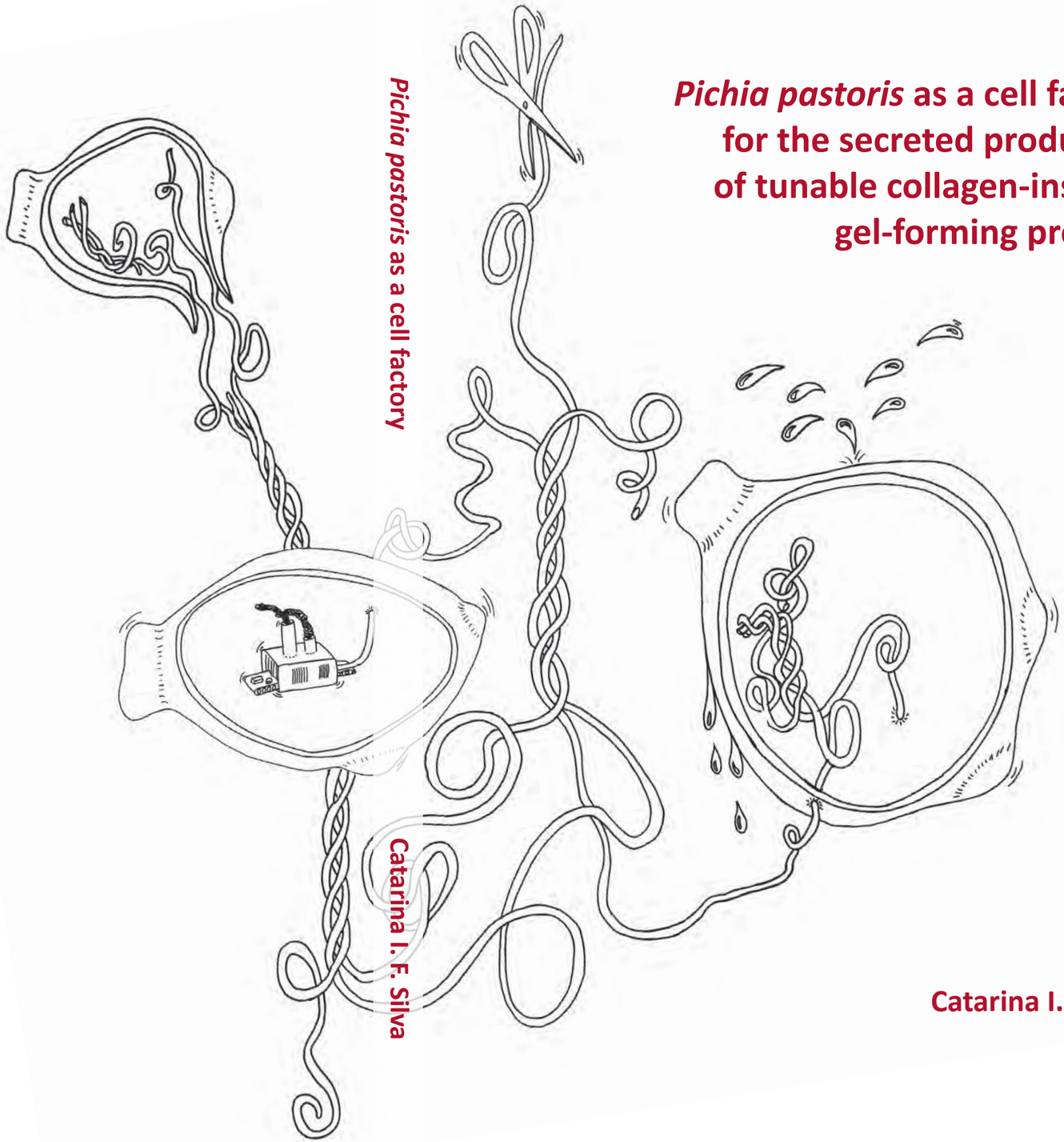
*Pichia pastoris* as a cell factory

Catarina I. F. Silva



# *Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins

Catarina I. F. Silva

## INVITATION

You are cordially invited for the public defense of my PhD thesis entitled:

***Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins**

On Friday, 11th of January 2013 at 13:30 h in the Aula of Wageningen University, Generaal Foulkesweg 1a, Wageningen.

Following the defence, there will be a reception at the Aula

**Catarina I.F. Silva**
catarinaifsilva@gmail.com

**Paranymphs**

Helena Teles
helenateles@hotmail.com

Aldana Ramirez
ramirezdana@yahoo.com

# *Propositions:*

1. Triple helix strength is *Pichia pastoris'* weakness. (this thesis, Chapter 2)

2. The simplest way for collagen-like proteins to travel along *P. pastoris'* secretory pathway is by avoiding interactions; the secretory pathway must be travelled alone. (this thesis, Chapters 2 and 3)

3. We are 99% microbial and 1% human; the fact that our immune system is able to distinguish friend from foe is one amazing evolutionary accomplishment.

4. Organic food cannot be considered healthier solely because it is grown in a natural way.

5. Materialism was never frowned upon until emerging countries started to be able to afford it.

6. 'Seeing is believing', ultimately results in seeing only what one believes in.

Propositions belonging to the thesis, entitled

'*Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins'.

Catarina I.F. Silva
Wageningen, 11 January 2013

# *Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins

Catarina I. F. Silva

**Thesis committee**

**Promotor**
Prof. dr. G. Eggink
Professor of Industrial Biotechnology
Wageningen University

**Co-promotor**
Dr. F.A. de Wolf
Senior scientist at Food and Biobased Research
Wageningen University

**Other members**
Prof. dr. J. van der Oost, Wageningen University
Prof. dr. H.A.B. Wösten, Utrecht University
Prof. dr. H.J. Bosch, Wageningen University
Dr. ir. J. van der Gucht, Wageningen University

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

# *Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins

Catarina I. F. Silva

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday, 11th January 2013
at 1:30 p.m. in the Aula.

Catarina I. F. Silva

*Pichia pastoris* as a cell factory for the secreted production of tunable collagen-inspired gel-forming proteins, 164 pages.

PhD thesis, Wageningen University, Wageningen, The Netherlands (2013)
With references, with summaries in Dutch and English

ISBN: 978-94-6173-457-0

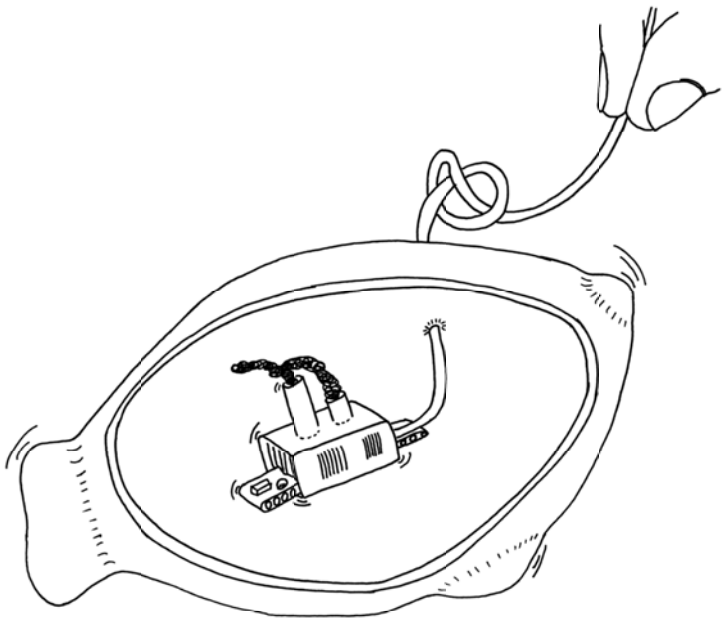
*Imagination is more important than knowledge.*
Albert Einstein

*There are more things in heaven and earth, Horatio,*
*Than are dreamt of in your philosophy.*
in *Hamlet* by William Shakespeare

## CONTENTS

# Chapter *1*

## General introduction



***Polymers** – from the greek polumerēs*
*(consisting of many parts).*

## 1.1 Protein polymers

Nature is a highly skilled protein polymer engineer. With a limited number of relatively simple building blocks, the 20 amino acids, it has been able to create a myriad of highly complex and specialized protein polymers. Proteins are normally subdivided into two groups, namely, fibrous or globular proteins. This distinction is not only conformational but also functional since there is a clear relation between a protein's conformation and its function. Globular proteins are relatively spherical in shape, as the name implies, and predominantly function as enzymes, transport (chaperones) or receptor proteins. Fibrous proteins are characterized by their elongated form, playing a structural role in both animal and plant tissues. Fibrous proteins have a multiple level organizational structure which allows energy absorption and material reinforcement when subjected to deformation or stress [1]. Thanks to this hierarchical structure, these polymers are capable of withstanding many complex and harsh environments while maintaining their mechanical properties, such as toughness, strength and adaptability. Collagen, silk and elastin are fibrous proteins that have received a great deal of attention due to their highly desirable properties (Table 1.1). The capability of these materials to unify contrasting mechanical properties such as strength, robustness and adaptability is of significant interest for tissue engineering, cell scaffolding and drug delivery systems [2]. It is within their relatively simple but highly repetitive amino acid sequence that resides the capacity to adopt the conformations responsible for their mechanical properties (Table 1.1). Despite their advantages, many of these

desirable materials can only be obtained in a cost-efficient manner from extracellular matrices such as bones, hairs and silk cocoons. This origin is not ideal since it poses the risk of pathogen transmission or allergic reactions [3]. Furthermore, the extraction of protein polymers from animal tissues results in batch-to-batch heterogeneity and irreproducibility of the final product characteristics [4]. The difficulty to characterize and predict the final product properties limits their applicability. The lack of control over mechanical properties, nanostructure, degradation rate [5] and strictly dependence of what is available in nature restricts their application in the pharmaceutical and medical fields.

**Table 1.1** Basic features of collagen, silk and elastin protein polymers.

| Biopolymer | Examples of typical amino acid sequence | Conformation | Main functions | Relevant features |
|---|---|---|---|---|
| Collagen | $(Gly\text{-}X_{aa}\text{-}Y_{aa})_n$ [a] | Triple helix with neighbouring molecules | Structural protein in tissues such as connective tissue, tendon, skin bone and cartilage. | Strength and flexibility |
| Silk | $(GA)_n$ or $(GGX_{bb})_n$ or $(GPGX_{bb}X_{bb})_n$ [b] | β-sheet and random coil | Building element of many arthropod nests, cocoon and prey traps. | Strength, stiffness and elasticity |
| Elastin | $(Gly\text{-}Val\text{-}Gly\text{-}Val\text{-}Pro)_n$ [c] | α-helical | Structural protein found predominantly in connective tissue of arteries, ligaments, skin and lung. | Strength, and elasticity |

[a] $X_{aa}$ normally proline and $Y_{aa}$ normaly 4-hydroxyproline.

[b] $X_{bb}$ can be either tyrosine, leucine or glutamine.

[c] The sequence presented is an idealized designer sequence. Natural elastin comprises a multitude of other related sequences.

It is therefore not surprising, that there is intensive research in the development of new systems for the production of polymers inspired by, or capable of, mimicking nature's diverse protein polymers properties.

## 1.2 Protein production

### 1.2.1 Synthetic proteins

Chemical synthesis can be an efficient method to fabricate short peptides in relatively small quantities. However, the solid state synthesis of peptide sequences with more than 35-40 amino acids is not viable due to a drop in yield and efficiency, coupled with an exponential increase in cost [6]. Furthermore, chemical synthesis of longer polypeptides by random polymerization process, originates a mixture of products with different chain lengths [7]. Although, synthetic materials offer significant tunability, which is difficult to achieve with harvested natural materials, concerns remain about the potential immunogenicity and toxicity of degraded fragments, crosslinkers and activating agents, used to manufacture synthetic polypeptides (Table 1.2) [5].

### 1.2.2 Recombinant proteins - inspired by and produced by nature

An alternative to the protein materials harvested from natural sources or synthetic protein polymers, is the use of recombinant protein biomaterials. Recombinant protein biomaterials have the advantage of being able to combine desirable properties from both natural and synthetic biomaterials (Table 1.2) [5]. Additionally, the biosynthesis of peptide polymers using natural pathways allows monodisperse and extremely sophisticated polymers to be synthetized [8]. The advantages of using recombinant DNA approaches include i) the formation of biopolymers with a defined primary sequence, stereochemistry and precise molecular weight by employing genetic templates; ii) reasonable quantities of polymers generated by utilizing recombinant systems; iii) targeted secondary and

tertiary structures by using *in vivo* folding machinery; iv) lack of harsh organic solvents used in synthesis and purification; and v) purification of recombinant proteins is accomplished with the use of water as solvent which is less harmful to the environment. Moreover, the absence of solvents and chemicals makes the proteins readily available for medical applications [9].

Recombinant DNA technology is particularly advantageous in the expression of large and repetitive proteins such as collagen and silk. Codon selection is important since the repetitive nature of the primary sequence presents an opportunity for loss of message integrity through looping out of sequence [10]. For many proteins the codon usage pattern in the native host is quite distinct from that of the microbial host, to the extent that inefficient synthesis and potential deletions or other errors in the product may occur [10]. An optimization of the codon usage according to the recombinant host is therefore advisable.

**Table 1.2.** Comparative advantages and challenges of different protein production methods (adapted from [5]).

| | Synthetic polypeptides | Natural polypeptides | Recombinant protein materials |
|---|---|---|---|
| Toxicity | Can be high | Low | Low |
| Tunability | High | Low | High |
| Yield | Peptide dependent | Low | Low |
| Bioactivity | Can be low | High | High |
| Batch-to-batch consistency | Can be low | Low | High |

Despite the obvious advantages, the design and synthesis of recombinant proteins biomaterials come with a specific set of challenges. So far, a universal system for protein production does not exist. For each recombinant protein a flexible cloning strategy must be designed, a recombinant gene must be synthesized to achieve translational efficiency, an appropriate expression system

must be chosen and purification protocols developed to obtain high yields of engineered protein.

In order to be able to produce protein materials and to determine which recombinant system can best perform the task, one needs to know exactly how they are produced in nature. The main focus of this thesis is the recombinant production of custom-made collagen-like proteins in *Pichia pastoris.* Hence, a deep understanding of how collagen is synthesised and of *P. pastoris* main characteristics, are of fundamental importance.

## 1.3 Collagen

Collagen is characterized by a unique tertiary structure, the triple helix, responsible for maintaining the structural integrity of tissues and organs. Due to this special conformation, all multicellular animals possess collagens as major structural proteins [3]. Collagens also regulate a number of other biological events, including cell attachment, migration and differentiation, tissue regeneration and animal development [3]. The specific functions of collagens are prompted by specific interactions of collagen-binding molecules (membrane receptors, soluble factors and other extracellular matrix components) with structures displayed on the collagen triple helices [3]. Collagen molecules may also contain non-collagenous (non-triple helical) domains involved in other more sophisticated functions than just the molecular architecture of basement membranes. The non-collagenous domains of collagen IV, XV and XVIII are known to inhibit angiogenesis and tumor growth [11, 12]. Proteolytic release of the soluble non-collagenous fragments stimulates migration, proliferation, apoptosis or survival of different cell types and suppresses various morphogenetic events [13]. To date, about 30 genetically distinct types of mammalian collagen molecules have been identified that are assembled into at least 19 collagen varieties found in different tissues in the same individual [14]. Collagen-like molecules have also been found in lower

eukaryotic, prokaryotic, and viral genomes [15]. 'Collagen' is therefore a generic term to cover a large family of distinct proteins, each with specific structures, functions and tissue distribution in the extracellular matrix. Nonetheless, regardless of the type of collagen and its origin one feature is key, viz. the triple helix. All collagen share the same triple helical structure where three parallel polypeptides, α-chains, coil around each other to form a right handed triple helix. In animals, these free collagen triple helices are known as tropocollagen and its hierarchical organization into more complex structures generates fibers and networks in tissues, such as bone, skin tendons, basement membranes and cartilage [16, 17].

The ability of the triple helix to assemble into supramolecular structures (Fig. 1.1) is the reason for collagen commercial uses in food industry and medical applications such as cosmetic surgery and tissue repair.

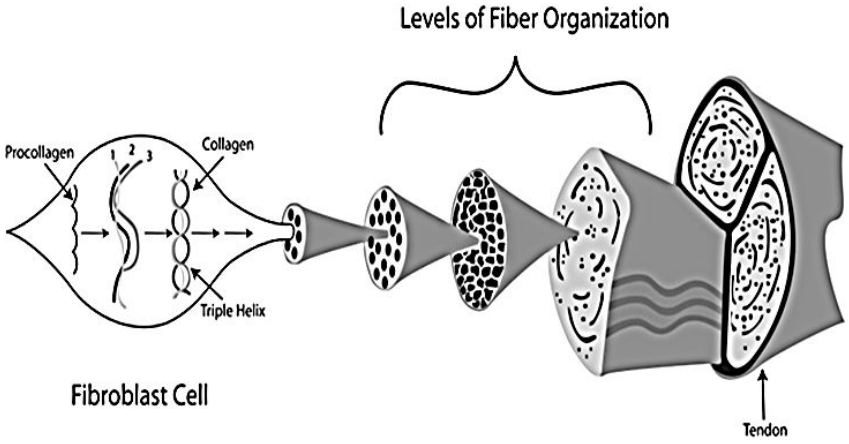



**Figure 1.1** The fibroblast is one of the sites of collagen synthesis. Procollagen molecules are produced individually and then developed into triple helices (tropocollagen) through a series of post-translational modifications in association with other procollagen molecules. The tropocollagen molecules associate to form fibrils. Fibrils are organized into fibers. Successive levels of fiber organization (secondary and tertiary bundles) culminate in the tendon substance. Reproduced with permission from ref [18].

## 1.3.1 Structure and biosynthesis of mammalian collagen

### 1.3.1.1 How is the triple helix conformation achieved?

As mentioned, the triple helix conformation is due to the unique amino acid composition and sequence of the collagen protein. All collagens are characterized by the repetition of triplets of the type (Gly-$X_{aa}$-$Y_{aa}$). It is the presence of glycine every third amino acid that allows the three polypeptide chains to coil around each other since only side chains the size of a hydrogen atom can be accommodated in the triple helix interior [3]. The three polypeptide chains in the triple helix are staggered by one-residue along the common axis of the helix and linked through interchain hydrogen bonds. The hydrogen donors are the peptide –NH groups of glycine residues and the hydrogen acceptors are the peptide –CO groups of residues on the $Y_{aa}$ positions [19]. In many mammalian collagens such as type I or type III, about one-third of the $X_{aa}$ and $Y_{aa}$ positions are occupied by proline and 4-(R)-hydroxyproline residues respectively, which stabilize the helix conformation due to the ring structure and stereoelectronic effects [20, 21]. To further stabilize the triple helix, some lysine residues are also converted to hydroxylysine, which can then react into covalent intermolecular crosslinks. Together with bridging water molecules, the covalent crosslinks, can participate in hydrogen bonds between the polypeptide chains [22-24]. The triple helix is further modified to meet the specialised needs of different tissues such as bone, skin, tendon, teeth, blood vessels and cartilage [19, 24, 25]. Collagen synthesis is a complex process that involves a number of intracellular and extracellular post-translational modifications.

### 1.3.1.2 Mammalian collagen biosynthesis – from triple helix to fibrils

Due to its special structure, and to the number of steps involved to achieve it, mammalian collagen production takes place in highly specialized cells, such as fibroblast, chondrocytes and osteoblasts [6]. Collagens are first synthesized in the

rough endoplasmic reticulum as long rod-like proteins composed of three proline rich chains known as procollagen [26]. In the ER, these chains are packed into a triple helix that is stabilized by hydrogen bonds between the residues in the chain. The folding of the procollagen triple helix requires the hydroxylation of specific prolines by prolyl hydroxylase because hydroxyproline is essential in creating stereoelectronic effects stabilizing the triple helix [21]. If prolyl hydroxylase is inhibited, procollagen does not completely fold and is retained in the ER [27]. It is the different hydroxyproline content of the collagen protein that adjusts the thermal stability of collagen monomers to the body temperature in warm-blooded animals or to the environment temperature in cold-blooded animals [28]. For the proper alignment and folding of the three chains and to regulate the time at which fibril formation is triggered, many procollagen molecules contain two globular propeptides situated at the N- and C-terminal ends [3, 26]. The N- and C-propeptides prevent the premature formation of fibrils and fibres inside the cell [19, 25]. The C-terminal propeptides promote nucleation of triple helix formation which then propagates in a zipper-like manner towards the N-terminus [19, 25].

After triple helix formation inside the ER, the oligomeric complex moves to the Golgi where it aggregates (Fig. 1.2). During transport from the ER to the Golgi, procollagen is found within tubular-saccular structures 300 nm in length [26]. Thus, this supramolecular cargo does not utilize the conventional 60-80 nm vesicle-mediated transport used by globular proteins to traverse the Golgi stacks [26]. Bonfanti *et al*. suggested that procollagen moves in the anterograde direction across the Golgi complex by a process involving progressive maturation of Golgi cisternae. Once in the extracellular space, the N- and C-domains are cleaved off and the tropocollagen subunits spontaneously assemble into larger structures (microfibrils and fibrils) (Fig. 1.2) [26].

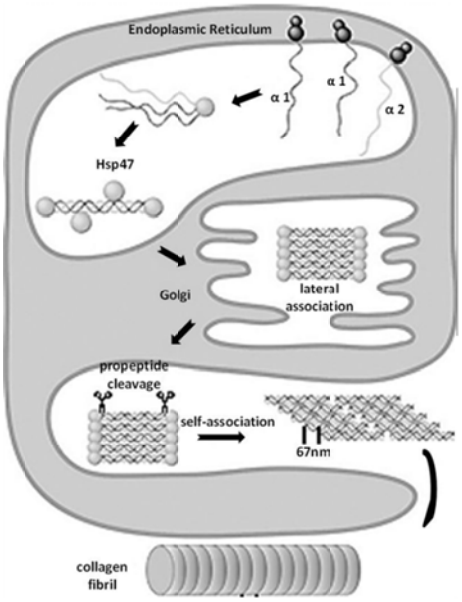



**Figure 1.2.** Schematic representation of the collagen biosynthesis in the endoplasmic reticulum, and of the transport through the Golgi and into the extracellular matrix. The collagen molecules reach their triple helical configuration in the endoplasmic reticulum with the help of peptidyl-prolyl-4-hydroxylase and chaperone Hsp47, and self-associate in various aggregates until a stable 67-nm periodic fibril is obtained. Reproduced with permission from ref [29].

## 1.4 How can collagen be harvested?

### 1.4.1 Collagen and gelatine from animal tissue

Today, the main sources of collagen and its denature counterpart, gelatine, are bovine and porcine tissues posing the risks of pathogen transmission or allergic reactions [3]. Additionally, the variability in composition and structure of animal-derived collagen and gelatine presents a significant challenge for those using these proteins in medical applications, where reliable and predictable materials are essential. Mammalian tissues contain more than one type of collagen and the ratio between the different types of collagen and the degree of crosslinks vary with the type and age of the tissue used, resulting in a lack of reproducibility, and

heterogeneity between batches [4]. The gelatine production process further adds to the heterogeneity of the material. The industrial production of gelatine consists of three main steps: pre-treatment of the raw material (normally skin and bones of cows and pigs), extraction of the gelatine, and purification and drying. The first step, the pre-treatment, consists of either an alkaline (liming) or acid treatment depending on the degree of covalent bonds in the collagen triple helix. Liming is mainly applied to animal tissues, in which collagen has a high degree of covalent links. It breaks the covalent cross-links and improves the efficiency of gelatine extraction. A smaller number of covalent bonds are broken during the acid treatment, which is thus limited to tissues of young animals, or tissues with small amounts of covalent bonds. After pre-treatment, the gelatine is extracted with hot water. The heat breaks the hydrogen and hydrophobic bonds within the triple-helix structure and leads to its dissociation into individual soluble chains, and small fragments. Cooling makes the chains rewind once again into a triple helical structure. However, because the chains have lost their initial register and properties, they do not align correctly and do not fold into complete collagen triple helices. Instead, they form interconnected triple helical junction zones in a random fashion. The helix formation is determined by the coincidental proximity of polypeptide domains and the potential thermostability of helices formed by the interaction of those domains [30]. The final product is composed of a mixture of collagen chains of different length and composition, with various individual melting properties. In view of the low degree of control over the rheological and melting properties that can be obtained for traditional gelatines, there is a need to find new sources of collagen and gelatine in which the properties can be better controlled.

**Table 1.3.** Comparison of critical parameters of animal-derived collagen and recombinant collagen (taken from [4]).

| Parameter | Animal-derived collagen | Recombinant collagen |
|---|---|---|
| Source | Collagen extracted from numerous animal carcasses and tissues | Genetically well-defined expression systems |
| Safety | Risk of contamination with prions and viruses<br>original source not easily identified | No risk of prion/animal virus contamination<br>Fully characterized, defined source |
| Predictability | Batch-to-batch variability high with mixture of type I and III chains. | Consistent and reproducible process leading to predictable performance |
| Structure | Derived from covalently crosslinked type I and type III collagen with degree of crosslinking dependent on animal age and tissue; homogeneous gelatine difficult to purify | Derived from mostly monomeric collagen or well-defined segment of a single chain, from which homogeneous, recombinant gelatine can be derived |
| Product recovery | Complex and variable tissue matrices, require harsh treatment for collagen extraction | Reproducible extracellular or intracellular product recovered by cell/tissue breakage under mild conditions, using pharmaceutical-type bioprocessing |
| Intellectual property | Limited | Property composition, formulation and processing, trademark possible |
| Customizability | Limited | Collagens of specific conformation and composition can be engineered to enhance or remove certain characteristics for the application of interest. |

### 1.4.2 Recombinant Collagen

Over the past 20 years, recombinant systems for the large scale-production of recombinant collagen have been developed and optimized. A comparison of critical parameters of animal-derived collagen and recombinant collagen is summarized in Table 1.3, which illustrates the many advantages of the recombinant production.

Recombinant collagen has been expressed in mammalian cells, insect cells, *Escherichia coli*, transgenic tobacco, mice and silkworm [4, 6]. Of these recombinant hosts, only the mammalian cells expressed collagen with 4-hydroxyproline content identical to native collagen. However, since the level of protein production was low (0.6-20 mg/L) and the process costs (media, etc.) were high, this system was not commercially viable [6]. Accumulation to commercially acceptable levels of various types of collagen has been successfully achieved using *P. pastoris* and *Saccharomyces cerevisae*, making microbial production using yeast the current method of choice for production of recombinant collagen [10, 31-33]. As host organisms, yeasts combine the molecular genetic manipulation and growth characteristics of prokaryotic organisms together with the subcellular machinery for performing post-translational protein modification of eukaryotes [34, 35]. Because yeast recombinant systems have no native proly-4-hydroxylase (P4H) the production of fully hydroxylated and therefore triple helical recombinant collagen relies on its co-expression with an animal or human P4H [32, 36, 37]. The use of a non-hydroxylating recombinant system results in unstable non-triple helical collagen molecules. These are susceptible to potential proteolytic degradation [33]. Furthermore, non-hydroxylated collagen is incapable of self-assembly into collagen fibrils [38], which makes it unsuitable for tissue engineering applications.

### 1.4.3 *Pichia pastoris* as a host system

*P. pastoris* has been widely used to produce recombinant proteins in both basic research and industrial biotechnological applications [39]. *P. pastoris* offers several advantageous features: ease of genetic manipulation; availability of an efficient host-vector system, including the tightly regulated, inducible and very strong *AOX1* promoter and the ability to produce properly folded proteins with correct disulphide bond formation and other eukaryotic posttranslational modifications [35, 39]. Because a methanol-inducible system based on the *AOX1* promoter allows

for tightly regulated high-level expression of foreign genes, the commonly used fermentation strategies are based on the use of glycerol as a substrate for biomass growth, followed by methanol feed for induction [34, 35]. *P. pastoris* is also well developed for industrial fermentation reaching high cell densities on low-cost media [34, 35]. Since *P. pastoris* only secretes few endogenous proteins and proteases [39], the purification of the heterologous product from the culture medium is simplified. Furthermore, *P. pastoris* genome has been recently sequenced [40] opening the door for its improvement as a cell factory through systems biology strategies. Given the obvious advantages, *P. pastoris* is frequently used as a cell factory for the production of a number of recombinant proteins [33-35, 39]. However, the use of this system for the secreted production of collagen triple helix has proven to be a challenge.

### 1.4.4 *Pichia pastoris* and the collagen triple helix

Recombinant type I and III collagen fragments ranging in size from 8.7 to 53 kDa have been shown to be expressed at high levels in *P. pastoris* and secreted into the culture medium as single-chain polypeptides by the yeast α-mating factor (α-MF) pre-pro sequence [33, 41, 42]. In contrast, recombinant full-length triple helical type I and III procolagen molecules are not secreted by neither the authentic collagen signal sequence nor the α-MF, but accumulate within the endoplasmic reticulum (ER) [37, 43-45]. This incapacity to secret full length triple helical procollagen is not unique to *P. pastoris* as it also occurs in *Sacharomyces cerevisae* and insect cells, where the procollagen chains are likewise translocated into the lumen of the ER [25, 32, 46, 47]. Transportation of collagen polypeptide through the secretory pathway of *P. pastoris* appears to be affected by its conformation [36]. Such results could be explained by the absence of collagen-processing and/or collagen-specific transport machinery in *P. pastoris*. It is noteworthy; that a full-length type I procollagen molecule is a rod-like protein 300 nm in length which

does not fit the conventional 60-90 nm transport vesicles of the *P. pastoris* secretory pathway.

*P. pastoris* could thus be used for selective expression of non-secreted triple-helical and secreted single chain fragments, depending on the application. However, the intracellular accumulation of triple helical collagen could cause considerable stress in the host cell and lead to production arrest and protein degradation. In view of the protein yield, it would be more appropriate to use *P. pastoris* for the secreted expression of single chain collagen and perform *in vitro* hydroxylation for the establishment of the triple helical conformation. Also, downstream purification costs account for 80% of the total production costs making the possibility to secrete these proteins extremely appealing from the economic point of view.

Taking this into consideration, our group has developed collagen-inspired block copolymers capable of forming triple helices but not along the full length of the polypeptide chain. Triple helix formation is restricted to certain chain segments (blocks) and does not require proline hydroxylation.

## 1.5 Collagen-like polymers – the non-hydroxylating system

An interesting aspect of the use of recombinant production systems is the possibility to design novel molecules based on the triple helix motif, but different from natural collagens. Our group has developed a yeast-based recombinant system for the production of collagen-inspired triblock polymers. These triblock polymers, known as collagen-like proteins compromise repetitions of the distinctive collagen (Gly-$X_{aa}$-$Y_{aa}$) amino acid domain, in order to establish triple helices and gelatin-like behaviour [31, 48].

Our triblock polymers consist of two thermo-responsive terminal (Pro-Gly-Pro)$_9$ end blocks capable of forming triple helices, denoted as "**T$_9$**" and a middle block, abbreviated as **P$_m$** or **R$_m$**, that assumes random coil conformation at any

temperature (Fig. 1.3) [31, 49]. Due to the high proline content [31] the (Pro-Gly-Pro)$_9$ (**T$_9$**) domain has all the information needed for the triple helix formation and no proline hydroxylation is required [31, 48]. The melting temperature ($T_m$) of the triple helices formed by [(Pro-Gly-Pro)$_9$]$_3$ is ~ 41 °C (at 1.1 mM of protein) meaning, that at this temperature half of the helical structure of the gelatine solution is lost [31]. The protein-protein interactions are done by weak, non-covalent hydrogen bonds, established between the peptide-NH of the glycine and the peptide-CO of the proline C-terminal to the glycine in an adjacent chain. The triple helices are thus reversible in nature, which allows thermoreversible gel formation and, furthermore allows the gels to restructure and flow.

The **P** and **R** blocks are identical with respect to molecular weight and amino acid composition, but have a different amino acid sequence (Fig. 1.3). Block **R** is a randomized version of **P**, in which glycine never occurs in the third position after a previous glycine [31]. As in collagen, glycine occurs every third position in the **P** block, but neither **P** nor **R** can form triple helices and remain in a random coil-like conformation [38] down to 0 °C.

Our **T$_n$P$_m$T$_n$** and **T$_n$R$_m$T$_n$** block copolymers consist therefore of short proline-rich triple helix-forming end blocks that define the $T_m$ of the gel, separated by long random coil spacer blocks that define the distance between the trimer-forming domain in the molecules. As the protein sequence is fully known and triple helix formation is exclusive to the (Pro-Gly-Pro)$_9$ domains, these proteins are capable of establishing a gel network with defined pore sizes and physical-chemical properties [48, 50]. Furthermore, their secreted production by the yeast *P. pastoris* is highly successful.

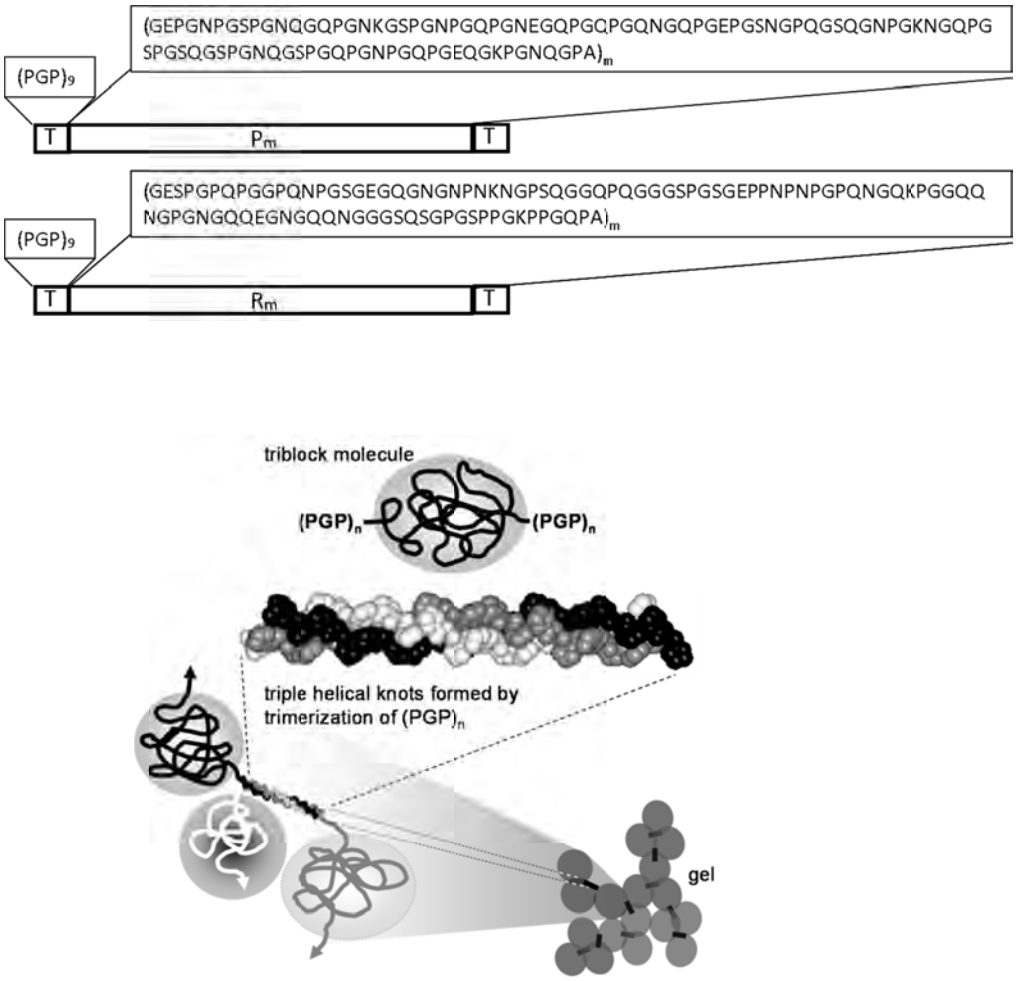



**Figure 1.3.** Structure of the collagen-inspired triblock polymers. (Upper image) Primary structure. The P and R monomers have exactly the same length and composition; with an overall glycine and proline content of 33 and 22%, respectively, no other aminoacids with hydrophobic side groups, and no hydroxyprolyne. P consists of 33 collagen-like (Gly-Xaa-Yaa) triplets and R has a randomized sequence, but both are unable to form triple helices and remain soluble random coils also at low temperature. (Lower image) Schematic representation of gel formation by assembly of triple-helical knots. Three middle blocks originate from each knot.

Our triblock copolymers with $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks and random coild midblocks were produced and secrete at a high yield (2-3 g/L) [31, 33, 49].

Although, these proteins can establish stable triple helices at the optimal *P. pastoris* growth temperature (30°C) this did not seem to cause any secretion impairment. The middle block length also didn't seem to have an effect on the secreted production yield since triblock copolymers with different midblock lengths were equally successfully produced [49].

### 1.5.1 Unveiling sequence-stability and structure stability

The above mentioned triblock system is ideally suited for applications involving the energetic and structural control of materials. It can be used as a model to investigate the relationship between the number of the individual repetitive motifs and their mechanical function in the triple helix.

Our collagen-inspired self-assembling materials were previously used as a model system for a study of physical networks [48, 50]. The well-defined junction multiplicity of three and the singularity of the forces involved, generated exclusively by hydrogen bonds, made it possible to develop a theoretical model based on classical gel theory and calculate the viscoelastic behaviour of the network [48]. It was concluded, that the gel structure and the resulting mechanical properties strongly depend on concentration, temperature and the molecular design of the polymer [48, 49]. An increased size of the middle block markedly increases the gel stiffness and decreases the erosion rate, as a result of decreased formation of intramolecular loops [48, 49].

The potential use of these proteins as drug delivery systems has been showed by Teles *et. al* [49]. However, in the investigated series, only the influence of the midblock length on the viscoelastic and slow release properties of the gel was studied. The networks were established by the exact same trimerizing (Pro-Gly-Pro)$_9$, domain having therefore the same $T_m$ [31]. As mentioned, this domain corresponds to a $T_m$ of 41°C (at 1.1mM), which is not suitable for all medical applications. Furthermore, the present networks erode within 2 days, which is too

fast for tissue engineering applications. The degradation should be synchronized with cellular repair in such a way, that tissue replaces the material within the desired time interval. It would be interesting to design a series of triblocks with custom-made $T_m$ for specific medical applications.

$T_m$ is expected to be dependent on the length [51] and amino acid sequence [52] of the helix-forming end-blocks. Triple helices formed by longer $(Pro\text{-}Gly\text{-}Pro)_n$ stretches will have a higher $T_m$, due to the higher number of hydrogen bonds involved (Fig. 1.3) [52, 53]. In the present thesis, new block copolymers with different lengths of (Pro-Gly-Pro) trimerizing blocks, were designed in order to study its influence on the $T_m$ and viscoelastic properties of the gel. The capability to independently design each block, by genetic engineering, allows the development of a new class of hydrogels with fine-tuned physicochemical properties and tailor made functionalities.

## 1.6 Aim and outline of this thesis

Collagen is a highly desired material in the biomedical field due to its special characteristics; it can provide both strength and flexibility. Although this natural protein, and its denatured counterpart, gelatine, can be easily harvested from animal tissues, the resulting products have unpredictable properties and suffer from batch-to-batch variations. The animal origin of these materials further raises safety concerns and restrains us to what is already available and in sufficient amounts in nature. Chemical synthesis could be an alternative for the production of short-length protein polymers, but for proteins with more than 35-40 aminoacids, the price costs would be too high. Recombinant technologies seem to be the answer to these short-comings. In theory, these new technologies allow the design of new protein materials with predictable and custom-made properties and the use of safe, reliable recombinant systems for its production.

The subject of this thesis is the use of *P. pastoris* as a cell factory for the production of recombinant collagen-inspired proteins with tunable composition, structure and melting temperatures. Our previous results showed that triblock polymers with (Pro-Gly-Pro)$_9$ trimer-forming end blocks, with a $T_m$ of 41 °C, are efficiently produced and secreted by *P. pastoris.* The fact that these polymers can form stable triple helices at the growth temperature (30 °C) of *P. pastoris*, did not seem to have an effect on their secretion. Based on these results, we designed a series of triblock polymers with different (Pro-Gly-Pro)$_n$ length, to study the influence on the $T_m$ and obtain gel networks with different properties. However, the results obtained seem to indicate that the (Pro-Gly-Pro) length influences not only the triple helix $T_m$ but also the protein secretion yield.

In Chapter 2, we describe the influence of the elongation of the trimer-forming end-block to (Pro-Gly-Pro)$_{16}$ on both the triple helix $T_m$ and secretion yield. The elongation of the trimer forming block did increase the $T_m$ but also resulted in a five times lower secretion yield, and partial protein degradation. In this chapter, we study the reasons behind this result by scrambling the amino acid sequence of the trimerizing end block such, that triple helix formation was suppressed. Scrutiny of the proteolysis profile allowed the discovery of the enzyme responsible for the protein degradation. The creation of a strain deficient in the specific protease allowed the secretion of the intact polymer which was then capable of forming self-supporting gels of high thermal stability. Although the proteolytic degradation was overcome, the production yield remained the same in the protease deficient strain. To further optimize *P. pastoris* as a cell factory for the production of collagen-like proteins with high thermal stability, the secretion bottleneck must be known. In Chapter 3, electron microscopy was used to study the intracellular distribution of the recombinant molecules with and without helix-forming capacity. Clear morphological differences between the cells expressing trimer-forming and non-trimerizing proteins could be observed. These observations allowed drawing conclusions on further optimization of *P. pastoris* as a cell factory

for the production of collagen-inspired gel-forming polymers with high thermal stability.

In Chapter 4, a series of triblock copolymers with different triple helix-forming $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks length were designed, and the product intactness compared. In this chapter, the possibility of tuning the $T_m$ independently of the mid-block through the variation of the $(Pro\text{-}Gly\text{-}Pro)_n$ length was investigated. The possible influence of the mid-block on the final $T_m$ was also studied by comparison of the $T_m$ of the triblock with the $T_m$ of free $(Pro\text{-}Gly\text{-}Pro)_n$ blocks of the same length. Differential scanning calorimetry and rheology experiments were performed to characterize the thermostability of the triple helices and the rheological and relaxation properties of the gel network formed. The need to use a *P. pastoris* protease deficient strains according to the $(Pro\text{-}Gly\text{-}Pro)_n$ length was also investigated.

In Chapter 5, the possibility of using a simple technique like SDS-PAGE to study differences in the amino acid sequence and the capacity to establish protein-protein interactions was investigated.

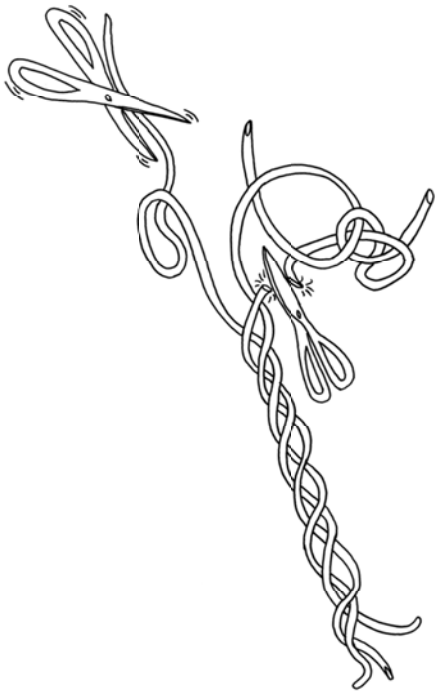
In Chapter 6, the most relevant results obtained in Chapter 2, 3, 4 and 5 are highlighted and conclusions drawn. Also, in this chapter, other relevant results that deserve further research will be discussed.

## References

[1] Johnson, J. C., Wanasekara, N. D., Korley, L. T., *Biomacromolecules* 2012.
[2] Buehler, M. J., Yung, Y. C., *Hfsp J* 2010, *4*, 26-40.
[3] Koide, T., *Philos Trans R Soc Lond B Biol Sci* 2007, *362*, 1281-1291.
[4] Yang, C., Hillas, P. J., Baez, J. A., Nokelainen, M., Balan, J., Tang, J., Spiro, R., Polarek, J. W., *BioDrugs* 2004, *18*, 103-119.
[5] Sengupta, D., Heilshorn, S. C., *Tissue Eng Part B-Re* 2010, *16*, 285-293.
[6] Gomes, S., Leonor, I. B., Mano, J. F., Reis, R. L., Kaplan, D. L., *Prog Polym Sci* 2012, *37*, 1-17.
[7] Kopecek, J., *Eur J Pharm Sci* 2003, *20*, 1-16.
[8] Rodriguez-Cabello, J. C., Girotti, A., Ribeiro, A., Arias, F. J., *Methods Mol Biol* 2012, *811*, 17-38.
[9] Rabotyagova, O. S., Cebe, P., Kaplan, D. L., *Biomacromolecules* 2011, *12*, 269-289.
[10] Ramshaw, J. A. M., Peng, Y. Y., Glattauer, V., Werkmeister, J. A., *J Mater Sci-Mater M* 2009, *20*, 3-8.
[11] Maeshima, Y., Colorado, P. C., Torre, A., Holthaus, K. A., Grunkemeyer, J. A., Ericksen, M. B., Hopfer, H., Xiao, Y., Stillman, I. E., Kalluri, R., *J Biol Chem* 2000, *275*, 21340-21348.
[12] Petitclerc, E., Boutaud, A., Prestayko, A., Xu, J., Sado, Y., Ninomiya, Y., Sarras, M. P., Jr., Hudson, B. G., Brooks, P. C., *J Biol Chem* 2000, *275*, 8051-8061.
[13] Ortega, N., Werb, Z., *Journal of cell science* 2002, *115*, 4201-4214.
[14] Voet, D., Voet, J. G., Pratt, C. W., *Fundamentals Of Biochemistry*, John Wiley & Sons, Inc 2008.
[15] Berisio, R., De Simone, A., Ruggiero, A., Improta, R., Vitagliano, L., *Journal of peptide science : an official publication of the European Peptide Society* 2009, *15*, 131-140.
[16] Heino, J., Huhtala, M., Kapyla, J., Johnson, M. S., *Int J Biochem Cell Biol* 2009, *41*, 341-348.
[17] Shoulders, M. D., Raines, R. T., *Annual review of biochemistry* 2009, *78*, 929-958.
[18] Johnson, M. T., *Front Biosci* 2009, *14*, 1505-1515.
[19] Prockop, D. J., Kivirikko, K. I., *Annual review of biochemistry* 1995, *64*, 403-434.
[20] Improta, R., Berisio, R., Vitagliano, L., *Protein Sci* 2008, *17*, 955-961.

[21] Kotch, F. W., Guzei, I. A., Raines, R. T., *J Am Chem Soc* 2008, *130*, 2952-2953.
[22] Ramachandran, G. N., Reddi, A. H., *Biochemistry of collagen*, Plenum Press, New York.
[23] Cooper, G. M., *The Cell: a molecular approach*, ASM Press, Washington, D. C 2000.
[24] Stryer, L., *Biochemistry*, Freeman, New York 1998.
[25] Myllyharju, J., Kivirikko, K. I., *Annals of medicine* 2001, *33*, 7-21.
[26] Bonfanti, L., Mironov, A. A., Jr., Martinez-Menarguez, J. A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H. J., Mironov, A. A., Luini, A., *Cell* 1998, *95*, 993-1003.
[27] Beck, K., Boswell, B. A., Ridgway, C. C., Bachinger, H. P., *Matrix Biol* 1996, *15*, 155-155.
[28] Leikina, E., Mertts, M. V., Kuznetsova, N., Leikin, S., *Proc Natl Acad Sci U S A* 2002, *99*, 1314-1318.
[29] Persikov, A. V., Brodsky, B., *Proc Natl Acad Sci U S A* 2002, *99*, 1101-1103.
[30] Aalsbersberg, W. Y., Hamer, R. J., Jasperse, P., Jongh, H. H. J., Kruif, C. G., Walstra, P., De Wolf, F. A. (Eds.), *Industrial proteins in perspective* Elsevier, Amsterdam 2003.
[31] Werten, M. W., Teles, H., Moers, A. P., Wolbert, E. J., Sprakel, J., Eggink, G., de Wolf, F. A., *Biomacromolecules* 2009, *10*, 1106-1113.
[32] Toman, P. D., Chisholm, G., McMullin, H., Gieren, L. M., Olsen, D. R., Kovach, R. J., Leigh, S. D., Fong, B. E., Chang, R., Daniels, G. A., Berg, R. A., Hitzeman, R. A., *J Biol Chem* 2000, *275*, 23303-23309.
[33] Werten, M. W., van den Bosch, T. J., Wind, R. D., Mooibroek, H., de Wolf, F. A., *Yeast* 1999, *15*, 1087-1096.
[34] Cereghino, J. L., Cregg, J. M., *FEMS Microbiol Rev* 2000, *24*, 45-66.
[35] Cregg, J. M., Cereghino, J. L., Shi, J., Higgins, D. R., *Mol Biotechnol* 2000, *16*, 23-52.
[36] Pakkanen, O., Pirskanen, A., Myllyharju, J., *Journal of Biotechnology* 2006, *123*, 248-256.
[37] Pakkanen, O., Hamalainen, E. R., Kivirikko, K. I., Myllyharju, J., *J Biol Chem* 2003, *278*, 32478-32483.
[38] Werten, M. W., Wisselink, W. H., Jansen-van den Bosch, T. J., de Bruin, E. C., de Wolf, F. A., *Protein Eng* 2001, *14*, 447-454.
[39] Bollok, M., Resina, D., Valero, F., Ferrer, P., *Recent Pat Biotechnol* 2009, *3*, 192-201.

[40] De Schutter, K., Lin, Y. C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., Van de Peer, Y., Callewaert, N., *Nat Biotechnol* 2009, *27*, 561-566.
[41] Olsen, D., Jiang, J., Chang, R., Duffy, R., Sakaguchi, M., Leigh, S., Lundgard, R., Ju, J., Buschman, F., Truong-Le, V., Pham, B., Polarek, J. W., *Protein Expr Purif* 2005, *40*, 346-357.
[42] Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., Carmichael, D., Perala, M., Hamalainen, E. R., Jarvinen, M., Polarek, J., *Adv Drug Deliv Rev* 2003, *55*, 1547-1567.
[43] Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., Kivirikko, K. I., *EMBO J* 1997, *16*, 6702-6712.
[44] Keizer-Gunnink, I., Vuorela, A., Myllyharju, J., Pihlajaniemi, T., Kivirikko, K. I., Veenhuis, M., *Matrix Biol* 2000, *19*, 29-36.
[45] Nokelainen, M., Tu, H., Vuorela, A., Notbohm, H., Kivirikko, K. I., Myllyharju, J., *Yeast* 2001, *18*, 797-806.
[46] Lamberg, A., Helaakoski, T., Myllyharju, J., Peltonen, S., Notbohm, H., Pihlajaniemi, T., Kivirikko, K. I., *J Biol Chem* 1996, *271*, 11988-11995.
[47] Olsen, D. R., Leigh, S. D., Chang, R., McMullin, H., Ong, W., Tai, E., Chisholm, G., Birk, D. E., Berg, R. A., Hitzeman, R. A., Toman, P. D., *J Biol Chem* 2001, *276*, 24038-24043.
[48] Skrzeszewska, P. J., de Wolf, F. A., Werten, M. W. T., Moers, A. P. H. A., Stuart, M. A. C., van der Gucht, J., *Soft Matter* 2009, *5*, 2057-2062.
[49] Teles, H., Skrzeszewska, P. J., Werten, M. W. T., van der Gucht, J., Eggink, G., de Wolf, F. A., *Soft Matter* 2010, *6*, 4681-4687.
[50] Teles, H., Vermonden, T., Eggink, G., Hennink, W. E., de Wolf, F. A., *Journal of Controlled Release* 2010, *147*, 298-303.
[51] Engel, J., Chen, H. T., Prockop, D. J., Klump, H., *Biopolymers* 1977, *16*, 601-622.
[52] Persikov, A. V., Ramshaw, J. A. M., Brodsky, B., *J Biol Chem* 2005, *280*, 19343-19349.
[53] Brodsky, B., Thiagarajan, G., Madhan, B., Kar, K., *Biopolymers* 2008, *89*, 345-353.



# Chapter *2*

# Secreted production of collagen-inspired gel-forming polymers with high thermal stability in *Pichia pastoris*

Catarina I. F. Silva, Helena Teles, Antoine P. H. A. Moers, Gerrit Eggink, Frits. A. de Wolf and Marc W.T. Werten, *Biotechnology and Bioengineering*, 2011, 108 (11), 2517-2525.

## Abstract

Previously, we have shown that gel-forming triblock proteins, consisting of random coil middle blocks and trimer-forming (Pro-Gly-Pro)$_9$ end blocks, are efficiently produced and secreted by the yeast *Pichia pastoris*. These end blocks had a melting temperature ($T_m$) of ~41°C (at 1.1 mM of protein). The present work reveals that an increase of $T_m$ to ~74°C, obtained by extension of the end blocks to (Pro-Gly-Pro)$_{16}$, resulted in a five times lower yield and partial endoproteolytic degradation of the protein. A possible cause may be that the higher thermostability of the longer (Pro-Gly-Pro)$_{16}$ trimers results in a higher incidence of trimers in the cell, and that this disturbs secretion of the protein. Alternatively, the increased length of the proline-rich (Pro-Gly-Pro)$_n$ domain may negatively influence ribosomal translation, or may result in e.g. hydrophobic aggregation or membrane-active behavior owing to the greater number of closely placed proline residues. To discriminate between these possibilities, we studied the production of molecules with randomized end blocks that are unable to form triple helices. The codon- and amino acid composition of the genes and proteins, respectively, remained unchanged. As these nontrimerizing molecules were secreted intact and at high yield, we conclude that the impaired secretion and partial degradation of the triblock with (Pro-Gly-Pro)$_{16}$ end blocks was triggered by the occurrence of intracellular triple helices. The degradation of the triple helix-forming molecule was overcome by using a yapsin 1 protease disruptant, and the intact secreted polymer was capable of forming self-supporting gels of high thermal stability.

## 2.1 Introduction

Collagen, and its denatured and partly degraded derivative, gelatine, are important biomaterials in regenerative medicine and tissue engineering [1]. The characteristic feature of collagen is its triple-helical structure, consisting of three polypeptide chains in $3_1$ (extended) helical conformation, coiled around each other. Each chain of the triple helix consists of repeating Gly-$X_{aa}$-$Y_{aa}$ triplets, where $X_{aa}$ and $Y_{aa}$ can be various amino acids. The presence of glycine every third amino acid is essential, as only amino acid side chains the size of a hydrogen atom can be accommodated in the triple helix interior [2]. About one-third of the $X_{aa}$ and $Y_{aa}$ positions is occupied by proline residues, which promotes the extended helix conformation. It is only because the prolines in the $Y_{aa}$ position are post-translationally modified to 4-hydroxyprolines by the enzyme peptidyl-prolyl-4-hydroxylase [P4H; 3, 4], that triple helices at this low proline content are thermally stable above 5-15°C.

Despite the importance of collagen and gelatin in biomedical applications, their principal source - animal tissue - poses the risk of immune reactions or transmission of pathogens. The development of recombinant systems for the production of custom-made artificial collagens offers the advantage of a safer material with predictable properties [5, 6]. Recombinant yeast systems such as *Saccharomyces cerevisiae* or *Pichia pastoris* have been successfully used for the production of single-chain (i.e unfolded) collagen-like proteins [7-10]. However, because of the absence of a native P4H in these yeasts, their use for the production of triple-helical collagen relies on the coexpression of heterologous P4H subunits and the collagen polypeptide chains [11-13]. It was found that the triple-helical conformation so obtained completely inhibited collagen secretion, leading to protein accumulation in the yeast cell [14, 15].

Our group has previously developed a nonhydroxylating system for the secreted production of gel-forming triblock copolymers [6]. These ‘gelatins’

consist of long random coils flanked on both ends by collagen-inspired (Pro-Gly-Pro)$_n$ domains. Because of their high proline content, these domains are capable of forming triple helices without prolyl-4-hydroxylation. In this system, the length and composition of the random coil domain, and the length and melting temperature ($T_m$) of the (Pro-Gly-Pro)$_n$ end blocks can be independently tuned, resulting in gels with custom-made properties [6, 16, 17]. The previously described '**T$_9$P$_4$T$_9$**' is such a triblock-copolymer, equipped with a highly hydrophilic ~37 kDa '**P$_4$**' middle block, flanked by two (Pro-Gly-Pro)$_9$ end blocks ('**T$_9$**'). It is capable of forming trimers with a $T_m$ of 41°C [as determined at a protein concentration of 1.1 mM; 6]. Although one might expect this protein to form intracellular triple helices or even gels during secretion, **T$_9$P$_4$T$_9$** was efficiently produced and secreted by *P. pastoris* [6].

Triple helices formed by longer (Pro-Gly-Pro)$_n$ stretches will have higher thermostability owing to the larger number of hydrogen bonds involved [18, 19]. So as to explore the possibilities for production and secretion of gel-forming triblock copolymers with higher $T_m$, we presently increased the trimer thermostability by extending the **T$_9$** end blocks to (Pro-Gly-Pro)$_{16}$, further denoted as '**T$_{16}$**'. This, however, led to a lower yield and partial degradation of the '**T$_{16}$P$_4$T$_{16}$**' product. Although in our polymers only the short end blocks can become triple-helical, the increased trimer thermostabilty of **T$_{16}$** may have triggered a response similar to the above-mentioned arrested secretion of fully triple-helical collagen. On the other hand, elongation of the end blocks also causes a higher demand for Gly and Pro tRNAs, and leads to increased hydrophobicity of the end blocks. Also these effects could potentially influence production levels. To determine whether triple helix formation is the main cause of the impaired secretion of **T$_{16}$P$_4$T$_{16}$**, we developed nontrimerizing variants of **T$_9$P$_4$T$_9$** and **T$_{16}$P$_4$T$_{16}$**. In these polymers, the end blocks were equal in length and amino acid composition to **T$_9$** and **T$_{16}$**, but their sequence was randomized so as to avoid the

$(Gly\text{-}X_{aa}\text{-}Y_{aa})_n$ configuration. On the DNA level the codon usage was kept the same. We also investigated whether the degradation of $\mathbf{T_{16}P_4T_{16}}$ could be remedied by using a protease knock-out strain previously developed in our laboratory [8].

## 2.2 Material and Methods

**Construction of Expression Vectors**

The gene encoding $\mathbf{T_{16}}$ was constructed by concatenating two copies of the **T** gene monomer present in vector pCR4-TOPO-$\mathbf{T_9}$ [6]. The vector was cut with *Rsr*II/*Eco*O109I to release the **T** monomer and, in a separate digestion, linearized with *Eco*O109I and dephosporylated. The released $\mathbf{T_9}$ block was ligated into the linearized and dephosphorylated vector, to yield pCR4-TOPO-$\mathbf{T_{16}}$. The vector pMTL23-$\mathbf{P_4}$ [10] contains a gene encoding the custom-designed, highly hydrophilic ~37 kDa random coiled protein '$\mathbf{P_4}$'. Addition of the $\mathbf{T_{16}}$ gene to both sides of the $\mathbf{P_4}$ gene via *Dra*III/*Van*91I, and the subsequent construction of expression vector pPIC9-$\mathbf{T_{16}P_4T_{16}}$ was exactly analogous to the construction of vector pPIC9-$\mathbf{T_9P_4T_9}$ described previously [6].

The monomeric gene $\mathbf{S_9}$ was designed by randomizing the sequence of the $\mathbf{T_9}$ gene monomer in such a way that not every third residue is a glycine and so, any $(Pro\text{-}Gly\text{-}Pro)_n$ sequence was abolished. This prevents the formation of collagen triple helices by the encoded protein, while the amino acid composition is kept the same. The codon usage of the $\mathbf{S_9}$ gene was also kept identical to that of the $\mathbf{T_9}$ gene. The $\mathbf{S_9}$ block was prepared by overlap extension PCR using the oligonucleotides5'-GAGTCTCACCCGGTGAGC-3'and 5'CCACCGGCTGGA GGTGGAGGACCCGGACCTGGACCTGGTCCTGGAGGTGGACCAGGTCCG GGTGGTGGTCCACCTGGAGGCGGACCGGGCTCACCGGGTGAGACTC-3'. The ~0.1 kb product was cloned into pCR4-TOPO (Invitrogen), resulting in vector pCR4-TOPO-$\mathbf{S_9}$. The $\mathbf{S_9}$ insert was then used for the construction of expression

vector pPIC9-$\mathbf{S_9P_4S_9}$, analogously to the construction of pPIC9-$\mathbf{T_9P_4T_9}$ described previously [6]. The same applies to the concatenation of $\mathbf{S_9}$ to $\mathbf{S_{16}}$ in pCR4-TOPO, and the subsequent generation of vector pPIC9-$\mathbf{S_{16}P_4S_{16}}$.

Transformation of *P. pastoris his4* strain GS115 (Invitrogen) and a derived *his4 yps1* protease disruptant [8] was as described previously [9]. All plasmids were linearized with *Sal*I to promote integration at the *his4* locus rather than the *AOX1* locus, thus enabling normal growth on methanol. Strains containing a single copy of the vector at the *his4* locus were selected by Southern blotting.

**Fermentation of *P. pastoris***

Fed-batch fermentations (initial volume ~1.4 L) were performed in Bioflo 3000 fermenters (New Brunswick Scientific), as described previously [6]. The cultures were inoculated with precultures grown to $OD_{600}$ ~2. The pH was maintained at 3.0 throughout the fermentation. Temperature was kept constant at 30°C and methanol levels were kept at 0.2% (w/v) using a gas sensor-controller.

**Protein Purification**

Purification of all types of triblock copolymers was done by ammonium sulfate precipitation followed by extensive dialysis, which results in a purity of ~99% [6].

**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 pre-stained molecular mass markers. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen).

To allow the quantification of the amount of recombinant protein in the cell-free broth, a calibration curve of 2, 5, 10 and 20 µg of freeze-dried pure protein was included on the same gel as the samples being quantified. A $\mathbf{T_9P_4T_9}$ calibration curve was used to quantify the amount of secreted $\mathbf{T_9P_4T_9}$ or $\mathbf{T_{16}P_4T_{16}}$, and an $\mathbf{S_{16}P_4S_{16}}$ calibration curve was used for $\mathbf{S_{16}P_4S_{16}}$ and $\mathbf{S_9P_4S_9}$. This was done

because, possibly as a result of the formation of triple helices upon removal of SDS from the gels during the washing and staining procedure, $\mathbf{T_9P_4T_9}$ and $\mathbf{T_{16}P_4T_{16}}$ stained relatively strongly compared to $\mathbf{S_9P_4S_9}$ and $\mathbf{S_{16}P_4S_{16}}$. Gel images were acquired using a GS-800 calibrated densitometer (Bio-Rad) and bands were quantified using Quantity One computer software.

N-terminal protein sequencing of SDS-PAGE bands blotted onto PVDF membrane was performed by Midwest Analytical (St. Louis, MO).

**Mass Spectrometry**

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed as previously described [6] on an Ultraflex mass spectrometer (Bruker), using a 2,5-dihydroxyacetophenone matrix and a 600 μm AnchorChip target (Bruker). Protein Calibration Standard II (Bruker) was used for external calibration.

**Differential Scanning Calorimetry**

Degassed 0.5 mL protein solutions (1.2 mM in 0.1 M sodium phosphate pH 7.0) were loaded into a MicroCal VP-DSC calorimeter at 20°C. Protein solutions were equilibrated for 15 h at 20°C to allow complete triple helix formation. Melting curves from 20 to 90°C were recorded at a scan rate of 15°C/h.

**Gel Formation and Inverted Tube Test**

Purified proteins were dissolved at 2.5% (w/v) in Milli-Q water at 85°C, and then allowed to form a gel overnight at room temperature. To test the thermostability, the microtube containing the gel was heated for 2 h at 60°C in a block heater equipped with a heated lid to prevent evaporation (Robbins TruTemp). The tube was then removed from the heater and immediately inverted.

## 2.3 Results

### Biosynthesis of Triblock Gelatins with Extended End Blocks

To produce triblock copolymers with higher thermal stabilities than our previously reported **$T_9P_4T_9$** [6], the '**$T_9$**' DNA that codes for the triple helix-forming block $(Pro\text{-}Gly\text{-}Pro)_9$ was extended to code for $(Pro\text{-}Gly\text{-}Pro)_{16}$. The resulting '**$T_{16}$**' DNA was then used to build the **$T_{16}P_4T_{16}$** triblock DNA, which was inserted into vector pPIC9 for methanol-induced expression and secretion. The vector was transfected into the yeast *P. pastoris* for stable genomic integration, and verified single-copy strains were grown in a bioreactor as previously described for **$T_9P_4T_9$** [6]. SDS-PAGE of cell-free samples of the broth taken ~48 h after methanol induction shows that the concentration of **$T_{16}P_4T_{16}$** in the fermentation broth (Fig. 2.1, lane 3) was considerably lower than that of **$T_9P_4T_9$** (Fig. 2.1, lane 2). To check the reproducibility of the production process and product levels, triplicate fermentations were performed under identical conditions with the **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$** expression strains. Samples were taken at similar time points to determine the wet biomass and protein concentration in the extracellular medium. For both strains, all three fermentations appeared to be highly reproducible in terms of cell growth (Left Fig. 2.2A) and substrate consumption. After ~48 h of methanol induction, a cell density (wet mass) of approximately 600 ± 6 g/L was reached in all three **$T_9P_4T_9$** fermentations and an average of 0.9 ± 0.1 L of methanol was consumed, while the corresponding values for the **$T_{16}P_4T_{16}$** fermentations were 500 ± 40 g/L and 1 ± 0.1 L of methanol. The growth curves of the **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$** strains were very similar, although the strains secreting **$T_{16}P_4T_{16}$** seemed to reach the stationary phase somewhat earlier than those producing **$T_9P_4T_9$** (Left Fig. 2.2A).

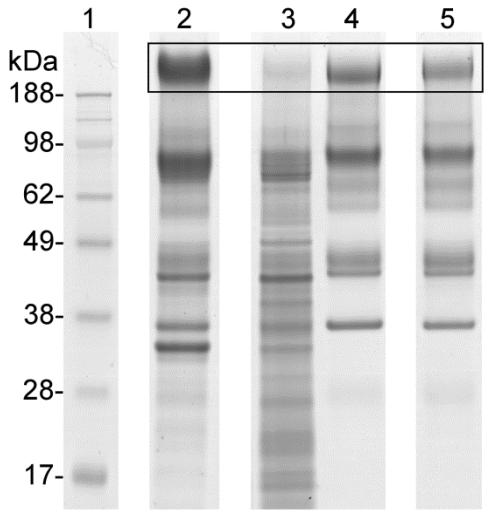



**Figure 2.1.** SDS-PAGE of extracellular medium after ~48 h of methanol induction. Lane 1, molecular weight marker; lane 2, $\mathbf{T_9P_4T_9}$; lane 3, $\mathbf{T_{16}P_4T_{16}}$, lane 4, $\mathbf{S_9P_4S_9}$; lane 5 $\mathbf{S_{16}P_4S_{16}}$. For clarity, a box was placed around the bands that correspond to the triblock copolymers. Owing to the highly hydrophilic nature of the $\mathbf{P_4}$ middle block, the polymers migrate at a very high apparent molecular weight [6, 10]. 10 μL of microfiltered cell-free broth was loaded for all samples.

The high reproducibility of the fermentations, especially for each separate strain, enabled us to compare the production yield of $\mathbf{T_9P_4T_9}$ and $\mathbf{T_{16}P_4T_{16}}$. Densitometric analysis of the gelatin bands after separation of the cell-free broth by SDS-PAGE, showed that at any time after induction, the extracellular concentration of $\mathbf{T_{16}P_4T_{16}}$ was three- to sixfold lower than that of $\mathbf{T_9P_4T_9}$ (Right Fig. 2.2 A). The final extracellular concentration of $\mathbf{T_9P_4T_9}$ and $\mathbf{T_{16}P_4T_{16}}$, after ~48 h of methanol induction, was 2.1 ± 0.2 g/L and 0.38 ± 0.02 g/L, respectively (Right Fig. 2.2A). In agreement, the weight of the final lyophilized products also suggested a more than threefold reduction in yield of $\mathbf{T_{16}P_4T_{16}}$, relative to $\mathbf{T_9P_4T_9}$, although the reproducibility of losses during purification was not strictly monitored.

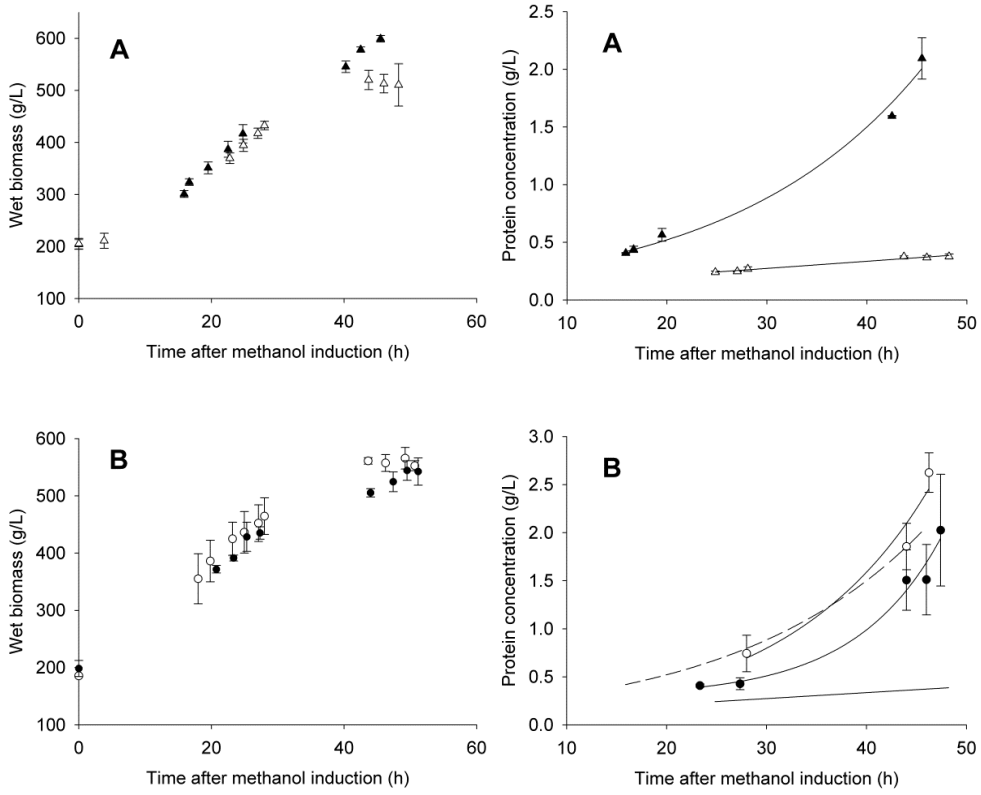



**Figure 2.2.** (**Left Image**) Comparison of growth curves of *P. pastoris* strains expressing different triblock copolymers. (A) $\mathbf{T_9P_4T_9}$ (▲) and $\mathbf{T_{16}P_4T_{16}}$ (△). (B) $\mathbf{S_9P_4S_9}$ (●) and $\mathbf{S_{16}P_4S_{16}}$ (○). The error bars indicate the sample standard deviation of three independent fermentations. (**Right Image**) Production of triblock copolymers as a function of time. The protein concentration in the extracellular medium was quantified by densitometric scanning of the polymer bands after SDS-PAGE. (A) Comparison between $\mathbf{T_9P_4T_9}$ (▲) and $\mathbf{T_{16}P_4T_{16}}$ (△). (B) Comparison between $\mathbf{S_9P_4S_9}$ (●) and $\mathbf{S_{16}P_4S_{16}}$ (○). For convenient comparison, the trend lines of $\mathbf{T_9P_4T_9}$ and $\mathbf{T_{16}P_4T_{16}}$ are repeated in B as dashed and continuous lines, respectively. The error bars indicate the sample standard deviation of three independent fermentations.

To compare the integrity of $\mathbf{T_9P_4T_9}$ and $\mathbf{T_{16}P_4T_{16}}$, purified samples were analyzed by MALDI-TOF. The major mass peak observed for $\mathbf{T_9P_4T_9}$ at 41,739 Da (Fig. 2.3A) corresponds, within experimental error, to its expected molecular weight (41,741 Da), in accordance with our previous work [6]. In the mass spectrum of $\mathbf{T_{16}P_4T_{16}}$ (Fig. 2.3A), the peak at 45,277 Da also corresponds well to the expected molecular weight for the intact protein (45,259 Da). However, several additional peaks of lower molecular weight can be observed. These exact peaks were consistently present in samples taken at different time points throughout several fermentations. The finding of these multiple peaks also matches the fact that the $\mathbf{T_{16}P_4T_{16}}$ band on SDS-PAGE. (Fig. 2.1, lane 3) was rather diffuse, suggestive of a mixture of species. After extended separation of purified $\mathbf{T_{16}P_4T_{16}}$ by SDS-PAGE, several bands could be distinguished (not shown), which were then subjected to N-terminal sequencing. In all cases, two major sequences were found in an approximate 3:2 molar ratio. The first sequence corresponded to the N-terminus of $\mathbf{T_{16}P_4T_{16}}$ as expected after processing of the α-factor prepro sequence. A minor fraction of the molecules had N-terminal Glu-Ala and Glu-Ala-Glu-Ala extensions, which commonly occurs because of partial processing of the α-factor prepro secretory signal by the *P. pastoris* dipeptidylaminopeptidase [9]. The second sequence - GSPGNP - corresponded to a sequence that occurs once in each **P** unit of the middle block. The fact that one distinct internal sequence was found shows that the protein was cleaved by a site-specific endoprotease. The inferred scissile bond is preceded in the protein sequence by a Lys residue. Because the $\mathbf{P_4}$ middle block is a tetramer of four identical **P** units, the obtained internal sequence, and thus also the potential cleavage site, occurs four times in $\mathbf{T_{16}P_4T_{16}}$. The above-mentioned additional peaks can now be explained as follows. Assuming each *individual* molecule is cut only once, at either the second, third or fourth site, then the expected N- and C-terminal halves would have masses of 15,272 / 30,005 Da, 24,381 / 20,896 Da, and 33,490 / 11,787 Da, respectively. Indeed, these masses

match very well with the actually observed secondary peaks in the MALDI-TOF spectrum of **$T_{16}P_4T_{16}$**: 15,249 / 29,984 Da, 24,362 / 20,877 Da, and 33,482 / 11,866 Da. Cleavage at the first site would theoretically result in an N-terminal fragment of 6,164 Da and a C-terminal fragment of 39,113 Da, but the former is below the 10 kDa cut-off of the MALDI-TOF analysis shown (although it was observed in measurements tailored towards a low mass range; not shown), and the latter is likely not detectable because the signal intensity of the various fragments was strongly decreasing with molecular mass (Fig. 2.3A).

**Possible Causes for the Compromised Production of $T_{16}P_4T_{16}$**

It has been reported that triple-helical hydroxylated collagen is not secreted by yeasts, but rather accumulates intracellularly [14, 15]. Although our molecules are not hydroxylated and only the smallproline-rich end blocks can form triple helices, it is conceivable that the secretion of **$T_{16}P_4T_{16}$** may be similarly impaired because the longer **$T_{16}$** blocks form more thermostable triple helices *in vivo* than the **$T_9$** block. To measure the melting temperature ($T_m$) of **$T_{16}P_4T_{16}$** in comparison to that of **$T_9P_4T_9$**, differential scanning calorimetry (DSC) was performed. At a protein concentration of 1.1 mM, **$T_9P_4T_9$** was previously shown to have a trimer $T_m$ of ~41°C [6].

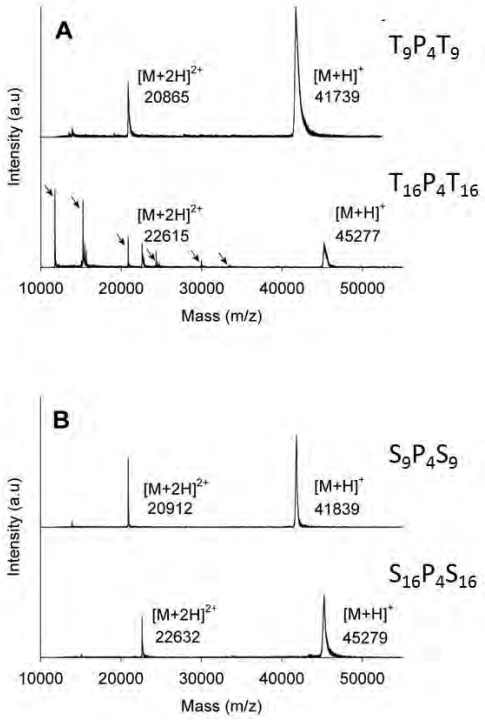



**Figure 2.3.** MALDI-TOF mass spectra of purified secreted triblock gelatins. (A) **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$** and (B) **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$**. Multiply charged charged molecular ions are indicated.

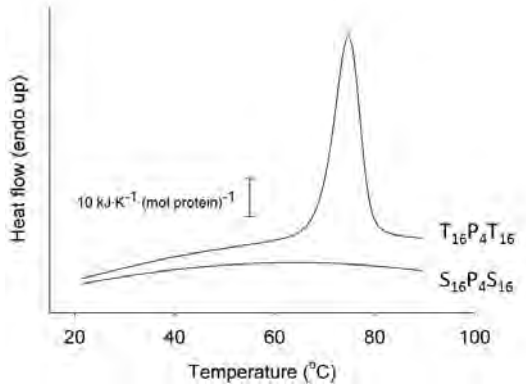



**Figure 2.4.** Thermal denaturation of **$T_{16}P_4T_{16}$** and **$S_{16}P_4S_{16}$** (1.2 mM), as reflected by DSC.

As expected, $\mathbf{T_{16}P_4T_{16}}$ was found to have a much higher $T_m$, namely ~74°C (Fig. 2.4). Because the $T_m$ of $(Pro\text{-}Gly\text{-}Pro)_n$ depends strongly on the peptide concentration [20, 21], and yet the local concentration of the triblock copolymers during transit through the secretory pathway is unknown, it is not possible to directly correlate the *in vitro* $T_m$ values with the growth temperature of the yeast (30°C). Nonetheless, these data do confirm that the $\mathbf{T_{16}}$ end blocks will have a stronger tendency to associate and to give rise to a higher proportion of stable triple helices intracellularly.

Apart from triple helix formation, also e.g. translational problems or increased end block hydrophobicity could be underlying the lower yield of $\mathbf{T_{16}P_4T_{16}}$ as compared to that of $\mathbf{T_9P_4T_9}$. To discriminate between triple helix formation and other mechanisms as the causative factor, modified versions of the trimerizing $\mathbf{T_{16}}$ and $\mathbf{T_9}$ blocks were designed. By randomizing the position of the glycine residues in the end blocks, thus avoiding collagen-like $(Gly\text{-}X_{aa}\text{-}Y_{aa})_n$ triplet repeats, triple helix- and gel formation was made impossible. The end blocks with scrambled amino acid sequence are denoted as '$\mathbf{S_{16}}$' (scrambled version of $\mathbf{T_{16}}$) and '$\mathbf{S_9}$' (scrambled version of $\mathbf{T_9}$). Because $\mathbf{S_{16}}$ and $\mathbf{T_{16}}$ have the same length and amino acid composition (33% glycine and 67% proline), their overall (average) hydrophobicity and/or tendency to deplete the local tRNA pool in the vicinity of the ribosomes during translation, will be equal. For the same purpose of minimizing the differences between the randomized and collagen-like end blocks, except for helix-forming capacity, identical codons were used for the $\mathbf{S_{16}}$ and $\mathbf{T_{16}}$ genes, and for the $\mathbf{S_9}$ and $\mathbf{T_9}$ genes.

**Biosynthesis of Nontrimerizing Triblock Copolymers**

DNA fragments encoding the $\mathbf{S_9}$ and $\mathbf{S_{16}}$ scrambled end blocks were placed on both sides of the $\mathbf{P_4}$ gene, so as to build the $\mathbf{S_9P_4S_9}$ and $\mathbf{S_{16}P_4S_{16}}$ genes, respectively. These genes were then inserted into vector pPIC9. The resulting vector was

transfected into *P. pastoris* and single-copy integrants were selected. Triplicate fermentations for the production of **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** were performed in exactly the same way as for **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$**. The fermentations were again highly reproducible in terms of cell growth (Left Fig. 2.2B) and substrate consumption. After ~48 h of methanol induction, a cell density (wet mass) of approximately 542 ± 23 g/L was reached in all **$S_9P_4S_9$** fermentations and approximately 1 ± 0.1 L of methanol was consumed, while the corresponding values for the **$S_{16}P_4S_{16}$** fermentations were 565 ± 19 g/L and 0.9 ± 0.1 L of methanol. The growth curves of the **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** strains were comparable (Left Fig. 2.2B), and in turn very similar to those obtained with the **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$** strains (Left Fig. 2.2A). This enabled us to directly compare the production of **$S_{16}P_4S_{16}$** to that of **$T_{16}P_4T_{16}$** (and to that of **$S_9P_4S_9$** and **$T_9P_4T_9$**). Densitometric analysis of the **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** bands after separation of the cell-free broth by SDS-PAGE (Fig. 2.1, lanes 4 and 5), showed that at any time after induction, the yield of secreted **$S_{16}P_4S_{16}$** was similar to that of **$T_9P_4T_9$** and **$S_9P_4S_9$**, and three- to sevenfold higher than that of **$T_{16}P_4T_{16}$** ( Right Fig. 2.2B). The final extracellular concentration of **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$**, after ~48 h of methanol induction, was 2.1 ± 0.6 g/L and 2.6 ± 0.2 g/L, respectively.

MALDI-TOF mass spectrometry of purified **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** showed that the proteins were intact. Their mass (41,741 and 45,289 Da, respectively, Fig. 2.3B) corresponds, within experimental error, to the expected molecular mass (41,739 Da and 45,259, respectively). N-terminal sequencing of purified **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** separated by SDS-PAGE revealed only the expected N-terminus (as before, with some traces of N-terminal Glu-Ala and Glu-Ala-Glu-Ala extensions).

Consistent with the scrambled sequence of the end blocks, **$S_{16}P_4S_{16}$** was unable to form triple helices, as evidenced by DSC (Fig. 2.4). Accordingly, and in contrast to **$T_9P_4T_9$** [6], 10% (w/v) solutions of **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** did not form gels at temperatures varying from 20 down to 4°C.

**Biosynthesis of $T_{16}P_4T_{16}$ in a Yapsin 1 Protease Disruptant**

Because the nontrimerizing **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$** were produced intact and at high yield, it can be concluded that the degradation and low yield of **$T_{16}P_4T_{16}$** are triggered by the *in vivo* formation of stable triple helices. If the responsible protease is known, it may be possible to tackle the observed proteolytic degradation. As mentioned above, the endoproteolytic cleavage site in each **P** monomer in **$T_{16}P_4T_{16}$** was preceded by a Lys residue. The yapsin family of aspartic proteases is characterized by the ability to cleave C-terminal of monobasic (and dibasic) residues [22]. Although not all Lys residues in **$T_{16}P_4T_{16}$** were cleaved, it may well be that the specific sites cleaved meet the substrate requirements of any or all of the yapsin proteases of *P. pastoris*. We previously reported the isolation of the *P. pastoris* yapsin 1 (*YPS1*) gene and its disruption [8]. Total yapsin activity in the disruptant was reduced by ~95% relative to the wild type, showing that yapsin 1 was the principal representative of this family of proteases in *P. pastoris* [8]. To determine whether yapsin 1 is involved in the degradation of **$T_{16}P_4T_{16}$**, we transformed our *P. pastoris* GS115 *yps1* strain with vector pPIC9- **$T_{16}P_4T_{16}$**, and the resulting strain was grown in a bioreactor. SDS-PAGE of the extracellular medium revealed the successful production of **$T_{16}P_4T_{16}$** (Fig. 2.5A, lane 2). According to densitometry, the product yield was similar to that of **$T_{16}P_4T_{16}$** in the wild type, and, thus, still ~five times lower than that of **$T_9P_4T_9$**. Just as the triblock copolymers above, the product could be efficiently purified by ammonium sulfate precipitation (Fig. 2.5A, lane 3). The weight of the final lyophilized product was similar to that obtained with the wild type.

MALDI-TOF analysis (Fig. 2.5B) showed that the protein was fully intact, in contrast to **$T_{16}P_4T_{16}$** produced in the wild type (Fig. 2.3A, lower panel). This conclusion was further supported by N-terminal sequencing, which showed exclusively the expected N-terminus. As with the other triblocks, this included a minor fraction of Glu-Ala- and Glu-Ala-Glu-Ala-extended species. This is

consistent with the slight heterogeneity seen in the main peak of the MALDI-TOF spectrum, as the observed masses near-perfectly match the 200 Da mass difference expected for each Glu-Ala pair. These peaks were discernible only in the spectrum of Fig. 2.5B, merely because of the improved resolution achieved in that particular analysis as compared to those of Fig. 2.4.

In contrast to nontrimerizing **$S_{16}P_4S_{16}$**, a 2.5% (w/v) solution of intact **$T_{16}P_4T_{16}$** formed a self-supporting gel at room temperature. Repeated melting at 85°C and cooling showed that the gel formation process was thermoreversible. After incubation at 60°C for 2 h it was still possible to invert the microtube without causing flow, demonstrating the high thermostability of the gel (Fig. 2.6). For comparison, gels of **$T_9P_4T_9$** at 10% (w/v) had a rheologically determined $T_m$ of ~37°C and were completely gone at ~40°C [6].

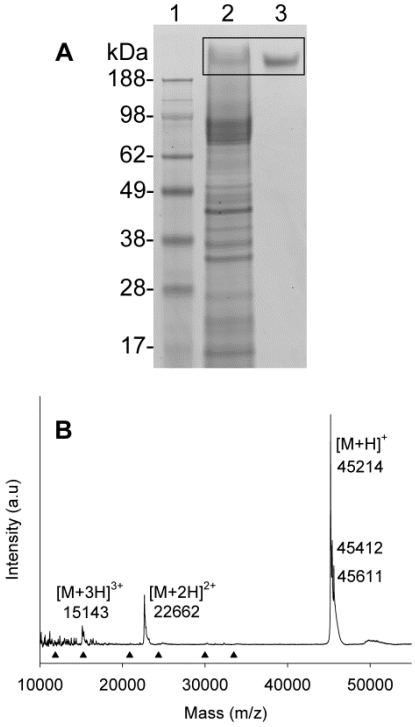



**Figure 2.5.** Characterization of **$T_{16}P_4T_{16}$** produced in *yps1* strain. (A) SDS-PAGE of 10 µl of microfiltered cell-free broth. Lane 1, molecular weight marker; lane 2, extracellular

medium after ~48 h of methanol induction; lane 3, purified protein. (B) MALDI-TOF mass spectrum of partially purified protein (baseline subtracted for clarity). Multiply charged molecular ions are indicated. Up-pointing triangles indicate the position of the fragments found when the protein is produced in wild type *P. pastoris* (compare with Fig. 2.3A, lower panel).

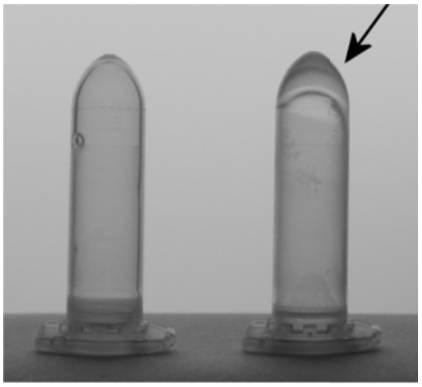


**Figure 2.6.** Inverted microtubes containing nonheated $\mathbf{S_{16}P_4S_{16}}$ and heated $\mathbf{T_{16}P_4T_{16}}$ (2 h 60°C).

## 2.4 Discussion

The previously described successful secreted production of the gel-forming $\mathbf{T_9P_4T_9}$ triblock gelatin [6] allowed us to design a follow-up series of variants with different middle or end blocks. While we have shown that both drug-releasing and viscoelastic properties can be tailored by changing the length of the middle block [17, 23], the melting temperature of the gels is determined by the length of the triple helix-forming end blocks [6, 18, 19]. In this work we explored the possibility to produce a variant with a much higher $T_m$ than before, by increasing the length of the $(Pro\text{-}Gly\text{-}Pro)_n$ end block.

While the $\mathbf{T_9P_4T_9}$ triblock gelatin with $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks had a $T_m$ of ~41°C [as determined by DSC at 1.1 mM of protein; 6], the present $\mathbf{T_{16}P_4T_{16}}$ polymer with $(Pro\text{-}Gly\text{-}Pro)_{16}$ end blocks had a much higher $T_m$ of ~74°C. This

shows that indeed the $T_m$ can be controlled via the length of the $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks. However, while the **$T_9P_4T_9$** triblock copolymer was completely monodisperse and produced at a yield comparable to that of the nontrimerizing middle block alone [6], the **$T_{16}P_4T_{16}$** polymer was not fully intact and was secreted at an appreciably lower yield.

Why would elongation of the **$T_9$** block to **$T_{16}$** have these effects? The observed protein fragments could be thought to represent C-terminally truncated molecules resulting from premature termination of transcription and/or translation. Hypothetically, premature termination of transcription could have been caused by an unfavorable structure specifically of the **$T_{16}$** DNA sequence. Premature termination of translation could have occured in response to the larger number of prolines and glycines in the longer **$T_{16}$** end blocks, as this may have resulted in depletion of the local pool of required tRNAs near the ribosome during translation and subsequent dissociation of the ribosome. However, MALDI-TOF and protein sequencing revealed that the observed protein fragments instead actually represent a mixture of N- and C-terminally truncated species generated by endoproteolytic cleavage at specific sites in the middle block. What remains is the issue of the low yield of **$T_{16}P_4T_{16}$** as compared to **$T_9P_4T_9$**. Possibly, the above-suggested depletion of tRNAs merely causes some ribosome stalling and consequently a lower production rate. Alternatively, the larger number of prolines in the elongated end blocks implies a larger number of prolyl $(CH_2)_3$ side groups and, thus, an increase in hydrophobicity. This may in turn cause protein aggregation intracellularly and arrested secretion, or possibly the relatively hydrophobic **$T_{16}$** end blocks could act as lipophilic segments that perturb cellular membranes. Finally, the higher $T_m$ of the **$T_{16}$** end blocks relative to that of the **$T_9$** blocks could give rise to more stable triple helices at the growth temperature of the yeast. Possibly, the yeast cells are not able to properly secrete the resulting trimeric molecules. Note that it cannot be predicted whether the triblock copolymers will actually form triple-helices *in vivo*,

as the $T_m$ of $(Pro\text{-}Gly\text{-}Pro)_n$ peptides depends strongly on their concentration [20, 21], while their local concentration during secretion is unknown. Moreover, other intracellular compounds could possibly promote or hinder triple helix formation.

The relative importance of increased triple helix formation as a cause of the impaired production of **$T_{16}P_4T_{16}$**, as compared to hydrophobicity or tRNA depletion, was investigated by creating **$S_9P_4S_9$** and **$S_{16}P_4S_{16}$**. The amino acid sequences of the **$S_9$** and **$S_{16}$** end blocks are randomized versions of **$T_9$** and **$T_{16}$**, so as to avoid any $(Gly\text{-}X_{aa}\text{-}Y_{aa})_n$ structure. In this manner the amino acid composition and hydrophobicity of **$T_9$** and **$T_{16}$** are maintained, while rendering trimer formation impossible. At the DNA level, these modified end blocks have exactly the same codon composition as the trimer-forming end blocks, so that also tRNA requirements will remain the same. Since the nontrimerizing **$S_{16}P_4S_{16}$** and **$S_9P_4S_9$** were produced intact and at essentially the same yield as **$T_9P_4T_9$**, increased triple helix formation by the **$T_{16}$** end blocks relative to **$T_9$** must be regarded as the primary cause for both the lower yield and the partial degradation of **$T_{16}P_4T_{16}$**. While the triple helix- and gel-forming capacity of **$T_9P_4T_9$**, which has only short helix-forming domains, is by itself not deleterious to efficient secretion of the intact polymer [6], the length and thermostability of the helix-forming domains apparently cannot be increased without limit. In addition to triple helix formation itself, it may be that possibly resulting supramolecular structures actually cause the impaired secretion. These structures could be gelatin networks [6, 16], and/or aggregates of triple-helical **$T_{16}$** blocks. Such aggregation could occur because triple-helical end blocks are expected to have a more hydrophobic surface than their unfolded single-chain counterparts.

It has been reported that full-length triple-helical collagen is not secreted by yeast and instead accumulates in the endoplasmic reticulum [11, 12, 14, 24, 25]. Also shorter hydroxylated 45 and 9 kDa human α1(I) collagen fragments, fused to a trimer-forming ‘foldon’ domain derived from bacteriophage T4 fibritin, were

only partially secreted by *P. pastoris* [15]. The secreted fragments appeared to be monomeric, while intracellularly retained molecules were entirely triple-helical. Possibly, the lower yield of the potentially trimer-forming **$T_{16}P_4T_{16}$**, as compared to the other triblock configurations, is likewise caused by partial retention in the endoplasmic reticulum or at least retarded secretion. Although we did not observe appreciable amounts of intracellularly accumulated **$T_{16}P_4T_{16}$** in SDS-PAGE of whole cell extracts (not shown), expression of the gene could be subject to negative feedback regulation, or any retained trimeric protein may be proteolytically degraded. The latter likely does not occur with natural hydroxylated collagen sequences as they can become entirely triple-helical and protease-resistant [26], in contrast to our triblock copolymers with large mandatory random coil middle blocks.

Although the random coil **$P_4$** middle block is inherently highly accessible to proteases, it was secreted fully intact when produced on its own [10], or as part of **$T_9P_4T_9$**, **$S_9P_4S_9$**, and **$S_{16}P_4S_{16}$**. In contrast, the **$P_4$** middle block as part of secreted **$T_{16}P_4T_{16}$** was found to be degraded. Thus, proteolysis of the middle block is apparently triggered by the occurrence *in vivo* of triple-helical end blocks. The cleavage in the middle blocks of **$T_{16}P_4T_{16}$** took place C-terminal of specific single Lys residues, the smallest uniquely identifying sequence motif being either Asn-Lys▼ or Lys▼Gly (where the down-pointing triangle indicates the scissile bond). This is not to suggest that the protease involved actually recognizes these particular dipeptide motifs, as its substrate specificity is likely complex. Nonetheless, given the presence of a Lys residue immediately N-terminal of the scissile bond, it seems plausible that a basic residue-specific protease is involved, such as a member of the yapsin family of aspartic proteases [22]. Indeed, when we produced **$T_{16}P_4T_{16}$** in our previously reported *yapsin 1* protease disruptant [8], it was fully intact and able to form gels. The total product level had not improved, showing that the lower yield of **$T_{16}P_4T_{16}$** as compared to **$T_9P_4T_9$** is not somehow a consequence of yapsin-

mediated proteolysis, but rather a separate issue such as discussed in the previous paragraph. Although, for **$T_{16}P_4T_{16}$**, we thus did not obtain the same exceptional g/L yields as before for **$T_9P_4T_9$** [6], a (nonoptimized) secreted yield in the range of several hundreds mg/L of cell-free broth is still reasonably high and offers prospects towards future production of biomedical materials.

Yapsin 1 is a periplasmic protease attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor [22]. Although it has been suggested that, in *S. cerevisiae*, Yapsin 1 may be transiently active already in the late secretory pathway [27, 28], recent data suggest that activation of the zymogen takes place at the cell surface in the periplasmic space [29]. Why does Yapsin 1 exclusively cleave **$T_{16}P_4T_{16}$**? We believe this is most likely a matter of kinetics. If trimeric **$T_{16}P_4T_{16}$** is secreted at a relatively low rate (in terms of traveling the secretory pathway and/or crossing the cell wall), it would be more exposed to Yapsin 1 activity (be it in the late secretory pathway, or in the periplasmic space). Another explanation may be found in the fact that Yapsin 1 is involved in the activation and/or shedding of periplasmic enzymes and proteins during cell wall stress and remodeling [29, 30]. Possibly, any arduous traversal of the cell wall by trimeric **$T_{16}P_4T_{16}$** triggers the cell wall integrity response, which may in turn, as has been found for various forms of cell wall stress [30], lead to vastly induced *YPS1* expression.

Given that only **$T_{16}P_4T_{16}$** was proteolytically degraded, and in view of the periplasmic localization of the protease responsible, it is tempting to speculate whether the secreted triblock molecules, contrary to natural hydroxylated collagens, traveled the secretory pathway in trimeric form. However, it cannot be ruled out that **$T_{16}P_4T_{16}$** only formed trimers upon reaching the periplasmic space, or that the presence of any triple-helical molecules possibly retained in the endoplasmic reticulum indirectly triggered yapsin-mediated degradation of secreted monomeric molecules.

In practical terms, clearly our collagen-inspired triblock design and the use of the *yps1* disruptant allow the secreted production of intact polymers capable of forming triple-helices and gels of high thermal stability. Thus, insights gained in the present study will further the development of tailor-made biomedical gels with an expanded range of functional properties.

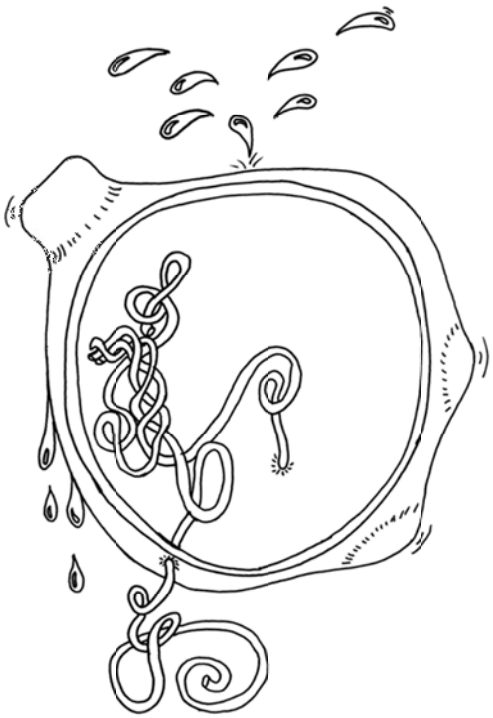
## Acknowledgement

We thank Paulina Skrzeszewska for help with DSC measurements. This work 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] Ramshaw, J. A. M., Peng, Y. Y., Glattauer, V., Werkmeister, J. A., *Journal of materials science Materials in medicine* 2009, *20 Suppl 1*, S3-8.
[2] Koide, T., *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2007, *362*, 1281-1291.
[3] Lamande, S. R., Bateman, J. F., *Seminars in cell & developmental biology* 1999, *10*, 455-464.
[4] Olsen, D. R., Leigh, S. D., Chang, R., McMullin, H., Ong, W., Tai, E., Chisholm, G., Birk, D. E., Berg, R. A., Hitzeman, R. A., Toman, P. D., *The Journal of biological chemistry* 2001, *276*, 24038-24043.
[5] Baez, J., Olsen, D., Polarek, J. W., *Applied microbiology and biotechnology* 2005, *69*, 245-252.
[6] Werten, M. W. T., Teles, H., Moers, A. P. H. A., Wolbert, E. J. H., Sprakel, J., Eggink, G., de Wolf, F. A., *Biomacromolecules* 2009, *10*, 1106-1113.
[7] Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., Carmichael, D., Perala, M., Hamalainen, E.-R., Jarvinen, M., Polarek, J., *Advanced drug delivery reviews* 2003, *55*, 1547-1567.
[8] Werten, M. W. T., de Wolf, F. A., *Appl Environ Microbiol* 2005, *71*, 2310-2317.
[9] Werten, M. W. T., van den Bosch, T. J., Wind, R. D., Mooibroek, H., de Wolf, F. A., *Yeast* 1999, *15*, 1087-1096.
[10] Werten, M. W. T., Wisselink, W. H., Jansen-van den Bosch, T. J., de Bruin, E. C., de Wolf, F. A., *Protein engineering* 2001, *14*, 447-454.
[11] Toman, P. D., Chisholm, G., McMullin, H., Giere, L. M., Olsen, D. R., Kovach, R. J., Leigh, S. D., Fong, B. E., Chang, R., Daniels, G. A., Berg, R. A., Hitzeman, R. A., *The Journal of biological chemistry* 2000, *275*, 23303-23309.
[12] Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., Kivirikko, K. I., *EMBO J* 1997, *16*, 6702-6712.
[13] Vaughan, P. R., Galanis, M., Richards, K. M., Tebb, T. A., Ramshaw, J. A. M., Werkmeister, J. A., *DNA Cell Biol* 1998, *17*, 511-518.
[14] Keizer-Gunnink, I., Vuorela, A., Myllyharju, J., Pihlajaniemi, T., Kivirikko, K. I., Veenhuis, M., *Matrix biology : journal of the International Society for Matrix Biology* 2000, *19*, 29-36.
[15] Pakkanen, O., Pirskanen, A., Myllyharju, J., *Journal of Biotechnology* 2006, *123*, 248-256.
[16] Skrzeszewska, P. J., de Wolf, F. A., Werten, M. W. T., Moers, A. P. H. A., Stuart, M. A. C., Gucht, J. v. d., *Soft Matter* 2009, *5*, 2057-2062.
[17] Teles, H., Skrzeszewska, P. J., Werten, M. W. T., Gucht, J. v. d., Eggink, G., de Wolf, F. A., *Soft Matter* 2010, 4681-4687.
[18] Persikov, A. V., Ramshaw, J. A. M., Brodsky, B., *The Journal of biological chemistry* 2005, *280*, 19343-19349.

[19] Brodsky, B., Thiagarajan, G., Madhan, B., Kar, K., *Biopolymers* 2008, *89*, 345-353.
[20] Engel, J., Chen, H. T., Prockop, D. J., Klump, H., *Biopolymers* 1977, *16*, 601-622.
[21] Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bächinger, H. P., Engel, J., *J Mol Biol* 2001, *308*, 1081-1089.
[22] Gagnon-Arsenault, I., Tremblay, J., Bourbonnais, Y., *FEMS Yeast Res* 2006, *6*, 966-978.
[23] Teles, H., Vermonden, T., Eggink, G., Hennink, W. E., de Wolf, F. A., *J Controlled Release* 2010, *147*, 298-303.
[24] Nokelainen, M., Tu, H., Vuorela, A., Notbohm, H., Kivirikko, K. I., Myllyharju, J., *Yeast* 2001, *18*, 797-806.
[25] Pakkanen, O., Hamalainen, E.-R., Kivirikko, K. I., Myllyharju, J., *The Journal of biological chemistry* 2003, *278*, 32478-32483.
[26] Bruckner, P., Prockop, D. J., *Anal Biochem* 1981, *110*, 360-368.
[27] Ash, J., Dominguez, M., Bergeron, J. J., Thomas, D. Y., Bourbonnais, Y., *J Biol Chem* 1995, *270*, 20847-20854.
[28] Cawley, N. X., Olsen, V., Zhang, C. F., Chen, H. C., Tan, M., Loh, Y. P., *J Biol Chem* 1998, *273*, 584-591.
[29] Gagnon-Arsenault, I., Parise, L., Tremblay, J., Bourbonnais, Y., *Mol Microbiol* 2008, *69*, 982-993.
[30] Krysan, D. J., Ting, E. L., Abeijon, C., Kroos, L., Fuller, R. S., *Eukaryot Cell* 2005, *4*, 1364-1374.



# Chapter 3

# Intracellular localization of extracellular targeted collagen-inspired gel-forming polymers with high thermal stability in *Pichia pastoris*

Catarina I.F. Silva, Gerrit Eggink and Frits A. de Wolf

## Abstract

Helix stability in collagen-inspired proteins (CIP), with small triple helix-forming domains, defines the intactness and yield of CIPs secreted by the yeast *Pichia pastoris*. Yapsin 1, a periplasmic protein, seems to selectively degrade CIPs capable of establishing triple helices with high thermal stability. The aim of this study was to determine the difference in subcellular distribution of CIPs capable of establish highly stable triple helices and virtually identical molecules in which, helix formation was suppressed by scrambling part of the amino acid sequence. Electron microscopy was used to study the intracellular distribution of the recombinant molecules with and without helix-forming capacity. The presence of expanded endoplasmic reticulum (ER) and specific membranous vesicular compartments was found exclusively in cells secreting CIPs with the ability to form triple helices and gels. This shows that formation of highly stable triple helices occurs already in the ER and possibly blocks or retards protein secretion. In order to optimize the yield of proteins capable of forming intermolecular complexes and/or supramolecular assemblies, the formation of the intermolecular triple helices in the ER must be avoided.

## 3.1 Introduction

Mammalian collagen triple helix prodution and assembling, requires special cellular machinery and a transportation system that only seems to exist in highly specialized cells, such as the fibroblats. The collagen molecule consists of three polypeptide chains in extended helical conformation ($3_1$) coiled around each other. Each chain of the triple helix is made of repeating Gly-$X_{aa}$-$Y_{aa}$ triplets, where the presence of glycine every third amino acid is essential, as only side chains the size of a hydrogen atom can be accommodated in the triple helix interior [1-3]. Formation of triple helix is favored by the presence of proline in either position and by hydroxyproline in the $Y_{aa}$ position. Triple helix formation takes place already in the lumen of the fibroblats' endoplasmic reticulum (ER) and the collagen molecule is transported along the secretory pathway as a triple helix [4]. Due to the size of this supramolecular structure (collagen type I is ~ 1,5x300nm) the transport from the ER to the Golgi does not involve the conventional vesicle-mediated (60-80 nm) transport, normally found in yeasts, but tubular-saccular structures of greater length (>300 nm in length). According to Bonfanti *et al.* triple helical collagen molecules move therafter across the Golgi complex of the secretory pathway by cisternal maturation.

Collagen and its denatured counterpart gelatine, are highly desirable biomaterials for biomedical applications such as drug delivery or tissue engineering. However, the only way to obtain such materials in a commercial effective way is by extracting them from animal tissues. Animal origin materials pose the risk of immune reactions and transmission of pathogens. Therefore, there is a need for the development of reliable sources of collagen and gelatine proteins. The recent advances in gene and protein engineering, have allowed the use of recombinant systems as cell factories for the production of natural and custom-made proteins.

The yeast *Pichia pastoris* has been widely used as a recombinant system for the production of secreted mammalian heterologous proteins since its secretory pathway exhibits much of the structure and function of the mammalian secretory system [5-10]. Secreted production is preferred, as proteins produced by *P. pastoris* are usually folded correctly and constitute the majority of total protein in the extracellular medium, facilitating downstream processing and significantly reducing the production costs [11, 12]. However, the use of this system for the secretion of the collagen triple helix has proven to be a challenge. The presence of fully triple-helical collagen molecules completely inhibited secretion, leading to intracellular protein accumulation and ER stress [13, 14]. Pakkanen *et al*. observed that the conformation of collagen polypeptides had a marked effect on secretion, as induction of triple helix formation led to an accumulation of triple-helical molecules inside the cells, despite the presence of the α-mating factor secretory signal. Only single-chain (i.e unfolded) collagen-like proteins were found extracelullarly [5, 15-17].

In Chapter 2, it was shown that the secretion of collagen-inspired proteins capable of forming triple helices with a ~74ºC melting temperature ($T_m$), 44ºC above the growth temperature of *P. pastoris* (30ºC), is inefficient. These polymers, consist of a long ~37 kDa random coil block spacer (**$P_4$**) and two, considerably smaller ~ 4 kDa collagen-inspired $(Pro\text{-}Gly\text{-}Pro)_{16}$ end-blocks (referred as **$T_{16}$**), capable of forming highly stable triple helices. **$T_{16}P_4T_{16}$** was found to be secreted at a low yield (~ 0,3 mg/L) and partially degraded by the periplasmic protease, Yapsin 1. Contrary to the collagen proteins mentioned above, **$T_{16}P_4T_{16}$** cannot form full lenght triple helices and therefore **$P_4$** can be easily accessible by proteases (Fig. 3.1).

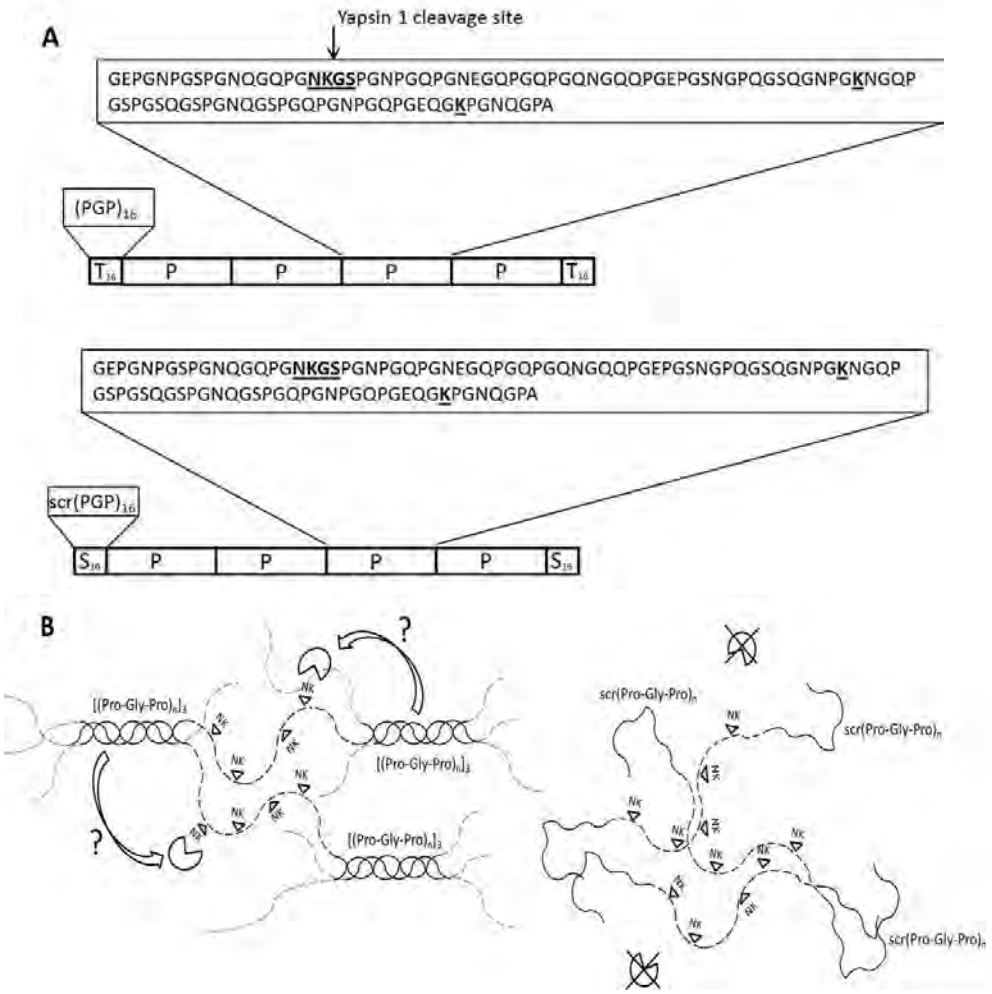



**Figure 3.1.** (A) Schematic representation of $\mathbf{T_{16}P_4T_{16}}$ and $\mathbf{S_{16}P_4S_{16}}$ triblock copolymer structure. Yapsin 1 cleaves C-terminally to monobasic sites containing either Lys or Arg. Despite the presence of three lysines (K) in each **P** block sequence (underlined), Yapsin 1 only cleaves once after the NK aminoacid sequence and in only one concameter. (B) Schematic representation of supramolecular structures that can be formed by $\mathbf{T_{16}P_4T_{16}}$ and $\mathbf{S_{16}P_4S_{16}}$. Full lines (—) are intended to represent end-blocks and dashed lines (- - -) the $\mathbf{P_4}$ middle block. Only $\mathbf{T_{16}P_4T_{16}}$ has $(Pro\text{-}Gly\text{-}Pro)_n$ triple helix-forming end-blocks that can self-assemble and give rise to gels. The scrambled sequence in $\mathbf{S_{16}}$ precludes the formation of triple helices. Protein degradation by Yapsin 1 was only observed in $\mathbf{T_{16}P_4T_{16}}$ proteins (see Chapter 2). Thus, proteolysis of the middle block is either triggered by the *in vivo* occurrence of triple-helical end-blocks or by longer residence times of $\mathbf{T_{16}P_4T_{16}}$ as compared to $\mathbf{S_{16}P_4S_{16}}$.

The low secretion yield and partial degradation of $\mathbf{T_{16}P_4T_{16}}$ by Yapsin 1 was totally abolished by replacing $\mathbf{T_{16}}$ with a variant of $\mathbf{T_{16}}$, the $\mathbf{S_{16}}$ block, in which trimer formation was prevented by scrambling the amino acid sequence and avoiding the presence of glycine every third position (see Chapter 2). Degradation by Yapsin 1 and lower secretion yields seemed therefore to be exclusively related to the production of proteins with end-blocks capable of forming stable triple helices. Intracellular formation of triple helices could either trigger yapsin-mediated degradation and/or simply retard the secretion rate increasing the exposing time of $\mathbf{T_{16}P_4T_{16}}$ to Yapsin 1.

In order to locate where triple helix formation is occurring, the cellular morphology of *P. pastoris* cells producing triple helix and non-triple helix proteins was compared by electron microscopy. Our findings suggest triple helix formation occurs already within the ER and gives rise to expanded ER and extra-cytoplasmic vesicular compartments.

## 3.2 Material and Methods

### Strains and Growth Conditions

Generation of *P. pastoris* strains expressing the triple helix-forming $\mathbf{T_{16}P_4T_{16}}$ and the non-triple helix forming $\mathbf{S_{16}P_4S_{16}}$, have already been described in Chapter 2.

Batch growth was performed in 20 mL shake flasks. The strains were first grown over night at 30°C and 300 rpm in Buffered Minimal Glycerol medium (BMG) at pH6 (Invitrogen) until a $OD_{600}$ of about 4. Expression of heterologous genes was induced in a Buffered Minimal Methanol medium (BMM) at pH6.

### Electron Microscopy

Sample preparation, ultramicrotomy and transmission electron microscopy (TEM) were performed as previously described [18].

## 3.3 Results

### Analysis of the Cellular Morphology of the Different Recombinant *P. pastoris* Strains

The cellular morphology of the methanol-induced recombinant strains expressing triple helix forming **$T_{16}P_4T_{16}$** and non-triple helix forming **$S_{16}P_4S_{16}$** was studied by transmission electron microscopy (TEM). The methanol-induced wild-type strain was used as a control. Analysis of the two recombinant strains revealed the usual subcellular organelles such as nuclei, mitochondria, vacuoles and peroxisomes. Additional membranous vesicular structures (Fig. 3.2A and B) and ER expansion (Fig. 3.3B) were observed solely in cells expressing the **$T_{16}P_4T_{16}$** triple helix forming protein. Some of the membranes of this cytoplasmic tubular network were closely associated with the nuclear membrane (Fig. 3.2B, dash arrow). These vesicular structures and expanded ER were invariably absent, both in wild-type cells and in cells expressing the non-triple helix forming **$S_{16}P_4S_{16}$** (Fig.3.2C and 3.2D). This indicates that these structures had developed specifically in response to triple helix formation which correlates with the reported five-fold lower yield and partial degradation of **$T_{16}P_4T_{16}$** described in Chapter 2.

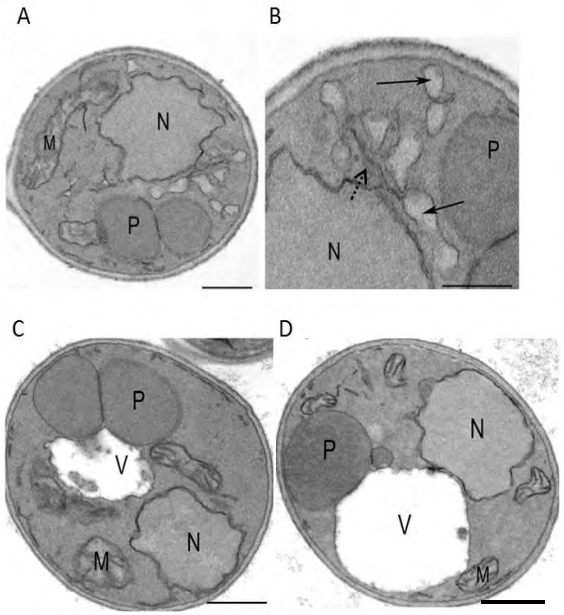



**Figure 3.2.** Analysis of the morphology of recombinant cells producing triple helix-forming $\mathbf{T_{16}P_4T_{16}}$ (A and B), or the non-triple helix forming $\mathbf{S_{16}P_4S_{16}}$ (C), and morphology of wild-type cells (D). The bar in Figure A, C and D represent 500 nm. The bar in Figure B represents 200 nm. The black arrows in (B) indicate the vesicular membranes. The dash arrow in (B) indicates the membranes of swollen ER. N, nucleus: P, peroxisome; M, mitochondrion; V, vacuole.

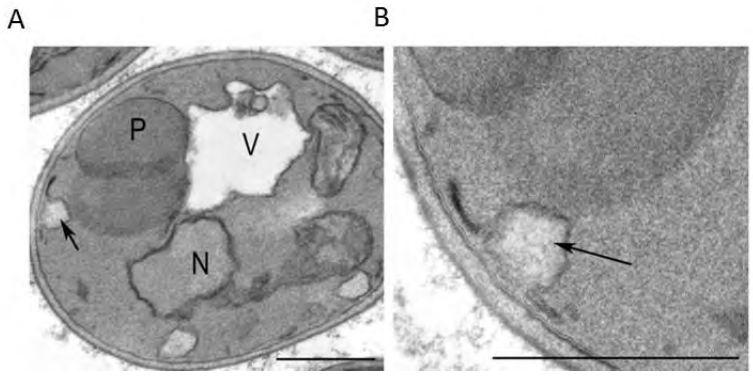



**Figure 3.3** Analysis of the cellular morphology of recombinant $\mathbf{T_{16}P_4T_{16}}$. Image (B) corresponds to the 2 times magnification of image (A). The bar represents 500 nm. The black arrows in (A) and (B) indicate ER expansion. N, nucleus: P, peroxisome; V, vacuole.

## 3.4 Discussion

In eukaryotic cells, all proteins that need to be secreted must pass through the ER to be properly folded and modified. The transitional ER (tER) is a specialized ER subdomain in which proteins destined for the Golgi apparatus are packaged into transport vesicles [19]. In *P. pastoris* such tER sites can only emerge at specific places (Fig. 3.4A) and not along the entire ER membrane like in *Saccharomyces cerevisae* [20]. Thus, at each exit point, a local accumulation of heterologous proteins can occur (Fig. 3.4A). In the particular case of our collagen-inspired proteins, if the concentration needed to form triple helices and gels is reached, such accumulation could lead to local protein aggregation and gel formation (Fig. 3.4B).

In the present study, we observed ER expansion, formation of vesicles and partial protein degradation only when gel-/triple helix forming proteins with a high $T_m$ were being secreted. Since *P. pastoris* does not express autologous collagen-like molecules, it is possible that this yeast lacks the proper transportation system for the adequate secretion of proteins with a triple helical conformation. It is known that the accumulation of misfolded proteins leads to ER stress and activates the unfolded protein response (UPR). The UPR maintains ER homeostasis by two intimately connected but distinct mechanisms: (1) by providing new ER-folding machinery and (2) by providing more ER surface area and lumenal space [21]. ER expansion could be the direct result of protein aggregation caused by triple helix formation or, could be a way to decrease the concentration of accumulated $\mathbf{T_{16}P_4T_{16}}$ and avoid protein aggregation by increasing the ER volume. Dilution caused by an increase of membrane area could help avoid detrimental interactions. This effect is also observed during normal folding of native proteins [21]. During folding, the hydrophobic residues that are normally buried inside the protein, in their final conformations, are exposed. The ER volume is therefore increased to lower the concentration of folding intermediates and avoid aggregate formation.

Although none of the possibilities are mutually exclusive, it is probable that the formation of harmful stable triple helices within the ER causes the expansion of the ER (Fig. 3.4B).

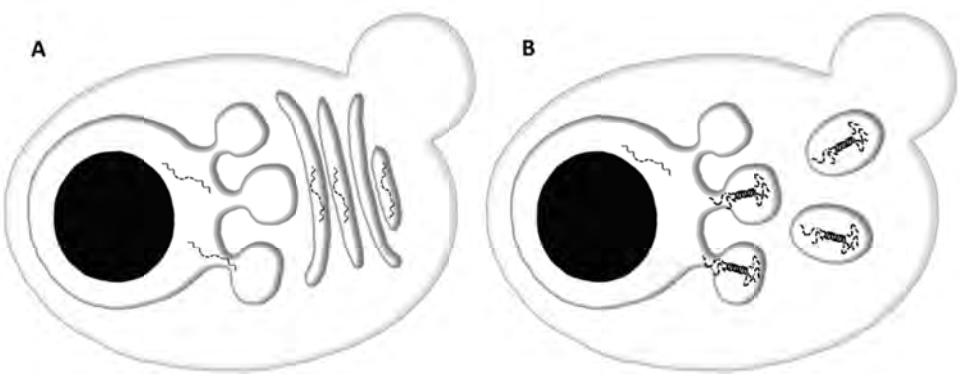



**Figure 3.4.** Schematic representation of the proposed intracellular accumulation of triple helix-forming proteins. (A) *P. pastoris* contains discrete tER sites that produce coherent Golgi stacks. (B) Intracellular accumulation of triple helix-forming proteins leads to ER expansion and vesicle formation.

For the proper optimization of *P. pastoris* as a cell factory, two separate parameters need to be taken into consideration: the quality and the quantity of the recombinant proteins to be produced. In Chapter 2, it was shown that the partial protein degradation of $\mathbf{T_{16}P_4T_{16}}$ can be completely avoided by using a *yapsin 1* mutant strain. However, the secretion yield of $\mathbf{T_{16}P_4T_{16}}$ in the *yapsin 1* mutant was not increased to the level of the non-triple helix forming, $\mathbf{S_{16}P_4S_{16}}$. Yapsin 1 activity seemed only to influence the product quality and not the quantity. This result seems to be in accordance with the observations of this chapter. Yapsin 1 is a zymogen and its activation takes place at the cell surface and is regulated by the environmental pH [22]. If triple helix formation and production arrest seems to occur already within the ER then the deletion of a periplasmic enzyme would not affect the production yield only the quality of the secreted protein.

As it was shown, in Chapter 2, there is no significant increase of triple helix-forming proteins in the fermentation broth throughout the fermentation. The formation of triple helices, already within the ER, could result in aggregation of the relatively hydrophobic triple helices and network (gel) formation. In turn, this could result in different possible effects. For example, a total secretion arrest of triple helical proteins in the ER would allow only monomeric (non-crosslinked, non-network) molecules to leave the compartment and be secreted. The monomeric molecules could represent a minority of the total number of $\mathbf{T_{16}P_4T_{16}}$ molecules in the ER explaining the low secretion yield. Alternatively, a general retardation of the secretion process of $\mathbf{T_{16}P_4T_{16}}$, possibly accompanied by a decreased transcription and/or translation, could be triggered by the intracellular occurrence of triple helices or networks (gels), decreasing the production rate and yield of the heterologous product. Regardless of the mechanism involved, the results obtained in this study imply that the problems associated with the production of high-$T_m$ gelatins may be solved by temporarily suppress triple helix formation already within the ER. HSP47 is a human collagen chaperone thought to bind to non-hydroxylated $(Pro\text{-}Gly\text{-}Pro)_n$ and prevent aggregation of triple helices. However, co-expression of HSP47 and $\mathbf{T_{16}P_4T_{16}}$ was shown to be ineffective in our hands (data not shown). Due to the artificial nature of our proteins it is probable that the use of chaperones might be unsuccessful since there are no natural recognition sites. It is therefore a better option to construct the protein in such a way that triple helix formation would not take place. The attachment of mutually repulsive, highly soluble and / or charged fusion proteins to the $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks (i.e. to the N- and C-terminus of the triblock gelatin) could possibly avoid triple helix formation during transit from the ER to the extracellular space (Fig. 3.5). The extensions that are to be temporarily fused to the end blocks should of course be chosen among proteins that are known to be efficiently secreted by yeast and should be cleaved off after secretion so as to generate gel-forming materials.

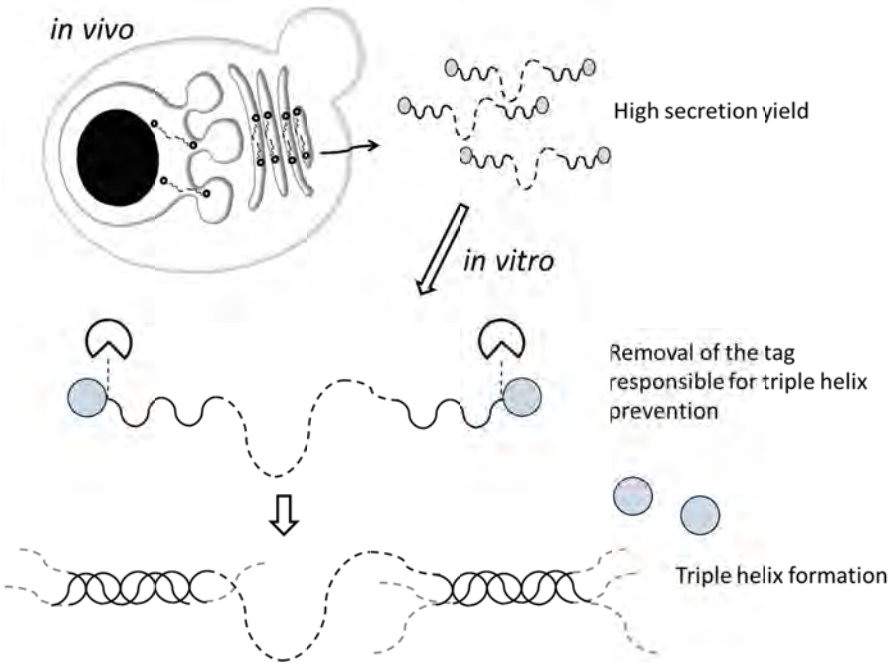



**Figure 3.5**. Schematic representation of the secreted production optimization of collagen-inspired gel-forming polymers in *P. pastoris.* Gel-forming proteins able to establish highly stable triple helices should be designed with a fusion tag capable to prevent triple helix formation. Such tags can then be removed, *in vitro*, by enzymatic cleavage. Triple helix formation should only occur once the tags are removed.

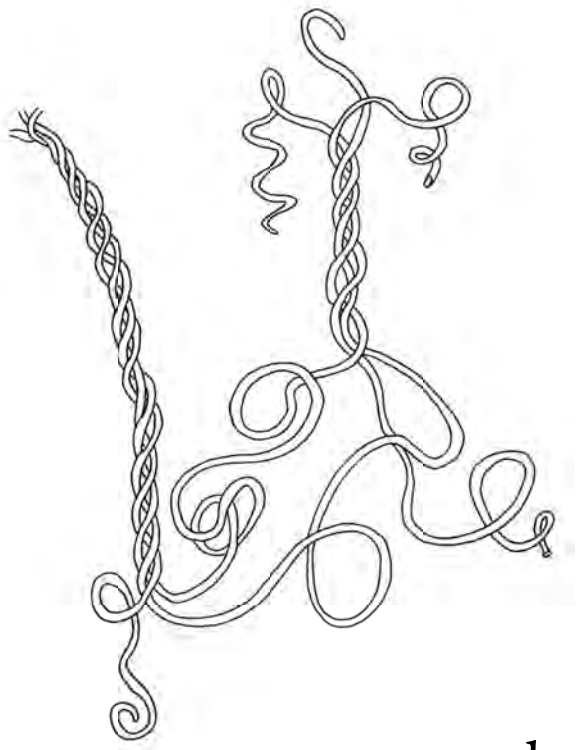
## Acknowledgement

The authors would like to thank to Prof. Ida van der Klei for insightful discussions and Rinse de Boer and Anita Kram for their expert technical assistance. This work 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] Engel, J., Chen, H. T., Prockop, D. J., Klump, H., *Biopolymers* 1977, *16*, 601-622.
[2] Brodsky, B., Persikov, A. V., *Adv Protein Chem* 2005, *70*, 301-339.
[3] Persikov, A. V., Ramshaw, J. A., Brodsky, B., *The Journal of biological chemistry* 2005, *280*, 19343-19349.
[4] Bonfanti, L., Mironov, A. A., Jr., Martinez-Menarguez, J. A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H. J., Mironov, A. A., Luini, A., *Cell* 1998, *95*, 993-1003.
[5] Werten, M. W., van den Bosch, T. J., Wind, R. D., Mooibroek, H., de Wolf, F. A., *Yeast* 1999, *15*, 1087-1096.
[6] Baez, J., Olsen, D., Polarek, J. W., *Appl Microbiol Biotechnol* 2005, *69*, 245-252.
[7] Xie, T., Liu, Q., Xie, F., Liu, H., Zhang, Y., *Prep Biochem Biotechnol* 2008, *38*, 308-317.
[8] Richter, S., Nieveler, J., Schulze, H., Bachmann, T. T., Schmid, R. D., *Biotechnol Bioeng* 2006, *93*, 1017-1022.
[9] Yu, M., Lange, S., Richter, S., Tan, T., Schmid, R. D., *Protein Expr Purif* 2007, *53*, 255-263.
[10] Xie, J., Zhang, L., Ye, Q., Zhou, Q., Xin, L., Du, P., Gan, R., *Biotechnol Lett* 2003, *25*, 173-177.
[11] Bollok, M., Resina, D., Valero, F., Ferrer, P., *Recent Pat Biotechnol* 2009, *3*, 192-201.
[12] Cereghino, J. L., Cregg, J. M., *Fems Microbiol Rev* 2000, *24*, 45-66.
[13] Pakkanen, O., Pirskanen, A., Myllyharju, J., *J Biotechnol* 2006, *123*, 248-256.
[14] Keizer-Gunnink, I., Vuorela, A., Myllyharju, J., Pihlajaniemi, T., Kivirikko, K. I., Veenhuis, M., *Matrix Biol* 2000, *19*, 29-36.
[15] Werten, M. W., Wisselink, W. H., Jansen-van den Bosch, T. J., de Bruin, E. C., de Wolf, F. A., *Protein Eng* 2001, *14*, 447-454.
[16] Werten, M. W., de Wolf, F. A., *Appl Environ Microbiol* 2005, *71*, 2310-2317.
[17] Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., Carmichael, D., Perala, M., Hamalainen, E. R., Jarvinen, M., Polarek, J., *Adv Drug Deliv Rev* 2003, *55*, 1547-1567.
[18] Waterham, H. R., Titorenko, V. I., Haima, P., Cregg, J. M., Harder, W., Veenhuis, M., *The Journal of cell biology* 1994, *127*, 737-749.
[19] Palade, G., *Science* 1975, *189*, 867.

[20] Rossanese, O. W., Soderholm, J., Bevis, B. J., Sears, I. B., O'Connor, J., Williamson, E. K., Glick, B. S., *The Journal of cell biology* 1999, *145*, 69-81.
[21] Schuck, S., Prinz, W. A., Thorn, K. S., Voss, C., Walter, P., *J Cell Biol* 2009, *187*, 525-536.
[22] Gagnon-Arsenault, I., Tremblay, J., Bourbonnais, Y., *FEMS yeast research* 2006, *6*, 966-978.



# Chapter *4*

# Tuning of collagen triple-helix stability in recombinant telechelic polymers

Catarina I.F. Silva, Paulina J. Skrzeszewska, Monika D. Golinska, Marc W. T. Werten, Gerrit Eggink and Frits A. de Wolf, *Biomacromolecules*, 2012, *13* (5), 1250–1258.

## Abstract

The melting properties of various triblock copolymers with random coil middle blocks (100-800 amino acids) and triple helix-forming $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks (n=6-16) were compared. These gelatin-like molecules were produced as secreted proteins by recombinant yeast. The investigated series shows that the melting temperature ($T_m$) can be genetically engineered to specific values within a very wide range by varying the length of the end block. Elongation of the end blocks also increased the stability of the helices under mechanical stress. The length-dependent melting free energy and $T_m$ of the $(Pro\text{-}Gly\text{-}Pro)_n$ helix appear to be comparable for these telechelic polymers and for free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides. Accordingly, the $T_m$ of the polymers appeared to be tunable independently of the nature of the investigated non-crosslinking middle blocks. The flexibility of design and the amounts in which these non-animal biopolymers can be produced (g/L range) create many possibilities for eventual medical application.

## 4.1 Introduction

Designer block copolymers built from various functional protein modules (blocks) are potentially valuable biomaterials with a wide variety of biofunctional and mechanical properties. Tailor-made functional properties can be created through the appropriate design of the amino acid sequence of each block and of the order of the blocks. Thus, the local distribution of hydrophobic and charged groups, the patterns of folding, and the occurrence of supramolecular interactions can in principle be defined [1-7].

One particularly interesting module for such polymers is a collagen-like block with a built-in tendency to form triple helices (trimers) with neighboring molecules [8-11]. Natural collagen consists of three peptide chains folded into an 'extended' (polyproline-II-type) triple helix with three amino acid residues per turn. Glycine, the smallest amino acid, occurs as every third residue, in agreement with the limited space available in the interior of the triple helix. The three chains are staggered by one amino acid along the helix axis, such that the peptide-NH of the glycines can form hydrogen bonds with the peptide-CO of the residues C-terminal to glycine in an adjacent chain. The side chains of the non-glycine residues point away from the central axis of the triple helix and as a result, most amino acids can be readily accomodated in these positions. Formation of the triple helix is favored by the presence of proline in either position, and by hydroxyproline or arginine in the position N-terminal to glycine [12-16]. The φ-angle, which is fixed by the proline side ring, is compatible with the extended helix and cannot contribute much to a gain of entropy when the helix is unfolded. Accordingly, $(Pro\text{-}Gly\text{-}Pro)_n$-based collagen-like peptides of sufficient length readily form triple helices and have been used as model collagen-triple helices [12, 17-19].

Block copolymers with designed amino acid sequence and structure can be produced by recombinant DNA technology. The possibilities provided by materials

produced in this manner contrast with those offered by chemically synthesized polymers or by traditional (natural) protein materials such as gelatine. Chemical synthesis normally yields polydisperse material without a defined monomer sequence, and sequential polypeptides prepared by solid state synthesis cannot be produced in large quantities at reasonable cost. Traditional gelatins are isolated from animal tissues as denatured, partially degraded and modified natural collagen. Since the natural folding pattern of collagen cannot be reconstituted in such gelatins, they follow a frustrated triple helix folding characterized by random intra- and intermolecular interactions and subsequent formation of a random polymer network (gel) [20, 21]. A molecular understanding of the behavior of such gels is hard to achieve, because the properties vary with the composition of the gelatin batch (which essentially consists of a multitude of different molecules) and with the thermal history of the sample (which essentially leads to a multitude of intra- and supramolecular arrangements [22, 23]). Furthermore, concerns regarding products of animal origin make these materials less suitable for medical applications.

Recently, a yeast-based recombinant production system for monodisperse gelatin-like polymers with tailor-made properties was developed by our group [8]. These polymers consist of terminal $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks capable of forming triple helices with a melting temperature ($T_m$) of ~41°C (at 1.1 mM), and middle blocks of defined sequence and length (400-800 amino acids) that assume random coil conformation at any temperature [8, 24]. Because only the short end blocks interact with neighboring molecules to form triple helices and gels, the molecules are telechelic. The rheological and slow-release properties of gels formed by such triblock polymers with different types and sizes of randomly coiled middle blocks have been characterised by us in detail [5, 9, 10, 24, 25]. It was found that, as a result of decreased formation of intramolecular loops, an increased size of the middle block markedly increases the gel stiffness and decreases the rate of erosion.

While several functional properties are determined by the middle block, the $T_m$ of these telechelic polymers is expected to be dependent on the length [12] and amino acid sequence [16] of the helix-forming end blocks. This would allow to tune the $T_m$ through the design of the end blocks. Building on Chapter 2 and previous work [8, 24], we presently compare the triple helix stability under thermal and mechanical stress of polymers having $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks with n-values ranging from 6 to 16, and with $T_m$ values differing by as much as 70°C, at polymer concentrations around 1 mM. Even the highest $T_m$ values were reached without having to insert chemical (covalent) crosslinks that could confer toxicity or compromize biodegradablity. To investigate whether $T_m$ and middle block function can be designed independently, a series of molecules with different middle blocks is included in the analysis, and the $T_m$ of the telechelic molecules is compared to that of free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides.

## 4.2 Material and Methods

### Construction of Expression Vectors

The abbreviated designations $\mathbf{T_n}$, $\mathbf{P_n}$, $\mathbf{R_n}$ and $\mathbf{E_n}$ of the protein polymer blocks used in this work are defined in Figure 4.1. Vectors that contain DNA coding for these blocks are denoted accordingly.

A DNA fragment encoding $\mathbf{T_6}$ (Supporting Information, Table S4.1) was synthesized by GenScript (Piscataway, NJ) and supplied in cloning vector pUC57. The gene was released by digesting the vector with *Dra*III/*Van*91I. The previously-described vector pMTL23-$\mathbf{R_4}$ [8] was linearized with *Dra*III and dephosphorylated. The released $\mathbf{T_6}$ block was ligated into the digested pMTL23-$\mathbf{R_4}$ to yield pMTL23-$\mathbf{T_6R_4}$. This plasmid was then linearized with *Van*91I and dephosphorylated, after which the previously released $\mathbf{T_6}$ block was inserted to yield pMTL23-$\mathbf{T_6R_4T_6}$. The synthetic DNA fragment encoding $\mathbf{T_{12}}$ (Supporting Information, Table S4.1) was

likewise obtained from GenScript and the subsequent construction of pMTL23-**$T_{12}R_4T_{12}$** was analogous to the above. Construction of pMTL23-**$T_{16}R_4T_{16}$** involved the same cloning steps using the **$T_{16}$** block previously described in Chapter 2). For the construction of triblock copolymer genes containing the **$E_{20}$** middle block, an alternative **$T_6$** end block (Supporting Information, Table S4.1) with different restriction sites was synthesized by DNA 2.0 (Menlo Park, CA) and supplied in cloning vector pJ204. The vector, referred to as pJ204-**$T_6$**, was digested with *Bsm*BI and dephosphorylated. The **$E_{20}$** block was released from the previously published vector pMTL23-AIII-**$E_{20}$** [26] with *Fok*I/*Bsm*FI and then ligated into the linearized pJ204-**$T_6$** vector. The resulting pJ204-**$T_6E_{20}$** vector was then digested with *Bsa*I/*Bsm*BI and the released insert was ligated into the dephosphorylated *Bsa*I site of pJ204-**$T_6$** to yield vector pJ204-**$T_6E_{20}T_6$**. All of the above triblock genes were released from the pMTL23 or pJ204 cloning vectors via *Xho*I/*Eco*RI and cloned into the corresponding sites of the *P. pastoris* expression vector pPIC9 (Invitrogen). The resulting vectors pPIC9-**$T_6R_4T_6$**, pPIC9-**$T_{12}R_4T_{12}$**, pPIC9-**$T_{16}R_4T_{16}$** and pPIC9-**$T_6E_{20}T_6$** were then linearized with *Sal*I and transfected into *P. pastoris his4* strain GS115 (Invitrogen) as described previously [27]. Other GS115-based strains used in this study are: **$T_9P_4T_9$**, **$T_{16}P_4T_{16}$** (Chapter 2), **$T_9P_8T_9$**, **$T_9R_4T_9$**, and **$T_9R_8T_9$** (previously denoted as **TP4T** [8], **TP8T** [5, 24], **TR4T** [8, 9], **TR8T** [5, 24] respectively). As mentioned in Chapter 2, a GS115-derived *yps1* protease mutant was use for the expression of **$T_{16}P_4T_{16}$**.

**Fermentation of *P. pastoris***

Fed-batch fermentations (initial volume ~1.4L) were performed in Bioflo 3000 fermenters (New Brunswick Scientific), as described previously [8]. The cultures were inoculated with precultures grown to $OD_{600}$ ~2. The pH was maintained at 3.0 for all strains except **$T_6E_{20}T_6$**, which requires pH 6.0 for optimal protein production [26]. Temperature was kept constant at 30°C and methanol levels were kept at 0.2 % (w/v) using a gas sensor-controller.

**Protein Purification**

Purification of all types of triblock copolymers except $\mathbf{T_6E_{20}T_6}$ was done by ammonium sulfate precipitation as previously described [8], followed by extensive dialysis. The purity of the batches, determined as described previously [3, 8], was ~99 %. The $\mathbf{T_6E_{20}T_6}$ protein was purified from the cell-free broth by the ITC method [26] using 5M NaCl and temperature cycling between 65°C and 4°C. The purified products were dialyzed after which the desalted product was lyophilized.

**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 pre-stained molecular mass markers. Gels were stained with Coomassie SimplyBlue SafeStain (Invitrogen). To allow the quantification of the amount of recombinant protein in the cell-free broth, a calibration curve of 2, 5, 10 and 20 µg of freeze-dried pure protein was included on the same gel as the samples being quantified. Gel images were acquired using a GS-800 calibrated densitometer (Bio-Rad) and bands were quantified using Quantity One computer software. N-terminal sequencing of proteins, either directly or after blotting of SDS-PAGE bands onto PVDF membrane, was performed by Midwest analytical (St. Louis, MO).

**Mass Spectrometry**

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed as previously described [8] on an Ultraflex mass spectrometer (Bruker), using a 2,5-dihydroxyacetophenone matrix and a 600 µm AnchorChip target (Bruker). Protein Calibration Standard II (Bruker) was used for external calibration.

**Differential Scanning Calorimetry (DSC)**

Degassed 0.5 mL protein solutions in 0.1 M sodium phosphate pH 7.0 were loaded into a MicroCal VP-DSC calorimeter at 20°C. Protein solutions were equilibrated for 15 h at 20°C for triblock gelatins with **$T_9$**, **$T_{12}$** and **$T_{16}$** end blocks, and 2°C for **$T_6R_4T_6$**, to allow complete triple helix formation. **$T_6E_{20}T_6$** was equilibrated for 5 hours at 20ºC followed by 5 hours at 15ºC, 5 hours at 10ºC and finalized with 10-15 hours at 5ºC. Melting curves were recorded at a scan rate of 15°C/h. The concentrations used were 2.5 mM **$T_6R_4T_6$**, 16.6 mM **$T_6E_{20}T_6$**, 0.43-2.4 mM **$T_9R_4T_9$**, 0.45-2.2 mM **$T_9P_4T_9$**, 0.68 mM **$T_{12}R_4T_{12}$**, 1.3 mM **$T_{16}R_4T_{16}$** and 0.2-1.3 mM **$T_{16}P_4T_{16}$**. The calorimetric transition enthalpy was obtained by integration of the area under the excess heat capacity peak. Generally, the flooring of the thermogram and the determination of the area under the endotherm involved the fitting of the thermogram baseline with a third-order polynominal and the application of Simpons's rule.

**Rheology**

A 0.7 mM **$T_{16}P_4T_{16}$** sample was prepared in 10 mM phosphate buffer (pH 7) and then heated at 85°C for 15 minutes, allowing the protein to dissolve completely under conditions where no triple helices are formed. The rheological measurements were made using an Anton Paar Physica MCR 301 rheometer equipped with a cone and plate geometry of 50 mm diameter. The temperature was controlled by a Peltier system, which allowed fast heating and cooling. A solvent trap was used to minimize evaporation. Before addition of the heated protein solution, the plate was preheated to 85°C. After lowering the cone, the system was quenched to 20°C. Gel formation was monitored for 15 hours by applying a sinusoidal deformation to the system (f=1 Hz and γ=1 %) and determining the storage (*G'*) and loss (*G"*) moduli. To determine the stress relaxation, creep experiments were carried out at 20ºC, after $G_0$' reached its plateau value. A stress of 5 and 10 Pa was applied to the **$T_{16}P_4T_{16}$** and **$T_9P_4T_9$** gel, respectively, and the strain response was followed in

time. Both stress values were within the linear regime. The duration of the deformation phase was 1800 s, which was followed by 1800 s of recovery.

## 4.3 Results

### Production of Telechelic Polymers by *P. pastoris*

The structure of the telechelic protein polymers and their constituting $\mathbf{T_n}$, $\mathbf{P_n}$, $\mathbf{R_n}$ and $\mathbf{E_n}$ blocks is schematically illustrated in Figure 4.1. As examples of the production of telechelic polymers with various $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks, the production of $\mathbf{T_6R_4T_6}$ and $\mathbf{T_{12}R_4T_{12}}$ is shown in Figure 4.2. *P. pastoris* strains containing the genes encoding $\mathbf{T_6R_4T_6}$ and $\mathbf{T_{12}R_4T_{12}}$, with preprosequences targeting the products to the extracellular medium, were cultured in bioreactors. Cell-free fermentation broths and purified products were analyzed by SDS-PAGE and MALDI-TOF (Fig. 4.2). The purified proteins migrated as single bands in SDS-PAGE, confirming a high purity and intactness (see also Materials and Methods). As expected [8, 28], the proteins showed retarded migration, owing to the high polarity of the middle block [28] and the high content (relative to the marker proteins) of small amino acids residues. The molecular weights observed in MALDI-TOF were in good agreement with the expected values of 40.23 kDa and 43.25 kDa for $\mathbf{T_6R_4T_6}$ and $\mathbf{T_{12}R_4T_{12}}$, respectively. Furthermore, N-terminal sequencing revealed that the proteins were N-terminally intact. The intactness of $\mathbf{T_6R_4T_6}$ and $\mathbf{T_{12}R_4T_{12}}$ agrees with that of all triblock copolymers with $\mathbf{T_6}$ and $\mathbf{T_9}$ end blocks produced in *P. pastoris*, but contrasts with the previously observed partial degradation of $\mathbf{T_{16}P_4T_{16}}$, which originated from *in vivo* formation of stable $\mathbf{T_{16}}$ triple helices, and which could only be overcome by recurrence to a protease knockout mutant (Chapter 2). The yield of $\mathbf{T_{12}R_4T_{12}}$ (~0.6 g/L in the cell-free broth, according to densitometry) was intermediate between the yields of the polymers with $\mathbf{T_9}$ end blocks (several g/L in the cell-free broth [8]) and polymers with $\mathbf{T_{16}}$

end blocks (typically ~0.4 g/L in the cell free broth as seen in **Chapter 2**). The yield of $\mathbf{T_6R_4T_6}$ was comparable to that of the polymers with $\mathbf{T_9}$ blocks.

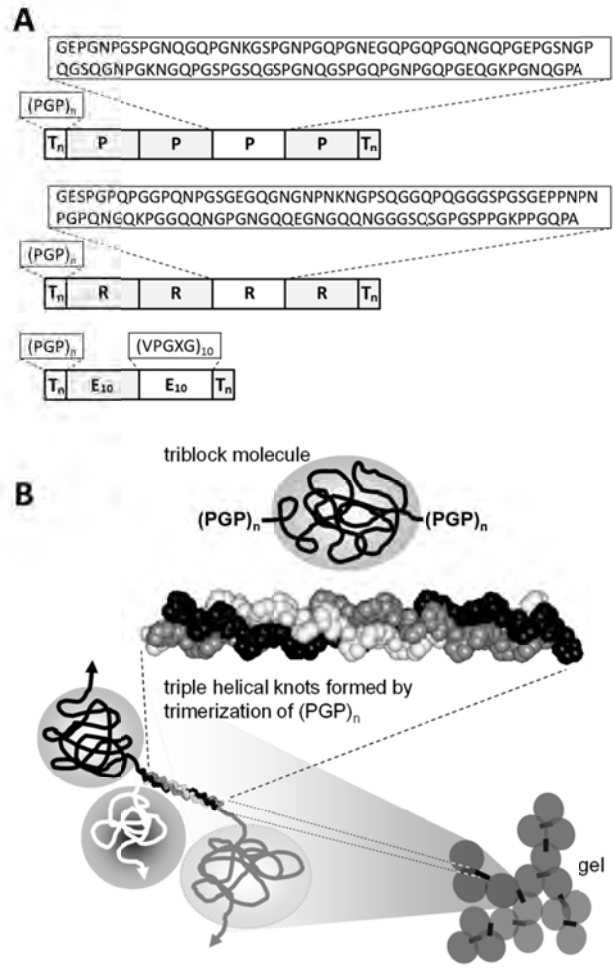



**Figure 4.1.** Structure of the presently-studied collagen-inspired telechelic polypeptides. (A) Primary structure. As examples of middle blocks, $\mathbf{P_4}$, $\mathbf{R_4}$ and $\mathbf{E_{20}}$ are shown, but octamers of **P** and **R** were also used. The **P** and **R** monomers have exactly the same length and composition, with an overall glycine and proline content of 33 and 22 %, respectively, no other amino acids with hydrophobic side groups, and no hydroxyproline. **P** consists of 33 collagen-like (Gly-$X_{aa}$-$Y_{aa}$) triplets and **R** has a randomized sequence, but both are unable

to form triple helices and remain soluble random coils, owing to the absence of sufficient helix-promoting amino acids (no hydroxyproline and no high proline content [8]). The monomer $\mathbf{E_{10}}$ consists of ten Val-Pro-Gly-Xaa-Gly peptide repeats with a 30, 20 and 50 % incidence of Gly, Ala and Val, respectively, in the $X_{aa}$ position [26]. At the temperatures and salt (buffer) concentrations used in this work, the free $\mathbf{E_{20}}$ molecule without end blocks is below its lower critical solution temperature and forms a soluble random coil [29], but preliminary data suggest that in gel form, some turbidity may occur. (B) Schematic representation of gel formation by assembly of triple-helical knots. Reproduced with permission from Ref. 8. Copyright 2009 American Chemical Society. Three mid-blocks originate from each knot. Each mid block also has a $(Pro\text{-}Gly\text{-}Pro)_n$ block at its other end (not shown for clarity).

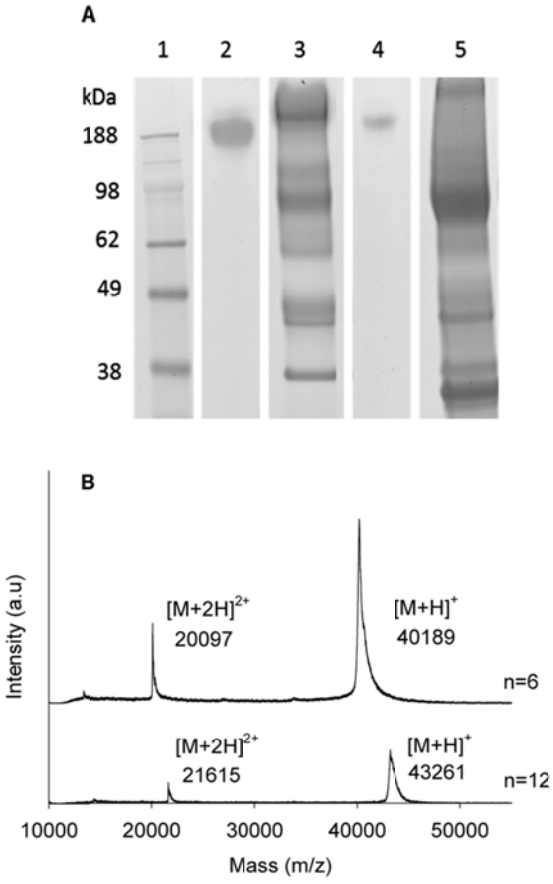



**Figure 4.2.** Production of two examples of telechelic polymers with different end block lengths. (A) SDS-PAGE: lane 1, molecular weight marker; lane 2, purified $\mathbf{T_6R_4T_6}$; lane 3,

cell-free $\mathbf{T_6R_4T_6}$ fermentation broth; lane 4, purified $\mathbf{T_{12}R_4T_{12}}$ and lane 5, cell-free $\mathbf{T_{12}R_4T_{12}}$ fermentation broth. (B) MALDI-TOF of purified $\mathbf{T_6R_4T_6}$ and $\mathbf{T_{12}R_4T_{12}}$. Single- and double-charged molecular ions are indicated. SDS-PAGE gels and MALDI-TOF spectra of purified triblocks with $\mathbf{T_9}$ [8] and $\mathbf{T_{16}}$ end blocks (Chapter 2) were previously shown.

### Thermostability of Triple Helices as Determined by End Block Length

The melting profiles of the triple-helical end blocks were measured by DSC. Endothermic transitions occurred upon melting of all four triblocks. At increasing length of the end block, the maximum of the excess heat capacity ($\Delta C_p$ peak) occurred at a higher temperature, and the melting enthalpies (area under the peaks) increased accordingly (Fig. 4.3). Repeated DSC experiments on various samples resulted in the average enthalpy values shown in Figure 4.4.

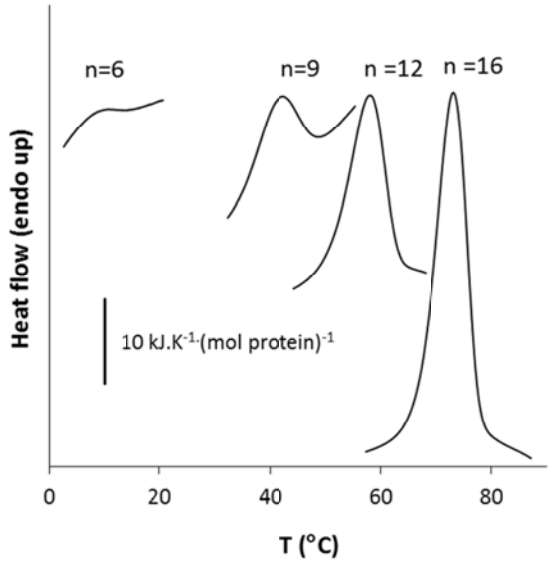



**Figure 4.3.** Examples of thermal denaturation of $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks with different lengths (n) in telechelic polymers, as reflected by DSC. The protein concentrations were 2.4, 1.4, 0.68 and 1.3 mM for n=6, 9, 12 and 16, respectively. These molecules all had an $\mathbf{R_4}$ middle block.

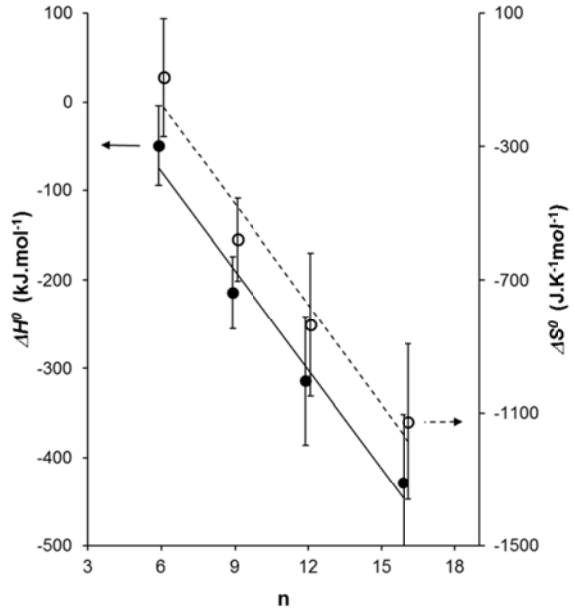



**Figure 4.4**. Average enthalpy and entropy of triple helix formation (expressed per mole of triple helix) as a function of the length (n) of the $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks in telechelic polymers. Closed symbols / left axis: $\Delta H°$, open symbols / right axis: $\Delta S°$. The sample sizes were 2, 10, 2 and 4 for n = 6, 9, 12 and 16, respectively. $\mathbf{T_nP_mT_n}$ as well as $\mathbf{T_nR_mT_n}$ with m = 4 or 8, and also $\mathbf{T_6E_{20}T_6}$ were included, as the middle block appeared to be of little influence on $\Delta H°$ and $\Delta S°$. The protein concentrations varied from 0.2 to 2.4. mM. The sample standard deviation is indicated. To prevent overlapping error bars, the data points were slightly shifted away from their integer x-axis positions. The correlation coefficients for $\Delta H°$ and $\Delta S°$ were 0.99 and 0.98, respectively. For the data of individual samples, see the online supplementary information, Table S4.2.

The $T_m$ is defined as the temperature where the fraction of end blocks taking part in triple helices is 0.5, which is more or less equal to the temperature where the $\Delta C_p$ peak occurs. Triple helix formation of short $(Pro\text{-}Gly\text{-}Pro)_n$ stretches is generally considered to behave approximately like a two-state transition characterized by an equilibrium constant [12].

$$K = \frac{[helix]}{[unfolded\ coil]^3} \quad (4.1)$$

From the definition of $T_m$, it follows that at a temperature $T_m$, the concentration of unfolded coils can be substituted by ½$C_e$, and the triple helix concentration by (½$C_e$)/3 (one helix consists of three chains), where $C_e$ is the end block concentration (equal to twice the polymer concentration). Noting that $\Delta G° = \Delta H° - T\Delta S° = -RT \cdot lnK$, which can then be written as $\Delta H° - T_m\Delta S° = -RT_m \cdot \ln(\frac{C_e}{6}/\frac{C_e^3}{8})$ at $T_m$, the entropy ($\Delta S°$) can be easily calculated from the measured values of the enthalpy ($\Delta H°$) and $T_m$ according to equation (4.2), where R is the gas constant.

$$\Delta S° = \frac{\Delta H°}{Tm} - R \cdot ln({}^{3C_e^2}/_4) \quad (4.2)$$

The resulting $\Delta S°$ values are also shown in Figure 4.4.

It is seen that the absolute $\Delta H°$ and $\Delta S°$ values increase with the length of the end block. The values for n=6 are probably less accurate and subject to an increased risk of underestimation of $\Delta H°$ and $T_m$, because (i) the area under the peak is relatively small (Fig. 4.3), (ii) relatively large concentrations are needed, leading to high viscosities, less accurate pipetting and risk of inclusion of air, and (iii) a relatively long pre-incubation is required at low temperature, possibly because of slow *cis-trans* isomerization. Nevertheless, an apparently linear relationship is obtained between the length of the end block and both the enthalpy and entropy, in good agreement with previous observations on free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides of various lengths [12].

The $T_m$ values of the various triblock polymer samples, which were obtained at various protein concentrations, are plotted as a function of the length of the end blocks in Figure 4.5A (black dots). When these $T_m$ values are compared to

literature data obtained with free (Pro-Gly-Pro)$_n$ peptides at various concentrations (Fig. 4.5A, '+' markers), the triblock polymers seem to have somewhat higher $T_m$ values at first sight. Possibly, this difference can be explained, at least in part, by concentration differences. Rearrangement of equation (4.2) into equation (4.3) below highlights the concentration dependence of $T_m$ at fixed n-value.

$$T_m = \frac{\Delta H^\circ}{(\Delta S^\circ + R \cdot ln\left({}^{3C_e^2}/_{4}\right))} \tag{4.3}$$

Since $\Delta H^\circ$ and $\Delta S^\circ$ were known and are independent of concentration, we could convert the $T_m$ values in Figure 4.5A, each measured at one particular concentration, to values expected for any other concentration. Figure 4.5B shows the values for an arbitrarily chosen concentration of 1 mM triblock polymer, equal to 2 mM end blocks. A similar operation for the free peptides yielded only a few data points, since in several cases the concentrations and/or enthalpy values were not specified in the publications concerned. For clarity, these peptide data points are not shown in Figure 4.5B, since they overlapped with the triblock gelatin data points. Because the enthalpy and entropy are linearly related to the length of the end blocks or free peptides (Fig. 4.4), it is possible to derive the enthalpy and entropy value per interchain hydrogen bond (*i.e.*, per Pro-Gly-Pro tripeptide; the derivation will be addressed below). Using these values per H-bond or tripeptide, which will be further denoted as $\Delta h^\circ$ and $\Delta s^\circ$, one can calculate $\Delta H^\circ$, $\Delta S^\circ$ and also $T_m$ (see Equation 4.3) for any triple helix length of interest, even if experimental $T_m$ data for specific n values are missing. This principle was used to construct the continuous curves in Figure 4.5B. It is seen that the concentration-corrected curves for triblock gelatins and free peptides are very close to each other. Different ways of calculation (see below) resulted in slightly different curves for each of these two

groups and in view of this variability; the difference between triblock copolymers and free peptides is most likely insignificant.

The values of $\Delta h°$ and $\Delta s°$ were derived in three different ways. In the first approach, the values were derived by averaging over the entire set of measurements. For that purpose, and in harmony with the observed linear dependence of $\Delta H°$ on n, the $\Delta H°$ values obtained in each individual measurement were divided by the expected number (3n - 2) [12] of hydrogen bonds, formed between the peptide-NH of glycines in one end block and the peptide-CO of residues C-terminal to glycine in another end block taking part in the same trimeric knot. The $\Delta S°$ values were divided by the number of (Pro-Gly-Pro) tripeptides per triple helix (3n). The $\Delta h°$ and $\Delta s°$ values thus obtained could be averaged over the whole data set. In the second approach, the values were derived by iterative optimization of $\Delta h°$ and $\Delta s°$ such that the differences between the measured $T_m$ values and the values calculated according to equation (4.3) from $\Delta h°$ and $\Delta s°$ were minimized. In the third approach, $\Delta h°$ and $\Delta s°$ were derived from linear least squares fitting of a plot of (3n - 2)/$T_m$ versus 3n - 2 (in analogy of Hanes-Woolf or Klotz plots, which are more familiar in classical enzymology or ligand binding studies, respectively). If $\Delta h°$ and $\Delta s°$ are defined according to $\Delta H°$= (3n - 2)·$\Delta h°$ and $\Delta S°$ = 3n·$\Delta s°$, then equation (4.3) can be rearranged into equation (4.4).

$$\frac{(3\mathrm{n}-2)}{T_m} = (3\mathrm{n}-2)\cdot\left(\frac{\Delta s°}{\Delta h°}\right) + \left(\frac{2\Delta s° + R\cdot ln\left(3C_e^2/4\right)}{\Delta h°}\right) \qquad (4.4)$$

With (3n - 2)/$T_m$ on the Y-axis and 3n - 2 on the X-axis (Fig. 4.6), a slope $\Delta s°/\Delta h°$ is obtained, and an intercept ($2\Delta s°$+R·ln($3C_e^2$/4))/$\Delta h°$), from which both $\Delta h$ and $\Delta s°$ can be easily calculated.

The $\Delta h°$ and $\Delta s°$ for the free $(\text{Pro-Gly-Pro})_n$ peptides were similarly derived using the values of $\Delta H°$, $T_m$ and n available in the literature [12, 16, 17,

30]. All $\Delta h°$ and $\Delta s°$ values thus obtained for triblock gelatins and free peptides are shown in Table 4.1.

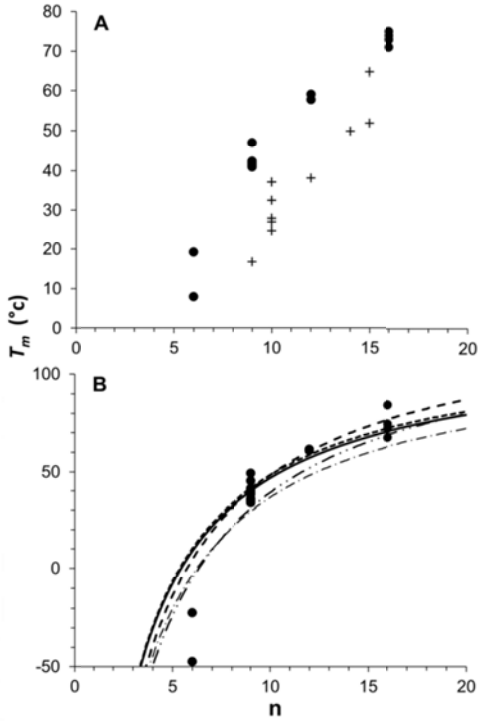



**Figure 4.5.** Melting temperature ($T_m$) as a function of the length, n, of the $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks present in telechelic polymers, in comparison to free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides. (A) Experimental values of triblock gelatins (●) and free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides (+), obtained at concentrations ranging from 0.2 to 2.4 mM. The values for the free peptides were taken from literature [12, 16, 17, 30]. (B) Corresponding values expected for a concentration of 1 mM triblock polymer or 2 mM free $(Pro\text{-}Gly\text{-}Pro)_n$ peptide. The data points (●) were obtained by conversion of the triblock polymer values shown in (A) to the equivalent values expected for 2 mM end blocks (1 mM triblock polymer), using equation (4.3). For the data of individual samples, see the online supplementary information, Table S4.2. The top three continuous curves are calculated using the $\Delta h°$ and $\Delta s°$ values for triblock polymers shown in Table 4.1, based on averaging (dotted curve), iterative error minimization (dashed curve), or linear regression (solid curve). The lower two curves are calculated using the $\Delta h°$ and $\Delta s°$ values for free peptides in Table 4.1, similarly based on

averaging or iterative fitting (indiscernible; both represented by the double dot-dash curve), or on linear regression (single dot-dash curve).

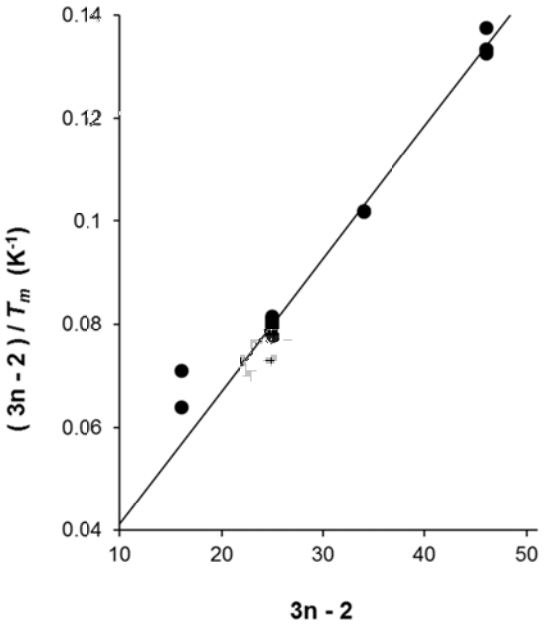



**Figure 4.6.** Plot for derivation of triblock $\Delta h°$ and $\Delta s°$ from the dependence of $T_m$ on end block length (n). The individual experimental $T_m$ values were first converted to those expected for a nominal concentration of 1 mM triblock gelatin, equivalent to 2 mM end blocks, according to the procedure explained in the text. The correlation coefficient after conversion was 1.00. According to equation (4.4), the $\Delta h°$ and $\Delta s°$ values can be calculated from the slope, which was $2.58 \cdot 10^{-3}$ K$^{-1}$, and from the intercept, which was $1.55 \cdot 10^{-2}$ K$^{-1}$. The resulting $\Delta h$ and $\Delta s°$ values are shown in Table 4.1. The data for n = 6 (3n - 2 = 16) were disregarded, for reasons explained in the text in connection to Figure 4.4.

**Table 4.1** $\Delta h°$ and $\Delta s°$ values of triblock gelatins and free peptides derived in three different ways.

| Method of derivation | $\Delta h°$ (kJ.mol$^{-1}$) | $\Delta s°$ (J.K$^{-1}$.mol$^{-1}$) | Correlation coefficient[a] |
|---|---|---|---|
| | | Triblock gelatins**:** | |
| Averaging of $\Delta H°/(3n - 2)$ values | -8.8 (1.6)[b] | -22 (4.5)[b] | 0.95 |
| Iterative error minimization[c] | -10 | -26 | 0.96 |
| Linear regression[d] | -10 | -26 | 0.96 |
| | | Free $(Gly\text{-}Pro\text{-}Pro)_n$ peptides: | |
| Averaging of $\Delta S°/3n$ values | -7.2 (1.0)[b] | -18 (3.1)[b] | 0.88 |
| Iterative error Minimization[c] | -8.4 | -22 | 0.88 |
| Linear regression[e] | -5.6[f] | -13[f] | 0.88 |

[a] Correlation between experimental and calculated $T_m$ values according to equation 4.3, using the $\Delta h°$ and $\Delta s°$ values shown in this table.

[b] Data for n=6 were disregarded. The sample standard deviation is shown in parentheses. With the n=6 data included, the $\Delta h°$ and $\Delta s°$ values were ~ -8.2 ± 2.5 kJ.mol$^{-1}$ and ~ -20 ± 7.4 J.K$^{-1}$.mol$^{-1}$, respectively.

[c] Iterative optimization of $\Delta h°$ and $\Delta s°$ using the Microsoft Excel Solver function, such that the differences between the measured $T_m$ values and the values calculated from $\Delta h°$ and $\Delta s°$ (according to equation 4.3) were minimized. The same values were obtained with and without inclusion of the n=6 data.

[d] According to Figure 4.6, i.e. after concentration normalization of individual $T_m$ values, and disregarding n=6 data. Without normalization, the $\Delta h°$ and $\Delta s°$ values were ~ -12 kJ.mol$^{-1}$ and ~ -31 J.K$^{-1}$.mol$^{-1}$, respectively. With normalization and with n = 6 data included, they were ~ -6.2 kJ.mol$^{-1}$ and ~ -15 J.K$^{-1}$.mol$^{-1}$, respectively.

[e] The method for free peptides was analogous to that shown in Figure 4.6 for triblock gelatins.

[f] These values are less reliable, because the available literature values were obtained at only two different n-values (4 values at n=10 and 1 value at n=15). The fit of the $(3n - 2)/T_m$-versus-$(3n - 2)$ plot had a correlation coefficient of 0.98. For other n-values, the necessary data combination of $\Delta H°$ and concentration was not available and the conversion of all $T_m$ values to those expected for a single nominal concentration could not be performed. Regression with mutually incompatible data obtained at different concentrations, and without such a conversion, yielded a $\Delta h°$ of -6.3 kJ.mol$^{-1}$, and a $\Delta s°$ of -16 J.K$^{-1}$.mol$^{-1}$.

A comparison of the three very different calculation methods gives insight into the stability of the data and thus, their reliability. The first method solely relies on the experimental measurement of $\Delta H°$ for the determination of $\Delta h°$. The second and third methods solely rely on experimental determination of $T_m$ (most obvious from Figure 4.6), but each have a different sensitivity to the spread of the experimental data. Table 4.1 shows that similar $\Delta h°$ and $\Delta s°$ values were obtained using the three different methods. Note that insufficient literature data were available for a reliable linear regression of the $T_m$ values of free $(Gly\text{-}Pro\text{-}Pro)_n$ peptides. The $\Delta h°$ and $\Delta s°$ values of triblock gelatins and free $(Gly\text{-}Pro\text{-}Pro)_n$ peptides were also similar and, in view of the spread of the data and its effect on the calculations, they could not be distinguished. The $\Delta h°$ values of triblock gelatins and free $(Gly\text{-}Pro\text{-}Pro)_n$ peptides are in agreement with the enthalpy of an N-H-O hydrogen bond (~8 $kJ.mol^{-1}$).

**Density and Stress Relaxation of Triple-helical Crosslinks, as Determined by End Block Length**

In previous work, we showed that triblock gelatins with $\mathbf{T_9}$ end blocks form a gel (network) through triple-helical interchain connections that are formed exclusively by the $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks [8, 9, 24]. The temperature-dependent relaxation time of $\mathbf{T_9}$ triple helix dissociation under mechanical stress was also measured [9, 10, 25].

In the present chapter, we investigated the effect of *increased* thermal stability (i.e. increased end block length) on the rheological and relaxation properties. First, the gel-forming behavior of $\mathbf{T_{16}P_4T_{16}}$, followed by dynamic rheology after cooling from 85°C to 20ºC, was compared to that of $\mathbf{T_9P_4T_9}$. Because of the higher thermostability of $\mathbf{T_{16}P_4T_{16}}$, the temperature of measurement is further below the $T_m$ of $\mathbf{T_{16}P_4T_{16}}$ than below the $T_m$ of $\mathbf{T_9P_4T_9}$, under typical conditions. For example, a measurement at 20°C and 1 mM polymer would be ~54°C below the $T_m$ of $\mathbf{T_{16}P_4T_{16}}$ (~74°C), but only 21°C below the $T_m$ of $\mathbf{T_9P_4T_9}$

(~41°C, see Fig. 4.5B above and Table S4.2 of the online supplementary information). In typical measurements, the triple helix association/dissociation equilibrium is therefore shifted more towards the helix form in **$T_{16}P_4T_{16}$** than in **$T_9P_4T_9$**, at equal concentration, and the density of triple-helical crosslinks (and the elastic modulus) will be higher in **$T_{16}P_4T_{16}$** gels than in **$T_9P_4T_9$** gels. This effect is expected to be detectable at measuring temperatures down to 20-30°C below the (concentration-dependent) $T_m$ of **$T_9P_4T_9$**, because the equilibrium percentage of end blocks in triple helix form deviates form 100 % in that region [9].

Figure 4.7 shows the development in time of the storage modulus *G'* and the loss modulus *G"*, which reflect the elastic and viscous behavior, respectively. In the beginning of the gelation process, the viscous properties were predominant but with time, as the network developed, the elastic properties prevailed. A steady-state storage modulus was reached after 10 - 15 hours. To maximize the difference between **$T_9P_4T_9$** and **$T_{16}P_4T_{16}$**, a low concentration (0.7 mM) was used to generate the **$T_{16}P_4T_{16}$** gel. This concentration was previously shown [9] to be too low for **$T_9P_4T_9$** to form a well-detectable gel at room temperature, in view of the lower $T_m$ of **$T_9P_4T_9$**. In Figure 4.7, the time-dependent development of *G'* of 0.7 mM **$T_{16}P_4T_{16}$** is therefore compared to that of 1.2 mM **$T_9P_4T_9$**, at which concentrations both materials yield a similar steady state value of *G'* and thus, a similar density of elastically active chains and triple-helical junctions. Note that each trimolecular helix represents one crosslink, and that the contribution of each helical junction to the elasticity of the network is essentially the same in **$T_9P_4T_9$** and in **$T_{16}P_4T_{16}$** gels, because the helix-forming end blocks are much shorter than the entire triblock copolymer chain. The *kinetics* of gel formation are influenced by several factors including the time to complete the zipper-like folding of the triple helix, the formation of premature network junctions consisting of incompletely folded helices, and the relaxation time of unfolding, which are all significantly different for longer helices. Under the conditions used, these factors are apparently balanced

such that the kinetics of gel formation of **$T_{16}P_4T_{16}$** are similar to, or maybe slightly faster than those of **$T_9P_4T_9$**.

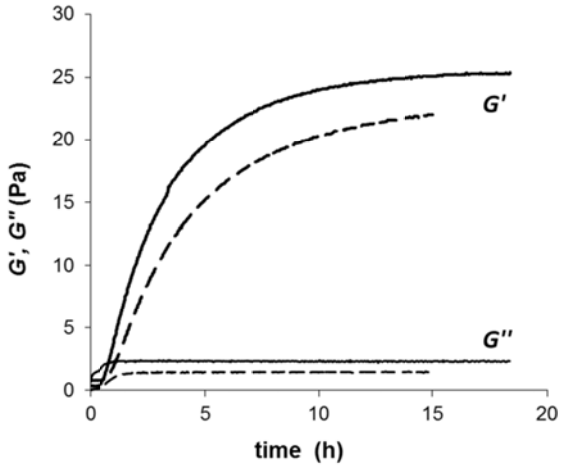



**Figure 4.7.** Time-dependent network (gel) development by 0.7 mM **$T_{16}P_4T_{16}$** (solid lines) and 1.2 mM **$T_9P_4T_9$** (dashed lines) at 20°C, as monitored by the storage modulus *G'* (top two curves) and loss modulus *G*" (lower two curves).

Triple helix dissociation is an important parameter determining the suitability of the gels for various applications [5]. To monitor the relaxation of the network under mechanical stress, we carried out creep tests as described previously for **$T_9R_4T_9$** [9]. After *G'* had reached a plateau value ($G_0$') of 20 – 25 Pa (at 0.7 mM **$T_{16}P_4T_{16}$** and 1.2 mM of **$T_9P_4T_9$**, see Figure 4.7), we compared the creep of the **$T_{16}P_4T_{16}$** and **$T_9P_4T_9$** gels under constant mechanical stress **(**Fig. 4.8). Clearly, **$T_{16}P_4T_{16}$** gels creep much slower than **$T_9P_4T_9$** gels, which show that the relaxation time of **$T_{16}P_4T_{16}$** is indeed much slower than that of **$T_9P_4T_9$**. As previously published [9, 24], the creep does apparently not depend very much on the type of middle block. Although **$T_9P_4T_9$** and **$T_9R_4T_9$** have different hydrodynamic radii and a different tendency to form intramolecular loops, both are characterized by a relaxation time $\tau$ of ~2-3 ks [9, 24], which corresponds to a relaxation time $\tau_0$ of dissociation of individual triple helices in the network of ~10 ks [9]. The values for

**$T_{16}P_4T_{16}$** were more difficult to determine in view of the small slope of the creep curve and the limited amount of material available for rheology.

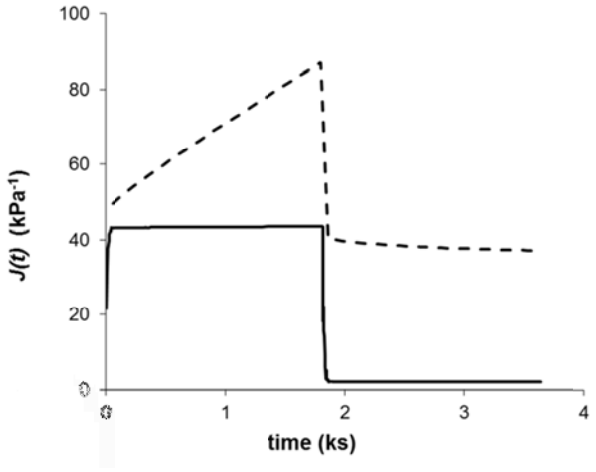



**Figure 4.8.** Creep experiments with **$T_{16}P_4T_{16}$** (solid line) and **$T_9P_4T_9$** (dashed line) gels at 20°C. The concentration of **$T_{16}P_4T_{16}$** and **$T_9P_4T_9$** was 0.7 and 1.2 mM, respectively, and the $G_0$' of both gels was in the range of 20 to 25 Pa at the onset of creep.

## 4.4 Discussion

In earlier work, we produced and characterized various telechelic gelatin-like polymers with equal collagen-like end blocks and equal $T_m$, but different types of middle blocks and different mechanical and slow release properties [5, 8, 24]. We showed that the elastic modulus of the gels can be increased, and the erosion rate can be decreased, by increasing the length and / or the persistence length of the middle block, without any effect on $T_m$ [5, 24]. Presently, we addressed a series of such polymers in which the melting temperature can be tuned in a wide range of possible $T_m$ values by varying the length of the end blocks. It should be noted that in addition to the length, also the amino acid composition and sequence of the end blocks can be changed to tune the helix stability [16]. In first approximation, the $T_m$

appears to be independent of the middle block design. This is illustrated in the present work by the observation that (1) the general trends shown in Figures 4.4 to 4.6 can be obtained with a set of telechelic polymers with different middle blocks, and (2) the concentration- and length-dependent $T_m$ of free $(Gly\text{-}Pro\text{-}Pro)_n$ peptides and $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks in triblock gelatins is similar (see Figure 4.5B). Apparently, the entropic penalty for a $(Pro\text{-}Gly\text{-}Pro)_n$ block to be part of a larger molecule is small. As middle blocks, we presently used (1) ~400 or ~800 amino acid long stretches of hydrophilic collagen-like $(Gly\text{-}X_{aa}\text{-}Y_{aa})_m$ sequences, (2) randomized versions of the aforementioned collagenous sequences with the same amino acid composition and length, yet an ~8 % higher gyration radius than the non-randomized sequences [24], and (3) an ~200 amino acid-long elastin-like sequence that develops hydrophobic interactions and optical turbidity in a temperature region just below the $T_m$ of the gel. Likely, a wide variety of functional properties can be built into the middle blocks without loosing the possibility of independently tuning the $T_m$ and the helix stability under mechanical stress. For obvious reasons, this is subject to the condition that the middle blocks will not give rise to stable intermolecular bonds, especially if these bonds occur close to the endblocks, since that would decrease the $\Delta S°$ of triple helix formation significantly.

Previously, we developed and published a model [9] relating the temperature- and concentration-dependent intra- and intermolecular interactions of $\mathbf{T_9X_mT_9}$ to the experimentally-determined mechanical properties of the gel. In that work, in which **X** was **P** [24] or **R** [9] and m was 4 or 8, very good agreement between experimental data and model predictions was shown for an extensive set of data and conditions [9, 24]. The model reflects the strong influence of temperature on the percentage of end blocks taking part in triple helices, in the region where the difference $T_m$ - T is smaller than ~20 to 30°C. The model further reflects the influence of the polymer concentration on (1) the helix content and (2) the balance between (a) trimolecular helices (network junctions) and (b) bimolecular helices that result from intramolecular loops. The trimolecular junctions increase the

elastic modulus of the gel and the loops disturb the network and lower the modulus. The model now also appears to correctly predict a storage modulus of ~20 Pa for $\mathbf{T_{16}P_4T_{16}}$ at 0.7 mM (Fig. 4.9A) and $\mathbf{T_9P_4T_9}$ at 1.2 mM (Fig. 4.9B), at 20°C (vertical dotted line in Figure 4.9), in agreement with the experimental values shown in Figure 4.7. Similarly in agreement with our observations (data not shown), a very low storage modulus is predicted for polymers with $\mathbf{T_6}$ end blocks (Fig. 4.9A-B, dotted line with diamonds), except at very high polymer concentration (Fig. 4.9C). Based on this model and on the $T_m$ data obtained at present, the elastic properties of the gel can in principle be predicted for any end-block (examples in Figure 4.9) and thus, the gel properties can be tuned to suit particular applications at widely different operating temperatures and concentrations. This can be done without recurring to chemical crosslinking, which circumvents the poor degradability and risk of toxicity in the human body associated with covalently crosslinked gels.

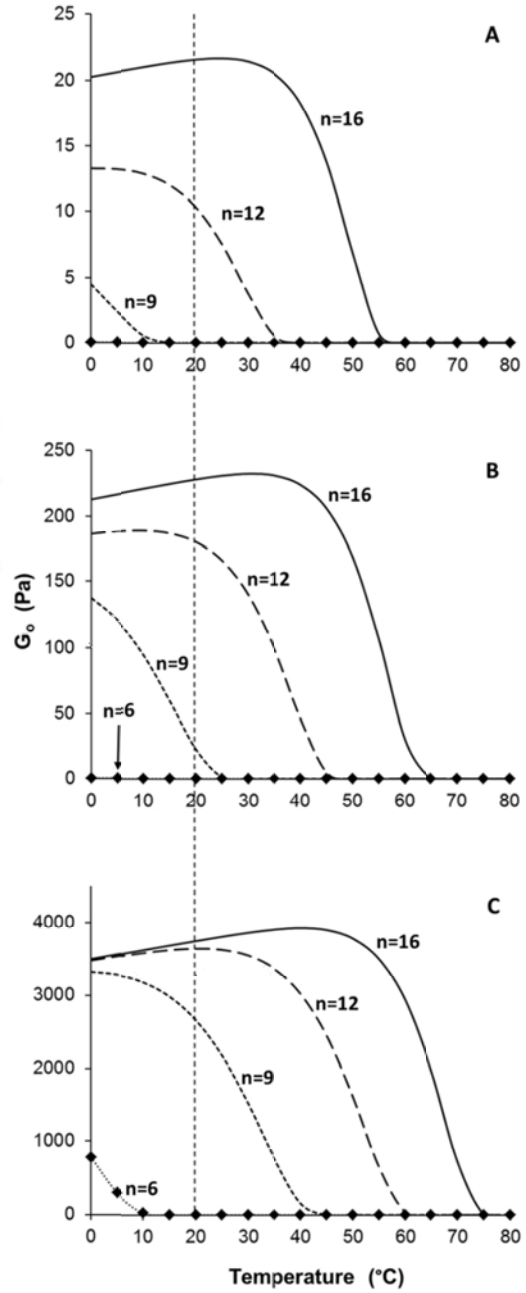



**Figure 4.9.** Temperature-dependent plateau storage modulus as predicted by the model described previously[9] in combination with the $T_m$ values determined at present. The

predictions shown are for polymers with end blocks of length n = 6 to 16 and middle blocks of ~400 amino acids. (A) Triblock concentration 0.7 mM as used in Figure 4.7 for $\mathbf{T_{16}P_4T_{16}}$. (B) triblock concentration 1.2 mM as used in Figure 4.7 for $\mathbf{T_9P_4T_9}$. (C) High triblock concentration (4.8 mM), where also $\mathbf{T_6R_4T_6}$ is predicted to form a gel below 10°C. Experimentally, we have indeed observed $\mathbf{T_6R_4T_6}$ gel formation at high concentration and ~4 - 5°C. The dotted line marked with ♦ refers to n=6, the dotted line to n=9, the dashed line to n=12 and the solid line to n=16. Values of -8 kJ.mol$^{-1}$ and -20 J.K$^{-1}$.mol$^{-1}$ were used for $\Delta h°$ and $\Delta s°$, respectively. Critical overlap concentrations of 1.13, 1.05, 0.980 and 0.897 mM were used for polymers with $\mathbf{T_6}$, $\mathbf{T_9}$, $\mathbf{T_{12}}$ and $\mathbf{T_{16}}$ end blocks, respectively, based on their respective radii of gyration [31].

The modular (block copolymeric) genetic design of these collagen-inspired molecules enables easy and independent adaptation of the individual modules or blocks. Also the number of blocks can be easily increased, to expand the range of functional properties and/or facilitate the combination of certain properties such as higher stiffness and lower $T_m$. Finally, bioactive and / or biointeractive domains could be introduced into the middle block independently from changes in $T_m$. The construction of new block copolymers from modules with a record of high-yield production in a given production host enhances the chance that these copolymers can be successfully produced. The modular approach thus maximizes the span of functional properties, while minimizing the effort of design and risk of production failure.

Our collagen-inspired materials expectedly offer excellent biocompatibility. The $\mathbf{P_m}$ blocks were previously shown to be biocompatible in several *in vivo* and other tests [32]. Furthermore, the current middle blocks could be replaced with, for example, human-identical collagen sequences. These also will keep a random coil conformation down to very low temperatures, if they remain unhydroxylated as in our yeast production system [27]. Finally, $(\text{Pro-Gly-Pro})_n$ sequences naturally occur in human and animal collagen, where they serve to nucleate the zipper-like folding of the triple helix. Previously, we showed that $\mathbf{T_9R_nT_9}$ offers erosion rates compatible with short- to medium-term drug delivery applications (i.e., spanning a few days) [5]. Our present results show that the elongation of the $(\text{Pro-Gly-Pro})_9$

end-blocks from n=9 to n=16 resulted in a much higher relaxation time of dissociation. This principle can be used to engineer gels with a lower rate of erosion and a longer lifetime in the body.

The flexible, versatile design and the tunable triple helix stability of these non-animal gelatin-like polymers will expectedly speed up developments towards application in the biomedical field. The high levels of production of these proteins in yeast [8], and the possibilities for future scale-up to industrial production can only facilitate this.

## 4.5 Conclusion

Recombinant telechelic polymers with various middle blocks and collagen-like triple helix-forming (Pro-Gly-Pro)n end blocks were produced as secreted proteins in the yeast Pichia pastoris. The thermomechanical stability of the triple helices could be tuned by varying the length n of the $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks through genetic engineering. By varying n from 6 to 16, the $T_m$ could be varied in a range of about 70°C, at concentrations around 1 mM, without having to recur to chemical (covalent) crosslinking. The variation of the thermostability can be understood in terms of a simple linear relationship between the length of the end block and the free energy of helix formation, or its constituting entropic and enthalpic components. This relationship appeared to be quantitatively comparable for the end blocks of the triblock gelatins and for free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides, and it was also not influenced by the nature of the investigated middle blocks.

All triblock copolymers studied formed stable hydrogels, and creep experiments showed that an increase of the end block length resulted in a higher stability under mechanical stress. Also in view of their biocompatibility, these materials can probably be tuned to be profitably applied in a variety of medical applications.

**Associated content**

**Supporting Information.** The DNA sequences of the triple-helix forming end blocks used in this study can be found in Table S4.1, and the data underlying Figures 4.4-4.6 in table S4.2

## Acknowledgement

The authors would like to thank Dr. J. van der Gucht for helpful discussions. This work 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] Rabotyagova, O. S., Cebe, P., Kaplan, D. L., *Biomacromolecules* 2011, *12*, 269-289.
[2] Li, F., de Wolf, F. A., Marcelis, A. T., Sudholter, E. J., Cohen Stuart, M. A., Leermakers, F. A., *Angew Chem Int Ed Engl* 2010, *49*, 9947-9950.
[3] Martens, A. A., Portale, G., Werten, M. W. T., de Vries, R. J., Eggink, G., Stuart, M. A. C., de Wolf, F. A., *Macromolecules* 2009, *42*, 1002-1009.
[4] Shen, W., Zhang, K., Kornfield, J. A., Tirrell, D. A., *Nat Mater* 2006, *5*, 153-158.
[5] Teles, H., Vermonden, T., Eggink, G., Hennink, W. E., de Wolf, F. A., *J Control Release* 2010, *147*, 298-303.
[6] Xu, C., Breedveld, V., Kopecek, J., *Biomacromolecules* 2005, *6*, 1739-1749.
[7] Wu, X. Y., Sallach, R., Haller, C. A., Caves, J. A., Nagapudi, K., Conticello, V. P., Levenston, M. E., Chaikof, E. L., *Biomacromolecules* 2005, *6*, 3037-3044.
[8] Werten, M. W. T., Teles, H., Moers, A. P. H. A., Wolbert, E. J. H., Sprakel, J., Eggink, G., de Wolf, F. A., *Biomacromolecules* 2009, *10*, 1106-1113.
[9] Skrzeszewska, P. J., de Wolf, F. A., Werten, M. W. T., Moers, A. P. H. A., Stuart, M. A. C., Gucht, J. v. d., *Soft Matter* 2009, *5*, 2057-2062.
[10] Skrzeszewska, P. J., Sprakel, J., de Wolf, F. A., Fokkink, R., Stuart, M. A. C., van der Gucht, J., *Macromolecules* 2010, *43*, 3542-3548.
[11] Stahl, P. J., Romano, N. H., Wirtz, D., Yu, S. M., *Biomacromolecules* 2010, *11*, 2336-2344
[12] Engel, J., Chen, H. T., Prockop, D. J., Klump, H., *Biopolymers* 1977, *16*, 601-622.
[13] Josse, J., Harrington, W. F., *J Mol Biol* 1964, *9*, 269-287.
[14] Mizuno, K., Hayashi, T., Peyton, D. H., Bachinger, H. P., *J Biol Chem* 2004, *279*, 38072-38078.
[15] Vitagliano, L., Berisio, R., Mazzarella, L., Zagari, A., *Biopolymers* 2001, *58*, 459-464.
[16] Persikov, A. V., Ramshaw, J. A., Brodsky, B., *J Biol Chem* 2005, *280*, 19343-19349.
[17] Sutoh, K., Noda, H., *Biopolymers* 1974, *13*, 2391-2404.
[18] Engel, J., Kurtz, J., Katchals.E, Berger, A., *J. Mol. Biol.* 1966, *17*, 255-272.
[19] Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bachinger, H. P., Engel, J., *J. Mol. Biol.* 2001, *308*, 1081-1089.
[20] Harrington, W. F., Rao, N. V., *Biochemistry* 1970, *9*, 3714-3724.

[21] Busnel, J. P., Morris, E. R., Rossmurphy, S. B., *Int J Biol Macromol* 1989, *11*, 119-125.
[22] Djabourov, M., in: Aalsbersberg, W. I., Hamer, R. J., Jasperse, P., Jongh, H. H. J. d., Kruif, C. G. d., Walstra, P., Wolf, F. A. d. (Eds.), *Progress in Biotechnology*, Elsevier Science, Amsterdam 2003, pp. 133-218.
[23] Nijenhuis, K. T., *Colloid Polym Sci* 1981, *259*, 1017-1026.
[24] Teles, H., Skrzeszewska, P. J., Werten, M. W. T., van der Gucht, J., Eggink, G., de Wolf, F. A., *Soft Matter* 2010, *6*, 4681-4687.
[25] Skrzeszewska, P. J., de Wolf, F. A., Stuart, M. A. C., van der Gucht, J., *Soft Matter* 2010, *6*, 416-422.
[26] Schipperus, R., Teeuwen, R. L., Werten, M. W. T., Eggink, G., de Wolf, F. A., *Appl Microbiol Biotechnol* 2009, *85*, 293-301.
[27] Werten, M. W. T., van den Bosch, T. J., Wind, R. D., Mooibroek, H., de Wolf, F. A., *Yeast* 1999, *15*, 1087-1096.
[28] Werten, M. W. T., Wisselink, W. H., Jansen-van den Bosch, T. J., de Bruin, E. C., de Wolf, F. A., *Protein Eng* 2001, *14*, 447-454.
[29] Schipperus, R., Eggink, G., de Wolf, F. A., *Biotechnology Progress* 2011, 242-247.
[30] Persikov, A. V., Xu, Y. J., Brodsky, B., *Protein Sci* 2004, *13*, 893-902.
[31] Fitzkee, N. C., Rose, G. D., *Proc Natl. Acad. Sci USA* 2004 *101*, 12497-12502.
[32] *US Patent 20050119170,* 2005.

## Support Information

**Table S.4.1** DNA sequences of the triple-helix forming end blocks.

| Gene | Sequence |
|---|---|
| $T_6$[a] | CACCCGGTGCTCCCGGTCCGCCTGGTCCACCTGGTCCACCTGGTCCAC<br>CGGGTCCTCCCG<br>GACCAGCCGGTGG |
| $T_{12}$ | CACCCGGTGCTCCCGGTCCGCCTGGTCCACCTGGTCCACCTGGTCCAC<br>CAGGACCACCAG<br>GTCCACCTGGACCACCGGGTCCGCCTGGTCCACCTGGTCCACCGGGTC<br>CTCCCGGACCAG<br>CCGGTGG |
| $T_6$[b] | CTCGAGAAAAGAGAGGCTGAAGCTGGGTCTCCTGGTCCACCTGGTCC<br>ACCTGGTCCACCA<br>GGACCACCAGGTCCACCTGGACCACCTGGTTGAGACGGAATTC |

[a] end-block used together with $\mathbf{R_4}$ middle block

[b] end-block used together with $\mathbf{E_{20}}$ middle block

**Table S 4.2** Thermodynamical parameters of individual samples.

| n | mid block | $T_m$ | $\Delta H°$ (kJ.K$^{-1}$.mol$^{-1}$) | $\Delta S°$ (kJ.K$^{-1}$.mol$^{-1}$) | [triblock] (mM) | [(PGP)$_n$] (mM) | $T_m$ at 1 mM triblock or 2 mM (PGP)$_n$ | Average-based[a] $T_m$ at 1 mM triblock or 2 mM (PGP)$_n$ |
|---|---|---|---|---|---|---|---|---|
| 6 | $E_{20}$ | 20 | -81 | -218 | 16.57 | 33.15 | -23 | 4.9 |
| 6 | $R_4$ | 8 | -17 | 30 | 2.44 | 4.87 | -47 | 4.9 |
| 9 | $R_4$ | 41 | -175 | -465 | 2.19 | 4.38 | 34 | 40.2 |
| 9 | $R_4$ | 42 | -192 | -505 | 1.10 | 2.21 | 41 | 40.2 |
| 9 | $R_4$ | 41 | -319 | -896 | 0.43 | 0.86 | 45 | 40.2 |
| 9 | $R_4$ | 41 | -192 | -511 | 1.40 | 2.80 | 38 | 40.2 |
| 9 | $R_4$ | 47 | -192 | -509 | 2.40 | 4.80 | 39 | 40.2 |
| 9 | $R_8$ | 42 | -209 | -573 | 2.40 | 4.80 | 35 | 40.2 |
| 9 | $P_4$ | 42 | -204 | -557 | 2.25 | 4.49 | 35 | 40.2 |
| 9 | $P_4$ | 42 | -224 | -608 | 1.12 | 2.25 | 41 | 40.2 |
| 9 | $P_4$ | 42 | -204 | -526 | 0.45 | 0.90 | 49 | 40.2 |
| 9 | $P_8$ | 42 | -234 | -651 | 2.40 | 4.80 | 36 | 40.2 |
| 12 | $R_4$ | 58 | -263 | -682 | 0.68 | 1.36 | 61 | 60.2 |
| 12 | $R_4$ | 59 | -365 | -988 | 0.68 | 1.36 | 61 | 60.2 |
| 16 | $R_4$ | 73 | -418 | -1107 | 1.29 | 2.58 | 72 | 76.4 |
| 16 | $R_4$ | 73 | -441 | -1173 | 1.29 | 2.58 | 72 | 76.4 |
| 16 | $P_4$ | 75 | -522 | -1399 | 1.29 | 2.58 | 74 | 76.4 |
| 16 | $P_4$[b] | 74 | -334 | -830 | 0.20 | 0.40 | 84 | 76.4 |
| Free (pro-Gly-Pro)$_n$ peptides: | | | | | | | | |
| 5[c] | - | n.p.[d] | -282 | n.p.[d] | - | n.p.[d] | - | -23.8 |
| 9[e] | - | 17 | n.p.[d] | n.p.[d] | - | n.p.[d] | - | 31.3 |
| 10[f] | - | 28 | -156 | -386 | - | 0.40 | 45 | 30.3 |
| 10[f] | - | 32 | -179 | -453 | - | 0.40 | 47 | 39.3 |
| 10[g] | - | 27 | n.p.[d] | n.p.[d] | - | 1.19 | - | 39.3 |
| 10[c] | - | n.p.[d] | -238 | -648 | - | n.p.[d] | 44 | 39.3 |
| 10[c] | - | 25 | -215 | -630 | - | 1.00 | 20 | 39.3 |
| 10[c] | - | 37 | -215 | -583 | - | 1.40 | 40 | 39.3 |
| 10[c] | - | 37 | -199 | -528 | - | 1.30 | 41 | 39.3 |
| 12[g] | - | 38 | n.p.[d] | n.p.[d] | - | 0.99 | - | 51.9 |
| 14[g] | - | 50 | n.p.[d] | n.p.[d] | - | 0.85 | - | 61.3 |
| 15[g] | - | 52 | n.p.[d] | n.p.[d] | - | 0.79 | - | 65.2 |
| 15[c] | - | 65 | -314 | -810 | - | 0.92 | 70 | 65.2 |

[a] Calculated from the average $\Delta h°$ and $\Delta s°$ values in Table 4.1.

[b] From protease knockout.

[c] from literature reference 12.

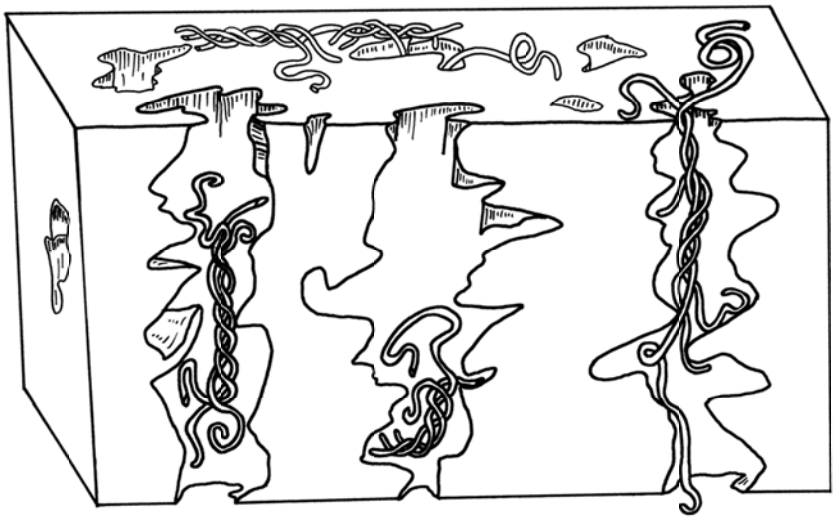
n.p., value not provided by the literature source.

[e] from literature reference 16.
[f] from literature reference 31.
[g] from literature reference 17.

# Chapter 5

# Anomalous behavior of recombinant triple-helix forming proteins in SDS-PAGE

Catarina I.F. Silva, Gerrit Eggink and Frits A. de Wolf, submitted



## Abstract

The focus of this work was on SDS-PAGE of collagen-inspired proteins (CIP). The present CIP were produced in yeast as expression products of synthetic recombinant genes. CIP with equal amino acid composition, but different sequence, migrated at different rates in SDS-PAGE. During destaining, CIP capable of forming triple helical crosslinks in solution, diffused out of the gel less easily than variants incapable of triple helix formation. Thus, SDS-PAGE displays the molecular conformation of these proteins as well as protein-protein interactions of recombinant CIP.

## 5.1 Introduction

In polyacrylamide gel electrophoresis (PAGE) of proteins, a polyacrylamide gel matrix is used to separate mixtures of proteins (e.g., from cells, subcellular fractions, column fractions, or immunoprecipitates) according to their electrophoretic mobility. Electrophoretic mobility is a function of the length and charge of a polypeptide chain. The migration rate is therefore determined by the combination of the pore size of the PA gel, and the charge, size, and shape of the protein. In SDS-PAGE, sodium dodecyl sulphate (SDS) is added to disrupt inter- and intra-molecular bonds so that the migration rate is determined by the molecular weight/charge ratio. SDS causes the unfolding of the proteins by binding to hydrophobic protein sites. This results in a "reconstructive denaturation" where, proteins adopt a conformational mixture of α-helix and random coil [1]. In most proteins, the binding of SDS to the polypeptide chain imparts a fixed number of charges per unit mass of the protein. And as a result, the migration and separation of the proteins during electrophoresis is approximately according to size.

Contrary to most proteins, fibrillar proteins such as silk, elastin and collagen are rich in light amino acids with small side groups, such as glycine and alanine. The low average mass per amino acid residue of these proteins implies that their chain length and number of residues is higher than that of globular proteins of the same molecular mass [2, 3]. Consequently, collagens migrate in SDS-PAGE at an apparent molecular weight ($Mw_{ap}$, as calibrated with globular Mw marker proteins) approximately 1.4 times higher than the true Mw [4].

In addition to that, we have previously shown that collagen-like proteins (CIP) with an unnatural (designed), very hydrophilic amino acid composition show an even greater aberrant migration behaviour [5]. It was shown that this additional effect was due specifically to the highly hydrophilic nature of these CIP [6] and

probably, to poor binding of charge-providing SDS molecules since the interaction of SDS with proteins is mainly of a hydrophobic nature [7].

The ~ 37 kDa recombinant protein previously denoted as **$P_4$** is such a highly hydrophilic protein based on a synthetic (designed) gene [5]. Like natural gelatins, which consist of denatured and partly degraded natural collagen, **$P_4$** consists of a repetition of (Gly-$X_{aa}$-$Y_{aa}$) triplets, but in contrast to natural gelatine, it contains no hydrophobic amino acids other than proline, and it is specifically rich in asparagine and glutamine [5]. **$R_4$** has the same amino acid composition as **$P_4$**, but its protein sequence is scrambled so as to avoid glycine as every third residue [8, 9]. Because they both have the same amino acid composition and therefore the same Mw and length, it was expected that both proteins would migrate at the same velocity in SDS-PAGE. However, as described in the following, we discovered that this was not the case.

Gel boxes and 10% NuPAGE Bis-Tris gels were purchased from Invitrogen. Gels were run at 200 V for 50 minutes under reducing conditions in MES running buffer, pH 7, supplemented with antioxidant, following the manufacturer's protocol. To avoid interference with the staining process and remove SDS, the gels were washed 3 times with Milli Q (MQ) water for 5 minutes. Proteins visualization was achieved by staining the gels for 1 hour with the non-standard, water-based SimplyBlue™ SafeStain, and SeeBlue® Plus2 Pre-Stained Standard was used as a molecular marker. To allow a better visualization of the gel bands and remove non-specific associations the gels were destained for 1 hour with MQ water.

As expected, both proteins migrated poorly in the acrylamide gel network. However, the protein with the randomized sequence (**$R_4$**) migrated faster than **$P_4$** (Fig. 5.1). Fitzkee *et al*. have shown that even a polypeptide chain that assumes random coil conformation may contain small secondary structure elements that contribute to the overall flexibility of the chain [10].

Kubo *et al*. observed that a collagen alpha 2 chain binds a higher number of SDS molecules than an alpha 1 chain and consequently migrates faster in a polyacrylamide gel [11], in spite of the almost identical length of both chains.

It is not known whether **$R_4$** and **$P_4$** bind different amounts of SDS and the different behavior of the proteins in SDS-PAGE gels could be a consequence of i) local secondary structure elements, ii) different SDS binding or iii) a combination of both. **$R_4$** contains longer stretches of contiguous proline residues than **$P_4$**, which can possibly result in increased SDS binding and faster migration.

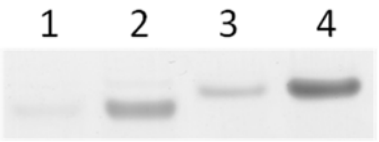



**Figure 5.1.** SDS-PAGE of purified **$R_4$** and **$P_4$** proteins. Lane 1, 2 μg **$R_4$**, Lane 2, 20 μg **$R_4$**, Lane 3, 2 μg **$P_4$** and Lane 4, 20 μg **$P_4$**. A 1-hour destaining was applied.

In previous work [5, 8] and in Chapters 2 and 4, we have investigated chimeric proteins consisting of **$P_4$** or **$R_4$** fused to N- and C-terminal $(Pro\text{-}Gly\text{-}Pro)_n$ domains capable of forming collagen-like triple helices (trimers) with neighboring molecules. The trimer-forming capability of these $(Pro\text{-}Gly\text{-}Pro)_n$ domains, denoted "**$T_n$**", confers network-forming (i.e. gel-forming) properties to the fusion proteins. In Chapter 4 a series of such **$T_nP_4T_n$** and **$T_nR_4T_n$** fusions was created with n=6, 9, 12, and 16. All of these showed the expected anomolous migration behavior, resulting from the high polarity of the **$P_4$** or **$R_4$** middle block [5] and the high content of small amino acid residues [9]. The identity and intactness of each protein was confirmed by N-terminal sequencing and MALDI-TOF mass spectrometry ([9] and Chapter 2).

When compairing **$T_nP_4T_n$** and **$T_nR_4T_n$** with **$P_4$** and **$R_4$**, it appeared that **$P_4$** and **$R_4$** have a tendency to diffuse out of the gel during destaining. Normally, the destaining step for protein visualization lasted 1 to 1.5 hours. However, it was

observed that after overnight (ON) destaining, **$P_4$** and **$R_4$** proteins were no longer visible in the gel (compare Figure 2 and 3, lanes 7-10; Table 5.1). A new staining step did not result in the reappearance of the protein bands (data not shown). Remarkably, the attachment of the trimer-forming **$T_n$** end-blocks resulted in retardation (n=6) or complete abolishment (n=9 or larger) of the protein diffusion out of the gel during destaining (compare Fig. 5.2 and 5.3, lanes 1, 2, 5, 6; Table 5.1).

We have previously shown that the thermostability of the triple helices is determined by the end-block length [14]. A higher number of (Pro-Gly-Pro) triplets results in a higher number of hydrogen bonds established between neighboring $(Pro\text{-}Gly\text{-}Pro)_n$ chains, and therefore, a higher (concentration-dependent) melting temperature ($T_m$) [12] (Table 5.1). For example, **$T_9$** and **$T_{16}$** end blocks have a melting temperature of ~ 41 and 74 °C, respectively, at a concentration of 1.2 mM.

To investigate a possible correlation between $T_m$ and the stability of the protein in the PAGE gel during destaining, a gel with samples containing either **$T_9P_4T_9$** or **$T_{16}P_4T_{16}$** was destained overnight and subsequently heated in a MQ water bath at 37°C for 1 hour. It was observed that in such conditions, **$T_{16}P_4T_{16}$** ($T_m$ ~ 74 °C) stays longer in the gel than **$T_9P_4T_9$** ($T_m$ ~ 41 °C) (compare, in Fig. 5.2-4, lanes 1 & 2, and lanes 5 & 6). In a general sense, it appeared that the proteins with trimer-forming end blocks migrate out of the gel only when heated and apparently, the diffusion rates are directly correlated with the $T_m$ and size of the end-blocks. Without heating, all trimer-forming proteins were retained in the gel, as shown above (Fig. 5.3, lanes 1, 2, 5, 6; Table 5.1).

These observations indicate that the presence of trimer-forming end-blocks inhibited the diffusion out of the gel. To confirm that the retention was indeed due to the formation of $[(Pro\text{-}Gly\text{-}Pro)]_3$ trimers (possibly leading to protein networks in the SDS-PAGE gel), two additional tests were performed. As a first test, the effect of 8M urea, which should inhibit triple helix formation, was investigated. It was indeed observed that the presence of urea abolished the retention of proteins

with trimer-forming capability during overnight (ON) destaining. Re-staining did not reinstall the visibility of the protein bands. As second test, the behavior of mutant versions of the **$T_9$** and **$T_{16}$** end-blocks, mentioned in Chapter 2 and denoted as "**$S_9$** and **$S_{16}$**", respectively, was investigated (Table 5.1). **$S_9$** and **$S_{16}$** have exactly the same length and amino acid composition as **$T_9$** and **$T_{16}$**, but have a randomized sequence in which the repetitive (Pro-Gly-Pro) motive is no longer present (i.e. Gly does no longer occur in the third position after a preceding Gly). Therefore, triple helix formation can no longer take place in **$S_9$** and **$S_{16}$** (Fig. 5.5).

**Table 5.1.** Triple helix stability in relation to disappearance of CIP's from SDS-PAGE gels under various destaining conditions.

| Protein | End-blocks | $T_m$ (ºC) | Ability to form a gel | Condition required for disappearance |
|---|---|---|---|---|
| $R_4$ | - | - [a] | no | 1 – 1.5 hour |
| $P_4$ | - | - [a] | no | 1 – 1.5 hour |
| $T_6R_4T_6$ | $(PGP)_6$ | 10 | yes | ON [b] |
| $T_9P_4T_9$ | $(PGP)_9$ | 41 | yes | heating |
| $T_{12}P_4T_{12}$ | $(PGP)_{12}$ | 59 | yes | heating |
| $T_{16}P_4T_{16}$ | $(PGP)_{16}$ | 74 | yes | no disappearance |
| $S_9P_4S_9$ | $scr(PGP)_9$ [c] | - [a] | no | 1 – 1.5 hour |
| $S_{16}P_4S_{16}$ | $scr(PGP)_{16}$ [c] | - [a] | no | 1 – 1.5 hour |

[a] these proteins do not have trimer-forming blocks and are not able to form gels. Hence, there is no $T_m$.

[b] ON, overnight (not shown in the figures).

[c] with scrambled sequence, so as to avoid Gly in the third position after the previous Gly and obtain incompatibility with triple helix formation.

It was observed that proteins with non-trimer-forming **S** end-blocks are no longer detectably present in the SDS-PAGE gel after ON destaining (compare Fig. 2 and 3, lanes 3,4, and compare Fig. 5.6A and B). The detectability of the proteins in the gel after ON destaining thus appeared to correlate with the trimer- and network-forming capability of the molecules in solution. To know whether the non-

detected proteins were diffusing out of the gel or just losing their stain (Fig. 5.6A and B), the ON destaining water was collected and freeze dried, and the dried material was loaded in concentrated form on a new gel, which was stained and destained during 1 hour as previously mentioned. The protein bands were again observed (Fig. 5.6C), showing that these proteins had originally indeed diffused out of the gel during ON destaining and were now recovered again from the destaining water. Only the incubation of the gel for 45 minutes with 5% glutaraldehyde in 0.2 M phosphate buffer (pH 7.5) prevented the diffusion of the non-triple helix-forming proteins out of the gel (data not shown).

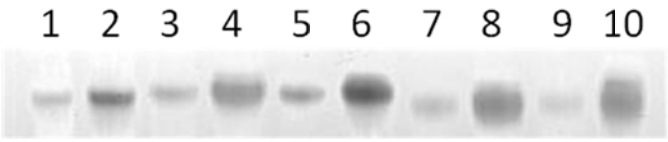



**Figure 5.2.** SDS-PAGE gel of purified recombinant proteins after 1.5 hour destaining. Lane 1, 5 µg **$T_{16}P_4T_{16}$**, Lane 2, 20 µg **$T_{16}P_4T_{16}$**, Lane 3, 5 µg **$S_{16}P_4S_{16}$**, Lane 4, 20 µg **$S_{16}P_4S_{16}$**, Lane 5, 5 µg **$T_9P_4T_9$**, Lane 6, 20 µg **$T_9P_4T_9$**, Lane 7, 5 µg **$R_4$**, Lane 8, 20 µg **$R_4$**, Lane 9, 5 µg **$P_4$** and Lane 10, 20 µg **$P_4$**.

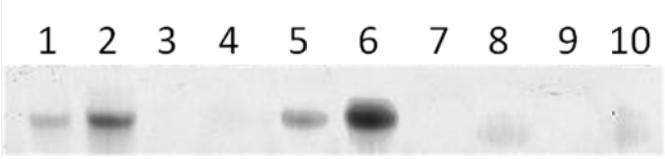



**Figure 5.3.** SDS-PAGE gel of purified recombinant proteins after 24 hour destaining. Lane 1, 5 µg **$T_{16}P_4T_{16}$**, Lane 2, 20 µg **$T_{16}P_4T_{16}$**, Lane 3, 5 µg **$S_{16}P_4S_{16}$**, Lane 4, 20 µg **$S_{16}P_4S_{16}$**, Lane 5, 5 µg **$T_9P_4T_9$**, Lane 6, 20 µg **$T_9P_4T_9$**, Lane 7, 5 µg **$R_4$**, Lane 8, 20 µg **$R_4$**, Lane 9, 5 µg **$P_4$** and Lane 10, 20 µg $P_4$.

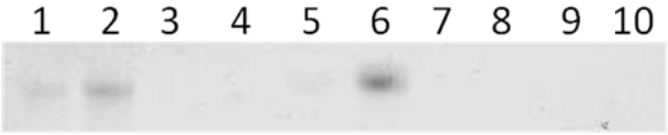



**Figure 5.4.** SDS-PAGE gel of purified recombinant proteins after 24 hour destaining and 1 hour in a water bath at 37 °C. Lane 1, 5 µg **$T_{16}P_4T_{16}$**, Lane 2, 20 µg **$T_{16}P_4T_{16}$**, Lane 3, 5 µg

**$S_{16}P_4S_{16}$**, Lane 4, 20 µg $S_{16}P_4S_{16}$, Lane 5, 5 µg **$T_9P_4T_9$**, Lane 6, 20 µg **$T_9P_4T_9$**, Lane 7, 5 µg **$R_4$**, Lane 8, 20 µg **$R_4$**, Lane 9, 5 µg **$P_4$** and Lane 10, 20 µg **$P_4$**.

.

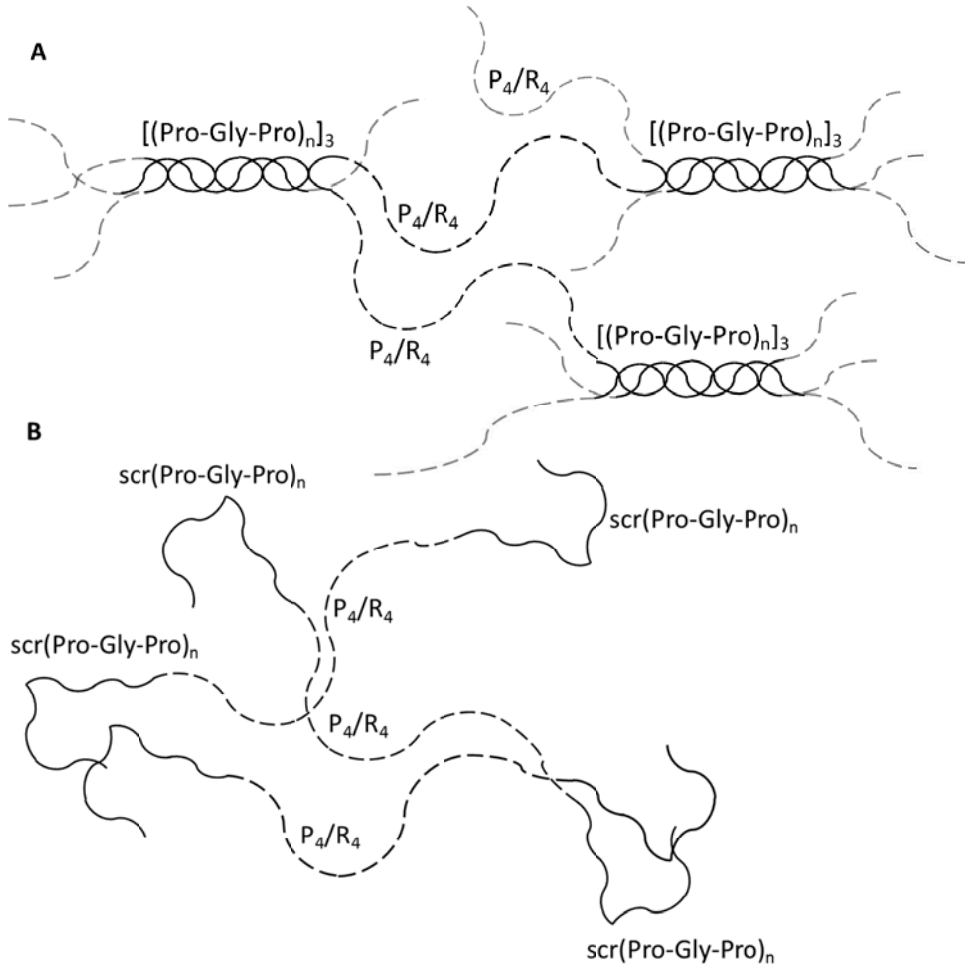



**Figure 5.5.** Schematic representation of protein interactions. A) Triblock copolymers with $(Pro\text{-}Gly\text{-}Pro)_n$ end blocks can self-assemble into networks through formation of $[(Pro\text{-}Gly\text{-}Pro)_n]_3$ triple helices by the end blocks. B) Triblocks with non-triple helix-forming end blocks consisting of a scrambled sequence of 33 % Gly and 67 % Pro cannot form networks. Full lines (—) are intended to represent end-blocks and dashed lines (- - -) the **$P_4$** or **$R_4$** middle blocks.

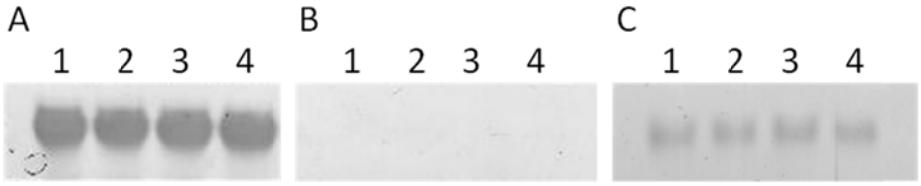



**Figure 5.6.** SDS-PAGE of purified $\mathbf{S_{16}P_4S_{16}}$ recombinant protein. A) Lane 1-4, purified 20 μg $\mathbf{S_{16}P_4S_{16}}$ after 1 hour destaining. B) Lane 1-4, 20 μg $\mathbf{S_{16}P_4S_{16}}$ after 20 hours destaining. C) Lane 1-4, 10 μL of destaining water from gel B, concentrated by freeze drying and subsequently dissolving in 1 mL. Gel C was destained for 1 hour.

Together, these results show that self-assembly of triple helices and/or (resulting from that) formation of supramolecular protein networks within the acrylamide gel during destaining is responsible for the retention of the proteins in the gel. Triple helix formation occurs due to the presence of non-covalent hydrogen bonds and is therefore reversible in nature. Although the proteins are heated until 75°C, before being loaded in the gel, triple helices can again be formed once the temperature is below their $T_m$. Furthermore, during the process of staining and destaining the SDS is washed out and in situ formation of triple helix and networks could occur.

In conclusion, our results show that denaturing SDS-PAGE can monitor the ability of CIP (and possibly also other proteins, especially of the fibrillar type) to form supramolecular assemblies. Furthermore, subtle differences in migration speed can reveal changes in the amino acid sequence of random coils.

## Acknowledgement

This work 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] Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., Deber, C. M., *P Natl Acad Sci USA* 2009, *106*, 1760-1765.
[2] Freytag, J. W., Noelken, M. E., Hudson, B. G., *Biochemistry* 1979, *18*, 4761-4768.
[3] Noelken, M. E., Wisdom, B. J., Jr., Hudson, B. G., *Anal Biochem* 1981, *110*, 131-136.
[4] Butkowski, R. J., Noelken, M. E., Hudson, B. G., *Method Enzymol* 1982, *82*, 410-423.
[5] Werten, M. W. T., Wisselink, W. H., Jansen-van den Bosch, T. J., de Bruin, E. C., de Wolf, F. A., *Protein Engineering* 2001, *14*, 447-454.
[6] Werten, M. W. T., van den Bosch, T. J., Wind, R. D., Mooibroek, H., de Wolf, F. A., *Yeast* 1999, *15*, 1087-1096.
[7] Reynolds, J. A., Tanford, C., *P Natl Acad Sci USA* 1970, *66*, 1002-&.
[8] Skrzeszewska, P. J., de Wolf, F. A., Werten, M. W. T., Moers, A. P. H. A., Stuart, M. A. C., Gucht, J. v. d., *Soft Matter* 2009, *5*, 2057-2062.
[9] Werten, M. W. T., Teles, H., Moers, A. P. H. A., Wolbert, E. J. H., Sprakel, J., Eggink, G., de Wolf, F. A., *Biomacromolecules* 2009, *10*, 1106-1113.
[10] Fitzkee, N. C., Rose, G. D., *Proc Natl Acad Sci U S A* 2004, *101*, 12497-12502.
[11] Kubo, K., Takagi, T., *Collagen Rel Res* 1984, *4*, 201-208.
[12] Engel, J., Kurtz, J., Katchals.E, Berger, A., *J Mol Biol* 1966, *17*, 255-272.

# Chapter *6*

## General discussion

*Divide et impera*

*Latin expression meaning to Divide and Conquer*

Collagen and gelatine array of applications is quite extensive, ranging from gelling agents in food, emulsifiers in photographic films, fillers in cosmetics and structural networks in drug delivery and tissue engineering systems. The key element for their broad spectrum of applications is the ability to establish triple helices and assemble into supramolecular structures. However, till today, these biomaterials are harvested from animal tissues generating safety concerns due to the potential transmission of animal-related infectious agents and induction of immunological reactions [1]. Additionally, due to this origin, the final product consists of a heterogeneous material made of different types of molecules that suffer from batch-to-batch variations [2], making its properties and performances difficult to predict. The contribution of collagen and gelatine to the highly demanding biomedical field could be broader if they were produced in an animal-free, reproducible and custom-made way.

Currently, recombinant microbial systems are the only available technology able to produce customized collagen and gelatine in an animal-free and cost-effective way. Recombinant microbial systems can produce retraceable products with consistent quality. The existing DNA technologies allow the design of novel collagen-inspired proteins with tailor-made functionalities, architectures [3-5] and enhanced biological properties [6].

## 6.1. Secreted expression of collagen-inspired proteins by *Pichia pastoris*

Our group has previously shown that the yeast *P. pastoris* can be successfully used for the secreted production of a new class of monodisperse, biodegradable and biocompatible gel-forming block copolymers [3, 7]. These polymers consist of triple helix-forming $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks (**$T_9$**) flanking a long highly hydrophilic random coil midblock (**$R_m$** or **$P_m$**) (Fig. 6.1) [7, 8]. Our triblock system was designed with the aim to independently tune the middle block and triple helix-forming end blocks, so that gels with predictable rheological properties could be produced. Previous work done by our group has proven that the middle block length can be independently modified from the end blocks as to originate gels with different rheological properties. Teles *et al*. have shown that the middle block length influenced gel strengths [8] and drug release profiles [4]. The authors also demonstrated that the middle block length could be designed so as to tailor drug release profiles [4]. Furthermore, it was shown that the triple helix formation occurs exclusively due to the presence of the trimerizing $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks (Fig. 6.1) [3, 7, 9]. The contribution of the middle block to triple helix formation was found to be negligible since triple helices formed by three $(Pro\text{-}Gly\text{-}Pro)_9$ chains always corresponded to a melting temperature ($T_m$) of ~ 41 °C (at a protein concentration of 1.1 mM) regardless of the middle block length and amino acid sequence [7]. On the biological production aspect, it was shown that, although longer middle blocks render higher gel strengths their secreted production was not affected. Gel-forming proteins with 400 (**$P_4$** or **$R_4$**) or 800 (**$P_8$** or **$R_8$**) amino acids long mid-blocks were secreted at the same yield (Table 6.1) [3, 8]. What is more, even though these polymers can indeed establish stable triple helices and eventually gels at *P. pastoris* growth temperature (30°C), no secretion impairment was observed and the proteins were secreted intact and at a high yield (Table 6.1) [3, 7].

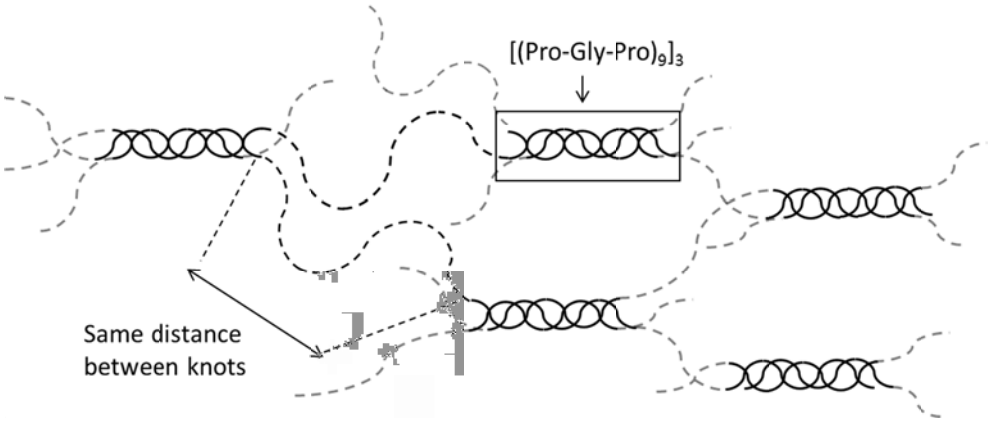



**Figure 6.1.** Schematic representation of hydrogels from collagen-inspired triblock polymers. Triple helix formation only occurs due to the presence of trimerizing (Pro-Gly-Pro) end blocks resulting in gels with defined distances between the crosslinks.

Hence, the possibility to use a recombinant system such as *P. pastoris* capable of growing in a defined, low cost medium and able to secrete high yields of custom-made proteins seemed to be ideal for the biomedical industry. Accordingly, the goal of this thesis was to prove the possibility to independently design the triblock system $T_m$ by changing the (Pro-Gly-Pro) end blocks length, as well as the use of *P. pastoris* for its secreted expression. End blocks with different lengths were genetically designed to assemble into triple helices with different thermostabilities. However, as it was shown in the previous chapters, the elongation of the triple helix forming end blocks seemed to influence more than just the triple helix $T_m$.

### 6.1.1. Triple helix strength is *P. pastoris'* weakness.

In Chapters 2 and 4, it was shown that the triple helix $T_m$ can be tuned by varying the (Pro-Gly-Pro) triple helix-forming domain length. As expected, longer (Pro-

Gly-Pro)$_n$ stretches yielded triple helices with higher thermostabilities due to a larger number of hydrogen bonds involved [10, 11]. However, the elongation of the (Pro-Gly-Pro) end blocks seemed to influence the protein secretion yields and intactness. While the secreted expression yield of **$T_6R_4T_6$** was comparable to the **$T_9P_4T_9$** polymer, the yield of **$T_{12}R_4T_{12}$** was intermediate between the yields of the polymers with **$T_9$** end blocks and polymers with **$T_{16}$** end blocks (see Chapter 4). Furthermore, the secreted production of the polymer with **$T_{16}$** end blocks seemed to have reached the limits of *P. pastoris* secretion capabilities and the **$T_{16}P_4T_{16}$** polymer was secreted at a 5-fold lower yield and partially degraded (Chapter 2). The polymer degradation could only be overcome by recurrence to a Yapsin 1 protease knockout mutant. This partial protein degradation was found to be exclusive to the polymer that can establish highly stable triple helices (**$T_{16}P_4T_{16}$**). Polymers with **$T_6$**, **$T_9$** and **$T_{12}$** end blocks corresponding to a $T_m$ of 10, 41 and 54°C respectively, were secreted intact, albeit with different yields (Chapter 4 and Table 6.1). Although, the occurrence of intracellular triple helix established by the other collagen-inspired polymers cannot be discarded, it can be suggested that such helices are not as stable as the ones formed by **$T_{16}$** end blocks, resulting in a balance between triple helix formation and dissociation. The *in vivo* occurrence of stable triple-helices by **$T_{16}$** end blocks could directly trigger proteolysis or simply allow it to occur due, to the probable longer intracellular residence times of **$T_{16}P_4T_{16}$** when comparing with other proteins. The results obtained in Chapter 2 and 4 seem to indicate that there is an inverse relationship between triple helices stability and the polymers secreted expression yields.

A similar relationship between yield and stability was observed on the secreted production of elastin-like polypeptides (ELPs) [12]. ELPs are polymers composed of a repetition of short peptides inspired by the hydrophobic regions of natural elastin. Like the collagen-inspired polymers, ELPs are thermo-responsive but in an inverse way. While collagen-inspired polymers agglomerate when the temperature is lower than their $T_m$, ELPs agglomerate only when the temperature is

higher than their transition temperature ($T_t$) [13]. When a solution of ELPs is heated above its $T_t$, the hydrophobic regions cause contraction, desolvation, and protein aggregation. Schipperus *et al*. observed that ELPs with low $T_t$ are produced at a lower yield than polymers with higher $T_t$ [13]. The authors hypothesized that ELPs with lower $T_t$ are more prone to meet, in *P. pastoris* secretory pathway, conditions at which, they aggregate or stick to other cell components due to their highly hydrophobic content [13].

In this thesis, we were able to discard the $\mathbf{T_{16}P_4T_{16}}$ low secretion yields and partial degradation, due to hydrophobic interactions. This, thanks to the design of a randomized version of equal hydrophobic content but unable to form triple helices, the $\mathbf{S_{16}}$ block. The $\mathbf{S_{16}P_4S_{16}}$ proteins were secreted intact and at high yields despite having the same amino acid composition and hence hydrophobic content (Table 6.1 and Chapter 2). It could be concluded that the impaired secretion and partial protein degradation of $\mathbf{T_{16}P_4T_{16}}$, was triggered by the intracellular occurrence of stable triple helices.

Together, these observations suggest that there is an inverse relationship between the ability to establish stable supramolecular structures and the protein secretion yields. In order to overcome the secretion bottleneck of triple helix forming polymers, several attempts were made to optimize *P. pastoris*' growth medium or avoid triple helix formation itself.

### 6.1.2 Growth medium optimization for collagen-inspired proteins secreted expression by *P. pastoris*

Growth conditions are known to influence protein production yields and quality [14-17]. To obtain optimal production yields, fermentation protocols must be optimized according to the protein being expressed. So far, there is no such thing has an universal growth medium.

In Chapter 2 it was demonstrated that **$T_{16}P_4T_{16}$** partial degradation could only be avoided by the use of a *P. pastoris* Yapsin 1 deficient strain. Although, the use of a Yapsin 1 knockout strain allowed the production of intact collagen-inspired proteins with high thermal stability, the high secretion yields were not re-established. Schipperus *et al*. have shown that the ELPs production and degradation was pH dependent [12]. Within the experimental pH range of 3 to 7, an increasing secretion yield was found in cultures grown at higher pH, with an optimum at pH 6. As the pH of the medium is known to influence the protein expression yields [14,15,18], the optimization of the fermentation medium pH was attempted. Contrary to the previous authors, it was observed that the expression of **$T_{16}P_4T_{16}$** at pH 6 resulted in lower cell growth (Fig. 6.2). SDS-PAGE of microfiltrated cell-free broth at pH 6 revealed the presence of extra protein bands that could be associated with cell lysis (Fig. 6.3). Cellular lysis and growth arrest were only observed after methanol induction, suggesting it resulted from **$T_{16}P_4T_{16}$** expression. The extracellular pH influence seemed to be restricted to **$T_{16}P_4T_{16}$** since, the growth rate of cells expressing **$T_9P_4T_9$** was not influenced by the pH of the medium (Fig. 6.2). The inverted effect of the medium pH on the optimal growth conditions for the expression of the ELPs and collagen-inspired proteins, could be explained by two different mechanisms. On the cellular level, the higher (extracellular) pH could cause changes in the chaperones availability or on the efficiency of the proteins transport across the cell wall, resulting in aggregation or mislocalization. Simply, the cellular machinery at pH 3 could be more beneficial for the **$T_{16}P_4T_{16}$** transport while the conditions at pH 6 could be better for ELPs secretion. Accordingly, it was shown in Chapter 3 that at pH 6 and under batch conditions, **$T_{16}P_4T_{16}$** accumulates in cellular compartments associated with the ER. Formation of highly stable triple helices seems to occur already in the ER, possibly blocking or retarding protein secretion. However, these observations can only be stated for cells grown under batch conditions and at pH 6. Future research on the

impact of the extracellular pH on cells grown under fed-batch and batch conditions should be performed, to clearly understand the pH influence on the $\mathbf{T_{16}P_4T_{16}}$ secreted expression.

On the protein level, the exposure of the ELPs to low pH in the periplasm could trigger an increase of its hydrophobicity causing intracellular protein agglomeration and secretion arrest. Secretion arrest could also be caused by the ELPs interaction with the cellular lipid membranes due to their higher hydrophobicity.

For $\mathbf{T_{16}P_4T_{16}}$, the presence, at pH 6, of positively charged lysines and negatively charge glutamic acids could lead to electronic interactions between the $\mathbf{P_4}$ middle blocks, obstructing the access of Yapsin 1 or other protease to their restriction site. The degradation profile of $\mathbf{T_{16}P_4T_{16}}$ polymers secreted at pH 6 was similar to the one at pH 3, so the lack of protease activity at the higher pH was not the cause for cell lysis. But perhaps, was not as efficient. Yapsins are zymogen proteins requiring an acidic pH (pH4) for their activation [19]. The fact that the external pH was 6 could mean that this specific set of enzymes was not active and degradation was probably performed by other protease. To see if there are other pH specific proteases, it could be interesting to grow the yapsin 1 mutant at pH 6 and see if $\mathbf{T_{16}P_4T_{16}}$ is once again partially degraded.

Curiously, the attachment of the enhanced Green Fluorescence Protein (eGFP) molecule to both ends of the $\mathbf{T_{16}P_4T_{16}}$ molecule, restored the normal cellular growth rates at pH 6 (Fig. 6.2). The product presence in the fermentation medium was confirmed by measuring the eGFP fluorescence levels in the fermentation supernatant. Fluorescence levels increased throughout the fermentation and no significant levels of cell lysis were observed (Fig. 6.4). This could mean that the protein present in the fermentation medium resulted from secretion and not to cell lysis. The attachment of eGFP, a 27 kDa globular protein, to $\mathbf{T_{16}P_4T_{16}}$ seemed to have resulted in less cellular stress. There could be several explanations for this observation. On one hand, the presence of eGFP could have

altered the secretion pathway and simply by coincidence, resulted in a new secretion pathway that caused less stress to the cells secreting **eGFP-$T_{16}P_4T_{16}$-eGFP** at pH 6.

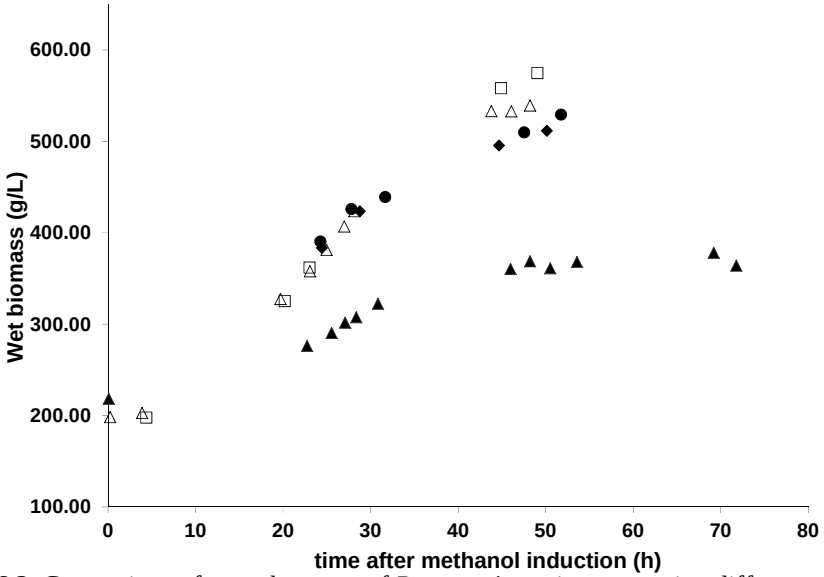



**Figure 6.2.** Comparison of growth curves of *P. pastoris* strains expressing different triblock copolymers at different pH. At pH 3, (△) **$T_{16}P_4T_{16}$** and (□) **$T_9P_4T_9$**. At pH 6, (●) **$T_9P_4T_9$** and (▲) **$T_{16}P_4T_{16}$** (◆) **eGFP-$T_{16}P_4T_{16}$-eGFP**.

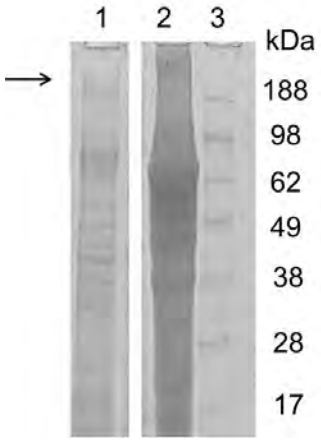



**Figure 6.3.** SDS-PAGE of extracellular medium of *P. pastoris* cells expressing **$T_{16}P_4T_{16}$** at different extracellular pH. Lane 1, pH 3; Lane 2, pH 6; Lane 3, molecular weight marker.

For all samples, 10 μL of microfiltered cell-free broth after ~ 48 h of methanol induction was loaded.

In fact, it has been showed that eGFP is not a passive reporter but rather forms oligomers in the secretory pathway, which could alter the intracellular transport of the protein [20]. Also, the attachment of eGFP to the triple helix-forming end blocks could have delayed or obstructed triple helix formation and avoid cellular stress. In Chapter 2, it was demonstrated that production problems of $\mathbf{T_{16}P_4T_{16}}$ are due to intracellular triple helix formation suggesting, that the temporarily suppression of triple helix formation could resolved the problem. Unfortunately, it was not possible to purify **eGFP-$\mathbf{T_{16}P_4T_{16}}$-eGFP** neither by ammonium sulfate precipitation nor by anti-GFP immunoprecipitation, so nothing can be concluded about the protein intactness.

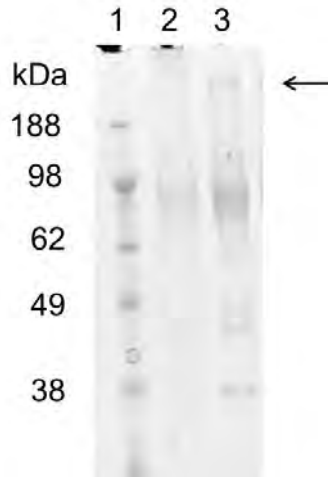



**Figure 6.4.** SDS-PAGE of extracellular medium of *P. pastoris* cells expressing $\mathbf{T_{16}P_4T_{16}}$ or **eGFP-$\mathbf{T_{16}P_4T_{16}}$-eGFP**. Lane 1, molecular weight marker; Lane 2, $\mathbf{T_{16}P_4T_{16}}$; Lane 3, **eGFP-$\mathbf{T_{16}P_4T_{16}}$-eGFP**. For both samples, 10 μL of microfiltered cell-free broth after ~ 48 h of methanol induction was loaded.

### 6.1.3 Avoiding intracellular triple helix formation

In order to hinder triple helix formation and secrete a protein that could be easily purified from the fermentation medium, half of the **$P_4$** mid-block was added to both ends of the **$T_{16}P_4T_{16}$**, originating the **$P_2T_{16}P_4T_{16}P_2$** polymer. The new addition indeed re-established the production and secretion yields (Fig. 6.4 and Table 6.1) but was not sufficient to inhibit partial protein degradation (data not shown). **$P_2T_{16}P_4T_{16}P_2$** was degraded in a similar manner as **$T_{16}P_4T_{16}$** suggesting that Yapsin 1 was again responsible for protein degradation. Yet, SDS-PAGE analysis of **$P_2T_{16}P_4T_{16}P_2$** supernatant fermentation broth shows that, there are fewer *P. pastoris* proteins in the fermentation medium indicating that there is less cellular lysis (Fig. 6.5). This suggests that the secreted expression of this new construct seemed to have caused less cellular stress. As a control, the length of the mid-block was doubled to ~ 800 amino acids (**$P_8$**) creating a protein with the same number of blocks as **$P_2T_{16}P_4T_{16}P_2$** but different sequence, the **$T_{16}P_8T_{16}$** polymer. Although this elongation also reduced the relative length of each of the **$T_{16}$** end-blocks to ~5% of the entire molecule like in **$T_9P_4T_9$**, it did not improve the secretion yield (Table 6.1). In reality, **$T_{16}P_8T_{16}$** still has two end blocks capable of supramolecular assembly, indicating that the strength of the triple helix is indeed the cause for secretion impairment.

Based on these results, further research could be done on the secreted expression of **$P_2T_{16}P_4T_{16}P_2$** in a Yapsin 1 knockout strain. The expression of **$P_2T_{16}P_4T_{16}P_2$** in a Yapsin 1 deficient strain might yield an intact polymer. It would be interesting to discern if the **$P_2$** blocks attached to the **$T_{16}$** triple helix-forming end-blocks could indeed interfere with triple helix formation. DSC analysis should be performed to verify if the triple helices formed would be less stable than the ones established in the **$T_{16}P_4T_{16}$** polymer.

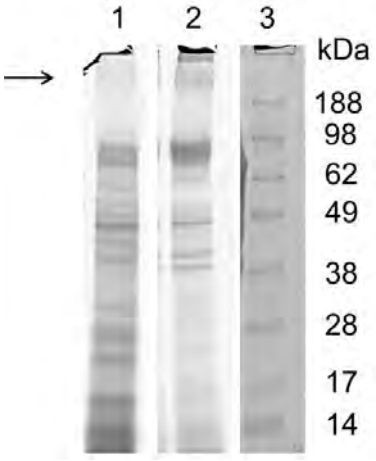



**Figure 6.5.** SDS-PAGE of extracellular medium of *P. pastoris* cells expressing $\mathbf{T_{16}P_4T_{16}}$ and $\mathbf{P_2T_{16}P_4T_{16}P_2}$. Lane 1, $\mathbf{T_{16}P_4T_{16}}$; Lane 2, $\mathbf{P_2T_{16}P_4T_{16}P_2}$; Lane 3, molecular weight marker. For both samples, 10 μL of microfiltered cell-free broth after ~ 48 h of methanol induction was loaded.

### 6.1.4 The secretory pathway must be traveled alone

Although it was not unequivocally proven that intracellular triple helix hinderance could re-establish high secretion yields, the results obtained seem to point on that direction. Therefore, a universal *P. pastoris* production system could possibly be designed as to achieve high secretion yields. Primarily, intracellular triple helix formation should be avoided by attaching a protein known to be properly secreted, easily purified and preferably globular. The presence of a globular protein should act as an obstacle to triple helix formation (Fig. 3.5 in Chapter 3). During the design of the chimeric protein a cleavage site could be added between the collagen-inspired polymer protein and the triple-helix hindrance protein (tag) in order to easily remove it when it is no longer necessary. Once a high secretion yield is achieved, proteins could be purified from the fermentation broth and *in vitro* removal of the tag protein could be done by enzymatic reaction, so that triple helix formation could occur (see Chapter 3 Fig. 3.5).

**Table 6.1.** Comparison of secretion yield, gel formation capabilities and melting temperatures of triple helix and non-triple helix forming collagen-inspired proteins.

| | Protein | Tm (°C)[a] | Protein quality | Secretion yield (g/L) | End-blocks | Reference |
|---|---|---|---|---|---|---|
| Trimerizing proteins | $T_6R_4T_6$ | 10 | intact | ~ 4 [b] | $(Pro\text{-}Gly\text{-}Pro)_6$ | Chapter 4 |
| | $T_9P_4T_9/T_9R_4T_9$ | 41 | intact | 2.1 ± 0.2 | $(Pro\text{-}Gly\text{-}Pro)_9$ | [7, 9] |
| | $T_9P_8T_9/T_9R_8T_9$ | 41 | intact | 2.0 ± 0.2 | $(Pro\text{-}Gly\text{-}Pro)_9$ | [3] |
| | $WT_9R_4T_9C$ | 41 /50 | partially degraded | n.a | $(Pro\text{-}Gly\text{-}Pro)_9$ | [21] |
| | $T_{12}R_4T_{12}$ | 54 | intact | ~ 2 [b] | $(Pro\text{-}Gly\text{-}Pro)_{12}$ | Chapter 4 |
| | $T_{16}P_4T_{16}/T_{16}R_4T_{16}$ | 74 | partially degraded | 0.38 ± 0.02 | $(Pro\text{-}Gly\text{-}Pro)_{16}$ | Chapter 2/This chapter |
| | $P_2T_{16}P_4T_{16}P_2$ | n.m | partially degraded | 2.0 ± 0.03 | $(Pro\text{-}Gly\text{-}Pro)_{16}$ | This chapter |
| | $T_{16}P_8T_{16}$ | n.m | partially degraded | 0.38 ± 0.02 | $(Pro\text{-}Gly\text{-}Pro)_{16}$ | This Chapter |
| Non-Trimerizing proteins | $S_9P_4S_9$ | -[c] | intact | 2.1 ±0.6 | randomized version of $(Pro\text{-}Gly\text{-}Pro)_9$ | Chapter 2 |
| | $S_{16}P_4S_{16}$ | -[c] | intact | 2.6 ±0.2 | randomized version of $(Pro\text{-}Gly\text{-}Pro)_{16}$ | Chapter 2 |

[a] concentrations at which $T_m$ was measured can be seen in the respective chapters
[b] yield based on a single data point.
[c] not applicable
n.a – data not available
n.m – not measured

Besides the protein engineering solution, the use of other recombinant systems could also be considered. For the production of collagen-inspired proteins that can establish highly stable triple helices it would be advisable to use an organism that can grow at high temperatures. The filamentous fungi *Aspergillus niger* is able to grow at temperatures as high as 50°C and could therefore be an option [22, 23]. Several mammalian cell lines have also been used to produce novel

engineering collagens [24]. This last system has the advantage of having adequate levels of endogenous collagen processing proteins (such as P4H and HSP47) facilitating the recombinant collagen folding and transport along the secretory pathway [24, 25].

## 6.2 Tunability of the thermostability

One of the aims of this thesis was to prove that the different blocks of our collagen-inspired triblock system can be independently tuned and originate gel networks with predictable properties. Natural collagen and gelatine preparations are highly complex and therefore it is extremely difficult to predict their supramolecular conformations and stability based on their amino acid sequence. In our triblock system, triple helices are established exclusively by hydrogen bonds between a known $(Pro\text{-}Gly\text{-}Pro)_n$ sequence making the prediction of the its thermostability possible (Chapter 4). It was shown, that the variation of the thermostability could be understood in terms of i) a simple linear relationship between the length of the end block and the free energy of helix formation, or ii) its constituting entropic and enthalpic components (Chapter 4). Furthermore, the use of reversible ligations, like hydrogen bonds, allows the adaptability of our gel networks to external stimuli such as pH, temperature or salinity. When designed in such a way, these materials, known as smart materials, are able to change their conformation according to the external conditions. However, if reversibility is not a requesite, other aproaches could be followed to vary triple helices thermostability. Enzymatic reactions like hydroxylation of the proline residues could be performed as to increase the triple helix thermostability and further mimick collagen molecules but, this possibility requires the use of mammalian cells with a native P4H [2, 24] or the use of a recombinant system with a heterologous P4H [1, 26, 27]. In Chapter 4, it was demonstrated that by increasing the number of (Pro-Gly-Pro) domains through genetic engineering, the thermomechanical stability of our polymers triple helices

could be increased. In a similar manner, the trimerizing blocks could be designed as to incorporate aminoacids able to establish covalent ligations. For example, cysteine residues can establish disulphide bonds originating an irreversible covalent ligation. Our group has designed a collagen-inspired polymer based on the **$T_9R_4T_9$** polymer to which a tryptophan (W) and a cysteine (C) amino acid were added to the N- and C-terminal ends, respectively, originating the **$WT_9R_4T_9C$** polymer [21]. Differential scanning calorimetry (DSC) of the **$WT_9R_4T_9C$** polymer revealed a biphasic profile with one peak at 41°C and a second peak with a maximum at 50°C (Fig. 6.6). The first peak corresponds to the formation of triple helices by the **$T_9$** end block; triple helices formed by this domain have a $T_m$ of 41 °C (at 1.1 mM protein concentration). The second peak could be attributed to the triple helices formed by **$T_9$** end block stabilized by a disulphide bond which could increase the triple helices $T_m$. The formation of a disulphide bridges will decrease the translational (diffusive) entropy component in the non-triple helical state. Therefore, there will be a lower entropy penalty for trimer formation. This directly stabilizes the helix form, stirring the equilibrium in the direction of the helices, originating helices with higher $T_m$. Curiously, these polymers were secreted at a low yield and partially degraded by *P. pastoris*. Despite having a lower melting temperature than the properly secreted **$T_{12}R_4T_{12}$**, it was necessary to use a Yapsin 1 knockout strain to produce intact polymers. Several scenarios could explain such result. Formation of triple helices occurs in two steps: nucleation and propagation [5]. Formation of helix nuclei requires the simultaneous collision of three mutually aligned T blocks. At low concentrations, when such a simultaneous collision is rather infrequent, nucleation is rate limiting [5]. However, in the **$WT_9R_4T_9C$** polymer, disulfide bonds can be established and stable dimmers are formed increasing the possibilities for trimer formation. Formerly, when trimer formation relied exclusively on weak, reversible hydrogen bonds the chance of three helices meeting at the same time was lower. As it was demonstrated in this thesis, intracellular formation of stable triple helix leads to protein degradation.

**$WT_9R_4T_9C$** polymers could be degraded simply by the higher chance of its end blocks establishing stable triple helices. Also, the presence of the disulphide bonds could lead to the formation of hexamers or other possible supramolecular structures which could also explain the $T_m$ increase and protein degradation. So degradation could simply be a reaction to the formation of supramolecular structures and not directly to trimer stability. Regardless of the reasons behind protein degradation, it was demonstrated that the addition of an amino acid able to establish covalent ligations influenced the thermostability of the triple helix. Throughout this thesis it was demonstrated that our polymers properties can be precisely tuned and predicted through genetic engineering design which, together with their biocompatibility, makes these materials suitable for a variety of biomedical applications.

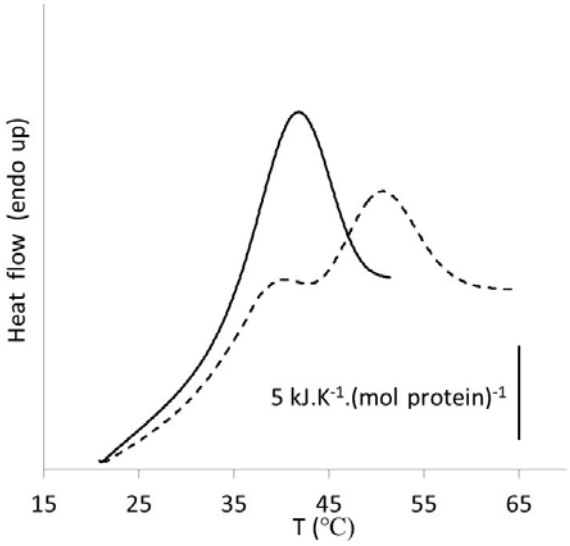



**Figure 6.6.** Thermal denaturation of **$T_9R_4T_9$** and **$WT_9R_4T_9C$** (1.2 mM), as reflected by DSC. The DSC profile of **$T_9R_4T_9$** (full line) shows a single peak with a maximum at ~ 41°C. **$WT_9R_4T_9C$** DSC profile (dashed line) shows a biphasic thermogram [21]. The low temperature peak coincides with $T_m$ of the triple helices formed exclusively by hydrogen bonds and the higher temperature peak could be attributed to the triple helices in which a disulphide bond is involved.

### 6.2.1 Biomedical applications

Collagen and collagen-based biomaterials are widely used in the fields of drug delivery, tissue engineering and wound dressing. To date, most of the present research is aimed at the enhancement of the mechanical strength, biodegradability or delivery characteristics of these materials. For tissue engineering applications it is known that, the elasticity of the cell growth matrix directs stem cell development to different lineages [28]. Soft networks (0.1-1 kPa) mimic brain tissue promoting neuron development, stiffer scaffolds (8-17 kPa) allow muscular tissue growth, and the growth of bone cells are promoted by gel networks with and elastic modulus of 24-40 kPa [28]. Teles *et al.* have shown that the middle block length, of our triblock system, can be changed as to originate gel networks with specific drug delivery profiles making them interesting drug delivery and wound dressing materials. However, the authors have shown that at physiologic temperature (37°C) the gels erode in 2 days which for tissue engineering purposes would not suffice since it is too fast for tissue growth. In the previous chapters, it was demonstrated that the end blocks length influenced both the gel strength and its thermal stability with longer end-blocks originating sturdier gel structures with slower erosion rates. Our triblock system can hence be designed to create gel networks with the desired elasticity and erosion rates for specific cell growth lines. Therefore, the recently developed collagen-inspired polymers have potential in the biomedical field and should be further studied to analyse the tunability of drug delivery profiles and cell growth promotion.

## 6.3 Conclusions and future prospects

Over the course of this project we designed and produced, with different degrees of success, gel-forming collagen-inspired triblock copolymers using the yeast *P. pastoris*. It was shown that, through genetic engineering the thermostability of the

collagen-inspired polymers triple helix can be tuned and predicted, with longer end blocks having higher melting temperatures due to the higher number of hydrogen bonds involved. However, when using a recombinant organism like *P. pastoris* there seems to be an inverse relationship between protein secretion yields and triple helix thermostability and formation kinetics. Polymers with high triple helix thermostability were secreted at lower yields than polymers with lower triple helix thermostabilities. It was shown that, intracellular formation of stable triple helices leads to cellular stress and protein degradation and should be avoided if possible. In case such occurrence cannot be avoided, a *P. pastoris yapsin 1* mutant can be used as a cell factory for the secreted production of intact collagen-inspired proteins that can form highly stable triple helices albeit the lower secretion yields.

It was demonstrated that the use of triple helix-forming $(Pro\text{-}Gly\text{-}Pro)_n$ domains allows the predictability of the triple helices stability since both enthalpic and entropic contributions of each hydrogen bond are known. This know-how will permit the design of gel networks with specific rheological properties and applications. This system also allows the introduction of different blocks with other characteristics and functionalities, such as human sequences or cell binding sites. Although the applicability of these polymers in the biomedical field was a major concern throughout this thesis that, should not be a constraint and other fields and alternatives should be explored. For example, biorecognition and biomineralization sequences could be incorporated for the development of new types of conductors that can be used in the field of electronics, optics and sensors devices [29].

## References

[1] Pakkanen, O., Pirskanen, A., Myllyharju, J., *J Biotechnol* 2006, *123*, 248-256.
[2] Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., Carmichael, D., Perala, M., Hamalainen, E.-R., Jarvinen, M., Polarek, J., *Adv Drug Deliv Rev* 2003, *55*, 1547-1567.
[3] Teles, H., Skrzeszewska, P. J., Werten, M. W. T., van der Gucht, J., Eggink, G., de Wolf, F. A., *Soft Matter* 2010, *6*, 4681-4687.
[4] Teles, H., Vermonden, T., Eggink, G., Hennink, W. E., de Wolf, F. A., *J Control Release* 2010, *147*, 298-303.
[5] Skrzeszewska, P. J., de Wolf, F. A., Stuart, M. A. C., van der Gucht, J., *Soft Matter* 2010, *6*, 416-422.
[6] Majsterek, I., McAdams, E., Adachi, E., Dhume, S. T., Fertala, A., *Protein Sci* 2003, *12*, 2063-2072.
[7] Werten, M. W., Teles, H., Moers, A. P., Wolbert, E. J., Sprakel, J., Eggink, G., de Wolf, F. A., *Biomacromolecules* 2009, *10*, 1106-1113.
[8] Teles, H., Vermonden, T., Eggink, G., Hennink, W. E., de Wolf, F. A., *Journal of Controlled Release* 2010, *147*, 298-303.
[9] Skrzeszewska, P. J., de Wolf, F. A., Werten, M. W. T., Moers, A. P. H. A., Stuart, M. A. C., van der Gucht, J., *Soft Matter* 2009, *5*, 2057-2062.
[10] Persikov, A. V., Ramshaw, J. A., Brodsky, B., *J Biol Chem* 2005, *280*, 19343-19349.
[11] Brodsky, B., Thiagarajan, G., Madhan, B., Kar, K., *Biopolymers* 2008, *89*, 345-353.
[12] Schipperus, R., Teeuwen, R. L., Werten, M. W., Eggink, G., de Wolf, F. A., *Appl Microbiol Biotechnol* 2009, *85*, 293-301.
[13] Schipperus, R., Eggink, G., de Wolf, F. A., *Biotechnology Progress* 2011, In press.
[14] Jahic, M., Gustavsson, M., Jansen, A. K., Martinelle, M., Enfors, S. O., *J Biotechnol* 2003, *102*, 45-53.
[15] Cereghino, J. L., Cregg, J. M., *Fems Microbiol Rev* 2000, *24*, 45-66.
[16] Plantz, B. A., Sinha, J., Villarete, L., Nickerson, K. W., Schlegel, V. L., *Appl Microbiol Biotechnol* 2006, *72*, 297-305.
[17] Sinha, J., Plantz, B. A., Inan, M., Meagher, M. M., *Biotechnology and Bioengineering* 2005, *89*, 102-112.
[18] Cregg, J. M., Cereghino, J. L., Shi, J., Higgins, D. R., *Mol Biotechnol* 2000, *16*, 23-52.

[19] Gagnon-Arsenault, I., Parise, L., Tremblay, J., Bourbonnais, Y., *Mol Microbiol* 2008, *69*, 982-993.
[20] Jain, R. K., Joyce, P. B., Molinete, M., Halban, P. A., Gorr, S. U., *Biochem J* 2001, *360*, 645-649.
[21] Skrzeszewska, P. J., personal communication.
[22] Abd El-Rahim, W. M., El-Ardy, O. A. M., *Desalin Water Treat* 2011, *25*, 39-46.
[23] Abrashev, R., Dolashka, P., Christova, R., Stefanova, L., Angelova, M., *J Appl Microbiol* 2005, *99*, 902-909.
[24] Bulleid, N. J., John, D. C., Kadler, K. E., *Biochem Soc Trans* 2000, *28*, 350-353.
[25] Kyle, S., Aggeli, A., Ingham, E., McPherson, M. J., *Trends Biotechnol* 2009, *27*, 423-433.
[26] Pakkanen, O., Hamalainen, E.-R., Kivirikko, K. I., Myllyharju, J., *J Biol Chem* 2003, *278*, 32478-32483.
[27] Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., Kivirikko, K. I., *EMBO J* 1997, *16*, 6702-6712.
[28] Engler, A. J., Sen, S., Sweeney, H. L., Discher, D. E., *Cell* 2006, *126*, 677-689.
[29] Bai, H. Y., Xu, F., Anjia, L., Matsui, H., *Soft Matter* 2009, *5*, 966-969.

# Summary

*Summary*

It is the ability to establish triple helices and assemble into supramolecular structures, which makes collagen and its denature counterpart, gelatine, interesting for the food and biomedical industry. Collagen and gelatine array of applications is quite extensive, ranging from gelling agents in food, emulsifiers in photographic films, fillers in cosmetics and structural networks in drug delivery and tissue engineering systems. Despite their vast use, the animal origin of these protein materials, mainly bovine and porcine tissues, poses the risk of pathogen transmission or allergic reactions. However, mammalian tissues contain more than one type of collagen resulting in a lack of reproducibility and heterogeneity between batches. The variability in composition and structure of animal-derived collagen and gelatine presents a significant challenge for those using these proteins in medical applications, where reliable, predictable and traceable materials are essential. Additionally, the harvest of collagen and gelatine from animal tissues restrains us to the use of what is already available in nature and in commercial quantities.

Thanks to the development of DNA recombinant technologies, the capability to design genes and obtain proteins with custom made conformations and functionalities has become a current day activity. In accordance, a myriad of recombinant systems have been developed to produce high yields of high quality heterologous proteins. However, there is no universal recombinant system that can be used by default. The choice between recombinant systems depends of the protein being expressed, the need for post translational modifications, the concentration of protein required and the costs associated to its production. In the present thesis, *Pichia pastoris* was the recombinant system of choice due to its many advantages. It is able to reach high cell densities on defined low-cost media, has a tight regulated AOX methanol induced promoter and can produce properly folded proteins with correct disulphide bond formation and other eukaryotic

*Summary*

posttranslational modifications. The ability to secrete high amounts of heterologous proteins and low amounts of endogenous proteases is further considered as a plus by the biomedical industry, as downstream processing costs can be considerably reduced when compared to other industrial strains as *Saccharomyces cerevisae* or *Escherichia coli*.

The possibility to custom design collagen and gelatine-inspired biomaterials and use a reliable production system like *P. pastoris*, is highly appealing for the safety demanding biomedical field.

Our group has previously used this system for the secreted production of collagen-inspired proteins formed by two triple helix-forming $(Pro\text{-}Gly\text{-}Pro)_9$ end blocks and one long random coil middle block. The triblock system allows the formation of gel networks with defined properties and porosities since i) triple helix is exclusively established by the (Pro-Gly-Pro) end blocks, connected by ii) a random coil middle block of known length. It was shown that the middle block can be independently tuned to originate gel networks with predictable rheological properties and drug delivery profiles and had no effect over the triple helices melting temperature. Only the collagen-inspired $(Pro\text{-}Gly\text{-}Pro)_9$ end block domain seemed to contribute to the triple helix $T_m$ which was ~ 41°C at a 1.1.mM protein concentration. Despite the possibility to establish triple helices ($T_m$ ~ 41°C) at *P. pastoris* growth temperature (30°C), the secreted expression of these proteins was not impaired. Also, the elongation of the middle block proved to be no burden for *P. pastoris* secretory capability, since similar secretion yields were found for triblocks with middle blocks of ~37 or ~73 kDa. These positive results, paved the way to the design of a series of triblock polymers with different end block lengths and hence $T_m$. However, every system as a limit and *P. pastoris* is no exception.

In Chapter 2, it was demonstrated that the elongation of the collagen-inspired $(Pro\text{-}Gly\text{-}Pro)_9$ domain to $(Pro\text{-}Gly\text{-}Pro)_{16}$ resulted in a considerable lower secretion

*Summary*

yield and partial protein degradation by protease Yapsin 1. This result could have been explained by i) higher prolyl-tRNA turnover requirements, ii) higher hydrophobic content resulting in hydrophobic interactions or iii) incapacity to process highly stable triple helical conformations ($T_m$ of ~74°C). The restitution of the secretion yields and intact product secretion by designing a polymer with non-collagen-like end blocks, revealed that the protein secretion impairment was due to intracellular triple helix formation. The non-collagen-like endblocks have the same amino acid composition but randomized sequence as to avoid triple helix formation. The intact gel-forming polymer could only be obtained by recurring to a *P. pastoris* strain deficient in the periplasmic protease Yapsin 1. Although the use of Yapsin 1 knockout strain allowed the secretion of intact polymer it did not restore the high secretion yields. To know where the secretion bottleneck was occurring, a closer look at *P. pastoris* cells was conducted.

In Chapter 3, electron microscopy was performed to observe where triple helix formation was occurring. ER expansion and vesicle formation were observed only when gel-/triple helix-forming proteins with a high $T_m$ were being expressed. Such observations were absent in cells expressing proteins with mutant end-blocks, endblocks with equal amino acid composition but unable to establish triple helices. The presence of extra organelles could have resulted from protein aggregation due to triple helix formation or ER volume enlargement as an attempt to minimize protein aggregation. This result suggests that triple helix formation should be avoided in order to achieve high secretion yields of proteins able to establish highly stable triple helices.

In Chapter 4, it was shown that the thermomechanical stability of the triple helices could be tuned by varying the length of the (Pro-Gly-Pro) end blocks through genetic engineering. An increase of the triple helix $T_m$ could be achieved by

*Summary*

increasing the number of hydrogen bonds involved without the need to resort to chemical (covalent) crosslinking. All triblock copolymers studied formed stable hydrogels, and, an increase of the end block length resulted in higher stabilities under mechanical stress. The variation of the thermostability could be understood in terms of i) a simple linear relationship between the length of the end block and the free energy of helix formation, or ii) its constituting entropic and enthalpic components. The comparison with triple helices formed by free $(Pro\text{-}Gly\text{-}Pro)_n$ peptides seemed to indicate that this relationship was not influenced by the nature of the middle blocks. Regarding the use of *P. pastoris* as a cell factory, it was observed that there is an inverse relation between triple helix stability and secreted production yield. Proteins able to establish triple helices with high thermostablity were secreted at a lower yield.

In Chapter 5, it was shown that denaturing SDS-PAGE gels can monitor the ability of collagen-inspired proteins to form supramolecular assemblies. During the destaining process proteins that cannot establish triple helices diffuse out of the acrylamide gel, while triple helix-forming collagen-inspired proteins are still visible after destaining over-night. Furthermore, it was observed that the diffusion speed from the gel is related to the triple helix $T_m$. Proteins that can establish triple helices with high thermostability diffuse slower from the gel. In addition, the migration of the middle blocks in a SDS-PAGE gel revealed that subtle differences in migration speed can expose changes in the amino acid sequence of random coil proteins.

During the course of this project several attempts were made to increase the secretion yield of collagen-like proteins with high thermo stability. While not all experiments were successful, their results do yield additional hypothesis for further research that could lead to a further improvement of *P. pastoris* as a cell factory for

the production of collagen-like proteins. These experiments are discussed in detail in Chapter 6.

# Samenvatting

*Samenvatting*

Collageen en zijn gedenatureerde vorm, gelatine, zijn van belang voor de levensmiddelen en biomedische industrie omdat beide polymeren via triple helix vorming een supramoleculaire structuur kunnen creëren. De toepassingen van collageen en gelatine zijn uitgebreid en variëren van gelerende verbindingen in levensmiddelen, emulgatoren ten behoeve van fotografische films, vulmiddelen en dragermateriaal in cosmetica en medicijnen. Ondanks hun brede gebruik brengt de dierlijke herkomst van deze eiwitten risico's met zich mee, zoals overdracht van besmettelijk ziektes en allergische reacties. Daarnaast bevatten dierlijke cellen verschillende typen collageen wat resulteert in variatie in kwaliteit en samenstelling van de verschillende producten. Vooral voor medische toepassingen is deze variatie een grote uitdaging omdat betrouwbaarheid en voorspelbaarheid ten aanzien van de eigenschappen en belangrijk kwaliteitskenmerk zijn.

Dankzij de ontwikkeling van recombinant DNA technologie is de mogelijkheid ontstaan om genen te ontwerpen die coderen voor eiwitten die heel specifieke eigenschappen en functionaliteiten hebben. In deze proefschrift, is de gist *Pichia pastoris* gebruikt als expressie systeem voor synthetische genen die coderen voor de verschillende collageen varianten. Met *P. pastoris* kunnen eenvoudig hoge celdichtheden worden bereikt en via een methanol induceerbare promotor kunnen recombinant eiwitten tot hoge expressie worden gebracht. Bovendien beschikt *P. pastoris* over een efficiënt secretie mechanisme zodat de eiwitten intact door de gistcel kunnen worden uitgescheiden. Hierdoor kunnen de eiwitten relatief simpel en goedkoop tot hoge zuiverheid worden geïsoleerd .

Onze groep heeft dit systeem al veelvuldig gebruikt voor de productie van collageen-geïnspireerde eiwitten bestaande uit twee triple helix vormende eindblokken en een lang "random coil" midden blok. Het zogenaamde tri-blok systeem maken gel vormende netwerken mogelijk die precies gedefinieerde eigenschappen hebben. Het midden blok kan zo ontworpen worden dat het gelvormende netwerk voorspelbare reologische eigenschappen heeft zonder dat de

*Samenvatting*

smelt temperatuur van de triple helix eindblokken daardoor wordt beïnvloed. Ondanks het feit dat de smelttemperatuur ($T_m$ 41°C)van de triple helix vorm van deze eiwitten aanzienlijk hoger is dan de cultivatie temperatuur (30°C) van *P. pastoris* werden deze eiwitten normaal uitgescheiden zonder dat daarbij triple helix vorming optrad. Deze positieve resultaten opende de weg naar het ontwerp van nieuwe tri-blok eiwitpolymeren waarbij de lengte van de eindblokken werd gevarieerd en daarmee naar verwachting de smelttemperatuur.

In Hoofdstuk 2, wordt aangetoond dat verlenging van de op collageen gelijkende eindblokken van $(Pro\text{-}Gly\text{-}Pro)_9$ naar $(Pro\text{-}Gly\text{-}Pro)_{16}$ resulteert in een aanmerkelijk lager opbrengst van het uitgescheiden eiwit en gedeeltelijke afbraak van het eiwit door het protease Yapsin 1. Dit resultaat kan mogelijk verklaard worden uit i) noodzaak van hogere prolyl-tRNA beschikbaarheid, ii) toegenomen hydrofobiciteit van het eiwit leidt tot meer hydrofobe interacties, iii) de gistcel is niet in staat om gevormde stabiele triple helix structuren ($T_m$ van 74°C) te verwerken. Met een eiwit waarvan de eindblokken wat betreft aminozuursamenstelling identiek is aan bovengenoemd eiwit werd een normale opbrengst en secretie van het eiwit gevonden. Hieruit hebben we geconcludeerd dat bij het eiwit met collageenachtige eindblokken intracellulaire triple helix vorming optreedt wat secretie van het eiwit door de gistcel van het eiwit verhinderd. Gebruik van een Yapsin (protease) mutant van gist resulteerde in secretie van een intact eiwit product, de opbrengst was alleen nog veel lager dan voor de eiwitten met de kortere collageenachtige eindblokken. Om te bepalen waar het probleem zich voordoet in de uitscheiding van het eiwit zijn *P. pastoris* cellen nader onderzocht.

In Hoofdstuk 3, is elektronen microscopie uitgevoerd aan gist cellen om te zien waar triple helix vorming van de collageen achtige eiwitten zich voordoet. Expansie van het endoplasmatische reticulum (ER) en vorming van membraan vesicles werden alleen waargenomen wanneer triple helix vormende eiwitten met een hoge smelttemperatuur tot expressie werden gebracht in *P.pastoris*. De

*Samenvatting*

aanwezigheid van extra organellen kan het resultaat zijn van eiwit aggregatie ten gevolge van triple helix vorming of vergroting van ER volume is een poging van de cel om eiwit aggregatie te verminderen. Dit resultaat is een aanwijzing dat triple helix vorming vermeden moet worden om efficiënte secretie opbrengsten van deze eiwitten te realiseren.

In Hoofdstuk 4, hebben we laten zien dat de thermo-mechanische stabiliteit van de triple helix structuren gericht veranderd kunnen worden de lengte van (Pro-Gly-Pro) eindblokken te variëren. Een verhoging van de $T_m$ van deze op collageen gebaseerde eiwitten ontstaat door een toename van het aantal waterstofbruggen. Alle door ons bestudeerde tri-blok co-polymeren (eiwitten) vormen stabiele hydrogelen en toenemende lengte van de eind-blokken resulteert in een hogere stabiliteit tijdens mechanische stress. De variatie in thermostabiliteit kan verklaard worden uit een eenvoudige lineaire relatie tussen lengte van de eind blokken en de vrije energie van helix vorming. Een vergelijking met triple helices gevormd uit “vrije” $(Pro\text{-}Gly\text{-}Pro)_n$ peptiden suggereert dat de relatie tussen lengte en helix vorming niet wordt beïnvloed door samenstelling van de midden-blokken. Voor het gebruik van *P.pastoris* als “fabriek” voor de productie van deze eiwitten is duidelijk aangetoond dat, naar mate de thermostabiliteit van de triple helix toeneemt, de productie afneemt.

In Hoofdstuk 5 heb ik laten zien dat denatureerde SDS-PAGE gelen gebruikt kunnen worden om vorming van supramoleculaire structuren van op collageen gebaseerde eiwitten aan te tonen. Tijdens het ontkleuringsproces van de gel zullen eiwitten die geen triple helix kunnen vormen uit de acrylamide gel diffunderen, terwijl triple helix vormende eiwitten na overnacht ontkleuring nog steeds zichtbaar zijn in de SDS-PAGE gel. Verder is aangetoond dat eiwitten die een triple helix vormen met een hogere thermostabiliteit langzamer uit de gel diffunderen.

Tijdens dit onderzoek is op verschillende manieren geprobeerd om de secretie opbrengst te vergroten van op collageen gelijkende eiwitten met een hoge

*Samenvatting*

thermostabiliteit. Resultaten van deze experimenten hebben nieuwe inzichten gegeven over de wijze waarop *P. pastoris* als fabriek voor de productie van op collageen gelijkende eiwitten kan worden gebruikt. Experimenten voor dit verdere onderzoek worden in meer detail besproken in Hoofdstuk 6.

*Samenvatting*

# Acknowledgements

*Acknowledgements*

*Acknowledgements*

This PhD thesis has been quite a journey. When I look back, I remember it with a big smile because I had the good fortune of having the right travel mates by my side. They were always cheering me on and sometimes I even had the feeling they were actually carrying me to the finish line. Without certain people I would have never made it so far. To them I would like to say thank you.

On the scientific level I would like to thank my promoter Gerrit Eggink for the freedom and trust he gave me to conduct my research. To my daily supervisor, Frits de Wolf, many thanks for the always lively scientific discussions and for the critical revision of this manuscript, you helped putting my ideas in order and for that I am thankful. To Marc Werten and Antoine Moers I thank you for the insightful contribution to this thesis both practical and theoretically. Thank you Emil Wolbert and Jan Springer for your scientific input and skilled aid in the laboratory. Most importantly, thank you for the entertaining laboratory environment. You always made science seem fun. Thank you Jan Springer and Carmen Boeriu for your open door policy and always answering my last minute questions. Thank you to Prof. Ida van der Klei for receiving me in her group and helping me uncover the cell's mysteries.

To Paulina Skrzeszewska I would like to thank the help with the rheological measurements and scientific discussions. To Jasper van der Gucht I would like to thank the availability and readiness to answer my question on the proteins' mechanical properties.

To Prof. Hans Tramper, I would like to express my deepest appreciation. Hans you have made sure that this trip would end in the best possible way, with a graduation

*Acknowledgements*

party. Thank you for giving me the guidance and support I needed, you are a true Ph.D whisperer.

But I would never been able to put my scientific findings in paper if it weren't for the mental support of my family and friends.

Thank you to Sabina and Ishay for adopting me. For allowing me to make your house my own and for trusting me with your most precious belongings, your children and dog. To Ben, Julia and Rafi Rebe Raz thank you for being such great play mates and sharing your Tom and Jerry episodes with me after a hard day's work/play. Ben, thank you so much for the Dutch lessons and making sure the babysitter knew what to do. Although, I will not forget the time you told me that your mother allowed you to eat 4 donuts on a row ;). Julia, thank you for sharing (once in a while) your princess crowns and dolls with me. I now have a crown of my own so we don't need to fight over it anymore. Rafi, your timing was perfect.

Thank you Baci, for the long walks, warm welcome home and Hebrew command words. It was always fun to see people's faces when I said "Die Baci, Die".

To the girls, Anastasia, Sabina, Barbara, Elsa and Aldana I would like to thank the wonderful time spent in Wageningen. You made this city seem like home to me and for that I am eternally grateful. Anastasia and Sabina, thank you for the lunch breaks at the Futurum. Not only was a great way to relax during the day but it was also the place where great ideas were born. Elsaki, minha fada do lar, obrigada pelos cafézinhos, bolinhos e brunchs matinais. Obrigada por me teres recebido em tua casa e me teres ajudado nesta fase final. És a personificação de uma casa Portuguesa: "com uma promessa de beijos e dois bracos a minha espera".

*Acknowledgements*

Davidinho, muito obrigada por seres quem és. Sem ti isto não teria sido tão divertido. Obrigada pela camaradagem, pelas festas, pelas alegrias e pelas risadas. Um muito obrigada pela sabedoria Colombiana. Contigo aprendi que o que se leva desta vida é mesmo o que se come, o que se bebe e o que se baila ☺

Charla, chica!!!!!. Gracias por tudo. Por la amistad, cenas, buenisima musica y todavia mejores conversaciones…… Te quiero muchísimo y espero verte pronto en tu amado México.

Jerome, merci beaucoup for the wonderful food and great conversations. With you Aldana and Charla I learnt that sharing a good meal and wine is a spiritual experience and something to be cherished. Thank you for the quiche recipe which I hope one day to cook for you…. just don't really know when….

Laurix, Leilocas, Leninha e Evinha obrigada por terem partilhado comigo esta montanha russa. Como bem sabem sem a vossa ajuda não teria conseguido. Obrigada pelas longas conversas, pelos high-teas em Utrecht, enfim, por tudo o que foi e que continuará a ser ☺. Vocês serão sempre tripulantes desta minha viagem.

Patricinho, dank je wel for welcoming me into your country. You always made sure I would not be overwhelmed by Dutch bureaucracy and that I took care of all the paperwork properly. Thank you for the nice Dutch food (you actually made me enjoy stamppot) and for teaching me (over and over again) how to change a flat tire. But most of all, thank you for your friendship.

Arno, thank you so much for the beautiful art work and your capacity to draw exactly what I had in mind. Thank you for having the unbelievable patience to read

*Acknowledgements*

my thesis and correct it. Most of all, thank you for giving me the inspiration I needed to finish this thesis, promising me a trip to Kenya (btw, I'm still waiting!).

To Billie Holiday, Sarah Vaughan, Chavela Vargas and Concha Buika, thank you for the company during the late hours of writing.

Aos meus pais, irmãos e avó gostaria de agradecer o apoio e amor incondicional. Sei que a distância tem sido difícil e que gostariam que estivesse perto, mas tambem sei que entendem que as asas são para voar. Mãe e Pai, obridada por nos terem ensinado que se deve lutar pelo que se acredita.

To my beautiful paranymphs, Helena Teles and Aldana Ramirez thank you for keeping me motivated throughout this journey. Especially for telling me what I needed to hear and not always what I wanted to hear. Aldana, thank you for taking me in when times were rough. If you hadn't given me shelter I would have never had the time nor the peace of mind to find what I was looking for. I hope one day you will let me do the same for you. Thank you for the wonderful trips that have shaped our lives. Helena, you always put up with my incessant discussions about the content of this thesis, you where my toughest reviewer because you were always right. Thank you so much for keeping me on the right and realistic track when it comes to science. I find it quite remarkable that such a spiritual person can be also so much salt-of-the-earth at the same time. Thank you for showing me that every trip is both physical and spiritual.

Thank you to the University of Wageningen for bringing all these people into my life.

*Curriculum vitae*

Publications

Training Activities

*Curriculum vitae*

# *Curriculum vitae*

Catarina I. F. Silva was born in Leiria, Portugal, on the 24th of April, 1980. Upon completing her secondary education at Escola Secondária Francisco Rodrigues Lobo (Leiria) in 1998 she moved to Porto and initiated a degree in Biology at Universidade do Porto. After 1 semester she decided to change and moved to Lisbon to study Biological Engineering at Instituto Superior Técnico (IST), Technical University of Lisbon (UTL). In 2005, Catarina got an ERASMUS scholarship to finish her M.Sc studies at Technical University of Denmark (DTU). Her M.Sc thesis, entitled "The impact of two *Bacillus subtilis* UDP-Glucose Dehydrogenase on Exopolysaccharide", was carried out at the Microbial Physiology and Genetics Group. After completion of her MSc, she received a scholarship from the Portuguese Ministry of Science to engineer *Lactococcus lactis* for the production of Mannitol at the Cell Physiology and NMR group at Instituto de Tecnologia Química e Biológica, Portugal. Immediately after, she started her Ph.D at Bioprocess Engineering at Wageningen University. The research goal was the optimization of *Pichia pastoris* as a cell factory for the secreted production of collagen-like proteins with the ability to self-assemble into triple helices. The outcome of her research can be found in this thesis. In parallel with her Ph.D program, Catarina attended the Shenzhou Open University of Traditional Chinese Medicine, in Amsterdam, where she followed a 3 year program of Acupuncture studies. On the 16th of March, Catarina started working as a Postdoctoral fellow at the Royal Netherlands Institute for Sea Reasearch, NIOZ, where she is studying the influence of light on the coral microbial community.

*Curriculum vitae*

*Publications*

# *Publications*

**Catarina I.F. Silva**, Gerrit Eggink, and Frits A. de Wolf, *Anomalous behavior of recombinant triple-helix forming proteins in SDS-PAGE*, submitted.

**Catarina I.F. Silva**, Paulina J. Skrzeszewska, Monika D. Golinska, Marc W. T. Werten, Gerrit Eggink, and Frits A. de Wolf, *Tuning of Collagen Triple-Helix Stability in Recombinant Telechelic Polymers*, Biomacromolecules, 2012, *13* (5), 1250–1258.

**Catarina I.F. Silva**, Helena Teles, Antoine P. H. A. Moers, Gerrit Eggink, Frits. A. de Wolf and Marc W.T Werten, *Secreted production of collagen-inspired gel-forming polymers with high thermal stability in Pichia pastoris*, Biotechnology and Bioengineering, 2011, 108 (11), 2517-2525

Dina Petranovic, Ole Michelsen, Ksenija Zahradka, **Catarina Silva**, Mirjana Petranovic, Peter R. Jensen, Ivan Mijakovic, *Bacillus subtilis strain deficient for the protein-tyrosine kinase PtkA exhibits impaired DNA replication***,** Molecular Microbiology**,** 2007, 63(6), 1797-1805.

*Publications*

*Training activities*

# *Overview of completed training activities*


**Discipline specific activities**

| | |
|---|---|
| Bioinformatics: A user's approach | 2007 |
| Advance course in *P. pastoris* technology | 2007 |
| Gene transcription in yeast | 2008 |
| Visualization of cellular processes | 2008 |
| Sustainable process, product and system design | 2009 |
| B-Basic symposium | 2007/08/09/10 |
| PEGS Symposium | 2008/09 |

**General courses**

| | |
|---|---|
| VLAG Ph.D week | 2006 |
| Techniques for writing and presenting a scientific paper | 2006 |
| Innovation presentation | 2007 |
| Leo Petrus innovation Challenge | 2009 |

**Optionals**

| | |
|---|---|
| Bioprocess Engineering brainstorm days | |
| Genomics researchers event | 2007 |
| Bioprocess Engineering Ph.D study tour to Japan | 2008 |
| Brain&Games | 2008 |
| IRSR symposium | 2010 |

Art work by Arno Kang’eri

Printed by Ridderprint BV