapsin activity in the *kex2* single disruptant was increased by 95% relative to the wild type.

Table 3.1. Kex2 and yapsin activities of wild-type and disruptant strains^a

Strain	Relative activity (%) of:	
	Kex2	Yapsin
Wild type	100 ± 2	100 ± 1
<i>kex2</i>	1 ± 1	200 ± 4
<i>yps1</i>	96 ± 0.4	5 ± 3
<i>kex2 yps1</i>	1.0 ± 0.5	11 ± 3
Wild type ^b	2.4 ± 0.4	6 ± 4

^a Values are means ±SD (*n* = 3) and are expressed as percentages of the mean wild-type activities (1.7 and 0.037 nmol/min per mg of solubilized membrane protein for the Kex2 and yapsin assay, respectively).

^b Control reactions in which the Kex2 assay contained the chelator EGTA instead of the Ca²⁺ cofactor, and in which the yapsin assay contained the aspartic protease inhibitor pepstatin. The accessory aminopeptidase M used in the yapsin assay was basically unaffected by pepstatin (not shown).

3.3.4. Phenotype of protease disruptants

A number of phenotypic defects have been described for *kex2* mutants of different yeast species [6, 34-36]. We did not observe obvious aberrant morphologies of cells grown in YPD, nor did we notice any growth arrest at 16 °C as reported for *S. cerevisiae* [6]. The generation time of *P. pastoris kex2* and *kex2 yps1* disruptants grown in YPD medium at 30 °C was increased by ~50% relative to the wild type (data not shown). The morphology and growth rate of the *yps1* disruptant in YPD were indistinguishable from the wild type. In practical terms, all strains had good viability and were amenable to high-cell-density fermentation (~350 to 550 g/L wet weight).

3.3.5. Gelatin production in protease disruptants

Use of the *kex2* disruptant for the production of Colla1-2 prevented the predominant proteolysis that occurs in the wild-type strain. While major fragments resulting from cleavage at Met-Gly-Pro-Arg are seen in the wild type (Figure 3.2, lane 1, bands 4 to 6), these fragments were absent in the *kex2* disruptant (Figure 3.2, lane 3). This shows that Kex2 is responsible for the cleavage observed in the wild type. The banding pattern of the remaining minor degradation products in the *kex2*

disruptant was very similar to that of the mutated Col1a1-2* gelatin produced in the wild type (Chapter 2), which lacks the proteolyzed Met-Gly-Pro-Arg sites (Figure 3.2, lane 2). This result indicates that Kex2 is not an important contributor to the formation of these minor fragments.

Col1a1-2 produced by the *yps1* disruptant (Figure 3.2, lane 4) had a banding pattern virtually identical to that of the wild type (Figure 3.2, lane 1). Except for two diffuse bands (bands 1 and 2) just over the full-length band (band 3), gelatin produced in the *kex2 yps1* double disruptant (Figure 3.2, lane 5) was similar to that produced in the *kex2*-disrupted strain (Figure 3.2, lane 3). Disruption of the *YPS1* gene apparently has no significant effect on the integrity of Col1a1-2.

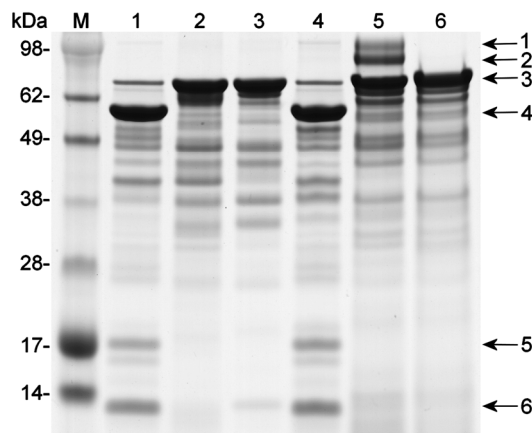


Figure 3.2. SDS-PAGE of gelatin produced in wild-type and disruptant strains. Lanes correspond to gelatin type and strain, respectively, as follows: lane 1, Col1a1-2/wild type (Chapter 2); lane 2, Col1a1-2*/wild type (Chapter 2); lane 3, Col1a1-2/*kex2*; lane 4, Col1a1-2/*yps1*; lane 5, Col1a1-2/*kex2 yps1*; lane 6, endoglycosidase H-treated Col1a1-2/*kex2 yps1*. Lane M, molecular mass marker. Arrows indicate protein bands referred to in the text. Gelatin was purified from the clarified broth by differential acetone precipitation (Chapter 2). The equivalent of 5 μ l of clarified broth was loaded for all samples, except for lanes 5 and 6, in which the equivalent of 15 μ l was loaded for approximate normalization. Note that gelatins migrate in SDS-PAGE at an apparent molecular mass ~40% higher than the true molecular mass (Chapter 2).

3.3.6. Characterization of α -factor leader processing

The major Col1a1-2 band produced by the *kex2* disruptant (Figure 3.2, lane 3, band 3) migrated at the same position as that of Col1a1-2* produced by the wild type (Figure 3.2, lane 2, band 3). As the α -factor propeptide of the latter is correctly processed (Chapter 2) and unprocessed precursors would have had a visibly higher molecular mass in SDS-PAGE, some form of α -factor leader processing must have occurred with the gelatin produced by the *kex2* disruptant. The N-

terminal sequence of the major Col1a1-2 band from the *kex2* disruptant (Figure 3.2, lane 3, band 3) was Glu-Ala-Glu-Ala-Pro. This corresponds to the sequence immediately C-terminal to the dibasic site in the α -factor propeptide/Col1a1-2 fusion, and indeed reveals accurate Kex2-like processing of the propeptide. The sequence also reveals a lack of processing of the α -factor derived (Glu-Ala)₂ spacer by the late Golgi protease dipeptidyl aminopeptidase A (DPAPase A) [37].

The N-terminal sequence of the full-length band obtained with the *yps1* disruptant (Figure 3.2, lane 4, band 3) was Glu-Ala-Pro-Met-Gly, which is identical to the sequences previously reported for the full-length bands of Col1a1-2 and Col1a1-2* produced in the wild type (Chapter 2). The sequence reveals correct cleavage at the dibasic site of the α -factor propeptide, and partial DPAPase A processing of the (Glu-Ala)₂ spacer (i.e., removal of one of the two Glu-Ala repeats).

The two diffuse bands appearing in the *kex2 yps1* double disruptant (Figure 3.2, lane 5, bands 1 and 2) disappeared following treatment of the protein with endoglycosidase H (Figure 3.2, lane 6). As Col1a1-2 produced in the wild type is not glycosylated (not shown), the effect of endoglycosidase H probably reflects the removal of oligosaccharides from small remnants of the α -factor propeptide, which is known to be glycosylated. N-terminal sequencing of the main and the diffuse (glycosylated) bands from the *kex2 yps1* disruptant (Figure 3.2, lane 5, bands 1 to 3) confirmed the presence of such remnants. The following sequences internal to the α -factor propeptide were found: band 1, Asp-Val-Ala-Val-Leu and Val-Ala-Val-Leu-Pro; band 2, Phe-Ile-Xaa-Thr-Thr; band 3, Glu-Glu-Gly-Val-Ser and Glu-Lys-Arg-Glu-Ala. Apparently, α -factor propeptide cleavage at the dibasic site does not occur in the *kex2 yps1* double disruptant, and the aberrant precursor maturation seen is probably caused by other proteases. This shows that Yps1 mediated the α -factor propeptide cleavage at the dibasic site observed in the *kex2* single disruptant.

3.3.7. Extracellular precursor processing

Yps1 activity in *S. cerevisiae* is detectable at the cell surface [38] and in the extracellular medium [9, 11], so extracellular Kex2-independent precursor processing in fermentations of *P. pastoris* was evaluated (Figure 3.3) by mixing cell-free supernatant of the *kex2 yps1* double disruptant (containing incompletely processed gelatin precursors) with either washed *kex2* cells or cell-free *kex2*

supernatant (harvested prior to induction of gelatin expression). The gelatin precursors produced by the *kex2 yps1* double disruptant appeared to be processed when incubated with washed *kex2* cells (Figure 3.3, lane 2) and when incubated with cell-free *kex2* supernatant (Figure 3.3, lane 3), but not in control incubations. The control incubations contained either *kex2 yps1* cell-containing broth alone (Figure 3.3, lane 1) or a mixture of cell-free *kex2 yps1* supernatant with cell-containing *kex2* fermentation broth in the presence of the aspartic protease inhibitor pepstatin (Figure 3.3, lane 4).

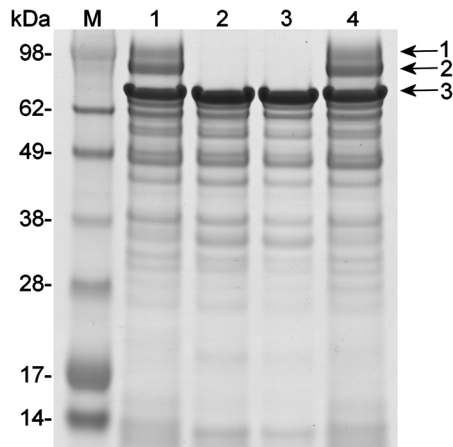


Figure 3.3. SDS-PAGE of extracellular precursor processing. Lanes correspond to the following incubations: lane 1, *kex2 yps1* cell-containing broth; lane 2, cell-free *kex2 yps1* supernatant mixed with *kex2* cells; lane 3, cell-free *kex2 yps1* supernatant mixed with *kex2* cell-free supernatant; lane 4, cell-free *kex2 yps1* supernatant mixed with *kex2* cell-containing broth in the presence of pepstatin. Lane M, molecular mass marker. Protein bands are indicated as follows: arrows 1 and 2, Col1a1-2 precursors; arrow 3: mature Col1a1-2. Gelatin was purified by differential acetone precipitation (Chapter 2). The equivalent of 15 μ l of *kex2 yps1* fermentation supernatant was loaded in each lane.

3.4. Discussion

We cloned *P. pastoris* homologs of the *S. cerevisiae* Kex2 and Yps1 basic residue specific endoproteases. Both deduced protein sequences were highly similar to the *S. cerevisiae* prototypes and contained all of the expected functional domains. Disruption of the putative *P. pastoris* KEX2 and YPS1 genes caused loss of most activity in corresponding enzyme assays, which confirms the identity of the genes. The near lack of yapsin activity in the *yps1* disruptant also suggests that Yps1 represents the major yapsin in *P. pastoris*, at least under the conditions used. The measured yapsin activity of the *kex2* disruptant was ~95% higher than that of the

wild-type strain. Yps1 is clearly the major contributor to this increased activity, given the low activity of the *kex2 yps1* disruptant. Accordingly, a preliminary assay showed that ~90% of the activity in the *kex2* disruptant was inhibited by the aspartic protease inhibitor pepstatin. We do not know whether the increased yapsin activity in the *kex2* disruptant is the result of a regulated, compensatory response to the lack of Kex2 activity.

Production of Col1a1-2 gelatin in the *kex2* disruptant prevented the prominent cleavage normally observed at the Met-Gly-Pro-Arg motifs, which confirms our prior hypothesis that Kex2 was responsible for this monobasic cleavage (Chapter 2). The banding pattern of low-level degradation products in the *kex2* strain was not significantly different from that in the wild type. We previously suggested that these minor fragments might have originated from limited cleavage by Kex2 at Arg residues within a suboptimal sequence context, because a tailored gelatin with no Arg residues at all was secreted completely intact (Chapter 4) and because even the relatively discriminative S₂ and S₄ subsites (Schechter & Berger nomenclature [39]) of *S. cerevisiae* Kex2 have some leeway in the substrate residues that they can accommodate [40, 41]. Apparently, however, few sites in Col1a1-2 other than Met-Gly-Pro-Arg meet the Kex2 protease's overall substrate requirements, as most of the minor bands were also present in the *kex2* disruptant.

The integrity of Col1a1-2 produced by the *yps1* disruptant was no better than that of gelatin produced by the wild type. The degradation seen in the *kex2 yps1* double disruptant was similar to that in the *kex2* disruptant. Thus, Yps1 does not seem to be an important contributor to the degradation of Col1a1-2. Yps1 was involved in the Kex2-independent processing of the α -factor propeptide, as the Col1a1-2 precursor was properly processed in a *kex2* disruptant but not in the *kex2 yps1* double disruptant. Col1a1-2 is secreted at 2-3 g/L in wild-type *P. pastoris* (Chapter 2), and the productivity of the *kex2* disruptant as judged by SDS-PAGE is similar (Figure 3.2). It was unexpected that a *P. pastoris kex2* disruptant can completely mature gelatin secreted at such high levels, since the *S. cerevisiae* KEX2 gene was first isolated by complementation of *kex2* mutants that were sterile due to their inability to produce mature α -factor [2]. Furthermore, maturation of α -factor/somatostatin precursors did not occur in a *kex2* mutant of *S. cerevisiae* [42]. Through the use of a sensitive halo assay, Egel-Mitani et al. [5] demonstrated production of mature α -factor by an *S. cerevisiae kex2* mutant when either Yps1 or

the prepropheromone itself was overexpressed. This result suggests that only a portion of the α -factor molecules were processed in this mutant, which is consistent with the partial processing reported for an α -factor/insulin hybrid in *S. cerevisiae* [5]. In comparison, the precursor processing observed during the fermentation of the *P. pastoris* *kex2* disruptant appears rather efficient.

While Kex2 is thought to recognize only the region N-terminal to the cleaved bond [42, 43], *S. cerevisiae* Yps1 is predicted to interact with at least three residues on the C-terminal side [44]. The particular protein fused to the α -factor propeptide may therefore influence precursor processing, and we cannot formally exclude the possibility that the Col1a1-2 precursor is more efficiently processed than other α -factor propeptide fusions. However, the Col1a1-2 precursor is not expected to be a uniquely well-suited substrate for Yps1, as it lacks the basic amino acids that Yps1 is proposed to favor among the first three [44] or six [45] residues C-terminal to the cleavage site.

S. cerevisiae Yps1 is localized to the extracellular side of the plasma membrane *via* its GPI-anchor. It may play a role in the proteolytic activation of secreted enzymes or the shedding of membrane receptors [9]. Before reaching the periplasm, Yps1 may be transiently active during its passage through the late secretory pathway [9, 10]. This hypothesis is supported by the secretion of biologically active α -factor by an *S. cerevisiae* *kex2* mutant [5] because the pheromone is active only when its N-terminal Glu-Ala repeats are removed by the *trans*-Golgi protease DPAPase A [37], following cleavage at Lys-Arg. In wild-type *P. pastoris*, only one of the two Glu-Ala repeats of Col1a1-2 was processed (Chapter 2), probably due to steric hindrance by the adjacent Pro residue. In the *kex2* disruptant, however, neither of the two Glu-Ala repeats was processed by DPAPase A. The high level of gelatin secreted may have exceeded the capacity of any transient Yps1, and the prerequisite dibasic cleavage of the α -factor propeptide might not have occurred until the gelatin was beyond the Golgi and its DPAPase A.

Yapsin activity is detectable at the cell surface of *S. cerevisiae* [38] and Yps1 is also present in the extracellular medium [9, 11]. The release of GPI-anchored Yps1 from the cell may be due to specific phospholipases [9]. In high-cell-density fermentations, at least some of the Yps1 released probably is due to cell lysis. To determine whether *P. pastoris* Yps1 can process the α -factor leader in the

fermentation broth, cell-free supernatant of the *kex2 yps1* disruptant was incubated with washed *kex2* cells or with cell-free *kex2* supernatant. In both cases, the precursor forms produced by the *kex2 yps1* double disruptant are processed, suggesting the presence of both cell surface-associated and free Yps1 activity. Be it cell associated or free, Yps1 activity is present in the fermentation broth at pH 3.0. Extracellular Yps1 probably is generally active in *P. pastoris* fermentations, as the pH range of 3.0 to 5.0 common for *P. pastoris* fermentations approximates the active pH range for *S. cerevisiae* Yps1 [26]. Yps1 activity in the broth may contribute to the efficient α -factor propeptide processing observed in the *kex2* disruptant, although the precursor molecules could already have been processed by Yps1 in the periplasmic space.

The *S. cerevisiae* Kex2 and Yps1 proteases have overlapping but distinct substrate specificities [8]. For example, the S₁ subsite of Kex2 is highly specific for Arg [46, 47], while Yps1 is also active toward substrates with Lys at the corresponding position [26, 45]. In the case of Colla1-2 production in *P. pastoris*, the Met-Gly-Pro-Arg motif was recognized only by Kex2, but the α -factor propeptide could be cleaved by both Kex2 and Yps1. The α -factor processing ability of both proteases, together with their divergent sequence specificities, suggests that *kex2* and *yps1* single-gene disruptants of *P. pastoris* should be evaluated for the expression of heterologous proteins that undergo degradation at basic sites. Whether cleavage of the propeptide by either protease actually occurs probably depends on the particular protein it is fused to, as bulky proteins may result in steric interference. Since the Glu-Ala repeats were not removed from Colla1-2 in the *kex2* disruptant, this often redundant spacer may need to be omitted from gene constructs with this strain if the native N-terminus of the foreign protein is required. In light of the many reports of proteolytic degradation at basic residues in yeasts, *kex2* and *yps1* disruptants of *P. pastoris* may be very useful for the production of secreted proteins in this powerful expression system.

Acknowledgments

We thank Jeroen Bastiaans, Patrick Janssen, Chantal Quik, and Valentina Tremaroli for technical assistance.

References

1. Brake AJ (1990). *Methods Enzymol* 185, 408-421.
2. Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984). *Cell* 37, 1075-1089.
3. Cereghino GP, Cereghino JL, Ilgen C, Cregg JM (2002). *Curr Opin Biotechnol* 13, 329-332.
4. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J (2003). *Adv Drug Deliv Rev* 55, 1547-1567.
5. Egel-Mitani M, Flygenring HP, Hansen MT (1990). *Yeast* 6, 127-137.
6. Komano H, Fuller RS (1995). *Proc Natl Acad Sci USA* 92, 10752-10756.
7. Olsen V, Cawley NX, Brandt J, Egel-Mitani M, Loh YP (1999). *Biochem J* 339, 407-411.
8. Bourbonnais Y, Germain D, Ash J, Thomas DY (1994). *Biochimie* 76, 226-233.
9. Ash J, Dominguez M, Bergeron JJ, Thomas DY, Bourbonnais Y (1995). *J Biol Chem* 270, 20847-20854.
10. Cawley NX, Olsen V, Zhang CF, Chen HC, Tan M, Loh YP (1998). *J Biol Chem* 273, 584-591.
11. Kang HA, Kim SJ, Choi ES, Rhee SK, Chung BH (1998). *Appl Microbiol Biotechnol* 50, 187-192.
12. Li SW, Khillan J, Prockop DJ (1995). *Matrix Biology* 14, 593-595.
13. Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM (2000). *Biotechnol Bioeng* 70, 1-8.
14. Rice P, Longden I, Bleasby A (2000). *Trends Genet* 16, 276-277.
15. Bateman A, Birney E, Cerruti L, Durbin R, Etwiler L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002). *Nucleic Acids Res* 30, 276-280.
16. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997). *Protein Eng* 10, 1-6.
17. Hirokawa T, Boon-Chiang S, Mitaku S (1998). *Bioinformatics* 14, 378-379.
18. Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F (2004). *J Mol Biol* 337, 243-253.
19. Triglia T, Peterson MG, Kemp DJ (1988). *Nucleic Acids Res* 16, 8186.
20. Chambers SP, Prior SE, Barstow DA, Minton NP (1988). *Gene* 68, 139-149.
21. Becker DM, Guarente L (1991). *Methods Enzymol* 194, 182-187.
22. Achstetter T, Wolf DH (1985). *EMBO J* 4, 173-177.
23. Erlanger BF, Kokowsky N, Cohen W (1961). *Arch Biochem Biophys* 95, 271-278.
24. Maruyama K, Sugano S (1994). *Gene* 138, 171-174.
25. Volloch V, Schweitzer B, Rits S (1994). *Nucleic Acids Res* 22, 2507-2511.
26. Azaryan AV, Wong M, Friedman TC, Cawley NX, Estivariz FE, Chen HC, Loh YP (1993). *J Biol Chem* 268, 11968-11975.
27. Katakura Y, Zhang W, Zhuang G, Omasa T, Kishimoto M, Goto Y, Suga K (1998). *J Ferm Bioeng* 86, 482-487.
28. Germain D, Dumas F, Vernet T, Bourbonnais Y, Thomas DY, Boileau G (1992). *FEBS Lett* 299, 283-286.
29. Brenner C, Bevan A, Fuller RS (1993). *Curr Biol* 3, 498-506.
30. Gluschkof P, Fuller RS (1994). *EMBO J* 13, 2280-2288.
31. Wilcox CA, Redding K, Wright R, Fuller RS (1992). *Mol Biol Cell* 3, 1353-1371.
32. Azaryan AV, Friedman TC, Cawley NX, Loh YP (1995). *Adv Exp Med Biol* 362, 569-572.
33. Hamada K, Terashima H, Arisawa M, Yabuki N, Kitada K (1999). *J Bacteriol* 181, 3886-3889.
34. Davey J, Davis K, Imai Y, Yamamoto M, Matthews G (1994). *EMBO J* 13, 5910-5921.
35. Enderlin CS, Ogrzydzak DM (1994). *Yeast* 10, 67-79.
36. Bader O, Schaller M, Klein S, Kukula J, Haack K, Muhlschlegel F, Korting HC, Schafer W, Hube B (2001). *Mol Microbiol* 41, 1431-1444.
37. Julius D, Blair L, Brake A, Sprague G, Thorner J (1983). *Cell* 32, 839-852.

38. Komano H, Seeger M, Gandy S, Wang GT, Krafft GA, Fuller RS (1998). *J Biol Chem* 273, 31648-31651.
39. Schechter I, Berger A (1967). *Biochem Biophys Res Commun* 27, 157-162.
40. Bevan A, Brenner C, Fuller RS (1998). *Proc Natl Acad Sci USA* 95, 10384-10389.
41. Suzuki Y, Ikeda N, Kataoka E, Ohsuye K (2000). *Biotechnol Appl Biochem* 32, 53-60.
42. Bourbonnais Y, Bolin D, Shields D (1988). *J Biol Chem* 263, 15342-15347.
43. Rockwell NC, Krysan DJ, Komiyama T, Fuller RS (2002). *Chem Rev* 102, 4525-4548.
44. Olsen V, Guruprasad K, Cawley NX, Chen HC, Blundell TL, Loh YP (1998). *Biochemistry* 37, 2768-2777.
45. Cawley NX, Chen HC, Beinfeld MC, Loh YP (1996). *J Biol Chem* 271, 4168-4176.
46. Rockwell NC, Fuller RS (1998). *Biochemistry* 37, 3386-3391.
47. Rockwell NC, Thorner JW (2004). *Trends Biochem Sci* 29, 80-87.

Chapter 4

Secreted production of a custom-designed, highly hydrophilic gelatin in *Pichia pastoris*

A custom-designed, highly hydrophilic gelatin was produced in *Pichia pastoris*. Secreted production levels in single-copy transformants were in the range of 3-6 g/L of clarified broth and purification to near homogeneity could be accomplished by differential ammonium sulfate precipitation. Despite the fact that gelatins are highly susceptible to proteolysis because of their unfolded structure, the recombinant protein was shown to be fully intact by SDS-PAGE, N-terminal sequencing, gel filtration chromatography and mass spectrometry. Owing to its highly hydrophilic nature, the migration of the synthetic gelatin in SDS-PAGE was severely delayed. Esterification of the carboxylic amino acid side chains resulted in normal migration. The high polarity of the synthetic gelatin also accounts for its negligible surface activity in water at concentrations up to 5% (w/v), as determined by tensiometry. Circular dichroism spectrometry showed that the non-hydroxylated gelatin did not form triple helices at 4 °C. The spectrum was even more representative of the random coil conformation than the spectrum of natural nonhydroxylated gelatins.

4.1. Introduction

Gelatin, in essence denatured and partially degraded collagen, is traditionally prepared by hot acid or alkaline extraction of animal bones and hides. Apart from its main use as a gelling agent in food [1], gelatin is also used in medical and industrial applications such as intravenous infusions [2], matrix implants [3], injectable drug delivery microspheres [4] and photographic film [5]. Despite the diversity of current uses of natural gelatin, recombinant gelatins may provide benefits for specific applications, in that the chemical composition and molecular weight can be precisely controlled and reproduced [6, 7]. Furthermore, recombinant gelatins do not bear the risk of associated infectious diseases such as bovine spongiform encephalopathy (BSE).

Several reports have described the production of recombinant gelatin-like proteins in *Escherichia coli*. Analogously to the natural amino acid sequence of the collagen triple helix-forming domain, synthetic genes are constructed from repeating (Gly-Xaa-Yaa)_n-encoding oligonucleotides, where Xaa and Yaa are often proline [8-11]. Gene instability problems are commonly observed with such highly repetitive genes [11]. Also, expression levels usually obtained in *E. coli* are rather low and purification of the intracellularly produced protein can be difficult. Recently, Kajino et al. [12] reported the use of *Bacillus brevis* for the expression of gelatin-like proteins. They used sequence stretches selected from natural collagen genes and polymerized them to form semi-synthetic gelatin. Prior to their report, we reported the use of the methylotrophic yeast *Pichia pastoris* as a superior host for the secretion of recombinant gelatins having natural amino acid sequences, at up to 14.8 g/L of clarified broth (Chapter 2).

Having established the suitability of *P. pastoris* for the expression of natural recombinant gelatins, we set out to investigate the possibilities of producing entirely custom-designed gelatins having novel physico-chemical properties. A monomeric gene encoding a highly hydrophilic 9 kDa gelatin was designed such, as to allow convenient polymerization into larger multimers. The monomeric gene is much longer than the single oligonucleotide monomers used in the expression of synthetic gelatins in *E. coli*, mentioned above. This offers more flexibility in the design of the amino acid sequence and a concomitant decrease in the overall repetitiveness of the gene. Here, we describe the high-level secretion of a fully

synthetic, highly hydrophilic and non-degraded 36.8 kDa gelatin by *P. pastoris* and its characterization.

4.2. Materials and methods

4.2.1. Vector construction

The monomeric gelatin gene (referred to hereafter as 'P' for 'Polar') was constructed by overlap extension PCR [13] of long oligonucleotides (underlined in Figure 4.1A). PCR was performed with a Perkin-Elmer GeneAmp 9700, using the proofreading enzyme *Pwo* DNA polymerase (Eurogentec). The 5' half of the gene was constructed by overlap extension of the first and second oligonucleotides and co-amplified by outer primers directed against nucleotides 1-26 (sense) and 197-174 (antisense). Likewise, the 3' half of the gene was constructed by overlap extension of the third and fourth oligonucleotide and co-amplified by primers directed against nucleotides 174-197 (sense) and 363-337 (antisense). The resulting PCR products were isolated from an agarose gel and were combined by another overlap extension PCR and co-amplified with the primers directed against nucleotides 1-26 (sense) and 363-337 (antisense). The resulting 0.3 kb PCR fragment was digested with *XhoI/EcoRI* and cloned in vector pMTL23 [14] to form vector pMTL23P. The sequence of the gene was verified by automated DNA sequencing of both strands.

The monomeric gene was released by digesting pMTL23P with *DraIII/Van91I* (Figure 4.1B). In a separate reaction the vector was linearized with *Van91I* and dephosphorylated. The *DraIII/Van91I* fragment was then inserted into this linearized vector to yield vector pMTL23P2. This process of insertional doubling can in principle be repeated to form multimers of any desired length, but was repeated only once here to form the vector pMTL23P4. The tetrameric gene (referred to hereafter as 'P4') was then cloned into the *XhoI/EcoRI* sites of vector pPIC9 (Invitrogen) that contains a *HIS4* selectable marker, an alcohol oxidase 1 (*AOX1*) promoter/terminator cassette and a *Saccharomyces cerevisiae* α -factor prepro secretory signal [15]. The Kex2 and dipeptidylaminopeptidase (DPAPase) cleavage sites of the α -factor prepro sequence are lost from pPIC9 when using the *XhoI* site, but are restored upon ligation of the gelatin gene by the sequence between *XhoI* and *DraIII* (Figure 4.1A).

4.2.2. Transformation of *P. pastoris*

Plasmid pPIC9P4 was linearized with *SalI* in order to obtain preferentially Mut⁺ transformants (i.e., by integration at the *his4* locus rather than the *AOX1* locus and thus allowing normal growth on methanol [16]). Transformation of *P. pastoris* strain GS115 (*his4*) [17] by electroporation and selection of Mut⁺ transformants was as described previously (Chapter 2).

4.2.3. Fermentative production of synthetic gelatin in *P. pastoris*

Fermentations were performed in 1-140 liter fermentors (Applikon) in minimal basal salt medium (Invitrogen) supplemented with 0.2% (v/v) PTM1 trace salts (Invitrogen). Methanol fed-batch fermentations were performed as described previously (Chapter 2), with the exception that no protease-inhibiting supplements such as casamino acids were added and that the pH during methanol fed-batch was maintained at 3.0 for all fermentations.

4.2.4. Small-scale purification of synthetic gelatin by differential acetone precipitation

Differential acetone precipitation was as described previously (Chapter 2). Chilled acetone was added to fermentation supernatant at 40% (v/v), after which endogenous proteins were pelleted by centrifugation. The acetone concentration in the supernatant was then increased to 80% (v/v) and the pellet obtained after centrifugation was washed with 80% acetone and air-dried.

4.2.5. Preparative purification of synthetic gelatin by differential ammonium sulfate precipitation

Preparative purification of synthetic gelatin from fermentation supernatant consisted of twice-repeated ammonium sulfate precipitation at 40% saturation (4 °C) and subsequent washing of the precipitate with 60% saturated ammonium sulfate. Depending on the scale of the purification, separation of the precipitate from the liquid was either by centrifugation or by depth filtration using AKS-4 sheets (USF Seitz-Schenk). The protein was subsequently desalted by diafiltration and lyophilized.

4.2.6. Bicinchoninic acid protein assay

A commercially available bicinchoninic acid (BCA) protein assay was used according to the manufacturer's recommendations (Pierce). The reaction was performed at 60 °C for 30 min. The calibration curve was prepared gravimetrically from lyophilized, desalted P4 gelatin, purified by differential ammonium sulfate precipitation (purity at least 98%).

4.2.7. SDS-PAGE

SDS-PAGE [18] was performed in a Mini-PROTEAN II system (Bio-Rad) under reducing denaturing conditions. Gels consisted of a 5% stacking and a 12.5% separating zone (2.7% crosslinking). Gels were stained using Coomassie PhastGel Blue R-350 (Amersham-Pharmacia Biotech) and were destained by heating in water using a microwave oven, similarly to Faguy et al. [19].

4.2.8. Gel filtration chromatography

Protein in 0.1 M sodium chloride was loaded on a 10 × 250 mm column packed with Superose 12 (Amersham-Pharmacia Biotech). Elution was carried out with 0.1 M sodium chloride at a flowrate of 0.2 mL/min, collecting 2 mL fractions and monitoring the absorbance at 214 nm.

4.2.9. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed at the Department of Biochemistry, Wageningen University, The Netherlands, using a Voyager DE-RP delayed extraction mass spectrometer (PerSeptive Biosystems). Samples were prepared by the dried droplet method, using sinapinic acid dissolved in 30% (v/v) acetonitrile, 4% (v/v) trifluoroacetic acid, as matrix. Measurements were made in the positive, linear mode and the accelerating voltage was 25,000 V. Cytochrome c and bovine serum albumin were used as external calibrants.

4.2.10. Chemical modification of gelatins

Esterification of carboxylic amino acid side chains was adapted from Wilcox [20]. A 100 µg amount of protein was incubated in 500 µL methanol, 0.1 M

hydrochloric acid at 4 °C for 72 h. The methanol was exchanged for 1 mM hydrochloric acid by diafiltration in a 3 kDa Microcon (Millipore). Removal of ester groups was performed by incubating the esterified protein in 100 mM Tris/HCl, pH 8.8 at 20 °C for 72 h.

Hydrazination of carboxylic amino acid side chains was performed according to [21]. A 50 µg amount of protein was dissolved in 20 µL of 50 mM sodium phosphate buffer, pH 7. After addition of 170 µL of 8 M urea, 1 M hydrazine and 0.1 M 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide, pH 4.5, the mixture was incubated at room temperature for 2 h.

4.2.11. Surface tension measurements

Surface tension at the liquid-air interface was measured according to the du Noüy ring method [22] using a Krüss K6 tensiometer. The temperature of the sample vessel was maintained at 20 °C. The actual measurements were performed 5 min after lowering of the ring onto the liquid surface, as suggested by [23]. Bovine serum albumin was used as a reference protein and raw data were corrected for the hydrostatic volume effect according to [24].

4.2.12. Circular dichroism spectrometry

Proteins were dissolved in Milli-Q water at 0.1 mg/mL. Measurements were performed at the Department of Biochemistry, Wageningen University, The Netherlands using a Jasco J-715 spectropolarimeter. The pathlength was 0.1 cm and spectra were recorded from 190 to 260 nm at 4 °C, using a scanning speed of 20 nm/min at a resolution of 0.1 nm.

4.3. Results

4.3.1. Design and construction of the synthetic P4 gelatin gene

The basic structure of natural gelatins consists of repeating Gly-Xaa-Yaa triplets, where Xaa and Yaa are often proline and hydroxyproline, respectively (the latter being postrationally modified proline). This structure maintains the open, unfolded conformation characteristic of gelatin. Owing to this unfolded conformation, gelatin is fairly hydrophilic because its hydrogen bonds are highly

exposed. Furthermore, only a small fraction of the protein is occupied by hydrophobic amino acids such as Trp, Tyr, Phe, Leu, Ile, Val and Met.

Our synthetic P4 gelatin design also provides the (Gly-Xaa-Yaa)_n structure. To increase its hydrophilicity relative to that of natural gelatins, we designed a gelatin without any hydrophobic amino acids other than proline and with a high content of the hydrophilic amino acids asparagine and glutamine (Table 4.1). To illustrate the high hydrophilicity of this synthetic gelatin compared with natural gelatins, the GRAVY values (grand average of hydropathy [25]) of P4, natural recombinant Col3a1 gelatin (Chapter 2), and cattle bone gelatin are indicated in Table 4.1.

The content of acidic and basic residues was modulated to give an isoelectric point similar to that of common limed bone gelatin (Table 4.1). Only Lys was used as a basic residue, because we anticipated proteolysis of mono-arginylic sites based on previous work (Chapter 2).

The P4 gene was constructed from four identical P monomers that were designed to have the codon usage of *P. pastoris* highly expressed genes [26]. The monomeric gene contains restriction sites for *Dra*III and *Van*91I. These enzymes allow the design of mutually complementary, non-palindromic overhangs that enable convenient elongation of the gene by insertional doubling in a fixed orientation (Figure 4.1). The process can be repeated until the desired polymer length has been achieved. This modular design offers flexibility in the construction of future gelatins, by allowing the combination of different types and lengths of polymerized gelatins *via* the *Dra*III/*Van*91I sites. The *Xho*I and *Eco*RI sites provided by the sequence allow the direct insertion of the final synthetic gene into *P. pastoris* expression vector pPIC9, resulting in a fusion to the α -mating factor prepro secretion signal. Thus, the combined four P modules were cloned into pPIC9 to yield vector pPIC9P4.

Table 4.1. Composition and basic physico-chemical parameters of synthetic and natural gelatins ^a

Amino acid	P4	Col3a1 ^b	Cattle bone gelatin ^c
Ala	1.0	7.9	12.0
Arg	0.0	4.8	5.0
Asn	12.0	3.5	-
Asp	0.0	3.1	-
Asx	-	-	4.3 ^d
Cys	0.0	0.0	0.0
Gln	16.0	3.5	-
Glu	4.0	3.5	-
Glx	-	-	6.7 ^d
Gly	33.7	33.8	33.4
His	0.0	1.8	0.5
Ile	0.0	1.3	1.2
Leu	0.0	1.8	2.4
Lys	3.0	3.5	2.6 ^e
Met	0.0	0.4	0.7
Phe	0.0	0.4	1.3
Pro	22.4	21.1	22.2 ^e
Ser	8.0	6.1	3.4
Thr	0.0	1.8	1.6
Trp	0.0	0.0	0.0
Tyr	0.0	0.4	0.2
Val	0.0	1.3	2.4
GRAVY	-1.77	-1.08	-0.75 to -1.09 ^e
Isoelectric point	4.9	9.7	4.7 to 5.4 ^f
Molecular weight (kDa)	36.8	20.6	65 to 300 ^f

^a Theoretical amino acid composition (mol%), GRAVY values, isoelectric point and molecular weight were calculated using the ProtParam tool available at the Expasy WWW server [27]. Experimental values were obtained as indicated below.

^b Natural non-hydroxylated recombinant gelatin produced in *P. pastoris* (Chapter 2).

^c The amino acid composition (mol%) of limed cattle bone gelatin is the mean of four measurements.

^d Asn/Asp and Gln/Glu cannot be distinguished by amino acid analysis and are given as Asx and Glx, respectively. This does not affect the GRAVY calculation.

^e Values for Lys and Pro include hydroxylysine and hydroxyproline, respectively. The hydropathy indices of these modified amino acids were assumed to lie between those of the respective unmodified amino acids and the lowest value of the Kyte and Doolittle hydropathy scale. The range of GRAVY values indicated for cattle bone gelatin represents both extremes.

^f Natural cattle bone gelatin is a heterogeneous mixture of molecules of different molecular weights and isoelectric points within the indicated ranges [28].

4.3.2. Production of synthetic P4 gelatin

Plasmid pPIC9P4 was used to transform *P. pastoris* GS115. Randomly chosen transformants were fermented and culture supernatants harvested throughout the fermentation were subjected to SDS-PAGE. The theoretical molecular weight of P4 is 36.8 kDa. Because collagenous proteins migrate in SDS-PAGE at an apparent molecular weight ~40% higher than the true molecular weight (Chapter 2) [29], one

would expect a band of ~52 kDa. Instead, however, the Coomassie Blue stained SDS-PAGE gel showed a faint blurry band at the top of the separating gel that had a tendency to diffuse from the gel during methanol-acetic acid destaining. Migration of the gelatin into the gel was improved by running the gel at 4 °C at twice the voltage recommended by the manufacturer of the electrophoresis system (i.e., 400 instead of 200 V). Diffusion of the protein from the gel during destaining was reduced by destaining the gel in water heated in a microwave oven, similarly to Faguy et al. [19], rather than performing the common lengthy incubations in methanol-acetic acid. Figure 4.2, lane 1 shows fermentation supernatant analyzed in this manner. N-terminal protein sequencing of this band (Sequencing Center Utrecht, The Netherlands) revealed the expected amino acid sequence (GPPGEPGNPG). There was no indication of incomplete processing of the α -factor derived Glu-Ala repeats by dipeptidylaminopeptidase, such as is occasionally observed when using this prepro sequence for secretion of heterologous proteins (Chapter 2) [30-32].

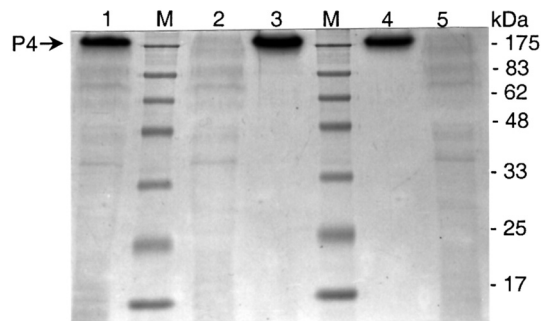


Figure 4.2. Purification of synthetic gelatin. Lane 1, fermentation supernatant; lane 2, precipitate obtained at 40% (v/v) acetone; lane 3, precipitate obtained by increasing the acetone concentration of the 40% (v/v) supernatant to 80% (v/v); lane 4, precipitate obtained at 40% of ammonium sulfate saturation; lane 5, supernatant obtained at 40% of ammonium sulfate saturation; lane M, molecular weight marker.

Recombinant gelatins with natural amino acid sequences can be purified from the fermentation broth by using differential acetone precipitation (Chapter 2). Endogenous extracellular proteins are precipitated at 40% (v/v) acetone, after which the gelatin is precipitated from the supernatant by addition of acetone to 80% (v/v). Figure 4.2, lanes 2 and 3 show that acetone precipitation was equally effective for the purification of synthetic gelatins. Amino acid analysis was performed (in triplicate) to estimate the purity of the precipitated gelatin. By linear

least-squares fitting of the observed data in terms of the P4 amino acid composition, the average relative contributions (vector moduli) of the P4 component and the residuals component were found to be 96.2 and 3.8%, respectively ($\pm 1.4\%$ SD). The purity of the preparation at the protein level is thus estimated to be $>96\%$, as stochastic fluctuations of measured values contribute to the residuals component. BCA protein analysis of acetone precipitates of different fermentations showed gelatin yields in the range of 3-6 g/L of clarified broth.

We previously found that recombinant gelatins with natural amino acid sequences can also be purified from *P. pastoris* fermentations using differential ammonium sulfate precipitation (unpublished data). Gelatinous proteins precipitate at 40% saturation, while endogenous extracellular *P. pastoris* proteins surprisingly do not precipitate at up to 80% saturation. In agreement with this, readily precipitable proteins such as β -lactoglobulin are rendered virtually unprecipitable upon mixing them with fermentation supernatant. We investigated if it was possible to purify P4 gelatin by differential ammonium sulfate precipitation. Indeed, Figure 4.2, lanes 4 and 5 show that synthetic gelatin is quantitatively precipitated at 40% ammonium sulfate saturation, while no endogenous proteins are visible. Based on amino acid analysis (in triplicate) and subsequent linear least squares fitting of the observed data, the purity at the protein level was estimated to be $>98.1\%$ ($\pm 0.7\%$ SD). Two-dimensional electrophoresis followed by silver staining showed virtually no contaminants (not shown). Gelatin yields determined by BCA analysis of ammonium sulfate precipitates of several fermentations (possible only after thorough desalting because ammonium sulfate reduces BCA reactivity [33]), were within 0.3 g/L of the values determined by acetone precipitation. This small difference ($<10\%$) is largely due to interference of the BCA assay by the small amount of reducing exopolysaccharides that co-purifies in the acetone precipitation procedure, while exopolysaccharides are virtually eliminated when using ammonium sulfate precipitation (data not shown). Compared with differential acetone precipitation, the overall higher purity obtained and the higher amenability to scale-up render differential ammonium sulfate precipitation the method of choice for preparative purification of P4 gelatin.

4.3.3. Establishing the molecular weight of synthetic P4 gelatin

A possible explanation for the aberrant molecular weight observed in SDS-PAGE could be that P4 is glycosylated. N-linked glycosylation can be ruled out because no susceptible sites are present in the amino acid sequence. However, *P. pastoris* is also able to perform O-glycosylation and the structural determinants for such an event are unclear [34]. To rule out this possibility, periodic acid - Schiff staining [35] and Alcian Blue staining [36] were performed on ammonium sulfate purified P4. No glycosylation was observed (not shown).

To determine whether the synthetic gelatins have in fact the correct molecular weight, but merely exhibit aberrant behavior in SDS-PAGE, analytical gel filtration chromatography was performed. The Superose 12 column was calibrated with a mixture of natural recombinant gelatin fragments (Chapter 2), giving a series of molecular weights of 53, 42, 28, 16, 12 and 8 kDa. Ammonium sulfate purified P4 was subjected to gel filtration chromatography. Only one significant peak was observed. N-terminal protein sequencing of this fraction in solution (Sequencing Center Utrecht) showed the correct N-terminus for P4. The molecular weight of P4 deduced from the chromatogram was 47 kDa. This is clearly much closer to the theoretical value of 36 kDa than the molecular weight apparent from SDS-PAGE, although the deviation is still significant.

Mass spectrometry was used to ultimately determine the molecular weight of P4. Materials purified by both ammonium sulfate precipitation and gel filtration chromatography were analyzed and the results were in good mutual agreement. Figure 4.3 shows the MALDI-TOF mass spectrum of P4 purified by ammonium sulfate precipitation. The observed molecular weight of 36,835 Da corresponds well with the theoretical value of 36,818 Da. This result shows that the apparent high molecular weight observed in SDS-PAGE is indeed the result of aberrant migration behavior. Furthermore, the SDS-PAGE and gel filtration chromatography results are confirmed, in that there is no presence of proteolytically degraded fragments.

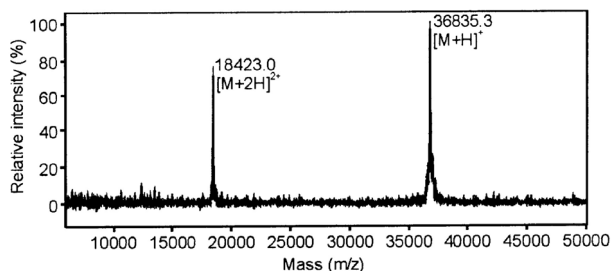


Figure 4.3. MALDI-TOF mass spectrum of synthetic gelatin purified by differential ammonium sulfate precipitation. The $[M + H]^+$ and $[M + 2H]^{2+}$ peaks correspond to singly and doubly charged molecular ions, respectively.

We speculated that the aberrant migration rate of synthetic gelatin compared with normal gelatin was due to low binding of SDS in view of its high hydrophilicity, as the interaction of SDS with proteins is mainly of a hydrophobic nature [37]. Esterification of the carboxylic amino acid side chains with methanol-hydrochloric acid would increase the protein's hydrophobicity and thus its SDS binding capacity and migration rate. While the migration of natural gelatins treated in this way was only slightly affected (i.e., a decrease in apparent molecular weight of about 2 kDa), the esterified synthetic gelatin migrated much faster than the unmodified protein (Figure 4.4).

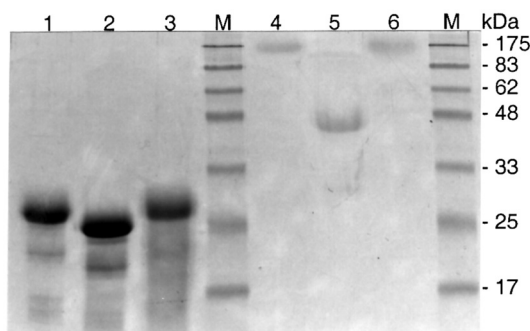


Figure 4.4. Esterification of natural recombinant Col3a1 and synthetic P4 gelatin. Lane 1, unmodified Col3a1 natural recombinant gelatin; lane 2, esterified Col3a1; lane 3, esterified Col3a1 after hydrolysis of the ester groups; lane 4, unmodified P4; lane 5, esterified P4; lane 6, esterified P4 after hydrolysis of the ester groups; lane M, molecular weight marker.

The molecular weight observed for esterified P4 was about 50 kDa. As natural gelatins migrate about 40% more slowly than common proteins (Chapter 2) [29], this value is in good agreement with the value expected for natural gelatins of the

same molecular weight as P4 (i.e., a 36.8 kDa natural gelatin would migrate at ~52 kDa). Removal of the ester groups at high pH restored the aberrant migration of P4 (Figure 4.4). Hydrazination of the carboxylic amino acid side chains of P4 did not result in an altered migration rate (not shown), indicating that the effect of esterification was not due to the mere removal of negative charge.

4.3.4. Characterization of synthetic P4 gelatin

In marked contrast with recombinant gelatins having natural sequences, we noticed that highly concentrated solutions of P4 showed essentially no foaming. A direct relationship exists between protein (surface) hydrophobicity, surface tension and foam stability [38]. Therefore, the surface activity of P4 relative to that of Col3a1 natural recombinant gelatin was determined, using the du Noüy ring method [22]. Figure 4.5 shows that P4 does not show any significant lowering of the surface tension of water at concentrations up to 5% (w/v), while Col3a1 already has an effect at 0.01%. Within the range up to 10% of protein, it was not possible to determine the apparent critical micelle concentration (CMC; i.e., the concentration whereby the surface tension curve reaches a plateau phase) for either of the gelatin types. For comparison, bovine serum albumin has an apparent CMC of about 0.003% [23].

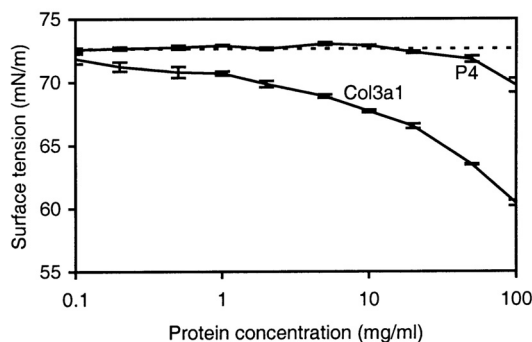


Figure 4.5. Surface tension of natural recombinant Col3a1 and synthetic P4 gelatin as a function of concentration. Error bars indicate the standard deviation of five measurements. The dotted line indicates the surface tension of water.

We previously showed that natural, non-hydroxylated recombinant gelatins produced in *P. pastoris*, do not form collagen triple helices even at 4 °C (Chapter 2). Synthetic P4 gelatin was also non-hydroxylated, as shown by the amino acid

analyses and N-terminal sequencing described above. Circular dichroism spectrometry was performed to see whether the greater hydrophilicity of P4 somehow influences triple helix formation. Figure 4.6 shows the spectrum of P4 and natural recombinant Col3a1 gelatin at 4 °C. Both proteins show a clear absence of the positive peak at about 221 nm characteristic of the collagen triple helix [39, 40]. The mean residue ellipticity of P4 at that position of the spectrum is lower than that of Col3a1.

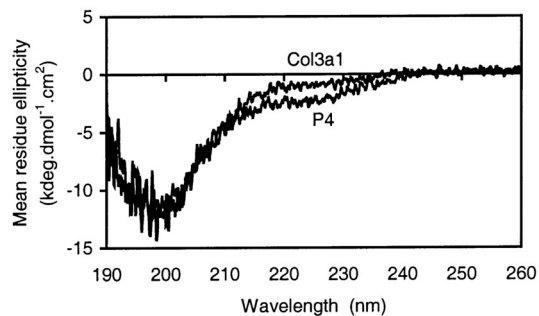


Figure 4.6. Circular dichroism spectra of natural recombinant Col3a1 and synthetic P4 gelatin at 4 °C.

4.4. Discussion

This paper describes the extracellular production of a recombinant, custom-designed and highly hydrophilic gelatin in *P. pastoris*. Yields in single-copy transformants were 3-6 g/L of clarified broth, which is comparable to the yields obtained in single-copy transformants for natural gelatins (Chapter 2).

Despite the obvious physico-chemical differences, both synthetic P4 and natural recombinant gelatin could be purified from the fermentation broth by both differential acetone precipitation and differential ammonium sulfate precipitation. The universality of the two purification techniques is probably due to the hydrophilicity and unfolded structure of gelatins in general. Especially differential ammonium sulfate precipitation allowed convenient large-scale purification of P4 gelatin to near homogeneity.

Secreted synthetic gelatin was fully intact, as evidenced by SDS-PAGE, gel filtration chromatography, N-terminal sequencing and mass spectrometry. This is in contrast with natural recombinant gelatins produced in *P. pastoris*, which were partly degraded (Chapter 2). Apart from the occurrence of some minor

background degradation, collagen type I derived natural recombinant gelatins were cleaved into several major bands by a Kex2-like protease. Cleavage occurred C-terminal of two occurrences of the mono-arginylic sequence Met-Gly-Pro-Arg. We speculated that the amino acids occupying the -2 and -4 positions in this motif (relative to the site of cleavage) were a major factor determining the cleavage efficiency (Chapter 2), which is in accordance with recent data on the substrate specificity of *S. cerevisiae* Kex2 [41, 42]. It may well be that the above-mentioned minor background degradation represented a limited extent of cleavage at Arg residues having 'suboptimal' residues at the -2 and -4 positions. Therefore, in the design of the synthetic gelatin described here, only Lys was used as a basic residue to control the isoelectric point. The finding that secreted P4 was completely intact does indeed suggest a general susceptibility of Arg residues in recombinant gelatins to proteolysis, and may thus have implications for the rational design of (partially) unfolded proteins to be expressed extracellularly in *P. pastoris*.

While the molecular weight of P4 as determined by mass spectrometry was in good agreement with the value deduced from the amino acid sequence, the molecular weight apparent from gelatin-calibrated gel filtration chromatography was about 10 kDa higher. Ionic interactions with Superose 12 are negligible in the presence of salt and only small hydrophobic peptides appear to show significant hydrophobic interactions with this matrix [43]. It is therefore not very likely that a lower degree of such interactions of P4 relative to natural gelatins causes the seemingly aberrant molecular weight. Possibly, the effect is due to an increased hydrodynamic size of P4 relative to natural gelatins, as a result of increased interaction of this highly hydrophilic protein with water.

Natural gelatins migrate about 40% more slowly in SDS-PAGE than expected. Several possible explanations for the aberrant migration behavior of gelatins have been suggested [29, 44-47]. It is most likely not the result of anomalously low SDS-binding, but is at least in part due to the low average residue molecular weight of gelatin, resulting in a relatively high number of residues (i.e., molecular length) per unit of molecular weight [29, 45, 47]. Most other reports on aberrant protein migration rates in SDS-PAGE involve highly acidic proteins that show reduced binding of SDS due to electrostatic repulsion by the protein's high negative net charge [21, 48-50]. SDS-PAGE showed that P4 migrates at a highly reduced rate, even much more slowly than natural gelatins. We showed here that esterification

of the carboxylic side chains of P4 restores its migration rate roughly to that expected for normal gelatins (i.e., about 40% more slowly than common proteins). In contrast, hydrazination did not affect the migration rate of P4. Hydrazination eliminates the negative charge of the same carboxylic residues as does esterification, but reduces the protein's hydrophobicity while esterification increases it. As the binding of SDS to proteins is primarily hydrophobic in nature [37], the extremely slow migration of the highly polar P4 gelatin in SDS-PAGE is therefore most likely the result of insufficient SDS binding and a concomitant low negative net charge. The finding that the resolution of the SDS-PAGE was improved by increasing the field strength to twice that recommended by the manufacturer of the electrophoresis system indicates that the higher field strength aids protein migration in overcoming diffusive forces.

Surface activity is a major determinant in a protein's function as a protective colloid (e.g., in photographic emulsions). Solutions of P4 showed essentially no foaming and tensiometric analysis of solutions of P4 in water showed only negligible surface activity. P4 thus represents a novel hydrocolloid combining some of the characteristics unique to gelatins and a low surface activity commonly expected only for the most hydrophilic polysaccharide hydrocolloids.

We previously showed that non-hydroxylated, natural recombinant gelatins do not show triple-helical structure in circular dichroism spectrometry (Chapter 2). It is a well-established fact that hydroxyproline residues play a crucial role in the stabilization of the collagen triple helix. This role is easily recognized when examining the amino acid compositions and thermal stabilities of natural collagens from different species [51]. X-ray crystallography showed that the triple helix is surrounded by a cylinder of hydration [52]. Although recently questioned [53, 54], the role of hydroxyproline in the stabilization of the triple helix is generally attributed to its hydrogen bonding with this water network [55]. In view of the high polarity of P4, we considered it prudent to investigate its conformation using circular dichroism spectrometry. No triple helical structure was observed at 4 °C and P4 gelatin is thus an essentially non-gelling gelatin. Non-gelling gelatins permit novel applications such as low temperature silver halide crystallization in the preparation of photographic emulsions [7]. Comparison of the circular dichroism spectrum of P4 with that of Col3a1 natural recombinant gelatin showed that the mean residue ellipticity of P4 at the discriminating wavelength of 221 nm

was actually lower than that of Col3a1. Although the latter is essentially in a random coil conformation, the higher hydrophilicity of P4 probably reduces minor intramolecular and intermolecular interactions, thereby resulting in a slightly lower ellipticity.

Current research is directed towards the production of other synthetic gelatins with distinct functionalities and combining them to form chimeric tailor-made biopolymers.

Acknowledgments

We thank Ir. Friso B.J.van Assema, Ir. Harald G.de Jonge, Ing. Pieter H.Lodder, Ir. Marcia P.M.Dielissen and Ing. Marjo Koets for assistance, Ir. Harrold A.van den Burg (Department of Biochemistry, Wageningen University, The Netherlands) for mass spectrometry, Dr. Ir. Carlo P.M.van Mierlo (Department of Biochemistry, Wageningen University, The Netherlands) for help with circular dichroism measurements and Dr. Ir. Johan M.H.Stoop for critical reading of the manuscript.

References

1. Asghar A, Henrickson RL (1982). *Adv Food Res* 28, 231-372.
2. Saddler JM, Horsey PJ (1987). *Anaesthesia* 42, 998-1004.
3. Pollack SV (1990). *J Dermatol Surg Oncol* 16, 957-961.
4. Rao KP (1995). *J Biomater Sci Polym Ed* 7, 623-645.
5. Courts A (1980), in *Applied Protein Chemistry*, (Grant BA, Ed.), Applied Science, London, pp. 1-29.
6. Van Heerde GV, van Rijn AC, Bouwstra JB, de Wolf FA, Mooibroek H, Werten MWT, Wind RD, van den Bosch TJ (1998). US Pat 6,150,081.
7. De Wolf FA, Werten MWT, Wisselink HW, Jansen-van den Bosch TJ, Toda Y, van Heerde GV, Bouwstra JB (2000). *Eur Pat Appl* 1 014 176.
8. Goldberg I, Salerno AJ, Patterson T, Williams JI (1989). *Gene* 80, 305-314.
9. Obrecht G, Lefèvre, J. F. and Meyrueis, P. (1991). *Fr Pat Appl* 91 16215.
10. Gardner K, Lock RL, O'Brien JP, Salemmme FR (1992). *Pat Appl PCT/US92/09655*.
11. Cappello J, Ferrari F (1994), in *Plastics from microbes*, (Mobley DP, Ed.), Carl Hanser Verlag, Munich, pp. 35-92.
12. Kajino T, Takahashi H, Hirai M, Yamada Y (2000). *Appl Environ Microbiol* 66, 304-309.
13. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989). *Gene* 77, 51-59.
14. Chambers SP, Prior SE, Barstow DA, Minton NP (1988). *Gene* 68, 139-149.
15. Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA (1991). *Gene* 105, 205-212.
16. Clare JJ, Rayment FB, Ballantine SP, Sreekrishna K, Romanos MA (1991). *Nat Biotechnol* 9, 455-460.
17. Cregg JM, Barringer KJ, Hessler AY, Madden KR (1985). *Mol Cell Biol* 5, 3376-3385.
18. Laemmli UK (1970). *Nature* 227, 680-685.
19. Faguy DM, Bayley DP, Kostyukova AS, Thomas NA, Jarrell KF (1996). *J Bacteriol* 178, 902-905.
20. Wilcox PE (1967). *Methods Enzymol* 11, 607-617.
21. Matagne A, Joris B, Frère JM (1991). *Biochem J* 280, 553-556.
22. Lecomte du Noüy P (1919). *J Gen Physiol* 1, 521-524.
23. Clarkson JR, Cui ZF, Darton RC (1999). *J Colloid Interface Sci* 215, 333-338.
24. Harkins WD, Jordan HF (1930). *J Am Chem Soc* 52, 1751- 1771.
25. Kyte J, Doolittle RF (1982). *J Mol Biol* 157, 105-132.
26. Sreekrishna K, Kropp KE (1996), in *Nonconventional Yeasts in Biotechnology. A Handbook.*, (Wolf K, Ed.), Springer, Berlin, pp. 203-253.
27. Appel RD, Bairoch A, Hochstrasser DF (1994). *Trends Biochem Sci* 19, 258-260.
28. Alleavitch J, Turner WA, Finch CA (1988), in *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A12, Wiley-VCH, Weinheim, pp. 307-317.
29. Butkowski RJ, Noelken ME, Hudson BG (1982). *Methods Enzymol* 82, 410-423.
30. Vedvick T, Buckholz RG, Engel M, Urcan M, Kinney J, Provow S, Siegel RS, Thill GP (1991). *J Ind Microbiol* 7, 197-201.
31. Briand L, Perez V, Huet JC, Danty E, Masson C, Pernollet JC (1999). *Protein Express Purif* 15, 362-369.
32. Goda S, Takano K, Yamagata Y, Katakura Y, Yutani K (2000). *Protein Eng* 13, 299-307.
33. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goetze NM, Olson BJ, Klenk DC (1985). *Anal Biochem* 150, 76-85.
34. Duman JG, Miele RG, Liang H, Grella DK, Sim KL, Castellino FJ, Bretthauer RK (1998). *Biotechnol Appl Biochem* 28 (Pt 1), 39-45.
35. Zacharius RM, Zell TE, Morrison JH, Woodlock JJ (1969). *Anal Biochem* 30, 148-152.
36. Wardi AH, Michos GA (1972). *Anal Biochem* 49, 607-609.

37. Reynolds JA, Tanford C (1970). *Proc Natl Acad Sci USA* 66, 1002-1007.
38. Horiuchi T, Fukushima D, Sugimoto H, Hattori T (1978). *Food Chem* 3, 35-42.
39. De Wolf FA, Keller RCA (1996). *Progr Colloid Polym Sci* 102, 9-14.
40. Rossi A, Zuccarello LV, Zanaboni G, Monzani E, Dyne KM, Cetta G, Tenni R (1996). *Biochemistry* 35, 6048-6057.
41. Bevan A, Brenner C, Fuller RS (1998). *Proc Natl Acad Sci USA* 95, 10384-10389.
42. Suzuki Y, Ikeda N, Kataoka E, Ohsuye K (2000). *Biotechnol Appl Biochem* 32, 53-60.
43. Andersson T, Carlsson M, Hagel L, Pernemalm P, Janson J (1985). *J Chromatogr* 326, 33-44.
44. Furthmayr H, Timpl R (1971). *Anal Biochem* 41, 510-516.
45. Freytag JW, Noelken ME, Hudson BG (1979). *Biochemistry* 18, 4761-4768.
46. Hayashi T, Nagai Y (1980). *J Biochem* 87, 803-808.
47. Noelken ME, Wisdom BJJ, Hudson BG (1981). *Anal Biochem* 110, 131-136.
48. Ohara O, Teraoka H (1987). *FEBS Lett* 211, 78-82.
49. Casarégola S, Jacq A, Laoudj D, McGurk G, Margaron S, Tempête M, Norris V, Holland IB (1992). *J Mol Biol* 228, 30-40.
50. McGrath KP, Fournier MJ, Mason TL, Tirrell DA (1992). *J Am Chem Soc* 114, 727-733.
51. Privalov PL (1982). *Adv Protein Chem* 35, 1-104.
52. Bella J, Eaton M, Brodsky B, Berman HM (1994). *Science* 266, 75-81.
53. Holmgren SK, Bretscher LE, Taylor KM, Raines RT (1999). *Chem Biol* 6, 63-70.
54. Nagarajan V, Kamitori S, Okuyama K (1999). *J Biochem* 125, 310-318.
55. Brodsky B, Shah NK (1995). *FASEB J* 9, 1537-1546.

Chapter 5

Biosynthesis of an amphiphilic silk-like polymer

An amphiphilic silk-like protein polymer was efficiently produced in the yeast *Pichia pastoris*. The secreted product was fully intact and was purified by solubilization in formic acid and subsequent precipitation of denatured host proteins upon dilution with water. In aqueous alkaline solution, the negatively charged acidic polymer assumed extended helical (silk III-like) and unordered conformations. Upon subsequent drying, it assumed a conformation rich in β -turns. In water at low pH, the uncharged polymer aggregated and the solution became turbid. Concentrated solutions in 70% (v/v) formic acid slowly formed gels. Replacement of the formic acid-water mixture with methanol and subsequent drying resulted in β -sheets, which stacked into fibril-like structures. The novel polymer instantaneously lowered the air-water interfacial tension under neutral to alkaline conditions, and reversed the polarity of hydrophobic and hydrophilic solid surfaces upon adsorption.

Published as: Werten MWT, Moers APHA, Vong T, Zuilhof H, van Hest JCM, de Wolf FA (2008). *Biomacromolecules* 9, 1705-1711.

5.1. Introduction

Material scientists have long been investigating ways of producing stimulus-responsive, nanostructured, and self-assembling polymers for high-end applications such as self-healing coatings, nanofibers, nanoelectronic assemblies, surgery, tissue engineering, and drug delivery. Because of the limited level of control over conventional polymerization processes, interest has increased for biological polymeric materials and proteins in particular. The monomers that constitute a protein polymer are monoenantiomeric (L- α) amino acids, and their sequence is exactly defined by an invariant genetic (DNA) template.

Proteins vary greatly in characteristics, such as their specific folding and 3D structure, their hydrophobicity and net charge, and their interaction with other molecules. Protein engineering allows to adapt the natural DNA templates or design novel templates and, thus, to produce monodisperse polymers, consisting of one or multiple blocks, each of which can be designed to be inert, stimulus-responsive, or promoting self-assembly into specific supramolecular structures. Protein polymers that have thus far been produced by genetic engineering include elastin-like, collagen-like, silk-like, leucine zipper-like, and newly designed sequences. These biomimetic materials have been excellently reviewed elsewhere [1-5].

Real-life application of biosynthetic polymers requires that a sufficiently high productivity is achieved. We previously reported the secreted production of collagenous polymers in the yeast *Pichia pastoris* at very high levels (Chapters 2 to 4). In the present work we studied the possibility of secreted production in this host of pH-responsive surface active polymers for coating applications. As a structural framework for the design of such polymers, we chose a silk-like polymer pioneered by the group of Tirrell [6], consisting of repeating (GA)₃GE octapeptides. The acidic glutamate residues, in combination with the self-assembling properties of silk Gly-Ala repeats [7], impart pH-responsive properties. When dried from certain solvents [6, 8, 9] it forms insoluble, silk crystal-like stacks of antiparallel β -sheets, bordered by Gly-Glu-Gly (3-residue) γ -turns, with successive glutamic acid residues on opposite sides of the sheet. To elicit amphiphilic characteristics, we replaced alternate glutamates with leucine, as a highly hydrophobic and otherwise inert, aliphatic amino acid, resulting in poly-[(GA)₃GE(GA)₃GL]. We describe here the efficient biosynthesis of this amphiphilic polymer in the yeast

P. pastoris, its purification, and the characterization of its physicochemical and structural properties.

5.2. Materials and methods

5.2.1. Generation of recombinant strains

A double-stranded DNA monomer EE, encoding the amino acid sequence (GA)₃GE(GA)₃GE, was produced by annealing of two complementary oligonucleotides EE-FW and EE-RV (Table 5.1). Likewise, a monomer EL encoding (GA)₃GE(GA)₃GL was produced by annealing of the complementary oligonucleotides EL-FW and EL-RV (Table 5.1). Both double-stranded adapters have *EcoRI/XhoI* overhanging ends. The adapters were ligated into an *EcoRI/XhoI* digested pMTL23 vector [10], which had been modified to remove the *BsaI* site normally present. The inserts were multimerized by digestion with *BsaI/BanI* and recursive directional ligation. Clones of similar size were selected for both protein types: EE24, bearing 24 repeats of the EE monomer, and EL28, bearing 28 repeats of the EL monomer. The inserts were cloned into the *EcoRI/NotI* sites of expression vector pPIC9 (Invitrogen, Breda, The Netherlands). The resulting vectors were linearized with *SalI* to promote homologous integration at the *his4* locus upon transformation of *P. pastoris* GS115 by electroporation, as described previously (Chapter 2). See Table 5.1 for the EE24 and EL28 amino acid sequences expected upon expression of the cloned genes.

Table 5.1. Oligonucleotide and protein sequences

Name	DNA or amino acid sequence
EE-FW ^a	AATTCGGTCTCGGTGCTGGTGCTGGTGCTGGTGAGGGAGCCGGTGCTGGAGCCGCGAAGGTGCCTAAGCGGCCGC
EE-RV ^a	TCGAGCGGCCGCTTAGGCACCTTCGCCGGCTCCAGCACCGGCTCCCTCACCAGCACCAGCACCAGCACCAGACCG
EL-FW ^a	AATTCGGTCTCGGTGCTGGTGCTGGTGCTGGTGAGGGAGCCGGTGCTGGAGCCGCTTGGGTGCCTAAGCGGCCGC
EL-RV ^a	TCGAGCGGCCGCTTAGGCACCCAAAGCCGGCTCCAGCACCGGCTCCCTCACCAGCACCAGCACCAGCACCAGACCG
EE24 ^b	YVEFGL[(GA) ₃ GE(GA) ₃ GE] ₂₄ GA
EL28 ^b	YVEFGL[(GA) ₃ GE(GA) ₃ GL] ₂₈ GA

^a DNA sequence (5'-3') of oligonucleotides used for construction of gene monomers.

^b Theoretical amino acid sequence of mature protein.

5.2.2. Fermentation of *P. pastoris*

Fed-batch fermentations were performed in 2.5-liter Bioflo 3000 fermentors (New Brunswick Scientific, Nijmegen, The Netherlands), essentially as described by Zhang et al. [11]. Minimal basal salts medium [11] was used and no protease-

inhibiting supplements were added. The pH throughout the entire fermentation was maintained at 5.0 (EE24) or 3.0 (EL28) by addition of ammonium hydroxide as base. The methanol fed-batch phase for protein production lasted three to four days. A homemade semiconductor gas sensor-controller, similar to that described by Katakura et al. [12], was used to monitor the methanol level in the off-gas and to maintain a constant level of ~0.2% (w/v) methanol in the broth. At the end of the fermentation, the cells were separated from the broth by centrifugation for 10 min at $10,000 \times g$ and 4 °C in a Sorvall SLA-3000 rotor (Thermo Scientific, Breda, The Netherlands), followed by microfiltration of the supernatant. For EE24, the pH of the culture was raised to 10.0 by addition of sodium hydroxide prior to centrifugation to prevent the formation of a gel layer on top of the cell pellet.

5.2.3. Purification of EE24 and EL28

All centrifugation was done for 30 min at $20,000 \times g$ and 4 °C, interchangeably in a Sorvall SLA-1500 or SLA-3000 rotor (Thermo Scientific).

The EE24 protein was precipitated from the cell-free broth by adjusting the pH to 3.0 with sulfuric acid followed by centrifugation. The precipitate was resuspended in an equal volume (relative to the cell-free broth) of Milli-Q water adjusted to pH 9.0 with sodium hydroxide. The pH was then adjusted to 3.0 with sulfuric acid, ammonium sulfate was added to 30% of saturation, and the sample was centrifuged. The ammonium sulfate precipitation was repeated once, and the final pellet was resuspended in 0.2 volumes (relative to the cell-free broth) of Milli-Q water and the pH was adjusted to 9.0 with ammonium hydroxide. Ethanol was added to 80% (v/v) and the sample was centrifuged. The supernatant containing a portion of the pure EE24 protein was collected. The pellet was subjected to several more rounds of resuspension and ethanol precipitation, to increase the recovery of the EE24 protein extracted in ethanol. The protein was then precipitated from the combined supernatants by addition of sodium chloride to 50 mM and centrifugation. The air-dried pellet was resuspended in Milli-Q water, and extensively desalted by repeated isoelectric precipitation using a volatile acid (formic acid) and/or dialysis against Milli-Q water. The final EE24 product was lyophilized.

The EL28 protein was precipitated from the cell-free broth by addition of acetone to 40% (v/v) and centrifugation. The pellet was resuspended in an equal

volume (relative to the cell-free broth) of 100% (v/v) formic acid and incubated overnight at room temperature. The liquid was diluted 15 times with Milli-Q water and centrifuged to pellet contaminating proteins. The EL28 protein was precipitated from the supernatant by addition of ammonium sulfate to 40% of saturation and centrifugation. This step was repeated once, and the final pellet was resuspended in an equal volume (relative to the cell-free broth) of Milli-Q water. The protein was then precipitated by addition of 40% (v/v) acetone and centrifugation. The air-dried pellet was resuspended, desalted, and lyophilized as described above for EE24.

5.2.4. SDS-PAGE, densitometry and N-terminal protein sequencing

The NuPAGE Novex system (Invitrogen) was used for SDS-PAGE, with 10% Bis-Tris gels, MES SDS running buffer and SeeBlue Plus2 prestained molecular mass markers. All protein samples were preincubated for 30 min with 6 M urea, pH 8.5, prior to mixing with NuPAGE LDS loading dye (Invitrogen), and were not heated before electrophoresis. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). Quantification of protein bands by densitometry was performed with a GS-800 calibrated densitometer (Bio-Rad, Veenendaal, The Netherlands), relative to a calibration curve of the corresponding purified protein (see Results and Discussion for purity). The standard solution used to make the calibration curve was prepared by weighing a known amount of the pure lyophilized protein. Blotting of proteins for N-terminal sequencing by Edman degradation was as described previously (Chapter 2). Protein sequencing was performed by Midwest Analytical (St. Louis, MO).

5.2.5. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed using an Ultraflex mass spectrometer (Bruker, Wormer, The Netherlands), following a procedure modified after Wenzel et al. [13]. Samples were prepared by the dried droplet method on a 600 μ m AnchorChip target (Bruker), using 5 mg/mL 2,5-dihydroxyacetophenone, 1.5 mg/mL diammonium hydrogen citrate, 25% (v/v) ethanol and 7% (v/v) trifluoroacetic acid as matrix. Measurements (50 shots at 20 Hz) were made in the positive, linear mode, with the following parameters: ion source 1, 20 kV; ion source 2, 18.45 kV;

lens 5 kV, pulsed ion extraction, 550 ns. Protein Calibration Standard II (Bruker) was used for external calibration.

5.2.6. Protein crystallization

Crystallization of proteins was performed in 70% (v/v) formic acid (Merck, reagent grade, ≤ 10 ppm nonvolatile matter) under vapor diffusion of methanol, similar to Cantor et al. [8]. Samples were incubated overnight at room temperature at 30 mg/mL in 70% (v/v) formic acid under gentle agitation. The cup containing the sample was then opened and placed in a bigger closed vessel containing a bottom layer of 100% (v/v) methanol. After incubation for 48 h at room temperature the protein gel was washed three times with 100% (v/v) methanol.

5.2.7. Infrared spectroscopy

Protein gels in 100% (v/v) methanol were smeared onto a zinc selenide attenuated total reflection (ATR) crystal and allowed to dry. Infrared spectra were recorded using a Vector 22 Fourier transform infrared spectroscopy (FTIR) spectrometer (Bruker) set as follows: resolution 4 cm^{-1} , scanner speed 10 kHz, number of scans 1024.

5.2.8. Circular dichroic spectroscopy

For measurement in solution the proteins were dissolved in 50 mM sodium borate, pH 9.0. Alternatively, proteins were dissolved in dilute ammonium hydroxide, pH 9.0, and dried as a film onto the outside of a cuvette before measuring. Circular dichroic (CD) spectra were recorded at 20 °C using a J-715 spectropolarimeter (Jasco, IJsselstijn, The Netherlands). Spectra were obtained as the average of 10 consecutive scans from 260 to 190 nm, using a scanning speed of 100 nm/min at a resolution of 0.2 nm. Data were smoothed by applying a moving average filter with a window size of 2 nm, and the spectra were normalized to similar peak intensities in view of their nonquantitative character.

5.2.9. Drop tensiometry

The liquid-air surface tension as a function of time was measured at room temperature using an automated drop tensiometer [14] (IT Concept, Longessaigne,

France). Each measurement started with a clean interface of a newly formed drop of protein solution (10-15 μL) in air. Proteins were freshly prepared in 50 mM sodium phosphate buffer at 2 g/L (pH 8.0) or 0.2 g/L (pH 3.0). Surface tension was determined by drop shape analysis.

5.2.10. Contact angle measurements

Polystyrene discs cut from petridishes (Greiner Bio-One, Alphen aan den Rijn, The Netherlands) or 0.03 mm aluminum discs (Wanit Universal GmbH & Co., Dietzenbach, Germany) were thoroughly washed with Milli-Q water and 100% (v/v) ethanol before use. After air drying, the discs were submerged for 48 h at room temperature in 70% (v/v) formic acid (blank) or in a 1 mg/mL protein solution in 70% (v/v) formic acid. The discs were then incubated for three hours at -20 °C and subsequently rinsed with 10 mM hydrochloric acid, followed by two agitated washes for 45 min in Milli-Q water and air drying. Sessile drop contact angles were determined using a VCA Optima XE surface analysis system (AST Products, Billerica, MA). Milli-Q water droplets of 2 μL (polystyrene) or 1.5 μL (aluminum) were applied, and contact angles (expressed as the average of the left and right values) were determined after a delay of five seconds. A total of 15 replicate droplets divided over three plates were measured for each sample.

5.2.11. Atomic force microscopy

Proteins were crystallized as described above, but at 1 mg/L protein. Si(100) wafers (n-type, 1-5 Ω/cm) modified with a carboxylic acid-terminated monolayer of undecylenic acid [15-17] were submerged in the protein solution during crystallization. The discs were then washed with Milli-Q water to remove nonadsorbed material and with 100% (v/v) methanol to facilitate air drying. Atomic force microscopy (AFM) was in tapping mode in air using a Dimension 3100 or MultiMode system, a NanoScope IV controller and NanoScope 6.12r2 software (Digital Instruments, Santa Barbara, CA), and NSG10 cantilevers (NT-MDT, Zelenograd, Russia). Image processing (flattening and plane fitting) and analysis (line profiles) were done using WSXM 4.0 software [18].

5.3. Results and discussion

5.3.1. Biosynthesis and purification

Fermentations were performed with *P. pastoris* strains expressing the EE24 nonamphiphilic prototype and the EL28 amphiphilic variant as extracellular proteins (see Table 5.1 for protein sequences). Initial cultures were done at pH 3.0, but for EE24 it was found necessary to culture at pH 5.0 because the secreted protein would otherwise form a gelly precipitate. Culture supernatants were analyzed by SDS-PAGE (Figure 5.1A, lanes 1-2). The samples subjected to SDS-PAGE contained urea (see Materials and Methods), because otherwise the EL28 protein was often not detectable on gel except as halo-like bands just below the wells (as confirmed by N-terminal sequencing). The urea probably solubilized nonelectrophoresable aggregates. EE24 and EL28 migrate slower in SDS-PAGE than would be expected on the basis of their molecular weight (28,248 and 32,366 Da, respectively), which is a known anomaly for this type of protein [19].

EE24 was purified essentially by a combination of isoelectric precipitation and subsequent removal of remaining host proteins by ethanol precipitation. We were unable to purify EL28 using the same or other common approaches (differential salt precipitation, filtration, ion exchange and hydrophobic interaction chromatography), in that the product consistently copurified with host proteins. In view of the above-mentioned aggregation problems during SDS-PAGE, it seems plausible that the amphipathic EL28 formed insoluble complexes with these contaminating proteins. Indeed, a successful purification method was found that relies on the use of pure formic acid as an effective organic solvent to disrupt intra- and intermolecular protein interactions. Upon dilution of the solubilized proteins with water the denatured host proteins precipitated nearly quantitatively, while the silk-like protein remained soluble, at least for some time. The purification methods developed for EE24 and for EL28 both resulted in highly pure and intact protein (Figure 5.1A, lanes 3-4). The purity of both polymers on the protein level was estimated to be at least 98%, based on amino acid analysis and subsequent linear least-squares fitting of the observed data to a combination of (1) the theoretical composition of the respective pure protein and (2) the composition determined for host-derived proteins present in the medium (not shown). Polysaccharides are the major nonprotein contaminant in these preparations, and were present at less than 1% (w/w), as determined by a carbohydrate assay [20].

MALDI-TOF mass spectrometry (Figure 5.1B) showed that the proteins have the expected molecular weights, and confirmed the monodispersity apparent from SDS-PAGE. N-terminal sequencing revealed the expected sequence (see Table 5.1) for both proteins (YVEFGLG), with no heterogeneity resulting from their proteolytic maturation in the host organism.

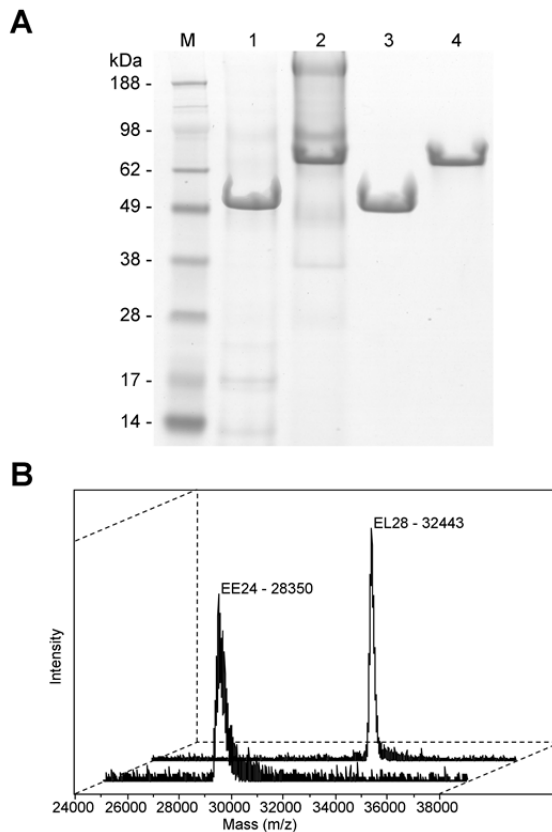


Figure 5.1. Production and purification of silk-like polymers. (A) SDS-PAGE of total extracellular protein and purified EE24 and EL28. Lane M, molecular weight marker; 1, EE24 culture supernatant (2 μ l); 2, EL28 culture supernatant (6 μ l); 3, purified EE24 (6 μ g); 4, purified EL28 (6 μ g). (B) MALDI-TOF mass spectra of purified EE24 and EL28.

To estimate the yield of EE24 and EL28 at the end of fermentation, an SDS-PAGE gel with final culture supernatants from duplicate fermentations and known quantities of the corresponding purified protein was analyzed by densitometric scanning. The yield calculated for EE24 was 2.9 ± 0.3 g/L of clarified broth, and for EL28 0.9 ± 0.2 g/L (\pm SD). With purified EL28 (as with both purified and unpurified

EE24), a linear relationship was found between the densitometric response and the amount of protein over a wide range, while for unpurified EL28, this range was very narrow. This indicates that the above-mentioned aggregation problems of EL28 culture supernatant during SDS-PAGE are not completely resolved by the presence of urea in the samples. It cannot be ruled out that even within the linear range used, part of the unpurified EL28 might be obscured, and the yield estimated for EL28 could be an underestimation. In any case, the range of roughly 1-3 g/L of secreted protein described here rates as high-yield expression for *P. pastoris* [21] or microbial hosts in general [22].

The differences encountered between EL28 and EE24 with respect to their aggregation behavior are noteworthy. While EE24 precipitated from the fermentation broth at pH 3.0, EL28 did not. In purified form, however, both proteins could be precipitated at low pH, suggesting that in the broth EL28 may have bound other components that prevent its self-aggregation. Moreover, the formation of such complexes may explain why only unpurified EL28 shows migration and detection problems in SDS-PAGE.

5.3.2. Secondary structure

Poly-[(GA)₃GE] self-organizes into crystalline stacks of antiparallel β -sheets when dried after gel formation in 70% (v/v) formic acid and methanol vapor diffusion [8]. We verified if EL28, which has alternate glutamic acids replaced by leucine, could also form such a structure under these conditions. Unfortunately, CD spectroscopy of the turbid dry films of EE24 and EL28 prepared under these conditions could not be used due to scattering artifacts. Therefore, the protein gels were dried onto an ATR crystal and ATR-FTIR was performed. Indeed, the spectra of both EE24 and EL28 (Figure 5.2) show amide I, II, and III vibrations around 1,622, 1,522, and 1,230 cm^{-1} , respectively, characteristic of the β -sheet conformation, while the weak amide I band at $\sim 1,697 \text{ cm}^{-1}$ is indicative of the antiparallel β -sheet conformation [6]. The amide I component at $\sim 1,652 \text{ cm}^{-1}$ may arise from a disordered conformation, chain folds, or a trace of silk I-like structures [6, 9], and has been shown to correlate with the level of hydration of the molecule [23]. This suggests that, given its lower absorbance at this frequency, a lower fraction of EL28 than of EE24 is present in disordered form. Accordingly, tentative fitting of the spectra to reference spectra and Lorentzian/Gaussian peak shapes proposed by

Venyaminov et al. [24] indicated that the contribution of β -sheet in EL28 was around 82-92% and approximately 1.3 times higher than that in EE24. This may result from packing constraints, or because in 70% (v/v) formic acid - 30% (v/v) water some hydrophobic effects play a role and a juxtaposition of dehydrated nonpolar leucine residues, as compared to H-bonding glutamic acid residues, promotes the transition to the crystalline state.

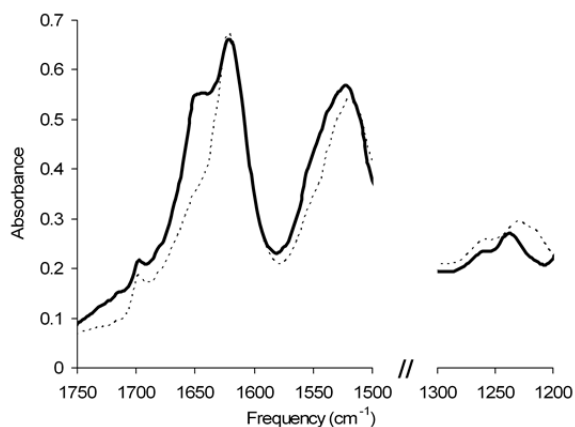


Figure 5.2. ATR-FTIR spectra of crystallized EE24 (solid line) and EL28 (broken line).

In view of the potential formulation of the protein in aqueous solutions or dispersions for coating applications, the behavior of EL28 under these conditions and after drying was analyzed next by CD spectroscopy and compared to EE24 (Figure 5.3). At low pH, both EL28 and EE24 aggregated and the resulting turbidity of the samples did not permit the registration of reliable UV-CD spectra. The spectra obtained in alkaline buffer suggest a mixture of random and extended (polyglycine II-like [25] or silk fibroin type III-like [26]) conformations, where the more positive relative ellipticity around 220 nm for EL28 indicates a higher contribution of extended conformation elements as compared to EE24. Drying the polymers dissolved in (volatile) alkaline buffer to a film resulted in uncommon spectra, having a minimum at about 208 nm and maxima around 195 and 220 nm. Lednev et al. [27] have recorded very similar spectra of the related protein poly-[(GA)₃GY(GA)₃GE(GA)₃GH(GA)₃GK]. Such spectra were also seen by Safar et al. [28], who showed that, depending on the conditions, two clearly different CD spectra could be recorded in dry poly-L-lysine films. One was a prototypical β -sheet spectrum, observed under conditions where FTIR also indicated mainly β -

sheet components. β -Sheet CD spectra of for example poly-L-lysine [29] and silk fibroin type II [7] have a single negative peak at ~ 217 nm and lack a positive band around 220–225 nm. The other spectrum observed by Safar et al. [28] in poly-L-lysine films was found under conditions where FTIR indicated non- β -sheet structures, and it closely resembled the unusual film spectra shown in Figure 5.3. As Crisma et al. [30] determined highly comparable spectra for the model peptides Z-Aib-L-Pro-Aib-L-Pro-Ome and Piv-L-Pro-Aib-NHMe when present in a mixed β -turn conformation, we tentatively ascribe the observed spectra of dried EE24 and EL28 to a conformation rich in β -turns. Although the exact types of β -turns in our material are not necessarily the same, it is interesting to note that also poly(Ala-Gly) in type II conformation [31] and *Bombyx mori* silk fibroin in type I conformation [32] are composed of repeated β -turns.

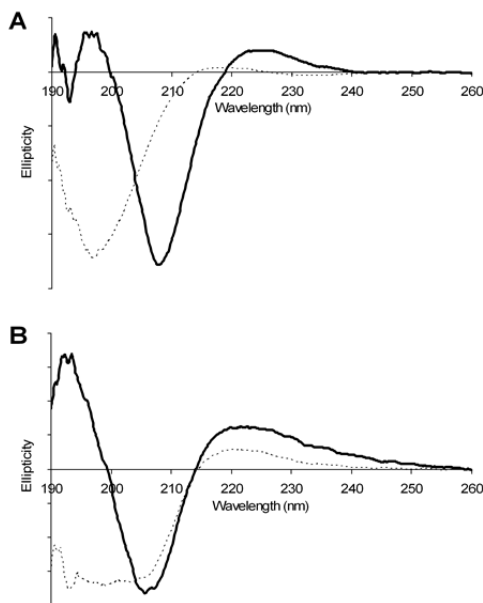


Figure 5.3. CD spectra of purified EE24 (A) and EL28 (B). Broken lines, aqueous solution at pH 9.0, solid lines, films dried from aqueous solution at pH 9.0. Spectra are normalized to similar peak intensities.

5.3.3. Supramolecular fibrils

Proteins, dissolved under crystallizing conditions, were adsorbed to surface-modified Si(100) wafers. The wafers were thoroughly rinsed and analyzed by AFM (Figure 5.4). Fibrils were observed with a length in the order of hundreds of

nanometers to several micrometers. To analyze the height and width distribution of the fibrils, 50 random small line profiles (each perpendicular to a fibril) were made, and the height relative to the baseline and the width of each peak was determined. The average fibril height was 2.4 ± 1.3 nm (\pm SD; range 0.7 to 5.9 nm) for EE24 and 2.7 ± 1.3 nm (\pm SD; range 1.0 to 7.6 nm) for EL28. Although these average heights are similar to the fold-to-fold distance of 2.8 nm calculated for poly-[(GA)₃GE] in antiparallel β -sheet conformation (with Glu present in the turns), or to the value of 3.6 nm obtained for this structure with x-ray diffraction [6], the thinnest EE24 and EL28 fibrils are considerably smaller. Topilina et al. [33] also observed lower fibril heights for the related protein poly-[(GA)₃GY(GA)₃GE(GA)₃GH(GA)₃GK], but they propose a different structure for that particular polymer. The apparent fibril width found for EE24 was 50 ± 11 nm (\pm SD; range 25-74 nm) and 49 ± 14 nm (\pm SD; range 23-87 nm) for EL28, although the true width of the fibrils will be smaller due to tip convolution (the tip used had a typical radius of 10 nm). For comparison, the expected length of a fully stretched monomolecular β -sheet structure (with 0.474 nm interstrand spacings, as determined for poly-[(GA)₃GE] [6]) is \sim 23 nm for EE24 and \sim 27 nm for EL28.

Krejchi et al. [6] observed the formation of needle-shaped lamellae upon crystallization of their poly-[(GA)₃GE]-type protein. Smeenk et al. [34] chemically crosslinked poly(ethylene glycol) (PEG) at its N- and C-termini, which resulted in the formation of fibrils upon crystallization. They found that the PEG groups prevent macroscopic crystallization, as a nonconjugated control showed only few fibrils due to the formation of larger aggregates. In this light it was surprising to find that EE24 and EL28 also formed fibrils. A possible explanation could be the fact that we used a 10,000-fold lower protein concentration during the crystallization (i.e., 1 mg/L), which may promote fibril formation over isotropic aggregation. Also the fact that we crystallized the proteins in the presence of the modified silicon wafers might be of influence. On the other hand, a remarkable morphological difference with the PEG-conjugated fibrils is evident. The fibers described here have a tortuous, crooked appearance, whereas the PEG-lined fibers [34] had a much stiffer appearance with a persistence length in the μ m range.

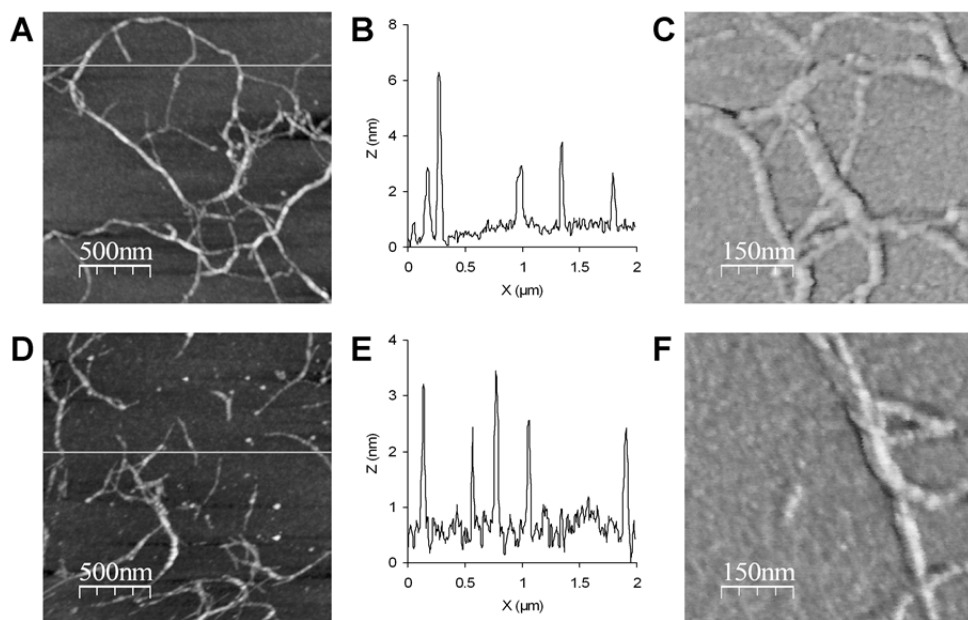


Figure 5.4. AFM of EE24 (A-C) and EL28 (D-F) fibrils deposited on modified Si(100) wafers under crystallizing conditions. (A, D) Topography image. (B, E) Topography profiles for the cross-sections indicated in A and D. (C, F) Detail of fibrils; shown are phase images to allow optimal resolution of fibril structure.

5.3.4. Amphiphilicity

To explore the anticipated amphiphilic character of EL28, as compared to EE24, the effect of both polymers on the surface tension of an air-water interface was analyzed, initially at pH 8.0 (Figure 5.5). Clearly, EL28 is more surface active than EE24. In view of the above CD results, both polymers are expected to be in random or extended conformation at high pH, unless other conformations are induced specifically at the air-water interface. The observed instantaneous effect of EL28 on the surface tension argues against such a conformational rearrangement. The amphiphilic character of EL28 under these conditions is therefore most likely inherent to its more hydrophobic composition, and not related to any specific tertiary or supramolecular structure. At pH 3.0, no effect of EL28 and EE24 was observed. Both polymers aggregated and, probably because of this, their availability at the air-water interface was negligible.

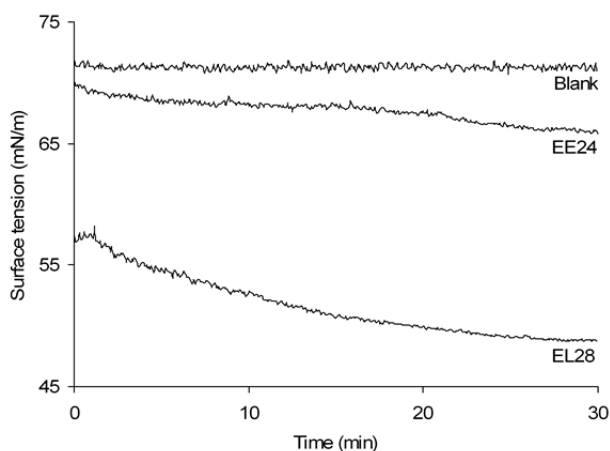


Figure 5.5. Drop tensiometry of EE24 and EL28 at 2 g/L and pH 8.0. For clarity, the results at pH 3.0 are not shown as they completely coincided with the blank.

To see if EL28 is capable of altering the surface hydrophobicity of solid substrates, a hydrophobic substrate (polystyrene) and a moderately hydrophilic substrate (aluminum) were submerged in solutions of each of the two polymers in 70% (v/v) formic acid, followed by thorough water washes to remove nonadhering material. As shown in Table 5.2, both substrates became more hydrophilic upon adsorption of EE24. The effect was small on polystyrene but very pronounced on aluminum, suggesting that the biopolymer may not have adhered well to the hydrophobic substrate. Interestingly, adsorbed EL28 rendered the hydrophobic substrate more hydrophilic and, conversely, the hydrophilic substrate markedly more hydrophobic. This finding confirms the amphiphilic nature of EL28, also in the solid state. In view of the ATR-FTIR results, β -sheet stacks of EL28 must have formed under the conditions applied. At present, it is unknown to what extent the leucine-containing β -sheet edges were all oriented to the same side of the stack, thus, possibly contributing to the amphiphilicity of EL28.

Table 5.2. Effect of protein on contact angle

Substrate	Mean contact angle (degrees; $n = 15$) ^a		
	Blank	EE24	EL28
Polystyrene	87 ± 4	83 ± 3 (6×10^{-3})	75 ± 7 (1×10^{-5})
Aluminum	58 ± 12	33 ± 8 (3×10^{-7})	78 ± 8 (9×10^{-6})

^a Standard deviations are indicated. Probability values of two-sided *t*-tests for lack of effect are shown in parentheses.

5.4. Conclusions

Efficient biosynthesis of the amphiphilic silk-like polymer [(GA)₃GE(GA)₃GL]₂₈ was demonstrated in the yeast *P. pastoris*. The polymer was monodisperse and was produced at the g/L level. The secreted polymer could be purified to near homogeneity making use of the much higher tendency of the contaminating host proteins to precipitate from solution under acidic conditions, after complete denaturation with formic acid.

The CD spectrum of the negatively charged polymer in aqueous solution at high pH indicates a dominance of random and extended helical (silk III-like) structures, whereas the spectrum of a coating prepared from such a solution shows an unusual combination of a minimum at about 208 nm and maxima around 195 and 220 nm, indicating a conformation rich in β -turns.

The polymer is pH-responsive, in that it self-aggregates in water at low pH. It forms β -sheets and fibril-like structures upon crystallization in 70% (v/v) formic acid under vapor diffusion of methanol, as shown by ATR-FTIR and AFM.

The amphiphilic nature of the polymer is clear from tensiometry, and sessile drop contact angle measurements. Upon adsorption it renders a hydrophobic solid substrate hydrophilic, and a hydrophilic substrate hydrophobic. Given the biocompatibility of silk-like materials, the polymer may be of interest for biomedical applications, such as the coating of surgical implants, or pH-responsive controlled drug release.

Acknowledgments

We thank Patrick van Doeveren and Louise van Zeeland for technical assistance, and Guus Frissen, Renate Ganzevles, Paulien Harmsen, Barend van Lagen, and Adrie Westphal for support in the performance of various analytical measurements. We are indebted to Arie van der Bent and Martien Cohen Stuart for helpful discussions, and to Gerrit Eggink for the realization and coordination of the B-Basic programme Performance Materials in which part of this work was carried out. All the aforementioned are from Wageningen UR. We thank Arend van Buul and Kasper Nørgaard (Institute for Molecules and Materials, Radboud University Nijmegen) for help with AFM imaging.

This research was financially supported, in part, by a grant of the Dutch Programme EET (Economy, Ecology and Technology), a joint initiative of the Ministries of Economic Affairs, Education, Culture, and Sciences, and of Housing, Spatial Planning, and the Environment (EETK98020 & EETK01031) and, in part, by The Netherlands Ministry of Economic Affairs and the B-Basic partner organizations (www.b-basic.nl) through B-Basic, a public-private NWO-ACTS programme (ACTS = Advanced Chemical Technologies for Sustainability).

References

- Heslot H (1998). *Biochimie* 80, 19-31.
- Van Hest JC, Tirrell DA (2001). *Chem Commun* 19, 1897-1904.
- Sanford K, Kumar M (2005). *Curr Opin Biotechnol* 16, 416-421.
- Woolfson DN, Ryadnov MG (2006). *Curr Opin Chem Biol* 10, 559-567.
- Zhao X, Zhang S (2007). *Macromol Biosci* 7, 13-22.
- Krejchi MT, Atkins ED, Waddon AJ, Fournier MJ, Mason TL, Tirrell DA (1994). *Science* 265, 1427-1432.
- Iizuka E, Yang JT (1968). *Biochemistry* 7, 2218-2228.
- Cantor EJ, Atkins ED, Cooper SJ, Fournier MJ, Mason TL, Tirrell DA (1997). *J Biochem* 122, 217-225.
- Parkhe AD, Cooper SJ, Atkins ED, Fournier MJ, Mason TL, Tirrell DA (1998). *Int J Biol Macromol* 23, 251-258.
- Chambers SP, Prior SE, Barstow DA, Minton NP (1988). *Gene* 68, 139-149.
- Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM (2000). *Biotechnol Bioeng* 70, 1-8.
- Katakura Y, Zhang W, Zhuang G, Omasa T, Kishimoto M, Goto Y, Suga K (1998). *J Ferm Bioeng* 86, 482-487.
- Wenzel T, Sparbier K, Mieruch T, Kostrzewa M (2006). *Rapid Commun Mass Spectrom* 20, 785-789.
- Benjamins J, Cagna A, Lucassen-Reynders EH (1996). *Colloids Surf, A* 114, 245-254.
- Sun QY, de Smet LC, van Lagen B, Wright A, Zuilhof H, Sudholter EJ (2004). *Angew Chem Int Ed Engl* 43, 1352-1355.
- Sun QY, de Smet LC, van Lagen B, Giesbers M, Thune PC, van Engelenburg J, de Wolf FA, Zuilhof H, Sudholter EJ (2005). *J Am Chem Soc* 127, 2514-2523.
- Faucheux A, Gouget-Laemmel AC, Henry de Villeneuve C, Boukherroub R, Ozanam F, Allongue P, Chazalviel JN (2006). *Langmuir* 22, 153-162.
- Horcas I, Fernandez R, Gomez-Rodriguez JM, Colchero J, Gomez-Herrero J, Baro AM (2007). *Rev Sci Instrum* 78, 013705.
- McGrath KP, Tirrell DA, Kawai M, Mason TL, Fournier MJ (1990). *Biotechnol Prog* 6, 188-192.
- Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F (1956). *Anal Chem* 28, 350-356.
- Cregg JM, Cereghino JL, Shi J, Higgins DR (2000). *Mol Biotechnol* 16, 23-52.
- Schmidt FR (2004). *Appl Microbiol Biotechnol* 65, 363-372.
- Chen CC, Krejchi MT, Tirrell DA, Hsu SL (1995). *Macromolecules* 28, 1464-1469.
- Venjaminov S, Kalnin NN (1990). *Biopolymers* 30, 1259-1271.
- Rippon WB, Walton AG (1971). *Biopolymers* 10, 1207-1212.
- Valluzzi R, Gido SP, Zhang W, Muller WS, Kaplan DL (1996). *Macromolecules* 29, 8606-8614.
- Lednev IK, Ermolenkov VV, Higashiya S, Popova LA, Topilina NI, Welch JT (2006). *Biophys J* 91, 3805-3818.
- Safar J, Roller PP, Ruben GC, Gajdusek DC, Gibbs CJ, Jr. (1993). *Biopolymers* 33, 1461-1476.
- Townend R, Kumosinski TF, Timasheff SN, Fasman GD, Davidson B (1966). *Biochem Biophys Res Commun* 23, 163-169.
- Crisma M, Fasman GD, Balaram H, Balaram P (1984). *Int J Pept Protein Res* 23, 411-419.
- Lotz B, Keith HD (1971). *J Mol Biol* 61, 201-215.
- Asakura T, Ohgo K, Komatsu K, Kanenari M, Okuyama K (2005). *Macromolecules* 38, 7397-7403.
- Topilina NI, Higashiya S, Rana N, Ermolenkov VV, Kossow C, Carlsen A, Ngo SC, Wells CC, Eisenbraun ET, Dunn KA, et al. (2006). *Biomacromolecules* 7, 1104-1111.
- Smeenk JM, Otten MB, Thies J, Tirrell DA, Stunnenberg HG, van Hest JC (2005). *Angew Chem Int Ed Engl* 44, 1968-1971.

Chapter 6

Precision gels from collagen-inspired triblock copolymers

Gelatin hydrogels find broad medical application. The current materials, however, are from animal sources, and their molecular structure and thermal properties cannot be controlled. This study describes recombinant gelatin-like polymers with a general design that inherently offers independent tuning of the cross-link density, melting temperature, and biocompatibility of the gel. The polymers contain small blocks with thermoreversible trimerization capacity and defined melting temperature, separated by hydrophilic nontrimerizing blocks defining the distance between the knot-forming domains. As an example, we report the secreted production in yeast at several g/L of two nonhydroxylated ~42 kDa triblock copolymers with terminal trimerizing blocks. Because only the end blocks formed cross-links, the molecular architecture of the gels is much more defined than that of traditional gelatins. The novel hydrogels had a ~37 °C melting temperature, and the dynamic elasticity was independent of the thermal history. The concept allows to produce custom-made precision gels for biomedical applications.

Published in modified form as: Werten MWT, Teles H, Moers APHA, Wolbert EJH, Sprakel J, Eggink G, de Wolf FA (2009). *Biomacromolecules* 10, 1106-1113.

6.1. Introduction

Stimulus-responsive polymer gels are promising materials for chemomechanical systems, sensors, surgery, regenerative medicine, and pharmaceuticals [1]. Biocompatibility is an essential requirement for medical gels [2] and, therefore, gelatin [3], which is derived from the natural matrix protein collagen [4], is ideally suited. To the benefit of millions of patients, several kilotons [5] of gelatin with a billion dollar market value are used on an annual basis and novel gelatin materials are being developed for bone [6], retinal [7], tissue-enhancing [8], and other implants, in vascular prostheses [9], sponge embolization therapy [10], scaffolds for tissue and cell culture [11, 12], medical glues [13], blood supplementation fluids [14, 15], drug carriers [11, 16], wound dressings [17], and vaccines [18].

For such applications, the animal origin of gelatin poses the risk of contamination with transmissible disease agents (e.g., prions) and immune responses [15, 18-20]. Furthermore, the extraction of gelatin from bones and hides causes uncontrolled degradation and results in a variable multitude of molecular species [3]. Recombinant production allows to circumvent these problems and, accordingly, several groups including our own have produced single-component natural gelatins (Chapter 2) [5, 20] and nonnatural gelatin-like designer proteins (Chapter 4) [21-23]. We realized secreted production of gelatins in the yeast *Pichia pastoris* at very high levels, as well as simple and scalable purification procedures (Chapters 2 and 4). In contrast to animal gelatins, these recombinant gelatins are naturally nongelling.

Nongelling recombinant gelatins are advantageous when used in blood supplementation fluids and vaccines, where the current use of animal gelatins requires suppression of unwanted gel formation by chemical modification [14, 15, 24] or degradation [18, 20], respectively. However, most medical applications of gelatin do require gel formation. The basis of the thermoreversible knots in a gelatin gel is the characteristic triple-helical structure of collagen [3]. It consists of repetitive Gly-Xaa-Yaa triplet repeats, commonly containing approximately 22% proline [3]. At this low proline content, triple helices are not thermally stable above 5-15 °C unless, as occurs in animals, the prolines in the Yaa position are post-translationally modified to 4-hydroxyprolines by the enzyme peptidyl-prolyl-4-hydroxylase (P4H) [25-27]. Because microbial hosts generally lack this enzyme,

recombinant production of thermally stable triple-helical collagen and gelatin requires coexpression of the two subunits of the mammalian P4H [25, 26, 28].

For applications such as controlled drug delivery, *independent* control over the melting temperature (T_m), cross-link density, and biocompatible properties of the gelatin gel would be highly advantageous. Thus far, however, no such material has been conceived. Manipulating the T_m of recombinant gelatins is already difficult in itself, as it would require control over the degree of hydroxylation. Proline-rich trimer-forming gelatin-like (collagenous) peptides with adjustable, length-dependent T_m have been synthesized chemically [29-31], but these are too short to form gels and cannot be produced in an economically feasible way. Polycondensated (and, thus, polydisperse) peptides long enough to form a gel can neither provide a homogeneous, controllable, and biomedically relevant T_m , nor a controllable cross-link density.

We describe here recombinant nonhydroxylated gelatin-like protein polymers capable of forming tuned precision gels. The block copolymers consist of short proline-rich triple helix-forming blocks that define the T_m , separated by long random coil spacer blocks that define the distance between the knot-forming domains in the molecules. The currently used spacers are extremely hydrophilic (Chapter 4). At least one of them is capable of attracting human cells in culture [32], and has favorable biocompatibility as compared to animal products in blood applications [15]. Gel-forming triblock protein polymers have been described before, making use of coiled-coils [33-38], elastin-mimetic motifs [39], or the fluorescent protein DsRed [37] as the cross-link-forming domain. The present work is the first report of such polymers where the knots are formed by (trimerizing) collagenous blocks. This choice of cross-linking domain offers thermoreversible melting and a defined junction multiplicity [40] (number of associative groups per knot) of exactly three. Notably, also the random coil spacer blocks are of a collagenous (gelatin-like) nature. The polymers, thus, truly represent a novel class of *gelatins*, offering defined functionality (number of associative groups per molecule) and tunable T_m , while expectedly retaining the hallmark biocompatibility of traditional collagenous materials.

6.2. Materials and methods

6.2.1. Construction of expression vectors

Triple helix-forming block **T** was prepared by PCR using the oligonucleotides T-FW and T-RV (Appendix Table A6.1). The ~0.1 kb product was cloned into vector pCR4-TOPO (Invitrogen), resulting in vector pCR4-TOPO-**T**. The previously described vector pMTL23-**P4** (Chapter 4) contains a gene encoding the custom-designed, highly hydrophilic 36.8 kDa collagenous protein '**P4**'. This vector was digested with *Dra*III (5' to the **P4** gene) and dephosphorylated. Vector pCR4-TOPO-**T** was digested with *Dra*III/*Van*91I. The released **T** block was ligated into the linearized and dephosphorylated vector, to yield vector pMTL23-**TP4**. This vector was then digested with *Van*91I (3' to the **P4** gene) and dephosphorylated, and a second *Dra*III/*Van*91I digested **T** block was inserted to yield vector pMTL23-**TP4T**. The **TP4T** gene was cloned into *P. pastoris* expression vector pPIC9 (Invitrogen) via *Xho*I/*Eco*RI.

The gene encoding **R4** was constructed by concatenating four copies of an **R** gene monomer. The monomeric gene was designed by randomizing the sequence of the **P** gene monomer (Chapter 4) in such a way that not every third residue of the encoded protein is glycine, preventing the formation of collagen triple helices while maintaining the same amino acid composition. The **R** gene monomer was constructed by overlap extension PCR [41] using oligonucleotides RA-FW and RA-RV for the 5' half of the gene, and oligonucleotides RB-FW and RB-RV for the 3' half (Appendix Table A6.1). The products of these reactions were combined by overlap extension PCR to generate the entire **R** gene monomer. The monomer was cloned into vector pMTL23 [42] via *Xho*I/*Eco*RI and multimerized via *Dra*III/*Van*91I, as described previously for **P4** (Chapter 4), resulting in vector pMTL23-**R4**. The **R4** fragment was cloned into expression vector pPIC9 via *Xho*I/*Eco*RI. The construction of pMTL23-**TR4T** and subsequent subcloning into pPIC9 was analogous to the procedures described for **TP4T**.

The DNA sequences (and translated amino acid sequences) of the *Xho*I/*Eco*RI fragments encoding **P4** (Chapter 4), **R4**, **TP4T**, and **TR4T** have been deposited in GenBank under accession numbers EU834225-EU834228. Although, for clarity, the **T** blocks are referred to as (Pro-Gly-Pro)₉ in the main text, the cloning procedure results in a Gly-Pro-Pro-Gly-Ala extension at the N-terminus, and an Ala-Gly-Gly extension at the C-terminus.

6.2.2. Transformation of *P. pastoris*

Expression vectors were linearized with *SalI* to promote integration at the *his4* locus rather than the *AOX1* locus, thus enabling normal growth on methanol [43]. Transformation of *P. pastoris his4* strain GS115 [43] and selection of transformants was as described previously (Chapter 2).

6.2.3. Fermentation of *P. pastoris*

Fed-batch fermentations were performed in 2.5-L Bioflo 3000 fermentors (New Brunswick Scientific), essentially as described by Zhang et al. [44]. Minimal basal salts medium [44] was used and no protease-inhibiting supplements were added. The pH was maintained at 3.0 throughout the fermentation by addition of ammonium hydroxide as base. The methanol fed-batch phase for protein production lasted two to three days. A homemade semiconductor gas sensor-controller, similar to that described by Katakura et al. [45] was used to monitor the methanol level in the off-gas and to maintain a constant level of ~0.2% (w/v) methanol in the broth. At the end of the fermentation, the cells were separated from the broth by centrifugation for 10 min at $10,000 \times g$ (RT) in an SLA-3000 rotor (Sorvall), and the supernatant was microfiltered.

6.2.4. Protein purification

All centrifugation steps were performed in an SLA-1500 or SLA-3000 rotor (Sorvall) for 30 min at $20,000 \times g$, and resuspension of protein pellets was always in Milli-Q water at 65 °C.

As a precaution, the cellfree broth was heated for 30 min at 65 °C to melt possible gel structures formed by the recombinant protein. The pH was raised to 8.0 by addition of sodium hydroxide to allow precipitation of medium salts by centrifugation (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 (4 °C). This precipitation procedure was repeated once, and 50 mM sodium chloride and 40% (v/v) acetone were added to the final resuspended pellet. After centrifugation (4 °C), the pellet was discarded and acetone was added to the supernatant up to 80% (v/v). The protein pellet obtained

after centrifugation was air-dried, resuspended, and desalted by extensive dialysis against Milli-Q water. The final product was lyophilized.

6.2.5. SDS-PAGE and N-terminal protein sequencing

The NuPAGE Novex system (Invitrogen) was used for SDS-PAGE, with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 prestained molecular mass markers. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). Blotting of proteins for N-terminal sequencing by Edman degradation was as described previously (Chapter 2). Protein sequencing was performed by Midwest Analytical (St. Louis, MO).

6.2.6. Mass spectrometry

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry was performed using an Ultraflex mass spectrometer (Bruker). Samples were prepared by the dried droplet method on a 600 μm AnchorChip target (Bruker), using 5 mg/mL 2,5-dihydroxyacetophenone, 1.5 mg/mL diammonium hydrogen citrate, 25% (v/v) ethanol, and 1% (v/v) trifluoroacetic acid as matrix. Measurements (50 shots at 20 Hz) were made in the positive, linear mode, with the following parameters: ion source 1, 20.0 kV; ion source 2, 18.45 kV; lens, 5.0 kV; pulsed ion extraction, 550 ns. Protein Calibration Standard II (Bruker) was used for external calibration.

6.2.7. Circular dichroism spectroscopy

Proteins were dissolved at 100 mg/mL in 0.2 M sodium phosphate pH 3.0 and heated to 45 °C. Gels were allowed to form overnight inside the cuvette (path length 0.01 mm) at room temperature. Circular dichroism (CD) spectra were recorded from 260 to 210 nm using a J-715 spectropolarimeter (Jasco) set at the desired temperature. Spectra were obtained as the average of ten consecutive scans using a scanning speed of 100 nm/min at a resolution of 0.2 nm. Melting curves from 20 to 65 °C were derived by increasing the temperature in steps of 5 or 10 °C, followed by equilibration at each temperature for 10 min prior to scanning. To allow recording of full spectra from 260 to 195 nm, measurements were also performed at 0.2 mg/mL in a cuvette with 1 mm path length.

6.2.8. Differential scanning calorimetry

Degassed 0.5 mL protein solutions (0.4–2.2 mM in 0.2 M sodium phosphate pH 3.0) were loaded into a MicroCal VP-DSC calorimeter at 45 °C. Protein solutions were equilibrated for 10 h at 20 °C to allow complete triple helix formation. For melting curves, the temperature was increased from 5 to 65 °C at a scan rate of 15 °C/h.

6.2.9. Rheometry

An Anton Paar Physica MCR301 rheometer, equipped with a stainless steel CC17 Couette geometry, gap size, 0.71 mm; bob radius, 8.3 mm; and sample volume 3 mL, was operated at an angular frequency of 1 Hz and a strain of 0.1%. Protein solutions (2.2 mM in 0.2 M sodium phosphate, pH 3.0) were heated to 45 °C, introduced in the geometry, and subsequently quenched to 20 °C to induce gel formation. Measurements were done at 20 and 5 °C. Melting of the gel was studied by increasing the temperature from 20 to 65 °C in steps of 2.5 °C that lasted 10 min each, and these data were smoothed using a moving average filter over 21 data points (10 min).

6.2.10. Erosion and release studies

The erosion rate of **TR4T** gels was measured using a method similar to that reported previously [36]. **TR4T** was dissolved at 200 mg/mL in PBS (10 mM sodium phosphate, 150 mM sodium chloride, pH 7.4) at 60 °C. Gel layers of 1 mm thickness were made by pipetting 44 µL of the solution into cylindrical polypropylene containers of 7.5 mm diameter and 5 mm height. The containers were sealed in plastic foil to prevent drying and the gels were allowed to set overnight at room temperature. The containers were unpacked and submerged in 2.2 mL of PBS and incubated at either 20 or 37 °C under gentle tilting at 20 rpm with an amplitude of ~6 mm. Erosion of **TR4T** was determined by monitoring the protein concentration in the supernatant by measuring the absorbance at 230 nm (for lack of absorbance at 280 nm owing to the absence of aromatic amino acids in **TR4T**) and comparison with a calibration curve of **TR4T**. Enzyme release from **TR4T** gels was studied similarly, using gels where 10 mg/mL lysozyme was added to the molten gel prior to preparing the gel layers. The release of lysozyme was quantified by determining the tryptophan fluorescence (Ex = 295 nm, Em = 240 nm)

of the supernatant over time against a calibration curve of lysozyme (eroded **TR4T** did not cause significant background fluorescence as it does not contain aromatic residues). Both the erosion and the release experiments were performed in triplicate, and absorbance and fluorescence measurements were done using a Safire microplate reader (Tecan).

6.3. Results

6.3.1. Rationale for nonhydroxylated precision gels

As an example of the proposed concept, two ABA type triblock copolymers designated **TP4T** and **TR4T** are described, but many other functional multiblock configurations are conceivable. The amino acid sequences of all protein polymers described in this work are available through GenBank (accession numbers ACF33476-ACF33479).

The triple helix-forming **T** blocks at both ends of the triblock copolymers consist of (Pro-Gly-Pro)_{*n*} homopolymeric stretches (Figure 6.1A). On the basis of helix melting studies with chemically synthesized (Pro-Pro-Gly)_{*n*} [30, 31, 46], the length of *n* was tentatively chosen to be nine to provide a melting point in a biomedically relevant range.

The middle section of the molecule acts as a hydrophilic spacer (Figure 6.1A) and consists of either of two random coiled block variants. The first is **P4**, a synthetic gelatin-like molecule previously developed by our group (Chapter 4). It is extremely hydrophilic, acts as a cytophilic protein in human cell culture [32], and shows outstanding biocompatibility as a plasma expander [15]. Despite **P4**'s collagenous primary structure, it does not form detectable triple helices at 4 °C because of the absence of 4-hydroxyprolines (Chapter 4). It cannot, however, be excluded beforehand that **P4** might play a minor role in network formation in the presence of the proline-rich trimer-forming **T** blocks. Therefore, a second type of mid-block, **R4**, was constructed. It has the same amino acid composition as **P4**, but its protein sequence is quasi-random in that it does not have glycine as every third residue. Thus, **R4** by definition cannot form triple helices.

Because both the end blocks and mid blocks have a defined length and composition, and because trimerization of the mid blocks is unlikely (**P4**) or impossible (**R4**), the resulting gels are expected to contain only cross-links made-up of (Pro-Gly-Pro)₉ (Figure 6.1B,C). This is in stark contrast with traditional gels

prepared from animal gelatin, where all molecules have a different makeup, and the entire chain engages in network formation.

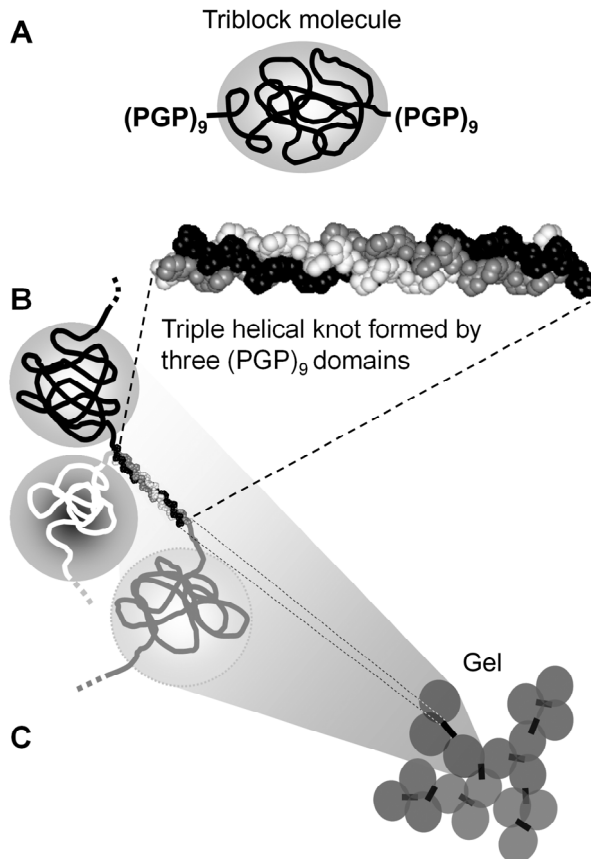


Figure 6.1. Schematic representation of gel formation by triblock copolymers. (A) Triblock copolymer with $(Pro-Gly-Pro)_9$ end blocks and random coil mid block. (B) Detail of a $(Pro-Gly-Pro)_9$ end block in triple-helical form and of a trimeric knot in the gel. Three mid blocks originate from the knot, each of which also has a $(Pro-Gly-Pro)_9$ block at its other end (not shown for clarity). (C) Gel consisting of triblock copolymers. The circles represent the time-averaged space occupied by the random coiled mid blocks. The dark bars represent a random subset of trimeric knots formed between three neighboring chains.

6.3.2. Biosynthesis of triblock copolymers

P. pastoris strains expressing TR4T and TP4T as extracellular proteins were constructed and grown in bioreactors. Culture supernatants were analyzed by SDS-PAGE (Figure 6.2A). The proteins were purified from the cell-free broth essentially by differential ammonium sulfate precipitation, similarly to the purification of P4 (Chapter 4). The purity of the proteins was estimated to be at

least 99%, based on amino acid analysis and subsequent linear least-squares fitting to the observed data of (1) the theoretical composition of the respective pure protein and (2) the composition determined for host-derived proteins present in the medium. Purified **TR4T** and **TP4T** migrated as single bands in SDS-PAGE (Figure 6.2A), indicating their purity and intactness.

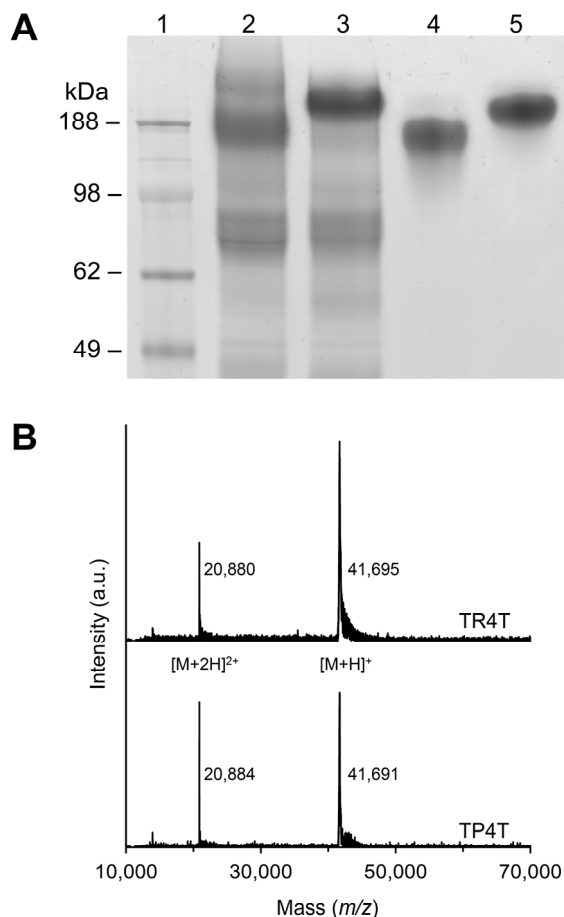


Figure 6.2. Production of protein polymers. (A) SDS-PAGE: lane 1, molecular weight marker; lane 2, culture supernatant of **TR4T**; lane 3, culture supernatant of **TP4T**; lane 4, purified **TR4T**; lane 5, purified **TP4T**. For culture supernatants, 5 μ L was loaded, and for purified proteins, \sim 20 μ g. A full-length gel (including **R4** and **P4** control proteins) is shown in Appendix Figure A6.1. (B) MALDI of purified **TR4T** and **TP4T**. Singly and doubly charged molecular ions are indicated.

MALDI mass spectrometry confirms this observation and shows that the molecular weight of the proteins is within experimental error of the expected value of 41,741 Da (Figure 6.2B). Clearly, the proteins migrate aberrantly in SDS-PAGE,

as was demonstrated previously for **P4** (Chapter 4). N-terminal sequencing of the bands separated by SDS-PAGE further confirmed the identity of the products. A minor fraction (~15%) of the molecules had a single N-terminal Glu-Ala extension, which is known to occasionally occur because of partial processing of the α -factor prepro secretory signal by the *P. pastoris* dipeptidylaminopeptidase (Chapter 2). Judging from the intensity of the bands in SDS-PAGE, the volumetric productivity of **TR4T** and **TP4T** appears comparable to that of **P4**, which is produced at 3-6 g/L of clarified broth (Chapter 4).

6.3.3. Trimer-forming functionality is restricted to proline-rich end blocks.

The occurrence of trimeric knots is essential for the generation of a thermoreversible gelatin network. In the triblock design, only the proline-rich **T** blocks at both ends of the triblock copolymers fulfill this role, and the mid blocks act as spacers. To confirm that the combined blocks in the triblock copolymer retain these distinct functions, UV-CD spectroscopy was performed at a protein concentration of 100 mg/mL, where the chance of intermolecular interactions is high. Both **TR4T** and **TP4T** formed optically clear dispersions and yielded very similar spectra, with a consistently negative ellipticity below 260 nm and a negative peak around 205-200 nm, reflecting mainly random structure [29, 47]. Spectra obtained with **TR4T** at 20 and 65 °C are shown in Figure 6.3 (see Appendix Figure A6.2 for measurements at low concentration that permitted registration of spectra down to 195 nm). The predominance of random structure is in agreement with the random coil propensity of the large mid block, making up 88% of the molecular length. To investigate whether the secondary structure of free mid and end blocks can fully account for that of **TR4T** and **TP4T**, or whether the mid and end blocks influence each other's structure when coupled, the triblock spectra were compared to those of the individual blocks. Linear least-squares fitting of spectra of isolated **R4** (Appendix Figure A6.2) and Ac-(Gly-Pro-Pro)₁₀-NH₂ in triple-helical or heat-denatured form [48] (data kindly provided by H.P. Bächinger) to the spectra of **TR4T** resulted in a close fit (Figure 6.3). A poly-(Gly-Pro-Pro) content of ~9% was obtained both at 20 and 65 °C, in reasonable agreement with the 12% length ratio of the end blocks relative to the entire block copolymer. This shows that the end and mid blocks in **TR4T** indeed assume the same secondary structure as the isolated components and, thus, that at 20 °C, the end blocks, like Ac-(Gly-

Pro-Pro)₁₀-NH₂ [48], are largely in helical conformation. The spectral contribution of the end blocks can be appreciated by subtraction of the fitted contribution of **R4** from the spectra of **TR4T** at 20 and 65 °C (see inset Figure 6.3). Similarly, the secondary structure of **TP4T** could be fully accounted for by that of isolated **P4** and Ac-(Gly-Pro-Pro)₁₀-NH₂ (Appendix Figure A6.2), and its poly-(Gly-Pro-Pro) content of ~14% closely matched the expected value. These results show that, in both polymers, selectively, the proline-rich end blocks behave like Ac-(Gly-Pro-Pro)₁₀-NH₂ [48] and, thus, take part in helix formation.

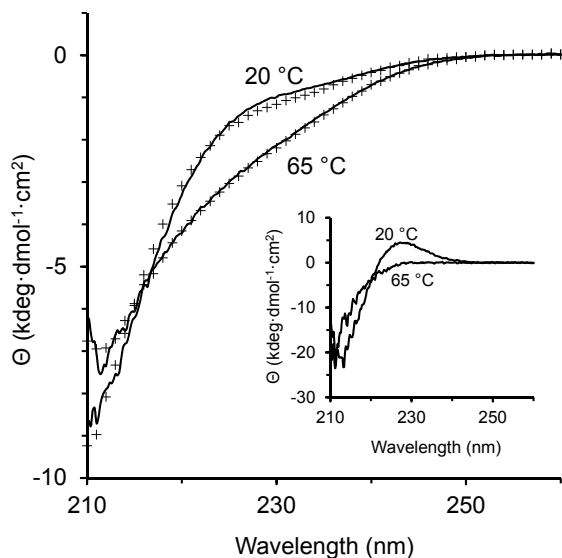


Figure 6.3. Study of the secondary structure at 20 and 65 °C, as reflected by CD. Spectra of **TR4T** (100 mg/mL) are shown, along with fitted spectra (+) composed of contributions of **R4** and (Gly-Pro-Pro)_n in isolated form. Inset: spectra of the proline-rich end blocks, as deduced by subtracting the fitted contribution of **R4** from the spectra of **TR4T**.

6.3.4. Melting behavior of triple-helical end blocks

The average ellipticity from 230 to 225 nm (at 100 mg/mL), as an indicator of the presence of 3₁ helices [47], was determined by CD spectroscopy at various temperatures between 20 and 65 °C. Figure 6.4A shows the thermal response of the **R4** (monoblock) polymer, which, in view of the above-described absence of crosstalk between end and mid blocks, reflects the properties of the mid block in **TR4T** (see Appendix Figure A6.3 for the melting curve of **P4**). Its near-linear temperature dependence is characteristic of a gradual repopulation of local

conformations of individual single chain sections [47, 49, 50]. The behavior of the end blocks can be deduced by subtracting the contribution of the free **R4** monoblock polymer, as determined by component fitting (Figure 6.3), from the melting curve of **TR4T**. The observed sigmoidal temperature dependence (Figure 6.4A; see Appendix Figure A6.3 for the melting curve deduced for the end blocks in **TP4T**), is characteristic of cooperative unfolding of triple helices [29, 31, 46]. The apparent T_m for the proline-rich end blocks deduced from the inflection point of these curves was 47 °C for both **TR4T** and **TP4T**. The corresponding van't Hoff enthalpies were calculated to be 188 and 223 kJ/mol, respectively. Based on the calorimetric enthalpy obtained for (Pro-Pro-Gly)₁₀ trimers by Frank et al. [31], the melting enthalpy expected for a (Pro-Gly-Pro)₉ end block trimer is 223 kJ/mol. The van't Hoff enthalpies obtained for **TR4T** and **TP4T** are 0.85 and 1.0 times this value, respectively. This shows that the cooperatively melting unit is approximately equal to the end blocks, providing further evidence that triple helix formation occurs specifically through the proline-rich end blocks.

The thermal denaturation of the polymers between 5 and 65 °C was also studied by differential scanning calorimetry (DSC; see Figure 6.4B for **TR4T** and Appendix Figure A6.3 for **TP4T**). While **R4** and **P4** polymer solutions did not show endothermic transitions, and thus, no triple helices were present, **TR4T** and **TP4T** had ΔC_p maxima at 41 ± 0.5 °C (\pm SD, $n = 3$) and 42 ± 0.4 °C (\pm SD, $n = 3$), respectively. This confirms the conclusion drawn from CD that selectively the end blocks in the triblock copolymers are responsible for trimerization. The somewhat higher T_m values obtained with CD probably result from a systematic offset from the true temperature, owing to inefficient heat transfer in the spring-loaded CD cuvette assembly. The enthalpies of **TR4T** and **TP4T** were 154 ± 56 kJ/mol (\pm SD, $n = 3$) and 140 ± 8 kJ/mol (\pm SD, $n = 3$), respectively, in good agreement with the expected enthalpy of about 148 kJ/mol per set of two (Pro-Gly-Pro)₉ end blocks (i.e., two-thirds of the value of an end block trimer). This indicates once more that nearly all (Pro-Gly-Pro)₉ blocks in **TR4T** and **TP4T** had formed triple helices. The shape of the thermograms of **TR4T** and **TP4T** reveals van't Hoff enthalpies of 260 ± 39 kJ/mol (\pm SD, $n = 3$) and 321 ± 79 kJ/mol (\pm SD, $n = 3$), respectively. This corresponds to, respectively, 1.2 ± 0.2 (\pm SD, $n = 3$) and 1.4 ± 0.4 (\pm SD, $n = 3$) times the calorimetric enthalpy of a (Pro-Gly-Pro)₉ trimer. Therefore, as was concluded from CD, the cooperatively melting unit roughly corresponds to the end blocks. A thermal

denaturation cycle repeated on the same samples (after 10 h at 20 °C) resulted in calorimetric and van't Hoff enthalpies that deviated less than 3 and 6%, respectively, from the above values obtained with freshly prepared solutions, showing complete reversibility of trimer formation.

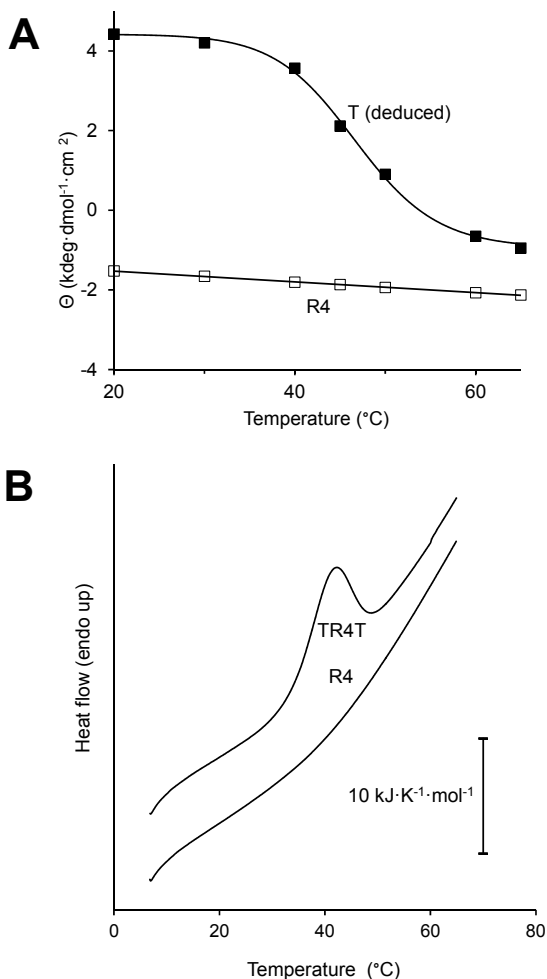


Figure 6.4. Thermal denaturation as reflected by (A) the average ellipticity from 230 to 225 nm or by (B) DSC. (A) Temperature profile of **R4** produced as a separate monoblock polymer and of the proline-rich end blocks as deduced by subtracting the fitted contribution of **R4** from the melting curve of **TR4T**. (B) DSC thermograms of **R4** (1.4 mM) and **TR4T** (1.1 mM), recorded after equilibration for 10 h at 20 °C.

6.3.5. Triblock copolymers form gels

Initial observations showed that **TR4T** and **TP4T**, unlike **R4** and **P4** (Chapter 4), formed optically clear gels when preheated solutions were cooled to room

temperature. For example, at 7.8 mM equiv of trimer-forming end block, a stainless steel bead with a diameter of 4.8 mm and a weight of 0.45 g, was supported by a **TR4T** gel during at least several hours until the gel was molten by heating.

Gel formation was further investigated by means of dynamic rheology, at a protein concentration of 2.2 mM, corresponding to 4.4 mM or ~10 mg/mL equiv of (Pro-Gly-Pro)₉ end blocks (see Figure 6.5 for **TR4T** and Appendix Figure A6.4 for **TP4T**). Upon quenching from 45 to 20 °C, the storage modulus (dynamic elasticity) steeply increased within minutes, while the loss modulus (dynamic loss) remained low. After roughly five hours at 20 °C, the storage modulus of **TR4T** and **TP4T** reached a final value of approximately 380 and 190 Pa, respectively. This range is in good agreement with values obtained in a variety of animal gelatins at a corresponding triple helix concentration [51], which confirms that most of the end blocks took part in helix-forming knots. No further changes of the storage modulus were observed during the 28-hour measurement. The gel strength increased 5 (**TR4T**) to 25% (**TP4T**) upon cooling from 20 to 5 °C.

The melting behavior of the **TR4T** and **TP4T** gels was studied by increasing the temperature from 20 to 65 °C and following the decrease of the storage modulus (see inset of Figure 6.5 for **TR4T** and Appendix Figure A6.4 for **TP4T**). The apparent inflection point in the melting curves of both gels is at ~37 °C, which is below the T_m values obtained with DSC. This is in agreement with the expectation that melting of the gels occurs before the majority of the trimeric knots unfold.

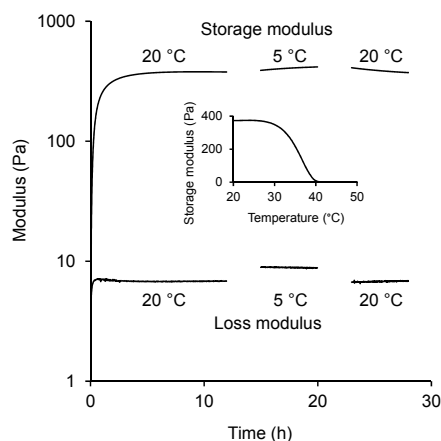


Figure 6.5. Characterization of a **TR4T** gel by dynamic rheology. The time-resolved storage and loss modulus were measured at the temperatures indicated, at a protein concentration of 2.2 mM. For each new temperature, the sample was allowed to equilibrate for three hours in the geometry before monitoring. Inset: storage modulus as a function of temperature.

6.3.6. Enzyme release and hydrogel erosion

To get an impression of the suitability of the present gels for release of, for example, protein drugs, the release from a **TR4T** gel (200 mg/mL) of the 21 kDa model protein lysozyme, and the erosion of the gel itself (both in a 50-fold excess volume of PBS buffer) were investigated in a pilot study (Figure 6.6). At 20 °C, the release of lysozyme was higher than the erosion of the hydrogel, indicating that release of the enzyme was caused both by erosion of the gel and by diffusion of the enzyme through the polymer network. Assuming erosion at the surface of the gel, the erosion rate at this temperature was $4.6 \times 10^{-3} \text{ mg cm}^{-2} \text{ min}^{-1}$, as calculated from the slope of the linear regression line indicated in Figure 6.6. At physiological temperature the release of lysozyme was virtually equal to the erosion of the gel. This bulk erosion concurs with the observation that **TR4T** gels (200 mg/mL) at this temperature exist in a semimolten state, in accordance with the apparent T_m of ~ 37 °C as determined by rheology (at 100 mg/mL).

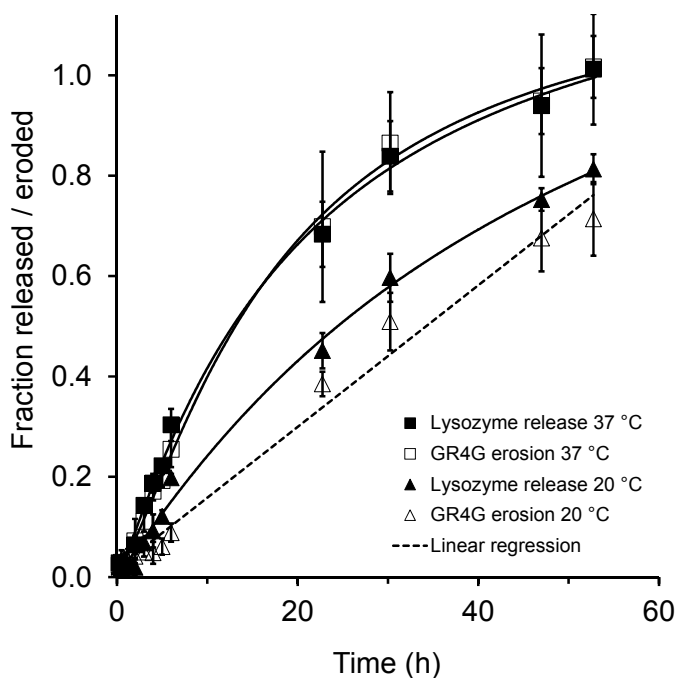


Figure 6.6. Erosion of a **TR4T** gel (200 mg/mL), and release of lysozyme (10 mg/mL) at 20 and 37 °C. Error bars represent SD ($n = 3$).

6.4. Discussion

Gelatins are important biomedical materials but currently have the disadvantage of being animal-derived, polydisperse, and chemically ill-defined. Furthermore, it is not possible to independently control the T_m , the distance between domains involved in triple-helices (forming the thermoreversible knots in the gel), and the biocompatibility. In this study, we met these challenges through the development and efficient production of recombinant gelatin-like block copolymers.

A prerequisite for the formation of a gelatin network is the formation of triple-helical knots. The proline-rich end blocks in our triblock design serve this function. The mid blocks represent random coil spacer regions that cannot readily (**P4**), or at all (**R4**), participate in trimer formation. We have shown here that the (Pro-Gly-Pro)₉ end blocks in **TP4T** and **TR4T** are both exclusively and near-quantitatively involved in trimerization, as both the calorimetric enthalpies obtained with DSC and the van't Hoff enthalpies obtained with DSC and CD were in good agreement with values expected for free (Pro-Gly-Pro)₉. The cross-links formed by the end blocks, thus, essentially have a fixed length and amino acid composition and, accordingly, a fixed T_m . Given this concurrence in enthalpic values the alignment of the end blocks is probably even largely in register, as is known for chemosynthetic (Pro-Pro-Gly)₁₀ peptides in trimeric form [52]. In any case, the triblock gelatins are highly defined in that the cross-links are formed strictly by the end blocks. This is in striking contrast with animal gelatin molecules, which form triple helices along their entire chain. As a consequence, the cross-links in animal gelatin have a wide variety of lengths, compositions, and melting temperatures [51]. The fact that traditional gelatins comprise a wide range of different molecules further increases their undefined nature.

Because animal gelatin gels are such highly complex systems, they have very broad melting trajectories. During aging, the storage modulus continues to increase in proportion to the logarithm of time [51, 53]. Furthermore, these gels show a pronounced dependence on the thermal history [51, 53]. For example, Te Nijenhuis [53] found that the stiffness of animal gelatins preaged at 17.4 °C strongly increased upon lowering the temperature, as additional helices of low thermal stability were formed. Our triblock gelatins behave very differently. Only the end blocks in **TP4T** and **TR4T** gels form helices, and these have distinct thermal

stability. Consequently, the storage modulus reached a plateau within only a few hours and the stiffness of the gels after preaging at 20 °C hardly increased upon cooling to 5 °C. The T_m of the gels is defined and can be directly controlled by varying the length of the terminal (Pro-Gly-Pro) $_n$ stretches, as known from studies with chemically synthesized peptides [30, 31, 46, 54]. Indeed, preliminary DSC experiments showed that, under identical conditions, triblock copolymers with (Pro-Gly-Pro) $_{16}$ end blocks have a much higher T_m (~57 °C) than **TR4T** (~41 °C).

Although the T_m values of **TR4T** and **TP4T** may appear rather high as compared to published T_m values for chemosynthetic (Gly-Pro-Pro) $_n$ [30, 46], the T_m is known to depend highly on the peptide concentration [30, 31]. Accordingly, the relatively high concentrations used here are probably the major determinant for the T_m values found. The minor extensions of the end blocks resulting from the cloning procedure (see Experimental Section), or the possible involvement in trimerization of a few residues of the mid blocks (immediately adjacent to the end blocks) may provide an additional increase in T_m . These effects will, however, be limited given the good agreement between the enthalpies obtained and those expected for (Pro-Gly-Pro) $_9$. Lastly, it is probable that the end blocks as part of the triblock copolymer have a higher T_m than that of free (Pro-Gly-Pro) $_9$ because of decreased entropy and because the end blocks each only have one free terminus. Although the T_m may, thus, not be determined *only* by the length of the end blocks, the *variability* in T_m is. Besides using the length of the end blocks, the T_m could also be controlled by changing their amino acid composition. The relation between amino acid sequence and thermal stability of the triple helix has been described in detail [54].

The length of both the knot-forming and spacer regions in the triblock copolymers are exactly defined. In trimeric form the length of the knots is around 8 nm, as calculated from the known size of (Pro-Pro-Gly) $_{10}$ trimers [55]. On the basis of data from a set of 33 unfolded proteins [56], the radius of gyration for the present mid blocks is expected to be around 7 nm, and the root-mean-square distance between the end blocks on opposite sides of the molecule is around 18 nm (or slightly bigger owing to the tendency of moderately proline-rich sequences to form transient local polyproline-II-like structure elements). In a homogeneous gel, this size range seems suitable for slow release of, for example, protein drugs. The trial study on release of lysozyme indeed suggests these gels could be further

developed for such applications. For example, longer or shorter mid blocks could be used to alter the distance between the cross-links at both ends of the polymer and, thus, indirectly, the cross-link density and release characteristics of the gel.

The erosion rate of the gels may also need to be controlled. Coiled-coil-based triblock polymer gels in excess fluid have been shown to erode quickly at a constant rate from the surface, owing to loop formation between the end blocks and the release of disengaged clusters from the network [36]. In principle, loops can also form in our system, depending on the length of the mid block. It took two days for ~70% of the 1 mm thick **TR4T** gel to erode at 20 °C. Although this seems fairly stable, applications requiring release over extended periods would still require chemical or enzymatic cross-linking. Shen et al. [36] showed it was possible to reduce the erosion rate of coiled-coil-based triblock gels by more than one hundredfold by employing dissimilar end blocks that do not readily associate with each other and, thus, cannot form loops. A means of achieving this in our collagen-based system may possibly be found in the work of Gauba and Hartgerink, who demonstrated preferential self-assembly of heterotrimers over homotrimers upon mixing oppositely charged chemosynthetic collagenous peptides [57, 58]. To provide the desired charge, Pro or Hyp residues in each peptide were replaced with either Asp, Glu, Lys or Arg. Applying this approach to our system, triblocks with positively charged end blocks could be mixed with triblocks that have negatively charged end blocks. Loop formation would be prevented by mutual charge repulsion of the end blocks, while a gel could still form through the formation of heterotrimers between oppositely charged end blocks of different molecules. Moreover, different types of mid blocks could be combined with each type of end block, allowing a unique level of control over the molecular architecture of the gels. This could be of interest particularly for biomedical applications, as there is considerable freedom to insert or, conversely, avoid specific cell-binding, protein-binding, protease-sensitive, or bioactive amino acid sequences in the mid blocks.

6.5. Conclusions

Triblock copolymers consisting of terminal (Pro-Gly-Pro)₉ blocks and a random coil spacer have been efficiently produced as extracellular proteins in recombinant yeast. Specifically the end blocks form collagen-like triple helices, contrasting

traditional gelatins where cross-links form along the entire chain of each molecule. Unlike animal gelatins, the dynamic elasticity of the gels is independent of the thermal history, and the T_m is inherently defined by the length of the end blocks. Together, these characteristics offer novel possibilities for controlled drug release and other biomedical applications.

Acknowledgments

We thank Jeroen Bastiaans, Claudia Patrignani, Bas Simon, Patrick van Doeveren, Martijn Hendriksen, Adrie Westphal, Paulina Skrzyszewska, and Jacqueline Donkers (all from Wageningen UR) for technical assistance, Frans van Hoesel (Centre for High Performance Computing and Visualisation, University of Groningen) for generating the 3D impression of the gel structure, and Hans Peter Bächinger (Oregon Health and Science University, Department of Biochemistry and Molecular Biology) for kindly providing CD spectral data of Ac-(Gly-Pro-Pro)₁₀-NH₂.

This research was financially supported, in part, by The Netherlands Ministry of Economic Affairs and the B-Basic partner organizations (www.b-basic.nl) through B-Basic, a public-private NWO-ACTS programme (ACTS = Advanced Chemical Technologies for Sustainability).

References

1. Sangeetha NM, Maitra U (2005). *Chem Soc Rev* 34, 821-836.
2. Hutmacher D, Hurzeler MB, Schliephake H (1996). *Int J Oral Maxillofac Implants* 11, 667-678.
3. Asghar A, Henrickson RL (1982). *Adv Food Res* 28, 231-372.
4. Myllyharju J, Kivirikko KI (2001). *Ann Med* 33, 7-21.
5. Baez J, Olsen D, Polarek JW (2005). *Appl Microbiol Biotechnol* 69, 245-252.
6. Rohanizadeh R, Swain MV, Mason RS (2007). *J Mater Sci Mater Med* 19, 1173-1182.
7. Watts RL, Raiser CD, Stover NP, Cornfeldt ML, Schweikert AW, Allen RC, Subramanian T, Doudet D, Honey CR, Bakay RA (2003). *J Neural Transm Suppl*, 215-227.
8. Huss FR, Junker JP, Johnson H, Kratz G (2007). *J Plast Reconstr Aesthet Surg* 60, 543-555.
9. Mattens E, Engels P, Hamerlijnck R, Kelder J, Schepens M, de Valois J, Vermeulen F, Wijers L (1999). *Cardiovasc Surg* 7, 432-435.
10. Takebayashi S, Hosaka M, Kubota Y, Ishizuka E, Iwasaki A, Matsubara S (1998). *J Urol* 159, 696-701.
11. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J (2003). *Adv Drug Deliv Rev* 55, 1547-1567.
12. Bouwstra J, van Es A, Toda Y (2006). *US Pat* 2006/0241032.
13. Chen T, Janjua R, McDermott MK, Bernstein SL, Steidl SM, Payne GF (2006). *J Biomed Mater Res, Part B* 77, 416-422.
14. Bouwstra J, Toda Y (2005). *US Pat* 2005/0101531.
15. Bouwstra J, Toda Y (2005). *US Pat* 2005/0119170.
16. Sutter M, Siepmann J, Hennink WE, Jiskoot W (2007). *J Control Release* 119, 301-312.
17. Balakrishnan B, Mohanty M, Umashankar PR, Jayakrishnan A (2005). *Biomaterials* 26, 6335-6342.
18. Bouwstra J, Toda Y, Van Es A (2006). *US Pat* 20060204511.
19. Hassoun S, Sabbah A (1998). *Allerg Immunol* 30, 80-82.
20. Olsen D, Jiang J, Chang R, Duffy R, Sakaguchi M, Leigh S, Lundgard R, Ju J, Buschman F, Truong-Le V, et al. (2005). *Protein Express Purif* 40, 346-357.
21. Goldberg I, Salerno AJ, Patterson T, Williams JI (1989). *Gene* 80, 305-314.
22. Capello J (1990). *Trends Biotechnol* 8, 309-311.
23. Kajino T, Takahashi H, Hirai M, Yamada Y (2000). *Appl Environ Microbiol* 66, 304-309.
24. Saddler JM, Horsey PJ (1987). *Anaesthesia* 42, 998-1004.
25. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997). *EMBO J* 16, 6702-6712.
26. Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA, Toman PD (2001). *J Biol Chem* 276, 24038-24043.
27. Myllyharju J (2003). *Matrix Biol* 22, 15-24.
28. Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, et al. (2000). *J Biol Chem* 275, 23303-23309.
29. Engel J, Kurtz J, Katchalski E, Berger A (1966). *J Mol Biol* 17, 255-272.
30. Engel J, Chen HT, Prockop DJ, Klump H (1977). *Biopolymers* 16, 601-622.
31. Frank S, Kammerer RA, Mechling D, Schulthess T, Landwehr R, Bann J, Guo Y, Lustig A, Bächinger HP, Engel J (2001). *J Mol Biol* 308, 1081-1089.
32. Rozkiewicz DI, Kraan Y, Werten MW, de Wolf FA, Subramaniam V, Ravoo BJ, Reinhoudt DN (2006). *Chem Eur J* 12, 6290-6297.
33. Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA (1998). *Science* 281, 389-392.
34. Xu C, Breedveld V, Kopeček J (2005). *Biomacromolecules* 6, 1739-1749.
35. Mi L, Fischer S, Chung B, Sundelacruz S, Harden JL (2006). *Biomacromolecules* 7, 38-47.

36. Shen W, Zhang K, Kornfield JA, Tirrell DA (2006). *Nat Mater* 5, 153-158.
37. Wheeldon IR, Barton SC, Banta S (2007). *Biomacromolecules* 8, 2990-2994.
38. Cao Y, Li H (2008). *Chem Commun*, 4144-4146.
39. Wright ER, McMillan RA, Cooper A, Apkarian RP, Conticello VP (2002). *Adv Funct Mater* 12, 149-154.
40. Tanaka F, Stockmayer WH (1994). *Macromolecules* 27, 3943-3954.
41. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989). *Gene* 77, 51-59.
42. Chambers SP, Prior SE, Barstow DA, Minton NP (1988). *Gene* 68, 139-149.
43. Cregg JM, Barringer KJ, Hessler AY, Madden KR (1985). *Mol Cell Biol* 5, 3376-3385.
44. Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM (2000). *Biotechnol Bioeng* 70, 1-8.
45. Katakura Y, Zhang W, Zhuang G, Omasa T, Kishimoto M, Goto Y, Suga K (1998). *J Ferment Bioeng* 86, 482-487.
46. Sutoh K, Noda H (1974). *Biopolymers* 13, 2391-2404.
47. De Wolf FA, Keller RCA (1996). *Prog Colloid Polym Sci* 102, 9-14.
48. Mizuno K, Hayashi T, Peyton DH, Bächinger HP (2004). *J Biol Chem* 279, 38072-38078.
49. Bruckner P, Rutschmann B, Engel J, Rothe M (1975). *Helv Chim Acta* 58, 1276-1287.
50. Makarov AA, Adzhubei IA, Protasevich II, Lobachov VM, Esipova NG (1993). *J Protein Chem* 12, 85-91.
51. Djabourov M (2003), in *Progress in Biotechnology*, Vol. 23 (Aalbersberg WI, Hamer RJ, Jasperse P, de Jongh HHJ, de Kruif CG, Walstra P, de Wolf FA, Eds.), Elsevier Science, Amsterdam, pp. 133-218.
52. Berg RA, Olsen BR, Prockop DJ (1970). *J Biol Chem* 245, 5759-5763.
53. Te Nijenhuis K (1981). *Colloid Polym Sci* 259, 1017-1026.
54. Persikov AV, Ramshaw JA, Brodsky B (2005). *J Biol Chem* 280, 19343-19349.
55. Kramer RZ, Vitagliano L, Bella J, Berisio R, Mazzarella L, Brodsky B, Zagari A, Berman HM (1998). *J Mol Biol* 280, 623-638.
56. Fitzkee NC, Rose GD (2004). *Proc Natl Acad Sci USA* 101, 12497-12502.
57. Gauba V, Hartgerink JD (2007). *J Am Chem Soc* 129, 2683-2690.
58. Gauba V, Hartgerink JD (2007). *J Am Chem Soc* 129, 15034-15041.

Appendix

Table A6.1. Oligonucleotide sequences

Name	DNA sequence (5'-3')
T-FW	GAGTCTCACCCGGTGCTC
T-RV	CCACCGGCTGGTCCGGGAGGACCCGGTGGTCCAGGTGGACCTGGTGGTCTGGTGGACCAGGTGGACC AGGTGGACCAGGCGGACCGGGAGCACCGGGTGAGACTC
RA-FW	GCGCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGAGTCACCAGGTCTCAGCCTGGTGGTCCA CAAAACCCAGGTTCCGGTGAAGGTCAGGGAAACGGTAACCCT
RA-RV	TGGGCCAGGGTTGGGTTTGGTGGTTCACCGGAACCTGGAGAACCACCTCCTTGTGGTTGACCACCTTGA GATGGTCCATTCTTGTTAGGGTTACCGTTTCCCTGACC
RB-FW	CAAACCCCAACCCTGGCCACAGAACGGTCAAAAGCCTGGTGGTCAACAAAACGGTCTGGTAATGGTCAA CAAGAGGGAAACGGTCAACAAAACGGTGGT
RB-RV	GCGTCTGCAGTACGAATTCTATTAGCCACCGGCTGGCTGACCTGGAGGTTTTCTGGTGGAGATCCAGGTC CGGACTGAGAACCAACCGTTTTTGTGACCGTTTC

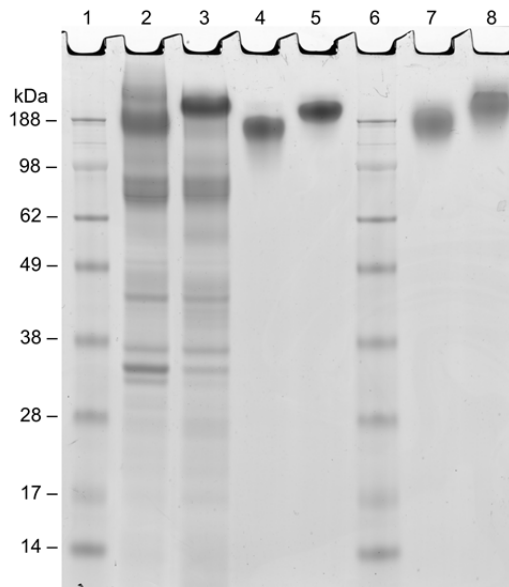


Figure A6.1. Full-length SDS-PAGE gel for Figure 6.2A. Lane 1, 6, molecular weight marker; lane 2, culture supernatant of **TR4T**; lane 3, culture supernatant of **TP4T**; lane 4, purified **TR4T**; lane 5, purified **TP4T**; lane 7, purified control protein **R4**; lane 8, purified control protein **P4**. For culture supernatants, 5 μ l was loaded, and for purified proteins ~20 μ g.

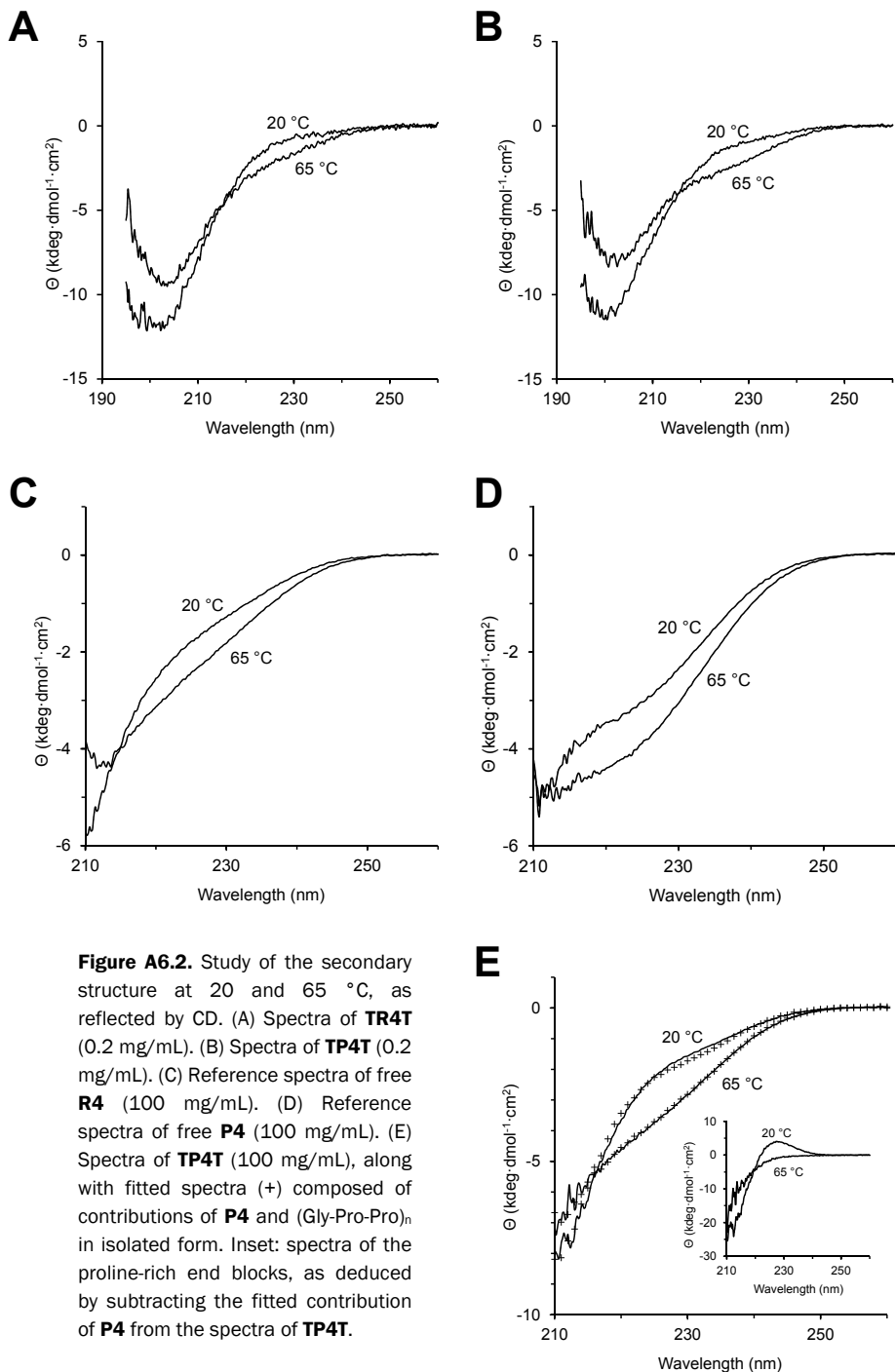


Figure A6.2. Study of the secondary structure at 20 and 65 °C, as reflected by CD. (A) Spectra of **TR4T** (0.2 mg/mL). (B) Spectra of **TP4T** (0.2 mg/mL). (C) Reference spectra of free **R4** (100 mg/mL). (D) Reference spectra of free **P4** (100 mg/mL). (E) Spectra of **TP4T** (100 mg/mL), along with fitted spectra (+) composed of contributions of **P4** and $(\text{Gly-Pro-Pro})_n$ in isolated form. Inset: spectra of the proline-rich end blocks, as deduced by subtracting the fitted contribution of **P4** from the spectra of **TP4T**.

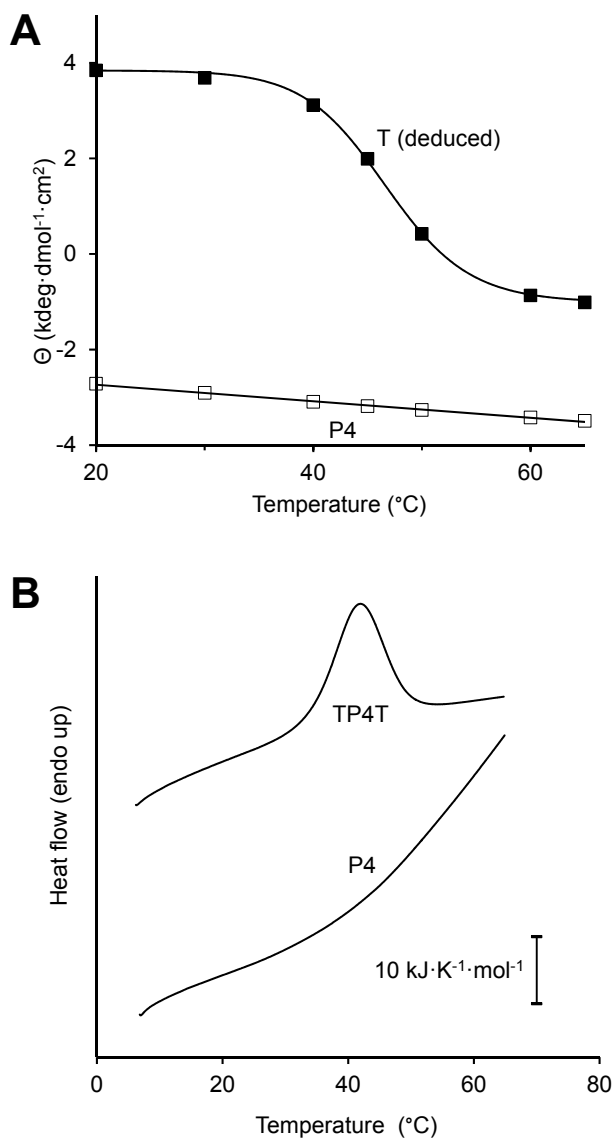


Figure A6.3. Thermal denaturation as reflected by (A) the average ellipticity from 230 to 225 nm or by (B) DSC. (A) Temperature profile of **P4** produced as a separate monoblock polymer, and of the proline-rich end blocks as deduced by subtracting the fitted contribution of **P4** from the melting curve of **TP4T**. (B) DSC thermograms of **P4** (1.4 mM) and **TP4T** (1.1 mM), recorded after equilibration for 10 h at 20 $^{\circ}\text{C}$.

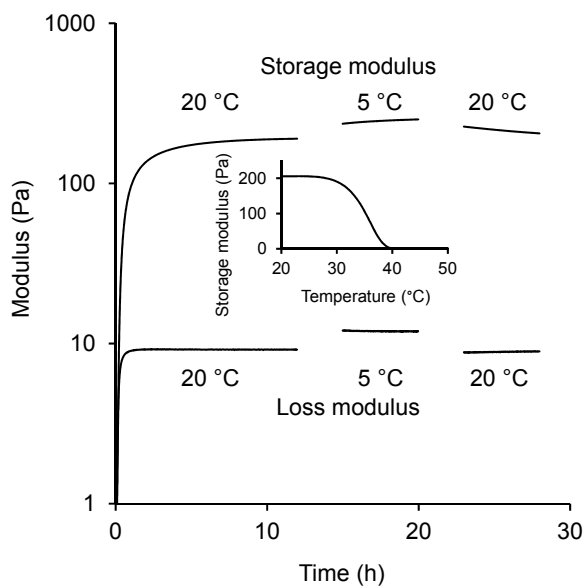


Figure A6.4. Characterization of a **TP4T** gel by dynamic rheology. The time-resolved storage and loss modulus were measured at the temperatures indicated, at a protein concentration of 2.2 mM. For each new temperature, the sample was allowed to equilibrate for three hours in the geometry before monitoring. Inset: storage modulus as a function of temperature.

Chapter 7

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

Specific coupling of *de novo* designed recombinant protein polymers for the construction of precisely structured nanomaterials is of interest for applications in biomedicine, pharmaceuticals 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 noninteracting 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 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 $\sim 3 \mu\text{M}$, regardless of glycosylation state of the PPxY domain. Binding was confirmed by isothermal titration calorimetry, with a K_d of $\sim 9 \mu\text{M}$. 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 (* equal contribution): Domeradзка NE*, Werten MWT*, de Vries R, de Wolf FA (2016). *Microb Cell Fact* 15, 105.

7.1. 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, 6].

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 noncovalent 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 modules with blocks 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 C^P_4 protein polymer previously developed by us. The 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 (Chapter 4). 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.

7.2. Materials and methods

7.2.1. Construction of expression vectors and strains

The double-stranded gene fragments encoding D^{WW} and D^{PPxY} were assembled via overlap extension PCR [30] from the oligonucleotides shown in the Appendix, Table A7.1. The gene fragments were digested with *XhoI/EcoRI* and cloned into *XhoI/EcoRI*-digested vector pMTL23 Δ BsaI [31], in order to obtain two vectors pMTL23 Δ BsaI- D^{WW} and pMTL23 Δ BsaI- D^{PPxY} . The vector pMTL23- C^P_4 contains the sequence encoding C^P_4 (previously referred to as 'P4'; Chapter 4), and was opened at the 3' end of the C^P_4 gene with *Van91I/EcoRI*. The newly prepared constructs pMTL23 Δ BsaI- D^{WW} and pMTL23 Δ BsaI- D^{PPxY} were digested with *DraIII/EcoRI* to release inserts D^{WW} and D^{PPxY} . The inserts were ligated into the opened pMTL23- C^P_4 vector, resulting in pMTL23- C^P_4 - D^{WW} and pMTL23- C^P_4 - D^{PPxY} .

A Ser¹² \rightarrow Ala mutant of the D^{PPxY} module was prepared by annealing of a pair of largely complementary oligos (Appendix Table A7.1), and is referred to as

D^{PPXY}*. This double-stranded adapter with *Dra*III/*Eco*RI overhangs was ligated into vector pMTL23-**C^P₄** previously digested with *Van*91I/*Eco*RI (at the 3' end of the **C^P₄** gene), resulting in pMTL23Δ*Bsa*I-**C^P₄**-**D^{PPXY}***.

The **C^P₄**-**D^{WW}**, **C^P₄**-**D^{PPXY}**, and **C^P₄**-**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-**C^P₄**-**D^{WW}**, pPIC9-**C^P₄**-**D^{PPXY}** and pPIC9-**C^P₄**-**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].

7.2.2. 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 [32] was used. The cultures were always inoculated with precultures grown to similar OD₆₀₀. 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 at the end of the glycerol fed-batch phase, and ~500 g/L 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.

7.2.3. 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 $20,000 \times g$ (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.

7.2.4. 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.

7.2.5. Treatment of proteins with α -mannosidase or phosphatase

For α -mannosidase digestions, 30 μ g 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 μ L of 20 mM sodium acetate, 0.4 mM zinc chloride, pH 5. Dephosphorylation involved incubation of 30 μ g of glycoprotein for 24 h at 37 °C under mild agitation with 60 U of calf intestinal alkaline phosphatase (NEB, Ipswich, MA) in 60 μ L 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).

7.2.6. 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 μ m AnchorChip target (Bruker, Leiderdorp, The Netherlands), using 5 mg/mL 2,5-dihydroxyacetophenone, 1.5 mg/mL 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 nonoverlapping 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.

7.2.7. 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 $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}/\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$, and $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}^*}/\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ were conducted in triplicate at RT. A 500 μ L aliquot of 10 μ M $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ was pipetted into a quartz fluorescence cuvette (Sigma-Aldrich, Zwijndrecht, The Netherlands). A 500 μ M solution of ligand was stepwise added to the cuvette, up to a final ligand concentration of ~24 μ M. The time interval between additions was 30 min. The final volume of added ligand solution did not exceed 5% of the starting volume of $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$. Curve fitting for K_d determination was done on averaged data of triplicate titrations.

7.2.8. 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 C^P_4 - D^{PPxY} or C^P_4 - 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 C^P_4 - 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 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.

7.3. Results and Discussion

7.3.1. Protein production and purification

The WW and PPxY domains are referred to hereafter as D^{WW} and D^{PPxY} , respectively (D for Dimerization). For amino acid sequences see Table 7.1. The domains were cloned at the 3' end of the gene encoding the previously reported C^P_4 protein polymer (Chapter 4). The encoded proteins C^P_4 - D^{WW} and C^P_4 - 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 C^P_4 block typically results in a purity of ~99% at the protein level [19, 34]. The gravimetrically determined yields, expressed in g per L of cell-free broth, were 2.2 g/L for C^P_4 - D^{WW} , and 2.3 g/L for C^P_4 - 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, 35, 36]. To our knowledge, production of WW and PPxY domains in *P. pastoris* has not been reported before. The nonoptimized yields are in the same g/L range as for C^P_4 (Chapter 4), showing that the D^{WW} and D^{PPxY} modules are not a significant

bottleneck. This offers good prospects towards their further use in the construction of complex supramolecular architectures and biomaterials.

Table 7.1. Amino acid sequences of the C-terminal **D^{WW}**, **D^{PPxY}** and **D^{PPxY*}** modules

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

^a All sequences end C-terminally in the cloning-derived amino acid sequence PAGG (not indicated).

7.3.2. Protein characterization

C^{P₄}-D^{WW} and **C^{P₄}-D^{PPxY}** were characterized by SDS-PAGE (Figure 7.1). Based on the well-established aberrant migration behavior of the **C^{P₄}** block in SDS-PAGE (Chapters 4 and 6) [23, 34], 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. This is due to the highly hydrophilic character and consequent low SDS-binding capacity of the **C^{P₄}** block (Chapter 4). The 13-residue **D^{PPxY}** block resembles the proline-rich nature of the **C^{P₄}** block, and as such is not expected to much affect the mobility of the protein in SDS-PAGE relative to that of **C^{P₄}** alone. On the other hand, the 37-residue hydrophobic **D^{WW}** block likely will, because attachment to **C^{P₄}** of peptides that improve SDS binding is known to increase the mobility of the polymer [23, 34]. Indeed, the proteins migrated accordingly in SDS-PAGE (Figure 7.1): **C^{P₄}-D^{PPxY}** migrates as slowly as the control **C^{P₄}** protein, and **C^{P₄}-D^{WW}** migrates faster than **C^{P₄}**. Although SDS-PAGE is not informative about the molecular mass of the polymers, it does show the proteins are relatively pure and intact.

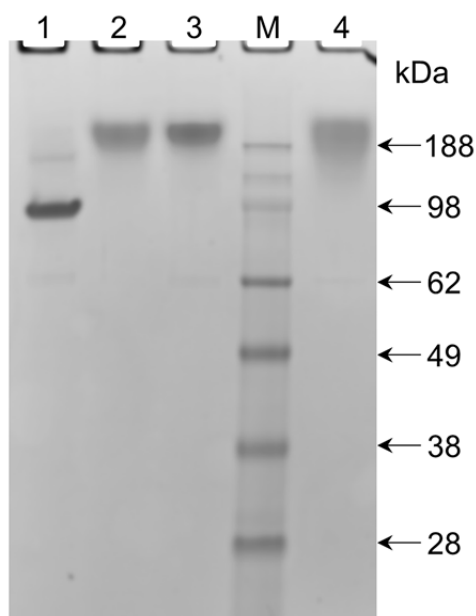


Figure 7.1. SDS-PAGE analysis of the purified protein polymers. Lane 1, $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$; lane 2, $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$; lane 3, $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY*}}$; lane M, protein molecular weight marker; lane 4, control protein C^{P_4} . Samples were loaded at 13 μg of protein per lane.

The molecular weight distribution of purified $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ and $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ was further investigated by MALDI-TOF mass spectrometry. A peak at m/z 41,440 was observed in the spectrum of $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ (Figure 7.2A). 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-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 $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ protein is pure and intact. The MALDI-TOF spectrum for $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ (Figure 7.2B), 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 $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ is glycosylated (Figure 7.3). Although the sequence of $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ contains no N-x-[ST] N-glycosylation motifs, *P. pastoris* is also capable of O-glycosylation [37]. Because C^{P_4} , as a separate protein, is known to be nonglycosylated (Chapter 4)

(see also Figure 7.3) most likely the single Ser residue in the added \mathbf{D}^{PPXY} block has been modified with *O*-glycans.

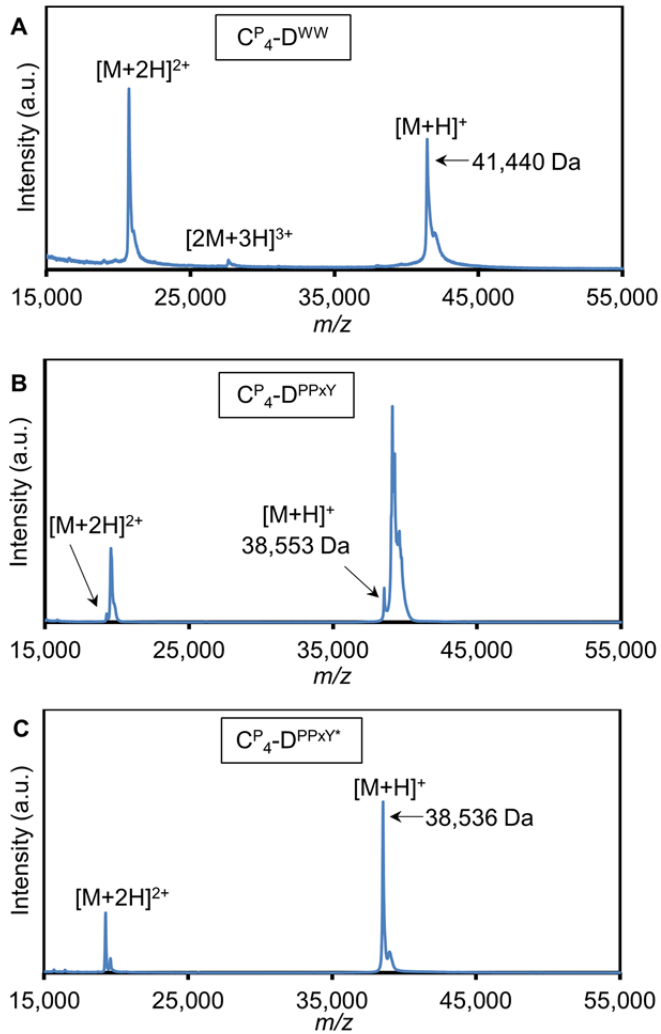


Figure 7.2. MALDI-TOF analysis of the purified protein polymers. (A) $\mathbf{C}^{\text{P}}_4\text{-D}^{\text{WW}}$, (B) $\mathbf{C}^{\text{P}}_4\text{-D}^{\text{PPXY}}$, (C) $\mathbf{C}^{\text{P}}_4\text{-D}^{\text{PPXY}^*}$.

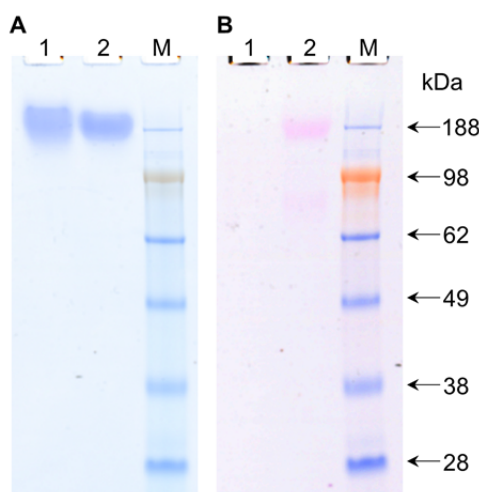


Figure 7.3. Detection of glycosylation by SDS-PAGE. Lane 1, **CP₄** control; lane 2, **CP₄-D^{PPXY}**; lane M: protein molecular weight marker. (A) Stained with Coomassie Brilliant Blue. (B) Stained using Periodic acid-Schiff staining. Equal amounts of protein were loaded for all samples.

7.3.3. Construction of a Ser¹² → Ala PPxY mutant

Glycosylation of therapeutic proteins by *P. pastoris* may cause adverse immune responses [38]. 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 **CP₄-D^{PPXY}**, we constructed a variant with a Ser¹² → Ala mutation in the PPxY module (Table 7.1). This variant, denoted **CP₄-D^{PPXY*}**, was produced in *P. pastoris* and purified as described above. SDS-PAGE shows a single band (Figure 7.1), and the yield of purified protein was 2.5 g/L of cell-free broth. MALDI-TOF of **CP₄-D^{PPXY*}** no longer showed the extensive pattern of glycosylated species (Figure 7.2C). 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 **CP₄-D^{PPXY*}** shows that indeed Ser¹² of the PPxY module in **CP₄-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 [39]. Although both **CP₄** and **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 [40], and with its inhibition when proline is the C-terminal neighbor [41]. However, this should not be taken to imply that such simple motifs are sufficient to determine O-glycosylation.

7.3.4. 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 [37], with dimers and trimers being the most abundant species [42, 43]. However, longer glycan chains of up to nine mannose residues have been described [44], as well as a phosphorylated Man₆ O-glycan [43]. The glycosidic links between the mannose units are mostly in α 1-2 arrangement [42, 43], although also α 1-3 and α 1-6 links have been reported [44, 45], as well as β 1-2 links [43].

For closely studying the glycan mass distribution, the MALDI-TOF analysis of $\mathbf{C^P_4-D^{PPxY}}$ was repeated with optimal settings for the m/z range of the relevant $[\mathbf{M+H}]^+$ ions (Figure 7.4A). Table 7.2 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 $\mathbf{C^P_4-D^{PPxY}}$ 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 $\mathbf{C^P_4-D^{PPxY}}$, and because both $\mathbf{C^P_4}$ (Chapter 4) and the nonglycosylated $\mathbf{C^P_4-D^{PPxY}^*}$ are not phosphorylated, such hydroxyl amino acid phosphorylation would need to have occurred specifically in the glycosylated fraction of the $\mathbf{C^P_4-D^{PPxY}}$ molecules. This unlikely notion can be excluded because treatment with alkaline phosphatase did not result in significant changes to the MALDI-TOF spectrum (Appendix Figure A7.1). 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 Man₆ O-glycan [43].

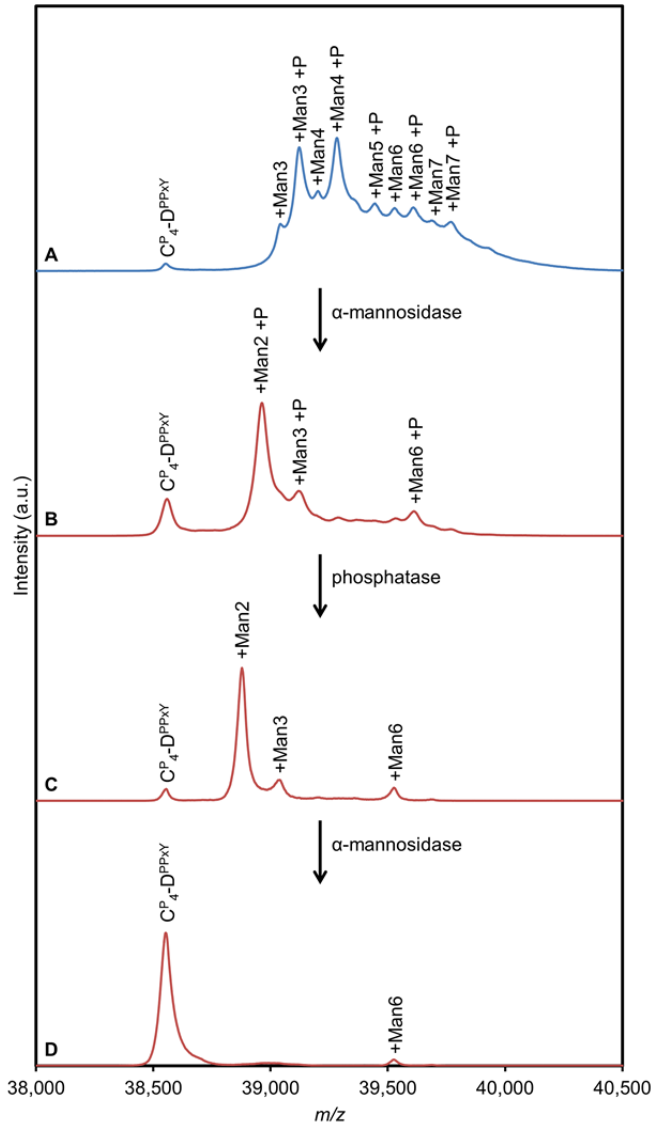


Figure 7.4. MALDI-TOF of $C^P_4-D^{PPxY}$ treated with α (1-2,1-3,1-6) mannosidase and phosphatase. (A) Untreated $C^P_4-D^{PPxY}$. The protein was digested with (B) α -mannosidase, then with (C) phosphatase, and finally digested again with (D) α -mannosidase. Enzymes were thermally inactivated between consecutive digestions. See Table 7.2 for observed m/z values and theoretical masses of the indicated glycoforms.

Table 7.2. 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
<i>CP₄-D^{PPXY}</i>				
	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
<i>CP₄-D^{PPXY} digested with α-mannosidase</i>				
	38,558	-	-	-
	38,963	405	Man ₂ + P	404
	39,121	563	Man ₃ + P	566
	39,611	1053	Man ₆ + P	1053
<i>CP₄-D^{PPXY} digested with (1) α-mannosidase, and (2) phosphatase^e</i>				
	38,553	-	-	-
	38,878	325	Man ₂	324
	39,039	486	Man ₃	486
	39,526	973	Man ₆	973
<i>CP₄-D^{PPXY} digested with (1) α-mannosidase, (2) phosphatase, and (3) α-mannosidase^e</i>				
	38,553	-	-	-
	39,526	973	Man ₆	973

^a See Figure 7.4 (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

When **CP₄-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 (Figure 7.4B, Table 7.2). 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 (Figure 7.4B; Table 7.2). Furthermore, a peak corresponding to Man₂+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 [46-48]. Thus, the digestion likely halted upon generation of the phosphorylated species observed in Figure 7.4B (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 (Figure 7.4C; Table 7.2), confirming that they were present as phosphodiester 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 (Figure 7.4D; Table 7.2). Possibly, the minor amount of remaining Man_6 is capped with a resistant β 1-2 Man disaccharide, as has been described for *P. pastoris* [43].

A detailed study of O-glycans chemically released from $\text{C}_4\text{-D}^{\text{PPxY}}$ 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* [47-49]. 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. [43]. Interestingly, the phosphorylated Man_6 previously reported was described as only a minor component [43], whereas most of the oligosaccharides on $\text{C}_4\text{-D}^{\text{PPxY}}$ are phosphorylated.

7.3.5. 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 $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ 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 $\text{C}^{\text{P}_4}\text{-D}^{\text{WW}}$ (10 μM) was thus followed during titration with a concentrated stock solution of either $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ or $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}^*}$.

An excitation wavelength of 295 nm was used to prevent excitation of tyrosines [50], which are present in the PPxY domain. With increasing concentration of the

ligand, the maximum emission wavelength of $\text{CP}_4\text{-D}^{\text{WW}}$ decreased, indicating the transition of at least one tryptophan of $\text{CP}_4\text{-D}^{\text{WW}}$ from a solvent-exposed to a more hydrophobic environment (Figure 7.5). As expected, this was accompanied by an increased fluorescence quantum yield. From the titration plots in Figure 7.5, a K_d of $2.5\ \mu\text{M}$ was calculated by nonlinear regression for $\text{CP}_4\text{-D}^{\text{PPxY}}$, and a K_d of $2.7\ \mu\text{M}$ for $\text{CP}_4\text{-D}^{\text{PPxY}^*}$. Apparently, the WW domain binds equally well to glycosylated and nonglycosylated $\text{CP}_4\text{-D}^{\text{PPxY}}$.

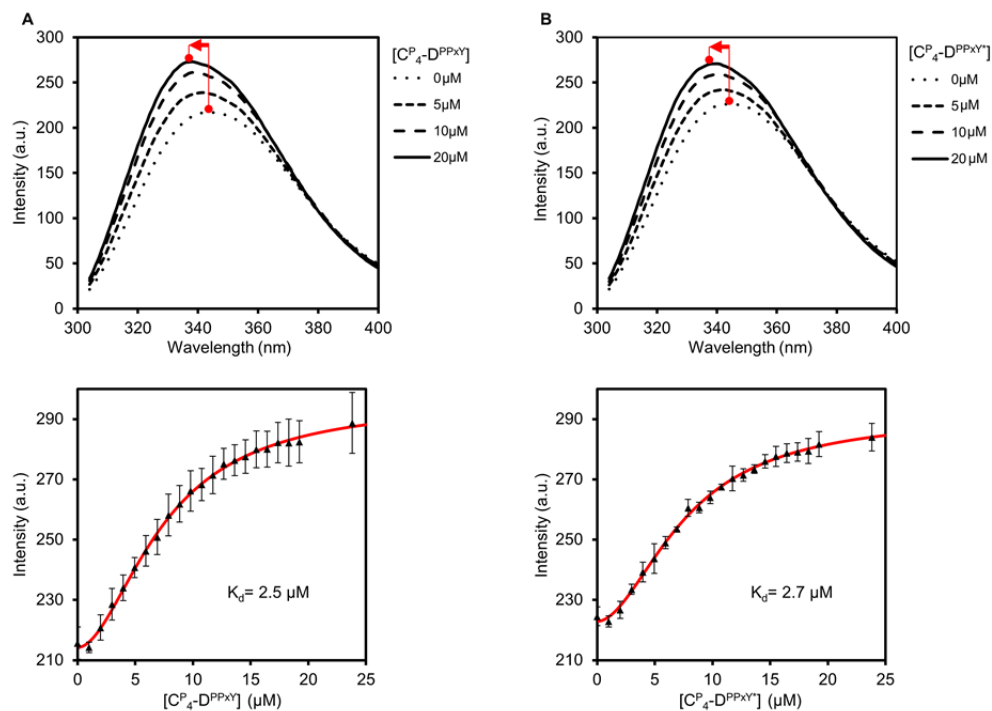


Figure 7.5. Steady-state tryptophan fluorescence of $\text{CP}_4\text{-D}^{\text{WW}}$ upon titration with (A) $\text{CP}_4\text{-D}^{\text{PPxY}}$ and (B) $\text{CP}_4\text{-D}^{\text{PPxY}^*}$. The top graphs present tryptophan fluorescence spectra obtained for $\text{CP}_4\text{-D}^{\text{WW}}$ upon addition of different concentrations of $\text{CP}_4\text{-D}^{\text{PPxY}}$ and $\text{CP}_4\text{-D}^{\text{PPxY}^*}$. The bottom graphs show titration of $\text{CP}_4\text{-D}^{\text{WW}}$ with $\text{CP}_4\text{-D}^{\text{PPxY}}$ and $\text{CP}_4\text{-D}^{\text{PPxY}^*}$, where fluorescence was monitored at 340 nm. Error bars represent SD ($n = 3$).

Binding affinities were further established using isothermal titration calorimetry. A fixed amount of $\text{CP}_4\text{-D}^{\text{WW}}$ ($200\ \mu\text{M}$) was titrated with concentrated ligand stock solution (Figure 7.6). Control experiments where buffer was titrated with ligand, resulted in a relatively small and constant heat of dilution (Appendix Figures A7.2A and B). Also $\text{CP}_4\text{-D}^{\text{WW}}$ titrated with control protein CP_4 showed only heat of dilution (Appendix Figure A7.2C). The binding isotherms of both $\text{CP}_4\text{-D}^{\text{PPxY}}$

and $\text{C}^{\text{P}}_4\text{-D}^{\text{PPxY}^*}$ showed an immediate decrease in differential power for each consecutive injection (Figure 7.6). The K_d values derived from the integrated heat plots in Figure 7.6 are 9.3 and 9.2 μM for $\text{C}^{\text{P}}_4\text{-D}^{\text{PPxY}}$ and $\text{C}^{\text{P}}_4\text{-D}^{\text{PPxY}^*}$, respectively. Both values are similar, and in reasonable agreement with the fluorescence spectroscopy results.

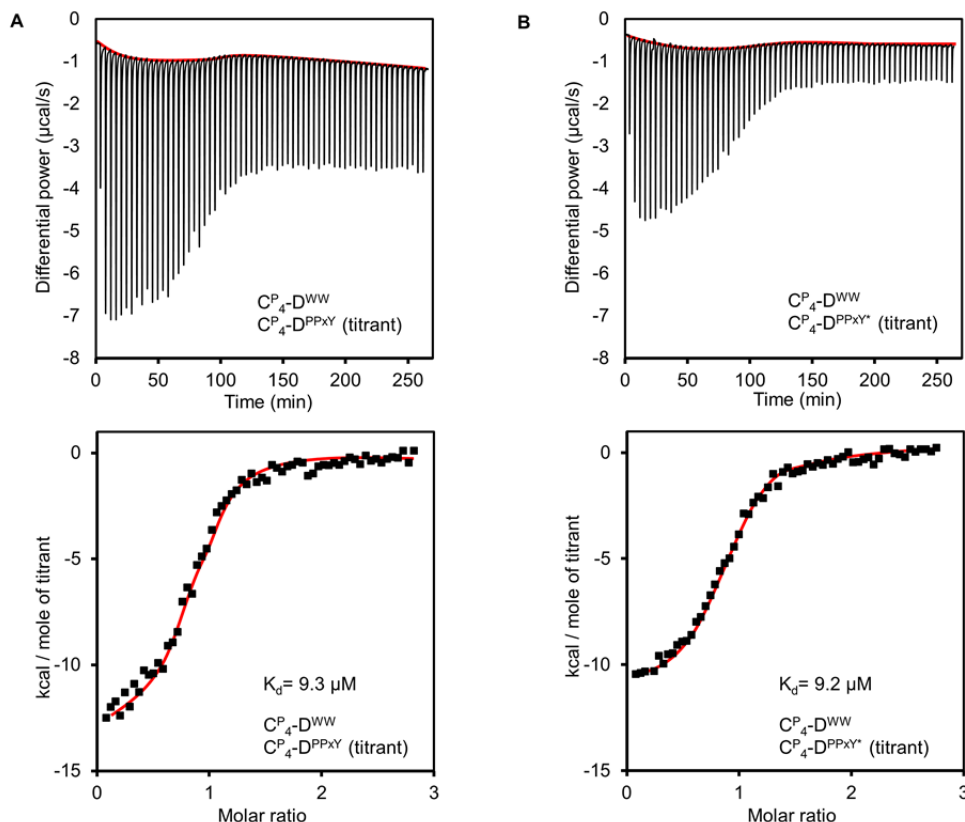


Figure 7.6. Binding study using isothermal titration calorimetry. (A) $\text{C}^{\text{P}}_4\text{-D}^{\text{PPxY}}$ titrant, (B) $\text{C}^{\text{P}}_4\text{-D}^{\text{PPxY}^*}$ titrant. The top graphs show the heat response upon titration of $\text{C}^{\text{P}}_4\text{-D}^{\text{WW}}$ with the proline-rich ligands. Bottom graphs represent the integrated peak areas per mole of ligand.

In general, the K_d of various WW-domains and their proline-rich ligands are in the high nM to low μM range [51]. According to Russ et al., the PPxY peptide, also used in our D^{PPxY} block, binds the CC43 WW domain with $K_d = 1.7 \mu\text{M}$, and the Nedd4.3 WW domain with $K_d = 11.2 \mu\text{M}$ [52]. 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 $9\ \mu\text{M}$) 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 $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}}$ and $\text{C}^{\text{P}_4}\text{-D}^{\text{PPxY}^*}$ 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.

7.4. 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.

Acknowledgments

We thank Ésió Bessa Ramos for his help in the design of the steady-state fluorescence spectroscopy experiment. The work of N.E.D. is supported by NanoNextNL, a micro and nanotechnology consortium of the government of The Netherlands and 130 partners. Financial support to M.W.T.W. and F.A.W. is provided by ERC Advanced Grant 267254 'BioMate'.

References

1. Rabotyagova OS, Cebe P, Kaplan DL (2011). *Biomacromolecules* 12, 269-289.
2. Ghandehari H, Cappello J (1998). *Pharm Res* 15, 813-815.
3. Tessmar JK, Göpferich AM (2007). *Adv Drug Deliv Rev* 59, 274-291.
4. Brea RJ, Reiriz C, Granja JR (2010). *Chem Soc Rev* 39, 1448-1456.
5. Matson JB, Stupp SI (2012). *Chem Commun* 48, 26-33.
6. Gagner JE, Kim W, Chaikof EL (2014). *Acta Biomater* 10, 1542-1557.
7. Price R, Poursaid A, Ghandehari H (2014). *J Control Release* 190, 304-313.
8. Shi P, Gustafson JA, MacKay JA (2014). *Int J Nanomedicine* 9, 1617-1626.
9. Heslot H (1998). *Biochimie* 80, 19-31.
10. Annabi N, Mithieux SM, Camci-Unal G, Dokmeci MR, Weiss AS, Khademhosseini A (2013). *Biochem Eng J* 77, 110-118.
11. Liu L, Busuttill K, Zhang S, Yang Y, Wang C, Besenbacher F, Dong M (2011). *Phys Chem Chem Phys* 13, 17435-17444.
12. Kopeček J, Yang J (2012). *Angew Chem Int Ed* 51, 7396-7417.
13. Fichman G, Gazit E (2014). *Acta Biomater* 10, 1671-1682.
14. Kim H, Siu K-H, Raeeszadeh-Sarmazdeh M, Sun Q, Chen Q, Chen W (2015). *Biotechnol Bioeng* 112, 1495-1505.
15. Walper SA, Turner KB, Medintz IL (2015). *Curr Opin Biotechnol* 34, 232-241.
16. Domeradzka NE, Werten MWT, de Wolf FA, de Vries R (2016). *Curr Opin Biotechnol* 39, 61-67.
17. Werten MWT, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA (1999). *Yeast* 15, 1087-1096.
18. Werten MWT, Wisselink WH, Jansen-van den Bosch TJ, de Bruin EC, de Wolf FA (2001). *Protein Eng* 14, 447-454.
19. Werten MWT, Teles H, Moers APHA, Wolbert EJH, Sprakel J, Eggink G, de Wolf FA (2009). *Biomacromolecules* 10, 1106-1113.
20. Schipperus R, Eggink G, de Wolf FA (2012). *Biotechnol Prog* 28, 242-247.
21. Beun LH, Storm IM, Werten MWT, de Wolf FA, Cohen Stuart MA, de Vries R (2014). *Biomacromolecules* 15, 3349-3357.
22. Włodarczyk-Biegun MK, Werten MWT, de Wolf FA, van den Beucken JJJ, Leeuwenburgh SCG, Kamperman M, Cohen Stuart MA (2014). *Acta Biomater* 10, 3620-3629.
23. Domeradzka NE, Werten MWT, de Vries R, de Wolf FA (2015). *Biotechnol Bioeng* 113, 953-960.
24. Wang H, Heilshorn SC (2015). *Adv Mater* 27, 3717-3736.
25. Chan G, Mooney DJ (2008). *Trends Biotechnol* 26, 382-392.
26. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P (1995). *FEBS Lett* 369, 67-71.
27. Macias MJ, Hyvonen M, Baraldi E, Schultz J, Sudol M, Saraste M, Oschkinat H (1996). *Nature* 382, 646-649.
28. Wong Po Foo CTS, Lee JS, Mulyasasmita W, Parisi-Amon A, Heilshorn SC (2009). *Proc Natl Acad Sci USA* 106, 22067-22072.
29. Pirozzi G, McConnell SJ, Uveges AJ, Carter JM, Sparks AB, Kay BK, Fowlkes DM (1997). *J Biol Chem* 272, 14611-14616.
30. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989). *Gene* 77, 51-59.
31. Golinska MD, Włodarczyk-Biegun MK, Werten MWT, Stuart MAC, de Wolf FA, de Vries R (2014). *Biomacromolecules* 15, 699-706.
32. Zhang W, Bevins MA, Plantz BA, Smith LA, Meagher MM (2000). *Biotechnol Bioeng* 70, 1-8.
33. Zacharius RM, Zell TE, Morrison JH, Woodlock JJ (1969). *Anal Biochem* 30, 148-152.
34. Moers APHA, Wolbert EJH, de Wolf FA, Werten MWT (2010). *J Biotechnol* 146, 66-73.
35. Bao W-J, Gao Y-G, Chang Y-G, Zhang T-Y, Lin X-J, Yan X-Z, Hu H-Y (2006). *Protein Express Purif* 47, 599-606.
36. Martinez-Rodriguez S, Bacarizo J, Luque I, Camara-Artigas A (2015). *J Struct Biol* 191, 381-387.

37. Bretthauer RK, Castellino FJ (1999). *Biotechnol Appl Biochem* 30, 193-200.
38. Daly R, Hearn MTW (2005). *J Mol Recognit* 18, 119-138.
39. Herscovics A, Orlean P (1993). *FASEB J* 7, 540-550.
40. Lehle L, Bause E (1984). *Biochim Biophys Acta* 799, 246-251.
41. Strahl-Bolsinger S, Tanner W (1991). *Eur J Biochem* 196, 185-190.
42. Duman JG, Miele RG, Liang H, Grella DK, Sim KL, Castellino FJ, Bretthauer RK (1998). *Biotechnol Appl Biochem* 28, 39-45.
43. Trimble RB, Lubowski C, Hauer CR, Stack R, McNaughton L, Gemmill TR, Kumar SA (2004). *Glycobiology* 14, 265-274.
44. Gustafsson A, Sjöblom M, Strindelius L, Johansson T, Fleckenstein T, Chatzissavidou N, Lindberg L, Ångström J, Rova U, Holgersson J (2011). *Glycobiology* 21, 1071-1086.
45. Boraston AB, Sandercock LE, Warren RAJ, Kilburn DG (2003). *J Mol Microbiol Biotechnol* 5, 29-36.
46. Hernández LM, Ballou L, Alvarado E, Gillece-Castro BL, Burlingame AL, Ballou CE (1989). *J Biol Chem* 264, 11849-11856.
47. Bergwerff AA, Stark W, Fendrich G, Knecht R, Blommers MJJ, Maerki W, Kragten EA, van Oostrum J (1998). *Eur J Biochem* 253, 560-575.
48. Nakayama K-i, Feng Y, Tanaka A, Jigami Y (1998). *Biochim Biophys Acta* 1425, 255-262.
49. Jars MU, Osborn S, Forstrom J, MacKay VL (1995). *J Biol Chem* 270, 24810-24817.
50. Weljie AM, Vogel HJ (2002). *Methods Mol Biol* 173, 75-87.
51. Kay BK, Williamson MP, Sudol M (2000). *FASEB J* 14, 231-241.
52. Russ WP, Lowery DM, Mishra P, Yaffe MB, Ranganathan R (2005). *Nature* 437, 579-583.

Appendix

Table A7.1. Oligonucleotides used in gene construction

Module	Oligonucleotide sequence (5' - 3')
D^{WW}	GC GCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTTTGCCTTCTGGTTGGGAA
	GGTCAACGTAATAAGTACGACCGTGAGGATCCTTTCTTTGTTCCCAACCAGAAGGCAAAG
	GTCGTACTTATTACGTTGACCATAACACTAGGACTACCACATGGGAAAGACCACAGCCATTGCCA
	GTACGAATTCTATTAGCCACCGGCTGGTGTCCAGGTGGCAATGGCTGTGGTCTT
D^{PPxY}	GC GCTCGAGAAAAGAGAGGCTGAAGCTGGTCCACCCGGTGCTGAATACCCTCCATACCCACCAC
	GTACGAATTCTATTAGCCACCGGCTGGACCAGATGGATAAGGAGGTGGTGGGTATGGAGGGTA
D^{PPxY*}	GTGCTGAATACCCTCCATACCCACCACCTCCTTATCCAGCTGGTCCAGCCGGTGGCTAATAG
	AATTCTATTAGCCACCGGCTGGACCAGCTGGATAAGGAGGTGGTGGGTATGGAGGGTATTCAGCACCGG

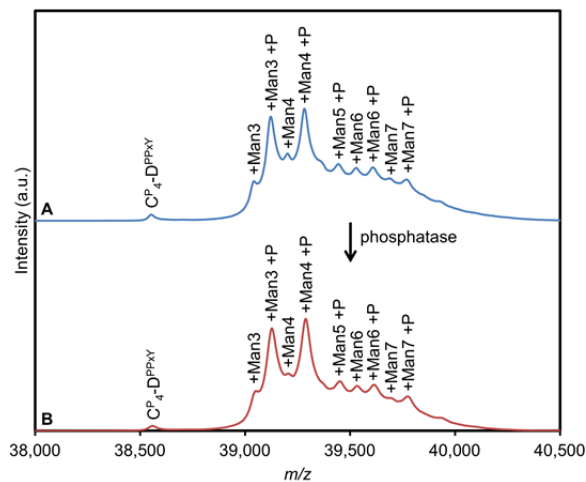


Figure A7.1. MALDI-TOF of C₄-D^{PPXY} before (A) and after (B) phosphatase treatment. The same glycoforms are seen in both spectra.

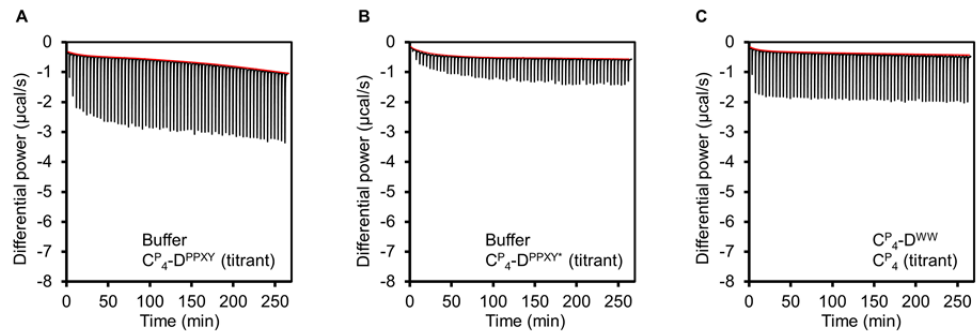


Figure A7.2. Control ITC measurements. (A) Titration of buffer with C₄-D^{PPXY}, (B) titration of buffer with C₄-D^{PPXY*}, (C) titration of C₄-D^{WW} with C₄ (protein without binding domain).

Chapter 8

Nanofibrillar hydrogel scaffolds from recombinant protein-based polymers with integrin- and proteoglycan-binding domains

This study describes the design, production, and testing of functionalized variants of a recombinant protein-based polymer that forms nanofibrillar hydrogels with self-healing properties. With a view to bone tissue engineering applications, we equipped these variants with N-terminal extensions containing either (1) integrin-binding (RGD) or (2) less commonly studied proteoglycan-binding (KRSR) cell-adhesive motifs. The polymers were efficiently produced as secreted proteins using the yeast *Pichia pastoris*, and were essentially monodisperse. The pH-responsive protein-based polymers are soluble at low pH and self-assemble into supramolecular fibrils and hydrogels at physiological pH. By mixing functionalized and nonfunctionalized proteins in different ratios, and adjusting pH, hydrogel scaffolds with the same protein concentration but varying content of the two types of cell-adhesive motifs were readily obtained. The scaffolds were used for the two-dimensional culture of MG-63 osteoblastic cells. RGD domains had a slightly stronger effect than KRSR domains on adhesion, activity, and spreading. However, scaffolds featuring both functional domains revealed a clear synergistic effect on cell metabolic activity and spreading, and provided the highest final degree of cell confluency. The mixed functionalized hydrogels presented here thus allowed to tailor the osteoblastic cell response, offering prospects for their further development as scaffolds for bone regeneration.

Published as (* equal contribution): Włodarczyk-Biegun MK*, Werten MWT*, Posadowska U, Storm IM, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Cohen Stuart MA, Kamperman M (2016). *J Biomed Mater Res A* 104, 3082-3092.

8.1. Introduction

In natural tissues, cells are surrounded by the extracellular matrix (ECM), which ensures the mechanical support and signaling necessary for regulation of cell metabolism and functions such as migration, proliferation, differentiation and apoptosis [1-4]. The ECM can be seen as a complex hydrogel containing multiple functionalities and bioactive groups at different length scales [5], which synergistically influence cell response. Mimicry of the properties and structure of the ECM remains a key objective in biomaterial design [2, 6, 7], which is why hydrogels are considered particularly promising candidates for use as scaffolds in tissue engineering [1, 8-10].

As highly hydrated polymers, hydrogels are usually biocompatible, noncytotoxic, well permeable to oxygen and nutrients, and often biodegradable [5, 8]. Hydrogels based on natural proteins and carbohydrates of the ECM, as well as a variety of synthetic polymers, have been investigated as biomaterials [5]. Despite their generally weak mechanical strength, hydrogels have not only shown promise in soft tissue engineering, but also in bone regeneration applications [11]. In applications not requiring gap filling, the hydrogel must primarily facilitate delivery of cells and/or release of growth factors [11]. Coating of surgical implants with hydrogels that provide biological cues, growth factors, or antibiotics can promote osseointegration [12]. Composite hydrogels complexed with ceramics or other additives have been devised to improve the mechanical properties [13], although no study has yet applied hydrogels in a load-bearing bone defect model [11]. Injectable hydrogels, that is, hydrogels that form *in situ* by chemical or physical crosslinking, are also of interest. They conform to the shape of the defect, and allow delivery of cells by mixing cells and polymer solution prior to crosslinking [5]. Self-healing hydrogels are especially suitable for this purpose, in that they can recover from strain-induced damage during application to the desired tissue [14], or afterwards.

In the design of biomaterial scaffolds as matrices for anchor-dependent cells such as osteoblasts, recognition sites for cell adhesion are crucial. The most commonly used peptide to improve cell adhesion is the Arg-Gly-Asp (RGD) motif derived from ECM proteins such as bone sialoproteins, vitronectin and fibronectin [15]. RGD-mediated adhesion is based on the interaction with multiple integrins and promotes adhesion of different cell types. Osteoblasts also attach through

another mechanism, less explored in the literature [16, 17], which involves the interaction between heparan sulfate proteoglycans on the cell membrane and heparin-binding sites in the ECM. Based on sequences found in bone adhesive proteins, Cardin and Weintraub [16] proposed that the amino acid sequence basic-basic-nonbasic-basic, in terms of protonation, binds heparan sulfate. One such sequence, Lys-Arg-Ser-Arg (KRSR), selectively improved adhesion of osteoblasts on model surfaces [17-20].

In recent years, a few studies used combinations of both RGD and KRSR adhesive motifs, in an effort to possibly create conditions more similar to the environment of the multifunctional native ECM [21-27]. However, these studies involved the post-production modification of scaffold surfaces with chemically synthesized RGD and KRSR peptides. In the present study, we report hydrogels self-assembled from genetically engineered proteins that contain these cell-adhesive motifs. Genetically engineered protein-based polymers used as scaffolds represent an emerging technology in the field of tissue engineering [2, 28-34]. Because amino acid sequence and molecular weight are directly defined by the genetic design, these polymers offer exquisite control over the chemistry, and desired bioactive modules can be incorporated without chemical modification [2, 32]. Protein-based polymers are in principle monodisperse, possible biological complications during production aside, which is difficult to achieve using conventional chemistry. These properties allow the generation of precisely defined, tailored materials. To our knowledge, this report presents the first recombinant protein-based hydrogel containing KRSR domains.

The hydrogel framework is based on the genetically engineered pH-responsive polymer $C_2S^{H_{48}}C_2$ developed by our group [35, 36]. At pH 6, or above, this protein self-assembles into stable, physical nanofibrillar hydrogels, thereby resembling nanostructural elements of the ECM. $C_2S^{H_{48}}C_2$ hydrogels are furthermore noncytotoxic and possess self-healing properties [37]. The material could thus potentially be used for example as an injectable scaffold, forming hydrogels in situ under physiological conditions, without use of chemical crosslinkers. We describe here the high-yield biotechnological production of functionalized variants of this protein-based polymer, and analyze the potential of the resulting fibrillar hydrogels for osteoblast cell culture. By mixing functionalized proteins, containing either RGD or KRSR motifs, and non-functionalized proteins in different ratios,

hydrogel scaffolds were conveniently prepared with precisely controlled overall concentration of the two motifs. We show that this tailoring allowed to control osteoblastic cell adhesion, activity, spreading, and formation of a confluent layer.

8.2. Materials and methods

8.2.1. Construction of recombinant strains

The production strain for triblock copolymer $C_2S^{H_{48}}C_2$ has been described by us previously [36]. Two variants of the corresponding gene were designed; one encoding N-terminal RGD-motifs ($B^{RGD}C_2S^{H_{48}}C_2$), and another encoding N-terminal KRSR-motifs ($B^{KRSR}C_2S^{H_{48}}C_2$). These genes and corresponding strains were constructed as follows.

A double-stranded adapter B^{RGD} (short for cell-Binding block with RGD motifs) with *HindIII* and *BsaI* overhangs was prepared by annealing of oligonucleotides: 5'-AGCTTGAATTCGGTAGAGGAGATTCTCTGGTGGATCCGGTGGAGGTCTGGAGGTGGATCTGGTGAAGAGGTGACTACCCAGGTCTCG-3' and 5'-GCACCGAGACCTGGTGAGTCACCTCTTCCACCAGATCCACCTCCAGAACCTCCACCGGATCCACCAGGAGAATCTCTCTACCGAATTCA-3'. This adapter encodes two GRGDSP motifs separated by a (GGSG)₃ flexible spacer. Likewise a double-stranded adapter B^{KRSR} (short for cell-Binding block with KRSR motifs) encoding three repeats of the sequence PKRSRGGG was prepared by annealing of oligonucleotides: 5'-AGCTTGAATTCGGTCCAAAGAGATCTAGAGGAGGTGGACCTAAACGTTCCAGAGGTGGAGGTCCAAAGAGGTCTAGAGGAGGTGGTCTCG-3' and 5'-GCACCGAGACCACCTCCTCTAGACCTCTTTGGACCTCCACCTCTGGAACGTTTAGGTCCACCTCCTCTAGATCTCTTTGGACCGAATTCA-3'. These adapters were inserted into the *HindIII*/*BsaI* sites of vector pMTL23- $C_2S^{H_{48}}C_2$ [36], resulting in vectors pMTL23- $B^{RGD}C_2S^{H_{48}}C_2$ and pMTL23- $B^{KRSR}C_2S^{H_{48}}C_2$, respectively. These were then digested with *EcoRI*/*NotI*, and the released inserts were cloned into the *EcoRI*/*NotI* sites of *Pichia pastoris* expression vector pPIC9 (Thermo Fisher Scientific). The vectors pPIC9- $B^{RGD}C_2S^{H_{48}}C_2$ and pPIC9- $B^{KRSR}C_2S^{H_{48}}C_2$ thus obtained were linearized with *SalI* and used to transform *P. pastoris* GS115 by electroporation as described previously (Chapter 2).

8.2.2. Fermentation and protein purification

Methanol fed-batch fermentations of the $B^{RGD}C_2S^{H_{48}}C_2$ and $B^{KRSR}C_2S^{H_{48}}C_2$ strains at 30 °C and pH 3 were performed in 2.5-L Bioflo 3000 bioreactors (New

Brunswick Scientific) as described previously for $C_2S^{H_{48}}C_2$ [37]. At the end of fermentation, the cells were separated from the broth by centrifugation for 10 min at $10,000 \times g$ (RT) in an SLA-3000 rotor (Sorvall), and the supernatant was micro-filtered.

Proteins were purified from the supernatant as previously described for $C_2S^{H_{48}}C_2$ [37], with the exception that the acetone precipitation step was omitted. Shortly, proteins were selectively precipitated twice at 45% ammonium sulfate saturation and subsequently dialyzed, filter sterilized ($0.2 \mu m$), and lyophilized.

8.2.3. SDS-PAGE, N-terminal protein sequencing, amino acid analysis, and carbohydrate determination

The NuPAGE Novex system (Thermo Fisher Scientific) was used for SDS-PAGE, with 10% Bis-Tris gels, MES SDS running buffer, and SeeBlue Plus2 prestained molecular mass markers. Gels were stained with Coomassie SimplyBlue SafeStain (Thermo Fisher Scientific).

Purified proteins in solution were N-terminally sequenced via Edman degradation by Midwest Analytical (St. Louis, MO). Amino acid composition analysis after acid hydrolysis was performed by Ansynth Service B.V. (Roosendaal, The Netherlands). Protein purity was estimated by linear least-squares fitting to the obtained amino acid composition of (1) the theoretical composition of the respective pure protein and (2) the composition determined for host-derived proteins present in the medium.

Because polysaccharides are the major nonprotein contaminant in the extracellular medium of *P. pastoris* cultures, the carbohydrate content in the purified protein was determined using a phenol-sulfuric acid assay as previously described [37]. Proteins contained $<2\%$ (w/w) of polysaccharides.

8.2.4. Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was performed using an ultrafleXtreme mass spectrometer (Bruker). Samples were prepared by the dried droplet method on a $600 \mu m$ AnchorChip target (Bruker), using 5 mg/mL 2,5-dihydroxyacetophenone, 1.5 mg/mL diammonium hydrogen citrate, 25% (v/v) ethanol and 1% (v/v) trifluoroacetic acid as matrix. Spectra were derived from ten 500-shot ($1,000 \text{ Hz}$) acquisitions taken at

nonoverlapping 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.

8.2.5. Atomic force microscopy

Protein samples of (A) pure $C_2S^{H_{48}}C_2$, and (B) $B^{KRSR}C_2S^{H_{48}}C$ / $B^{RGD}C_2S^{H_{48}}C_2$ mixed in a 1:1 ratio were dissolved in 10 mM HCl (final pH 2 to 3). The pH of the ~2 mL samples was adjusted to 7.4 ± 0.2 by addition of 0.1M NaOH, using a pH meter with a micro electrode (Mettler-Toledo InLab Micro). Milli-Q was then added to obtain protein solutions between 0.05% and 0.1% (w/v). After 24 h of incubation at room temperature, solutions were pipetted onto silica wafers and left for 20 min to allow fibril deposition. Wafers were subsequently rinsed with Milli-Q to remove salts. The samples were dried by removing the bulk of the solution with a tissue and/or using a mild stream of nitrogen. The dry samples were analyzed with a Nanoscope V (Veeco), in Scan Asyst mode, using a nonconductive silicon nitride probe (NP-10, Bruker) with a spring constant of 0.32 N/m. Images were recorded at a frequency of 0.997 Hz and processed with NanoScope Analysis 1.20 software (Veeco).

8.2.6. Rheology

Solutions of the purified proteins (A) $C_2S^{H_{48}}C_2$, (B) $B^{RGD}C_2S^{H_{48}}C_2$, and (C) $B^{KRSR}C_2S^{H_{48}}C_2$ were prepared in 10 mM HCl and vortexed for 1.5 h using a multi-tube holder. The pH of the ~2 mL samples was adjusted to 7.4 ± 0.2 by addition of 0.1M NaOH, using a pH meter with a micro electrode (Mettler-Toledo InLab Micro). PBS (phosphate buffered saline, pH 7.4, Sigma-Aldrich) was then added to reach a 2% (w/v) protein concentration. Solutions were loaded immediately into the rheometer (Physica MCR 501 Rheometer, Anton Paar) with a Couette CC10/T200 geometry (bob diameter 10.002 mm, cup diameter 10.845 mm). Each sample was placed in the measuring cell on top of 500 μ L of perfluorinated fluid (Galden HT-70, Solvay Specialty Polymers) to decrease the sample volume, and covered with paraffin oil to avoid evaporation. Gel formation was analyzed by recording storage moduli over time under sinusoidal deformation (frequency (f) = 1 Hz and strain (γ) = 1%). After 15 h, gels were broken (f = 1 Hz and logarithmically

increasing deformation $\gamma = 0.1$ to 100%). Self-healing behavior was then monitored by recording storage moduli under sinusoidal deformation as before. All measurements were performed at 37 °C.

8.2.7. Preparation of scaffolds for cell culture

Freeze-dried proteins were separately dissolved at 4% (w/v) in 10 mM HCl for at least 2 h by vortexing with a multi-tube holder. Solutions were then mixed to obtain 1 mL samples of the required protein compositions. To allow simultaneous preparation of multiple scaffolds without pH measurement, one volume of 0.1 M NaOH was added to adjust the protein concentration to 2% (w/v) and to reach a pH above neutral in all samples. This ensures proper gel formation across samples independent of the precise pH, as self-assembly occurs at pH 6 or above [36]. The protein solutions were immediately transferred to 48-well cell culture plates (250 μ L per well) and left for 2 h at room temperature to allow full setting of the gels. These 2 h are a conservative approach, as gels already form (crossover point of storage modulus and loss modulus) within minutes (Appendix Figure A8.1). To ensure a pH value of 7.4 for cell seeding, the gels were further incubated in 0.5 mL of PBS for 3 h and then in 0.5 mL of cell culture medium (α -MEM, 22571, Thermo Fisher Scientific) with 10% (v/v) fetal bovine serum (FBS) for 15 h (with one medium change after 12 h).

8.2.8. Cell pre-culture

Cryo-preserved MG-63 cells (ATCC CRL-1427; LGC Standards GmbH), passage 90, were cultured for 6 days in proliferation medium (α -MEM 22571, Thermo Fisher Scientific), supplemented with 10% (v/v) FBS, at 37 °C, 95% relative humidity and 5% CO₂. Prior to seeding, cells were washed twice with PBS, enzymatically detached for 5 min with trypsin-EDTA (0.25% (w/v) trypsin, 0.02% (w/v) EDTA), and resuspended at 36,000 cells/mL in osteogenic medium. The medium was composed of α -MEM supplemented with 10% (v/v) FBS, 100 U/mL penicillin and 10 μ g/mL streptomycin (Thermo Fisher Scientific), 50 mg/L ascorbic acid (Sigma-Aldrich), 10 nM dexamethasone (Sigma), 10 mM β -glycerophosphate disodium salt hydrate (Sigma). Next, 500 μ L of cell suspension was added per well (final seeding density: 20,000 cells/cm²). Scaffolds incubated in osteogenic medium without cells were included as controls. All samples were prepared in triplicate

($n = 3$). The entire study was run for 23 days in an incubator at 37 °C, 95% relative humidity and 5% CO₂. The culture medium was refreshed twice a week for all samples.

8.2.9. Cell adhesion

Cell adhesion was analyzed 4 h after cell seeding. The medium was removed from a dedicated set of wells to take out nonadherent cells. Gel surfaces were imaged by randomly taking light microscopy pictures ($n = 3$ per sample) using an inverted phase-contrast microscope (Leica DMIL). Pictures were taken close to the center of the gels to ensure that the imaged area was relatively flat. The number of cells on the area of 2,048 $\mu\text{m} \times 1,536 \mu\text{m}$ was counted in manual mode using ImageJ 1.49o software. Cell adhesion was calculated as the number of cells per mm².

8.2.10. Cell metabolic activity

To assess cell metabolic activity, an alamarBlue (Thermo Fisher Scientific) assay was performed on a dedicated set of wells on days 1, 3, 10, and 23, following the protocol of the manufacturer. Shortly, the cell culture medium in the wells was exchanged for fresh medium containing 10% (v/v) of alamarBlue reagent, followed by incubation at 37 °C for color development. In view of high cell activity on day 23, an incubation time of only 1.5 h was used for these samples, whereas for the earlier days 4 h was used. The medium containing the reagent was subsequently removed for measuring fluorescence (ex 560 nm, em 590 nm; FLx800 reader, Bio-Tek Instruments), and all values were normalized to activity per hour. To compare cell activity on different scaffolds, the average fluorescence of the medium was calculated ($n = 3$).

8.2.11. Cell viability and spreading

To study cell viability, and to visualize changes over time in cell distribution and morphology (spreading), a LIVE/DEAD Viability/Cytotoxicity Kit (L3224, Thermo Fisher Scientific) was used on days 1 and 3, and only the calcein-AM compound from the kit on days 10, and 23. Fluorescent staining of F-actin and

nuclei was performed on samples from days 1, 3, and 23 for further confirmation of cell distribution and spreading.

For the LIVE/DEAD assay, three samples of each gel type ($n = 3$) were washed twice with PBS and afterwards incubated in 0.5 mL of PBS with 2 mM calcein-AM and, optionally, 4 mM ethidium homodimer, in the dark, for 30 min at 37 °C. Next, scaffolds were washed twice with PBS and cells were visualized using a fluorescence microscope (Axio Imager Microscope Z1; Carl Zeiss Micro Imaging GmbH). Overlay images were prepared, with dead cells appearing in red (ex 517 nm, em 617 nm) and live cells in green (ex 488 nm, em 517 nm).

Samples for fluorescent staining of F-actin and nuclei were prepared as follows. First, the cell culture medium was removed from the cell culture well-plate. Subsequently, the scaffolds in the wells were washed gently with PBS, fixed for 1.5 – 4 h with 0.7 mL per well of 3.3% PFA (paraformaldehyde) in PBS, and then kept in 70% (v/v) ethanol until the day of imaging (for a maximum of 5 days). Prior to imaging the gels, ethanol was removed, scaffolds were washed twice in PBS, and cells were permeabilized for 20-30 min with 0.8 mL per well of 1% Triton X-100 in PBS containing 1% FBS. Scaffolds were then incubated for 2 h with Alexa Fluor 568 conjugated phalloidin (Thermo Fisher Scientific) for F-actin staining and with DAPI (Thermo Fisher Scientific) for visualization of nuclei. Both reagents were diluted in PBS containing 1% FBS (1:200 and 1:2,500, respectively), and 0.5 mL of this mixture was added per well. After washing in PBS for 10 min, scaffolds were examined using an Axio Imager Microscope Z1 (Carl Zeiss Micro Imaging GmbH).

8.2.12. Statistical analysis

Statistical analysis was performed using SPSS Statistics software. Statistical differences were estimated using one-way analysis of variance followed by an LSD post-hoc test. Differences were considered significant at $p < 0.05$, if not stated differently. Results were reported as the mean \pm standard deviation.

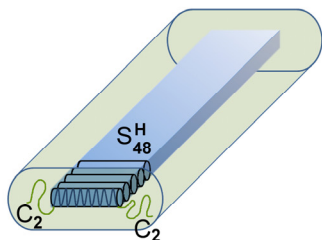
8.3. Results

8.3.1. Polymer design

We previously reported the molecular design of the fibril-forming protein $C_2S^{H_{48}}C_2$ [35, 36]. It features an $S^{H_{48}}$ (silk-inspired) middle block consisting of 48 repeats of the octapeptide GAGAGAGH. Whereas at low pH the proteins are soluble owing to repulsion by the positively charged histidines, at neutral pH the charges are neutralized [36] and the $S^{H_{48}}$ domains form β -rolls or β -sheets [38-40]. These, in turn, form the fibril core through stacking, mediated by hydrophobic interactions. The $S^{H_{48}}$ block is flanked on both sides by hydrophilic C_2 (random coil) blocks, each consisting of two 99 amino acid-long unfolded domains in tandem, which form a corona surrounding the core and provide colloidal stability (Figure 8.1A). At higher protein concentrations (above several g/L), gels are formed, most likely as a result of the entanglement of long fibrils [41], and possibly the occurrence of weak physical crosslinks [36].

Here, we designed two functionalized variants of $C_2S^{H_{48}}C_2$ with N-terminal cell-binding extensions (**B** domains, for short). One variant, $B^{RGD}C_2S^{H_{48}}C_2$, contains integrin-binding RGD motifs [42], and the other, $B^{KRSR}C_2S^{H_{48}}C_2$ contains proteoglycan-binding KRSR motifs [17] (Figure 8.1B). For the design, the following considerations were taken into account: (1) To potentially increase cell interaction, the design should preferably feature several copies of these motifs. (2) The length of the extensions should be limited, so as to minimize possible effects on the well-characterized fibril-forming properties of the $C_2S^{H_{48}}C_2$ core component. (3) The repeated cell-binding motifs should be separated by suitable flexible linkers to allow sufficient conformational freedom.

A



B

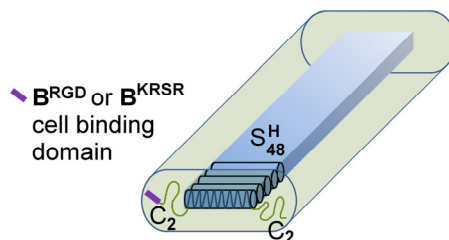


Figure 8.1. Tentative structure of fibrils at neutral pH (not to scale). (A) unmodified $C_2S^{H_{48}}C_2$; (B) functionalized variants. Representation based on the model proposed for the related $C_2S^{F_{48}}C_2$ [38].

In $\mathbf{B}^{\text{RGD}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$, a flexible Gly/Ser-rich sequence was chosen to separate the RGD motifs because this type of spacer is widely used as a constituent of engineered proteins produced in various hosts including *P. pastoris* [43, 44]. A spacer length of 12 residues in the form of three GGSG repeats was used. Two RGD motifs were incorporated, where the specific motif used was GRGDSP from fibronectin [42]. The N-terminal extension relative to $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ thus became: GRGDSPGGSGGGSGGGSGGRGDSP, which adds ~3% to the length of the protein sequence. This sequence is N-terminally preceded by the cloning-derived sequence YVEF, as is the case in $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ [38].

In the design of $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$, poly-glycine stretches were used as spacers between the KRSR motifs, in view of a study that reports osteoblast adhesion onto glass surfaces with C-terminally immobilized KRSR, KRSRG₃, or KRSRG₆ peptides [17]. However, as the authors found similar adhesion efficiency for all peptides tested, spacer length does not appear to be critical. Rather, because a tetravalent poly-lysine peptide core with four attached KRSR peptides performed significantly better, the authors suggest peptide density may be more important. We, therefore, incorporated three KRSRG₃ motifs, while maintaining the same small length of the N-terminal extension, namely 24 residues, as used in $\mathbf{B}^{\text{RGD}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$. Another critical design consideration was to prevent proteolysis of Lys-Arg in the KRSR motifs by the *P. pastoris* homolog of Kex2 protease. This serine endoprotease was first identified in *S. cerevisiae* as the protease required for maturation of the α -mating factor pheromone, via cleavage at the C-terminal side of Lys-Arg [45]. As we have shown previously (Chapters 2 and 3), *P. pastoris* Kex2 can also digest particular mono-arginyl sites, as governed to a large extent by the residues at positions -2 to -4 relative to the scissile bond. Under the assumption that *P. pastoris* and *S. cerevisiae* Kex2 have similar substrate specificity, the extent of putative cleavage at the C-terminal side of the KRSR sequences is likely low, in view of the moderately unfavorable Ser at position -2 relative to the scissile bond [46]. As for potential digestion of the canonical Lys-Arg motif in KRSR, we reasoned it might be possible to minimize this by placing highly unfavorable residues at positions -4 and -3. We thus arrived at the following sequence design for the N-terminal \mathbf{B}^{KRSR} extension: GPKRSRGGGPKRSRGGGPKRSRGG (again N-terminally preceded by the cloning-derived sequence YVEF, and C-terminally followed by the $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ sequence starting with Gly). For all three Lys-Arg motifs in this sequence, position -4 relative

to the scissile bond is occupied by Gly, and position -3 by Pro. Both amino acids rank as highly unfavorable in these positions, at least for *S. cerevisiae* Kex2 [47].

8.3.2. Protein synthesis and characterization

The polymers were produced as secreted proteins in methanol fed-batch bioreactor cultures of *P. pastoris*, and purified as previously described [37]. Based on the determined amino acid composition, the purity of **B^{RGD}C₂S^{H₄₈}C₂** was estimated at 99.8%, and that of **B^{KRSR}C₂S^{H₄₈}C₂** at 99.3%. The gravimetrically determined yields of the purified, desalted, and lyophilized material, expressed per volume of cell-free fermentation broth, were 3.2 g/L for **B^{RGD}C₂S^{H₄₈}C₂** and 3.5 g/L for **B^{KRSR}C₂S^{H₄₈}C₂**.

The functionalized proteins and unmodified **C₂S^{H₄₈}C₂** were analyzed by SDS-PAGE (Figure 8.2). The main bands of all three proteins migrated in SDS-PAGE at an apparent molecular weight of ~105 kDa, although the calculated masses are ~66 to ~68 kDa. As described previously for **C₂S^{H₄₈}C₂** [36, 37], this discrepancy is due to the high content of hydrophilic and small amino acids in this protein. The faint upper band and low mass smear visible in all proteins are probably SDS-PAGE artifacts, as has previously been shown for **C₂S^{H₄₈}C₂** [37]. N-terminal sequencing of **B^{RGD}C₂S^{H₄₈}C₂** showed the expected YVEFGRGD sequence without any secondary reactions, confirming that the protein was N-terminally intact. Likewise, N-terminal sequencing of **B^{KRSR}C₂S^{H₄₈}C₂** revealed mainly the expected YVEFGPKRSR sequence, which shows that most of the N-terminal **B^{KRSR}** extension was N-terminally intact. However, besides spurious background signals, a minor secondary species, most likely corresponding to SRGG, was also observed in this sample. This sequence may have resulted from a limited extent of Kex2 cleavage at the C-terminal side of the dibasic GPKR sequences.

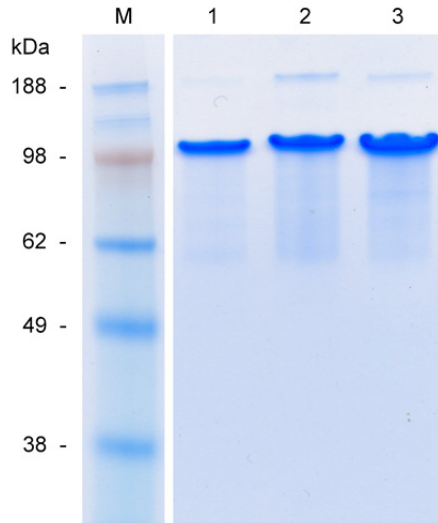


Figure 8.2. SDS-PAGE of purified proteins. Lane 1: $C_2S^{H_{48}}C_2$; lane 2: $B^{RGD}C_2S^{H_{48}}C_2$; lane 3: $B^{KRSR}C_2S^{H_{48}}C_2$; lane M: molecular weight marker. For all samples $\sim 5 \mu\text{g}$ of protein was loaded.

The molecular mass of the proteins was verified by MALDI-TOF (Figure 8.3). For both $B^{RGD}C_2S^{H_{48}}C_2$ and $B^{KRSR}C_2S^{H_{48}}C_2$, the observed masses of the main peaks (67,932 and 68,412 Da, respectively) correspond, within experimental error, to the calculated masses of the intact proteins (68,061 and 68,535 Da, respectively). The MALDI-TOF spectra furthermore show that $B^{RGD}C_2S^{H_{48}}C_2$ was fully monodisperse, and $B^{KRSR}C_2S^{H_{48}}C_2$ nearly so. The slight shoulder to the left of the main peak in the spectrum of $B^{KRSR}C_2S^{H_{48}}C_2$ (indicated by an arrow in Figure 8.3) contains three peaks whose masses correspond to that of the protein after N-terminal truncation following either the first, second, or third GPKR motif. This result is in agreement with the minor SRGG species observed in N-terminal sequencing. Additionally, three minor peaks are present, with masses that could correspond to N-terminal truncation following either the first, second, or third KRSR motif. Apparently, Kex2 cleavage occurred at the C-terminal side of the dibasic GPKR and monobasic KRSR motifs, but only at a very low level. The shoulder to the right side of the main peak of both proteins (Figure 8.3) corresponds to Glu-Ala extended species in a small portion of the molecules, as observed previously for $C_2S^{H_{48}}C_2$ [37], and results from incomplete processing of the α -mating factor propeptide (Chapter 2).

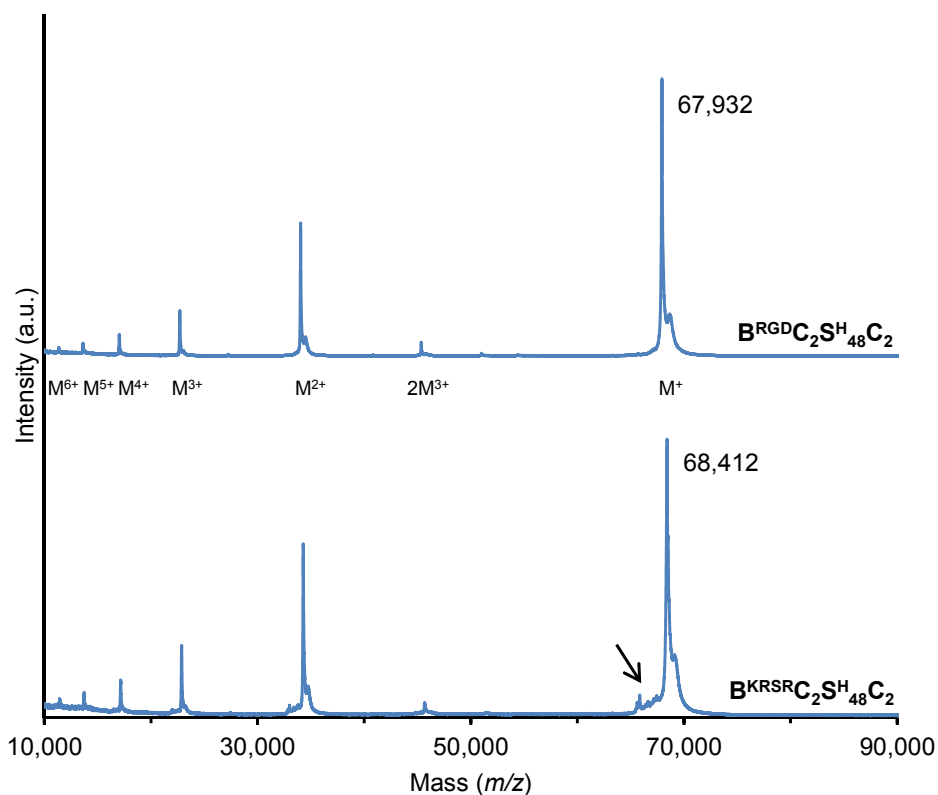


Figure 8.3. MALDI-TOF of purified $\mathbf{B}^{\text{RGDC}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ and $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$. Molecular ions are indicated. The arrow points to a minor shoulder discussed in the text.

8.3.3. Material characterization

AFM and rheology were performed to compare the fibril formation and mechanical properties of $\mathbf{B}^{\text{RGDC}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ and $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ with those previously established for $\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ [36, 37].

For AFM (Figure 8.4), dilute solutions in 10 mM HCl were prepared of the unmodified $\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ protein, and of a 1:1 mixture of the functionalized proteins $\mathbf{B}^{\text{RGDC}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ and $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$. After raising the pH to a physiological level of 7.4 and 24 h of incubation, both samples showed self-assembled fibrils in AFM with a length in the micrometer range, as described previously for $\mathbf{C}_2\mathbf{S}^{\text{H}}_{48}\mathbf{C}_2$ [36, 37]. This indicates that the fibril-forming capacity was not affected by functionalization of the molecules.

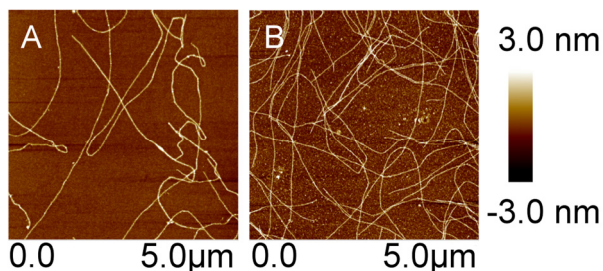


Figure 8.4. AFM of 0.1% (w/v) protein solutions after 24 h of incubation at pH \sim 7.4 and room temperature: (A) $C_2S^{H_{48}}C_2$, and (B) $B^{KRSCR}C_2S^{H_{48}}C_2$ / $B^{RGD}C_2S^{H_{48}}C_2$ mixed in a 1:1 ratio.

A rheological study (Figure 8.5) revealed that gel formation is as well relatively unaffected by the addition of functional domains to $C_2S^{H_{48}}C_2$. After adjusting the pH to a physiological level of 7.4, 2% (w/v) solutions of $C_2S^{H_{48}}C_2$, $B^{RGD}C_2S^{H_{48}}C_2$, and $B^{KRSCR}C_2S^{H_{48}}C_2$ all developed storage moduli in the range of 500-1,500 Pa. The observed differences in the final storage modulus and in the kinetics of gel formation can be attributed to batch-to-batch variations. Also, minor differences in pH adjustment could play a role, in view of the known influence of pH on the storage modulus of $C_2S^{H_{48}}C_2$ [36]. All proteins achieved a storage modulus above 200 Pa already after 1-2 h, and reached a plateau after 5 h. Gelation, defined as the crossover point of storage modulus and loss modulus, occurred within minutes after adjusting pH (Appendix Figure A8.1). As has been shown for $C_2S^{H_{48}}C_2$ [36], also $B^{RGD}C_2S^{H_{48}}C_2$ and $B^{KRSCR}C_2S^{H_{48}}C_2$ retained the ability to quickly recover after breaking, reaching storage moduli of 58 - 83% of the initial values.

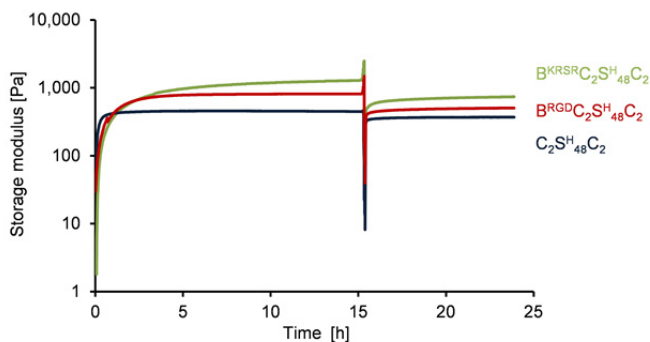


Figure 8.5. Rheometry of 2% gels of $C_2S^{H_{48}}C_2$, $B^{RGD}C_2S^{H_{48}}C_2$ and $B^{KRSCR}C_2S^{H_{48}}C_2$. Development of storage moduli as a function of time during gel formation, and recovery after strain-induced breakage of the gel at $t \sim 15$ h.

8.3.4. Cell culture

Scaffolds for two-dimensional cell culture studies were obtained by preparing 2% (w/v) gels with six different protein compositions, as listed in Table 8.1. This design allowed to determine the influence of the incorporated functional domains on cell behavior. By mixing $\mathbf{B}^{\text{RGD}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ and $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$, hydrogels are obtained that possess both RGD and KRSR motifs. By mixing $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ with either $\mathbf{B}^{\text{RGD}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ or $\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$, a reduced content of these cell-adhesive domains is obtained.

Table 8.1. Protein composition of 2% (w/v) cell culture scaffolds

Scaffold	Relative composition		
	$\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$	$\mathbf{B}^{\text{RGD}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$	$\mathbf{B}^{\text{KRSR}}\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$
0R/0K	100%	-	-
50R	50%	50%	-
100R	-	100%	-
50K	50%	-	50%
100K	-	-	100%
50R/50K	-	50%	50%

As we aim at developing materials for future use as bone cell scaffolds, and investigate the influence of KRSR as bone cell-specific adhesive motifs, MG-63 osteoblastic cells were used in the cell culture studies. The influence of the active domains on early cell response was studied by evaluating cell adhesion (Figure 8.6). As compared to the unmodified 0R/0K gel, the number of adherent cells was significantly higher on the functionalized gels, except for 50K. The simultaneous presence of both RGD and KRSR did not have a synergistic effect on cell adhesion.

The largest variation in the number of adherent MG-63 cells was found in unmodified 0R/0K. Apparently, the distribution of cells on the surface of the scaffold was not uniform. In accordance with our previous study, where rat bone marrow mesenchymal stem cells (MSC) were grown on nonfunctionalized $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ [37], this suggests that scaffolds containing $\mathbf{C}_2\mathbf{S}^{\text{H}_{48}}\mathbf{C}_2$ without adhesive domains are not particularly favorable for cell attachment. In contrast, the smallest variation was found for the 100R scaffold, which indicates a consistent positive effect from the RGD domains.

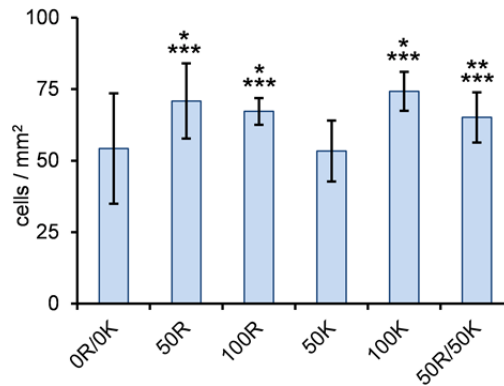


Figure 8.6. Number of cells adherent to the scaffold per mm² after 4 h of incubation. * $p < 0.05$, compared to 0R/0K; ** $p = 0.07$ compared to 0R/0K; *** $p < 0.05$ compared to 50K; ($n = 3$).

The effect of addition of RGD and KRSR domains on cell metabolic activity, as a measure of viability and possible proliferation, was tested with the alamarBlue assay (Figure 8.7). In this assay, the oxidized reagent resazurin is added to the culture medium, which after entering viable cells is reduced to resorufin by mitochondrial enzymes. This reduction can be observed as a change in color [48].

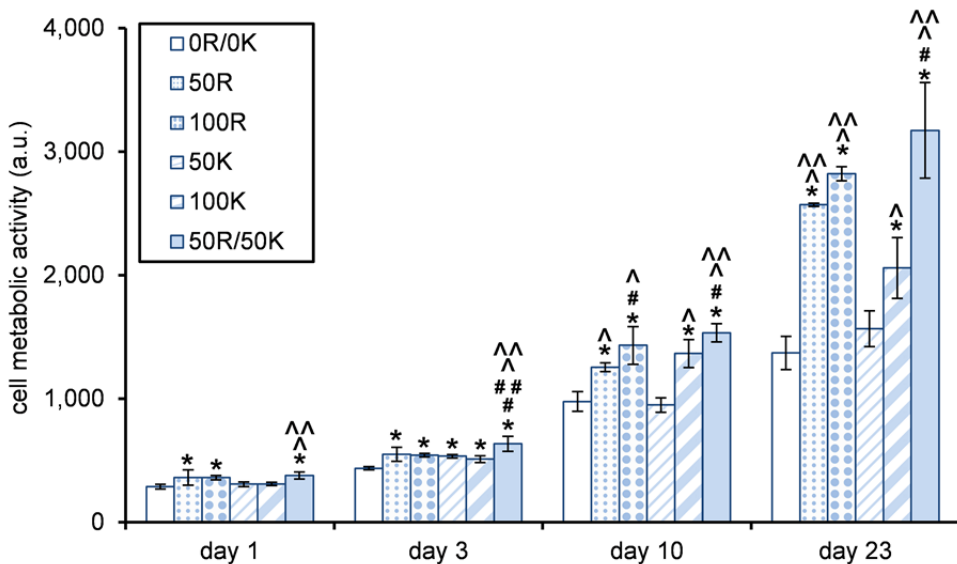


Figure 8.7. Cell metabolic activity as determined by the alamarBlue assay. * significantly higher compared to 0R/0K; # significantly higher compared to 50R; ## significantly higher compared to 100R; ^ significantly higher compared to 50K; ^^ significantly higher compared to 100K; ($n = 3$, $p < 0.05$).

Already after one day of cell culture, significant differences in cell activity were observed. Cells on the 50R, 100R, and 50R/50K scaffolds (i.e., scaffolds containing RGD domains) were metabolically more active in comparison with unmodified material. On day 23, 50R and 100R showed significantly more impact on cell activity than 50K and 100K, respectively, suggesting a stronger influence of RGD than of KRSR domains. Nonetheless, a synergistic effect of the scaffold modified with both functionalities (KRSR and RGD) became apparent from day 1 onward. With significantly higher metabolic activity of cells on 50R/50K compared to most of the other gel variants at each time point, there is a clear tendency for superior metabolic responses to our material when it contains both functionalities.

Cell survival was assessed after 1 and 3 days with a LIVE/DEAD assay. As visible in Figure 8.8, for all scaffold types, only a few individual cells were identified as dead (red-stained nuclei), and the rest remained alive (green cells). For $C_{2}S^{H_{48}}C_2$, this result is in agreement with previous findings for scaffolds prepared from that protein [37]. As expected, modification with RGD and/or KRSR motifs did not negatively affect the cytocompatibility of the gels.

The LIVE/DEAD images (Figure 8.8) also revealed changes over time in terms of cell morphology and cell distribution. Already on day 1, cells seeded on the scaffolds containing RGD motifs (50R, 100R, 50R/50K) started to show a more stretched morphology. On day 3, some cells with elongated shape were observed also on the gels exclusively containing KRSR domains (50K, 100K), although to a lesser extent than on scaffolds containing RGD. On day 3, the scaffold containing both RGD and KRSR motifs (50R/50K) yielded the highest degree of confluency and thus seemed most profitable for the cells. As of day 10, these samples showed a fully confluent layer. Also the 100K scaffold highly favored cell spreading and formation of a confluent layer on days 10 and 23. Interestingly, 50K scaffolds instead led to formation of clusters, and did not show cell spreading until day 23. A similar overall trend as apparent in Figure 8.8 was observed with fluorescent phalloidin/DAPI staining for the visualization of actin filaments and nuclei (Appendix Figure A8.2).

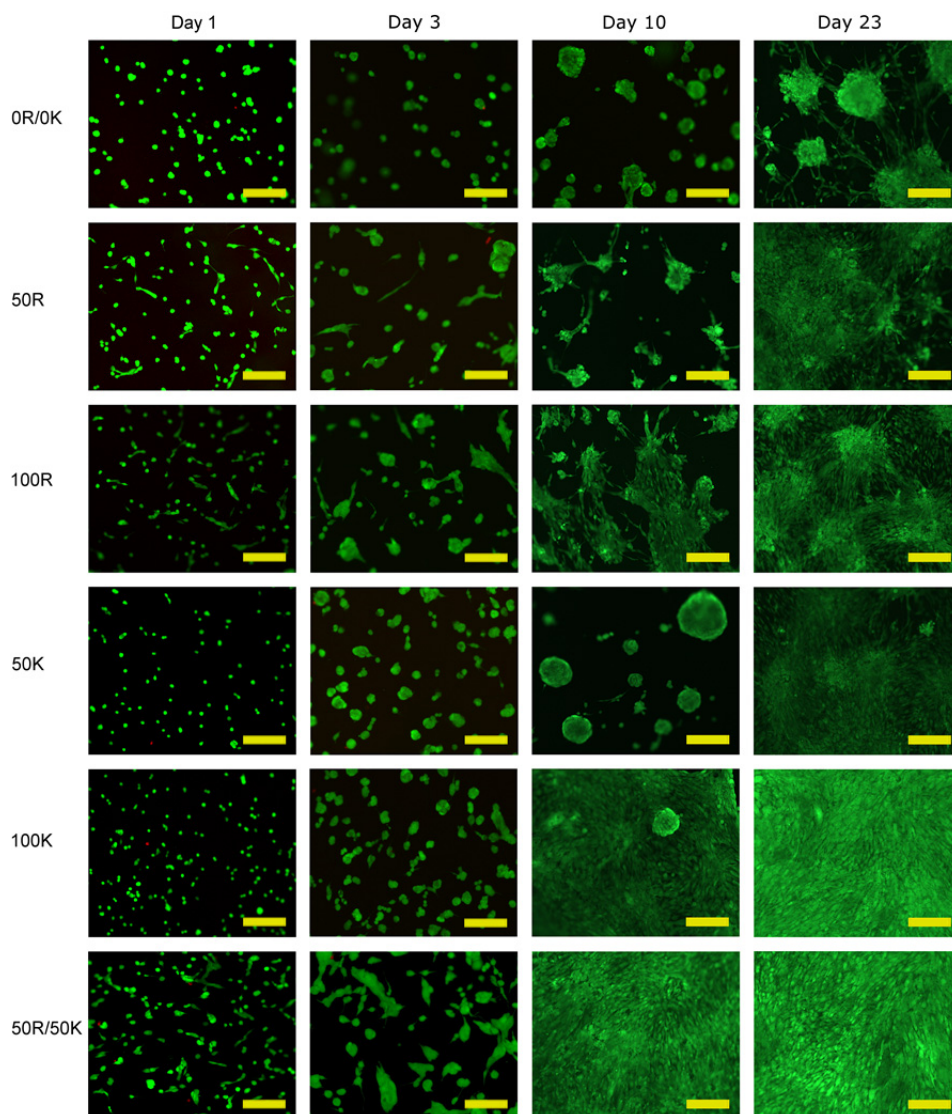


Figure 8.8. Cell survival, morphology, and distribution as monitored by the LIVE/DEAD assay. Viable cells show green fluorescence, whereas dead cells appear red (tested on days 1 and 3 only). Scale bar = 200 μm .

8.4. Discussion

In previous work, we have shown that the $\text{C}_2\text{S}^{\text{H}_{48}}\text{C}_2$ protein-based polymer can form self-assembling, noncytotoxic fibrillar gels allowing to maintain viable cells for at least 21 days [37]. However, the cells did not reveal a spread morphology

over time, and did not proliferate to form a confluent layer. To further develop our material toward use as a scaffold in bone cell culture, we created in this study two variants of the protein with a genetic design that includes either integrin-binding RGD motifs, or proteoglycan-binding bone cell-specific KRSR motifs.

The polymers were efficiently produced as secreted proteins in *P. pastoris* at up to 3.5 g/L of cell-free broth. The possibility to obtain multiple grams of these polymers from benchtop bioreactors readily allows their use in preclinical research. The **B^{RGD}C₂S^{H₄₈}C₂** polymer produced was completely monodisperse. Also **B^{KRSR}C₂S^{H₄₈}C₂** was found to be nearly intact, apart from a very minor level of proteolytic cleavage in the KRSR motif C-terminal of Lys-Arg, and probably also C-terminal of Ser-Arg. As described in Results, proteolytic processing of these motifs in the N-terminal random-coil C₂ block by the endogenous Kex2 protease of *P. pastoris* had been anticipated. The fact that this protease is responsible for processing at Lys-Arg of the α -mating factor prepro peptide, typically used to drive high-level secretion of heterologous proteins in this host, only attests to the enzyme's processivity and/or abundance. In our sequence design, we therefore attempted to particularly minimize cleavage to the C-terminal side of the canonical Lys-Arg motif in **KRSR** by placing expectedly unfavorable residues at position -4 and -3, relative to the scissile bond. Because only a negligible level of degradation was observed, this approach was apparently remarkably effective. To our knowledge, this represents the first reported case where Kex2 cleavage of a dibasic Lys-Arg motif in a recombinant protein was largely prevented by the careful design of the preceding sequence context. Importantly, the essentially monodisperse nature of the polymers adds to the level of control gained over the properties of the scaffolds produced from them.

By mixing nonfunctionalized **C₂S^{H₄₈}C₂** with **B^{RGD}C₂S^{H₄₈}C₂** and/or **B^{KRSR}C₂S^{H₄₈}C₂** in a chosen ratio (Table 8.1), tailored fibrillar scaffolds for cell culture were obtained with different concentrations of the functional groups, independent of the total polymer concentration, and thus, in principle independent of properties such as nominal pore size and storage modulus. The storage moduli of gels prepared from either of the functionalized proteins were indeed in the same range as that of **C₂S^{H₄₈}C₂**, also in accordance with the observation that functionalized and nonfunctionalized proteins showed similar fibril-forming properties in AFM. Apparently, the functional domains have little or no effect on the physico-chemical

characteristics of the proteins. This suggests that also other domains could be incorporated in the future, allowing convenient tuning of multiple functionalities by mixing. $\mathbf{B}^{\text{RGD}}\text{C}_2\text{S}^{\text{H}_{48}}\text{C}_2$ and $\mathbf{B}^{\text{KRSR}}\text{C}_2\text{S}^{\text{H}_{48}}\text{C}_2$ gels also retained the self-healing characteristics of unmodified $\text{C}_2\text{S}^{\text{H}_{48}}\text{C}_2$ [37]. This may be a useful property for potential biomedical use, allowing recovery of the gels if inadvertently broken during *in situ* application or afterwards. While the mixing approach allows to instantaneously prepare hybrid *scaffolds* containing both RGD and KRSR domains, we do not claim that hybrid *fibrils* are certainly formed. Stochastically, however, this will most likely be the case, as each fibril consists of thousands of polymer molecules. To the cells it is probably not relevant whether both domains are present in the same or separate fibrils, as each cell is in contact with countless nanofibrils. In either case there will likely be a certain optimal adhesive domain content, which can be reproduced at will by mixing the proteins in the appropriate ratio.

We previously observed that proliferation and mineralization of primary rat bone MSC were relatively low on nonfunctionalized $\text{C}_2\text{S}^{\text{H}_{48}}\text{C}_2$ in comparison with a collagen control, and anticipated that the cell response may be improved by functionalization of the polymer with cell-adhesive motifs [37]. Indeed, cell adhesion, metabolic activity, and spreading studied here using MG-63 osteoblastic cells were all enhanced on the scaffolds containing either of the functional domains (KRSR or RGD). RGD domains had a slightly more pronounced influence than KRSR. Based on similar observations, Rezania et al. [49] suggested that the principle binding domain is RGD, and that heparin-binding sites are secondary.

A clear synergistic effect on cell metabolic activity and spreading was apparent from day 1 onwards when RGD and KRSR domains were simultaneously present in the scaffolds. Incorporating both integrin-binding and heparin-binding domains may create conditions more similar to the environment of the multifunctional native ECM and may thus be more profitable for the cells. Couchman and Woods suggested that integrins and proteoglycans strongly cooperate during regulation of cell adhesion and spreading [50].

Interestingly, although cell activity and spreading in our study clearly benefited from the presence of both functional domains, no synergistic effect was seen for adhesion. Possibly, proteoglycan-binding motifs influence the cell response only after initial adhesion and particularly play a role in mediating cell

spreading [21]. Similarly, cell adhesion, activity, and spreading in our study did not appear to have a simple causal relation to the degree of confluency. As mentioned above, RGD had a slightly more pronounced influence than KRSR on cell adhesion, activity, and spreading. However, the highest degree of confluency on days 10 and 23 was seen on scaffolds 100K and 50R/50K.

8.5. Conclusions

We described the design and production of functionalized variants of the genetically engineered polymer $C_2S^{H_{48}}C_2$, and evaluated the performance of osteoblastic cells on self-assembled fibrillar hydrogels formed by these proteins. Functionalization involved incorporation of either integrin-binding (RGD) or bone cell-specific heparin-binding (KRSR) sequences into the genetic design. Both proteins, efficiently secreted by *Pichia pastoris* at g/L levels, were essentially monodisperse, and formed self-supporting gels under physiological conditions. Control over the overall concentration of functional groups, independently of the total polymer concentration, was readily achieved by mixing functionalized and nonfunctionalized proteins in various ratios. Functionalization promoted attachment, metabolic activity, and spreading of osteoblastic cells. The combined cell culture studies suggest that the scaffold containing both bioactive domains was most profitable. Overall, these results hold promise for further development of the material toward bone tissue engineering applications.

Acknowledgments

The research for this paper was financially supported by the Netherlands Institute for Regenerative Medicine (NIRM). Financial support to M. W. T. Werten, I. M. Storm, F. A. de Wolf and M. A. Cohen Stuart is provided by ERC Advanced Grant 267254 'BioMate'. The authors thank Prof. E. Pamuła for the cooperation with AGH University of Science and Technology and C. J. Slingerland for assistance with AFM imaging.

References

- Patterson J, Martino MM, Hubbell JA (2010). *Mater Today* 13, 14-22.
- Sengupta D, Heilshorn SC (2010). *Tissue Eng Part B Rev* 16, 285-293.
- Kim SH, Turnbull J, Guimond S (2011). *J Endocrinol* 209, 139-151.
- Cai R, Nakamoto T, Kawazoe N, Chen G (2015). *Biomaterials* 52, 199-207.
- Jia X, Kiick KL (2009). *Macromol Biosci* 9, 140-156.
- Shin H, Jo S, Mikos AG (2003). *Biomaterials* 24, 4353-4364.
- Langer R, Tirrell DA (2004). *Nature* 428, 487-492.
- Drury JL, Mooney DJ (2003). *Biomaterials* 24, 4337-4351.
- Geckil H, Xu F, Zhang X, Moon S, Demirci U (2010). *Nanomedicine* 5, 469-484.
- Zhu J, Marchant RE (2011). *Expert Rev Med Devices* 8, 607-626.
- Gibbs DMR, Black CRM, Dawson JI, Oreffo ROC (2016). *J Tissue Eng Regen Med* 10, 187-198.
- Agarwal R, García AJ (2015). *Adv Drug Deliv Rev* 94, 53-62.
- Short AR, Koralla D, Deshmukh A, Wissel B, Stocker B, Calhoun M, Dean D, Winter JO (2015). *J Mater Chem B* 3, 7818-7830.
- Tseng TC, Tao L, Hsieh FY, Wei Y, Chiu IM, Hsu SH (2015). *Adv Mater* 27, 3518-3524.
- Ruoslahti E (1996). *Annu Rev Cell Dev Biol* 12, 697-715.
- Cardin AD, Weintraub HJR (1989). *Arteriosclerosis* 9, 21-32.
- Dee KC, Andersen TT, Bizios R (1998). *J Biomed Mater Res A* 40, 371-377.
- Dettin M, Conconi MT, Gambaretto R, Pasquato A, Folini M, Di Bello C, Parnigotto PP (2002). *J Biomed Mater Res A* 60, 466-471.
- Hasenbein ME, Andersen TT, Bizios R (2002). *Biomaterials* 23, 3937-3942.
- Palchesko RN, Romeo JD, McGowan KA, Gwalt ES (2012). *J Biomed Mater Res A* 100A, 1229-1238.
- Sawyer AA, Hennessy KM, Bellis SL (2007). *Biomaterials* 28, 383-392.
- Schuler M, Hamilton DW, Kunzler TP, Sprecher CM, de Wild M, Brunette DM, Textor M, Tosatti SGP (2009). *J Biomed Mater Res B* 91B, 517-527.
- Heuts J, Salber J, Goldyn AM, Janssen R, Moller M, Klee D (2010). *J Biomed Mater Res A* 92A, 1538-1551.
- Kim JW, Ki CS, Park YH, Kim HJ, Um IC (2010). *Macromol Res* 18, 442-448.
- Bell BF, Schuler M, Tosatti S, Textor M, Schwartz Z, Boyan BD (2011). *Clin Oral Implant Res* 22, 865-872.
- Shin JH, Lee JW, Jung JH, Cho DW, Lim G (2011). *J Mater Sci* 46, 5282-5287.
- Broggini N, Tosatti S, Ferguson SJ, Schuler M, Textor M, Bornstein MM, Bosshardt DD, Buser D (2012). *J Biomed Mater Res A* 100A, 703-711.
- Maskarinec SA, Tirrell DA (2005). *Curr Opin Biotechnol* 16, 422-426.
- Nettles DL, Chilkoti A, Setton LA (2010). *Adv Drug Deliv Rev* 62, 1479-1485.
- Qiu W, Huang Y, Teng W, Cohn CM, Cappello J, Wu X (2010). *Biomacromolecules* 11, 3219-3227.
- Spiess K, Lammel A, Scheibel T (2010). *Macromol Biosci* 10, 998-1007.
- Romano NH, Sengupta D, Chung C, Heilshorn SC (2011). *Biochim Biophys Acta* 1810, 339-349.
- Gomes S, Leonor IB, Mano JF, Reis RL, Kaplan DL (2012). *Prog Polym Sci* 37, 1-17.
- Werkmeister JA, Ramshaw JAM (2012). *Biomed Mater* 7, 012002.
- Yan Y, Martens AA, Besseling NAM, de Wolf FA, de Keizer A, Drechsler M, Stuart MAC (2008). *Angew Chem Int Ed* 47, 4192-4195.
- Golinska MD, Włodarczyk-Biegun MK, Werten MWT, Cohen Stuart MA, de Wolf FA, de Vries R (2014). *Biomacromolecules* 15, 699-706.
- Włodarczyk-Biegun MK, Werten MWT, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Kamperman M, Cohen Stuart MA (2014). *Acta Biomater* 10, 3620-3629.

38. Martens AA, Portale G, Werten MWT, de Vries RJ, Eggink G, Cohen Stuart MA, de Wolf FA (2009). *Macromolecules* 42, 1002-1009.
39. Razzokov J, Naderi S, van der Schoot P (2014). *Soft Matter* 10, 5362-5374.
40. Zhao B, Cohen Stuart MA, Hall CK (2016). *Soft Matter* 12, 3721-3729.
41. Raghavan SR, Douglas JF (2012). *Soft Matter* 8, 8539-8546.
42. Pierschbacher MD, Ruoslahti E (1984). *Nature* 309, 30-33.
43. Whyteside G, Alcocer MJC, Kumita JR, Dobson CM, Lazarou M, Pleass RJ, Archer DB (2011). *PLoS One* 6, e22692.
44. Niu B, Huang Y, Zhang S, Wang D, Xu H, Kong D, Qiao M (2012). *Protein Expr Purif* 83, 92-97.
45. Julius D, Brake A, Blair L, Kunisawa R, Thorner J (1984). *Cell* 37, 1075-1089.
46. Bevan A, Brenner C, Fuller RS (1998). *Proc Natl Acad Sci USA* 95, 10384-10389.
47. Suzuki Y, Ikeda N, Kataoka E, Ohsuye K (2000). *Biotechnol Appl Biochem* 32, 53-60.
48. Nociari MM, Shalev A, Benias P, Russo C (1998). *J Immunol Methods* 213, 157-167.
49. Rezania A, Healy KE (1999). *Biotechnol Prog* 15, 19-32.
50. Couchman JR, Woods A (1999). *J Cell Sci* 112, 3415-3420.

Appendix

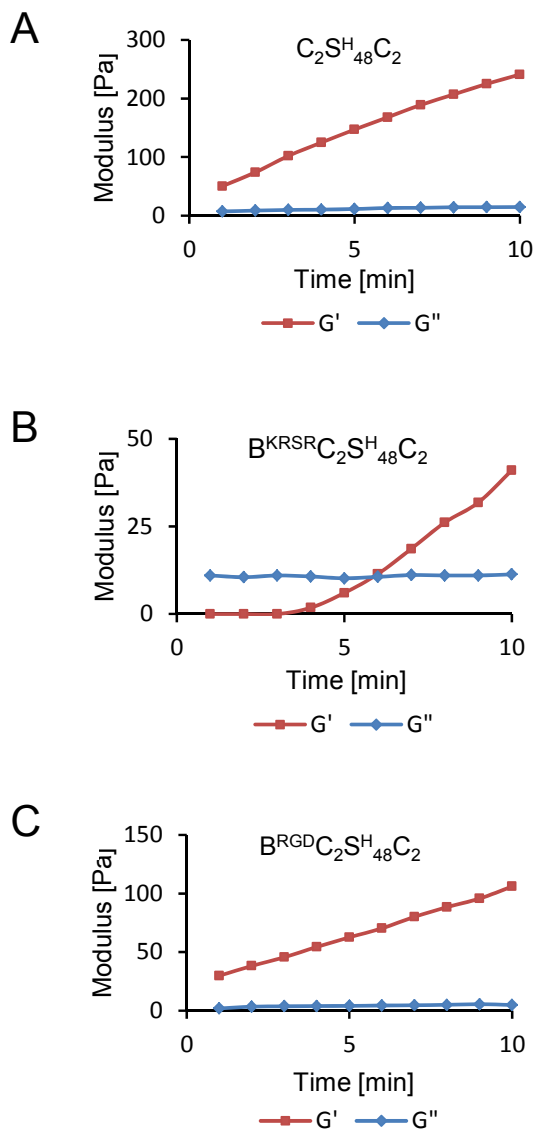


Figure A8.1. Rheometry of 2% gels of $C_2S^{H_{48}}C_2$ (A), $B^{KRSC_2}S^{H_{48}}C_2$ (B) and $B^{RGDC_2}S^{H_{48}}C_2$ (C). Development of storage and loss modulus as a function of time during the first 10 min of the measurement. Gelation occurs when the storage modulus (G') exceeds the loss modulus (G''). The time involved in titration of the pH to 7.4, which induces self-assembly, and loading of the sample into the rheometer prior to the actual measurement is less than 5 min.

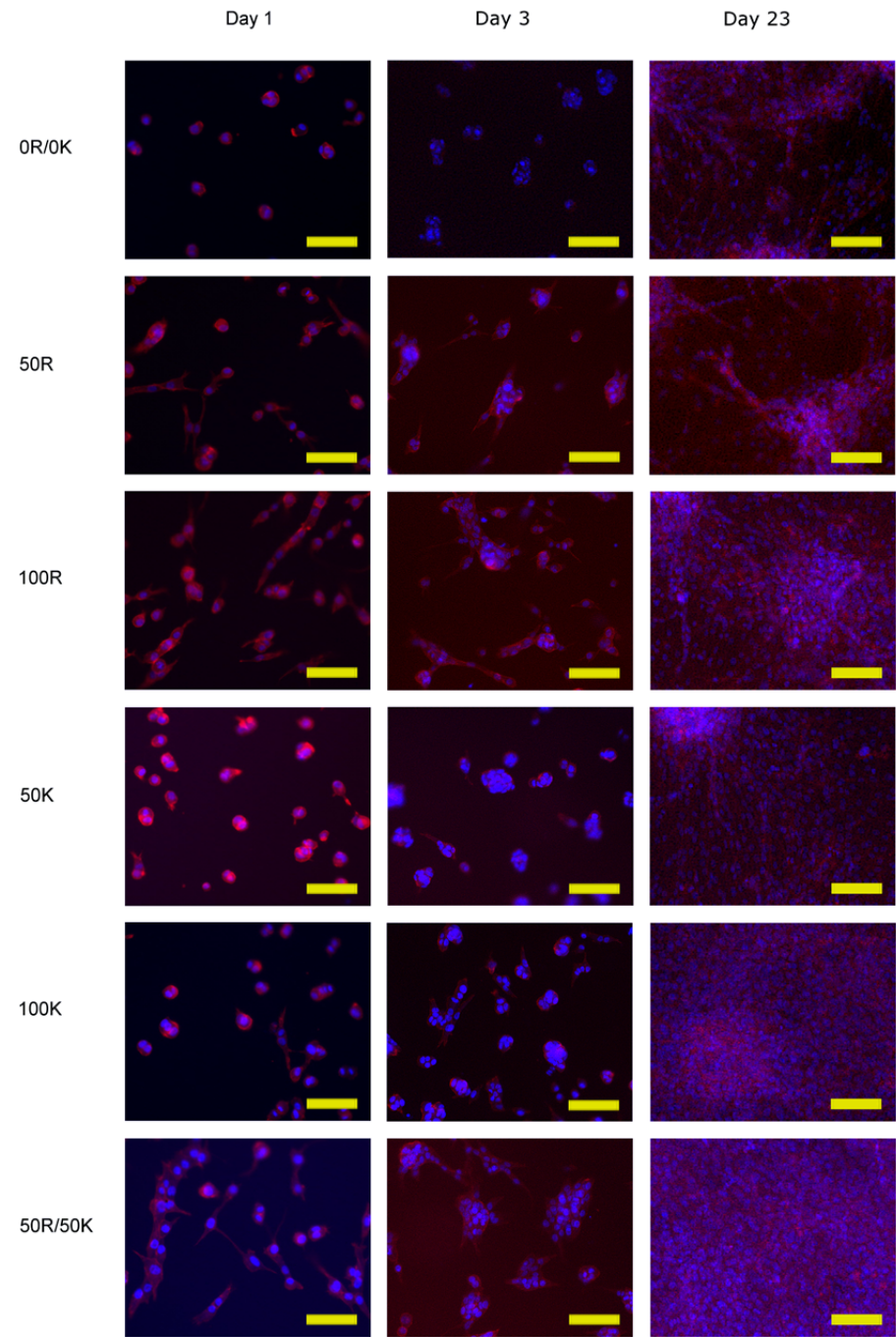


Figure A8.2. Cell morphology of fluorescently stained MG-63 cells on 0R/0K, 50R, 100R, 50K, 100K, and 50R/50K scaffolds after 1, 3, and 23 days of cell culture. Red: actin. Blue: nuclei. Scale bar = 100 μm .

Chapter 9

General discussion

Submitted in modified form as part of a review article:
Werten MWT, Eggink G, Cohen Stuart MA, de Wolf FA.

In the preceding chapters we have seen that a variety of functional protein polymers could be produced in *P. pastoris* at high yields. This includes gelatin-like proteins (Chapters 2 to 4), silk-like proteins (Chapter 5), and block copolymers capable of self-assembling into supramolecular structures (Chapters 6 to 8). Particularly this last category of proteins yielded sophisticated materials with novel features that are of potential interest for biomedical applications.

In this chapter we will place these results in the wider context of protein polymers produced in *P. pastoris*, both by our laboratory and others. We will then proceed to discuss problems that may occur when producing these atypical proteins, particularly in secretory mode, also drawing from what is generally known about protein expression bottlenecks. Possible solutions to these issues will be addressed. Finally, we will provide overall conclusions and an outlook to the future.

9.1. Protein polymers produced in *P. pastoris*

Table 9.1 provides an overview of protein polymers produced in *P. pastoris*, which will be discussed in this section. For each protein type, if applicable, we will first describe the properties of the natural polymer, and shortly outline the historic context in terms of production in recombinant *E. coli*. Completely *de novo*-designed block copolymers produced by our laboratory in *P. pastoris* have been summarized in Section 9.1.5.

Table 9.1. Protein polymers produced in *P. pastoris* ^a

Protein	MW (kDa) ^b	Targeting	Culture	Yield (g/L) ^c	Reference
<i>Silk-like proteins</i>					
<i>Nephila clavipes</i> MaSp1 mimics	64 / 127 ^d	I	S / B	0.2 – 1	[1]
<i>N. clavipes</i> MaSp1 mimics	64 / 127	E ^e	N/A	≤6 µg/OD ₆₀₀	[2]
<i>N. clavipes</i> MaSp1 mimics ±COL	~15 - 30	E ^e	S	N/A	[3]
<i>Nephila</i> sp. MaSp1 / MaSp2 mimics	94 / 113	I	B	N/A	[4, 5]
<i>N. clavipes</i> MaSp1 / MaSp2 mimics +NTD +CTD	44	E	B	N/A	[6]
<i>Euprosthrops australis</i> MaSp1 mimic +CTD	32	E ^e	S	N/A	[7]
Silk-like octapeptide repeats	28 - 32	E	B	0.9 - 2.9 ^f	Chapter 5
<i>Gelatins (nonhydroxylated)</i>					
Murine α1(I) and α1(III) fragments	21 - 53	E	B	2 - 15 ^f	Chapter 2
Artificial hydrophilic gelatin CP ₄	37	E	B	3 - 6 ^f	Chapter 4
Human α1(I), α1(II), and α1(III) fragments	5 - 90	E	B	N/A	[8]
Human α1(I) fragment	9	E	B	1.5 ^f	[9]
<i>Collagen (hydroxylated)</i>					
Human type III	~400 ^g	E ^e	S	15 mg/L	[10]
Human type I	~400 ^g	E ^e	B	0.2 - 0.5	[11]
Human type I, II, and III	~400 ^g	E ^e	B	0.2 - 0.6	[12]
Human type I, II, and III	~400 ^g	E ^e	B	0.7 - 1.5	[8]
Human α1(I) fragments	36 - 226 ^g	E ^e	S	N/A	[13]
<i>Chondrosia reniformis</i> nonfibrillar collagen	210 ^g	E ^e	S	N/A	[14]
<i>Elastin-like proteins</i>					
V ₅ A ₂ G ₃ -90, V ₅ A ₂ G ₃ -40, V ₅ L ₂ G ₃ -40	16 - 36	E	B	0.2 - 0.8 ^{f,h}	[15, 16]
V ₄ E ₁ -105	47	E	S	2.5 mg/L ^h	[17]
<i>Other</i>					
Human titin PEVK-region mimic	54	E	B	>1 ^h	[18]
Various block copolymers	38 - 115	E	B	0.4 - 6.5 ^f	Section 9.1.5

^a Abbreviations: MW, molecular weight; MaSp, major ampullate spidroin; COL, amino-terminal *Meloidogyne incognita* collagen helical domain; NTD, amino-terminal spidroin domain; CTD, carboxy-terminal spidroin domain; X_iY_jZ_k-n, elastin-like protein with n Val-Pro-Gly-Xaa-Gly pentapeptides, where the capital letters specify the guest residue Xaa in the ELP sequence, and the subscripts indicate the number of each corresponding guest residue encoded by the DNA monomer fragment [19]; I, intracellular targeting; E, extracellular targeting; S, shake flask; B, bioreactor; N/A, not applicable.

^b Refers to the expected molecular weight of the mature protein. For some articles this was calculated or estimated on the basis of the genetic details provided.

^c Except where noted, yields are given as the amount of heterologous protein expressed per L of culture (for some of the cited articles this definition was assumed).

^d Owing to gene rearrangements during transformation, proteins in the size range of 48-261 kDa were obtained.

^e Retained or largely retained intracellularly despite extracellular targeting (not clearly specified for *C. reniformis* collagen in *P. pastoris*).

^f Yield per L of cell-free broth.

^g Molecular weight of the collagen trimer (trimer formation not demonstrated for *C. reniformis* collagen in *P. pastoris*).

^h Yield of purified recombinant protein.

9.1.1. Silk-like proteins

9.1.1.1. Spider silks

Spider silks represent a fascinating family of fibrous proteins involved in prey capture and reproduction. At least seven types of silks are found in the golden orb-weaver spider, *Nephila clavipes*, each with different functions such as the orb web frame, the adhesive spiral, material for prey wrapping, etc. [20]. Dragline silk in particular is of interest, as it is thought to be nature's toughest polymer. On an equal weight basis it is tougher than steel and Kevlar, combining strength and extensibility [21]. The mechanics of silk fiber strength have been extensively studied [22-24].

Dragline silk is secreted by specialized cells of the spider's major ampullate gland, and contains two proteins, MaSp1 and MaSp2 for *N. clavipes* (formerly called spidroin 1 and 2) [25]. Both of these silks have a molecular mass of more than 300 kDa, and consist of three domains: a large repetitive middle domain responsible for the mechanical properties of the fiber, and nonrepetitive N-terminal and C-terminal domains [21]. The middle domain consists of repeats of highly similar modular units that each in turn consist of a crystalline and a less ordered domain (Figure 9.1). Consensus sequences of these units were derived from cDNA sequences: GGAGQGGYGGLGGQGAGRGLGGQ(GA)₂A₅ (MaSp1) and (GPGGYGPGQQ)₂GPSGPSA₈ (MaSp2). The crystalline domains consist of β -sheet-forming poly-Ala (MaSp1 and MaSp2) or poly(Gly-Ala) sequences (MaSp1), and the more elastic domains consist of Gly-Gly-Xaa (MaSp1) or Gly-Pro-Gly-Xaa-Xaa (MaSp2) sequences [20].

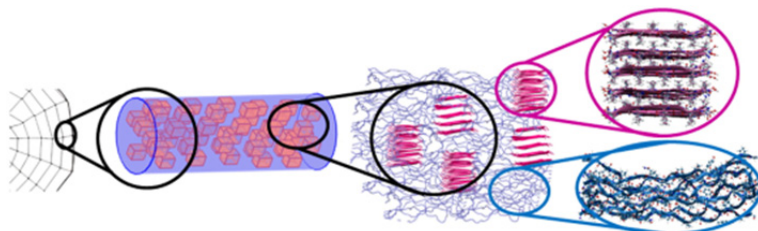


Figure 9.1. Schematic representation of spider silk architecture at macro- to nanoscale. Spider silk fibers consist of crystalline β -sheet domains (pink), and disordered elastic domains (blue). Reprinted from Cetinkaya et al. [23] with permission from Elsevier.

The remarkable mechanical features of spider silk, as well as the biocompatibility of silk-based materials [26, 27], have sparked the interest to create

recombinant spider silks. Several studies report the production of synthetic dragline silks in *E. coli*, consisting of repeated spidroin consensus sequences. Prince et al. [28] reported successful expression of *N. clavipes* dragline silk, in the size range of 15 to 41 kDa. MaSp1 oligomers with different repeat numbers were produced, as well as three block copolymers of MaSp1 and MaSp2. Typical yields were in the range of 2 to 15 mg/L of culture. Circular dichroism showed that indeed the proteins had significant β -sheet structure in aqueous solutions. Lewis et al. [29] performed similar work, and produced up to 32 repeats (112 kDa) of the MaSp2 consensus sequence. A 16 unit repeat was recovered at up to 10 mg/g wet cell weight, with typical yields of 1-2 mg/g. Fahnestock and Irwin [30] of DuPont have produced 65-163 kDa analogs of MaSp1 and MaSp2 in *E. coli*. The authors oligomerized relatively long (>300 bps) DNA monomers, each encoding three to four variants of either consensus repeat, in an effort to minimize the repetitiveness of the overall gene. Product yield in shake flask cultures was ~300 mg/L, or ~15% of the total protein. The fraction of fully intact products was limited to 48% for an 8-mer of the (codon-optimized) monomer and only 8.5% for a 16-mer, probably as a result of premature translation termination. Genetic instability also occurred, which produced internal gene deletions in roughly 10% of the genes in the population during a typical experiment, and a minor degree of gene expansions.

The same group also describes the intracellular production of MaSp1 analogs in *P. pastoris* [1], which represents the first example of the use of this yeast for the production of a protein polymer. Contrasting the authors' results using *E. coli*, genes of more than 3,000 codons could be expressed without truncated synthesis. Multicopy strains produced various lengths of the protein, differing by an integral number of repeats of the DNA monomer. Presumably this was due to homologous recombination during transformation, where additional vector copies inserted out of phase into the repetitive silk sequence of the first copy that had already integrated into the genome. To our knowledge, there are no other examples in the literature of such an effect in *P. pastoris*, and also in our laboratory we have never seen this with similarly repetitive sequences. This discrepancy may be due to differences in the transformation procedures used. Whereas Fahnestock and Bedzyk [1] used PEG-mediated spheroplast transformation and targeted for transplacement of the *AOX1* locus, we use electrotransformation and usually target

for insertion into the *his4* locus. The latter approach has a much lower chance of spontaneous multicopy generation [31], and thus less risk of artifacts such as described. Nonetheless, once transformed, the MaSp1 analog-producing strains were genetically very stable, in that there was no alteration of the protein size distribution during growth for at least 100 doublings [1]. Thus, unlike what the authors observed in *E. coli*, internal deletions or expansions of the encoding genes apparently did not occur. Protein purification from bead-milled cells involved precipitation of host proteins consecutively at pH 5 and at 60 °C, followed by differential ammonium sulfate precipitation. The yield in shake flask cultures was in the range of 0.2 - 1 g/L. The authors concluded that *P. pastoris* is a superior host for the production of large spider silks, and possibly also for other long repetitive sequences.

Fahnestock et al. [2] additionally reported secretory production of MaSp1 analogs, using three different secretion signals: the acid phosphatase signal sequence (hereafter Pho1) [32], the prepro region of the *S. cerevisiae* α -factor mating pheromone (hereafter ppaF) [33], and the pre region (signal peptide) of ppaF. In all cases some level of secretion was achieved, but most of the product accumulated intracellularly. The full ppaF leader and its mere signal peptide gave similar amounts of secreted protein, which appeared higher than obtained when using the Pho1 signal peptide. The reported secretory yield was less than 6 μ g of recombinant protein per OD₆₀₀ unit, which at best corresponds to mg/L levels.

Teulé et al. [3] briefly reported on their efforts to produce ~15 kDa proteins consisting of repeated parts of the *N. clavipes* MaSp1 consensus. These proteins were constructed in such a way that the content of the alanine-rich motifs was normal, reduced, or zero. Although details were not provided, the culturing of *P. pastoris* was probably done in shake flasks. Despite the fact that ppaF was used to drive secretion, all product was retained intracellularly. According to Western blot analysis, secretion was achieved only in cases where the spider silk sequence was fused to an ~15 kDa natural nematode-derived collagen sequence (see Section 9.1.2.1. for a description of nontrimeric collagen), which suggests improved overall solubility of the fusion proteins relative to the exclusively silk-like proteins. Protein purification was by ammonium sulfate precipitation or metal ion affinity chromatography. Yields were not reported.

Bogush et al. [4] constructed a synthetic gene monomer containing five typical repeating units selected from *N. clavipes* MaSp1 cDNA. A multimeric gene consisting of nine monomers was then constructed, encoding a 94 kDa MaSp1 analog referred to as 1F9. A 113 kDa MaSp2 analog, referred to as 2E12, was similarly constructed by multimerizing 12 monomers, each consisting of 5 quasi-repeats selected from *Nephila madagascariensis* cDNA. *P. pastoris* transformants were grown in methanol-fed fermentor cultures [5]. The intracellularly targeted proteins were purified from disrupted cells by ion-exchange chromatography and spontaneously formed nanofibrils and micelles of ~1 μm in aqueous solutions [4]. Fibers with a tensile strength of up to 0.15 GPa could be prepared by spinning the protein, dissolved in lithium chloride / formic acid, into an ethanol bath. Although no yields are provided, the authors mention that the production levels of MaSp2 in *P. pastoris* and *S. cerevisiae* were comparable [34]. However, whereas in the latter host 70-80% of MaSp2 was in the water-insoluble cellular fraction, in *P. pastoris* 60% was in the soluble fraction. The proteins are currently being developed into matrices for tissue engineering and drug delivery [4, 35, 36].

Gaines and Marcotte [6] very briefly described the production in *P. pastoris* of a 44 kDa mini-silk based on *N. clavipes* MaSp1. The protein consists of eight MaSp1 consensus repeats, as well as the N- and C-terminal domains, and was targeted for secretion using ppaF. The expected product, besides a minor amount of C-terminally truncated protein, was detected by Western blot analysis in the clarified broth of methanol-induced bioreactor cultures. Yields and possible intracellular retention were not reported. The His-tag-containing product could be purified from the broth using ion-exchange chromatography.

Jansson et al. [7] recently used *P. pastoris* for the production of the spider silk-like protein Z-4RepCT previously produced in *E. coli* [37]. The 32 kDa protein consists of an IgG-binding Z-domain, 4 consecutive silk-like poly-Ala/Gly-rich repeats from *Euprosthenops australis* MaSp1, the C-terminal CT domain from the same spidroin, and a C-terminal (His)₆ tag. The nonrepetitive CT domain had previously been shown to be a prerequisite for spontaneous fiber formation by 4RepCT [38]. Methanol-induced shake flask cultures were used. Although the ppaF signal was employed to achieve protein secretion, much of the protein was retained intracellularly. The secreted protein was partially degraded, which could be minimized to some degree by the use of protease inhibitors during culture and

purification (see also Section 9.2.3.1). It was additionally found to be N-glycosylated in the CT domain at the Asn-Xaa-Ser/Thr site present, as well as O-glycosylated at one or more serines in the sequence LSSPSAVSR, and possibly elsewhere. Probably because of this, secreted Z-4RepCT did not self-assemble into fibers, yet attempts to resolve this by enzymatic deglycosylation failed. The intracellularly retained protein was mostly present in the insoluble fraction, and less degraded than the secreted protein. Protein purified from the soluble intracellular fraction did self-assemble into fibers, but only after enzymatic deglycosylation. Protein purification from cell-free medium or cells disrupted by high pressure homogenization, was done using metal ion affinity chromatography. The protein yield was ambiguously defined, but reportedly similar to that obtained in *E. coli* [7].

9.1.1.2. Silk-like octapeptide repeats

Besides spider silk, also silk from the silk worm *Bombyx mori* has been produced using recombinant DNA technology. Cappello and Ferrari of Protein Polymer Technologies produced repeats of the hexapeptide GAGAGS from *Bombyx* silk fibroin in *E. coli* [39], and showed that the material produced adopted crystalline structures similar to the β -sheet structures of natural silk. The Tirrell group produced repeat proteins consisting of an artificially designed silk-inspired (GA)₃GE octapeptide in *E. coli* [40]. The repeating Ala-Gly dipeptides were mimicked after *Bombyx mori* silk and other alanyl-glycine-rich proteins known to adopt β -sheet structures in the solid state. The choice of three Ala-Gly repeats was based on dimensional restrictions known from the crystallization behavior of synthetic polyamides, and on the fact that a fibrous cross- β protein from lacewing folds with a periodicity of eight amino acids. Glutamic acid was inserted because its large size sterically prevents inclusion into the lamellar interior, its polarity and charge favor strong interaction with the environment, and because it is considered the weakest β -sheet former of all natural amino acids. The protein, containing 36 repeats of the octapeptide, was expressed in *E. coli* and purified from the soluble fraction of the cell lysate by a procedure involving sequential pH adjustments. Purity level and yield were not reported. After dissolving the protein in 70% formic acid and subsequent crystallization in methanol, a crystalline anti-parallel β -sheet structure could be demonstrated. A lamellar structure is formed by lateral

stacking of the sheets, where the polymer chain folds back at the lamellar surfaces, as dictated by the Glu residues. In a later paper [41], the authors studied the effect of amino acid side-chain volume by substituting Glu with Asn, Phe, Ser, Val, or Tyr. They found that, although the basic structure remains the same, the average intersheet stacking distance increased linearly with the volume of the substituting residue.

As described in Chapter 5, we chose the $(\text{GA})_3\text{GE}$ design as a structural framework for the design of pH-responsive polymers for coating purposes. The silk-like nature and glutamate residues together convey self-assembling behavior at low pH. Using the ppαF leader, secretory production was obtained of the 28 kDa EE24 protein with 48 repeats of the $(\text{GA})_3\text{GE}$ motif, referred to hereafter as S^{E}_{48} , in methanol fed-batch bioreactor cultures. The product was fully intact and was produced at ~3 g/L of cell-free broth. The amphiphilic EL28 variant, which has 56 octapeptides where every other glutamate residue is substituted with the highly hydrophobic amino acid leucine, was also monodisperse and produced at ~1 g/L of clarified broth or more. A recently produced variant is S^{H}_{48} , which encodes 48 repeats of the octapeptide $(\text{GA})_3\text{GH}$. It was purified by ammonium sulfate precipitation, resulting in a recovery of ~2 g/L of cell-free broth (unpublished). Our group has used several $[(\text{GA})_3\text{GX}]_n$ variants as a self-assembling module in various block copolymers (Section 9.1.5).

9.1.2. Collagen-like proteins

Although the family of collagens consists of multiple members, we will discuss only the most abundant, fibril-forming, collagens (types I-III, V, and XI). These extracellular matrix proteins provide structure and strength to connective tissues such as skin, bones, tendons, as well as internal organs. Characteristic of collagen is its triple-helical structure (Figure 9.2), where three polypeptide chains in extended helical conformation are coiled together [42]. Each of the three chains consist of about 338 tandemly repeated Gly-Xaa-Yaa triplets, where Xaa and Yaa can be occupied by different amino acids. The occurrence of glycine as every third residue is a prerequisite, because only this smallest of amino acids can fit inside the triple helix interior. The Xaa and Yaa positions are frequently occupied by prolines, which limits rotation of the protein chains [43]. About 50% of the prolines in the Yaa position are posttranslationally modified to 4-hydroxyprolines by the enzyme

peptidyl-prolyl-4-hydroxylase (P4H) [44], which allows the formation of additional hydrogen and water bridges, thereby stabilizing the triple helix at physiological temperature. Each collagen chain also contains N- and C-terminal propeptides, which allow for mutual recognition and in-register alignment of the three chains for subsequent zipper-like folding into the collagen triple helix. After secretion, the propeptides are cleaved off by specific proteases, upon which collagen self-assembles into fibrils [43].



Figure 9.2. Three-dimensional structure of the collagen triple helix. Shown is a trimeric (Gly-Pro-Hyp)₉ collagen model peptide [45], where Hyp is 4-hydroxyproline. PDB ID: 3B0S; visualized using Cn3D [46].

Gelatin is essentially denatured collagen, and is prepared by hot acid or alkaline extraction of animal tissues such as bones and hides [47]. The well-known gelling properties of heated and subsequently cooled gelatin are due to the random formation of triple-helical patches between neighboring molecules.

Both collagen and gelatin are important biomaterials in a variety of medical applications and as a material for the delivery of drugs [8, 48]. However, their animal origin poses risks for allergic reactions or transmission of disease-causing agents. Furthermore, since natural collagen fibrils are extensively crosslinked, the age and physiology of the animals used as source affect the physical properties of the material, leading to high batch-to-batch variation. In the case of gelatins, the extracted material consists of a variable mixture of chemically modified peptides covering a wide mass range [47].

Several groups, therefore, endeavored to produce recombinant alternatives. Early work involved the use of *E. coli*. For example, Goldberg [49] produced a collagen-like sequence (Gly-Pro-Pro)₃₂. Expression levels were not reported but appear to be very low. Similar approaches for the production in *E. coli* of tandemly

repeated short collagen-like sequences were followed by Obrecht [50] and Gardner [51]. Hori [52] produced fragments of natural bovine $\alpha 2$ (I) collagen up to 245 amino acids using *E. coli*. Yields were not reported, but were at least sufficient for use in ELISA tests. It should be noted that, because of the absence of a native P4H in prokaryotes, these fragments cannot form stable triple helices and are thus only collagen-like with respect to their sequence; they are single-chain collagen fragments.

9.1.2.1. Gelatin

As described in Chapter 2, we produced single-chain fragments of murine $\alpha 1$ (I) and $\alpha 1$ (III) collagen, referred to as Col1a1 and Col3a1, respectively. We termed these proteins ‘gelatins’ because they resemble denatured collagen, although they do not actually form gels owing to the absence of endogenous prolyl-4-hydroxylation in *P. pastoris*. A strain with an estimated 15 copies of the vector produced the Col3a1 gelatin at a level of ~15 g/L of cell-free broth, which is among the highest heterologous protein yields reported for *P. pastoris*. Interestingly, no Col3a1 product could be detected when we attempted intracellular production (unpublished). Originally conceived and developed by us for Fujifilm Corporation with a view to photography applications [53-55], the company used this technology as a basis for developing their commercial product for regenerative medicine, cellnest [56-58].

In Chapter 4 we presented a fully *de novo*-designed random coil gelatin P4, hereafter referred to as C^P₄, which consists entirely of the Gly-Xaa-Yaa triplets typical for collagen, but was designed to be highly hydrophilic. It consists of four repeated 99-residue long C^P blocks, and the secreted ~37 kDa protein was completely monodisperse and expressed at 3-6 g/L of clarified broth. The protein acts as a cytophilic protein in human cell culture [59], and shows excellent biocompatibility as a plasma expander [60]. The protein could be efficiently purified from the broth to >98% purity by simple ammonium sulfate precipitation. We found this to be a generally applicable strategy in *P. pastoris* for the purification of gelatins, including Col3a1 and Col1a1 mentioned above, or block copolymers that have C^P₄-derived modules as a random coil constituent (Section 9.1.5). This is because, at least when using basal salts medium with PTM1 trace elements [61], secreted *P. pastoris* host proteins and exopolysaccharides do not precipitate at up to

80% saturation, whereas gelatin-like proteins precipitate efficiently at around 40% saturation. At pilot-scale, we use crossflow membrane filtration instead of centrifugation to harvest the precipitate, which is then washed with ammonium sulfate solution, eluted with water, and subsequently concentrated and desalted using crossflow ultra-/diafiltration (unpublished).

FibroGen Inc. followed the same approach as outlined above for murine Col3a1 and Col1a1 for the production in *P. pastoris* of gelatins based on human $\alpha 1(I)$ collagen chains in the range of 5 to 90 kDa, having different pI from 4.6 to 10.0, as well as $\alpha 1(II)$ and $\alpha 1(III)$ collagen fragments [8]. Most of their studies involve a small gelatin $\alpha 1(I)$ fragment of 8.5 kDa with a pI of 9.4, for use as a stabilizer for biologics such as vaccines [8, 9, 62]. The protein was produced in bioreactors at ~1.5 g/L and purified by ion-exchange chromatography [9].

9.1.2.2. Collagen

To obtain triple-helical collagen, incorporation of hydroxyprolines is required. The Myllyharju group has done extensive work on the production of triple-helical collagen in *P. pastoris* through coexpression of human P4H [10]. P4H is an $\alpha 2\beta 2$ tetramer, where the β subunit is the ER-resident protein disulfide isomerase (PDI). Expression of only the α subunit did not lead to P4H activity, indicating that the α subunit does not form an active enzyme with *P. pastoris* PDI. When both the α and β subunits were expressed, low levels of active tetramer were obtained. However, on simultaneous coexpression also with full-length prepro- $\alpha 1(III)$ collagen chains (including the native signal peptide, as well as N-terminal and C-terminal pro- and telopeptides), 10-fold higher P4H assembly levels were obtained. The collagen molecules were retained intracellularly for 80-90%, and were expressed at a level of ~15mg/L in shake flask cultures. Trimeric collagen was purified from the soluble fraction of cell lysates by digestion with pepsin, which digests propeptides and host proteins but not triple-helical $\alpha 1$ chains, followed by size-exclusion chromatography. The degree of prolyl-4-hydroxylation relative to that of natural type III collagen was ~85%.

To study the nonsecretion of trimeric collagen, the authors used immunoelectron microscopy [63]. This showed that procollagen accumulated in membranous vesicular compartments closely associated with the nuclear membrane. The authors conclude that triple-helical collagen accumulates in the ER

and cannot proceed further down the secretory pathway. In an attempt to solve this, the authors replaced the native human signal peptide of type III collagen with the generally very efficient ppaF secretory leader. However, this led to only a minor increase in the fraction secreted, and was accompanied by a roughly 50% lower total amount of product.

Upon coexpression of the pro α 1 and pro α 2 chains of type I procollagen, correctly self-assembled trimeric type I procollagen was obtained with the correct 2:1 chain ratio [11]. As was expected, a fraction of the molecules also formed α 1 homotrimeric molecules. Chains lacking the N-propeptide likewise correctly formed trimeric collagen, with a 1.5-3-fold increase in yield. Production levels of trimeric procollagen were 200-500 mg/L in bioreactors, where molecules lacking the N-propeptide were at the higher end of this range. As before, the procollagen produced was intracellularly retained. The use of bioreactors with pure oxygen supplementation led to increased hydroxylation levels as compared to the above-mentioned earlier work with shake flasks, and the determined 4-hydroxyproline content matched that of natural type I collagen. Mature collagen produced from the recombinant procollagen molecules by pepsin digestion formed fibrils *in vitro* similarly to native collagen. The group also reported the production of type II collagen [12], and of shortened α 1(I) chains with only the C-terminal propeptide or a C-terminal foldon domain from bacteriophage T4 fibritin for trimerization [13]. See also Section 9.2.6.

A review by FibroGen Inc. [8], a company that partly funded the above work of the Myllyharju group, reports that the initial yield of 15 mg/L originally reported for type III collagen in *P. pastoris* [10] was successfully increased *via*, unspecified, genetic means and process improvements. Collagen types I, II, and III expression levels eventually reached 1.1, 0.7, and 1.5 g/L, respectively. An unusually high fermentation temperature of 32 °C was found to be important to achieve full hydroxylation, although this also lowered the overall collagen yield [48].

Merrett et al. [64] conducted a phase 1 clinical study using cell-free hydrogel implants prepared from Fibrogen's recombinant collagen by carbodiimide crosslinking. The implants were intended as mimics of the largely collagenous corneal extracellular matrix, to stimulate corneal regeneration. Over four years, the implants were stably integrated without the need for immunosuppression, and patients had improved vision from the preoperative level [65]. As one of few

examples of protein polymers that reached the clinical testing stage, this attests to the utility of *P. pastoris* in this field. An advantage of the yeast over *E. coli* in this context is the absence of endotoxins [66] in the former.

Recently, *P. pastoris* has also been used for the production of hydroxylated nonfibrillar collagen from the marine sponge *Chondrosia reniformis* [14]. Both the α and β subunits of P4H, as well as the collagen gene were derived from this organism, and were coexpressed in shake flask cultures of *P. pastoris*. The authors followed the example of Vuorela et al. [10] and replaced the native secretion signal of the β subunit of P4H with the ppaF leader. Hydroxylated collagen could be detected intracellularly, where 36% of all prolines were hydroxylated. Hydroxylation took place both in the Xaa and Yaa position of Gly-Xaa-Yaa triplets, which differs from the Yaa-specific activity of human P4H. Interestingly, the α subunit of P4H alone had an enzymatic activity of 54% relative to the tetramer, suggesting that it can either partially interact with the yeast's PDI, or does not require a β subunit. Collagen trimer formation and yields were not reported.

9.1.3. Elastin-like proteins

Elastin is a major component of the extracellular matrix, and forms the core of the elastic fibers that provide elasticity to tissues such as lungs, skin, and arteries [67]. It is initially produced as tropoelastin in elastogenic cells. This is a soluble protein of 60-72 kDa, depending on the isoform, and it consists largely of alternating hydrophobic and hydrophilic regions [68]. The hydrophobic regions contain mainly Gly, Val, Ala, and Pro, which are frequently present as tetra-, penta-, and hexapeptides [69, 70]. The hydrophilic domains are Ala-rich stretches with interspersed Lys residues, and function as crosslinking domains. Once synthesized and translocated into the lumen of the ER, tropoelastin is bound by an elastin-binding peripheral membrane protein (EBP) to prevent intracellular self-aggregation and degradation [71]. The elastin-EBP complex then moves through the Golgi to the cell surface, upon which the complex dissociates and tropoelastin is released [72]. Tropoelastin molecules then aggregate into globules on the cell surface through coacervation [73]. These globules are transferred to microfibrils in the extracellular space and fuse into larger fibrillar structures. In this process tropoelastin monomers are crosslinked by the enzyme lysyl oxidase *via* the lysine residues in the hydrophilic regions, resulting in the insoluble elastin polymer. The

elasticity of elastin, like that of rubber, is thought to be entropy-based: entropy decreases upon stretching, and recoil is driven by a return to maximal entropy [74].

The coacervation of tropoelastin monomers is caused by the protein's inverse temperature transition (ITT) behavior, or lower critical solution temperature (LCST) behavior. This behavior is well-known for many aqueous solutions of traditional chemically synthesized polymers such as poly(*N*-isopropylacrylamide) [75]. When a tropoelastin solution is heated above its transition temperature, phase separation occurs. This process is completely reversible upon cooling. At low temperature, water surrounds hydrophobic patches in a clathrate-like manner, keeping the protein unfolded. Upon a raise in temperature, the ordered water structure is disrupted and protein folding occurs through interactions between the hydrophobic domains. Although the protein structure becomes more ordered, the overall entropy increases because of the release of water [76].

To gain more understanding of the inverse temperature transition behavior of tropoelastin, Urry et al. [77] developed chemically synthesized model oligomers based on amino acid repeat sequences found in the hydrophobic domains of tropoelastin. Several of these synthetic polypeptides showed thermoresponsive behavior similar to that of tropoelastin [78, 79], and they follow general rules of polymer (LCST) phase behavior. The most studied repeat sequence is the pentapeptide VPGXG, where X can be any amino acid except proline. A hydrophobic residue at the X position, or high molecular mass, results in a low transition temperature, whereas a hydrophilic residue at the X position, or low molecular mass, results in a high transition temperature [80-82]. By using a charged amino acid at the X position, pH-dependent transition behavior is obtained [82]. The transition is also dependent on protein and salt concentration, where high concentrations of either have a lowering effect on the transition temperature [83, 84]. The dependence on protein concentration is reduced in longer molecules [84].

Recombinant elastin-like polypeptides (ELPs) are usually produced in *E. coli* [39, 85]. There is considerable biomedical interest in these materials in view of their biocompatible properties and tunable transition temperature [80, 86, 87]. The yield of ELPs in *E. coli* was markedly improved by not using standard Isopropyl β -D-1-thiogalactopyranoside (IPTG) induction, but rather relying on leaky transcription by the T7 promoter and culturing for extended periods [88]. In this manner, yields

in shake flasks of ~200-400 mg/L have been obtained [89]. Purification is mostly done by exploiting the reversible phase transition [90], in a procedure known as inverse transition cycling (ITC) [91]. This purification method involves repeated cycles of (i) aggregation of the ELPs at high temperature and/or high salt, followed by centrifugation, and (ii) resolubilization of the pellet at low temperature and low ionic strength, followed by centrifugal removal of any insoluble matter [90, 91]. Other hosts than *E. coli* have been employed, including tobacco cells [92], *Aspergillus nidulans* [93], tobacco plants [94], and *S. cerevisiae* [95]. In all these studies the ELPs were produced intracellularly.

Our group reported the first production of an elastin-like protein in *Pichia pastoris*, also representing the first example of secretory ELP production in any host [15]. The ELP gene introduced into the yeast encodes 90 VPGXG pentapeptide repeats, where the X position is occupied in a quasi-random order by valine, alanine, and glycine, at 50%, 20%, and 30%, respectively. Following the nomenclature of the Chilkoti group [19], we will hereafter refer to this ELP as V₅A₂G₃-90, and use analogous notation for all other ELPs. Bioreactor cultures were run in methanol fed-batch mode, and the ~36 kDa V₅A₂G₃-90 was purified from the cell-free broth by ITC. Within the experimental range of pH 3 to 7, product titers surprisingly increased from 0 to 410 mg/L with increasing culture pH, which was shown not to be related to proteolysis or ITC efficiency at different pH values. Because at pH 7 the protein was partially degraded, the optimal pH for production was established at 6, with a yield of purified intact V₅A₂G₃-90 of 255 mg per liter of cell-free broth. The transition temperature of the protein as determined by turbidity measurements was in good agreement with previously reported values.

We also produced two shorter ELPs in *P. pastoris* at pH 6 [16]. V₅A₂G₃-40 is an ELP with the same composition as V₅A₂G₃-90, featuring only 40 VPGXG repeats and consequently a higher transition temperature. V₅L₂G₃-40 is identical to V₅A₂G₃-40, except that all alanines in the X position of the pentapeptide repeats were replaced with the hydrophobic residue leucine. As expected, the transition temperature of V₅A₂G₃-40 measured over a wide concentration range under physiological conditions was considerably higher, by 10-20 °C, than that of both V₅A₂G₃-90 and V₅L₂G₃-40. Whereas V₅A₂G₃-40 was produced at 755 mg/L of cell-free broth, the yield of the hydrophobic V₅L₂G₃-40 was only 37 mg/L and it was partially degraded. Interestingly, at a low growth temperature of 20 °C the yield of

V₅L₂G₃-40 was roughly six times higher than at 30 °C and the protein was fully intact. See Section 9.2.6 for a discussion on the effects of both pH and transition temperature on ELP yields.

The Conticello group produced an ELP in *P. pastoris* consisting of 21 repeats of the amino acid sequence (VPGVG)₂VPGE(VPGVG)₂ [17]. This ~47 kDa V₄E₁-105 ELP includes a C-terminal c-Myc tag and a His-tag that was used for purification. Inverse transition behavior of the ELP produced was not studied, but under physiologically relevant conditions the protein is known to be present as a water soluble flexible polymer [17]. The yield after metal ion affinity chromatography was 2.5 mg/L. Analysis of the cell lysate revealed no significant amounts of intracellularly retained ELP.

9.1.4. Titin

The muscle protein titin, at ~4 MDa the largest protein known [96], provides elasticity and structure to the sarcomere. The elastic part of the protein contains two types of extensible segments: regions containing tandem repeats of ~100 residue immunoglobulin-like domains connected by short linkers, and intrinsically disordered regions including the so-called PEVK domain [97]. The PEVK domain consists of tandem repeats of 28-residue sequences consisting primarily of proline, glutamate, valine, and lysine. Depending on the muscle tissue the PEVK region varies in length from less than 200 to 2174 residues [96]. According to the most accepted model of titin elasticity, low stretch forces straighten the linkers between the tandemly repeated immunoglobulin-like domains, causing the immunoglobulin-like region to elongate, while intermediate to high physiological stretch forces mainly unfold and extend the PEVK segment [97]. At physiological force levels, unfolding of the immunoglobulin-like domains plays only a minor role. Although titin is often described as an entropic spring, where a force is needed to overcome entropic recoiling, nonentropic factors such as electrostatic and hydrophobic interactions of the PEVK region also play a role in the protein's elasticity [97].

Tsai et al. [18] produced a PEVK analog in *P. pastoris*, as well as in *E. coli* and Sf9 insect cells (*Spodoptera frugiperda*). The protein consists of 15 identical repeats of a 28 residue long sequence (PEPPKEVVPEKKAPVAPPKKPEVPPVKV), derived from human titin exon 172. It is flanked at both termini by multiple tags that were

included for purification, detection, and to allow single-molecule techniques such as AFM and Laser Tweezers. Random recombination of the repetitive genes occurred during all stages of cloning in *E. coli*. During protein production in *E. coli*, the protein was found to be toxic to the cells, and the use of low growth temperature, inclusion of glucose in the media, and mid-log phase IPTG induction were required. The yield, in shake flasks, was only 2 mg/L. While SDS-PAGE of the protein produced in Sf9 insect cells showed only one band after 24 h of culture, massive ladder formation was observed at 48 and 72 h. These secondary species, of both lower and higher mass than expected, were attributed to recombination events.

The protein was produced in *P. pastoris* using both secretory and intracellular expression. The results of intracellular production are scarcely discussed, and the authors mention that secretory production was most successful. Indeed, in bioreactor cultures this approach resulted in a yield of purified protein of more than 1 g/L, and the protein appeared largely intact in SDS-PAGE. The protein was purified by ammonium sulfate precipitation and hydroxylapatite chromatography.

9.1.5. De novo-designed block copolymers

In a general sense, block copolymers are polymers consisting of two or more chemically distinct and covalently linked homopolymeric modules. Block copolymers have attracted much attention because the combination of different types of physical behavior in one molecule can result in interesting new properties. When the block types used are mutually poorly miscible, the block copolymer can undergo ‘microphase separation’ [98]. That is, two-phase macroseparation is not possible because the blocks are covalently linked, and instead periodic nanostructures may be formed. A well-known chemically synthesized commercial polymer with such properties is Kraton D SBS, a triblock copolymer with polystyrene end blocks and an elastomeric polybutadiene middle block [99].

Nature has devised its own protein block copolymers, just one intriguing example of which is the mussel byssus protein preCol-P, which consists of a central collagen domain, flanking elastic domains, and histidine-rich terminal domains [100]. Similarly, and contrasting common chemical synthesis, genetic engineering allows direct sequential ordering of the functional blocks in a protein polymer through sequential ordering of the encoding DNA fragments in the gene template

[101] (see also Chapter 1). The first example of recombinant protein block copolymers was reported by the group of Cappello and Ferrari [39]. The authors produced block copolymers in *E. coli* where crystalline (GAGAGS)_n blocks were alternated with noncrystalline GAAGY or elastin-like (VPGVG)_n blocks. Many more protein block copolymers have been produced using *E. coli*. Just a few examples are copolymers combining different elastin-like blocks [102, 103], elastin-like and biofunctional sequences [104], different spider silk sequences [28], and a random coil middle block with flanking associative coiled-coil domains [105, 106].

Below we will list *de novo*-designed block copolymers produced in *P. pastoris*, so far exclusively pursued by our group (Table 9.2). Unless otherwise noted, all proteins were produced in bioreactors in methanol fed-batch mode, at pH 3 and 30 °C in basal salts medium with PTM1 trace elements [61], using strain GS115. Also unless otherwise noted, purification was essentially based on ammonium sulfate precipitation (Section 9.1.2.1).

Table 9.2. *De novo*-designed block copolymers produced as secreted proteins using *P. pastoris*

Block copolymer ^a	Molecular weight (kDa)	Yield (g/L) ^b	Reference
<i>Silk-inspired and pH-responsive</i>			
C^P₂-S^E₄₈-C^P₂ / S^E₂₄-C^P₄-S^E₂₄	66	4.5 - 6.5 ^c	[107]
C^P₂-S^H₄₈-C^P₂ / S^H₂₄-C^P₄-S^H₂₄	66	6.1 ^c	[108-111]
C^P₂-S^H_n-C^P₂; n = 8, 16, 24	43 / 48 / 52	>2	[112]
C^P₂-S^K₄₈-C^P₂ / S^K₂₄-C^P₄-S^K₂₄	66	>1	[110; unpublished]
B^{RGD}-C^P₂-S^H₄₈-C^P₂ / B^{KRSR}-C^P₂-S^H_n-C^P₂	68	3.2 / 3.5	Chapter 8
<i>Collagen-inspired and thermoresponsive</i>			
T₉-C^P₄-T₉ / T₉-C^R₄-T₉	42	3 - 6 ^c	Chapter 6
T₉-C^P_n-T₉ / T₉-C^R_n-T₉; n = 8, 12	78 / 115	1 - 3 ^c	[113; unpublished]
T_n-C^P₄-T_n / T_n-C^R₄-T_n; n = 6, 12, 16	40 / 43 / 45	3-6 / 0.6 / 0.4 ^{c,d}	[114, 115]
<i>Multi-responsive</i>			
S^E₂₄-E₄₀ / S^E₁₂-C^P₄-E₄₀	30 / 60	>0.5	[116]
C^P₂-S^H₄₈-C^P₂-D^C	66	3	[117]
T₉-C^R₄-T₉-D^C	42	0.4	[118]
T₉-C^R₄-B^{H6}	40	0.5	[119]
T₉-C^R₄-B^{K6}	40	0.4	[120]
S^H₈-C^R₄-T₉	44	2	[121]
<i>DNA-binding</i>			
C^P₄-B^{K12} / C^P₄-B^{(HK)6}	38	1	[122]
C^P₈-B^{Sso7d}	80	0.7	[123]
C^P₄-S^Q_n-B^{K12}; n = 2, 4, 10, 14	40 - 47	0.5 ^e	[124]
<i>Featuring heterodimerizing modules</i>			
C^P₄-D^A / C^P₄-D^B / T₉-C^P₄-D^A / T₉-C^P₄-D^B	43 - 45	0.7 - 1.3 ^f	[125]
C^P₂-S^H₄₈-C^P₂-D^A / C^P₂-S^H₄₈-C^P₂-D^B	72	2.3 / 1.4	[126]
C^P₄-D^{WW} / C^P₄-D^{PPXY}	41 / 39	2.2 / 2.3	Chapter 7

^a Block nomenclature is as follows: C, random coil; S, silk-like; T, collagen trimer-forming; B, binding; E, elastin-like (V₅A₂G₃); D, dimerizing. Block subtypes indicated in superscript are described in the text. Subscripts indicate the number of repeats for each block, where the repeating units for C, S, T, and E consist of 99, 8, 3, and 5 residues, respectively. Different polymer names may have been used in the cited articles.

^b Except where noted, yields are given as the amount of recombinant protein per L of cell-free broth recovered after purification, dialysis, and lyophilization. Note that significant (>50%) protein loss occurs particularly during dialysis owing to the linear shape of the polymers.

^c Yield of recombinant protein expressed, as determined by SDS-PAGE and calibrated densitometry.

^d The three yields listed refer to the n = 6, 12, and 16 variants, respectively.

^e Estimated yield of recombinant protein expressed.

^f Equivalent to the yield of 0.5 - 0.6 g/L of culture mentioned in the cited article.

9.1.5.1. Silk-inspired pH-responsive block copolymers

The first block copolymer produced by our laboratory consists of two modules mentioned above, namely the silk-like protein S^E₄₈ consisting of 48 repeats of the sequence (GA)₃GE (Chapter 5, Section 9.1.1.2), and two repeats of the 99-residue

hydrophilic random coil block C^P (Chapter 4, Section 9.1.2.1). The protein, referred to as $C^P_2-S^{E_{48}}-C^P_2$, is a triblock where the silk-inspired sequence $S^{E_{48}}$ is used as the middle block, flanked by noninteracting C^P_2 blocks [107]. In other words, the triblock structure of $C^P_2-S^{E_{48}}-C^P_2$ is identical to that of $C^P_2-S^{H_{48}}-C^P_2$ described in Chapter 8 (see Figure 8.1A), but the middle block has $(GA)_3GE$ repeats instead of $(GA)_3GH$ repeats. At neutral pH, the $S^{E_{48}}$ middle block is soluble owing to repulsion by the negatively charged glutamates. At low pH approaching pI, the charges are neutralized and the silk-like domain forms a β -roll or β -solenoid structure with the hydrophobic side groups facing inwards [127]. These then, through stacking mediated by hydrogen bonds, form the core of micrometer-long fibrils. The hydrophilic random coil end blocks surround the core and provide colloidal stability. The $C^P_2-S^{H_{48}}-C^P_2$ variant mentioned in Chapter 8 features charge neutralization and fibril formation at neutral pH [108, 109, 111]. The fibrils form gels, probably *via* entanglement of long fibrils, and perhaps weak physical crosslinks. The histidine-containing triblock copolymer was also produced with middle blocks featuring only 8, 16, or 24 octapeptides, where the first two variants formed spherical micelles rather than fibrils [112]. A less studied triblock is $C^P_2-S^{K_{48}}-C^P_2$, which features lysine as the last residue of each octapeptide (unpublished). Additionally, the polymers $S^{E_{24}}-C^P_4-S^{E_{24}}$, $S^{H_{24}}-C^P_4-S^{H_{24}}$, and $S^{K_{24}}-C^P_4-S^{K_{24}}$ were produced, which feature reverse block order [107, 110]. All silk-inspired triblock copolymers were produced at g/L levels by *P. pastoris* as secreted proteins using the ppaF leader. Production was at pH 5 for the Glu variant and at pH 3 for the His or Lys variants to prevent possible self-assembly of the proteins in the broth.

Because the variant with histidines forms hydrogels at physiological pH, and in view of the general biocompatibility of silk-like materials [26, 27], the polymer is of potential interest for biomedical applications such as tissue engineering. We have shown that these hydrogels show long-term stability in erosion studies and display self-healing behavior [128]. The latter is of great interest for the development of, e.g., injectable hydrogels, because self-healing materials can recover from damage incurred during application to the tissue [129], or afterwards. Rat bone marrow cells remained fully viable in direct contact with this material, although proliferation was low. As discussed in Chapter 8, we also produced functionalized polymer variants that feature genetically incorporated general cell-

adhesive RGD motifs ($\mathbf{B}^{\text{RGD}}\text{-CP}_2\text{-S}^{\text{H}_{48}}\text{-CP}_2$), as well as osteoblast-specific KRSR motifs ($\mathbf{B}^{\text{KRSR}}\text{-CP}_2\text{-S}^{\text{H}_{48}}\text{-CP}_2$), with purified protein yields of 3.2 and 3.5 g/L of cell-free broth, respectively. By mixing functionalized and nonfunctionalized polymers in the desired ratio, the composition of cell-adhesive domains could be tailored independently of the material's mechanical properties. Cell adhesion, metabolic activity, spreading, and confluency of MG-63 osteoblastic cells were highly promoted on these mixed materials. Scaffolds with varying RGD content and overall protein concentration were also tested in three-dimensional culture, where the cells were embedded in the scaffold and revealed spreading [130]. The most favorable combination featured high RGD content, and low protein concentration (i.e., low stiffness). These results showcase the possibility to independently tune different properties of these hydrogels, and the potential for further developing them into cell culture scaffolds.

We furthermore used the ability of KRSR domains to bind the negatively charged linear polysaccharide heparin [131], showing that heparin could act as a binder for controlled bundling of $\mathbf{B}^{\text{KRSR}}\text{CP}_2\text{S}^{\text{H}_{48}}\text{CP}_2$ fibrils [132]. In the natural extracellular matrix (ECM), bundling of fibrils allows for a wide variety of stiffness and organization of the network. Similarly, by modulating the protein to heparin ratio, we were able to affect fibril bundling, leading to faster formation of stiffer hydrogels. This approach for controlled bundling thus has potential in the fabrication of materials that mimic the ECM.

9.1.5.2. Collagen-inspired thermoresponsive block copolymers

Another category of triblock copolymers we developed uses collagen-like (Pro-Gly-Pro) $_n$ end blocks for self-assembly. In Chapter 6 we presented, as an example, the proteins $\mathbf{T}_9\text{-C}^{\text{P}_4}\text{-T}_9$ and $\mathbf{T}_9\text{-C}^{\text{R}_4}\text{-T}_9$, which have (Pro-Gly-Pro) $_9$ end blocks, and where C^{P_4} and C^{R_4} serve as random coil middle blocks. Owing to their high proline content, the end blocks are sufficiently stable to form triple helices in the absence of prolyl-4-hydroxylation (Section 9.1.2). The melting point of (Pro-Gly-Pro) $_n$ peptides can be controlled by their length n . Thus, besides \mathbf{T}_9 end blocks, we also constructed \mathbf{T}_n end blocks with $n = 6, 12$, and 16 [114, 115]. The proteins with $n = 6$ or 9 were produced at several g/L of cell-free broth, and the proteins with $n = 12$ or 16 at 0.6 and 0.4 g/L, respectively. Although the polymer with $n = 16$ was initially compromised by proteolysis, this problem could be solved by using the *yps1*

protease-deficient strain reported in Chapter 3 (see also Sections 9.2.3.2.4 and 9.2.6). The tunability of the melting point known from free (Pro-Gly-Pro) $_n$ peptides was retained in the triblock design [115]. In fact, the melting free energy and melting points determined were similar to those of free peptides of equal length. At sufficiently high concentration the triblock copolymers can form thermoreversible hydrogels, in which exclusively the collagen-like end blocks form the trimeric knots of the network. The reversible transition of the equilibrium from the trimeric to monomeric state of the T_n blocks occurs over ~ 20 - 30 °C. The melting range shifts to increasing temperatures with increasing n . In line with this, T_{16} -C $_4$ - T_{16} gels melted at a higher temperature than T_9 -C $_4$ - T_9 gels and also showed greater stiffness at the same temperature and molar concentration, because the equilibrium is shifted toward higher trimer content. Besides varying the length of the T_n end blocks, we also produced variants with longer random coil middle blocks, i.e., T_9 -C P_n - T_9 and T_9 -C R_n - T_9 with $n = 8$ or 12 [113; unpublished]. These large protein polymers of 78 or 115 kDa, respectively, were completely intact and efficiently secreted at 1-3 g/L of cell-free broth. By varying polymer length and concentration, elastic properties, nominal mesh size, swelling behavior, and erosion could be tuned. Sustained release of encapsulated proteins was demonstrated, which offers prospects for use of these hydrogels as a drug delivery system [133].

The demonstrated tunability of the thermoresponsive behavior of these hydrogels, as well as their expectedly biocompatible (collagen-like/gelatin-like) nature, offers promise for their use as highly defined alternatives to animal gelatins in biomedical applications. In contrast to aged, crosslinked, animal gelatins, the block copolymers showed fast self-healing behavior (Figure 9.3) [134]. As mentioned in Section 9.1.5.1, self-healing hydrogels are highly desirable in biomedical applications. We also demonstrated that a composite hydrogel network formed by T_9 -C R_4 - T_9 together with the fibril-forming silk-inspired C^{P_2} -S $^{H_{48}}$ - C^{P_2} polymer (Section 9.1.5.1) offers synergistic enhancement of rigidity and rupture strength [135], thereby further expanding the potential of these materials for biomaterial applications.



Figure 9.3. Self-healing of a $T_9-C^P_8-T_9$ collagen-inspired hydrogel. An orange-colored piece of broken hydrogel (top) was placed in contact with a blue-colored piece (bottom) for 15 min at RT, after which the two strongly adhered together through newly-formed triple helices.

9.1.5.3. Multi-responsive block copolymers

To create multifunctional biomaterials, in particular for medical applications, research efforts are increasingly devoted to the development of block copolymers that combine different functionalities. Besides bioactive modules, also modules that control topology or trigger-responsiveness can be incorporated. This section will focus on the latter.

Two dual-responsive silk-elastin-like protein polymers, $S^{E_{24}}-E_{40}$ and $S^{E_{12}}-C^P_4-E_{40}$, were constructed [116]. Both are composed of an N-terminal silk-like block with 24 or 12 $(GA)_3GE$ octapeptide repeats (Chapter 5), and a C-terminal $V_5A_2G_3-40$ elastin-like block (Section 9.1.3). These blocks impart pH-responsiveness and temperature sensitivity, respectively. The $S^{E_{12}}-C^P_4-E_{40}$ polymer additionally has the flexible and inert ~ 37 kDa C^P_4 spacer (Chapter 4) between the silk-like and elastin-like modules, which may allow the end blocks to take part in spatially separated microphases. The triblock was purified by ammonium sulfate precipitation, and the diblock by ITC. Both proteins were found to be pure and intact. The recovery after purification was more than 500 mg/L of clarified broth. The combination of both pH- and temperature-responsive modules in a single polymer resulted in complex supramolecular self-assembly behavior. For $S^{E_{12}}-C^P_4-E_{40}$ this behavior was studied in detail, and found to be pathway-dependent [136]. Fibrils were formed upon first lowering the pH, which bundled after subsequent heating. In contrast, micelles

were formed upon first heating the solution, which seemed to aggregate into large structures after subsequent lowering of the pH. Both pathways resulted in the formation of hydrogels, although with different morphologies and mechanical properties.

Redox-responsiveness was introduced into the $\text{C}^{\text{P}_2}\text{-S}^{\text{H}_{48}}\text{-C}^{\text{P}_2}$ polymer (Section 9.1.5.1) by addition of a cysteine residue to the C-terminus of the sequence [117]. The polymer, referred to as $\text{C}^{\text{P}_2}\text{-S}^{\text{H}_{48}}\text{-C}^{\text{P}_2}\text{-D}^{\text{C}}$, was produced as an intact protein with a recovery after purification of 3 g/L of cell-free broth, which is in the same range as the yield obtained for the unmodified polymer. Gels formed under oxidizing conditions had a significantly increased storage modulus as a result of increased fibril rigidity. We substantiated that the formation of intra-fibril disulfide bridges leads to a slightly reduced lateral pressure in the hydrophilic corona, which probably affects the packing of the silk-like blocks in the fibril core.

A C-terminally cysteine-extended variant of the collagen-inspired polymer $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9$ (Chapter 6) was similarly constructed [118]. Unlike the unmodified monodisperse protein, the cysteine-extended variant, $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9\text{-D}^{\text{C}}$, was initially degraded by proteolysis. The *yps1* protease-deficient strain reported in Chapter 3 was used to solve this problem (see also Sections 9.2.3.2.4 and 9.2.6), resulting in ~400 mg of purified intact protein per liter of clarified broth. Under reducing conditions, the polymer behaved like unmodified $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9$ in terms of melting point in differential scanning calorimetry and storage modulus of the hydrogel. Under oxidizing conditions, however, the thermogram showed an additional melting point shifted up by ~11 °C. This shows that a fraction of the triple helices was stabilized by disulfide formation, most likely owing to an increase in local concentration of the T_9 blocks. That is, the occurrence of preexisting disulfide-bonded T_9 pairs promoted the subsequent formation of triple helices. The storage modulus of hydrogels formed by $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9\text{-D}^{\text{C}}$ under oxidizing conditions was higher than that of normal $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9$ by a factor two, which was attributed to a shifted equilibrium toward more triple helices, as well as the crosslinking of different triple helices by disulfide bridges.

Two other asymmetric block copolymers based on $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9$ were devised, where the C-terminal T_9 block was replaced with a cationic binding block consisting of either six histidines ($\text{T}_9\text{-C}^{\text{R}_4}\text{-B}^{\text{H}_6}$) [119], or six lysines ($\text{T}_9\text{-C}^{\text{R}_4}\text{-B}^{\text{K}_6}$) [120]. Both proteins were produced intact at 500 and 400 mg/L of clarified broth,

respectively. Protein $T_9-C^{R4}-B^{H6}$ was used in conjunction with the polyanion sodium polystyrene sulfonate, where electrostatic interaction between the C-terminal B^{H6} block and the polyanion causes the formation of micelles at a temperature above the melting point of the T_9 block. At low temperature, in the absence of polyanion, star polymers are formed through the formation of triple helices by the N-terminal T_9 block. Under conditions where both end blocks self-assemble (low temperature, in the presence of polyanion), a thermoreversible hydrogel is formed. This network formation can start either with the charge-driven micelles (by cooling), or with the thermoresponsive star polymers (by addition of polyanion). Similarly, the $T_9-C^{R4}-B^{K6}$ polymer was used in conjunction with the anionic polymer xanthan. Xanthan is a widely used water soluble food additive with shear-thinning and pseudo-plastic properties, and its functionalization is of potential interest in biomaterial applications. The composite xanthan/ $T_9-C^{R4}-B^{K6}$ system formed reversible thermosensitive hydrogels, where the storage modulus was significantly increased as compared to xanthan alone.

Another dual responsive polymer is $S^{H8}-C^{R4}-T_9$, which has a short N-terminal silk-like block that self-assembles at neutral pH, and a thermoresponsive collagen triple helix-forming end block [121]. This protein was produced intact and at 2.0 g/L of clarified broth. It was used in conjunction with $C^{P2}-S^{E48}-C^{P2}$, together forming a hybrid system. In a first step at high temperature and at a pH halfway between the pKa values of the histidine and glutamic acid side chains, coassembly occurred between the oppositely charged S^{H8} and S^{E48} silk-like domains of both proteins. Whereas the individual proteins do not appreciably self-assemble under these conditions, the mixture resulted in the formation of mixed coassembled fibrils. At high temperature, slow gel formation occurred. Upon cooling of the mixture, the storage modulus vastly increased. This can be explained by the formation of trimeric T_9 nodes that crosslink $C^{P2}-S^{E48}-C^{P2}$ fibrils. A hydrogel system with thermoreversible crosslinks was obtained that features temperature-switchable stiffness.

9.1.5.4. DNA-binding block copolymers and artificial viral coat proteins

DNA is a much-investigated nanomaterial [137]. For many applications, including gene delivery, coating of DNA with suitable binder moieties is an effective approach to alter its physical properties and interactions with the

environment. However, coating DNA with chemically synthesized polymers typically results in aggregates with random size and shape, such that there is a need for more controlled coatings. In its simplest form, a macromolecule for DNA coating purposes should contain a cationic module that can bind the negatively charged DNA molecule, and a module that provides colloidal stability [138, 139]. We have designed and produced protein polymers that meet these requirements [122]. They feature the hydrophilic random coil C^P_4 sequence (Chapter 4) as the module conferring colloidal stability, and a short C-terminal cationic block. The latter consists of either 12 Lys residues, or 6 His-Lys dipeptides, in proteins $C^P_4\text{-B}^{K12}$ and $C^P_4\text{-B}^{(HK)6}$, respectively. The ~38 kDa diblock copolymers were produced in *P. pastoris* as monodisperse polymers and were purified using ammonium sulfate precipitation, resulting in a recovery of ~1 g/L of clarified broth. The proteins were shown to complex with single DNA molecules into highly stable bottlebrushes [122, 140]. Inter- and intramolecular bridging of DNA molecules was prevented by virtue of the combination of a relatively small binding block and a large block for colloidal stability. The DNA bottlebrushes show liquid crystalline behavior at DNA concentrations roughly one order of magnitude below those for bare DNA [141]. In addition to these polymers, which feature simple electrostatic DNA binding, a variant was produced that uses the 7 kDa protein Sso7d as the nonsequence-specific DNA-binding domain [123]. The DNA binding properties of this protein from the thermoacidophilic archaeobacterium *Sulfolobus solfataricus* have been well characterized, and it has been used for example to improve the processivity of DNA polymerases. The length of the random coil domain of the polymer was doubled from C^P_4 to C^P_8 , so as to increase the solubility and rigidity of the resulting protein-DNA complexes. The resulting protein, $C^P_8\text{-B}^{Sso7d}$, was produced in *P. pastoris* as a secreted intact protein, and efficiently purified by ammonium sulfate precipitation. The yield of purified protein was 0.7 g/L of cell-free medium. The protein polymer was able to coat various one-dimensional DNA templates, as well as a two-dimensional DNA origami nanostructure, without structural distortion. The polymer coating offered moderate protection against DNase I digestion.

The self-assembly of viruses is a highly cooperative process, where initial binding of coat proteins to the encoding viral genome triggers the binding of further coat proteins. With gene delivery applications in mind, we engineered

supramolecular cooperativity into the DNA-coating protein $C^P_4-B^{K12}$ described above. This was achieved by including the silk-like sequence $S^Q_n = [(GA)_3GQ]_n$ between the C^P_4 domain that provides colloidal stability and the B^{K12} DNA-binding domain [124]. See Figure 9.4 for a representation of the resulting polymer.

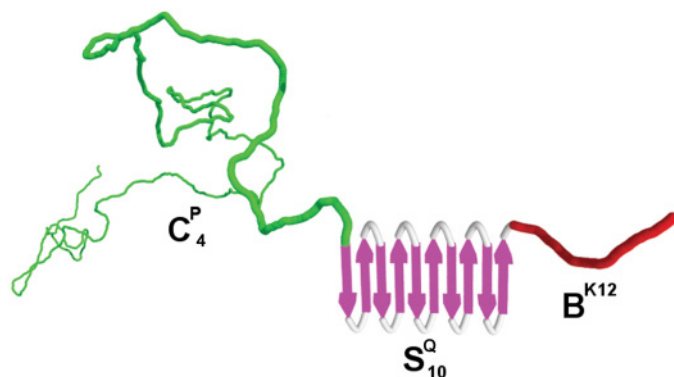


Figure 9.4. Design of viral coat protein $C^P_4-S^Q_n-B^{K12}$, with $n = 10$. C^P_4 is an N-terminal ~400-residue hydrophilic random coil block (green), S^Q_{10} constitutes the self-assembly block (pink) and is a tenfold repetition of the octapeptide $(GA)_3GQ$, and B^{K12} is the C-terminal dodecalysine DNA-binding block (red). Adapted from Hernandez-Garcia et al. [124].

This silk-like octapeptide is identical to that described in Chapter 5 and Section 9.1.5.1, except that Gln replaces the charged amino acids Glu, His, or Lys, so as to rule out electrostatic behavior of this block. The number of silk-like strands n determines the propensity of the S_n block to stack and, thereby, the degree of binding cooperativity. A series of $C^P_4-S^Q_n-B^{K12}$ polymers with $n = 0, 2, 4, 10$, and 14 was efficiently produced as intact proteins in *P. pastoris*. Fully cooperative assembly on a DNA template was obtained for $n = 10$ and 14. Rod-shaped virus-like particles were formed, in which the DNA was fully protected against high concentrations of DNase I. The particles were able to transfect HeLa cells, resulting in expression of a fluorescent reporter. These rationally designed artificial virus coat proteins, produced as triblock copolymers, thus provide mechanistic insight into viral self-assembly, and might be further developed as useful vectors in gene therapy.

9.1.5.5. Block copolymers with heterodimerizing modules

To expand the repertoire of possible topologies of nanostructured materials, a need exists to couple different protein polymers. This can be done by traditional

chemical crosslinking, but enzymatic and autocatalytic conjugation, as well as reversibly heterodimerizing modules are now also increasingly being investigated. We have reviewed these nonchemical approaches elsewhere [142]. Heterodimerizing modules are of particular interest, in that they allow the linking of different protein polymers or protein domains by the mere mixing of two separately produced components. A possible medical application could be injectable hydrogels that self-assemble *in situ* upon mixing of two protein polymers A and B.

A first category of heterodimerizing modules that we incorporated into protein polymers are leucine zipper peptides of 47 amino acids, capable of forming high affinity α -helical heterodimeric coiled coils. These peptides were first designed and studied by Moll et al. [143]. The relatively acidic peptide RR₁₂EE₃₄₅L will be denoted here as **D^A**, the relatively basic partner EE₁₂RR₃₄₅L as **D^B**. We added **D^A** and **D^B** domains to the C-terminus of **C^P₄** (Chapter 4) and **T₉-C^P₄** polymers (a diblock variant of **T₉-C^P₄-T₉**; Chapter 6). Although the four protein polymers were initially proteolyzed, optimized fermentation conditions (Section 9.2.3.1) led to the production of essentially intact protein polymers at 0.7 – 1.3 g/L of cell-free broth. The heterodimerizing modules were functional and mediated the coupling of the A and B counterparts. The **D^A** and **D^B** modules were also attached to the C-terminus of the fibril-forming triblock copolymer **C^P₂-S^H₄₈-C^P₂** (Section 9.1.5.1). The ensuing **C^P₂-S^H₄₈-C^P₂-D^A** and **C^P₂-S^H₄₈-C^P₂-D^B** polymers were purified from the cell-free broth at 2.3 and 1.4 g/L, respectively [126]. In mixtures of bare **C^P₂-S^H₄₈-C^P₂** and the **D^A/D^B**-containing variants, fibril bundling occurred to an extent controlled by the ratio of modified to unmodified polymer. This holds promise for the further development of biomaterials mimicking the ECM, as fibril bundling plays a key role in this environment.

Another category of heterodimerizing modules are the so-called WW-domain (~37 residues) and its complementary proline-rich ligand (~13 residues). The binding between the two occurs with high specificity and *K_d* values typically in the high nM to low μ M range [144]. As described in Chapter 7, we designed variants of the **C^P₄** polymer (Chapter 4) with C-terminal extensions consisting of (1) the p53-binding protein-2 PPxY sequence, or (2) a WW domain derived from human ubiquitin ligase, referred to as WWP1-1. The polymers, **C^P₄-D^{PPxY}** and **C^P₄-D^{WW}**, respectively, were produced at a purified protein yield of 2.2 – 2.5 g/L of cell-free

broth. The PPxY module was found to be O-glycosylated, and a large portion of the attached glycans was phosphorylated (see also Section 9.2.4). Glycosylation was abolished by mutating the single serine in the PPxY module to alanine.

9.2. Challenges

From Chapters 2 to 8 and the case studies described in Section 9.1 it is clear that *P. pastoris* is overall an excellent host for the production of protein polymers. Repetitive genes are stably maintained, product titers are often high, and the possibility of secretory production for many of these protein polymers allows efficient downstream processing. Nonetheless, some challenges are apparent also. We will discuss relatively common bottlenecks next, with a focus mainly on secretory production.

9.2.1. The need for bioreactors

Most molecular biology and biochemistry laboratories use shake flask cultures for protein expression in *E. coli*, for reasons of familiarity, convenience, and the minimal infrastructure required. Some researchers similarly trying *P. pastoris* using shake flasks, often with the aim of obtaining high yields, may abandon the system in disappointment. This may especially apply to the field of protein polymers, where relatively high amounts are needed for physical characterization of the resulting materials by rheometry, digital scanning calorimetry, tensile strength testing, etc. In our opinion, meaningful use of *P. pastoris* for the production of protein polymers, particularly in secretory mode, requires the use of well-controlled bioreactors. We routinely observe product levels in stirred tank fed-batch cultures that are higher than in shake flasks by a factor 20 to 100. A similarly large impact of the use of bioreactors is apparent from some of the protein polymer examples in Table 9.1, and heterologous proteins in general [145]. Unlike shake flasks, bioreactors offer strict control over feeding and dissolved oxygen (DO) levels, which enables the generation of very high biomass levels. Moreover, continuous controlled feeding minimizes cell lysis that can occur in shake flasks as a result of starvation and/or overdosing caused by intermittent manual feeding, particularly with methanol as the substrate. Whereas the pH in shake flasks is poorly controlled, typically by using buffered media, bioreactors offer continuous pH adjustment. Cell lysis and pH fluctuations can have a dramatic impact on

proteolysis (Section 9.2.3), and indeed we have often seen completely different degradation patterns and kinetics in shake flasks as compared to bioreactors. Although relatively uncontrolled cultures using shake flasks, multi-well plates, etc., can be useful for screening transformants, they have little relevance to eventual bioprocessing. In contrast, we find that bench-top bioreactor cultures of *P. pastoris* can be readily scaled-up to at least 1,000 L pilot-scale with only limited loss of productivity, and no impact on product integrity (unpublished). This is highly relevant for studies ultimately aiming at industrial production and commercial applications.

Although there are many ways of culturing *P. pastoris* in bioreactors, the most common approach is a process consisting of three stages as follows [61]. The first stage is a glycerol batch phase, intended to quickly generate biomass. This phase ends once glycerol is depleted, as evident from a sharp DO increase. The second stage is a glycerol fed-batch phase. It involves the feeding of glycerol at a growth-limiting rate, which serves to further increase biomass levels and simultaneously derepress the *AOX1* promoter. The third stage is a methanol fed-batch phase, in which the *AOX1* promoter is induced and actual protein production takes place. Many different approaches exist for the transition from glycerol to methanol, the methanol feeding regime, and DO management. We will comment on some of these process parameters in Section 9.2.3.1 in the context of proteolytic product degradation. For more details on bioprocess engineering of *P. pastoris*, particularly also in relation to productivity, the reader is referred to several outstanding reviews [146-150].

9.2.2. Secretory leader processing and efficiency

Many different secretory leaders have been used in *P. pastoris*. This includes both endogenous *P. pastoris* signal peptides such as the acid phosphatase signal sequence (Pho1) [32], and signal peptides taken from other organisms [151-153]. Often the native signal peptide of a given heterologous protein will function well in *P. pastoris* [32]. However, by far the most often used leader is the prepro region of the *S. cerevisiae* α -factor mating pheromone (ppaF). It is used in well-known *P. pastoris* vectors such as pPIC9 [154] and pPICZ α (Thermo Fisher Scientific). The efficiency of leader sequences appears protein-dependent, but overall the ppaF is considered to have the highest success rate [155]. Also most secreted protein

polymers were produced in *P. pastoris* using this leader (Section 9.1). The fact that we obtained g/L titers for rat Col3a1 gelatin using ppαF (Chapter 2), but could not detect any product when the Pho1 signal peptide was used (unpublished), illustrates the importance of leader choice. However, also ppαF is not without drawbacks, as it can lead to N-terminal heterogeneity of the target protein. Before elaborating on this, we will first describe the natural purpose of this leader, and its processing.

The *S. cerevisiae* α-factor pheromone is a 13-residue peptide secreted into the medium by haploid α cells as a prerequisite for mating with a haploids [156]. Figure 9.5A shows the structure and *in vivo* processing of the 165-residue precursor, which contains a 19-residue hydrophobic N-terminal signal peptide for ER translocation, a 64-residue pro-segment with three N-glycosylation sites, and four identical copies of the α-factor peptide. The pheromone peptides are separated by sequence stretches consisting of a dibasic Lys-Arg site and two to three Glu-Ala or Asp-Ala dipeptides. The pheromone peptides are released first by cleavage at the C-terminal side of Lys-Arg by the endopeptidase Kex2, followed by removal of the dipeptide repeats by the aminopeptidase DPAPase A, and removal of the Lys and Arg residues by the carboxypeptidase Kex1 [157]. These proteolytic steps all take place in the late Golgi.

Figure 9.5B depicts how the ppαF region is used to drive secretion of heterologous proteins in *S. cerevisiae* [33, 158] and *P. pastoris* [154]. The protein of interest is placed C-terminally of the dibasic Kex2 site, and typically also the (Glu-Ala)₂ spacer is included. The latter likely facilitates Kex2 processing by reducing steric interference caused by the fused protein [32]. However, the spacer itself is not always properly removed by DPAPase A [159-161]. We have observed removal of only one of the two Glu-Ala repeats with Col1a1 gelatin (Chapter 2), where the C-terminally flanking residue is proline. It has been suggested that proline in this position may interfere with DPAPase processing [162]. The gelatin-like polymer C^P₄ (Chapter 4), which has the small amino acid glycine as the first residue, revealed complete removal of the Glu-Ala repeats. With folded proteins, however, their spatial structure likely determines the efficiency of DPAPase processing to a large degree. As such, it is interesting to note that the T₉-C^P₄-T₉ and T₁₆-C^P₄-T₁₆ polymers, potentially capable of forming triple helices close to the Glu-Ala repeats, do show low levels of extended species (Chapter 6) [114]. Overexpression of

DPAPase A in *S. cerevisiae* led to increased maturation of α -factor pheromone [163], which suggests this may be a viable strategy for improving the processing of the ppaF leader when used to drive secretion of heterologous proteins. To our knowledge this approach has not yet been attempted in *P. pastoris*.

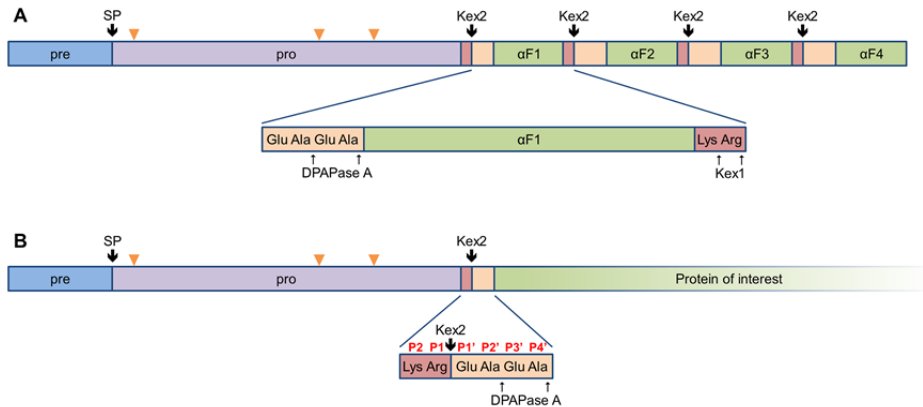


Figure 9.5. (A) Schematic structure and proteolytic processing of the *S. cerevisiae* α -factor mating pheromone precursor. (B) Use of the α -factor prepro secretory signal for secretion of heterologous proteins. The (Glu-Ala)₂ spacer is optional (see main text). Indicated in red font is an explanation of Schechter and Berger [164] nomenclature for protease substrate positions, as referred to throughout this chapter. SP: signal peptidase, α F1 – α F4: α -factor peptides, orange triangles: N-glycosylation sites, thick arrows: endopeptidase cleavage, thin arrows: exopeptidase cleavage.

There are many successful examples in *P. pastoris* where the ppaF leader was used without Glu-Ala spacer, thereby preventing N-terminal heterogeneity [165–169]. However, depending on the nature of the downstream protein, omitting the spacer may in turn result in compromised processing by Kex2. This then leads to short ppaF-derived extensions that are typically 9–11 residues in length [159, 160, 170, 171], and sometimes even larger glycosylated fragments corresponding to most of ppaF [160, 161, 172]. Although Kex2 is thought to positively recognize only the region N-terminal to the cleaved bond [173–175], *in vitro* Kex2 cleavage of a set of potential protein substrates suggested a preference for acidic residues at P1', P2', and P4' [176] (see Figure 9.5B for cleavage site nomenclature). Arg at P1' seems to disturb processing [177], or might result in downstream-shifted cleavage because it could itself function as the P1 position. Some other residues appeared unfavorable at P1' in an *in vitro* assay, particularly the β -carbon branched amino acids Ile, Val, Leu and Thr [178]. Pro at P1' was not included in the assay, but did appear to inhibit Kex2 cleavage according to other studies [179, 180]. Surprisingly, in a study

by Zheng et al. [181], Pro, together with Ile, actually resulted in the highest activity among all 20 P1' substitutions evaluated for improved secretion of an α -galactosidase. Yang et al. [182] likewise investigated the effect of different P1' residues on secretory yield in *P. pastoris*, but the results were highly variable among the proteins tested. All considered, we propose that P1' as such is not a major determinant, with the possible exception of Arg and Pro, but rather that efficient ppaF processing by Kex2 depends largely on the folded structure of the fused protein. A case has been reported where ppaF remained attached despite the presence of a Glu-Ala spacer [32], which suggests that the distance introduced by the spacer, as well as the presence of a native ppaF-derived acidic Glu residue at P1' were insufficient to overcome interference by the particular cargo protein. Intramolecular disulfide bond formation involving Cys¹ has been proposed to disturb Kex2 processing of interferon α -2a expressed in the methylotrophic yeast *H. polymorpha* [183]. Similarly hampered processing of interferon α -1 and α -2b with a disulfide-bonded Cys¹ has been observed in *P. pastoris* [184, 185]. Notably, also glycosylation of proteins fused to ppaF has been shown capable of disturbing processing by Kex2 [186, 187].

In an effort to improve Kex2 processing in *P. pastoris*, Yang et al. [182] overexpressed the enzyme, resulting in increased fluorescence in the supernatant by the Venus reporter protein. This strategy could be worthwhile for improving the processing of folded proteins that are recalcitrant to complete processing. However, we argue that Kex2-like processing capacity in *P. pastoris* is not normally limiting in the absence of steric interference, as we have not observed incomplete processing of the ppaF Kex2 site for the dozens of protein polymers produced in our laboratory at g/L levels. Most of these proteins tend to largely be in random coil conformation under physiological conditions, and contain a Glu-Ala spacer. Besides ppaF, several other secretory prepro leaders have been developed for *P. pastoris* that require cleavage of a dibasic Kex2 site [188-190]. A Glu-Ala spacer is most likely needed for some proteins also with these leaders.

As mentioned above, ppaF is generally a highly successful secretory leader. Although it is still not entirely clear why ppaF is so efficient, insight into the importance of particular sequence elements has been gained from improvements made through directed evolution [191] and site-directed mutagenesis [192, 193]. A cell-biological study by Fitzgerald and Glick [194] provides further important clues

regarding the efficiency of ppaF. First, the α -factor pro region promotes rapid ER exit in *P. pastoris* through its interaction with the ER export receptor Erv29. Second, and more importantly, whereas translocation into the ER by the α -factor pre signal alone is inefficient, this process is stimulated by the pro region. Like most yeast signal peptides, the α -factor signal peptide is thought to drive posttranslational translocation of proteins into the ER. This presumably employs a Brownian ratchet, where the polypeptide chain goes back and forth in the translocon, and is captured in the ER by binding the Hsp70 chaperone Kar2. Possibly, the α -factor pro region provides a long stretch of unfolded protein N-terminal to the folded protein of interest, allowing efficient ratcheting by Kar2. Additionally, the three N-linked glycans attached to the pro region may prevent back-sliding of the chain in the translocon [194]. Possibly in agreement, although other mechanisms may be involved, several studies have shown that the glycosylation of ppaF is indeed important for efficient secretion [195-197]. Fast posttranslational translocation may help minimize premature folding in the cytosol, which could block or retard translocation [194].

In relation to premature folding, the *S. cerevisiae* Ost1 signal peptide used in the study by Fitzgerald and Glick [194] deserves attention as a potentially useful leader for secretory production of recombinant proteins. Unlike most yeast signal peptides, the Ost1 peptide directs cotranslational translocation. As such, this would allow entry into the ER prior to folding, and thereby eliminate a possible roadblock for secretion. The authors also constructed a hybrid leader, consisting of the Ost1 prepeptide and the α -factor propeptide. This leader stimulated ER export in *P. pastoris* relative to the Ost1 signal peptide alone.

Following analyses of the *P. pastoris* secretome [198, 199], several promising endogenous signal peptides have recently been identified. These appear to offer similar efficiency as ppaF, but without the need for a propeptide, at least for the reporter proteins used [190, 200-202]. However, it has been suggested that the sequence directly downstream of the signal peptide may affect translocation efficiency [202], and thus the efficiency of these signal peptides may be more variable for different cargo proteins than when using prepro peptides such as ppaF. Interestingly, also a putative homolog of the *S. cerevisiae* α -factor mating pheromone gene has been identified in a search of the *P. pastoris* genome [203]. The encoded secretory leader has a prepro structure similar to that of *S. cerevisiae* ppaF,

and features the characteristic Kex2 site and Glu-Ala spacer. However, it notably differs from the *S. cerevisiae* prototype in lacking potential *N*-glycosylation sites. The new *P. pastoris* prepro leader reportedly offered similar efficiency as ppαF for the secretion of *Bacillus subtilis* levanase [203]. It will be interesting to see more future side-by-side comparisons of these recent endogenous pre and prepro leaders with *S. cerevisiae* ppαF for the secretory production of various proteins.

9.2.3. Proteolysis

The main groups of proteases in yeast are the vacuolar proteases, the proteasome, and the proteases of the secretory pathway [204]. The vacuolar proteases are involved in maintaining proteostasis by degrading excess or damaged proteins, and in the degradation of proteins under nutrient stress to allow recycling of amino acids [205]. Both the vacuolar proteases and the proteasome participate in the elimination of misfolded proteins from the cytosol and the ER [206]. See also Section 9.2.6. The proteases of the secretory pathway are responsible for the proteolytic maturation of secreted precursor proteins [207]. Whereas, e.g., filamentous fungi are known to also secrete a variety of proteases [208, 209], only few secreted (or cell-bound) proteases have been identified in *P. pastoris* (Chapter 3) [210]. Nonetheless, particularly vacuolar proteases are present in the medium, mainly as a result of cell lysis [211, 212]. Overall, degradation of secreted proteins is still one of the major challenges when using *P. pastoris* as a host, affecting not only product yield but importantly also product integrity. This particularly applies to protein polymers, as these are generally more accessible to proteolytic enzymes than globular proteins because of their often (partially) unfolded structure. In fact, it is because of such accessibility that gelatin (denatured collagen) is commonly used as a substrate in protease assays. Furthermore, owing to the repetitive sequence of protein polymers, if a proteolytically sensitive motif is present, it likely occurs multiple times. For these reasons, protein polymers represent particularly challenging targets for secretory production, yet as such also provide interesting clues for studying and tackling proteolysis in *P. pastoris*.

If a given heterologous protein suffers from degradation in secretory mode, it may be worthwhile to investigate its fate in intracellular expression. It is clear, however, that the requirement to disrupt cells for purification of intracellularly expressed proteins results in their exposure to a variety of intracellular proteases

[211]. To minimize proteolysis of secreted proteins, culture conditions may be optimized, and protease-deficient strains or appropriate sequence design could be employed. Below, we will address these options, surveying what is known from the degradation of various proteins in *P. pastoris*, including protein polymers.

9.2.3.1. Control of proteolysis by optimization of culture conditions

Degradation of secreted proteins by proteases present in the extracellular medium can often be resolved by adjusting culture conditions. This will usually have to be determined empirically, as the range of proteases involved is highly dependent on the target protein.

Perhaps the first variable to be evaluated is culture pH. *P. pastoris* grows well over a fairly broad pH range from 3 to 7, which offers considerable latitude for optimization [145]. Numerous examples exist where adjusting culture pH was essential for obtaining intact proteins, however, the optimal pH depends on the protein of interest. For example, a hookworm anticoagulant peptide was degraded at around pH 4, and its proteolysis was minimized at around pH 7 [213]. Similarly, severe degradation of human serum albumin occurred at pH 4.3, while the protein was produced intact above pH 5.9 [214]. On the other hand, pH 3.0 was required to minimize degradation of insulin-like growth factor [215] and cytokine growth-blocking peptide [216]. We likewise found pH 3.0 to be optimal for the production of rat Col3a1 gelatin (Chapter 2), and an artificial polymer with a heterodimerizing leucine zipper module (Section 9.1.5.5) [125]. Whereas not all enzymes may be stable at low pH, protein polymers lack catalytic activity and complex tertiary structure, and can as such handle a wide pH range.

Another common approach to minimize proteolysis is the addition to the medium of rich supplements such as casamino acids, peptone, soytone, and yeast extract (Chapter 2) [217, 218]. These supplements probably act mainly as competing substrates for the proteases in the broth. However, their role as a nitrogen source may also be involved, as nitrogen starvation has been implicated in the proteolysis of human serum albumin and hirudin [214, 219]. The ability of rich supplements to serve as a nitrogen and carbon source may also explain the often observed positive effect on viability and protein yield [220-224]. A recent study additionally describes a positive effect of supplementation with casamino acids on the ratio of secreted to intracellularly retained protein [224]. Because

peptone was found to have more effect on yield than ammonia or a mixture of amino acids, it has furthermore been suggested that the complex oligopeptides in peptone may play a bioactive role [225]. In terms of protein polymers, supplementation of the medium with casamino acids effectively minimized the proteolysis of natural mouse Col1a1 gelatins (Chapter 2), and the polymer with a leucine zipper domain (Section 9.1.5.5).

Besides complex amino acid supplements, also L-arginine has been used. It presumably acts as a substrate analog specifically for extracellular trypsin-like proteases [226, 227]. Other protease inhibitors may be used to reduce degradation caused by specific protease classes. For example, the addition of 5-10 mM of the metalloprotease inhibitor EDTA can improve product stability, although it may also cause the accumulation of exo- β -1,3-glucanase in the medium [32, 228]. Shi et al. [227] found that the medium of methanol-induced shake flask cultures showed protease activity peaks at pH 3.5-4.5, pH 5.5-7.0, and 10.0, which the authors ascribed to aspartic, cysteine, and serine-type proteases, respectively. At pH 7.5, inclusion of the serine protease inhibitor Pefabloc SC reduced total protease activity by 53%, and the aspartic protease inhibitor pepstatin by 30%. Interestingly, addition of EDTA, and especially of the cysteine protease inhibitor E-64, actually increased the overall proteolytic activity, presumably by reducing degradation of serine proteases. Sinha et al. [212] added the serine protease inhibitor PMSF, as well as pepstatin, EDTA, and E-64 to the cell-free medium of a methanol-grown interferon-producing bioreactor culture. The first three inhibitors helped reduce both general and interferon-specific proteolysis. The cysteine protease inhibitor E-64 had no effect on total protease activity. Jansson et al. [7] found that the secreted fraction of their spider silk-like Z-4RepCT molecules was partially degraded, which could be inhibited to some extent by addition to the culture of pepstatin and EDTA, or a commercial protease inhibitor cocktail (Section 9.1.1.1). Although specific protease inhibitors can thus sometimes alleviate proteolysis in small-scale cultures, and can help identify the type of proteases involved [229], the associated costs limit their applicability in industrial processes.

It has been suggested that the specific growth rate (μ) can be used to control proteolysis. Zhou and Zhang [230] found severe degradation of secreted hirudin at $\mu < 0.02 \text{ h}^{-1}$ in methanol-fed cultures, but only little degradation at $\mu = 0.02\text{-}0.05 \text{ h}^{-1}$. Because the specific production rate of intact hirudin was highest at $\mu = 0.02 \text{ h}^{-1}$,

this was considered the optimal process value. This specific growth rate was then maintained in a fed-batch culture *via* a limiting exponentially increasing methanol feed rate, which resulted in less degradation and higher yield than when not controlling the specific growth rate. Sinha et al. [231] determined the total protease production at different specific growth rates in methanol fed-batch cultures for the production of ovine interferon- τ . The specific protease production rate displayed an optimum curve with a maximum at $\mu = 0.035 \text{ h}^{-1}$. The specific production rate of interferon- τ displayed a similar curve with a maximum at $\mu = 0.02 \text{ h}^{-1}$. The authors argue there is a balance between protein synthesis and degradation, and that, for any given strain, the optimal specific growth rate should be determined. It is clear that the growth rate affects many cellular processes that together determine that optimal value. Rebnegger et al. [232] studied the transcriptome of a *P. pastoris* strain secreting human serum albumin at $\mu = 0.015\text{--}0.15 \text{ h}^{-1}$ in glucose-limited chemostats. They found that, with increasing specific growth rate, genes associated with translation, ER translocation, folding, and the proteasome were upregulated, whereas autophagy and ER-Associated Degradation (ERAD) [233] were downregulated.

The use of a low growth temperature during the induction phase is known to effectively reduce the proteolysis of many secreted proteins. The temperature used is often between 20 and 26 °C [218, 229, 234–237]. However, thanks to the fairly psychrotrophic nature of *P. pastoris*, temperatures as low as 10 to 15 °C have successfully been used [221, 238]. For the protein polymers with heterodimerizing leucine zippers we could almost completely prevent proteolysis in methanol fed-batch cultures by using an induction temperature of 20 °C (Section 9.1.5.5). To allow the cells to adapt and prevent accumulation of methanol or toxic metabolic intermediates, the temperature is usually slowly ramped down during the first hours after methanol induction [221, 238, 239], or shortly before [125]. At a normal growth temperature, methanol nonlimited conditions tend to quickly result in oxygen limitation, for wild-type Mut⁺ (methanol utilization plus) strains having an intact *AOX1* gene and consequently growing fast on methanol [240]. To reduce proteolysis, as well as to maintain constant DO levels despite using a constant methanol concentration, Jahic et al. [240] developed a temperature-limited fed-batch (TLFB) technique. The TLFB regime was realized by programming the temperature controller of the fermentor to lower the temperature setpoint

whenever the DO was below setpoint. The process resulted in higher protein yields, improved cell viability, and reduced proteolysis. The authors suggest proteolysis is most likely minimized at low temperature because of simple thermodynamics, and because there is less lysis and accompanying release of proteolytic enzymes into the medium. Although conceptually interesting, the approach has found little use in the literature, probably because feedback control between DO and the relatively slow cooling process is not straight-forward [241]. Also, there may not be a clear benefit in practice over simply implementing constant low temperature during the methanol fed-batch phase, after a programmed cooling trajectory.

Traditionally, *P. pastoris* fermentations are typically run in a methanol-limited mode, where DO levels are maintained above a certain level. An alternative strategy is to maintain a constant methanol level in the broth using a sensor-controller [242, 243]. As mentioned above, oxygen limitation will generally occur under this regime, after a period of time determined by the oxygen transfer capacity of the bioreactor [244, 245]. Protein polymers are usually produced in our laboratory in such a methanol-sufficient, sensor-controlled, mode. In our experience, this provides a more robust process than methanol-limited feeding, hence minimizing the risk of cell lysis and related proteolysis. Trentmann et al. [246] even found that, whereas the methanol-limited process resulted in poor cell integrity and complete loss of product activity as a result of proteolytic degradation, the methanol-saturated and oxygen-limited process yielded high levels of active product. Also Charoenrat et al. [244] found that cell viability was improved by the oxygen-limited process, and much less host proteins (potentially including proteases) were released into the medium. It has been suggested that, in terms of production levels, an optimal specific methanol uptake rate of $0.02 - 0.03 \text{ g g}^{-1} \text{ h}^{-1}$ should be maintained, and that under the oxygen-limited regime this can be achieved by controlling aeration of the bioreactor, or by using a prolonged glycerol fed-batch phase so as to reach high biomass levels prior to induction [247].

Sinha et al. [212] have compared the proteolytic activity in *P. pastoris* bioreactor cultures using both methanol and glycerol as the carbon source. Cultures grown on methanol showed lower viability, and higher vacuolar and total protease levels in the extracellular medium, than those grown on glycerol. Although only a small fraction of nonviable cells actually lyse, decreased viability on methanol suggests

increased lysis, and thereby increased release of intracellular proteases into the medium. The authors furthermore propose that growth on methanol causes cell stress and consequent higher protease activity, and that leaky cell membranes may result from the action of the hydrogen peroxide and formaldehyde formed. The duration of the methanol induction phase thus becomes an important parameter for optimization. Also, obtaining a high biomass on glycerol prior to methanol induction, followed by a short production phase, may reduce proteolysis. It has furthermore been suggested that the concentration of methanol should be kept below about 3.6 g/L, because higher concentrations are inhibitory to growth and may thus cause cell death and release of proteases [248]. Finally, it may be worthwhile to investigate the use of mixed-feed approaches [221, 249, 250] when recalcitrant proteolysis is observed, or to use a promoter that does not rely on cultivation on methanol, such as the constitutive *GAP* promoter [251].

The addition of ascorbic acid to the medium as an antioxidant can reportedly increase cell viability and reduce proteolysis of hirudin [252]. The underlying reasoning is that reactive oxygen species, generated by methanol metabolism, may accumulate intracellularly, and that this leads to cell death and release of proteases. Ascorbic acid supplementation did not reduce degradation of the protein polymers with leucine zipper domains (Section 9.1.5.5), but this may have been due to differences in fermentation conditions, or disparate proteolytic sensitivity of the respective target proteins.

By reducing the salt content of standard basal salts medium [61] to one-quarter, and subsequently maintaining a fixed conductivity by salt feeding, Jahic et al. [253] were able to lower the frequency of dead cells by a factor of four. Presumably this was due to the reduced osmotic pressure of the modified medium. The approach also resulted in a lower concentration of proteases and other host proteins in the medium.

Wang et al. [254] reported that a continuous culture strategy, involving constant methanol and cell concentrations, resulted in increased cell viability and reduced accumulation of proteases in the extracellular medium. Consequently, proteolysis of the protein of interest was strongly decreased. Goodrick et al. [255] obtained similar results in glucose-fed continuous cultures using the *GAP* promoter. They argue that the cells do not reach stationary phase in this mode of fermentation, resulting in the release of less proteases into the medium.

Furthermore, the recombinant protein in the outflow, if regularly harvested, is exposed to less concentrated proteases for a shorter time. Continuous culture may be an often overlooked option to reduce degradation by proteases in the extracellular medium.

9.2.3.2. Protease-deficient strains and sequence design

If optimizing culture conditions does not resolve the degradation of a given protein, protease-deficient strains could be evaluated, particularly when there are indications as to the nature of the proteases involved. Table 9.3 lists known protease deletions in *P. pastoris*.

Table 9.3. *P. pastoris* protease gene knockouts

Genotype	Protease	Sequence specificity ^a	Localization ^b	Original report ^c
<i>prc1</i>	Carboxypeptidase Y	C-terminal, mainly hydrophobic AA	Vacuole	[256]
<i>pep4</i>	Protease A (saccharopepsin)	Broad	Vacuole	[211]
<i>prb1</i> <i>pep4 prb1</i>	Protease B (cerevisin)	Broad	Vacuole	
<i>kex1</i>	Carboxypeptidase D	C-terminal basic AA	Late Golgi	[257]
<i>kex2</i>	Kexin	Lys-Arg↓, Arg-Arg↓, some Arg↓	Late Golgi	Chapter 3
<i>yps1</i> <i>kex2 yps1</i>	Yapsin 1	Basic AA	Periplasm	
<i>ste13</i>	Dipeptidyl aminopeptidase A	N-terminal Xaa-Ala↓ or Xaa-Pro↓	Late Golgi	[258]
<i>dap2</i>	Dipeptidyl aminopeptidase B	N-terminal Xaa-Ala↓ or Xaa-Pro↓	Vacuole	
<i>sub2</i>	Putative subtilisin	Presumably broad	Secreted	[210]
<i>yps2, -3, -7, mkc7</i>	Various yapsins	Basic AA	Periplasm	[259]
<i>yps', yps"</i>	Putative yapsins	Presumably basic AA	Periplasm	[260]
<i>ybr139w</i>	Putative carboxypeptidase	Presumably C-terminal AA	Vacuole	[261]

^a Sequence specificities based on *S. cerevisiae* prototypes (see main text). Downward arrows indicate site of proteolytic cleavage. Abbreviations: AA, amino acids.

^b Known or putative primary localization. Active protease may be present in the extracellular medium owing to cell lysis or other mechanisms.

^c Some of these gene deletions have been reproduced in later studies, in varying genetic backgrounds. See main text [262].

Below we will describe these proteases and their sequence specificity, and the utility of the corresponding deletion mutants. Means of preventing proteolysis by sequence design will be discussed where applicable. The reader should note that,

in relation to sequence specificity, full accessibility to the protease is assumed, such as in random coil regions of protein polymers. In globular proteins, potential proteolytic sites may be buried inside the folded structure.

9.2.3.2.1. Vacuolar proteases. Much of the proteolytic activity in yeast is derived from vacuolar proteases, particularly during nutrient starvation [263]. As described in Section 9.2.3.1, these proteases may be present in the extracellular medium, mainly as a result of cell lysis [212], which can lead to the degradation of secreted proteins. Gleeson et al. [211] constructed *P. pastoris* strains devoid of the vacuolar protease A (SMD1168 strain; *pep4*), protease B (SMD1165 strain; *prb1*), or both (SMD1163 strain; *pep4 prb1*). Protease A, an aspartic protease currently referred to as saccharopepsin, has a broad specificity resembling that of cathepsin D and pepsin, generally favoring large hydrophobic residues on both sides of the scissile bond [264]. Phe, Leu, and Glu are most preferred at P1, and Phe, Ile, Leu, and Ala at P1' (see Figure 9.5B for cleavage site nomenclature). Protease B, a serine protease currently referred to as cerevisin, also has a broad specificity, which has been described as mixed chymotrypsin-like and trypsin-like [265]. Protease A proteolytically activates both protease B and the vacuolar protease carboxypeptidase Y. Consequently, in *pep4* strains, protease B activity is reduced by about 50%, and carboxypeptidase Y activity is completely absent [211]. A strain disrupted in carboxypeptidase Y has also been described [256]. Carboxypeptidase Y is a serine exopeptidase with broad specificity mainly for hydrophobic amino acids at P1 and P1', but it also hydrolyzes the peptide bond of C-terminal basic amino acids at discernable rates [266]. Although *pep4* and/or *prb1* strains have been useful in reducing degradation of various heterologous proteins, these strains are not as robust as wild-type strains [211, 267]. Their use has generally not reduced degradation of protein polymers in our laboratory, and often resulted in poor growth, lower yields, and inefficient product purification owing to higher levels of extracellular host proteins (Chapter 2, unpublished) [125]. Very recently, a *P. pastoris* homolog of the *S. cerevisiae* gene encoding a putative vacuolar serine carboxypeptidase (Ybr139w) was disrupted in a *prb1* strain [261]. This deletion did not appear to offer benefits for the secretory production of a bacterial carboxylesterase, and there are currently also no examples of the involvement of Ybr139w in the degradation of heterologous proteins.

9.2.3.2.2. Kex1 protease. The Kex1 protease of *S. cerevisiae* is a membrane-bound serine carboxypeptidase involved in the precursor processing of α -factor pheromone and K1 killer toxin [268] (see also Section 9.2.2), and is localized in the late Golgi [269]. Its prototypical activity is the removal of the C-terminal basic amino acids Lys and Arg, although a product inhibition assay showed a minor effect also with His [270]. Heim et al. [271] confirmed the preference for basic residues of *S. cerevisiae* Kex1, but also found hydrolysis of the C-terminal residues Gln⁶⁵ and Leu⁶⁴ of the thrombin inhibitor hirudin. The authors showed that the preceding Tyr⁶³ residue promoted this atypical processing, and that also Phe can play this role. Boehm et al. [257] developed a *kex1* mutant of *P. pastoris*, which prevented the removal of the C-terminal lysine of endostatin, at least in fermentations up to 40 h. It was speculated that degradation observed in longer runs was caused by carboxypeptidase Y released through cell lysis. Later studies also showed reduced degradation in *P. pastoris kex1* strains of hirudin [272] and insulin glargine [273]. The degradation of hirudin in wild-type *P. pastoris* resembled the atypical Kex1 cleavage observed by Heim et al. [271] in *S. cerevisiae*, in that the nonbasic C-terminal residues Gln⁶⁵ and Leu⁶⁴ (and Tyr⁶³) were removed [272]. Olsen et al. [9] reported that ~30% of a 9 kDa human Col1a1 gelatin lacked its C-terminal Arg residue, probably as a result of Kex1 cleavage. Our group has produced protein polymers in *P. pastoris* containing a C-terminal cationic (Lys)₁₂ tail for DNA-binding purposes (Section 9.1.5.4) [122]. This stretch of basic amino acids is in principle a canonical target for Kex1-mediated proteolysis. By adding a C-terminal Gly to the sequence, as an otherwise minimally obstructive residue, we ensured that the essential cationic tails would remain intact [122].

9.2.3.2.3. Kex2 protease. As described in Section 9.2.2, the subtilisin-like serine protease Kex2 is of critical importance to the secretory production of many proteins, because it cleaves at the pair of basic amino acids (Lys-Arg) present between the commonly used pp α F secretory leader and the protein of interest. However, proteins containing susceptible sites may be subject to internal digestion by Kex2 [274-278]. Although Kex2 is mostly known for its processing of dibasic motifs, the mono-arginyl sequence Pro-Arg in yeast K1 killer preprotoxin is also a natural substrate [279]. This ability to process Pro-Arg was also found in *in vitro* assays [173, 280, 281]. We similarly observed cleavage of Met-Gly-Pro-Arg sites in

Col1a1 gelatins in *P. pastoris*, which was initially resolved by site-directed removal of these sites (Chapter 2). Because we suspected the involvement of Kex2, we later constructed a *kex2* disruptant (Chapter 3). Indeed, endoproteolysis of the Met-Gly-Pro-Arg motifs of Col1a1 was abolished in this strain. Surprisingly, correct processing at the dibasic site of the ppαF leader used to drive secretion of Col1a1 still occurred in the *kex2* strain (Chapter 3). The protease responsible for this alternative processing of ppαF was found to be yapsin 1. Although efficient yapsin-dependent processing of ppαF may well depend on the protein fused to it, the assumption by Govindappa et al. [201] that the use of *kex2* disruptants requires leader sequences that do not rely on processing of dibasic motifs is too general. In fact, we have found that two other protein polymers with different N-terminal sequences were correctly processed in a *kex2* mutant (unpublished), showing that Kex2-independent processing of ppαF is not unique to Col1a1. However, we did notice that both Glu-Ala repeats remained attached to Col1a1 produced in the *kex2* strain, whereas normally only one such dipeptide is present in this protein (Chapter 3). This was also the case for the other two protein polymers, which normally contain only a minor fraction of molecules with a single Glu-Ala extension. Possibly, yapsin-mediated processing did not occur until beyond the Golgi with its DPAPase A (Section 9.2.2). Omitting the Glu-Ala spacer may thus be needed when using the *kex2* strain if the fully native N-terminus of the heterologous protein is required. However, we do not currently know whether this in turn would negatively affect yapsin-mediated processing of the dibasic site of ppαF, as is the case for the processing of some proteins by Kex2 (Section 9.2.2). Although the *kex2* strain grows more slowly than the wild type (~50% increased generation time in YPD medium at 30 °C), it shows normal viability and is suitable for high-biomass fermentation (Chapter 3).

Upon screening the literature for heterologous proteins proteolyzed in *P. pastoris* at arginine residues, we found several cases likely caused by Kex2 that were not recognized as such. In particular potentially susceptible mono-arginyl sites appear often overlooked. Besides Pro-Arg, mentioned above, also other mono-arginyl motifs have been shown to be Kex2 substrates in an *in vivo* assay [282]. Additionally, K28 killer preprotoxin is processed in *S. cerevisiae* by Kex2 at mono-arginyl Glu-Arg and Ser-Arg sites [283], and seems to be accordingly processed also in *P. pastoris* [152]. On the other hand, definitely not all arginines are

substrates, even in the absence of any folded structure, as evident from noncleavage or very limited cleavage in Col1a1 of sites other than Met-Gly-Pro-Arg (Chapter 2). Even a canonical dibasic Lys-Arg motif in Col1a1 was not a major site of cleavage (Chapter 2), which implies that the sequence context plays an important role. Bevan et al. [282] assessed the effect on *S. cerevisiae* mating of all 19 substitutions of Lys at the P2 position in a truncated ppaF gene. Similarly, Suzuki et al. [284] determined the effect on *in vitro* cleavage by soluble *S. cerevisiae* Kex2 of substitutions at the P3 and P4 positions in a fusion protein with a dibasic Kex2 recognition site. It is clear from these combined studies that the P4-P1 residues together influence Kex2 cleavage efficiency. As mentioned in Section 9.2.2, the prime side residues appear relatively unimportant. Rockwell and Fuller 1998 showed that there is interplay between the subsites of Kex2 that interact with P4 and P1, in that a favorable P4 residue can partially compensate for an unfavorable P1 residue [285]. It remains to be seen to what extent similar crosstalk exists between the other subsites.

Ever since the observed degradation of the Met-Gly-Pro-Arg mono-arginyl motifs in Col1a1 (Chapter 2), we have routinely screened protein sequences to be produced in *P. pastoris* using a position weight matrix approach based on the data of Bevan et al. [282], Suzuki et al. [284], and others, to tentatively predict the likelihood of degradation by Kex2 (unpublished). Bader et al. [176] similarly used a position weight matrix approach for the prediction of likely Kex2 substrates in several fungal genomes, although their proposed matrix only takes dibasic sites into account. The caveat in using position weight matrix approaches for predicting the likelihood of Kex2 cleavage of individual proteins is that the relative importance of, and possible interplay between, the P4-P1 positions are not yet sufficiently understood to derive a truly reliable automated prediction tool. To detect such complex correlations, a neural network approach like that used by Duckert et al. [286] for the prediction of general (mainly furin-type) proprotein convertase cleavage sites is more suited, but at present too few known monobasic substrate sequences are available for training. Nonetheless, informed use of the data of Bevan et al. [282] and Suzuki et al. [284], also taking into account sequence information from known Kex2 substrates, can assist the rational design of arginine-rich protein polymer sequences. For example, in Chapter 8, we incorporated repeated proteoglycan-binding KRSR tetrapeptides separated by flexible Gly/Ser-

rich spacers into a silk-inspired polymer in such a way that unfavorable Gly and Pro residues occupied the P4 and P3 positions, respectively. This almost entirely prevented Kex2-mediated degradation, despite the presence of the canonical dibasic Lys-Arg motif.

9.2.3.2.4. Yapsins. The yapsin family of aspartic proteases consists of five members in *S. cerevisiae* (Yps1, -2, -3, -6, and -7), and homologs are found in various fungi [287]. These proteases are localized at the cell surface *via* a glycosylphosphatidylinositol (GPI) anchor [287]. Their natural function appears to be in the activation and/or shedding of periplasmic proteins during cell wall stress and remodeling [288, 289]. Yps1 (yapsin 1, previously referred to as Yap3) is the most studied yapsin. Although the protease is mainly active at the plasma membrane, it is also found in the extracellular medium and transiently active in the late secretory pathway [290-292]. It is involved in the degradation of several heterologous proteins produced in *S. cerevisiae* [292].

The substrate specificities of the *S. cerevisiae* aspartic yapsin proteases [293-295] overlap with those of Kex2 [296], in that they both cleave C-terminally of basic residues. In fact, Egel-Mitani et al. [293] showed that overexpression of yapsin 1 led to the production of mature α -factor in a *kex2* strain of *S. cerevisiae*. However, there are notable differences between the two protease classes [287]. First, Kex2 primarily recognizes residues on the N-terminal side of the scissile bond, whereas residues on both sides are important for yapsins. Second, whereas Kex2 is highly specific for Arg at P1, yapsins accept both Lys and Arg residues at this position. Third, whereas Kex2, despite its ability to process particular monobasic substrates, is primarily specific for dibasic sites, yapsins have a truly monobasic specificity. Nonetheless, additional nearby basic residues do appear to enhance catalytic efficiency. Pro at P2 is probably not tolerated, as cleavage by yapsin 1 of an N-terminal EEAEAEAK extension of insulin precursor was abolished by insertion of Pro before the Lys residue [172], and also yapsin 2 appears to exclude proline from P2 [297].

To establish whether yapsin 1 was involved in the formation of minor background bands seen in SDS-PAGE of Col1a1 gelatin (Chapter 2), we generated a *yps1* strain of *P. pastoris* (Chapter 3). Overall yapsin-like activity was decreased by 95% in this strain, suggesting that Yps1 is by far the major yapsin in the yeast.

As the pattern of proteolytic fragments of Col1a1 obtained in this strain was not significantly different from that observed in the wild type, yapsin 1 was apparently not involved in degradation of Col1a1. However, on other occasions the strain has proven to be essential in preventing degradation of protein polymers. Partial degradation of the collagen-inspired triblock $T_{16}-C^p_4-T_{16}$ (Section 9.1.5.2), as well as that of the disulfide bond-forming related polymer $T_9-C^R_4-T_9-Cys$ (Section 9.1.5.3), could be completely overcome by using the *yps1* strain. Interestingly, in both cases the observed proteolysis was most likely triggered by the high thermal stability of the collagen trimer-forming modules (Section 9.2.6). A number of publications confirm the usefulness of *YPS1* disruption for the production of various proteins in *P. pastoris* [298, 299]. Guan et al. [259] disrupted four more yapsin-like genes, but did not study their utility for the production of heterologous proteins. Wu et al. [260] disrupted the same set, as well as two more putative yapsin genes. However, reduced degradation of a serum albumin / parathyroid hormone fusion protein was only observed upon disruption of *YPS1*. The additional disruption of the *PEP4* gene resulted in further decreased degradation. A study by Sazonova et al. [299] confirmed our findings (Chapter 3) that disruption of *YPS1* does not result in phenotypic defects. Indeed, in our experience the strain is robust and suitable for high cell density fermentations at a normal growth rate. Note that a *kex2 yps1* double disruptant does show slow growth, and that the ppαF leader is aberrantly processed in this strain (Chapter 3).

9.2.3.2.5. Dipeptidyl aminopeptidase A. As described in Section 9.2.2, the Golgi-localized dipeptidyl aminopeptidase A (DPAPase A; a serine protease encoded by the *STE13* gene) is involved in the processing of the α-factor mating propheromone in *S. cerevisiae*, stepwise removing N-terminal Xaa-Ala dipeptides. Similarly, when using the ppαF secretory signal to drive secretion of heterologous proteins in *P. pastoris*, this yeast's DPAPase A homolog is essential for the removal of the Glu-Ala spacer commonly placed between the ppαF leader and the protein of interest. Unfortunately, DPAPase A may also cause undesired N-terminal degradation. Not only N-terminal Xaa-Ala dipeptides are a substrate of DPAPase A in *S. cerevisiae* and other yeasts, but also N-terminal Xaa-Pro dipeptides [156, 281, 300-302]. Indeed, elastin-like protein polymers produced in *P. pastoris* contained 80% of species lacking the first four residues GPVP [15]. A variant of *Plasmodium*

falciparum circumsporozoite protein similarly suffered deletion in *P. pastoris* of no less than 30 Asn-Ala and Asn-Pro dipeptides from its N-terminus [303], undoubtedly caused by DPAPase A. Interestingly, Prabha et al. [258] showed that the protease may also attack other than the canonical Xaa-Ala and Xaa-Pro dipeptides, as in their case an N-terminal His-Gly dipeptide was removed. This proteolysis was successfully abolished upon disruption of the *P. pastoris* *STE13* gene homolog [258]. The vacuolar dipeptidyl aminopeptidase B (a serine protease encoded by the *DAP2* gene), which has been shown a weak suppressor of *ste13* defects in *S. cerevisiae* with similar substrate specificity [156], did not seem to be involved in this degradation. Hopkins et al. [304] similarly found that DPAPase A is the major dipeptidase involved in the removal of N-terminal Xaa-Ala or Xaa-Pro dipeptides from three different proteins expressed in *P. pastoris*, but describe how for two of these it was necessary to disrupt both the *STE13* and *DAP2* genes to completely prevent truncation.

9.2.3.2.6. Extracellular subtilisin. Salamin et al. [210] found a 37 kDa serine protease in the extracellular medium of methanol-grown *P. pastoris* cells. The amino acid sequence of the protease referred to as Sub2 showed similarity to fungal subtilisins, but it was not a homolog of vacuolar cerevisin. The presence of a putative signal peptide suggests it is secreted from the cells, and not just present in the medium as a result of cell lysis. The protease was able to cleave the synthetic substrate N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, but its sequence specificity was not further investigated. A *sub2* strain of *P. pastoris* GS115 was created, but so far its use to prevent degradation of heterologous proteins has not been reported.

9.2.3.2.7. Overexpressed protease inhibitor. Although not strictly protease-deficient, an interesting strain was constructed that expresses a 55-amino acid BPTI-Kunitz-type protease inhibitor [305]. This inhibitor from the sea anemone *Stichodactyla helianthus* has broad specificity against serine, aspartic, and cysteine proteases. The authors cocultivated a strain producing the inhibitor with a strain producing the virus capsid protein of interest, and they also constructed a strain coexpressing both proteins. Both approaches strongly reduced degradation of the viral protein. Addition of the purified inhibitor to the medium of *P. pastoris* cultures also offered protection against degradation of the viral protein [305], and of a mini-proinsulin [306].

9.2.4. Protein glycosylation

The capacity of a heterologous host to perform glycosylation is important for the production of therapeutic glycoproteins [307] and in general for proteins that require glycosylation for stability and proper folding [308, 309]. Although *P. pastoris* and other yeasts are capable of glycosylation, they produce high-mannose structures, contrasting the more diverse composition of mammalian glycans [310]. Glycans attached by *P. pastoris* are significantly shorter than the hypermannosylated structures characteristic of *S. cerevisiae* [311]. Nonetheless, their nonhuman nature could potentially still render them immunogenic and could negatively affect the pharmacokinetic properties of therapeutic proteins [312]. Much effort has therefore been made to engineer *P. pastoris* toward the production of proteins with humanized glycan structures [313, 314].

For protein polymers, glycosylation is generally undesirable because this would lead to oligodisperse product and may perturb intended intra- or intermolecular interactions. *N*-glycosylation can in principle be prevented simply by ensuring that the motif Asn-Xaa-Ser/Thr (Xaa ≠ Pro) [315] does not occur in the sequence design of the polymer. It should be noted, however, that a very rare case of *N*-glycosylation in *P. pastoris* of the noncanonical motif Asn-Xaa-Cys has also been described [316].

O-glycosylation also occurs in *P. pastoris* [317]. Unfortunately no known consensus motif exists for *O*-glycosylation [318]. The predictive neural network NetOGlyc [318, 319] is only of limited help when designing protein sequences to be expressed in yeasts, because it has been trained on mammalian *O*-glycosylation sites. NetOGlyc 3.1 has been evaluated for fungal sequences and performs relatively poorly in terms of predicting individual *O*-glycosylation sites, and significantly overestimates the number of sites [320]. Additionally, depending on the version of the tool, NetOGlyc uses predictive algorithms for surface accessibility and secondary structure, which are likely to fail for nonglobular protein polymers. Preventing *O*-glycosylation through sequence design thus seems rather difficult.

Because disruption of individual *O*-mannosyltransferase genes results in serious growth defects, deleting the entire gene family is not a viable scenario to obtain proteins devoid of *O*-glycans [304, 321]. While it is possible to reduce the extent of *O*-glycosylation by addition of chemical *O*-mannosyltransferase

inhibitors to the fermentation medium, or to enzymatically remove *O*-glycans from proteins produced in appropriately glycoengineered strains [304, 321], these options are cost-prohibitive at large scale.

Returning to the issue of the relation between sequence and *O*-glycosylation, we note that serine/threonine-rich sequences are known to be relatively susceptible to *O*-glycosylation, particularly when prolines are in the vicinity of the hydroxyl residues [322]. While this observation may serve as a caution, of course it does not allow to actually predict *O*-glycosylation of a given protein sequence. For example, the protein polymer C^P_4 produced in *P. pastoris* is not glycosylated, despite having a serine/proline-rich sequence with ~22% Pro and ~8% Ser (Chapter 4). Interestingly, we found that the same polymer C-terminally extended with EYPPYPPPPYPSG (C^P_4 - D^{PPXY}) was *O*-glycosylated at the single serine in this peptide (Chapter 7). Because C^P_4 is a random coil (unfolded) protein, accessibility does not likely play a role in the exclusive glycosylation of serine in this C-terminal extension. It is striking that the single (glycosylated) serine in the C-terminal extension is preceded by proline, while all (nonglycosylated) serines in the C^P_4 sequence are followed by proline. This would seem to concur with the reported enhancement of *O*-mannosyl transfer in *S. cerevisiae* for peptides with proline at the N-terminal side of the hydroxyl amino acid [323], and with its inhibition when proline is at the C-terminal side [324]. However, in accordance with the above-described elusive specificity of *O*-glycosylation, such simple motifs are insufficient to determine glycosylation fate. Indeed, the variant of C^P_4 with randomized sequence, C^R_4 (Chapter 6), contains several Pro-Ser motifs like that in the D^{PPXY} module, yet the protein is not glycosylated. Vice versa, at least one protein produced in *P. pastoris* is *O*-glycosylated at a Ser residue followed by a potentially interfering Pro [325, 326]. As mentioned in Chapter 7, the population of glycan structures attached to serine in the C-terminal extension in C^P_4 - D^{PPXY} was found to be heavily phosphorylated, where phosphorylated mannose was present in diester linkage.

9.2.5. Protein phosphorylation

Protein phosphorylation represents the foremost regulatory mechanism of cellular processes in eukaryotes, and it is estimated that about 30% of the human proteome is phosphorylated [327]. In *S. cerevisiae*, there are 120 known kinases that

act on serine and threonine residues, and two that modify histidines [328]. Some of the former can also use tyrosine residues as a substrate [328]. The *P. pastoris* GS115 genome contains 152 annotated kinases [329]. With such an abundance of kinases one might expect that many heterologous proteins produced in yeasts could inappropriately become phosphorylated, but relatively few examples can be found in the literature. In part this may be explained by the fact that many studies do not characterize the proteins produced in sufficient detail to detect the addition of one or more phosphate groups. Additionally, cellular signaling pathway fidelity requires that substrate recognition by kinases is quite specific, and there are considerable differences between the kinase profiles of yeasts and mammals [330-332]. As with glycosylation, or any unintended posttranslational modification for that matter, phosphorylation of protein polymers would lead to undesired heterogeneity and altered physicochemical properties. There are several dozen computational tools that allow to predict phosphorylation of input sequences [331], including a yeast-specific one, NetPhosYeast [330]. Notwithstanding their important utility in the elucidation of cellular signaling pathways, in our experience these tools are of very limited value in the *de novo*-design of protein polymers because of dramatic overprediction.

In *P. pastoris*, only few studies reported protein phosphorylation of secreted or intracellularly produced heterologous proteins [9, 333-337]. Among these, the study by Olsen et al. [9] involves a protein polymer. The polymer is a ~9 kDa gelatin derived from human type I collagen $\alpha 1$ chain, produced in secretory fashion using the ppaF leader. About 50% of the protein was modified with a single phosphate, and ~10% contained two phosphates. The sites of phosphorylation were identified as the serines in the sequence (Gly-Ser-Pro)₂, where the second Ser residue was phosphorylated to a much larger degree than the first. The authors note that the synthetic gelatin C^P₄ (Chapter 4), which is likewise secreted using the ppaF leader, also contained many Gly-Ser-Pro triplets. Since these are not phosphorylated, the *P. pastoris* kinase responsible for phosphorylation of the collagen fragment probably recognizes a wider sequence context. Indeed, at least 9-12 substrate residues are thought to be in contact with the active site of kinases, and possibly many more neighboring residues are involved [331]. Olsen et al. do not speculate on the kinase responsible, but we note that, while the majority of kinases disfavor Pro following the phosphorylated

residue, the so-called proline-directed kinases primarily select for Pro at this position [338].

9.2.6. Self-assembly *in vivo*

Most protein polymers described in Section 9.1 are relatively hydrophilic and unfolded unless triggered to self-assemble. The latter ideally occurs only *in vitro* after production, typically in response to an environmental trigger such as pH or temperature. In fact, as protein folding is often considered to be the major rate-limiting bottleneck in secretory production [339], the lack of need for folding and corresponding interaction with chaperones may be why many of these polymers are secreted so efficiently, despite the demands placed on the translation machinery by their biased amino acid composition. However, some protein polymers may self-assemble into supramolecular structures already *in vivo*. If the protein polymer was targeted to the cytoplasm, then the formation of such structures may not necessarily impair product yield and quality. In fact, the protein may well sequester into stable inclusion bodies similar to those commonly seen in *E. coli*, as was the case in *P. pastoris* for the aggregation-prone green fluorescent protein variant VP1GFP [340]. If, however, the protein polymer was targeted for secretion, then self-assembly during passage through the secretory pathway is likely to result in its nonsecretion, as we will discuss below. Although not a protein polymer, hepatitis B surface antigen is an example of a protein similarly not secreted by *P. pastoris* because it assembled into large lamellar structures in extended ER [341].

It is clear from Section 9.1.1.1 that extracellular targeting of spider silks tends to result in intracellular retention of most of the product [2, 3, 7]. Self-assembly *in vivo* is a likely cause, given the strong tendency of silk-like sequences to form β -sheets and amyloid structures. It is all the more surprising that, e.g., the silk-like protein $S^{H_{48}}$, and the fibril-forming triblock copolymer $C^{P_2}-S^{H_{48}}-C^{P_2}$ are produced as secreted proteins at g/L levels (Sections 9.1.1.2 and 9.1.5.1). The pH of the yeast secretory pathway progressively drops from pH 7 in the ER to pH 6 in the late Golgi, through to pH 5.2 in the secretory vesicles [342]. Particularly the ER and Golgi are thus in a pH range conducive to self-aggregation ($S^{H_{48}}$) or fibril formation ($C^{P_2}-S^{H_{48}}-C^{P_2}$), as the silk-like block folds and stacks *in vitro* at pH >6 [111]. The hydrophilic end blocks and the slow kinetics of fibril formation likely play a role in

the case of $C^{P_2}-S^{H_{48}}-C^{P_2}$, but this does not explain why $S^{H_{48}}$, which readily aggregates at neutral pH *in vitro*, is secreted so efficiently. Apparently, there is a delicate balance between a protein's inclination to aggregate and its secretion, as governed by secretion kinetics, transient local concentrations *in vivo*, and the presence and possible interactions with biological cosolutes.

Collagen represents an example where self-assembly *in vivo* certainly occurs and causes nonsecretion. As described in Section 9.1.2.2, the Myllyharju group found that collagens produced in a strain overexpressing P4H are retained intracellularly, and that these appear to accumulate in the ER. They devised a series of collagen fragments to study this phenomenon more closely [13]. Using the ppaF leader in a *P. pastoris* strain expressing human P4H, collagen $\alpha 1(I)$ fragments were found to be secreted as single-chain polypeptides. In contrast, upon induction of triple helix formation by attachment of a C-terminal propeptide, the collagen molecules were fully retained intracellularly. Also when the small trimerizing foldon domain from bacteriophage T4 fibritin was used to induce trimerization, molecules accumulated inside the cells, and only a fraction of the molecules was secreted. This was even the case when only a small 9 kDa $\alpha 1(I)$ fragment with foldon domain was expressed. This shows that the triple-helical conformation prohibits secretion, independent of the size of the collagen fragment. The authors propose that, whereas in animal cells triple-helical collagen travels the Golgi complex by cisternal maturation, such a transport system may not be present in yeast. Indeed, also triple-helical collagen expressed in *S. cerevisiae* was retained intracellularly [343].

We similarly found poor secretion of the $T_{16}-C^{P_4}-T_{16}$ triblock copolymer with triple helix-forming collagenous end blocks (Section 9.1.5.2). Whereas $T_9-C^{P_4}-T_9$ is efficiently secreted and intact, the $T_{16}-C^{P_4}-T_{16}$ variant with longer end blocks, and thus higher melting temperature, is produced at a five times lower yield and partially degraded. In line with the above findings by the Myllyharju group, the formation of triple helices stable at growth temperature may prevent efficient secretion of $T_{16}-C^{P_4}-T_{16}$. Alternatively, it could be argued that the increased length of the proline-rich $(Pro-Gly-Pro)_n$ stretches may negatively impact ribosomal translation, or may, through increased hydrophobic interactions, result in protein aggregation. To discriminate between these possibilities, we produced molecules with randomized end blocks and otherwise unchanged amino acid composition

[114]. The fact that these, by definition nontrimerizing, molecules were secreted at high yield, shows that indeed trimer-formation disturbs secretion. Because $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ triblock copolymers can form gels upon trimerization of the end blocks, we investigated whether secretion is also compromised in diblock copolymers with only a single T_{16} block, which are by definition unable to form gels. To this end, we produced $\text{T}_{16}\text{-C}^{\text{P}_4}$ and $\text{C}^{\text{P}_4}\text{-T}_{16}$ diblock copolymers (unpublished). Although these were found to be degraded similarly as the triblocks, their yields were intermediate between those of $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ and the well-produced $\text{T}_9\text{-C}^{\text{P}_4}\text{-T}_9$. This suggests that, although triple helix formation as such is the main cause of poor secretion and degradation, gel formation does aggravate the condition further. The degradation of $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ was resolved by using the *yps1* strain reported in Chapter 3 (see also Section 9.2.3.2.4), which shows that trimer-formation in some way elicited yapsin 1-mediated proteolysis. If, like fully triple-helical hydroxylated collagen, some of the $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ molecules accumulate as trimers in the ER, this might indirectly trigger degradation of the secreted monomeric fraction by yapsin 1. A possible mechanism for this may be found in the fact that ER stress strongly induces expression of *YPS1* in *S. cerevisiae* [344]. Alternatively, if trimeric $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ molecules in fact do travel beyond the ER, these might directly trigger degradation by yapsin 1 as follows. Possibly, trimeric molecules travel the secretory pathway or cross the cell wall at a relatively low rate, which would increase their exposure to yapsin 1 activity in the late secretory pathway or in the periplasmic space. Alternatively, traversal of the cell wall by trimeric $\text{T}_{16}\text{-C}^{\text{P}_4}\text{-T}_{16}$ might trigger the cell wall integrity response, which in turn could trigger increased *YPS1* expression [288].

As mentioned in Section 9.1.5.3, we also found yapsin 1-mediated degradation of $\text{T}_9\text{-C}^{\text{R}_4}\text{-T}_9\text{-D}^{\text{C}}$. Apparently, the demonstrated triple helix-stabilizing effect of the C-terminal disulfide bridge-forming cysteine residue in this molecule triggers degradation in the same manner as when using long T_{16} end blocks [118].

Elastin-like proteins produced in *P. pastoris* with a low transition temperature [16] may represent another example of unfavorable self-assembly *in vivo*. As described in Section 9.1.3, the hydrophobic $\text{V}_5\text{L}_2\text{G}_3\text{-40}$ was produced at a much lower yield at a growth temperature of 30 °C than $\text{V}_5\text{A}_2\text{G}_3\text{-40}$, and was partially degraded. In cultures grown at 20 °C, the yield of $\text{V}_5\text{L}_2\text{G}_3\text{-40}$ was increased by a factor of six, and the protein was no longer degraded. We propose that at a low

growth temperature, V₅L₂G₃-40 is below its transition threshold and soluble during its travel down the secretory pathway. In contrast, at 30 °C, V₅L₂G₃-40 may undergo dehydration and transition to a more ordered β -strand-rich structure with exposed hydrophobic residues [345], which results in coacervation. This in turn could lead to retention of the protein in the secretory pathway, and/or trigger the unfolded protein response (UPR), which recognizes misfolded proteins through exposed hydrophobic regions normally buried in the protein core [346]. Degradation of the protein may then occur through ERAD [233], or ER-phagy [347]. For studies on the relation between secretion levels of aggregation-prone proteins in *P. pastoris* and the UPR, ERAD, and ER-phagy, see Kumita et al. [348] and Whyteside et al. [349]. Also see a proteomics study by Vanz et al. [350], where the highly hydrophobic hepatitis B surface antigen assembled into lamellar deposits in the ER and induced the UPR and ERAD pathways. Because the pp α F secretory signal used for V₅L₂G₃-40 directs posttranslational ER translocation (Section 9.2.2), it is also possible that the ELP coacervates at 30 °C already in the cytoplasm prior to, or during, translocation. The protein may then be targeted for proteasomal degradation [351], or proteolytically cleared from the translocon [352], respectively. The intermediate yield and nondegradation of V₅A₂G₃-90, which has a similar transition temperature *in vitro* as the more hydrophobic V₅L₂G₃-40 owing to its length, and the fact that growth at low temperature did not increase the yield of V₅A₂G₃-90 (Section 9.1.3), suggest that ELP chain length and hydrophobicity differently affect the *in vivo* transition temperature, intracellular retention, or stress responses.

The observed pH-dependence of ELP yield in *P. pastoris* (Section 9.1.3) may indirectly also be related to self-assembly *in vivo*. The yeast grows well in the pH range 3 – 7 (Section 9.2.3.1), and the synthesis of heterologous proteins as such is not normally dependent on extracellular pH. Hypothetically, *complete* degradation in the medium of *a fraction* of the molecules specifically at low pH might cause lower yield. However, incubation of intact ELP (produced at pH 6) with both cell-free and cell-containing medium from a non-ELP culture grown at pH 3 did not result in degradation. The effect can furthermore hardly be explained by pH-responsive behavior of the ELPs *in vivo*, as these do not contain any amino acids with charged side groups. Another hypothesis, namely reduced performance of the ITC purification procedure at low pH, was experimentally ruled out. A possible

explanation might be low permeability of the yeast cell wall at low pH. This is evident from studies with *S. cerevisiae* that showed reduced sensitivity at low pH to β -1,3-glucanase or Zymolyase [353, 354], which is a measure for the porosity of the cell wall [355-357]. A less permeable cell wall at low pH may slow down diffusion of the ELP from the periplasm into the extracellular medium, and ensuing high local polymer concentrations might trigger coacervation and intracellular retention. We are aware of the speculative nature of the foregoing; further study is needed.

For short self-assembling sequences, we found that their self-assembly may be prevented by fusion to a large hydrophilic carrier domain [358]. We employed this strategy in the production in *P. pastoris* of two strongly self-assembling undecapeptides developed by Aggeli et al. [359]. Initial attempts to secrete the free peptides using ppaF (Section 9.2.2) did not result in any detectable product. Subsequently, the peptides were produced as enterokinase-cleavable fusions to the C-terminus of the highly hydrophilic random coil protein polymer C^P₂ (i.e., half of C^P₄; Chapter 4). Also here, ppaF was used to drive secretion. The fused C^P₂ carrier allowed convenient detection and purification. The fusion proteins were recovered from the cell-free broth at 0.6-1.0 g/L, which corresponds to 45-75 mg/L of free peptide. Whereas the fusion protein did not self-assemble *in vitro*, the free peptide released by enterokinase digestion formed a self-supporting gel. This shows that the highly soluble C^P₂ domain efficiently suppressed self-assembly, which may be applicable also to other self-assembling peptides.

9.3. Conclusions and future prospects

Although the bacterium *E. coli* has been the main microbial workhorse for production of heterologous proteins including protein polymers, *P. pastoris* has emerged as an attractive eukaryotic alternative. The repetitive genes encoding protein polymers are stably integrated into the yeast's genome without the need for selective pressure, and many of the protein polymers produced in *P. pastoris* have been successfully expressed as secreted proteins. The latter has the major advantage of providing a first, very effective, purification step. This is a particularly important factor in the study of protein polymers, where physicochemical characterization of the resulting supramolecular materials requires considerable amounts. With relatively little purification effort, significant

amounts of protein polymers can often be obtained using bench-top bioreactors. Moreover, simple downstream processing is important in the context of commercial viability. As we have seen, interesting functional nanomaterials have been prepared from protein polymers produced in *P. pastoris*, some of which hold promise for further development toward biomedical applications.

Although obviously no single ideal host for all proteins exists, *P. pastoris* does seem particularly efficient for silk-like and gelatin-like proteins, in that it routinely provides g/L levels for most of such polymers. Elastin-like proteins with high transition temperatures are also expressed well in *P. pastoris*. However, the yeast seems currently less suited for the production of hydrophobic ELPs, as these appear to self-assemble during secretion (Sections 9.1.3 and 9.2.6). In this respect, intracellular expression of ELPs in *P. pastoris* would be worth evaluating, as well as fine-tuning expression levels. The possibility to accurately hydroxylate human collagenous sequences in appropriately engineered *P. pastoris* is a clear benefit over current bacterial systems, although the nonsecretion of trimeric collagen remains to be solved.

As discussed in Section 9.2, various solutions are available to challenges such as proteolysis and other undesired posttranslational modifications. With the availability of the *P. pastoris* genome and synthetic biology tools, many future improvements will expectedly be realized. Forthcoming increased understanding of secretion bottlenecks and the cellular stress responses evoked by protein overproduction [360] will be invaluable to increase the likelihood of successful production of any given protein, including protein polymers. Expanding the genetic code of heterologous hosts for the production of novel protein polymers with noncanonical amino acids is of great interest [361], and efforts in that direction with *P. pastoris* are encouraging [362]. Finally, also theoretical insight into the bioprocessing of the yeast is accumulating. These recent advances will undoubtedly further the utility of *P. pastoris* as an industrial host for the production of protein polymers.

References

1. Fahnestock SR, Bedzyk LA (1997). *Appl Microbiol Biotechnol* 47, 33-39.
2. Fahnestock SR, Yao Z, Bedzyk LA (2000). *J Biotechnol* 74, 105-119.
3. Teulé F, Aubé C, Ellison M, Abbott A (2003). *AUTEX Res J* 3, 160-165.
4. Bogush VG, Sokolova OS, Davydova LI, Klinov DV, Sidoruk KV, Esipova NG, Neretina TV, Orchanskyi IA, Makeev VY, Tumanyan VG, et al. (2009). *J Neuroimmune Pharmacol* 4, 17-27.
5. Sokolova OS, Bogush VG, Davydova LI, Polevova SV, Antonov SA, Neretina TV, Klinov DV, Debabov VG, Kirpichnikov MP (2010). *Mol Biol* 44, 150-157.
6. Gaines WA, Marcotte WR, Jr. (2011). *AATCC Rev* 11, 75-79.
7. Jansson R, Lau CH, Ishida T, Ramstrom M, Sandgren M, Hedhammar M (2016). *Biotechnol J* 11, 687-699.
8. Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J (2003). *Adv Drug Deliv Rev* 55, 1547-1567.
9. Olsen D, Jiang J, Chang R, Duffy R, Sakaguchi M, Leigh S, Lundgard R, Ju J, Buschman F, Truong-Le V, et al. (2005). *Protein Expr Purif* 40, 346-357.
10. Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI (1997). *EMBO J* 16, 6702-6712.
11. Nokelainen M, Tu H, Vuorela A, Notbohm H, Kivirikko KI, Myllyharju J (2001). *Yeast* 18, 797-806.
12. Myllyharju J, Nokelainen M, Vuorela A, Kivirikko KI (2000). *Biochem Soc Trans* 28, 353-357.
13. Pakkanen O, Pirskanen A, Myllyharju J (2006). *J Biotechnol* 123, 248-256.
14. Pozzolini M, Scarfi S, Mussino F, Salis A, Damonte G, Benatti U, Giovine M (2015). *J Biotechnol* 208, 28-36.
15. Schipperus R, Teeuwen RLM, Werten MWT, Eggink G, de Wolf FA (2009). *Appl Microbiol Biotechnol* 85, 293-301.
16. Schipperus R, Eggink G, de Wolf FA (2012). *Biotechnol Prog* 28, 242-247.
17. Sallach RE, Conticello VP, Chaikof EL (2009). *Biotechnol Prog* 25, 1810-1818.
18. Tsai WL, Forbes JG, Wang K (2012). *Protein Expr Purif* 85, 187-199.
19. Meyer DE, Trabbic-Carlson K, Chilkoti A (2001). *Biotechnol Prog* 17, 720-728.
20. Hinman MB, Jones JA, Lewis RV (2000). *Trends Biotechnol* 18, 374-379.
21. Tokareva O, Michalczychen-Lacerda VA, Rech EL, Kaplan DL (2013). *Microb Biotechnol* 6, 651-663.
22. Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999). *J Exp Biol* 202, 3295-3303.
23. Cetinkaya M, Xiao S, Markert B, Stacklies W, Grater F (2011). *Biophys J* 100, 1298-1305.
24. Blackledge TA, Perez-Rigueiro J, Plaza GR, Perea B, Navarro A, Guinea GV, Elices M (2012). *Sci Rep* 2, 782.
25. Vollrath F, Knight DP (2001). *Nature* 410, 541-548.
26. Altman GH, Diaz F, Jakuba C, Calabro T, Horan RL, Chen J, Lu H, Richmond J, Kaplan DL (2003). *Biomaterials* 24, 401-416.
27. Hakimi O, Knight DP, Vollrath F, Vadgama P (2007). *Compos Part B-Eng* 38, 324-337.
28. Prince JT, McGrath KP, DiGirolamo CM, Kaplan DL (1995). *Biochemistry* 34, 10879-10885.
29. Lewis RV, Hinman M, Kothakota S, Fournier MJ (1996). *Protein Expr Purif* 7, 400-406.
30. Fahnestock SR, Irwin SL (1997). *Appl Microbiol Biotechnol* 47, 23-32.
31. Romanos M, Scorer C, Sreekrishna K, Clare J (1998). *Methods Mol Biol* 103, 55-72.
32. Sreekrishna K, Brankamp RG, Kropp KE, Blankenship DT, Tsay JT, Smith PL, Wierschke JD, Subramaniam A, Birkenberger LA (1997). *Gene* 190, 55-62.
33. Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, Urdea MS, Valenzuela P, Barr PJ (1984). *Proc Natl Acad Sci USA* 81, 4642-4646.
34. Bogush VG, Sidoruk KV, Davydova LI, Zalunin IA, Kozlov DG, Moisenovich MM, Agapov, II, Kirpichnikov MP, Debabov VG

- (2011). *Dokl Biochem Biophys* 441, 276-279.
35. Agapov, II, Pustovalova OL, Moisenovich MM, Bogush VG, Sokolova OS, Sevastyanov VI, Debabov VG, Kirpichnikov MP (2009). *Dokl Biochem Biophys* 426, 127-130.
36. Moisenovich MM, Pustovalova OL, Arhipova AY, Vasiljeva TV, Sokolova OS, Bogush VG, Debabov VG, Sevastyanov VI, Kirpichnikov MP, Agapov, II (2011). *J Biomed Mater Res A* 96, 125-131.
37. Jansson R, Thatikonda N, Lindberg D, Rising A, Johansson J, Nygren PA, Hedhammar M (2014). *Biomacromolecules* 15, 1696-1706.
38. Stark M, Grip S, Rising A, Hedhammar M, Engstrom W, Hjalms G, Johansson J (2007). *Biomacromolecules* 8, 1695-1701.
39. Cappello J, Crissman J, Dorman M, Mikolajczak M, Textor G, Marquet M, Ferrari F (1990). *Biotechnol Prog* 6, 198-202.
40. Krejchi MT, Atkins ED, Waddon AJ, Fournier MJ, Mason TL, Tirrell DA (1994). *Science* 265, 1427-1432.
41. Cantor EJ, Atkins ED, Cooper SJ, Fournier MJ, Mason TL, Tirrell DA (1997). *J Biochem* 122, 217-225.
42. Gelse K, Poschl E, Aigner T (2003). *Adv Drug Deliv Rev* 55, 1531-1546.
43. Prockop DJ, Kivirikko KI (1995). *Annu Rev Biochem* 64, 403-434.
44. Myllyharju J (2003). *Matrix Biol* 22, 15-24.
45. Okuyama K, Miyama K, Mizuno K, Bachinger HP (2012). *Biopolymers* 97, 607-616.
46. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH (2000). *Trends Biochem Sci* 25, 300-302.
47. Asghar A, Henrickson RL (1982). *Adv Food Res* 28, 231-372.
48. Baez J, Olsen D, Polarek JW (2005). *Appl Microbiol Biotechnol* 69, 245-252.
49. Goldberg I, Salerno AJ, Patterson T, Williams JI (1989). *Gene* 80, 305-314.
50. Obrecht G, Lefèvre, J. F. and Meyrueis, P. (1991). *Fr Pat Appl* 91 16215.
51. Gardner K, Lock RL, O'Brien JP, Salemm FR (1992). *Pat Appl PCT/US92/09655*.
52. Hori H, Hattori S, Inouye S, Kimura A, Irie S, Miyazawa H, Sakaguchi M (2002). *J Allergy Clin Immunol* 110, 652-657.
53. Van Heerde GV, van Rijn AC, Bouwstra JB, de Wolf FA, Mooibroek H, Werten MWT, Wind RD, van den Bosch TJ (1998). *US Pat* 6,150,081.
54. De Wolf FA, Werten MWT, Wisselink WH, Jansen-van den Bosch TJ, Toda Y, van Heerde GV, Bouwstra JB (2000). *Eur Pat Appl* 1 014 176.
55. Toda Y, Mori F, Bouwstra J (2002). *J Soc Photogr Sci Technol Japan* 65, 381-389.
56. Tuin A, Kluijtmans SG, Bouwstra JB, Harmsen MC, Van Luyn MJ (2010). *Tissue Eng Part A* 16, 1811-1821.
57. Tuin A, Zandstra J, Kluijtmans SG, Bouwstra JB, Harmsen MC, Van Luyn MJ (2012). *Eur Cells Mater* 24, 320-330.
58. Parvizi M, Plantinga JA, van Speuwel-Goossens CA, van Dongen EM, Kluijtmans SG, Harmsen MC (2016). *J Biomed Mater Res A* 104, 503-516.
59. Rozkiewicz DI, Kraan Y, Werten MWT, de Wolf FA, Subramaniam V, Ravoo BJ, Reinhoudt DN (2006). *Chemistry* 12, 6290-6297.
60. Bouwstra J, Toda Y (2005). *US Pat* 2005/0119170.
61. Stratton J, Chiruvolu V, Meagher M (1998). *Methods Mol Biol* 103, 107-120.
62. Thyagarajapuram N, Olsen D, Middaugh CR (2007). *J Pharm Sci* 96, 3304-3315.
63. Keizer-Gunnink I, Vuorela A, Myllyharju J, Pihlajaniemi T, Kivirikko KI, Veenhuis M (2000). *Matrix Biol* 19, 29-36.
64. Merrett K, Fagerholm P, McLaughlin CR, Dravida S, Lagali N, Shinozaki N, Watsky MA, Munger R, Kato Y, Li F, et al. (2008). *Invest Ophthalmol Vis Sci* 49, 3887-3894.
65. Fagerholm P, Lagali NS, Ong JA, Merrett K, Jackson WB, Polarek JW, Suuronen EJ, Liu Y, Brunette I, Griffith M (2014). *Biomaterials* 35, 2420-2427.
66. Gorbet MB, Sefton MV (2005). *Biomaterials* 26, 6811-6817.
67. Van Eldijk MB, McGann CL, Kiick KL, van Hest JCM (2012). *Top Curr Chem* 310, 71-116.
68. Gacko M (2000). *Cell Mol Biol Lett* 5, 327-348.

69. Foster JA, Bruenger E, Gray WR, Sandberg LB (1973). *J Biol Chem* 248, 2876-2879.
70. Gray WR, Sandberg LB, Foster JA (1973). *Nature* 246, 461-466.
71. Hinek A, Rabinovitch M (1994). *J Cell Biol* 126, 563-574.
72. Mecham RP (1991). *Ann NY Acad Sci* 624, 137-146.
73. Kozel BA, Rongjish BJ, Czirok A, Zach J, Little CD, Davis EC, Knutsen RH, Wagenseil JE, Levy MA, Mecham RP (2006). *J Cell Physiol* 207, 87-96.
74. Vrhovski B, Weiss AS (1998). *Eur J Biochem* 258, 1-18.
75. Boutris C, Chatzi EG, Kiparissides C (1997). *Polymer* 38, 2567-2570.
76. Urry DW, Long MM (1977). *Adv Exp Med Biol* 79, 685-714.
77. Urry DW, Long MM, Cox BA, Ohnishi T, Mitchell LW, Jacobs M (1974). *Biochim Biophys Acta* 371, 597-602.
78. Urry DW (1988). *J Protein Chem* 7, 1-34.
79. Urry DW (1992). *Prog Biophys Mol Biol* 57, 23-57.
80. Urry DW, Parker TM, Reid MC, Gowda DC (1991). *J Bioact Compat Pol* 6, 263-282.
81. Urry DW, Luan CH, Parker TM, Gowda DC, Prasad KU, Reid MC, Safavy A (1991). *J Am Chem Soc* 113, 4346-4348.
82. Urry DW, Gowda DC, Parker TM, Luan CH, Reid MC, Harris CM, Pattanaik A, Harris RD (1992). *Biopolymers* 32, 1243-1250.
83. Luan CH, Parker TM, Prasad KU, Urry DW (1991). *Biopolymers* 31, 465-475.
84. Meyer DE, Chilkoti A (2004). *Biomacromolecules* 5, 846-851.
85. McPherson DT, Morrow C, Minehan DS, Wu J, Hunter E, Urry DW (1992). *Biotechnol Prog* 8, 347-352.
86. Chilkoti A, Christensen T, MacKay JA (2006). *Curr Opin Chem Biol* 10, 652-657.
87. Rodriguez-Cabello JC, Prieto S, Reguera J, Arias FJ, Ribeiro A (2007). *J Biomater Sci Polym Ed* 18, 269-286.
88. Guda C, Zhang X, McPherson DT, Xu J, Cherry JH, Urry DW, Daniell H (1995). *Biotechnol Lett* 17, 745-750.
89. Trabbic-Carlson K, Setton LA, Chilkoti A (2003). *Biomacromolecules* 4, 572-580.
90. McPherson DT, Xu J, Urry DW (1996). *Protein Expr Purif* 7, 51-57.
91. Meyer DE, Chilkoti A (1999). *Nat Biotechnol* 17, 1112-1115.
92. Zhang X, Guda C, Datta R, Dute R, Urry DW, Daniell H (1995). *Biotechnol Lett* 17, 1279-1284.
93. Herzog RW, Singh NK, Urry DW, Daniell H (1997). *Appl Microbiol Biotechnol* 47, 368-372.
94. Guda C, Lee S-B, Daniell H (2000). *Plant Cell Reports* 19, 257-262.
95. Toonkool P, Weiss AS (2001). *Acta Biotechnol* 21, 189-193.
96. Meyer LC, Wright NT (2013). *Front Physiol* 4, 368.
97. Linke WA, Hamdani N (2014). *Circ Res* 114, 1052-1068.
98. Lohse DJ, Hadjichristidis N (1997). *Curr Opin Colloid In* 2, 171-176.
99. Holden G, Bishop ET, Legge NR (1969). *J Polym Sci Pol Sym* 26, 37-57.
100. Coyne KJ, Qin XX, Waite JH (1997). *Science* 277, 1830-1832.
101. Rabotyagova OS, Cebe P, Kaplan DL (2011). *Biomacromolecules* 12, 269-289.
102. Meyer DE, Chilkoti A (2002). *Biomacromolecules* 3, 357-367.
103. Wright ER, McMillan RA, Cooper A, Apkarian RP, Conticello VP (2002). *Adv Funct Mater* 12, 149-154.
104. Straley KS, Heilshorn SC (2009). *Front Neuroeng* 2, 9.
105. Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA (1998). *Science* 281, 389-392.
106. Xu C, Breedveld V, Kopecek J (2005). *Biomacromolecules* 6, 1739-1749.
107. Martens AA, Portale G, Werten MWT, de Vries RJ, Eggink G, Cohen Stuart MA, de Wolf FA (2009). *Macromolecules* 42, 1002-1009.
108. Yan Y, Martens AA, Besseling NA, de Wolf FA, de Keizer A, Drechsler M, Cohen Stuart MA (2008). *Angew Chem Int Ed Engl* 47, 4192-4195.
109. Yan Y, de Keizer A, Martens AA, Oliveira CL, Pedersen JS, de Wolf FA, Drechsler M, Cohen Stuart MA, Besseling NA (2009). *Langmuir* 25, 12899-12908.
110. Beun LH, Beaudoux XJ, Kleijn JM, de Wolf FA, Stuart MA (2012). *ACS Nano* 6, 133-140.

111. Golinska MD, Włodarczyk-Biegun MK, Werten MWT, Cohen Stuart MA, de Wolf FA, de Vries R (2014). *Biomacromolecules* 15, 699-706.
112. Beun LH, Storm IM, Werten MWT, de Wolf FA, Cohen Stuart MA, de Vries R (2014). *Biomacromolecules* 15, 3349-3357.
113. Teles H, Skrzyszewska PJ, Werten MWT, van der Gucht J, Eggink G, de Wolf FA (2010). *Soft Matter* 6, 4681-4687.
114. Silva CIF, Teles H, Moers APHA, Eggink G, de Wolf FA, Werten MWT (2011). *Biotechnol Bioeng* 108, 2517-2525.
115. Silva CIF, Skrzyszewska PJ, Golinska MD, Werten MWT, Eggink G, de Wolf FA (2012). *Biomacromolecules* 13, 1250-1258.
116. Golinska MD, Pham TTH, Werten MWT, de Wolf FA, Cohen Stuart MA, van der Gucht J (2013). *Biomacromolecules* 14, 48-55.
117. Rombouts WH, Domeradka NE, Werten MWT, Leermakers FA, de Vries RJ, de Wolf FA, van der Gucht J (2016). *Biopolymers* 105, 795-801.
118. Pham TTH, Skrzyszewska PJ, Werten MWT, Rombouts WH, Cohen Stuart MA, de Wolf FA, van der Gucht J (2013). *Soft Matter* 9, 6391-6397.
119. Pham TTH, Wang J, Werten MWT, Snijkers F, de Wolf FA, Cohen Stuart MA, van der Gucht J (2013). *Soft Matter* 9, 8923-8930.
120. Pham TTH, Snijkers F, Storm IM, de Wolf FA, Cohen Stuart MA, van der Gucht J (2016). *Int J Polym Mater Polym Biomater* 65, 125-133.
121. Rombouts WH, de Kort DW, Pham TTH, van Mierlo CPM, Werten MWT, de Wolf FA, van der Gucht J (2015). *Biomacromolecules* 16, 2506-2513.
122. Hernandez-Garcia A, Werten MWT, Cohen Stuart MA, de Wolf FA, de Vries R (2012). *Small* 8, 3491-3501.
123. Hernandez-Garcia A, Estrich NA, Werten MWT, Van Der Maarel JRC, LaBean TH, de Wolf FA, Cohen Stuart MA, de Vries R (2016). *ACS Nano* 11, 144-152.
124. Hernandez-Garcia A, Kraft DJ, Janssen AFJ, Bomans PHH, Sommerdijk NAJM, Thies-Weesie DME, Favretto ME, Brock R, de Wolf FA, Werten MWT, van der Schoot P, Cohen Stuart MA, de Vries R (2014). *Nat Nanotechnol* 9, 698-702.
125. Domeradka NE, Werten MWT, de Vries R, de Wolf FA (2016). *Biotechnol Bioeng* 113, 953-960.
126. Domeradka NE, Werten MWT, de Wolf FA, de Vries R (2016). *Biomacromolecules* 17, 3893-3901.
127. Zhao B, Cohen Stuart MA, Hall CK (2016). *Soft Matter* 12, 3721-3729.
128. Włodarczyk-Biegun MK, Werten MWT, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Kamperman M, Cohen Stuart MA (2014). *Acta Biomater* 10, 3620-3629.
129. Tseng TC, Tao L, Hsieh FY, Wei Y, Chiu IM, Hsu SH (2015). *Adv Mater* 27, 3518-3524.
130. Włodarczyk-Biegun MK, Farbod K, Werten MWT, Slingerland CJ, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Cohen Stuart MA, Kamperman M (2016). *PLoS One* 11, e0155625.
131. Cardin AD, Weintraub HJ (1989). *Arteriosclerosis* 9, 21-32.
132. Włodarczyk-Biegun MK, Slingerland CJ, Werten MWT, van Hees IA, de Wolf FA, de Vries R, Stuart MA, Kamperman M (2016). *Biomacromolecules* 17, 2063-2072.
133. Teles H, Vermonden T, Eggink G, Hennink WE, de Wolf FA (2010). *J Control Release* 147, 298-303.
134. Skrzyszewska PJ, Sprakel J, de Wolf FA, Fokkink R, Cohen Stuart MA, van der Gucht J (2010). *Macromolecules* 43, 3542-3548.
135. Rombouts WH, Colomb-Delsuc M, Werten MWT, Otto S, de Wolf FA, van der Gucht J (2013). *Soft Matter* 9, 6936-6942.
136. Pham TTH, de Wolf FA, Cohen Stuart MA, van der Gucht J (2013). *Soft Matter* 9, 8737-8744.
137. Seeman NC (2010). *Annu Rev Biochem* 79, 65-87.
138. Kabanov VA, Kabanov AV (1998). *Adv Drug Deliv Rev* 30, 49-60.
139. Kakizawa Y, Kataoka K (2002). *Adv Drug Deliv Rev* 54, 203-222.
140. Zhang C, Hernandez-Garcia A, Jiang K, Gong Z, Guttula D, Ng SY, Malar PP, van Kan JA, Dai L, Doyle PS, et al. (2013). *Nucleic Acids Res* 41, e189.
141. Storm IM, Kornreich M, Hernandez-Garcia A, Voets IK, Roy B, Cohen Stuart MA,

- Leermakers FAM, de Vries R (2015). *J Phys Chem B* 119, 4084-4092.
142. Domeradzka NE, Werten MWT, de Wolf FA, de Vries R (2016). *Curr Opin Biotechnol* 39, 61-67.
143. Moll JR, Ruvinov SB, Pastan I, Vinson C (2001). *Protein Sci* 10, 649-655.
144. Kay BK, Williamson MP, Sudol M (2000). *FASEB J* 14, 231-241.
145. Cregg JM, Cereghino JL, Shi J, Higgins DR (2000). *Mol Biotechnol* 16, 23-52.
146. Zhang W, Inan M, Meagher MM (2000). *Biotechnol Bioprocess Eng* 5, 275-287.
147. Cos O, Ramon R, Montesinos JL, Valero F (2006). *Microb Cell Fact* 5, 17.
148. Potvin G, Ahmad A, Zhang Z (2012). *Biochem Eng J* 64, 91-105.
149. Valero F (2013), in *Protein Eng*, (Ogawa DT, Ed.), InTech.
150. Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, Fotiadis D, Kovar K (2015). *Biotechnol Adv* 33, 1177-1193.
151. Paifer E, Margolles E, Cremata J, Montesino R, Herrera L, Delgado JM (1994). *Yeast* 10, 1415-1419.
152. Eiden-Plach A, Zagorc T, Heintel T, Carius Y, Breinig F, Schmitt MJ (2004). *Appl Environ Microbiol* 70, 961-966.
153. Kottmeier K, Ostermann K, Bley T, Rodel G (2011). *Appl Microbiol Biotechnol* 91, 133-141.
154. Scorer CA, Clare JJ, McCombie WR, Romanos MA, Sreekrishna K (1994). *Nat Biotechnol* 12, 181-184.
155. Damasceno LM, Huang CJ, Batt CA (2012). *Appl Microbiol Biotechnol* 93, 31-39.
156. Julius D, Blair L, Brake A, Sprague G, Thorner J (1983). *Cell* 32, 839-852.
157. Brake AJ (1990). *Methods Enzymol* 185, 408-421.
158. Emr SD, Schekman R, Flessel MC, Thorner J (1983). *Proc Natl Acad Sci USA* 80, 7080-7084.
159. Vedvick T, Buckholz RG, Engel M, Urcan M, Kinney J, Provow S, Siegel RS, Thill GP (1991). *J Ind Microbiol* 7, 197-201.
160. Raemaekers RJ, de Muro L, Gatehouse JA, Fordham-Skelton AP (1999). *Eur J Biochem* 265, 394-403.
161. Zhao HL, He Q, Xue C, Sun B, Yao XQ, Liu ZM (2009). *FEMS Yeast Res* 9, 591-599.
162. Van den Bergh CJ, Bekkers AC, De Geus P, Verheij HM, de Haas GH (1987). *Eur J Biochem* 170, 241-246.
163. Barnes D, Blair L, Brake A, Church M, Julius D, Kunisawa R, Lotko J, Stetler G, Thorner J (1982). *Recent Adv Yeast Mol Biol* 1, 295-305.
164. Schechter I, Berger A (1967). *Biochem Biophys Res Commun* 27, 157-162.
165. Wagner SL, Siegel RS, Vedvick TS, Raschke WC, Van Nostrand WE (1992). *Biochem Biophys Res Commun* 186, 1138-1145.
166. Voza LA, Wittwer L, Higgins DR, Purcell TJ, Bergseid M, Collins-Racie LA, LaVallie ER, Hoeffler JP (1996). *Nat Biotechnol* 14, 77-81.
167. Oka C, Tanaka M, Muraki M, Harata K, Suzuki K, Jigami Y (1999). *Biosci Biotechnol Biochem* 63, 1977-1983.
168. Eurwilaichitr L, Roytrakul S, Suprasongsin C, Manitchotpisit P, Panyim S (2002). *World J Microbiol Biotechnol* 18, 493-498.
169. Ghosalkar A, Sahai V, Srivastava A (2008). *Protein Expr Purif* 60, 103-109.
170. Reverter D, Ventura S, Villegas V, Vendrell J, Aviles FX (1998). *J Biol Chem* 273, 3535-3541.
171. Emberson LM, Trivett AJ, Blower PJ, Nicholls PJ (2005). *J Immunol Methods* 305, 135-151.
172. Kjeldsen T, Brandt J, Andersen AS, Egel-Mitani M, Hach M, Pettersson AF, Vad K (1996). *Gene* 170, 107-112.
173. Rockwell NC, Wang GT, Krafft GA, Fuller RS (1997). *Biochemistry* 36, 1912-1917.
174. Bourbonnais Y, Bolin D, Shields D (1988). *J Biol Chem* 263, 15342-15347.
175. Rockwell NC, Krysan DJ, Komiyama T, Fuller RS (2002). *Chem Rev* 102, 4525-4548.
176. Bader O, Krauke Y, Hube B (2008). *BMC Microbiol* 8, 116.
177. Ledgerwood EC, George PM, Peach RJ, Brennan SO (1995). *Biochem J* 308, 321-325.
178. Rholam M, Brakch N, Germain D, Thomas DY, Fahy C, Boussetta H, Boileau G, Cohen P (1995). *Eur J Biochem* 227, 707-714.

179. Guisez Y, Tison B, Vandekerckhove J, Demolder J, Bauw G, Haegeman G, Fiers W, Contreras R (1991). *Eur J Biochem* 198, 217-222.
180. Xie YF, Chen H, Huang BR (2007). *J Biotechnol* 129, 472-480.
181. Zheng X, Fang B, Han D, Yang W, Qi F, Chen H, Li S (2016). *PLoS One* 11, e0161529.
182. Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, Jiang Z, Zhou Y, Wang Y, Xu A, et al. (2013). *PLoS One* 8, e75347.
183. Degelmann A, Muller F, Sieber H, Jenzelewski V, Suckow M, Strasser AW, Gellissen G (2002). *FEMS Yeast Res* 2, 349-361.
184. Liu PT, Ta TV, Villarete LH (2001). *Protein Expr Purif* 22, 381-387.
185. Salunkhe S, Soorapaneni S, Prasad KS, Raiker VA, Padmanabhan S (2010). *Protein Expr Purif* 71, 139-146.
186. Weiss HM, Haase W, Michel H, Reilander H (1998). *Biochem J* 330, 1137-1147.
187. Grünwald S, Haase W, Molsberger E, Michel H, Reilander H (2004). *Receptors Channels* 10, 37-50.
188. Laroche Y, Storme V, De Meutter J, Messens J, Lauwereys M (1994). *Nat Biotechnol* 12, 1119-1124.
189. Khasa YP, Conrad S, Sengul M, Plautz S, Meagher MM, Inan M (2011). *Yeast* 28, 213-226.
190. Heiss S, Puxbaum V, Gruber C, Altmann F, Mattanovich D, Gasser B (2015). *Microbiology* 161, 1356-1368.
191. Rakestraw JA, Sazinsky SL, Piatasi A, Antipov E, Wittrup KD (2009). *Biotechnol Bioeng* 103, 1192-1201.
192. Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, et al. (2013). *Gene* 519, 311-317.
193. Chahal S, Wei P, Moua P, Park SP, Kwon J, Patel A, Vu AT, Catolico JA, Tsai YF, Shaheen N, et al. (2017). *Gene* 598, 50-62.
194. Fitzgerald I, Glick BS (2014). *Microb Cell Fact* 13, 125.
195. Julius D, Schekman R, Thorner J (1984). *Cell* 36, 309-318.
196. Caplan S, Green R, Rocco J, Kurjan J (1991). *J Bacteriol* 173, 627-635.
197. Kjeldsen T, Andersen AS, Hach M, Diers I, Nikolajsen J, Markussen J (1998). *Biotechnol Appl Biochem* 27, 109-115.
198. Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, Kleinheinz M, Sauer M, Altmann F, Gasser B (2009). *Microb Cell Fact* 8, 29.
199. Huang CJ, Damasceno LM, Anderson KA, Zhang S, Old LJ, Batt CA (2011). *Appl Microbiol Biotechnol* 90, 235-247.
200. Liang S, Li C, Ye Y, Lin Y (2013). *Biotechnol Lett* 35, 97-105.
201. Govindappa N, Hanumanthappa M, Venkatarangaiah K, Periyasamy S, Sreenivas S, Soni R, Sastry K (2014). *J Microbiol Biotechnol* 24, 337-345.
202. Massahi A, Calik P (2016). *J Theor Biol* 408, 22-33.
203. Küberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T, Jaenicke S, Brinkrolf K, Goesmann A, Szczepanowski R, et al. (2011). *J Biotechnol* 154, 312-320.
204. Jones EW (1991). *J Biol Chem* 266, 7963-7966.
205. Hecht KA, O'Donnell AF, Brodsky JL (2014). *Cell Logist* 4, e28023.
206. Ding WX, Yin XM (2008). *Autophagy* 4, 141-150.
207. Bussey H (1988). *Yeast* 4, 17-26.
208. Van den Hombergh JP, van de Vondervoort PJ, Fraissinet-Tachet L, Visser J (1997). *Trends Biotechnol* 15, 256-263.
209. Yoon J, Maruyama J, Kitamoto K (2011). *Appl Microbiol Biotechnol* 89, 747-759.
210. Salamin K, Sriranganadane D, Lechenne B, Jousson O, Monod M (2010). *Appl Environ Microbiol* 76, 4269-4276.
211. Gleeson MA, White CE, Meininger DP, Komives EA (1998). *Methods Mol Biol* 103, 81-94.
212. Sinha J, Plantz BA, Inan M, Meagher MM (2005). *Biotechnol Bioeng* 89, 102-112.
213. Inan M, Chiruvolu V, Eskridge KM, Vlasuk GP, Dickerson K, Brown S, Meagher MM (1999). *Enzyme Microb Technol* 24, 438-445.
214. Kobayashi K, Kuwae S, Ohya T, Ohda T, Ohyama M, Ohi H, Tomomitsu K, Ohmura T (2000). *J Biosci Bioeng* 89, 55-61.
215. Brierley RA (1998). *Methods Mol Biol* 103, 149-177.

216. Koganawasa N, Aizawa T, Shimojo H, Miura K, Ohnishi A, Demura M, Hayakawa Y, Nitta K, Kawano K (2002). *Protein Expr Purif* 25, 416-425.
217. Clare JJ, Romanos MA, Rayment FB, Rowedder JE, Smith MA, Payne MM, Sreekrishna K, Henwood CA (1991). *Gene* 105, 205-212.
218. Li A, Crimmins DL, Luo Q, Hartupee J, Landt Y, Ladenson JH, Wilson D, Anant S, Dieckgraefe BK (2003). *Protein Expr Purif* 31, 197-206.
219. Yang J, Zhou X, Zhang Y (2004). *Biotechnol Lett* 26, 1013-1017.
220. Chauhan AK, Arora D, Khanna N (1999). *Process Biochem* 34, 139-145.
221. Woo JH, Liu YY, Stavrou S, Neville DM, Jr. (2004). *Appl Environ Microbiol* 70, 3370-3376.
222. Choi DB, Park EY (2006). *Process Biochem* 41, 390-397.
223. Zhao HL, Xue C, Wang Y, Yao XQ, Liu ZM (2008). *Appl Microbiol Biotechnol* 81, 235-241.
224. Kaushik N, Rohila D, Arora U, Raut R, Lamminmaki U, Khanna N, Batra G (2016). *BMC Biotechnol* 16, 12.
225. Pritchett J, Baldwin SA (2004). *J Ind Microbiol Biotechnol* 31, 553-558.
226. Tsujikawa M, Okabayashi K, Morita M, Tanabe T (1996). *Yeast* 12, 541-553.
227. Shi X, Karkut T, Chamankhah M, Alting-Mees M, Hemmingsen SM, Hegedus D (2003). *Protein Expr Purif* 28, 321-330.
228. Vad R, Nafstad E, Dahl LA, Gabrielsen OS (2005). *J Biotechnol* 116, 251-260.
229. Jahic M, Gustavsson M, Jansen AK, Martinelle M, Enfors SO (2003). *J Biotechnol* 102, 45-53.
230. Zhou X-S, Zhang Y-X (2002). *Biotechnol Lett* 24, 1449-1453.
231. Sinha J, Plantz BA, Zhang W, Gouthro M, Schlegel V, Liu CP, Meagher MM (2003). *Biotechnol Prog* 19, 794-802.
232. Rebnegger C, Graf AB, Valli M, Steiger MG, Gasser B, Maurer M, Mattanovich D (2014). *Biotechnol J* 9, 511-525.
233. Thibault G, Ng DT (2012). *Cold Spring Harb Perspect Biol* 4, a013193.
234. Li Z, Xiong F, Lin Q, d'Anjou M, Daugulis AJ, Yang DS, Hew CL (2001). *Protein Expr Purif* 21, 438-445.
235. Hong F, Meinander NQ, Jonsson LJ (2002). *Biotechnol Bioeng* 79, 438-449.
236. Fredericks D, Clay R, Warner T, O'Connor A, de Kretser DM, Hearn MT (2010). *Biotechnol Prog* 26, 372-383.
237. Jin H, Liu G, Dai K, Wang H, Li Z, Shi Z (2011). *Appl Biochem Biotechnol* 165, 559-571.
238. Wu D, Hao YY, Chu J, Zhuang YP, Zhang SL (2008). *Appl Microbiol Biotechnol* 80, 1063-1071.
239. Siren N, Weegar J, Dahlbacka J, Kalkkinen N, Fagervik K, Leisola M, von Weymarn N (2006). *Biotechnol Appl Biochem* 44, 151-158.
240. Jahic M, Wallberg F, Bollok M, Garcia P, Enfors SO (2003). *Microb Cell Fact* 2, 6.
241. Velut S, de Maré L, Hagander P (2007). *Control Eng Pract* 15, 135-147.
242. Guarna MM, Lesnicki GJ, Tam BM, Robinson J, Radziminski CZ, Hasenwinkle D, Boraston A, Jervis E, MacGillivray RT, Turner RF, Kilburn DG (1997). *Biotechnol Bioeng* 56, 279-286.
243. Katakura Y, Zhang W, Zhuang G, Omasa T, Kishimoto M, Goto Y, Suga K-I (1998). *J Ferment Bioeng* 86, 482-487.
244. Charoenrat T, Ketudat-Cairns M, Stendahl-Andersen H, Jahic M, Enfors SO (2005). *Bioprocess Biosyst Eng* 27, 399-406.
245. Khatri NK, Hoffmann F (2006). *Biotechnol Bioeng* 93, 871-879.
246. Trentmann O, Khatri NK, Hoffmann F (2004). *Biotechnol Prog* 20, 1766-1775.
247. Khatri NK, Hoffmann F (2006). *Appl Microbiol Biotechnol* 72, 492-498.
248. Jahic M, Veide A, Charoenrat T, Teeri T, Enfors SO (2006). *Biotechnol Prog* 22, 1465-1473.
249. Wang Z, Wang Y, Zhang D, Li J, Hua Z, Du G, Chen J (2010). *Bioresour Technol* 101, 1318-1323.
250. Gao MJ, Zhan XB, Gao P, Zhang X, Dong SJ, Li Z, Shi ZP, Lin CC (2015). *Appl Biochem Biotechnol* 176, 493-504.
251. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM (1997). *Gene* 186, 37-44.
252. Xiao A, Zhou X, Zhou L, Zhang Y (2006). *Appl Microbiol Biotechnol* 72, 837-844.
253. Jahic M, Knoblechner J, Charoenrat T, Enfors SO, Veide A (2006). *Biotechnol Bioeng* 93, 1040-1049.

254. Wang H, Li J, Liu L, Li X, Jia D, Du G, Chen J, Song J (2012). *Bioresour Technol* 124, 338-346.
255. Goodrick JC, Xu M, Finnegan R, Schilling BM, Schiavi S, Hoppe H, Wan NC (2001). *Biotechnol Bioeng* 74, 492-497.
256. Ohi H, Ohtani W, Okazaki N, Furuhata N, Ohmura T (1996). *Yeast* 12, 31-40.
257. Boehm T, Pirie-Shepherd S, Trinh LB, Shiloach J, Folkman J (1999). *Yeast* 15, 563-572.
258. Prabha L, Govindappa N, Adhikary L, Melarkode R, Sastry K (2009). *Protein Expr Purif* 64, 155-161.
259. Guan B, Lei J, Su S, Chen F, Duan Z, Chen Y, Gong X, Li H, Jin J (2012). *FEMS Yeast Res* 12, 969-979.
260. Wu M, Shen Q, Yang Y, Zhang S, Qu W, Chen J, Sun H, Chen S (2013). *J Ind Microbiol Biotechnol* 40, 589-599.
261. Marsalek L, Gruber C, Altmann F, Aleschko M, Mattanovich D, Gasser B, Puxbaum V (2017). *Biotechnol J*. 12, N/A, 1600584.
262. Ahmad M, Hirz M, Pichler H, Schwab H (2014). *Appl Microbiol Biotechnol* 98, 5301-5317.
263. Van Den Hazel HB, Kielland-Brandt MC, Winther JR (1996). *Yeast* 12, 1-16.
264. Dreyer T (1989). *Carlsberg Res Commun* 54, 85-97.
265. Kominami E, Hoffschulte H, Leuschel L, Maier K, Holzer H (1981). *Biochim Biophys Acta* 661, 136-141.
266. Remington SJ, Breddam K (1994). *Methods Enzymol* 244, 231-248.
267. Cereghino JL, Cregg JM (2000). *FEMS Microbiol Rev* 24, 45-66.
268. Dmochowska A, Dignard D, Henning D, Thomas DY, Bussey H (1987). *Cell* 50, 573-584.
269. Bryant NJ, Boyd A (1993). *J Cell Sci* 106, 815-822.
270. Cooper A, Bussey H (1989). *Mol Cell Biol* 9, 2706-2714.
271. Heim J, Takabayashi K, Meyhack B, Marki W, Pohlig G (1994). *Eur J Biochem* 226, 341-353.
272. Ni Z, Zhou X, Sun X, Wang Y, Zhang Y (2008). *Yeast* 25, 1-8.
273. Sreenivas S, Krishnaiah SM, Shyam Mohan AH, Mallikarjun N, Govindappa N, Chatterjee A, Sastry KN (2016). *Protein Expr Purif* 118, 1-9.
274. Ruitenbergh KM, Gilkerson JR, Wellington JE, Love DN, Whalley JM (2001). *Virus Res* 79, 125-135.
275. Shapiro RI, Wen D, Levesque M, Hronowski X, Gill A, Garber EA, Galdes A, Strauch KL, Taylor FR (2003). *Protein Expr Purif* 29, 272-283.
276. Chen Z, Chen H, Wang X, Ma X, Huang B (2007). *Protein Expr Purif* 52, 239-248.
277. Pyati P, Fitches E, Gatehouse JA (2014). *J Ind Microbiol Biotechnol* 41, 1237-1247.
278. Spiegel H, Schinkel H, Kastilan R, Dahm P, Boes A, Scheuermayer M, Chudobova I, Maskus D, Fendel R, Schillberg S, et al. (2015). *Biotechnol Bioeng* 112, 659-667.
279. Zhu YS, Zhang XY, Cartwright CP, Tipper DJ (1992). *Mol Microbiol* 6, 511-520.
280. Mizuno K, Nakamura T, Ohshima T, Tanaka S, Matsuo H (1989). *Biochem Biophys Res Commun* 159, 305-311.
281. Brenner C, Fuller RS (1992). *Proc Natl Acad Sci USA* 89, 922-926.
282. Bevan A, Brenner C, Fuller RS (1998). *Proc Natl Acad Sci USA* 95, 10384-10389.
283. Riffer F, Eisfeld K, Breinig F, Schmitt MJ (2002). *Microbiology* 148, 1317-1328.
284. Suzuki Y, Ikeda N, Kataoka E, Ohsuye K (2000). *Biotechnol Appl Biochem* 32, 53-60.
285. Rockwell NC, Fuller RS (1998). *Biochemistry* 37, 3386-3391.
286. Duckert P, Brunak S, Blom N (2004). *Protein Eng* 17, 107-112.
287. Gagnon-Arsenault I, Tremblay J, Bourbonnais Y (2006). *FEMS Yeast Res* 6, 966-978.
288. Krysan DJ, Ting EL, Abeijon C, Kroos L, Fuller RS (2005). *Eukaryot Cell* 4, 1364-1374.
289. Gagnon-Arsenault I, Parise L, Tremblay J, Bourbonnais Y (2008). *Mol Microbiol* 69, 982-993.
290. Ash J, Dominguez M, Bergeron JJ, Thomas DY, Bourbonnais Y (1995). *J Biol Chem* 270, 20847-20854.
291. Cawley NX, Olsen V, Zhang CF, Chen HC, Tan M, Loh YP (1998). *J Biol Chem* 273, 584-591.

292. Kang HA, Kim SJ, Choi ES, Rhee SK, Chung BH (1998). *Appl Microbiol Biotechnol* 50, 187-192.
293. Egel-Mitani M, Flygenring HP, Hansen MT (1990). *Yeast* 6, 127-137.
294. Komano H, Fuller RS (1995). *Proc Natl Acad Sci USA* 92, 10752-10756.
295. Olsen V, Cawley NX, Brandt J, Egel-Mitani M, Loh YP (1999). *Biochem J* 339, 407-411.
296. Bourbonnais Y, Germain D, Ash J, Thomas DY (1994). *Biochimie* 76, 226-233.
297. Komano H, Rockwell N, Wang GT, Krafft GA, Fuller RS (1999). *J Biol Chem* 274, 24431-24437.
298. Yao XQ, Zhao HL, Xue C, Zhang W, Xiong XH, Wang ZW, Li XY, Liu ZM (2009). *J Biotechnol* 139, 131-136.
299. Sazonova EA, Zobnina AE, Padkina MV (2013). *Russ J Genet* 49, 602-608.
300. Egel-Mitani M, Hansen MT (1987). *Nucleic Acids Res* 15, 6303.
301. Matoba S, Fukayama J, Wing RA, Ogrzydzak DM (1988). *Mol Cell Biol* 8, 4904-4916.
302. Kreil G (1990). *Trends Biochem Sci* 15, 23-26.
303. Zhang Q, Ding F, Xue X, Xu X, Pan W (2008). *Appl Microbiol Biotechnol* 78, 139-145.
304. Hopkins D, Gomathinayagam S, Lynaugh H, Stadheim TA, Hamilton SR (2014). *Appl Microbiol Biotechnol* 98, 2573-2583.
305. Fernandez E, Toledo JR, Mansur M, Sanchez O, Gil DF, Gonzalez-Gonzalez Y, Lamazares E, Fernandez Y, Parra F, Farnos O (2015). *Appl Microbiol Biotechnol* 99, 3875-3886.
306. Gil DF, Garcia-Fernandez R, Alonso-del-Rivero M, Lamazares E, Perez M, Varas L, Diaz J, Chavez MA, Gonzalez-Gonzalez Y, Mansur M (2011). *FEMS Yeast Res* 11, 575-586.
307. Li H, d'Anjou M (2009). *Curr Opin Biotechnol* 20, 678-684.
308. Narhi LO, Arakawa T, Aoki KH, Elmore R, Rohde MF, Boone T, Strickland TW (1991). *J Biol Chem* 266, 23022-23026.
309. Meldgaard M, Svendsen I (1994). *Microbiology* 140, 159-166.
310. Eklund EA, Freeze HH (2005). *Semin Pediatr Neurol* 12, 134-143.
311. Grinna LS, Tschopp JF (1989). *Yeast* 5, 107-115.
312. Sethuraman N, Stadheim TA (2006). *Curr Opin Biotechnol* 17, 341-346.
313. Hamilton SR, Zha D (2015). *Methods Mol Biol* 1321, 73-90.
314. Laukens B, De Wachter C, Callewaert N (2015). *Methods Mol Biol* 1321, 103-122.
315. Tanner W, Lehle L (1987). *Biochim Biophys Acta* 906, 81-99.
316. Chi YH, Koo YD, Dai SY, Ahn JE, Yun DJ, Lee SY, Zhu-Salzman K (2010). *Comp Biochem Physiol B Biochem Mol Biol* 156, 40-47.
317. Bretthauer RK, Castellino FJ (1999). *Biotechnol Appl Biochem* 30, 193-200.
318. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, et al. (2013). *EMBO J* 32, 1478-1488.
319. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S (1998). *Glycoconj J* 15, 115-130.
320. Gonzalez M, Brito N, Gonzalez C (2012). *BMC Microbiol* 12, 213.
321. Gomathinayagam S, Hamilton SR (2014). *Appl Microbiol Biotechnol* 98, 2545-2554.
322. Herscovics A, Orlean P (1993). *FASEB J* 7, 540-550.
323. Lehle L, Bause E (1984). *Biochim Biophys Acta* 799, 246-251.
324. Strahl-Bolsinger S, Tanner W (1991). *Eur J Biochem* 196, 185-190.
325. Mochizuki S, Hamato N, Hirose M, Miyano K, Ohtani W, Kameyama S, Kuwae S, Tokuyama T, Ohi H (2001). *Protein Expr Purif* 23, 55-65.
326. Hirose M, Kameyama S, Ohi H (2002). *Yeast* 19, 1191-1202.
327. Cohen P (2002). *Nat Cell Biol* 4, E127-E130.
328. Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, Smith D, Gerstein M, Reed MA, Snyder M (2000). *Nat Genet* 26, 283-289.
329. Shen W, Kong C, Xue Y, Liu Y, Cai M, Zhang Y, Jiang T, Zhou X, Zhou M (2016). *PLoS One* 11, e0167766.
330. Ingrell CR, Miller ML, Jensen ON, Blom N (2007). *Bioinformatics* 23, 895-897.

331. Trost B, Kusalik A (2011). *Bioinformatics* 27, 2927-2935.
332. Frades I, Resjo S, Andreasson E (2015). *BMC Bioinformatics* 16, 239.
333. Choi BK, Jimenez-Flores R (2001). *J Agric Food Chem* 49, 1761-1766.
334. Xu X, Kashima O, Saito A, Azakami H, Kato A (2004). *Biosci Biotechnol Biochem* 68, 1273-1278.
335. Barral P, Batanero E, Villalba M, Rodriguez R (2005). *Protein Expr Purif* 44, 147-154.
336. Brimau F, Cornard JP, Le Danvic C, Lagant P, Vergoten G, Grebert D, Pajot E, Nagnan-Le Meillour P (2010). *J Chem Ecol* 36, 801-813.
337. Freivalds J, Dislers A, Ose V, Pumpens P, Tars K, Kazaks A (2011). *Protein Expr Purif* 75, 218-224.
338. Ubersax JA, Ferrell JE, Jr. (2007). *Nat Rev Mol Cell Biol* 8, 530-541.
339. Idiris A, Tohda H, Kumagai H, Takegawa K (2010). *Appl Microbiol Biotechnol* 86, 403-417.
340. Rueda F, Gasser B, Sanchez-Chardi A, Roldan M, Villegas S, Puxbaum V, Ferrer-Miralles N, Unzueta U, Vazquez E, Garcia-Fruitos E, et al. (2016). *Microb Cell Fact* 15, 166.
341. Lünsdorf H, Gurramkonda C, Adnan A, Khanna N, Rinas U (2011). *Microb Cell Fact* 10, 48.
342. Orij R, Brul S, Smits GJ (2011). *Biochim Biophys Acta* 1810, 933-944.
343. Toman PD, Chisholm G, McMullin H, Giere LM, Olsen DR, Kovach RJ, Leigh SD, Fong BE, Chang R, Daniels GA, et al. (2000). *J Biol Chem* 275, 23303-23309.
344. Miller KA, DiDone L, Krysan DJ (2010). *Eukaryot Cell* 9, 1669-1679.
345. Li NK, Garcia Quiroz F, Hall CK, Chilkoti A, Yingling YG (2014). *Biomacromolecules* 15, 3522-3530.
346. Wickner S, Maurizi MR, Gottesman S (1999). *Science* 286, 1888-1893.
347. Bernales S, McDonald KL, Walter P (2006). *PLoS Biol* 4, e423.
348. Kumita JR, Johnson RJ, Alcocer MJ, Dumoulin M, Holmqvist F, McCammon MG, Robinson CV, Archer DB, Dobson CM (2006). *FEBS J* 273, 711-720.
349. Whyteside G, Alcocer MJ, Kumita JR, Dobson CM, Lazarou M, Pleass RJ, Archer DB (2011). *PLoS One* 6, e22692.
350. Vanz AL, Lunsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N, Rinas U (2012). *Microb Cell Fact* 11, 103.
351. Ast T, Schuldiner M (2011). *Curr Biol* 21, R692-695.
352. Ast T, Michaelis S, Schuldiner M (2016). *Cell* 164, 103-114.
353. Aguilar-Uscanga B, Francois JM (2003). *Lett Appl Microbiol* 37, 268-274.
354. Kapteyn JC, ter Riet B, Vink E, Blad S, De Nobel H, Van Den Ende H, Klis FM (2001). *Mol Microbiol* 39, 469-479.
355. Zlotnik H, Fernandez MP, Bowers B, Cabib E (1984). *J Bacteriol* 159, 1018-1026.
356. De Nobel JG, Klis FM, Munnik T, Priem J, van den Ende H (1990). *Yeast* 6, 483-490.
357. De Nobel JG, Klis FM, Priem J, Munnik T, van den Ende H (1990). *Yeast* 6, 491-499.
358. Moers APHA, Wolbert EJJ, de Wolf FA, Werten MWT (2010). *J Biotechnol* 146, 66-73.
359. Aggeli A, Nyrkova IA, Bell M, Harding R, Carrick L, McLeish TC, Semenov AN, Boden N (2001). *Proc Natl Acad Sci USA* 98, 11857-11862.
360. Puxbaum V, Mattanovich D, Gasser B (2015). *Appl Microbiol Biotechnol* 99, 2925-2938.
361. Link AJ, Mock ML, Tirrell DA (2003). *Curr Opin Biotechnol* 14, 603-609.
362. Young TS, Ahmad I, Brock A, Schultz PG (2009). *Biochemistry* 48, 2643-2653.

Summary

From a chemistry perspective, proteins can be thought of as polymers of amino acids, linked by amide bonds. They feature unsurpassed control over monomer sequence and molecular size. The amino acid sequence of proteins determines their three-dimensional folded structure, resulting in unique properties. Proteins such as collagen, elastin, and silk play a crucial structure-forming role in various tissues and animal architecture such as spider webs. These proteins are characterized by highly repetitive amino acid sequences, and can reversibly self-assemble into supramolecular structures through the formation of noncovalent bonds. The unique properties of these natural materials have sparked the interest of material scientists. Mimics of these natural proteins have been designed and produced as heterologous proteins in suitable expression systems.

The most commonly employed host for these so-called protein-based polymers, or protein polymers for short, is the bacterium *Escherichia coli*. In this thesis, we explored the use of an alternative platform, namely the methylotrophic yeast *Pichia pastoris* (*Komagataella phaffii*). This organism is well-known for its often relatively high yields, and offers a choice between intracellular or secretory production. Secretory production precludes the need to disrupt the cells, and provides a highly effective first purification step. We here evaluated the secretory production of a variety of protein polymers using this yeast system, including block copolymers featuring different functional modules.

In **Chapter 2**, we produced murine collagen fragments, or gelatins, without prolyl-4-hydroxylation. Despite the fact that gelatin is highly susceptible to proteolysis owing to its unfolded structure, degradation was minimized to a large degree by performing fermentations at pH 3 and by supplementation of the medium with casamino acids. Endoproteolytic cleavage by a putative Kex2-like protease was overcome by removal of the susceptible motifs from the protein sequence by site-directed mutagenesis. Using multicopy transformants, protein yields up to ~15 g/L of cell-free broth were obtained.

In **Chapter 3**, we developed *P. pastoris* strains deficient in the basic amino acid-specific endoproteases Kex2 and yapsin 1. Use of the *kex2* strain for the production of murine gelatins revealed that the endoproteolytic cleavage observed in Chapter 2 was indeed caused by Kex2. Yapsin 1 did not appear to be involved. Interestingly, although Kex2 is generally considered responsible for processing of the α -factor prepro leader used to drive secretion of heterologous proteins, yapsin

1 was found capable of completely taking over this task in the *kex2* mutant. Over the years, both protease-deficient strains have proven invaluable for the production of various otherwise degraded proteins.

In **Chapter 4**, we designed an artificial highly hydrophilic gelatin. The polymer was produced as a fully monodisperse secreted protein, at yields in the range of 3 to 6 g/L of clarified broth. Essentially pure product could be obtained from the cell-free broth by mere differential salt precipitation. The synthetic gelatin migrated very slowly in SDS-PAGE, which was demonstrably due to its high hydrophilicity. This high polarity also resulted in negligible surface activity in water. Circular dichroism spectrometry showed that the protein did not form collagen triple helices at 4 °C, and was in random-coil conformation.

In **Chapter 5**, a silk-like protein was produced consisting of (Gly-Ala)₃-Gly-Glu repeats, and a variant where every other octapeptide contains the hydrophobic amino acid Leu instead of Glu. Both protein polymers were secreted at g/L levels and fully monodisperse. Crystallization of proteins in 70% (v/v) formic acid under methanol vapor diffusion resulted in the formation of β -sheets and tortuous fibril-like structures. When used as a coating, the amphiphilic variant with both Glu- and Leu-containing octapeptides was able to reverse the polarity of hydrophobic and hydrophilic solid surfaces.

In **Chapter 6**, we developed hydrogel-forming triblock copolymers that consist of small collagen-inspired end blocks capable of thermoreversible trimerization, and a random coil middle block identical to, or derived from, the hydrophilic gelatin described in Chapter 4. This design inherently offers independent tuning of the crosslink density and melting temperature of the gel *via* the length of the middle and end blocks, respectively. As an example, ~42 kDa polymers were produced as secreted and fully intact proteins at several g/L. The resulting hydrogels had a ~37 °C melting temperature, and their dynamic elasticity, unlike that of animal gelatins, was independent of the thermal history. The tunability, slow release characteristics, and expected biocompatibility of these collagen-inspired hydrogels offer promise for further development towards biomedical applications.

In **Chapter 7**, we expand beyond homotypic interactions, such as occur between silk-like molecules (Chapter 5) and collagen-like molecules (Chapter 6), by introducing heterodimer-forming modules. An example of such modules are the

so-called WW-domains and their proline-rich ligands. Two C-terminally extended variants of the otherwise noninteracting random coil gelatin of Chapter 4 were prepared: one featuring a WW domain, and one with a proline-rich PPxY domain. Both secreted proteins were intact and produced at more than 2 g/L of clarified broth. The PPxY module was O-glycosylated, and a considerable fraction of the glycans contained phosphorylated mannose in diester linkage. Glycosylation was abolished by introduction of a Ser → Ala mutation in the PPxY module. The polymer with the WW domain was shown to heterodimerize with the polymers containing the PPxY domain, with similar K_d values in the low μM range for both glycosylation states of the latter. In further research, these highly specific coupling modules may be used to build complex supramolecular structures.

In **Chapter 8**, we report the production of functionalized variants of a silk-inspired triblock copolymer developed by our group. This triblock forms fibrils and hydrogels at physiological pH, and consists of random coil end blocks derived from the hydrophilic gelatin described in Chapter 4, and a silk-like middle block derived from the sequences described in Chapter 5. Functionalization involved the genetic incorporation of general integrin-binding motifs (RGD), as well as proteoglycan-binding osteoblast-specific cell-adhesive motifs (KRSR). The polymers were produced at several g/L of cell-free broth and were essentially intact owing to rational sequence design. By mixing functionalized and nonfunctionalized proteins in different ratios, scaffolds with varying content of the two types of cell-adhesive motifs could be prepared, independent of the total protein concentration, and thereby independent of nominal pore size and storage modulus. Different scaffolds resulted in a different response by the osteoblastic cells, demonstrating tailorability *via* the ratio of the functional groups. This offers promise for the further development of our silk-inspired polymer framework towards bone regeneration applications.

In **Chapter 9**, we provide context for Chapters 2 to 8 by presenting an overview of protein polymers produced in *P. pastoris* by our group and others, against the backdrop of pivotal studies involving *E. coli*. We draw conclusions from the results of this thesis and all other studies discussed in relation to challenges that may be encountered when producing protein polymers in *P. pastoris*, and suggest possible solutions. Opportunities for further development of this yeast as a host for the production of protein polymers are discussed.

Dankwoord

“Waarom sla je geen nietje door je artikelen?”, dat werd mij als publicerende laboratoriummedewerker de afgelopen twee decennia met enige regelmaat gevraagd. Tja, labwerk is het leukste dat er is, dus waarom promoveren? Maar goed, eind 2014 begon het dan toch te kriebelen en voelde het goed deze stap te nemen, hetgeen heeft geresulteerd in dit proefschrift. Op dat ‘nietje’ kom ik nog terug...

Martien, je hebt me in het kader van je ERC Advanced Grant ‘Biomate’ voorgesteld om te promoveren. Misschien wat onhandig dat ik dit pas zag zitten tegen het eind van dat project, maar ik vind het bijzonder eervol dat ik je laatste Wageningsse ‘AIO’ mag zijn. En dan nog een biotechie ook...

Gerrit, je hebt me lang geleden als toenmalig afdelingshoofd gevraagd of ik niet zou willen promoveren, en ook later als hoogleraar weer, in het kader van het B-Basic programma Performance Materials. Ik ben je zeer erkentelijk voor die blijk van waardering, en vind het fijn dat ook jij mijn promotor wilt zijn nu ik er dan eindelijk aan toe ben.

Frits, projectleider van de eiwitpolymeer projecten bij WFBR, en mijn copromotor, je hebt altijd mijn publicatiedrift ondersteund en mij herhaaldelijk de ‘P’ vraag gesteld. Als biofysicus en als biotechnoloog vullen we elkaar binnen het eiwitpolymeren thema uitstekend aan. Bij het schrijven van artikelen wisten we elkaar altijd op te porren totdat die ene alles verpletterende volzin er uiteindelijk uitrolde. Ik denk dat we trots mogen zijn op onze bijdrage als WFBR aan het eiwitpolymeren veld, voor een groot deel in samenwerking met PCC. Het potentieel is groot, dus hopelijk lukt het de komende tijd ook weer om bilaterale projecten met de industrie op te zetten.

I would like to thank the ‘protein polymer’ PhD students I have worked with over the years: Eric, Aernout, Rosalie, Helena, Roelof, Catarina, Paulina, Monika, Lennart, Armando, Thao, Wolf, Gosia, Inge, and Natalia. It was fun showing you a thing or two about protein engineering, and I have learnt from all of you in return. Your combined research efforts resulted in the wealth of information that allowed me to provide a nice broad perspective in the Discussion of this thesis and in a submitted review. Gosia and Natalia, thanks for the joint experimental work and cowriting of two papers presented in this thesis. You both have such cheerful and hard-working attitudes. Also thanks to Loiane and Mariam, our most recent PhD students, and to my trainees: Chantal, Patrick, Jeroen, Valentina, Claudia, and Bas.

Veel dank voor alles rondom de gezamenlijke AIO's aan de mensen van Physical Chemistry and Soft Matter: Renko, Marleen, Jasper, en Frans.

Antoine en Emil, jullie zijn mijn wapenbroeders op het laboratorium. Emil als mede-'molbi' en altijd enthousiaste positieveling. En Antoine (Toine, Twan ??) als oerdegelijke mede-fermenteerder en -eiwitzuiveraar in de 'goeie ouwe tijd' van blubber, vliesjes en mierenzuur. Los van werk natuurlijk ook bedankt voor de gezelligheid, en fijn dat jullie mijn paranimfen willen zijn!

Het onderzoek beschreven in dit proefschrift beslaat zoals gezegd meer dan twintig jaar, en ik heb heel wat mensen zien komen en gaan. Het is ondoenlijk hier iedereen te noemen die een rol heeft gespeeld bij het beschreven onderzoek, of op enig moment deel heeft uitgemaakt van de afdeling. Maar natuurlijk allemaal bedankt! In elk geval wil ik alle huidige nog niet genoemde medewerkers van Bioconversie bedanken voor hun collegialiteit en samenwerking: Aart, Ana, Anne, Astrid, Bwee, Carla, Chris, Hans, Heleen, Hetty, Jan S., Jan W., Janine, Jeroen, Lolke, Mamou, Marjo, Miriam, Pieternel, Ricardo, Ruud, Truus, en ook de overige collega's van BBP waar ik mee heb samengewerkt. Hoewel meerdere van jullie mij in de loop der jaren hebben 'lastiggevalen' met de suggestie dat ik zou moeten promoveren, wil ik Truus en Pieternel met name bedanken voor hun overtuigingskracht. Hans en Jan S., ik heb van jullie, de senior molbi's van Bioconversie, veel opgepikt m.b.t. gisten, schimmels en genetische trucjes. Tussen de regels door is daar vast het een en ander van terug te vinden in dit proefschrift.

Familie, schoonfamilie en vrienden bedankt voor jullie interesse en support, ook al zal misschien nog steeds niet duidelijk zijn wat er nou zo leuk is aan die 'microbeestjes'... Tot slot, lieve Cathelijne, zonder jou was ik hier nooit aan begonnen. Promoveren is een investering in de toekomst, maar het slaan van dat spreekwoordelijke nietje, naast mijn 'gewone' werk bij WFBR, was een zware klus waar we veel voor hebben moeten laten. Bedankt voor alle steun, zeker ook bij de zorg voor mijn ouders die er in deze periode onverwachts bij kwam, en voor het doen van de rotklusjes als ik weer avonden en weekenden achter elkaar zat te tikken. Ik ben heel blij dat we nu weer meer tijd voor elkaar hebben!

Marc

About the author

Marc Willem Theodoor Werten was born in Geffen (NB), The Netherlands, on May 13, 1972. After graduating at VWO-B 'Maaslandcollege' in Oss in 1990, he started the study Biotechnology at the University of Applied Sciences 'Hogeschool van Arnhem en Nijmegen'. He did his nine-month internship at the department of Animal Physiology at the University of Nijmegen, working on the molecular elucidation of the secretory pathway in peptide hormone-producing cells, under the supervision of Joost C. M. Holthuis and Prof. Dr. Gerard J. M. Martens. After his graduation in 1994, he served a year of civil duty at the Agrotechnological Research Institute (ATO-DLO) in Wageningen, as an alternative to military service. In this period, under the supervision of Dr. Miranda D. van de Rhee and Dr. Hans Mooibroek, he investigated mushroom browning by means of genetic transformation. He has worked at this institute (nowadays called Wageningen Food & Biobased Research and part of Wageningen University & Research) ever since, consecutively as a research assistant, senior research assistant, and scientist. He has mostly been involved with the genetic design, biotechnological production in *Pichia pastoris*, and basic characterization of protein-based polymers, which is the topic of this thesis. He has additionally produced enzymes and other proteins, both in *P. pastoris* and *E. coli*, and has been involved in the metabolic engineering and biotechnology of a variety of bacteria and fungi, as well as in general protein biochemistry. He has been responsible for the daily functioning of the molecular biology laboratory throughout most of his employment, and has regularly served as a reviewer for scholarly journals. After having guided many PhD students in their biotechnological efforts, he decided to pursue his own PhD degree at Wageningen University & Research as of January 2015, under the supervision of Prof. Dr. Martien A. Cohen Stuart, Prof. Dr. Gerrit Eggink, and Dr. Frits A. de Wolf. The results of this research are presented in this thesis.

Publications

This thesis

1. **Werten MWT**, Cohen Stuart MA, Eggink G, de Wolf FA, *submitted*.
2. Włodarczyk-Biegun MK*, **Werten MWT***, Posadowska U, Storm IM, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Cohen Stuart MA, Kamperman M (2016). *J Biomed Mater Res A* 104, 3082-3092 (* equal contribution).
3. Domeradзка NE*, **Werten MWT***, de Vries R, de Wolf FA (2016). *Microb Cell Fact* 15, 105 (* equal contribution).
4. **Werten MWT**, Teles H, Moers APHA, Wolbert EJH, Sprakel J, Eggink G, de Wolf FA (2009). *Biomacromolecules* 10, 1106-1113.
5. **Werten MWT**, Moers APHA, Vong T, Zuilhof H, van Hest JCM, de Wolf FA (2008). *Biomacromolecules* 9, 1705-1711.
6. **Werten MWT**, de Wolf FA (2005). *Appl Environ Microbiol* 71, 2310-2317.
7. **Werten MWT**, Wisselink WH, Jansen-van den Bosch TJ, de Bruin EC, de Wolf FA (2001). *Protein Eng* 14, 447-454.
8. **Werten MWT**, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA (1999). *Yeast* 15, 1087-1096.

Related work

9. Włodarczyk-Biegun MK, Slingerland CJ, **Werten MWT**, van Hees IA, de Wolf FA, de Vries R, Stuart MA, Kamperman M (2016). *Biomacromolecules* 17, 2063-2072.
10. Włodarczyk-Biegun MK, Farbod K, **Werten MWT**, Slingerland CJ, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Cohen Stuart MA, Kamperman M (2016). *PLoS One* 11, e0155625.
11. Rombouts WH, Domeradзка NE, **Werten MWT**, Leermakers FA, de Vries RJ, de Wolf FA, van der Gucht J (2016). *Biopolymers* 105, 795-801.
12. Hernandez-Garcia A, Estrich NA, **Werten MWT**, Van Der Maarel JRC, LaBean TH, de Wolf FA, Cohen Stuart MA, de Vries R (2016). *ACS Nano* 11, 144-152.
13. Domeradзка NE, **Werten MWT**, de Wolf FA, de Vries R (2016). *Biomacromolecules* 17, 3893-3901.

14. Domeradзка NE, **Werten MWT**, de Wolf FA, de Vries R (2016). *Curr Opin Biotechnol* 39, 61-67.
15. Domeradзка NE, **Werten MWT**, de Vries R, de Wolf FA (2016). *Biotechnol Bioeng* 113, 953-960.
16. Rombouts WH, de Kort DW, Pham TTH, van Mierlo CPM, **Werten MWT**, de Wolf FA, van der Gucht J (2015). *Biomacromolecules* 16, 2506-2513.
17. Włodarczyk-Biegun MK, **Werten MWT**, de Wolf FA, van den Beucken JJJP, Leeuwenburgh SCG, Kamperman M, Cohen Stuart MA (2014). *Acta Biomater* 10, 3620-3629.
18. Hernandez-Garcia A, Kraft DJ, Janssen AFJ, Bomans PHH, Sommerdijk NAJM, Thies-Weesie DME, Favretto ME, Brock R, de Wolf FA, **Werten MWT**, van der Schoot P, Cohen Stuart MA, de Vries R (2014). *Nat Nanotechnol* 9, 698-702.
19. Golinska MD, Włodarczyk-Biegun MK, **Werten MWT**, Cohen Stuart MA, de Wolf FA, de Vries R (2014). *Biomacromolecules* 15, 699-706.
20. Beun LH, Storm IM, **Werten MWT**, de Wolf FA, Cohen Stuart MA, de Vries R (2014). *Biomacromolecules* 15, 3349-3357.
21. Rombouts WH, Colomb-Delsuc M, **Werten MWT**, Otto S, de Wolf FA, van der Gucht J (2013). *Soft Matter* 9, 6936-6942.
22. Pham TTH, Wang J, **Werten MWT**, Snijkers F, de Wolf FA, Cohen Stuart MA, van der Gucht J (2013). *Soft Matter* 9, 8923-8930.
23. Pham TTH, Skrzyszewska PJ, **Werten MWT**, Rombouts WH, Cohen Stuart MA, de Wolf FA, van der Gucht J (2013). *Soft Matter* 9, 6391-6397.
24. Golinska MD, Pham TTH, **Werten MWT**, de Wolf FA, Cohen Stuart MA, van der Gucht J (2013). *Biomacromolecules* 14, 48-55.
25. Silva CIF, Skrzyszewska PJ, Golinska MD, **Werten MWT**, Eggink G, de Wolf FA (2012). *Biomacromolecules* 13, 1250-1258.
26. Hernandez-Garcia A, **Werten MWT**, Cohen Stuart MA, de Wolf FA, de Vries R (2012). *Small* 8, 3491-3501.
27. Silva CIF, Teles H, Moers APHA, Eggink G, de Wolf FA, **Werten MWT** (2011). *Biotechnol Bioeng* 108, 2517-2525.
28. Teles H, Skrzyszewska PJ, **Werten MWT**, van der Gucht J, Eggink G, de Wolf FA (2010). *Soft Matter* 6, 4681-4687.
29. Moers APHA, Wolbert EJH, de Wolf FA, **Werten MWT** (2010). *J Biotechnol* 146, 66-73.

30. Skrzyszewska PJ, de Wolf FA, **Werten MWT**, Moers APHA, Cohen Stuart MA, van der Gucht J (2009). *Soft Matter* 5, 2057-2062.
31. Schipperus R, Teeuwen RLM, **Werten MWT**, Eggink G, de Wolf FA (2009). *Appl Microbiol Biotechnol* 85, 293-301.
32. Martens AA, Portale G, **Werten MWT**, de Vries RJ, Eggink G, Cohen Stuart MA, de Wolf FA (2009). *Macromolecules* 42, 1002-1009.
33. Rozkiewicz DI, Kraan Y, **Werten MWT**, de Wolf FA, Subramaniam V, Ravoo BJ, Reinhoudt DN (2006). *Chemistry* 12, 6290-6297.
34. De Bruin EC, **Werten MWT**, Laane C, de Wolf FA (2002). *FEMS Yeast Res* 1, 291-298.

Other work

35. Weusthuis RA, Mars AE, Springer J, Wolbert EJH, van der Wal H, de Vrije TG, Levisson M, Leprince A, Houweling-Tan GB, Moers APHA, Hendriks SNA, Mendes O, Griekspoor Y, **Werten MWT**, Schaap PJ, van der Oost J, Eggink G (2017). *Metab Eng* 42, 66-73.
36. Van de Rhee MD, Mendes O, **Werten MWT**, Huizing HJ, Mooibroek H (1996). *Curr Genet* 30, 166-173.

The research described in this thesis was financially supported as indicated in the acknowledgments of each respective chapter. The writing of this thesis was in part financially supported by ERC Advanced Grant 267254 'BioMate'.

Financial support from Wageningen University & Research for printing this thesis is gratefully acknowledged.

Cover design by Marc W. T. Werten and Jeroen Advocaat.

Cover realization by Jeroen Advocaat.

Printed on FSC-certified paper by GVO drukkers & vormgevers, Ede.