CHAPTER 4

Recombinant Production of Self-Assembling Peptides

Michael J. McPherson[1,2,*], **Kier James**[1], **Stuart Kyle**[1], **Stephen Parsons**[2], and **Jessica Riley**[1]

Contents

[1] Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

[2] Centre for Plant Sciences, University of Leeds, Leeds LS2 9JT, United Kingdom

[*] Corresponding author.
E-mail address: m.j.mcpherson@leeds.ac.uk

Advances in Chemical Engineering, Volume 35
ISSN 0065-2377, DOI: 10.1016/S0065-2377(08)00204-4

1. INTRODUCTION

Self-assembly is ubiquitous in Nature; it describes the spontaneous association of molecular building blocks into coherent and well-defined structures without external instruction (Zhang, 2002). One aspect of bionanotechnology encompasses the synthesis of novel biomaterials based on an understanding of natural self-assembly. Proteins and peptides are of particular interest due to the number of their constituent amino acids, allowing the generation of a highly diverse family of biomolecules. Nature has taken full advantage of this diversity to produce a vast number of different structural proteins with well known examples including silk and collagen (Mitraki and van Raaij, 2005). Examples of structures that involve biomolecules complexed with other organic or inorganic components also exist, including bones (Giraud-Guille, 1988) and teeth (Paine and Snead, 1997). In all these cases Nature uses a "bottom-up" approach by building molecular assemblies atom by atom, molecule by molecule, and macromolecule by macromolecule. The concepts of this self-assembling procedure are well documented but are not easy to replicate (Zhang, 2003).

A fundamental principle of engineering molecular self-assembly is to manufacture the appropriate molecular building blocks that are capable of spontaneous stepwise interactions to generate larger assemblies. A key requirement for industrial scale implementation is our capacity to scale-up systems for the production of these molecular building blocks. Much of the work on self-assembling peptides has been based on chemically synthesized peptides. While useful in the discovery phase of identifying self-assembling sequences, this approach is costly, requires hazardous chemicals, and the size of peptide that can be synthesized is limited (Vandermeulen and Klok, 2004). To produce peptides on a larger, and ultimately an industrial scale, biological systems are being explored using recombinant DNA technology. Potential host organisms, include bacteria, yeasts, fungi, animal and plant cells, transgenic animals and plants. In this review we focus upon microbial and plant-based expression systems. First we consider general features of systems used to express recombinant proteins including the requirements for effective transcription and translation leading to production of the recombinant protein. We then consider aspects of protein recovery and the uses of various fusion and purification tags as well as mechanisms for their removal from the final product. These general issues are of importance for any expression system, although some specific details will vary depending upon the host organism.

2. RECOMBINANT PROTEIN PRODUCTION

2.1. Overview

Recombinant DNA technology allows a DNA sequence encoding the protein/peptide of interest to be cloned into an organism which acts as a

"biofactory" to produce substantial quantities of the recombinant product. The DNA coding sequence for the protein/peptide, which is often custom designed and manufactured synthetically, is cloned into an expression vector which provides the necessary regulatory signals for efficient expression of the protein by the host organism. The key components of an expression vector are (a) a promoter and terminator to direct transcription and termination of the mRNA, (b) a selectable marker, often for an antibiotic resistance trait to ensure retention of the vector by host cells, and (c) a multiple cloning site allowing the easy directional insertion of the coding region to be expressed. In addition, the coding region must also have a suitable ribosome binding region and an initiating codon (ATG). The precise origin of these components will depend upon the host organism as regulatory signals and selectable markers vary from host to host.

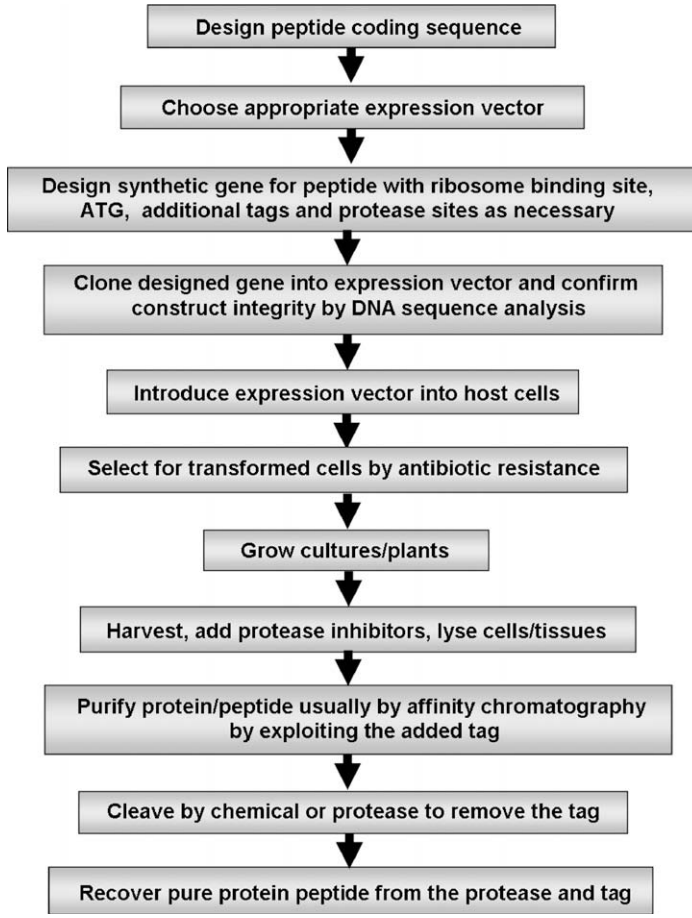
The resulting expression construct must then be introduced into the host cells by some form of chemical treatment, electroporation, or in plants biolistics, agrobacterium-mediated transfer or the use of viral infection. Following introduction of the expression vector DNA, transformed cells are grown in the presence of an appropriate antibiotic, which selects cells expressing the selectable marker protein. In bacteria and some yeasts as well as during transient expression in plant cells, the expression vector remains as an autonomous plasmid within the cell. In other cases, including some yeasts, fungi and stable plant lines the vector becomes integrated into the host genome. An outline of the processes for generating cells that express a recombinant protein and for subsequent recovery of the target recombinant product is shown in Scheme 1.

In addition to the basic features of an expression vector outlined above, other considerations include methods for inducing expression of the target gene, the subcellular location to which the protein/peptide is targeted, and fusion proteins or tags that can either enhance expression or solubility and/or facilitate detection, and purification of the recombinant product.

2.2. Expression and protein recovery

2.2.1. Transcription

Generally a key objective is to maximize the level of expression of the transgene to maximize the level of recombinant protein recovered. To achieve this, a promoter that can direct efficient transcription of the target DNA to produce the corresponding mRNA should be selected. Since recombinant products can sometimes be toxic to the host cells, it is common to employ an inducible promoter. This is normally a promoter taken from a gene that is highly expressed in the host organism, but which, under normal conditions, is either repressed by a repressor protein or is activated by an activator protein. The addition of a chemical inducer then either relieves repression or results in activation of the transcription process. During normal cell growth, the promoter is therefore inactive allowing cells to grow



Scheme 1 Outline of steps necessary for cloning a target-coding region, introducing the expression construct into the host, expression and product purification including removal of any fusion partner.

efficiently. At an appropriate time the inducing chemical is added to the cells so that they then begin to express the target protein/peptide.

2.2.2. Translation

The context of the mRNA sequence around the AUG codon translational start site, where the ribosome initiates translation of protein synthesis, is an important consideration. An appropriate consensus ribosome-binding region for the host organism should be used and potentially inhibitory RNA secondary structure, that may affect the ability of the ribosome to access this translation start site, avoided. Various RNA structure prediction programs such as

RNADraw (Matzura and Wennborg, 1996) can be useful in this regard. A method to alleviate this problem is to engineer silent mutations into the first few codons of the mRNA to reduce the stability of such RNA structures (Ivanovski et al., 2002). Translation of the protein by the ribosome relies upon the decoding of the mRNA in nucleotide triplets called codons. The frequency of specific codons in an organism reflects the availability of aminoacyl-charged tRNAs. With the exceptions of the amino acids Met (M) and Trp (W), all amino acids can be encoded by multiple codons, with Ser (S), Leu (L), and Arg (R) each having six possible codons. Different organisms, therefore, preferentially use different subsets of codons and so a coding region taken from one organism may not be appropriate for efficient translation by another host organism due to their differences in codon usage. For *Escherichia coli*, certain strains are available that contain additional copies of normally poorly represented tRNAs, and so these strains (such as CodenPlus* strains, Stratagene) can be useful for expressing heterologous sequences. However, the simplest approach, especially for short coding sequences such as those for peptides, is to design a DNA sequence that encodes the necessary amino acid sequence but which is optimized for the codon preference of the host organism, and to have this synthesized as an artifical gene. Codon optimization thus ensures that the availability of tRNAs does not reduce the efficiency of translation of a target protein. Coding sequences can be optimized for many organisms based on known genome sequences and useful web-based programs can be used to design optimized coding regions, for example, JAVA Codon Adaptation Tool (http://www.jcat.de/). Alternatively the commercial companies who provide synthetic genes can optimize the codon usage for a specificed host.

2.2.3. Protein recovery and purification

If a protein/peptide is secreted into the culture medium then it must be recovered from a dilute solution. This could involve precipitating the protein/peptide under nondenaturing conditions, such as using ammonium sulfate and resolubilizing in a smaller volume. Alternatively, a more convenient method is through the use of affinity materials (see below) with an appropriate matrix added to the supernatant to allow batch binding of the target protein.

Commonly, it is necessary to recover the protein from within host cells. This involves harvesting, usually involving some centrifugation or filtration step. Subsequently, the cells must be lysed, which can be done by physical approaches such as sonication or pressure systems including French press or cell disrupters. Simpler methods involve the use of detergent mixtures which can allow recovery of both soluble proteins in the supernatant and of inclusion bodies (Lee et al., 2004). Inclusion body proteins can also be recovered by other whole cell dissolution methods (Choe and Middelberg, 2001; Falconer et al., 1999). It is common to add protease inhibitor mixtures at the lysis stage to prevent damage to the target protein/peptide by proteases released from the host cells during the lysis process.

Subsequent protein purification usually relies upon the presence of an appropriate fusion tag incorporated at the stage of expression vector construction. Thus, the recombinant protein/peptide has an integral purification handle, in theory facilitating the rapid isolation of substantial amounts of recombinant product. Some purification tags require the use of specialist and expensive affinity reagents, and even though a substantial level of purification may be achieved, the protein/peptide may still be contaminated with other proteins and biomolecules. Alternatively, two tags could be used allowing an independent purification step, but this obviously results in increased processing times and cost. A common purification tag is a 6–8 histidine or His (H)-tag added at either the N- or the C-terminus of the protein. The cell extract containing the target protein/peptide can then be subjected to metal chelation chromatography, often using nickel or cobalt-charged affinity matrix. However, some endogenous host proteins often show a degree of affinity for such matrices, which in any case are rather expensive. An obvious approach is to use cheaper protein-binding materials such as cellulose which has been used, but not exploited to the extent one would expect (Reed et al., 2006; Rodriguez et al., 2004). Table 1 provides a summary of various protein fusion components and the method of purification.

It is usually desirable to design the expression construct so that the protein/peptide can be separated from the tag that may otherwise interfere with biological or physical function; this can be particularly important in the production of therapeutic reagents. The cleavage mechanism may be chemical or enzymatic. Unfortunately in many cases some additional amino acids remain associated with the target protein/peptide. For example, cyanogen bromide (CNBr) is a chemical that cleaves at peptide bonds C-terminal to methionine residues. This leaves a homoserine lactone residue at the C-terminus (Kaiser and Metzka, 1999). This is not ideal if trying to produce a therapeutic peptide sequence, although it could be useful for processes such as surface immobilization. However, CNBr is highly toxic and therefore may be undesirable for commercial use (Plunkett, 1976).

An alternative to chemicals for cleaving protein/peptides from fusion partners is enzymatic cleavage. Proteases can have high sequence specificity or substrate structure specificity although it can be difficult avoiding the incorporation of unwanted residues in the final product. Proteases tend to cleave towards the C-terminal end of their cleavage site, and therefore, it is advantageous to design constructs in which the protease site precedes the target protein/peptide. In this way often only one or two residues are added to the N-terminus of the product. Depending upon the protease and recognition sequence it may be even possible to produce natural N-terminii. By contrast, when the cleavage site is located at the C-terminal end of a protein, 5–6 additional residues from the recognition site can remain associated with the target protein. A widely used protease which can itself be produced recombinantly in high yield is tobacco etch virus (TEV) protease (Blommel

Table 1 Recombinant proteins/peptides used as fusion partners for recombinant protein/peptide expression

A. Fusion partners for recombinant protein/peptide purification

Fusion partner	Purification method	Size (kDa) or aa (sequence)
Glutathione-S-transferase (GST)	Glutathione-Sepharose	26 kDa
Maltose-binding protein (MBP)	Amylose resin	40 kDa
Thioredoxin	ThioBond resin	109 aa; 11.7 kDa
Chitin-binding domain	Chitin	~110 aa
Cellulose-binding domain	Cellulose	~110–160 aa
Protein A	IgG	14 kDa
FLAG™ peptide	Monoclonal antibody	7 aa (DYKDDDDK)
Poly-arginine	Cation exchange	6 aa (RRRRRR)
S.Tag	S.protein (RNAse)	15 aa (KETAAAKFERQHMDS)
His-tag	Divalent metals (Ni^{2+} or Co^{2+})	6 or 8 aa (HHHHHHHH)
StrepTag II	StrepTactin	8 aa (WSHPQFEK)
MAT-tag	Divalent metals (Ni^{2+} or Co^{2+})	7 aa (HNHRHKH)
AviD-tag	Avidin or Streptavidin	6 aa (DRATPY)
T7-tag	Monoclonal antibody	MASMTGGQQMG

Table 1 (*Continued*)

B. Fusion partners for other purposes		
Fusion partner	Property	Size (kDa) or aa (sequence)
Ketosteroid isomerase (KSI)	Produces insoluble fusion protein	125 aa
T7gene10	Produces insoluble fusion protein	260 aa
GFP	Fluorescent reporter protein	220 aa
NusA	Enhanced solubility	495 aa (54.8 kDa)
Ubiquitin	Enhanced solubility	76 aa
Sumo	Target for SUMO protease	117 aa
Ruby tag	Red reporter protein	53 aa (6.1 kDa)

Table 2 Approaches for specific cleavage of recombinant proteins

Cleavage agent	Cleavage specificity
Enzymatic	
Thrombin	-LVPR▼GS-
Factor Xa	-IDGR▼X-
Enterokinase	-DDDDK▼X-
Endoproteinase GluC	-E▼X-
Endoproteinase Lys-C	-K▼X-
TEV protease	-ENLYFQ▼(S,G)-
HRV 3C protease	-LEVLFQ▼GP-
Sumo protease	SUMO-GG▼XXX-
IGase	-PP▼YP-
Furin	-RX(R/K)R▼X-
Chemical	
Cyanogen bromide	-M▼X-
Formic acid	-D▼X-
Hydroxylamine	-N▼G-
2-nitro-5-thiocyanobenzoate (NTCB)	-C▼X-
1-cyano-4-dimethylaminopyridium tetrafluoroborate (CDAP)	-C▼X-
3-bromo-3-methyl-2-(2-nitrophenylthio)-3H-indole	-W▼X-
2-iodosobenzoic acid	-W▼X-

Amino acids are shown in single letter code; X is any amino acid; ▼denotes the cleavage site; - indicates the remainder of the protein/peptide chain.

and Fox, 2007; Dougherty and Parks, 1989; Dougherty et al., 1989; Fang et al., 2007). Table 2 provides examples of chemical and protease cleavage approaches.

3. HOST ORGANISMS FOR RECOMBINANT EXPRESSION

3.1. Bacteria

A range of bacteria have been used for heterologous protein expression (Terpe, 2006). *E. coli* is a popular bacterial expression host due to (a) its rapid growth rate on relatively simple growth media (doubling time ~20 min), (b) its well-characterized genome and genetics resulting in a wide range of expression host strains, and (c) the availability of a range of expression vectors. Yields of heterologous proteins can be potentially very

high as in the case of a repeat peptide system reported at 65% total cellular protein (Metlitskaia et al., 2004), although this is exceptional. *E. coli* can be grown in a fermenter, but a major advance for laboratory study is the recent introduction of the autoinduction process which can provide high cell-density cultures in shake flasks (Studier, 2005).

3.1.1. Autoinduction provides a new tool for high-density cultures of bacteria

Many inducible promoters in *E. coli*-based expression vectors are regulated by the lac operator/repressor system. Traditionally, the lac system has been induced by the addition of the expensive lactose analogue isopropyl-β-D-thiogalactopyranoside (IPTG) to the cell culture during mid-log phase when the cell density has reached a typical value of around 0.6, measured as optical density at 600 nm (OD_{600}). Under such condition the final OD_{600} of the culture would perhaps reach a value of 2–3. Autoinduction relies upon the natural induction of the lac system by conversion of lactose in the growth medium to allolactose. However, the growth medium also contains glucose, the preferred carbon source of the cells which allows initial growth under noninducing conditions. Following glucose depletion, a metabolic switch occurs to allow the cells to utilize lactose simultaneously inducing recombinant protein expression. The benefits of this approach are (a) no monitoring of cell growth to assess when to add inducer, (b) no addition of expensive IPTG is required, and (c) cultures reach very high cell densities, typically OD_{600} 20–30 resulting in higher potential yield of recombinant protein. The levels of cell density achievable by autoinduction in shake flasks are comparable with those that can be achieved by some fermentation processes. Further enhancements to the autoinduction system resulting in enhanced yields have recently been reported (Blommel et al., 2007).

3.1.2. Targeting protein production in bacteria

E. coli, a Gram-negative bacterium has both an inner and an outer cell membrane. Typically, proteins are expressed as cystoplasmic protein. The addition of a signal sequence to the N-terminus of a protein may allow it to be secreted into the periplasm: the compartment between the two membranes. Secretion into the periplasm represents a purification step as the cells separate the target protein from the cytoplasmic components. An example of a protein successfully secreted from *E. coli* is hirudin, a 65-amino acid anticoagulant peptide/protein produced in the salivary glands of medicinal leeches. Using an L-asparaginase II signal sequence around 60 mg/L were obtained (Tan et al., 2002).

Furthermore, other bacteria have been shown to have superior innate abilities to secrete proteins including *Bacillus subtilis*, a Gram-positive bacterium, which lacks membrane proteases. This indicates that prokaryotes could

be viable expression hosts for larger scale production of self-assembling peptides (Olmos-Soto and Contreras-Flores, 2003). Further evidence has been provided by Schmidt (2004) who has presented data on high levels of expression of proteins indicating achievable yields in the order of 1–2 g/L.

Even within the cytoplasm there are, in effect, two compartments. Soluble proteins accumulate in the cytoplasm, but for a number of heterologous proteins the lack of appropriate intracellular chaperones to prevent misfolding can often result in such proteins aggregating as insoluble inclusion bodies. Recovery of protein from inclusion bodies requires solubilization in a strong chaotropic agent such as 8M urea or 6M guanidinium hydrochloride. Usually some renaturation step is required for restoring normal structural and functional properties but this recovery process is not straightforward and can result in poor yields of functional protein (Marston and Hartley, 1990; Rai and Padh, 2001). However, in the case of peptides which do not form 3D structures, targeting to inclusion bodies can serve as a useful purification system. A number of fusion partners have been specifically used to direct recombinant peptides to inclusion bodies (discussed in Sections 5 and 6). Indeed, inclusion body localization can be important for protecting the peptides from degradation, and to prevent them exerting any toxic effects upon the host cell, for example, in the production of antimicrobial peptides.

3.1.3. Ralstonia eutropha

Ralstonia eutropha is a Gram-negative bacterium that has substantial capacity for bioremediation and is found in soil and water. This bacterium is being increasingly used as a host for recombinant protein production. For example, a soluble active organophosphohydrolase (OPH) was produced at levels greater than 10 g/L (Barnard et al., 2004). Previously, this enzyme formed inclusion bodies when heterologously expressed in *E. coli*. This high level of expression was achieved by randomly intergrating three copies of the gene into the bacterial chromosome under the control of a strong inducible promoter. It is likely to have good potential for recombinant protein production, but currently suffers from a lack of genetic tools particularly for development of autonomously replicating plasmids, which could be used as expression vectors.

3.1.4. Disadvantages of bacteria

Although bacterial hosts can be useful for protein production, *E. coli* and other prokaryotic expression systems have some limitations. For example, they are incapable of posttranslational modifications such as glycosylation and as a consequence many eukaryotic proteins produced in prokaryotic expression systems are nonfunctional (Rai and Padh, 2001). However, in the case of most peptides, including self-assembling peptides, such modifications are currently not required. Of course, this limitation may have implications for future

attempts to functionalize, or generate composite materials based on recombinant self-assembling peptides. Furthermore, a major issue with respect to downstream processing of *E. coli*-derived products is the potential release of endotoxins, derived from the cell wall lipopolysaccharides. Removal of these from the protein product is particularly important where the peptides or proteins are to be used for medical applications.

3.2. Yeasts

3.2.1. Advantages of yeasts

The yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* are useful expression hosts and others such as *Kluyveromyces lactis* are being developed (van Ooyen et al., 2006). Yeasts are popular for expression of recombinant eukaryotic proteins as levels often exceed those of other eukaryotic expression hosts such as mammalian and insect cells. As yeasts are eukaryotes they are capable of posttranslational modification of proteins such as *N*- and *O*-linked glycosylation, and can ensure correct folding and disulfide bond formation (Cereghino and Cregg, 2000; Cregg et al., 2000; Higgins and Cregg, 1998). Yeasts exhibit some of the advantages of prokaryotic expression systems such as a rapid doubling time and easy culturing to high cell densities on simple media (Rai and Padh, 2001). They are capable of secreting proteins to high levels given a suitable signal sequence. In addition, the genetics of commonly used species is relatively well understood. It has also been reported that yeasts are capable of coping with repetitive gene sequences, an important consideration for many peptides and structural proteins (Strausberg and Link, 1990).

On an industrial scale, safety is also an important consideration. Some yeast species have already been used extensively in the brewing and food industries, and as a result, these have been classified as generally recognized as safe (GRAS) by the US Food and Drugs Administration. Consequently, the guidelines for their use are less stringent, potentially reducing start-up costs. In addition, yeasts are good candidates for fermentation with appropriate fed-batch strategies in a bioreactor having the potential to greatly enhance the levels of target protein production. An example illustrating this is the production of 600 mg/L of human interferon α-2b in high cell-density cultures (Ayed et al., 2008). In the same study, exchanging and supplementing the medium also allowed protease degradation issues to be overcome. Evidence for large-scale protein production by a yeast has also been provided by Yamawaki et al. (2007) who concluded that continuous culture of *P. pastoris* with methanol feeding by a concentration control method held good industrial scale potential for recombinant protein production. Using this method yields of 810 mg/L of a single chain antibody (scFv) were achievable compared with only 198 mg/L by a dissolved oxygen control method, despite the methanol consumption rates being the same under both conditions.

The highest reported heterologous production of a protein in yeast comes from the work by Hasslacher et al. (1997). The enzyme hydroxynitrile lyase (Hnl) from the tropical rubber tree *Hevea brasiliensis* was reported to produce levels of 22 g/L intracellularly in *P. pastoris*. In the same study, *S. cerevisiae* and *E. coli* were tested in parallel experiments but were not competitive. Levels of proteins produced by yeasts have more typically been in the range of 1–15 g/L (Schmidt, 2004). A list of reported yields for expression of proteins in *P. pastoris* is provided by Cregg at http://faculty.kgi.edu/cregg/index.htm

3.2.2. Disadvantages of yeasts

There are also some disadvantages of yeasts. For instance, the procedures required for DNA transfection and lysis of yeasts are more difficult than for bacteria. They can also produce substantial quantities of proteases (Schmidt, 2004), requiring strategies to reduce the impact of proteolytic degradation on cellular, vacuolar, and secretory proteins (Zhang et al., 2007). There is considerable variation in the patterns of glycosylation between eukaryote species which can result in *O*- and *N*-linked glycosylation patterns that are considerably different to those of the native protein (Kukuruzinska et al., 1987). This glycosylation may not only have an impact on the activity of the protein but could also affect folding and secretion particularly if the protein is hyperglycosylated, a problem particularly associated with *S. cerevisiae* (Schmidt, 2004).

3.3. Filamentous fungi

3.3.1. Advantages of fungi

Fungi are potentially useful expression hosts. In contrast to yeasts fungi are capable of more complex posttranslational modifications with some proteins being produced with their natural glycosylation patterns, a good example being tissue plasminogen activator (t-PA) produced by *Aspergillus nidulans* (Schmidt, 2004). Their extensive use in the food and beverage industries has resulted in many species of the Aspergillus and Penicillium genera being granted GRAS status (Iwashita, 2002). Fungi offer excellent protein secretion potential with many species having the innate ability to secrete homologous proteins such as cellulases and amylases at levels of 30–40 g/L. As with yeast and bacteria, fungi also offer the potential of inducible gene expression systems. For example, the presence of pH regulatory systems such as that in *A. nidulans* where the *pacC* gene encodes a transcription factor and the *palA*, *palB*, *palC*, *palF*, *palH*, and *palI* genes encode components of a pH signal transduction pathway (Denison, 2000). This is a particularly appealing mechanism for controlling the expression of genes in fungi due to their ability to grow and thrive over a wide pH range (Arst and Penalva, 2003).

3.3.2. Disadvantages of fungi

Despite the potential for very high levels of secretion of homologous proteins, replicating these levels with heterologous proteins from other sources has yet to be realized. Further problems include the difficulty in transfecting fungi due to the degradation of foreign DNA and random genomic integration that can affect endogenous genes. Expression rates are also restricted by RNA instability, incorrect folding of mRNA, and incorrect folding and secretion of the resulting protein, which is influenced by glycosylation. If the resulting proteins are incorrectly glycosylated and/or secreted, then degradation by the host is often rapid (Punt et al., 2002). For fungal expression systems to be fully exploited, a better understanding of fungal genetics and physiology is required.

3.4. Transgenic plants

Transgenic plants are potent protein bio-factories. The ability of plants to produce complex heterologous proteins was demonstrated when a monoclonal IgM antibody was expressed and fully assembled in transgenic tobacco (During et al., 1990). The first peptide to be produced in a transgenic plant was cecropin B, a small antibacterial peptide from the giant silkmoth *Hyalophora ceropia* (Florack et al., 1995). It proved to be extremely sensitive to the plant endogenous proteases and was degraded within seconds in various plant cell extracts. Later, an analog carrying a valine (V) to methionine substitution was compared for protease degradation and was found to show a sixfold lower degradation rate indicating that the peptide sequence will significantly influence levels of degradation.

There are a number of examples of the development of plants for pharmaceutical protein production including antibodies, vaccines, and other bioactive proteins (Daniell et al., 2001; Daniell, 2006; Ma et al., 2005a, b, c; Twyman et al., 2003) as well as consideration of the issues surrounding regulatory issues of the use of transgenic plants for pharmaceutical protein applications (Ma et al., 2005b; Sparrow et al., 2007; Spok, 2007).

A few well-established crop species are preferentially used for transformation experiments, including tobacco, rice, corn, and soybean. From a therapeutic perspective producing pharmaceutical protein in crops that can be eaten raw, such as banana, has advantages particularly in developing countries (Ma et al., 2005a, c). Protein yield is an important factor, and so, plants that produce high biomass and high quantities of total protein are of particular interest. Soybean has a total protein content of 40%, which is higher than that of other crops meaning fewer plants are needed for the required protein quantity reducing handling costs and downstream processing (Kusnadi et al., 1997). A problem with soybean and many other crop plants is that they are food crops. In some regions, such as the EU, media attention has focussed on the possible contamination of traditional food lines with transgenic material in the U.S.A.

(Fox, 2003). Although there are arguments in favour of the use of crop plants, such as corn, for molecular farming purposes (Ramessar et al., 2008). By contrast, tobacco is a nonfood and nonfeed crop and so is not consumed as food by either humans or farm animals, thus transgenic materials could not inadvertently enter the food chain. Tobacco produces over 40 tons of leaf fresh weight per acre and also produces high seed quantities of up to one million seeds per plant (Daniell, 2006). Therefore, a single plant line can be very rapidly scaled-up which makes tobacco a good candidate for commercial scale recombinant protein production. As a result of these factors, tobacco is the most popular species for plant transformation with more transgenes being introduced into tobacco than into all other plant species combined (Daniell, 2006).

Four basic strategies for recombinant protein expression in plants are transient expression, stable nuclear transformed plants, chloroplast transformed plants, and suspension cultures derived from stable transgenic lines.

3.4.1. Transient expression

This can be achieved using infiltration of whole plants or plant parts (e.g., leaves) with *Agrobacterium tumefaciens* (Boehm, 2007; Sparkes et al., 2006; Vaquero et al., 1999). The transfer of T-DNA into plant cells occurs but in the absence of selective pressure there is no chromosomal integration of the T-DNA. After 2–3 days transient expression levels of the target protein can be examined by extraction of leaf protein and detection of the protein by Western blot analysis. This can be useful for assessing the most appropriate constructs to use for the development of stable transgenic lines, and as such is an analytical rather than production level system. A second transient method involves fusing the target gene to the coat protein gene of tobacco mosaic virus together with an endoplasmic reticulum retention signal, and using this construct for plant infection and propagation of the target protein (Spiegel et al., 1999). Due to the efficiency of viral infection, it is possible to generate substantial levels of recombinant proteins. For example a vector, based on a deconstructed TMV called a MagnICON viral vector, has been used to express hepatitis B surface antigen for use as a vaccine (Gleba et al., 2007). Using this system yields of nearly 300 μg/g leaf tissue in a tobacco species were obtained. This process could also be scaled up by the use of a leaf vacuum infiltration approach (Huang et al., 2008).

3.4.2. Stable nuclear transformed plants

The generation of stable transgenic plants normally involves Agrobacterium-mediated transfer of T-DNA into the plant cell where it undergoes random integration into the nuclear genome. Selective chemical pressure is applied, often hygromycin or G418, to select transformants expressing the selectable

marker protein. Due to its random nature, integration may occur in essential functional genes, whose disruption can lead to abnormal phenotypes, or alternatively into transcriptionally silent regions reducing transgene expression. It is therefore necessary to generate a large number of independent transgenic lines, to self-pollinate these and to select for stable integrants, and eventually homozygous lines. Screening of the lines for expression of the transgene, over several generations, to assess stability of the expression profile is then required. As a consequence, a significant time (6–12+ months) is necessary for the generation and characterization of suitable lines. However, once established these provide an almost unlimited and sustainable production capacity by standard farming practice.

Proteins produced in plants can be targeted to various subcellular organelles or cellular locations through the addition of suitable leader or retention sequences. For example, an N-terminal signal sequence plus C-terminal KDEL or HDEL signals direct and retain the protein in the endoplasmic reticulum (Gomord et al., 1997). Other locations are the apoplast, the vacuole, although this environment may be unsuitable due to its low pH and high proteolytic activity, the mitochondria, peroxisomes, and chloroplasts. Proteins can be expressed constitutively (using, for example, the CaMV 35S promoter or by use of appropriate spatially restricted promoters, can be expressed in specific plant organs).

Transgenic plant systems have the potential to produce recombinant proteins on a commodity scale (Kusnadi et al., 1997) due to the low cost of growing plants and because scale-up of production simply requires sewing seeds over a greater field area. As such they offer almost unlimited scalability (Giddings, 2001). It is estimated by Kusnadi et al. (1997) that transgenic plants can produce pharmaceutical proteins at between 10 and 50-fold lower cost than microbial fermentation systems, and 1,000 times lower than mammalian cell culture systems (Hood et al., 2002).

The process of protein production must be optimized for maximum yield while allowing the plant to function correctly without growth being adversely affected. When considering construct design one must consider carefully a number of factors as outlined below.

(a) *Transgene integration*. A common problem with nuclear transformed transgenic plants is gene silencing. This may be due to the methylation of the gene due to repeated homologous sequences (Meyer, 1996) and can drastically reduce the level of protein production by affecting mRNA transcript levels. Constructs must be designed to avoid silencing triggers such as prokaryotic DNA sequences, sequence repeats, and secondary RNA structure formation (Kohli et al., 2003). In addition, systems have been devised that inhibit the inherent plant-silencing mechanism (Johansen and Carrington, 2001; Voinnet et al., 2003).

(b) *Transcription.* In dicotyledonous plants such as tobacco, the strong constitutive cauliflower mosaic virus 35S (CaMV 35S) promoter (Odell et al., 1985) is commonly used to drive transcription. This may be appropriate if the plant is able to cope with continuous accumulation of protein in a range of its tissues, but for many proteins this could be detrimental to growth (Neumann et al., 2005). Use of a promoter that provides spatial and temporal control over transgene expression, and/or inducibility may be preferable. For example, inducible promoter systems respond to a chemical stimulus (Moore et al., 2006) such as ethanol (Caddick et al., 1998; Salter et al., 1998) or hormones (Padidam, 2003; Padidam et al., 2003; Zuo and Chua, 2000; Zuo et al., 2000). Ideally, the inducer should be a cheap and safe chemical that could be applied to the plants with minimal effort and which is environmentally benign. The mechanical gene activation (MeGA) system, developed by Cramer et al. (1999) for CropTech Corp., uses a promoter inducible by mechanical stress. The shearing of leaves when tobacco is harvested causes a rapid induction of protein expression. This means that any detrimental effect the protein may have had on the plant is negligible, and so one expects good biomass yield, although it is not clear whether appropriate levels of protein will accumulate within the damaged tissue postharvest. Therefore, it seems more appropriate to select promoters that will direct expression to appropriate locations where the product is stored safely and stably before harvest.

(c) *Targeting to subcellular locations*. Targeting a protein to specific compartments can increase yield and protein stability. Given the correct targeting signal, the reporter protein, green fluorescent protein (GFP), is accumulated to 3% in secreted fluid (Komarnytsky et al., 2000). Dual targeting of the same protein to both the chloroplast and the peroxisome, simultaneously, increased accumulation approximately twofold over targeting to either organelle alone (Hyunjong et al., 2006). In terms of posttranslational processing, proteins have different requirements. Complex proteins, such as antibodies, are best suited to the secretory pathway (Schillberg et al., 1999) where the addition of an ER retention H/KDEL C-terminal tag has been shown to enhance accumulation, and enable longer storage in harvested leaf material (Schillberg et al., 1999). ER retention also means the protein will not enter the Golgi apparatus where unwanted modifications such as glycosylations can occur.

3.4.3. Chloroplast transformation

Plastids provide a useful location for the production of human therapeutic proteins (Nugent and Joyce, 2005). It is possible to target integration within the plastid genome minimizing adverse effects that can occur through random integration into the host genomic DNA. Transformation of the chloroplast genome can result in 7–8,000 copies of the gene per cell and has resulted

in target proteins accounting for up to 47% of total soluble protein in leaf tissue (De Cosa et al., 2001). Proteins tend to fold properly within chloroplasts. These organelles also lack the ability to posttranslationally glycosylate proteins. Another major advantage of plastid transgenes is that in plants such as tobacco, choroplasts are maternally inherited meaning that there are no transgenes present in pollen thus enhancing the biological containment and reducing environmental risks (Daniell, 2006; Ruf et al., 2007).

3.4.4. Plant cell culture

Plant cell cultures are established from cells derived from stable transgenic plants. The level of recombinant protein that can potentially be generated by whole plant systems, whether by nuclear or by chloroplast transformation, cannot be rivaled by plant cell culture systems. However, cell culture systems offer some advantages including the ability to secrete protein into the culture supernatant, good control over cell growth conditions, reproducibility between batches, the ability to comply with GMP standards, and excellent containment of transgenes offering good commercial opportunities for certain types of recombinant protein production (Hellwig et al., 2004). Adventitious root and hairy root cultures can also be exploited for recombinant protein production in suitable bioreactors (Sivakumar, 2006).

Numerous examples of protein production from plant cell cultures exist with a good example being that of human-secreted alkaline phosphatase, which accumulated to 27 mg/L in the culture medium of transformed tobacco NT1 cells. Interestingly, the activity of the enzyme decreased during stationary phase not through protein degradation, but due to protein denaturation caused by uncharacterized factors generated by cell growth. Bacitracin, a protein stabilizer was shown to be effective at stabilizing activity over a 17-day culture period (Becerra-Arteaga et al., 2006). Tobacco was also used to produce a human monoclonal antibody against the rabies virus. The culture produced relatively low levels of functionally assembled antibody of 0.5 mg/L, but this was threefold greater than was produced by the original transgenic plant, and levels of production were maintained over a 3-month culture period (Girard et al., 2006). Sharma et al. (2006) also used tobacco to produce human interleukin-18 in the culture medium at a level of 166 µg/L and which proved to be bioactive. A range of contructs were tested for the expression of hepatitis B surface antigen (HBsAg) from tobacco cell suspension culture with maximum yields of 31 µg/L being achieved (Kumar et al., 2006).

Although tobacco is probably the most common host, a range of plant hosts have been shown to function in cell suspension cultures. For example, rice cell cultures produced human growth hormone in the medium at 57 mg/L and showed similar biological activity to human growth factor produced by *E. coli* (Kim et al., 2008b). Lee et al. (2007) produced biologically active human

cytotoxic T-lymphocyte antigen 4-immunoglobulin in a rice suspension culture at a level of 31.4 mg/L in the culture medium. One issue with secretion of recombinant proteins into the culture medium is the low concentration of the protein. McDonald et al. (2005) used a two-compartment membrane bioreactor in which secreted recombinant α-1-antitrypsin inhibitor was retained in the small volume cell compartment and was recovered at a concentration of up to 247 mg/L representing some 10% of total extracellular protein after 6 days. Multiple cycles of induction could be performed. Moss also shows potential for cell culture applications. The genome sequence has been determined (http://www.cosmoss.org/), and it has a haploid genome and a homologous recombination mechanism allowing targeted gene replacement or specific integration. To reduce immunogenicity issues in humans due to the differences in glycosylation patterns between plants and animals, two enzymes responsible for incorpotaion of xylose and fucose have been knocked out in moss leading to a "humanized" glycosylation pattern. With the ability for production of antibodies and the availability of photoreactors, moss promises to provide a useful system for future recombaint protein production (Deckor and Reski, 2007).

4. REPETITIVE AND SELF-ASSEMBLING PROTEINS AND PEPTIDES IN NATURE

4.1. Overview

In Nature, there are many examples of protein and peptide molecular self-assembly. Of the genetically engineered fibrous proteins, collagen, spider silks, and elastin have received attention due to their mechanical and biological properties which can be used for biomaterials and tissue engineering.

4.2. Recombinant production of self-assembling proteins

4.2.1. Silks

Silk proteins, with comparable mechanical and biological properties to native silk, are being produced recombinantly to improve cost-effective production. Expression of synthetic silk has been explored in many eukaryotic hosts including tobacco and potato plants (Scheller et al., 2001), as well as in mammalian epithelial cells (Lazaris et al., 2002) and transgenic goats (Williams, 2003). More recently, expression of spider dragline silk protein was demonstrated in the milk of transgenic mice to an estimated level of 11.7 mg/L (Xu et al., 2007a).

Cappello et al. (1990) pioneered the genetic engineering of repetitive building blocks. Working with the silkworm fibroin, they found that by starting with short oligonucleotide repeats, the size of the protein generated

was easily controlled. Synthetic genes encoding the dragline silk from *Nephila clavipes* have also been designed based upon native gene sequences. Highly repetitive regions exist containing consensus sequences, entitled NCMAG1 and NCMAG2 (formerly referred to as spidroin 1 and spidroin 2) (Foo and Kaplan, 2002). These consensus sequences are

NCMAG1: $_n$-GQGGYGGLGGQGAGRGGLGGQGAGA(A)$_n$GGA-c
NCMAG2: $_n$-GPGGYGPGQQGPGGYGPGQQGPSGPGS(A)$_n$-c

Prince et al. (1995) used the expression vector pQE-9 in which synthetic genes, generated from the consensus sequences, were placed under the control of the bacteriophage T5 promoter for expression in *E. coli*. A His-tag was added to the N-terminus of the recombinant protein to allow purification by nickel-affinity chromatography. However, the resulting yields of purified protein were low at 15 mg/L. By contrast, Fahnestock and Bedzyk (1997) achieved greater success using the methylotrophic yeast *P. pastoris* obtaining yields of 1 g/L prior to purification.

Yeast and bacterial systems often give low levels of expression of silks, and this has led to the development of production systems in tobacco and potato. Scheller et al. (2001) have shown that spider silk proteins can be produced in transgenic plants. They inserted synthetic spider silk protein (spidroin) genes into transgenic plants under the control of the CaMV35S promoter. Using this system they were able to demonstrate the accumulation of recombinant silk proteins to a level of at least 2% of total soluble protein in the endoplasmic reticulum of tobacco leaves, and potato tubers.

There are several reports of the use of silk in biomaterials. Kluge et al. (2008) provide a good overview of application of spider silks including recombinant versions. In addition to spider silks there are other types of silks that provide distinct and useful properties, such as those derived from mussels which will presumably become targets for recombinant protein production (Carrington, 2008).

4.2.2. Elastins

Elastin is an extracellular matrix protein found in connective tissue (Keeley et al., 2002), and as the name suggests provides elasticity and resilience to tissues which require extensibility and recoil. Elastin is a highly insoluble, cross-linked polymer synthesized as a soluble monomer, tropoelastin, which is cross-linked in the extracellular matrix by the enzyme lysyl oxidase to generate a polymeric structure. It is composed of hydrophobic domains rich in glycine (G), valine (V), and proline (P) residues, and cross-linking domains rich in alanine and lysine residues (Bellingham et al., 2001). Once polymerized elastin shows little degeneration with aging, and elastin formed in early development remains the same in the later stages of life (Urry, 1993).

Urry and colleagues pioneered the design and production of elastin-like proteins, which were (VPGVG)-based, with cell attachment sequences (GRGDSP) and showed their biocompatibility using immunogenicity studies and cytotoxicity studies with bovine aortic endothelial cells and fibroblasts (Nicol et al., 1991, 1992). Elastin-like materials of a polypeptide with the sequence G-$(VPGVG)_{19}$-VPGV fused to glutathione *S*-transferase has been recombinantly produced in *E. coli*. From a fermentation culture, the fusion protein was affinity purified and cleaved with protease Xa with a final yield of 1.15 mg/L of G-$(VPGVG)_{19}$-VPGV (McPherson et al., 1992). Guda et al. (1995) further reported on the hyper-expression of poly(GVGVP) in *E. coli* that exhibited inverse temperature transitions of hydrophobic folding and assembly in which the transitions ranged from hydrogel, elastic, and plastic conformations.

Bellingham et al. (2001) have investigated the self-assembling behaviour of recombinant human elastin polypeptides expressed in *E. coli*. They designed elastin polypeptide constructs as glutathione *S*-transferase (GST) fusion proteins, and introduced specific hydrophobic and cross-linking domains proposed to be important in elastin self-assembly. This elastin polypeptide was separated from the GST-fusion protein by CNBr cleavage or by thrombin cleavage. They found the hydrophobic domains, which contained a similar consensus sequence to human elastin (PGVGVA repeated 7 times), to be essential for elastin polypeptide self-assembly. It is very promising from a biomaterials perspective that recombinant elastin can undergo physiochemically induced transitions from monomer to fibrillar polymer, a property which is of great benefit at the biomaterial–tissue interface.

Girotti et al. (2004) have designed and bio-produced elastin-like protein polymers (ELP) which contain biofunctional motifs with cell adhesion sequences required for tissue-engineering applications. The protein polymer contained periodically spaced fibronectin CS5 domains enclosing the cell attachment sequence REDV. The overall sequence,

$(VPGIG)_2VPGKG(VPGIG)_2$EEIQIGHIP**REDV**DYHLYP
$(VPGIG)_2VPGKG(VPGIG)_2(VGVAPG)_3$

contains elastin-like units of the type $(VPGIG)_n$, with an adjacent VPGKG in which a lysine (K) was introduced in place of an isoleucine (I) so that cross-linking could occur. The CS5 human fibronectin domain was subsequently introduced (underscored) with a REDV recognition sequence (bold) with further VPGXG units completing the bioactive domain. These authors have also begun to exploit the properties of elastin-like polymers and their "smart" behavior in response to temperature. Below the transition temperature the noncross-linked polymer chains remained soluble in water, yet when the temperature was

increased above the transition temperature they formed an aggregated self-assembled complex. The cross-linking of lysine residues enabled the production of scaffold-like biomaterials for use in regenerative medicine. Recombinant approaches offer a more sustainable method for producing large quantities of ELPs with the incorporation of large block polymers and cell recognition sites, compared with production through solid-state synthesis.

5. RECOMBINANT PEPTIDE PRODUCTION

5.1. Overview

Despite burgeoning interest in the self-assembling peptide field, there are few published examples of attempts at recombinant production. This is probably because those working on self-assembling peptides are focussed more on the nature and properties of the peptides than on large-scale production. The production of short peptides in biological hosts commonly provokes intracellular degradation by proteases or if the peptide is antimicrobial or particularly hydrophobic they may be toxic to the host cell, resulting in low yields. Normally a fusion protein (Table 1) is used to both protect the peptide and/or the cell and allow recovery of the target peptide before being cleaved (Table 2) and removed in downstream processing steps to acquire pure active/functional peptide. These downstream processes can lead to a significant loss of target peptide. Recombinant production of peptides, therefore, is not trivial. Due to their potential for medical applications, therapeutic peptides such as antimicrobial peptides have been the subject of sustained attempts at recombinat production. Consideration of some recent examples provides some indication of useful aspects and potential difficulties for the design of self-assembling peptide systems.

Since the fusion protein partner is substantially larger than the size of the required peptide, overall peptide yield represents only a fraction of the purified fusion protein, even before losses due to subsequent purification. For example, for a fusion protein of 150 amino acids and a peptide of 15 amino acids, the relative levels of products are 91 and 9%, respectively. So, it is useful to increase the proportion of the final fusion protein that comprises peptide sequences. This can be achieved either by using a smaller fusion protein or by increasing the number of peptide sequences cloned in tandem with the fusion partner. As shown in the following examples both approaches have been adopted, but with variable results. Finally, the separation of the tandem repeats of peptides into monomers can be achieved by either chemical cleavage or enzymatic cleavage (Section 1.1.1.3).

5.2. Tandem repeat strategy

The concept of expression of tandem protein/peptide repeats interspersed by a single cleavable residue was first introduced by Shen (1984). Monomer, dimer, and trimer repeats of proinsulin were generated with Met residues between each repeat to allow CNBr cleavage. The trimer gave the greatest yield of proinsulin. Of more direct relevance for production of short peptides was the work of Kempe et al. (1985) who expressed tandem repeats of the neuropeptide Substance P (APKQQFFGLM-NH^+) joined to a β-galactosidase fusion partner. β-galactosidase was separated from the peptides by a Met residue, and each peptide contained a C-terminal Met, so upon CNBr cleavage the peptide monomers were released with a C-terminal homoserine lactone, which was then chemically converted to a homoserine amide. The fusion partner in this case is large at 1,024 amino acids and fusions of 4 (SP4), 16 (SP16), or 64 (SP64) repeats of the Substance P peptide were expressed. Due to the relative sizes of the β-galactosidase and the peptides, the theoretical levels of recovery would be expected to be 3.8%, 13.2% and 33.5% for the 4, 16 and 64 repeat constructs respectively. Following CNBr cleavage and high-performance liquid chromatography (HPLC) purification 95%, 79% and 70% recovery were reported. To put this into perspective, from a 10 mg sample of the SP64 a final yield of peptide of 2.36 mg was achieved compared with the theoretical yield of 3.35 mg.

A further example of effective expression of tandem repeats of peptide sequences associated with a fusion partner was provided by Kuliopulos and Walsh (1994). Their construct design was based on use of the 125 amino acid, hydrophobic protein fusion protein, ketosteroid isomerase (KSI). The vector pET31b is available from Novagen® (brand of EMD Chemicals Inc., and affiliate of Merck GmbH) and uses the T7 promoter to express a KSI-peptide repeat His-tag protein. Both the vector components of KSI and His-tag are separated from the peptide repeats by methionine residues. The peptide repeats are cloned into the vector by using the restriction enzyme *Alw*NI, which recognizes the sequence CAGNNN/CTG. The unspecified NNN sequence permits the use of ATG, which encodes the amino acid methionine, allowing subsequent CNBr cleavage. Figure 1 shows a schematic representation of the cloning of three tandem repeat sequences using this approach and subsequent protein/peptide products. Kuliopulos and Walsh, 1994 expressed a KSI-$(peptide)_5$-His tag protein in *E. coli* which was then purified by nickel-affinity chromatography and cleaved with CNBr. Cleavage resulted in the release of insoluble KSI, His-tag, and peptide monomers, with yields of peptide reported to be 50–55 mg/L following HPLC purification.

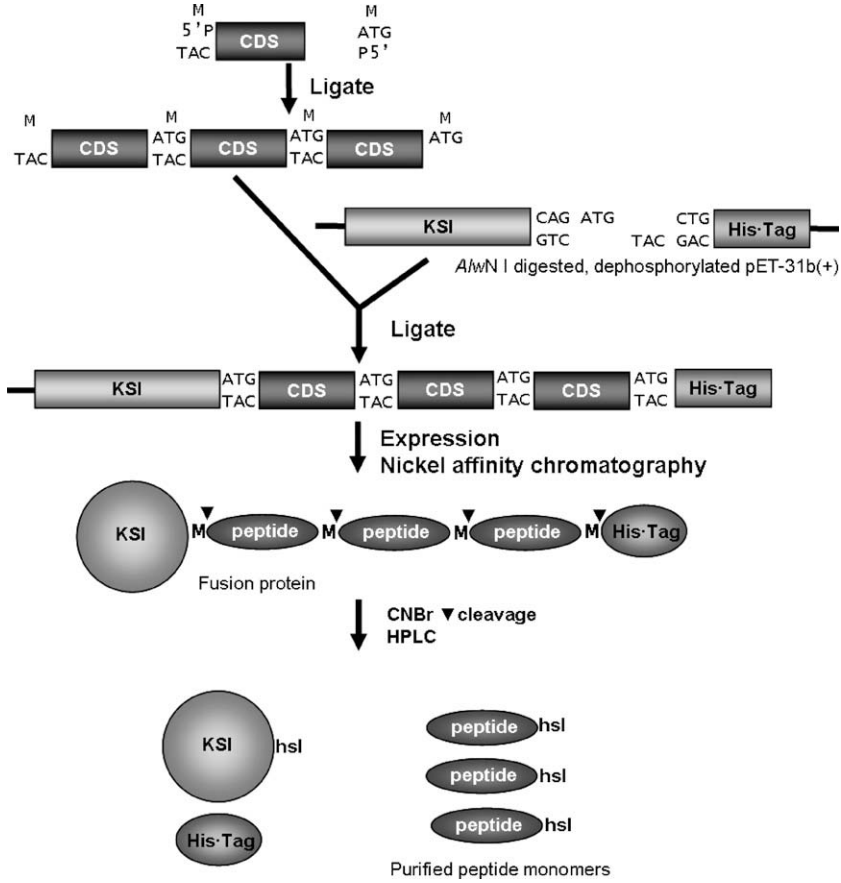


Figure 1 Strategy for cloning a peptide-coding sequence (CDS) as tandem repeats in the vector pET31b. The resulting fusion protein, comprising the ketosteroid isomerase (KSI), peptide repeats, and His-tag, is targeted to inclusion bodies. The fusion protein can be recovered and cleaved, in this case, with cyanogen bromide (CNBr) which acts at the methionine (M) residues allowing further separation of pure peptide from the other fusion components. The cleavage by CNBr results in a C-terminal homoserine lactone (hsl) on each peptide monomer.

5.3. Recombinant bioactive peptide examples

Recent examples of the production of bioactive peptides, often antimicrobials, in bacterial and yeast host systems highlight examples of various fusion and cleavage strategies.

5.3.1. Multiple antimicrobial peptide expression

Niu et al. (2008) used the methylotrophic yeast *P. pastoris* vector pPICZα-A in which transcription of the target protein is controlled by the very

powerful methanol-inducible alcohol oxidase 1 (AOX1) promoter. They expressed a hybrid protein comprising four different antimicrobial peptides. The peptides were Protegrin-1, 4 kDa Scorpion Defensin, Metalnikowin-2A, and Sheep Myeloid Antibacterial Peptide SMAP-29. The coding sequences were synthesized based on *P. pastoris* codon bias to enhance translational efficiency, and each peptide was preceded by a KEX2 protease cleavage site. The artificial protein had the sequence

EKR*RGGRLCYCRRRFCVCVGR**EKR***RGLRRLGRKIAHGVKKYG
PTVLRILRIAG**EKR***GFGCPLNQGACHRHCRSIRRRGGYCAGFF
KQTCTCYRN**EKR***VDKPDYRPRPWPRPNSR

with the recognition sequence (EKR) and cleavage site (*) for Kex-2 protease highlighted in bold. Upon expression and secretion into the growth medium the endogenous KEX2 activity catalyzed cleavage at the recognition sites to give the four peptides each with its native N-terminus. Results from expression trials indicated that KEX2 cleavage was not complete, with products larger than the size range (1.9–4.3 kDa) of the intended final cleaved peptide products. It was estimated that some 271 mg/L of the fusion protein was produced. The peptide material was not purified, but growth assays using diluted growth medium indicated inhibition of growth of both Gram-positive and Gram-negative bacteria, but no hemolytic activity on eukaryotic cells which was taken to indicate no toxic effect. This preliminary study is encouraging and demonstrates the potential for exploiting an endogenous proteolytic cleavage mechanism to generate a correct N-terminal sequence, although the C-terminus contains the additional cleavage site residues. Obviously, further refinement of the system and detailed characterization of purified products and their respective yields are required.

5.3.2. Cecropins

Cecropins are antimicrobial peptides first identified in insects; they are small cationic peptides produced in fat bodies and hemocytes in response to bacterial infections or injury. The peptide ABP-CM4 from the silkworm *Bombyx mori* is a 35-amino acid probable amphiphilic alpha-helical peptide that has been shown to kill bacteria, tumors, and fungi by permeabilization of cell membranes but does not display toxicity toward normal mammalian cells. Zhang et al. (2006) produced the synthetic coding region for the peptide optimized for *P. pastoris* codon usage, which was cloned directly into pPICZαA under the control of the AOX1 promoter. Under the most appropriate conditions, 20 °C, 2% Casamino acids, 0.5% methanol induced for 72 h some 40 mg/L of ABP-CM4 was secreted into the growth medium. Following a 30-min boiling step the peptide was further purified by Sephadex G-50 size-exclusion chromatography, resulting in a yield of

15 mg/L peptide. This purified peptide showed antimicrobial activity against both *E. coli* K12 D31 and the pathogenic fungi *Aspergillus niger*, *Trichoderma viride*, *Gibberella saubinetii*, and *Penicillium chrysogenum*.

In another study by Li et al. (2007a), the peptide ABP-CM4-coding sequence was synthesized and cloned into the *E. coli* vector pET32a to generate a fusion protein comprising Thioredoxin (Trx), a His-tag, and CM4. The fusion protein had the format **Trx-His-tag-Asp/Pro-CM4.** The CM4 being preceded by an Asp-Pro dipeptide which is the target for formic acid digestion (Table 2) resulting in release of a CM4 peptide with an N-terminal proline residue. The gene was not codon optimized for expression in *E. coli* and was expressed in *E. coli* BL21 (DE3) as a soluble protein at a level of 25 mg/L. The resulting protein was purified by nickel-affinity chromatography and cleaved by 1% formic acid (40 °C, 72 h). After lyophilization to remove the formic acid, a further nickel-affinity step to remove the fusion partner and any uncleaved protein was performed. Finally, the peptide sample was subjected to reverse-phase HPLC (RP-HPLC) yielding 1.2 mg/L of peptide which was shown to posess antimicrobial activities against strains of *E. coli* $K_{12}D_{31}$, *P. chrysogenum*, *A. niger*, and *G. saubinetii*.

Chen et al. (2008) have expressed a codon-optimized form of the CM4 peptide in *E. coli*. In this case, two alternative expression/purification systems were examined namely the widely used glutathione-*S*-transferase (GST) and a chitin-binding domain (CBD) system with associated intein splicing (New England Biolabs Inc.). The GST system failed to allow recovery of expressed fusion protein. In contrast, the CBD/intein system allowed the recovery of 110 mg/L fusion protein with a final RP-HPLC purification step yielding 2.1 mg/L of pure peptide which displayed antimicrobial activity against *E. coli* $K_{12}D_{31}$ and *Salmonella*.

Another study on expression of an insect cecropin in *E. coli* used a fusion to the thioredoxin protein (Xu et al., 2007b). In this case the *Musca domestica* (house fly) cecropin termed Mdmcec was cloned downstream of the thioredoxin with an intervening His-tag for purification and enterokinase (Table 2) cleavage site (bold) for release of peptide

TRX-HisTag-**DDDDK**/MKKIGKKIERVGQHTRDATIQTIGVAQ
QAANVAATLKG

The fusion protein was expressed in BL21 (DE3) cells by IPTG induction and subsequently purified by nickel-affinity chromatography, resulting in fusion protein yields of 48 mg/L. Following enterokinase cleavage and HPLC, 11.2 mg/L of pure recombinant Mdmcec was recovered. The recombinant peptide was shown to have the expected molecular mass by electrospray ionization-mass spectrometry and a predominantly helical conformation by circular dichroism. It also displayed antimicrobial activity against a range of Gram-negative and Gram-positive bacteria and

various fungi. It is proposed that this expression system will provide a reliable and simple method for production of different cationic peptides (Xu et al., 2007b).

5.3.3. Dermcidin

Most antimicrobial peptides are cationic but dermcidin is an anionic antimicrobial peptide secreted in human sweat (Cipakova et al., 2006). It comprises 110 amino acids and posesses a 19-amino acid N-terminal signal peptide which directs secretion. The 110-amino acid peptide is proteolytically processed to form C-terminal peptides of predominantly 48 or 47 amino acids in length which are responsible for antimicrobial activity. In this study, the expression of the 48-amino acid form was achieved by fusing the peptide-coding region to the C-terminus of the KSI sequence present in the expression vector pET31 (Novagen®) outlined previously (Figure 1).

```
KSI-MSSLLEKGLDGAKKAVGGLGKLGKDAVEDLE
SVGKGAVHDVKDVLDSVM-Histag
```

The peptide coding region was synthesized with *Alw*NI restriction enzyme sites encoding the amino acid methionine to allow subsequent CNBr cleavage. Following IPTG induction of the *E. coli* BL21(DE3)pLysS cells carrying the plasmid, the inclusion bodies were recovered by centrifugation, the pellet was dissolved in 6M guanidinum hydrochloride and the soluble fraction loaded onto a nickel-affinity column. The eluted protein was dialysed and the protein that precipitated was subjected to CNBr cleavage. Following lyophilization, the peptide was recovered by repeated extraction with distilled water. The recovered material showed antimicrobial activity against both bacterial and fungal samples. In this example, the dermicin extract would also be expected to contain the His-tag peptide as this would also be soluble.

5.3.4. Cathelicidins

Cathelicidins are host defense peptides found in mammals including LL-37, the only cathelicidin found in humans. Bacterial infection leads to release of the LL-37 precursor hCAP-18 which contains a conserved cathelin domain and C-terminal antimicrobial peptides. These peptides are referred to by their two N-terminal residues plus the length in amino acids. Thus, LL-37 starts with Leu-Leu and is 37 amino acids in length. Posttranslational processing of LL-37 in human sweat and skin generates further antimicrobial peptides, and this study investigated the expression of SK-29, KR-20, LL-29, and LL-23 which are related as shown.

```
LL-37 LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES
LL-23 LLGDFFRKSKEKIGKEFKRIVQR
LL-29 LLGDFFRKSKEKIGKEFKRIVQRIKDFLR
```

```
KR-20 KRIVQRIKDFLRNLVPRTES
SK-29 SKEKIGKEFKRIVQRIKDFLRNLVPRTES
```

In this study, Li et al. (2007b) used an *E. coli* BL21(DE3) host in conjunction with the vector pET32a to express a thioredoxin fusion of the format:

```
Trx-Histag-LVPR/GS-Stag-D/P-Peptide
```

Following expression, the cells were harvested and lysed, and to the cell lysate 1% v/v Triton was added and the soluble fusion protein purified by batch binding to cobalt-affinity resin. The collected resin was incubated with thrombin for 16 h to release the Stag-peptide segment which was then subjected to formic acid cleavage and HPLC to isolate the peptide. The yields were SK-29 (1.7 mg/L), KR-20 (0.7 mg/L), LL-29 (2.1 mg/L), and LL-23 (5.4 mg/L), and the peptides were shown to inhibit *E. coli*, although the least inhibitory, LL-23 was expressed to the highest level suggesting the peptides affected the host cells.

5.3.5. Piscidins

Piscidins are 22-amino acid long peptides with conserved N-termini, which represent a new family of antimicrobial peptides isolated from the mast cells of fish. Piscidin 1, derived from hybrid-striped bass, displays good inhibitory properties and has the amino acid sequence

```
FFHHIFRGIVHVGKTIHRLVTG
```

(Moon et al., 2007). This peptide is proposed to adopt an amphipathic, helical conformation (Silphaduang and Noga, 2001). The peptide coding region was fused to the C-terminus of a His-tag-ubiquitin-coding sequence. As the peptide codon usage was not optimized for *E. coli*, the construct was expressed in Rosetta (DE3)pLysS, a strain that produces normally poorly represented tRNAs. Following IPTG induction in a rich medium the protein was found in a soluble form, so the cells were lysed by a freeze-thaw method and the fusion protein isolated by nickel-affinity chromatography. In contrast, when growth was on minimal medium the protein aggregated into inclusion bodies, requiring the cells to be lysed by whole cell dissolution in 8 M urea. Following nickel-affinity binding, the fusion protein was washed sequentially in 6, 4, 2, 1, and finally 0 molar urea solutions to allow refolding of the ubiquitin. The eluted protein was then subject to cleavage using a recombinant form of yeast ubiquitin hydrolase to release the piscidin, which was then purified by RP-HPLC. The yields of fusion protein and purified peptide for the rich and minimal media were 15 and 1.5 mg/L, respectively. The isolation of this material was designed to allow NMR structural studies through isotopic labeling

during the minimal medium growth and so no bioactivity assays were performed (Moon et al., 2007).

5.3.6. Cationic peptide indolicidin derivative

To produce high levels of cationic antimicrobial peptides, a strategy was used by Metlitskaia et al. (2004) involving the generation of multimer repeats of the coding region, interspersed by a short anionic spacer, to counter the positive charge on the peptide. Additionally, between each coding sequence Met residues were incorporated to allow CNBr cleavage. The fusion partner in this case was a 96-amino acid region derived from the cellulose-binding domain (CBD) of *Clostridium celliulovarans*, termed C96. This domain was unable to bind cellulose, but was effective at targeting the fusion protein to inclusion bodies. The repeat unit of peptide-M-spacer-M therefore had the sequence

ILRWPWWPWRRK**M**AEAEPEAEPI**M**

where the cleaved Met residues (bold) would be converted to homoserine lactone. The most efficiently expressed construct in *E. coli* had 15 such repeat units associated with the C96 fusion and gave up to 65% total cellular protein. The inclusion bodies were recovered by centrifugation and then directly solubilized in 70% formic acid to which CNBr was added to cleave at the Met residues. Cation exchange followed by RP-HPLC chromatography steps were used to purify the peptide resulting in recovery of 100 mg/L, which is a good level of recovery. The recombinant peptide was shown to have similar biological activity to a synthetic version suggesting the C-terminal homoserine lactone (hsl) modification was not important for activity. This example highlights the benefits of targeting peptides to inclusion bodies where they can be protected from degradation and isolated from cellular components against which they may exhibit toxic effects.

5.3.7. Histonin

Another example of mutlimeric constructs giving high yields of antimicrobial peptides in *E. coli* was provided by Kim et al. (2008a) who were also keen to ensure the absence of any additional amino acid residues derived from fusion partners. Multimeric repeats of the histonin coding region

RAGLQFPVGKLLKKLLK**RLKR**

were fused to a truncated segment of the PurF protein, termed F4. Constructs containing between 2 and 32 repeats attached to F4 were generated with maximal levels of protein expression achieved with 12 repeats. As a result of the endogenous RLKR sequence (bold) at the C terminus of histonin, a furin cleavage step was possible allowing the recovery of native histonin monomers

with no additional sequence. Inclusion bodies were recovered, solubilized in 8 M urea, and following dialysis, furin cleavage was performed over a period of 24 h at 30 °C prior to cation exchange chromatography resulting in histonin yields of 167 mg/L of culture. The recombinant peptide displayed identical antimicrobial activity to a synthetic version.

5.3.8. Microbial peptide expression in plant systems

Several examples of production of antimicrobial peptides in plants have been reported, but these are predominantly from the perspective of protecting the plant against microbial insult, rather than use of the plant as a biofactory for the large-scale production of peptide.

An early attempt to produce a recombinant human 29-amino acid antimicrobial peptide SMAP29 in tobacco was reported by Morassutti et al. (2002) who used a chitin-binding domain for purification and an associated intein cleavage system to release the peptide. Expression levels appeared to be low, with the peptide associated with plant proteins. Upon electrophoretic separation, the peptide displayed the expected immunological and biological activites.

A further example is the production of proctolin, a neuropeptide with myotropic properties, in tobacco (Rao et al., 2004). The peptide has the sequence RYLPT and so a poly-proctolin gene was generated that had 10 repeats with dibasic cleavage sites between the repeats. These dibasic sites comprised two Arg residues (bold) so the repeats had the format

$$\mathrm{RYLPT}\mathbf{RR}\mathrm{YLPT}\mathbf{R}(\mathbf{R}\mathrm{YLPT}\mathbf{R})_n\mathbf{R}$$

Endogenous serine protease activity resulted in cleavage between the RR residues (bold) to give the sequence RYLPTR, from which the C-terminal Arg would then be removed by endogenous carboxypeptidase D activity to yield native peptide, which was detectable by mass spectrometry.

The examples outlined in this section serve to provide an indication that the production of peptides is not trivial, but that approaches are available for reasonable levels of recovery of some peptides. The challenge is to apply appropriate approaches for the production of self-assembling peptides.

6. RECOMBINANT EXPRESSION OF SELF-ASSEMBLING PEPTIDES

6.1. Overview

The production of short self-assembling peptides using recombinant technology is still in its infancy. Yields should be sufficiently high to make it cost effective, and one must be able to recover functional peptides from the fusion partner. As with any peptide, it is often difficult to produce a peptide

monomer with no additional amino acids following cleavage from its fusion partner (as shown by the cleavage sites in Table 2). Fortunately, additional residues are more likely to be acceptable with self-assembling peptides than they are for medically important bioactive peptides. Self-assembly is a process which is dependent on many physical and chemical properties, one important example being charge. For example, Aggeli et al. (2003) have designed peptides that exploit the pKs of various amino acids in order to promote pH responsiveness. In this instance, the distribution of charged residues and the overall net charge of the peptide are important considerations. With careful consideration of these factors, it has been possible to design peptides which reliably form antiparallel beta-sheet structures at defined pH. This is significant in terms of recombinant production as both enzymatic and chemical cleavage can alter charge. For example, the addition of a homoserine lactone following CNBr cleavage results in the removal of the negative charge at the C-terminus. This could be counteracted by blocking the N-terminus of the resulting peptide, but it is equally possible that the homoserine lactone could prove useful, for example, in surface immobilization of peptides.

6.2. Recombinant self-assembling peptide examples

6.2.1. $A\beta_{11\text{-}26}$ peptide

Using the KSI fusion, inclusion body approach, Sharpe et al. (2005) generated tandem repeats of the $A\beta_{11\text{-}26}$ peptide region by expression in *E. coli* under conditions where they could achieve uniform labeling with 13C and 15N for NMR structural studies of the aggregates formed by this peptide. They found that greatest yields of peptide were achieved with a triplet repeat. After CNBr cleavage and RP-HPLC purification they achieved reasonable levels of recovery, but not as high as expected. From the 80 mg/L fusion protein they expected to recover 27 mg/L peptide, but only achieved 10 mg/L. As a consequence of its pH responsiveness, the acidic conditions to which the Alzheimer's peptide $A\beta_{11\text{-}26}$ peptide was subjected prior to monomer purification probably contributed to the significant loss of peptide due to aggregation or fibril formation. This potential for aggregation is likely to be an issue for most self-assembling peptides, and so conditions will need to be discovered, which allow quantitiative recovery.

6.2.2. AG repeats

Panitch et al. (1997) produced impressive yields of 700 mg/L by *E. coli* fermentation of a simple small artificial protein comprising 60 repeats of an $(AG)_4$ motif with 23 and 33 amino acid N- and C-terminal fusions, respectively. The (AG)240 repeat units were not separated after expression, and the fusion sequences comprised just 10% of the whole fusion protein. This meant that loss of polypeptide material due to inefficient cleavage by

CNBr was negligible and resulted in around 78% yield recovery. The fed batch fermentation allowed the control of oxygen, nitrogen, and glucose sources, improving yields from tens of milligrams per liter previously obtained in standard batch fermentation.

6.2.3. RAD16

Reed et al. (2006) reported on the production of a self-assembling peptide using *R. eutropha.* This peptide, RAD16, was 16 amino acids in length and produced as tandem repeats with the sequence

$$(\text{RADARADARADARADA}\mathbf{E})_n$$

The introduction of a glutamic acid (E; bold), after each RAD16 repeat was made to allow cleavage by endoproteinase Glu-C, which cleaves C-terminal to glutamate. A cellulose binding domain (CBD) (Table 1) was selected as the affinity tag, as CBDs bind strongly and specifically to cellulose which is a relatively cheap and abundant purification matrix. The main problem encountered was the low level of peptide recovered. Theoretically, 1 g of fusion protein should give 267 mg of peptide, but only 10.1 mg of peptide was recovered after RP-HPLC.

6.2.4. Vesicle-forming peptides

The ability to generate peptide-derived vesicles may be useful for a range of applications such as the encapsulation and delivery of various pharmaceutical reagents including chemically synthesized drugs and biomolecules such as proteins and nucleic acids. van Hell et al. (2007) developed two peptides designed to self-assemble into vesicles in an aqueous solution. These peptides were amphiphilic possessing the same N-terminal hydrophobic tail, which was also acetylated after purification, but with alternative hydrophilic C-terminal regions comprising either 2 or 7 Glu residues. These Glu residues were intended to provide negatively charged domains at neutral pH which should have the property of forming a relatively large interfacial surface area favoring a vesicular structure. The sequences of the peptides were

```
SA2  Ac-Ala-Ala-Val-Val-Leu-Leu-Leu-Trp-Glu-
Glu-COOH
SA7  Ac-Ala-Ala-Val-Val-Leu-Leu-Leu-Trp-Glu-
Glu-Glu-Glu-Glu-Glu-Glu-COOH
```

The recombinant system for peptide production involved designing synthetic oligonucleotides, which could be annealed, strand extended, and then cloned into the pET-SUMO vector to generate a fusion protein. This fusion protein was expressed in *E. coli* BL21(DE3) cells by IPTG induction and then cells lysed by freeze-thaw lysis followed by sonication. The lysate was passed

through a Hi-Trap HP column to capture the His-tagged protein which was then eluted in 400 mM imidazole-containing buffer. Following cleavage of the SUMO from the peptide using a His-tagged SUMO protease for a period of 6 h, the samples were passed through another Ni-NTA matrix to separate the peptide from the SUMO protease and SUMO fusion protein. A 5-L fermentation yielded 300 mg of fusion protein and 30 mg of purified peptide.

The peptide was then N-terminally acetylated to increase hydrophobicity and to prevent any charge interactions with the Glu residues. The Trp was used as an integral probe for fluorescence anisotropy measurements to assess the critical aggregate concentration (CAC) of peptide. Vesicle formation was observed by electron microscopy studies, and the efficiency of vesicular structure was probed by entrapment of a small water-soluble fluorophore, calcein. The SA2 vesicles entrapped larger quantities of calcein than did the SA7 vesicles and the CAC of the former was also lower. It was suggested that the SA2 peptides may pack more densely within the vesicular structure enhancing their ability to retain the water-soluble probe. It was also demonstrated that the assembly of vesicles can be controlled by the pH of the surrounding solvent which may be useful for drug release regimes (van Hell et al., 2007).

6.2.5. β-sheet-forming peptides

(a) *mLac21.* Middelberg and colleagues described the expression of an oil–water interfacially active peptide, mLac21, a 21-amino acid peptide derived from the lac repressor. They expressed a single copy of this peptide in the pET31b system as a fusion to the inclusion body-forming protein KSI. Rather than using flanking Met residues for CNBr recovery, they used Cys residues with 1-cyano-4-dimethylaminopyridium tetrafluoroborate (CDAP) cleavage. Following ion exchange and RP-HPLC purification they recovered 2.92 mg peptide/L culture (Morreale et al., 2004).

(b) *P_{11}-2.* Subsequently, Hartmann et al. (2008) have used a similar approach to express tandem repeats of the peptide P_{11}-2, an 11-amino acid peptide CH_3CO-QQRFQWQFEQQ-NH_2 (Aggeli et al., 1997) that forms interpeptide hydrogen bonding in a pH- and concentration-dependent manner resulting in the formation of peptide hydrogels (Aggeli et al., 2003). Of course in a recombinant system the N- and C-terminal modifications of the synthetic peptide cannot be reproduced without post-synthesis manipulation. There are various applications of this peptide including bone defect treatment (Firth et al., 2006), surface coatings (Whitehouse et al., 2005), and even antimicrobial properties (Protopapa et al., 2006). Hartmann et al. (2008) generated KSI-fusions carrying 1, 2, 4, and 9 repeats of P_{11}-2 each separated by flanking Cys residues. The cultures were grown in minimal medium and induced by

IPTG addition, and interestingly, the level of production of fusion protein decreased with repeat number such that the 9-mer produced very low levels. The most efficient level of peptide could be recovered from the single repeat construct which produced some 85 mg fusion per litre culture. The cell pellets were dissolved in urea and the fusion protein captured on a nickel affinity column resulting in around 92% purity of fusion protein with a yield of 73% for the single repeat construct. The cleavage step with CDAP was not particularly efficient at around 49%, but an acetone precipitation step after cleavage resulted in efficient removal of contaminating proteins including KSI. The peptide was purified by RP-HPLC with a yield of 2.63 mg peptide per litre culture being obtained.

(c) *P_{11}-4.* We used a similar approach (Riley et al., submitted; Figure 1) to express tandem repeats of the related self-assembling peptide P_{11}-4 CH_3CO-QQRFEWEFEQQ-NH_2 that was designed to form β-sheets and nematic gels at low pH (Aggeli et al., 2003). Details of the properties of P_{11}-4 are provided by Carrick et al. (2007). This peptide has applications in hard tissue treatment (Firth et al., 2006) and bone/dental remineralization (Kirkham et al., 2007). P_{11}-4 was expressed as 1–6 tandem repeats with Met residues between the KSI fusion and each peptide repeat. An important aspect of this work was the use of the autoinduction system (Section 3.1.1) developed by Studier (2005) for the high-level production of recombinant proteins from the T7 promoter in pET31b. This system allows high cell densities to be achieved before induction. As was observed by Hartmann et al. (2008) for P_{11}-2 fusions, a reduction in yield was observed with the longer repeats of 4–6 copies. However, in this case the more productive construct contained three copies of P_{11}-4, in keeping with the finding of Sharpe et al. (2005) with $A\beta_{11\text{-}26}$. A strategy was used of isolating inclusion bodies by detergent-based cell lysis followed by centrifugation to recover the inclusion bodies which were then dissolved in urea then dialysed against water before CNBr cleavage. The most inefficient step in the process was the RP-HPLC purification. Nonetheless, the very high cell densities and efficient production of recombinant protein allowed recovery of 2.5 g fusion protein/L culture for the $(P_{11}\text{-}4)_3$ version which should have provided a theoretical yield of 530 mg peptide/L. The actual yield was around 90 mg/L which is still a further improvement on many other reports and further optimization should provide substantial improvement. The recovered peptide identity was confirmed by mass spectrometry and MS sequencing. The recombinant form demonstrated very similar properties to the chemically synthesized version, with characterization by circular dichroism, atomic force microscopy, and transmission electron microscopy revealing β-sheet structure and the formation of fibrillar structures as expected (Riley et al., submitted).

7. PERSPECTIVE

The examples of recombinant peptide production in bacterial and yeast systems indicate that improvements are being made in the levels of peptide that can be purified. With the use of autoinduction approaches and fermentation these systems have the potential for producing substantial quantities of peptides. With some bacteria there is the issue of potential endotoxin contamination. It seems likely that research will lead to an ever-increasing range of applications of self-assembling peptides such as new designed sequences with potential to replace many oil-based materials. In turn this will generate a demand for large-scale, inexpensive, and safe sources of recombinant peptides. Potentially, one of the most promising routes to meet such demand will be through a molecular farming approach in which transgenic crops are used as the production vehicle. Currently, there appear to be no reports of the development of such systems, although this is an area of active investigation in our laboratory.

ACKNOWLEDGMENTS

KARG, SP, and JMR acknowledge studentship support from the Biotechnology and Biological Sciences Research Council, SK achnowledges support from the Wellcome Trust. We thank The Dow Chemical Company for CASE studentship support (SP and JMR).

REFERENCES

Aggeli, A., Bell, M., Boden, N., Keen, J.N., Knowles, P.F., McLeish, T.C.B., Pitkeathly, M., and Radford, S.E. *Nature* **386**, 259 (1997).

Aggeli, A., Bell, M., Carrick, L.M., Fishwick, C.W.G., Harding, R., Mawer, P.J., Radford, S.E., Strong, A.E., and Boden, N. *J. Am. Chem. Soc.* **125**, 9619 (2003).

Arst, H.N., and Penalva M.A. *Fungal Gen. Biol.* **40**, 1 (2003).

Ayed, A., Rabhi, I., Dellagi, K., and Kallel, H. *Enz. Microb. Tech.* **42**, 173 (2008).

Barnard, G.C., Henderson, G.E., Srinivasan, S., and Gerngross, T.U. *Protein Expr. Purif.* **38**, 264 (2004).

Becerra-Arteaga, A., Mason, H.S., and Shuler, M.L. *Biotechnol. Prog.* **22**, 1643 (2006).

Bellingham, C.M., Woodhouse, K.A., Robson, P., Rothstein, S.J., and Keeley, F.W. *Biochim. Biophys. Acta* **1550**, 6 (2001).

Blommel, P.G., Becker, K.J., Duvnjak, P., and Fox, B.G. *Biotechnol. Prog.* **23**, 585 (2007).

Blommel, P.G., and Fox, B.G. *Protein Expr. Purif.* **55**, 53 (2007).

Boehm, R. *Biol. Emerg. Viruses* **1102**, 121 (2007).

Caddick, M.X., Greenland, A.J., Jepson, I., Krause, K.P., Qu, N., Riddell, K.V., Salter, M.G., Schuch, W., Sonnewald, U., and Tomsett, A.B. *Nat. Biotechnol.* **16**, 177 (1998).

Cappello, J., Crissman, J., Dorman, M., Mikolajczak, M., Textor, G., Marquet, M., and Ferrari, F. *Biotechnol. Prog.* **6**, 198 (1990).

Carrick, L.M., Aggeli, A., Boden, N., Fisher, J., Ingham, E., and Waigh, T.A. *Tetrahedron* **63**, 7457 (2007).

Carrington, E. *Trends Biotech.* **26**, 55 (2008).

Cereghino, J.L., and Cregg, J.M. *FEMS Microbiol. Rev.* **24**, 45 (2000).

Chen, Y.Q., Zhang, S.Q., Li, B.C., Qiu, W., Jiao, B., Zhang, J., and Diao, Z.Y. *Protein Expr.Purif.* **57**, 303 (2008).
Choe, W.S., and Middelberg, A.P.J. *Biotech. Bioeng.* **75**, 451 (2001).
Cipakova, I., Gasperik, J., and Hostinova, E. *Protein Expr. Purif.* **45**, 269 (2006).
Cramer, C.L., Boothe, J.G., and Oishi, K.K. *Plant Biotechnol.* **240**, 95 (1999).
Cregg, J.M., Cereghino, J.L., Shi, J.Y., and Higgins, D.R. *Molec. Biotech.* **16**, 23 (2000).
Daniell, H. *Biotechnol. J.* **1**, 1071 (2006).
Daniell, H., Streatfield, S.J., and Wycoff, K. *Trends Plant Sci.* **6**, 219 (2001).
De Cosa, B., Moar, W., Lee, S.B., Miller, M., and Daniell, H. *Nat. Biotechnol.* **19**, 71 (2001).
Deckor, E.L., and Reski, R. *Curr. Opin. Biotechnol.* **18**, 393 (2007).
Denison, S.H. *Fungal Genet. Biol.* **29**, 61 (2000).
Dougherty, W.G., Cary, S.M., and Parks, T.D. *Virology* **171**, 356 (1989).
Dougherty, W.G., and Parks, T.D. *Virology* **172**, 145 (1989).
During, K., Hippe, S., Kreuzaler, F., and Schell, J. *Plant Molec. Biol.* **15**, 281 (1990).
Fahnestock, S.R., and Bedzyk, L.A. *Appl. Micro. Biotechnol.* **47**, 33 (1997).
Falconer, R.J., O'Neill, B.K., and Middelberg, A.P.J. *Biotechnol. Bioeng.* **62**, 455 (1999).
Fang, L., Jia, K.Z., Tang, Y.L., Ma, D.Y., Yu, M., and Hua, Z.C. *Protein Expr. Purif.* **51**, 102 (2007).
Firth, A., Aggeli, A., Burke, J.L., Yang, X.B., and Kirkham, J. *Nanomedicine* **1**, 189 (2006).
Florack, D., Allefs, S., Bollen, R., Bosch, D., Visser, B., and Stiekema, W. *Transgenic Res.* **4**, 132 (1995).
Foo, C.W.P., and Kaplan, D.L. *Adv. Drug Deliv. Rev.* **54**, 1131 (2002).
Fox, J.L. *Nat. Biotechnol.* **21**, 3 (2003).
Giddings, G. *Curr. Opin. Biotechnol.* **12**, 450 (2001).
Girard, L.S., Fabis, M.J., Bastin, M., Courtois, D., Petiard, V., and Koprowski, H. *Biochem. Biophys. Res. Comm.* **345**, 602 (2006).
Giraud-Guille, M.M. *Calcified Tissue Internat.* **42**, 167 (1988).
Girotti, A., Reguera, J., Rodriguez-Cabello, J.C., Arias, F.J., Alonso, M., and Testera, A.M. *J. Mater. Sci.* **15**, 479 (2004).
Gleba, Y., Klimyuk, V., and Marillonnet, S. *Curr. Opin. Biotechnol.* **18**, 134 (2007).
Gomord, V., Denmat, L.A., Fitchette-Laine, A.C., Satiat-Jeunemaitre, B., Hawes, C., and Faye, L. *Plant J.* **11**, 313 (1997).
Guda, C., Zhang, X., McPherson, D.T., Xu, J., Cherry, J.H., Urry, D.W., and Daniell, H. *Biotechnol. Lett.* **17**, 745 (1995).
Hartmann, B.M., Kaar, W., Falconer, R.J., Zeng, B., and Middelberg, A.P.J. *J. Biotechnol.* **135**, 85 (2008).
Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Luckl, J., Griengl, H., Kohlwein, S.D., and Schwab, H. *Protein Expr. Purif.* **11**, 61 (1997).
Hellwig, S., Drossard, J., Twyman, R.M., and Fischer, R. *Nat. Biotechnol.* **22**, 1415 (2004).
Higgins, D.R., and Cregg, J.M. In "Pichia Protocols. Methods in Molecular Biology" (D.R. Higgins and J.M. Cregg, Eds.), p. 1. Humana Press, Totowa, NJ (1998).
Hood, E.E., Woodard, S.L., and Horn, M.E. *Curr. Opin. Biotechnol.* **13**, 630 (2002).
Huang, Z., LePore, K., Elkin, G., Thanavala, Y., and Mason, H.S. *Plant Biotechnol. J.* **6**, 202 (2008).
Hyunjong, B., Lee, D.S., and Hwang, I.W. *J. Exper. Botany* **57**, 161 (2006).
Ivanovski, G., Gubensek, F., and Pungercar, J. *Toxicon* **40**, 543 (2002).
Iwashita, K. *J. Biosci. Bioeng.* **94**, 530 (2002).
Johansen, L.K., and Carrington, J.C. *Plant Physiol.* **126**, 930 (2001).
Kaiser, R., and Metzka, L. *Anal. Biochem.* **266**, 1 (1999).
Keeley, F.W., Bellingham, C.M., and Woodhouse, K.A. *Phil. Trans. Royal Soc. Series B* **357**, 185 (2002).
Kempe, T., Kent, S.B.H., Chow, F., Peterson, S.M., Sundquist, W.I., Litalien, J.J., Harbrecht, D., Plunkett, D., and Delorbe, W.J. *Gene* **39**, 239 (1985).
Kim, J.M., Jang, S.A., Yu, B.J., Sung, B.H., Cho, J.H., and Kim, S.C. *Appl. Microbiol. Biotechnol.* **78**, 123 (2008a).

Kim, T.G., Baek, M.Y., Lee, E.K., Kwon, T.H., and Yang, M.S. *Plant Cell Rep.* **27**, 885 (2008b).
Kirkham, J., Firth, A., Vernals, D., Boden, N., Robinson, C., Shore, R.C., Brookes, S.J., and Aggeli, A. *J. Dental Res.* **86**, 426 (2007).
Kluge, J.A., Rabotyagova, U., Leisk, G.G., and Kaplan, D.L. *Trends Biotechnol.* **26**, 244 (2008).
Kohli, A., Twyman, R.M., Abranches, R., Wegel, E., Stoger, E., and Christou, P. *Plant Molec. Biol.* **52**, 247 (2003).
Komarnytsky, S., Borisjuk, N.V., Borisjuk, L.G., Alam, M.Z., and Raskin, I. *Plant Physiol.* **124**, 927 (2000).
Kukuruzinska, M.A., Bergh, M.L.E., and Jackson, B.J. *Annu. Rev. Biochem.* **56**, 915 (1987).
Kuliopulos, A., and Walsh, C.T. *J. Am. Chem. Soc.* **116**, 4599 (1994).
Kumar, G.B.S., Srinivas, L., Ganapathi, T.R., and Bapat, V.A. *Plant Cell Tissue and Organ Cult* **84**, 315 (2006).
Kusnadi, A.R., Nikolov, Z.L., and Howard, J.A. *Biotechnol. Bioeng.* **56**, 473 (1997).
Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J.F., Duguay, F., Chretien, N., Welsh, E.A., Soares, J.W., and Karatzas, C.N. *Science* **295**, 472 (2002).
Lee, C.T., Morreale, G., and Middelberg, A.P.J. *Biotechnol. Bioeng.* **85**, 103 (2004).
Lee, S.J., Park, C.I., Park, M.Y., Jung, H.S., Ryu, W.S., Lim, S.M., Tan, H.K., Kwon, T.H., Yang, M.S., and Kim, D.I. *Protein Expr. Purif.* **51**, 293 (2007).
Li, B.C., Zhang, S.Q., Dan, W.B., Chen, Y.Q., and Cao, P. *Biotechnol. Lett.* **29**, 1031 (2007a).
Li, Y.F., Li, X., and Wang, G.S. *Protein Expr. Purif.* **55**, 395 (2007b).
Ma, J.K.C., Barros, E., Bock, R., Christou, P., Dale, P.J., Dix, P.J., Fischer, R., Irwin, J., Mahoney, R., Pezzotti, M., Schillberg, S., Sparrow, P., Stoger, E., and Twyman, R.M. *EMBO Reports* **6**, 593 (2005a).
Ma, J.K.C., Chikwarnba, R., Sparrow, P., Fischer, R., Mahoney, R., and Twyman, R.M. *Trends Plant Sci.* **10**, 580 (2005b).
Ma, J.K.C., Drake, P.M.W., Chargelegue, D., Obregon, P., and Prada, A. *Vaccine* **23**, 1814 (2005c).
Marston, F.A.O., and Hartley, D.L. *Meth. Enzymol.* **182**, 264 (1990).
Matzura, O., and Wennborg, A. *Computer Applications in the Biosciences* **12**, 247 (1996).
McDonald, K.A., Hong, L.M., Trombly, D.M., Xie, Q., and Jackman, A.P. *Biotechnol. Prog.* **21**, 728 (2005).
McPherson, D.T., Morrow, C., Minehan, D.S., Wu, J.G., Hunter, E., and Urry, D.W. *Biotechnol. Prog.* **8**, 347 (1992).
Metlitskaia, L., Cabralda, J.E., Suleman, D., Kerry, C., Brinkman, J., Bartfeld, D., and Guarna, M.M. *Biotechnol. Appl. Biochem.* **39**, 339 (2004).
Meyer, P. *Biol. Chem. Hoppe-Seyler* **377**, 87 (1996).
Mitraki, A., and van Raaij, M.J. *Protein Nanotech.* **300**, 125 (2005).
Moon, W.J., Hwang, D.K., Park, E.J., Kim, Y.M., and Chae, Y.K., *Protein Expr. Purif.* **51**, 141 (2007).
Moore, I., Samalova, M., and Kurup, S. *Plant J.* **45**, 651 (2006).
Morassutti, C., De Amicis, F., Skerlavaj, B., Zanetti, M., and Marchetti, S. *FEBS Lett.* **519**, 141 (2002).
Morreale, G., Lee, E.G., Jones, D.B., and Middelberg, A.P.J. *Biotechnol. Bioeng.* **87**, 912 (2004).
Neumann, K., Stephan, D.P., Ziegler, K., Huhns, M., Broer, I., Lockau, W., and Pistorius, E.K. *Plant Biotechnol. J.* **3**, 249 (2005).
Nicol, A., Gowda, C., and Urry, D.W. *J. Vascular Surg.* **13**, 746 (1991).
Nicol, A., Gowda, D.C., and Urry, D.W. *J. Biomed. Mater. Res.* **26**, 393 (1992).
Niu, M.F., Li, X., Wei, J.C., Cao, R.B., Zhou, B., and Chen, P.Y. *Protein Expr. Purif.* **57**, 95 (2008).
Nugent, J.M., and Joyce, S.M. *Curr. Pharmaceut. Design.* **11**, 2459 (2005).
Odell, J.T., Nagy, F., and Chua, N.H. Nature **313**, 810 (1985).
Olmos-Soto, J., and Contreras-Flores, R.*Appl. Microbiol. Biotechnol.* **62**, 369 (2003).
Padidam, M. *Curr. Opin. Plant Biol.* **6**, 169 (2003).
Padidam, M., Gore, M., Lu, D.L., and Smirnova, O. *Transgenic Res.* **12**, 101 (2003).

Paine, M.L., and Snead, M.L. *J. Bone Mineral Res.* **12**, 221 (1997).
Panitch, A., Matsuki, K., Cantor, E.J., Cooper, S.J., Atkins, E.D.T., Fournier, M.J., Mason, T.L., and Tirrell, D.A. *Macromolec.* **30**, 42 (1997).
Plunkett, E.R. "Handbook of Industrial Toxicology". Chemical Publishing Co., Inc., New York (1976).
Prince, J.T., McGrath, K.P., Digirolamo, C.M., and Kaplan, D.L. *Biochemistry* **34**, 10879 (1995).
Protopapa, E., Aggeli, A., Boden, N., Knowles, P.F., Salay, L.C., and Nelson, A. *Med. Eng. Phys.* **28**, 944 (2006).
Punt, P.J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J., and van den Hondel, C. *Trends Biotechnol.* **20**, 200 (2002).
Rai, M., and Padh, H. *Curr. Sci.* **80**, 1121 (2001).
Ramessar, K., Sabalza, M., Capell, T., and Christou, P. *Plant Sci.* **174**, 409 (2008).
Rao, R., Breuer, M., Tortiglione, C., Malva, C., Baggerman, G., Corrado, G., De Loof, A., and Pennacchio, F. *Biotechnol. Lett.* **26**, 1413 (2004).
Reed, D.C., Barnard, G.C., Anderson, E.B., Klein, L.T., and Gerngross, T.U. *Protein Expr. Purif.* **46**, 179 (2006).
Rodriguez, B., Kavoosi, M., Koska, J., Creagh, A.L., Kilburn, D.G., and Haynes, C.A. *Biotechnol. Prog.* **20**, 1479 (2004).
Ruf, S., Karcher, D., and Bock, R. *Proc. Nat. Acad. Sci USA* **104**, 6998 (2007).
Salter, M.G., Paine, J.A., Riddell, K.V., Jepson, I., Greenland, A.J., Caddick, M.X., and Tomsett, A.B. *Plant J.* **16**, 127 (1998).
Scheller, J., Guhrs, K.H., Grosse, F., and Conrad, U. *Nat. Biotechnol.* **19**, 573 (2001).
Schillberg, S., Zimmermann, S., Voss, A., and Fischer, R. *Transgenic Res.* **8**, 255 (1999).
Schmidt, F.R. *Appl. Microbiol. Biotechnol.* **65**, 363 (2004).
Sharma, N., Kim, T.G., and Yang, M.S. *Biotechnol. Bioprocess Eng.* **11**, 154 (2006).
Sharpe, S., Yau, W.M., and Tycko, R. *Protein Expr. Purif.* **42**, 200 (2005).
Shen, S.H. *Proc. Nat. Acad. Sci USA* **81**, 4627 (1984).
Silphaduang, U., and Noga, E.J. *Nature* **414**, 268 (2001).
Sivakumar, G. *Biotechnol. J.* **1**, 1419 (2006).
Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C. *Nat. Protocols* **1**, 2019 (2006).
Sparrow, P.A.C., Irwin, J.A., Dale, P.J., Twyman, R.M., and Ma, J.K.C. *Transgenic Res.* **16**, 147 (2007).
Spiegel, H., Schillberg, S., Sack, M., Holzem, A., Nahring, J., Monecke, M., Liao, Y.C., and Fischer, R. *Plant Sci.* **149**, 63 (1999).
Spok, A. *Trends Biotechnol.* **25**, 74 (2007).
Strausberg, R.L., and Link, R.P. *Trends Biotechnol.* **8**, 53 (1990).
Studier, F.W. *Protein Expr. Purif.* **41**, 207 (2005).
Tan, S.H., Wu, W.T., Liu, J.J., Kong, Y., Pu, Y.H., and Yuan, R.Y. *Protein Expr. Purif.* **25**, 430 (2002).
Terpe, K. *Appl. Microbiol. Biotechnol.* **72**, 211 (2006).
Twyman, R.M., Stoger, E., Schillberg, S., Christou, P., and Fischer, R. *Trends Biotechnol.* **21**, 570 (2003).
Urry, D.W. *Angew. Chem. Int. Ed.* **32**, 819 (1993).
van Hell, A.J., Costa, C., Flesch, F.M., Sutter, M., Jiskoot, W., Crommelin, D.J.A., Hennink, W.E., and Mastrobattista, E. *Biomacromol.* **8**, 2753 (2007).
van Ooyen, A.J.J., Dekker, P., Huang, M., Olsthoorn, M.M.A., Jacobs, D.I., Colussi, P.A., and Taron, C.H. *FEMS Yeast Res.* **6**, 381 (2006).
Vandermeulen, G.W.M., and Klok, H.A. *Macromolec. Biosci.* **4**, 383 (2004).
Vaquero, C., Sack, M., Chandler, J., Drossard, J., Schuster, F., Monecke, M., Schillberg, S., and Fischer, R. *Proc. Nat. Acad. Sci USA* **96**, 11128 (1999).
Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. *Plant J.* **33**, 949 (2003).
Whitehouse, C., Fang, J.Y., Aggeli, A., Bell, M., Brydson, R., Fishwick, C.W.G., Henderson, J.R., Knobler, C.M., Owens, R.W., Thomson, N.H., Smith, D.A., and Boden, N. *Angew. Chem. Int. Ed.* **44**, 1965 (2005).
Williams, D. *Med. Device Technol.* **14**, 9 (2003).

Xu, H.T., Fan, B.L., Yu, S.Y., Huang, Y.H., Zhao, Z.H., Lian, Z.X., Dai, Y.P., Wang, L.L., Liu, Z.L., Fei, J., and Li, N. *Animal Biotechnol.* **18**, 1 (2007a).
Xu, X.X., Jin, F.L., Yu, X.Q., Ji, S.X., Wang, J., Cheng, H.X., Wang, C., and Zhang, W.Q. *Protein Expr. Purif.* **53**, 293 (2007b).
Yamawaki, S., Matsumoto, T., Ohnishi, Y., Kumada, Y., Shiomi, N., Katsuda, T., Lee, E.K., and Katoh, S. *J. Biosci. Bioeng.* **104**, 403 (2007).
Zhang, J., Zhang, S.Q., Wu, X., Chen, Y.Q., and Diao, Z.Y. *Process Biochem.* **41**, 251 (2006).
Zhang, S.G. *Biotechnol. Adv.* **20**, 321 (2002).
Zhang, S.G. *Nat. Biotechnol.* **21**, 1171 (2003).
Zhang, Y.W., Liu, R.J., and Wu, X.Y. *Annals Microbiol.* **57**, 553 (2007).
Zuo, J.R., and Chua, N.H. *Curr. Opin. Biotechnol.* **11**, 146 (2000).
Zuo, J.R., Niu, Q.W., and Chua, N.H. *Plant J.* **24**, 265 (2000).