
Fusion polypeptides for manufacturing of cyclic peptides

The present invention relates to fusion polypeptides for manufacturing of cyclic peptides, nucleic acid molecules encoding such fusion polypeptides and genetically modified cells comprising such nucleic acid molecules. Additionally, the present invention relates to a method for preparing a cyclic target peptide and target peptide mixtures. Further aspects
5 of the present invention become apparent when studying the attached patent claims and the specification, including examples.

New classes of cyclic peptides with beneficial therapeutical and interesting properties gained significance over the recent years. Naturally occurring cyclic peptides such as cyclotides are not available to the markets to this date because the assembly of the
10 required multistep organic synthesis procedure needs to be tailored to any new cyclic peptide. Cyclic peptides can be found in several classes of antibiotics. For example, vancomycin and streptomycin are antibiotics that were inspired by naturally occurring antibiotics such as penicillin. These substances are manufactured in optimized industrial processes and allow for an economically efficient production of certain cyclic peptides.
15 However, the modular processes cannot be transferred from one cyclic peptide to another cyclic peptide. Biosynthetic methods to express cyclic peptides in a recombinant way do exist, but they are not suitable for scale-up.

A method to screen cyclic peptides for their ability to inhibit certain polypeptide activities in a host has been developed in the early 2000s. It is called SCICLOPPS (split intein mediated circular ligation of peptides and polypeptides). This method is used to genetically encode dozens of varieties of circular peptide and test their inhibitory effects on polypeptides that were co-expressed with the polypeptide to be tested. Cyclic peptides are released in nature via a split-intein mechanism. Inteins are domains in a polypeptide that encode for their own release from the polypeptide.

It was possible to use this split-intein mechanism to produce cyclic peptides *in vivo*. Scott et al. (Production of cyclic peptides and polypeptides *in vivo*, *Proceedings of the National Academy of Sciences in the USA*, 96(24) (13638 – 13643), 1999) reported the use of a split-intein domain for the intracellular catalysis of peptide backbone cyclisation. "Cyclisation" as used herein means bringing linear (poly-)peptides in a cyclic (circular) conformation. The cyclisation was performed to generate peptides and polypeptides, which are stabilized against cellular catabolism in *E. coli*.

Also, other peptides, which are not naturally cyclic, are interesting targets for cyclisation. One example is the Green Fluorescent Polypeptide (GFP), which was reported to be cyclized *in vivo* in a paper of Iwai et al. (*Cyclic Green Fluorescent Polypeptide Produced in Vivo Using an artificially split PI-Pful Intein from Pyrococcus furiosus*, *The Journal of Biological Chemistry*, 276(19) (16548 - 16554), 2001). The amounts of cyclized GFP were reported to be on milligram scale.

All of the cyclic peptides, which were reported to be produced in the literature are only produced in small amounts, a reliable method for scale-up has been not reported. Also, it was reported that the split-intein mechanism is functional as an *in vivo* method to cyclize peptides and polypeptides, but it was not reported that this mechanism also works *in vitro*.

Also, the known split intein methods do not allow for the cyclization of sequences without N-terminal cysteine. Products of split inteins can only be released and circularized when the linear precursor of the cyclic product that lies between the two split intein parts carries an N-terminal cysteine. The cysteine becomes part of the product's sequence. Cyclic peptides comprising other N-terminal amino acids than cysteine are produced using sortases to circularize their backbones. The sortase based process cannot be easily transferred from one cyclic peptide to another as different sortases are needed for the production of different cyclic peptide. Thus, there is no general tool currently available for producing cyclic peptides with variable N-termini.

The international patent application WO2019138125 discloses fusion polypeptides for the manufacturing of linear target polypeptides with an autoprotease domain from N^{PRO}. These fusion polypeptides can not be used to produce cyclic target polypeptides and exhibit certain disadvantages in terms of stability of the proteins and scalability.

- 5 Thus, the primary task of the present invention was to provide a tool for the large-scale production of a variety of different cyclic peptides and also for the production of cyclized peptides and polypeptides, which have a natural linear configuration.

The primary task was solved by providing a fusion polypeptide comprising or consisting of, in direction from the N-terminus to the C-terminus:

- 10 (i) a split C-intein domain,
- (ii) an autoprotease domain,
- (iii) a target polypeptide sequence,
- (iv) a split N-intein domain, and
- (v) a purification domain embedded into the split N-intein domain after aspartate
- 15 at position 71 of the split N-intein domain,

wherein the purification domain (v) binds to a carbohydrate matrix.

- A split intein domain is a domain which can be a split C-intein domain or a split N-intein domain, referring to the corresponding positions (C-terminal or N-terminal) of the target peptide sequence to be cyclized. The two split-intein domains are flanking the autoprotease
- 20 domain and the target peptide sequence, wherein the purification domain is embedded into the split N-intein domain after the aspartate at position 71 of the split N-intein domain. It was surprisingly found, that even if the sequence of the N-intein was interrupted, this domain is still active.

- The C-intein and N-intein cyclize the fusion polypeptide as a whole at high pH values,
- 25 preferably above pH 9.0. In this conformation, the purification domain is still capable of binding to a carbohydrate matrix. The N-intein is C-terminal to the target peptide or target polypeptide and catalyses the release of the C-terminus of the target peptide or target polypeptide and enables the cyclization of the target sequence.

The autoprotease domain exhibits the function of an autoproteolytic cleavage site, which separates the target peptide from the C-intein and the autoprotease domain. This domain is activated under certain pH values. The autoprotease domain according to the invention is necessary to release the N-terminus of the target polypeptide in order to prepare it for cyclization (Figure 1).

The purification domain confers the binding of the fusion polypeptide to a carbohydrate matrix. The purification domain is capable to be active in basic environments. This capability is influenced by the ambient domains. The inclusion body signal and the autoprotease domain can influence the fusion polypeptides' deposition in inclusion bodies.

The target peptide domain, which is to be cyclized, comprises or consists of an amino acid sequence of a target peptide or polypeptide to be produced. The domain can consist of any amino acid sequence having between 6 and more than 1000 amino acids. Preferably, the target peptide consists of an amino acid sequence of 6 to 1000 amino acids, preferably 6 to 500 amino acids, more preferably of 6 to 100 amino acids, especially preferably of 6 to 50 amino acids. In one embodiment of the present invention, the target peptide may have an amount of hydrophobic amino acids of $\geq 10\%$, based on the total number of amino acids, more preferably of $\geq 20\%$, especially preferably of $\geq 30\%$ and even more preferably of $\geq 40\%$. In another embodiment, the target peptide may have an amount of hydrophilic amino acids of $\geq 10\%$, preferably of $\geq 20\%$, especially preferably of $\geq 30\%$ and even more preferably of $\geq 40\%$, again based on the total number of amino acids. In yet another embodiment, the target peptide may have an amount of hydrophobic and hydrophilic amino acids of $\geq 10\%$, more preferably of $\geq 20\%$, especially preferably of $\geq 30\%$ and even more preferably of $\geq 40\%$, based on the total number of amino acids.

The split C-intein domain and the split N-intein domain form a peptide bond at the N-terminus of the split C-intein domain and the C-terminus of the split N-intein domain resulting in a cyclic polypeptide. Subsequently, the C-terminus of the target polypeptide sequence is released by the autocatalytic site of the split N-intein domain and the autoprotease domain. The N-terminus of the target polypeptide sequence is released by forming a peptide bond between the target polypeptide N- and C-terminus. The cyclized target polypeptide is released and the linear remaining fusion polypeptide is retained at the carbohydrate matrix.

The unmodified split inteins are restrained regarding the N-terminal amino acids. The native split intein only allows for threonine, serine or cysteine to be at the N-terminus. Replacing block G of the C-split intein with the autoprotease domain according to the present invention

allows for all amino acids except for proline to be the N-terminal amino acid, which enhances the number of possible cyclic sequence enormously.

It was surprisingly found that the replacement of the C-terminal part of the C-split intein with a different autocatalytic domain leads to a wide variety of different cyclic sequences, which can be produced and are not limited to the ones that native split inteins form. The native C-split requires the N-terminal amino acid of the cyclic product to have a cysteine at the N-terminus. This N-terminal cysteine is then released into the product. As a consequence, the native split inteins will only produce cyclic peptides with an internal cysteine at the former (previous to cyclization) N-terminal position. This is remedied by inserting a new autocatalytic domain replacing Block C of the C-split intein. This sequence modification affects the release mechanism and the autocatalytic cysteine will remain as a part of the fusion polypeptide after the cyclic peptide product is released. Thus, the present invention enables access to a wide variety of cyclic peptides, which could not be produced with methods according to the state of the art, e.g. SCICLOPPS.

It is preferred in terms of the present invention that the produced cyclic peptides do not require an N-terminal cysteine at the N-terminal position of the corresponding linear sequence.

It is possible with the fusion polypeptides according to the present invention to produce a variety of different cyclic target polypeptides, wherein these target polypeptides do not need to be cyclic polypeptides in nature. Also, target polypeptides, which have a linear natural conformation can be cyclized using the fusion polypeptides according to the present invention.

In addition, the fusion polypeptides according to the present invention can be used to produce cyclic target polypeptides at an economically efficient large scale.

In one preferred embodiment, the fusion polypeptide according to the present invention has a purification domain, which comprises or consists of an amino acid sequence according to SEQ ID No.: 1 to SEQ ID No.: 3 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 1 to SEQ ID No.: 3.

The purification domains can consist of different carbohydrate binding modules (CBMs) of different organisms. The binding strength of the purification domain to the carbohydrate matrix can be enhanced by combining single building blocks of carbohydrate binding modules (CBM) with each other. The binding can be stabilized under specific reaction

conditions (e.g. a high ionic strength) and the size of the purification domain can be varied to fit the desired target peptide domain. The purification domains according to SEQ ID No.: 1 to SEQ ID No.: 3 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 1 to SEQ ID No.: 3 show preferential properties in terms of binding strength and optimized interaction with the other domains of the fusion polypeptide.

Whenever the present disclosure relates to the percentage of sequence identity of nucleic acid or amino acid sequences to each other these values define those values as obtained by using the EMBOSS Water Pairwise Sequence Alignments (nucleotide) program for nucleic acids or the EMBOSS Water Pairwise Sequence Alignments (polypeptide) program for amino acid sequences. Alignments or sequence comparisons as used herein refer to an alignment over the whole length of two sequences compared to each other. Those tools provided by the European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute (EBI) for local sequence alignments use a modified Smith-Waterman algorithm. When conducting an alignment, the default parameters defined by the EMBL-EBI are used. Those parameters are (i) for amino acid sequences: Matrix = BLOSUM62, gap open penalty = 10 and gap extend penalty = 0.5 or (ii) for nucleic acid sequences: Matrix = DNAfull, gap open penalty = 10 and gap extend penalty = 0.5. The skilled person is well aware of the fact that, for example, a sequence encoding a polypeptide can be "codon-optimized" if the respective sequence is to be used in another organism in comparison to the original organism a molecule originates from.

Another preferred embodiment of the present invention relates to a fusion polypeptide according to the invention, wherein the autoprotease domain (ii) comprises or consists of an amino acid sequence according to SEQ ID No.: 8 to SEQ ID No.: 12 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 8 to SEQ ID No.: 12.

Preferably (and advantageously, in particular in connection with preferred embodiments as described herein), the autoprotease domain is activated at a pH value of 6.8 or above (i.e. is not activated below), more preferably at a pH value of from 6.8 to 7.2. The autoprotease was inspired by the pestivirus autoprotease N^{pro} and is modified such that the pH of its environment is the activating trigger rather than the chaotrope concentration. The autoprotease domain according to SEQ ID No.: 8 to SEQ ID No.: 12 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 8 to SEQ ID No.: 12 show preferential properties in terms of a targeted activation of the autoproteolytic activity at specific pH conditions as defined above.

The autoproteolytic activity of the autoproteolytic domain is based on the catalytic diade of histidine and cysteine in the active site of autoprotease enzymes. These enzymes are the basis for an autoproteolytic domain according to the invention. The basis for such an autoproteolytic domain can be the autoprotease N^{pro} from the pestivirus or an autoprotease
5 from a potyvirus, picornavirus or any other viral autoprotease. Through targeted recombination or re-design of these sequences, autoprotease domain building blocks can be designed, which exhibit alone or in combination several advantages over their natural counterpart. On the one hand, the pH sensitivity of the autoprotease can be adjusted precisely. This exhibits the advantage, that the activity of the autoprotease can be
10 controlled to fit the desired reaction conditions. Either with a very tight pH value range to precisely activate the autoprotease at the desired pH and avoid the early release of the target peptide or also at harsh pH values, where naturally occurring autoproteases are not stable anymore.

Yet another preferred embodiment of the present invention relates to a fusion polypeptide
15 according to the invention, wherein the fusion polypeptide additionally comprises an inclusion body sequence selected from SEQ ID No.: 13 to SEQ ID No.15 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 13 to SEQ ID No.: 15.

Preferably, the inclusion body domain is embedded into a C-split intein domain, having the
20 following sequence from N- to C-terminus: a first portion of a C-split intein, preferably having an amino acid sequence selected from SEQ ID No.: 4 or SEQ ID No.: 6, an inclusion body sequence, preferably having an amino acid sequence according to any one of SEQ ID No.:13 to SEQ ID No.: 15 and a second portion of a C-split intein, preferably having an amino acid sequence selected from SEQ ID No. 5 or SEQ ID No.: 7.

25 Surprisingly it was found, that these sequences do not only control inclusion body promotion but also direct the refolding process in strongly basic environments.

Signal sequences are always selected based on their influence on the reprocessing of the target peptide. In one embodiment, an inclusion body signal sequence is used, guiding the target polypeptide to inclusion bodies.

30 One preferred embodiment of the present invention relates to a fusion polypeptide according to the invention, wherein the split C-intein domain (i) and the split N-intein domain (ii) are derived from a natural occurring DnaE split intein sequence, preferably derived from the organism *Senyochocystes sp* PCC6803 or *Nostoc punctiformes*.

Another preferred embodiment relates to a fusion polypeptide according to the invention, wherein the split C-intein domain (i) and the split N-intein domain (ii) comprise or consist of amino acid sequence pairs selected from the group consisting of SEQ ID No.: 17 and SEQ ID No.: 18, SEQ ID No.: 19 and SEQ ID No.: 20, SEQ ID No.: 21 and SEQ ID No.: 22, SEQ ID No.: 23 and SEQ ID No.: 24, SEQ ID No.: 25 and SEQ ID No.: 26, SEQ ID No.: 27 and SEQ ID No.: 28, or amino acid sequence pairs having having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 17 and SEQ ID No.: 18, SEQ ID No.: 19 and SEQ ID No.: 20, SEQ ID No.: 21 and SEQ ID No.: 22, SEQ ID No.: 23 and SEQ ID No.: 24, SEQ ID No.: 25 and SEQ ID No.: 26, SEQ ID No.: 27 and SEQ ID No.: 28,

wherein the first SEQ ID No. relates to the C-intein domain (i) and the second SEQ ID No. relates to the N-intein domain (ii).

In a preferred embodiment, the present invention relates to a fusion polypeptide having an amino acid sequence according to any of the sequences SEQ ID No.: 29 to 44 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 29 to 44, wherein the target polypeptide is inserted after the N-terminal cysteine of the autoprotease domain.

Another aspect of the present invention relates to a nucleic acid molecule encoding a fusion polypeptide according to the present invention.

In a preferred embodiment, the present invention relates to a nucleic acid molecule encoding a fusion polypeptide having a sequence according to SEQ ID No.: 77 to 92 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 77 to 92, wherein the nucleic acid sequence encoding the target polypeptide is inserted after the tgc triplet at position 588 in nucleic acid sequence SEQ ID No.: 77 and 85, after the tgc or tgt triplet at position 573 in nucleic acid sequence SEQ ID No.: 78 and 86, after the tgc or tgt triplet at position 588 in nucleic acid sequence SEQ ID No.: 79, 80, 82, 87, 88 and 90, after the tgc or tgt triplet at position 1014 in nucleic acid sequence SEQ ID No.: 81 and 89, after the tgc or tgt triplet at position 582 in nucleic acid sequence SEQ ID No.: 83 and 91, after the tgc or tgt triplet at position 336 in nucleic acid sequence SEQ ID No.: 84 and after the tgc or tgt triplet at position 678 in nucleic acid sequence SEQ ID No.: 92.

Another preferred embodiment of the present invention relates to a fusion polypeptide having an amino acid according to SEQ ID No.: 37 to 44 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or

more to SEQ ID No.: 37 to 44, wherein the fusion polypeptide carries a target peptide, which is inserted after the N-terminal cysteine of the autoprotease domain at position 196 of SEQ ID NO.: 37 to 44.

Yet another aspect of the present invention relates to a genetically modified cell, including
5 a recombinant nucleic acid molecule according the present invention, wherein the cell is capable of expressing a fusion polypeptide according to the present invention.

In one preferred embodiment of the genetically modified cell according to the invention, the cell is selected from the group consisting of *Escherichia coli*, *Vibrio natrigens*, *Saccharomyces cerevisiae*, *Aspergillus niger*, green algae, microalgae, HEK T293 and
10 Chinese hamster ovary cells (CHO).

Another aspect of the present invention relates to a method of preparing a target peptide comprising the steps of:

- (a) providing a genetically modified cell as defined above,
- (b) culturing the cell under conditions suitable for expression of a fusion polypeptide
15 according to the present invention,
- (c) obtaining the fusion polypeptide and optionally, unfolding of the obtained fusion polypeptide and directed refolding of said fusion polypeptide,
- (d) contacting the fusion polypeptide obtained in step (c) with a carbohydrate matrix,
- (e) circulizing the fusion polypeptide via the split C-intein domain (i) and the split N-intein
20 domain (ii),
- (f) cleaving the fusion polypeptide by activating the autoprotease domain of the fusion polypeptide, thereby obtaining a cyclic target peptide,
- (g) collecting a mixture comprising the cyclic target peptide.

Step (a) comprises providing a genetically modified cell expressing a fusion polypeptide.
25 Such cell is obtainable by introducing a nucleic acid molecule including a sequence encoding a fusion polypeptide, preferably in the form of a vector, into the cell by known methods such as for example by transfection or transformation.

In step (b), the cell is cultured under conditions suitable for expressing a fusion polypeptide according to the invention, preferably in a high-density culture. Culture conditions and especially conditions to achieve a high-density culture and corresponding media are well known to the person skilled in the art. In one embodiment of the present invention, the expression of the fusion polypeptide is achieved with a subsequent transport to inclusion bodies using a suitable signal sequence.

Step (c) comprises obtaining the fusion polypeptide from the culture broth and optionally unfolding of the obtained fusion polypeptide and directed refolding of said fusion polypeptide, if the fusion polypeptide is present in inclusion bodies. Solubilization conditions for the processing of inclusion bodies and conditions for the directed refolding are well known in the art. Preferably, inclusion bodies are solubilized by using 6 M guanidinium chloride, 8 M urea or 2 wt.-% sodium dodecyl sulfate and are refolded under neutral or mildly basic conditions. The directed refolding is initialized when the chaotrope or detergent concentration in the composition is below 1 wt.-%.

In step (d), the solubilized fusion polypeptide is contacted with a carbohydrate matrix such that the fusion polypeptide binds to the matrix by its purification domain (v). Cyclization of the fusion polypeptide backbone occurs under environmental conditions found in detergent and chaotrope conditions. The purification domain (v) is not negatively impacted by this cyclization. The cyclic fusion polypeptide binds to the starch mixture.

This step is performed under conditions, wherein the autoprotease domain (ii) is inactive, preferably by controlling the pH value rather than the chaotrope or denaturant concentration in order to avoid premature cleavage of the target peptide domain (iii) on the one hand and induce activity of the purification domain (v) on the other hand. Parallel, the split intein ((i),(iv)) is initialized under environmental conditions that keep the autoprotease inactive on the one hand but allow for an active purification domain (v) on the other hand. Under these conditions, the amount of cleaved fusion polypeptide is preferably <10 %, more preferably <5 %, especially preferably <3 % or even more preferably <1 %, based on the total amount of fusion polypeptide.

The underlying mechanism of the inactivity of the autoproteolytic domain can be described with two cases:

- (1) Conditions, wherein the autoprotease domain is constitutionally inactive and is only activated by a change of the environmental conditions, such as by an adaption of the temperature, the pH and/or the ionic strength, preferably by adapting the pH;

- (2) Conditions wherein the autoprotease domain is constitutionally active, however, having insufficient activity to achieve a premature cleavage of the target peptide domain during the period of time necessary for performing the method step (d), i.e. is kinetically inactive, preferably for up to 10 min, more preferably for up to 20 min and especially preferably up to 30 min.

In one embodiment, step (d) is performed under native conditions, i.e. under conditions wherein the autoprotease is constitutionally active. Surprisingly, it was found that even if the fusion polypeptide is present in its native state, the autoprotease domain remains sufficiently inactive during step (d). This effect was especially present when using an autoprotease having an amino acid sequence according to SEQ ID NO.: 7 to 11.

Preferably, an insoluble carbohydrate matrix is used in step (d), which facilitates the separation of impurities. Controlling the release of the N-terminus in step (d) mediated by the catalytic site of the pH sensitive autoprotease is paramount for the following cyclization process of the target sequence. Prior to cyclization and release of the target sequence, the cyclic fusion polypeptide binds to the carbohydrate matrix.

In step (e), the fusion polypeptide is cleaved by the autoprotease domain (ii) and the catalytic site of the split N-intein (iv) and the cyclized target polypeptide (iii) is released. Cleavage of the fusion polypeptide may result from addition of an autoproteolysis buffer, i.e. a buffer providing conditions under which the autoprotease is active, i.e. acidic conditions.

Step (f) results in obtaining a mixture by eluting the cleaved target peptide from the column. Preferably, the elution is done by using a buffer selected from the group consisting of HEPES, PBS, and TrisHCl at concentrations between 1 and 100 mM and at a pH of 6.5 to 7.5. Furthermore, the preferred buffer may be supplemented by arginine at a concentration of 10 to 100 mM or by sucrose at a concentration of 2 to 20 mM.

One preferred embodiment of the method according to the present invention relates to a method, wherein the carbohydrate matrix in step (d) consists of or comprises a substance selected from the group consisting of starch, lignin, carbohydrate polymers, copolymers with alpha-1,4- and alpha-1,6 glycosidic bonds of glucose or other sugars and mixtures thereof and is preferably present in a packed column, as a packed substrate or as starch grains consisting of amylose and amylopectin.

Starch is a complex mixture of carbohydrates from different sugar polymers. Plant cells collect the sugars they produce in a storage organelle called a vacuole. When the cells and organelles are mechanically destroyed, the starch granules are released. Depending on the plant species, there are differences in the raw starch. Starches can have different grain
5 sizes ranging from less than 25 μm to more than 100 μm in diameter. The higher the proportion with diameters of over 75 μm , the higher the probability of non-specific adsorption and thus the retention of impurities in the products after starch purification. In addition, there are starch granules, such as wheat, which are porous and can absorb amylases in internal channels. Starch consists of the components amylose and
10 amylopectin. In contrast to amylopectin, amylose is water-soluble. The swelling behavior of the respective starch in water also depends on the proportions of the two species. Thus, unpurified cornstarch in water acquires a cement-like consistency, whereas table potato starch remains water-permeable. All carbohydrate-binding enzymes have a high affinity to their substrate, which is also present under harsh conditions. Preferably, the starch grains
15 are insoluble in water. It is furthermore preferred, if the soluble amylose parts and polypeptides have been removed from the starch.

Another preferred embodiment of the method according to the present invention relates to a method, wherein the circularizing of the fusion polypeptide via the split C-intein domain (i) and the split N-intein domain (ii) in step (e) is performed at a pH of above 7.5, preferably of
20 above 9.0

Yet another preferred embodiment of the method according to the present invention relates to a method, wherein the activation of the autoprotease domain in step (e) is performed at pH 6 to pH 8, preferably at pH 6.5 to 7.5, especially preferably at pH 7 to pH 7.4.

One aspect of the present invention relates to a recombinant nucleic acid molecule,
25 encoding a fusion polypeptide according to the present invention and a cloning site for incorporation of a recombinant nucleic acid molecule according to the present invention, optionally operatively linked to an expression control sequence.

Another aspect of the present invention relates to a mixture comprising or consisting of a cyclic target peptide, preferably of a synthetic cyclic target peptide and a total amount of
30 0.001 to 1 wt.-% sodium and/or potassium, based on the total weight of the sum of sodium (if present), potassium (if present) and target peptide, wherein the mixture is obtained or obtainable by a method according to the invention.

Preferably, the cyclic target peptide, which is comprised in a mixture according to the invention, does not require an N-terminal cysteine at the N-terminal position of the corresponding linear sequence.

The invention is further characterized by illustrative, non-limiting examples.

Short description of sequences

SEQ ID No.: 1 to SEQ ID No.: 3 are artificial amino acid sequences encoding purification domains.

SEQ ID No.: 4 to SEQ ID No.: 7 are artificial amino acid sequences encoding C-split intein sequence variants.

SEQ ID No.: 8 to SEQ ID No.: 12 are artificial amino acid sequences encoding autoprotease domains.

SEQ ID No.: 13 to SEQ ID No.: 15 are artificial amino acid signal sequences for intracellular targeting of the fusion polypeptide according to the invention.

SEQ ID No.: 16 to SEQ ID No.: 28 are artificial and non-artificial amino acid sequences encoding pairs of split C-inteins and split N-inteins. The sequences with SEQ ID No. 17, 19, 21, 23, 25, and 27 encode split C-inteins and the sequences with SEQ ID No. 16, 18, 20, 22, 24, 26 and 28 encode split N-inteins.

SEQ ID No.: 29 to SEQ ID No.: 44 are artificial amino acid sequences encoding preferred fusion polypeptides, wherein the target polypeptide sequence can be inserted after the N-terminal cysteine of the autoprotease domain.

SEQ ID No.: 45 to SEQ ID No.: 48 are amino acid sequences of target peptides to be cyclized.

SEQ ID No.: 49 to SEQ ID No.: 51 are artificial nucleic acid sequences encoding purification domains.

SEQ ID NO.: 52 to SEQ ID No.: 55 are artificial nucleic acid sequences encoding C-split intein sequence variants.

SEQ ID No.: 56 to SEQ ID No.: 60 are artificial nucleic acid sequences encoding autoprotease domains.

SEQ ID No.: 61 to SEQ ID No.: 63 are nucleic acid sequences encoding signal sequences for intracellular targeting of the fusion polypeptide according to the invention.

SEQ ID No.: 64 to SEQ ID No.: 76 are artificial and non-artificial nucleic acid sequences encoding pairs of split C-inteins and split N-inteins. The sequences with SEQ ID No.: 65, 67, 69, 71, 73 and 75 encode split C-inteins and the sequences with SEQ ID No.: 64, 66, 68, 70, 72, 74 and 76 encode split N-inteins.

- 5 **SEQ ID No.: 77 to SEQ ID No.: 92** are artificial nucleic acid sequences encoding preferred fusion polypeptides, wherein the target polypeptide sequence can be inserted after the N-terminal cysteine of the autoprotease domain.

SEQ ID Nos.: 93 to SEQ ID No.: 96 are nucleic acid sequences encoding target peptides to be cyclized.

Short description of Figures

Figure 1 shows a schematic drawing of the cyclization reaction leading to a cyclized product. The reaction starts with the formation of a peptide bond at the N-terminus of the split C-intein and the C-terminus of the split N-intein resulting in a circular polypeptide. The next steps occur in an organized way. The C-terminus of the product sequence is released by the autocatalytic site of the split N-intein and the N-terminus of the product sequence is released by forming a peptide bond between the N- and C-terminus of the target peptide. This way, the cyclic product is released and the linear sequence is kept at the starch matrix.

Figure 2 A shows the results of the kinetic evaluation of the release from a fusion polypeptide according to SEQ ID No.: 39 at pH 7.0 at 421 nm and 397 nm. The upper curve shows the absorption at 421 nm and the lower curve the absorption at 397 nm. GFP has an absorption maximum at 397 nm in its linear form and at 421 nm in its cyclic form. It is shown that most of the product released from the fusion peptide is cyclic GFP, wherein the amount of linear GFP released is low. **Figure 2 B** is a non-purified sample of a fusion protein according to SEQ ID No.: 39 carrying GFP as target peptide to be cyclized after activation. The product cyclic GFP has a size of 28.5 kDa. **Figure 2 C** shows a purified sample, wherein the fusion polypeptide according to SEQ ID No.: 39 carrying GFP as target peptide is subjected to a starch matrix and incubated (binding). The autoprotease is activated and the target peptide cyclic GFP eluted. The sample is then analysed on a tricine gel. It can be seen, that the sample is almost free of impurities. The bands in the first lane of Figure C at 75 and 25 kDa are the bands of the marker.

Figure 3 A shows the results of the kinetic evaluation of the release from a fusion polypeptide according to SEQ ID No.: 40 at pH 7.0 at 421 nm and 397 nm. The upper curve shows the absorption at 421 nm and the lower curve the absorption at 397 nm. GFP has an absorption maximum at 397 nm in its linear form and at 421 nm in its cyclic form. It is shown that most of the product released from the fusion peptide is cyclic GFP, wherein the amount of linear GFP released is low. **Figure 3 B** is a non-purified sample of a fusion protein according to SEQ ID No.: 40 carrying GFP as target peptide to be cyclized after activation. The product cyclic GFP has a size of 28.5 kDa.

Figure 4 A shows the results of the kinetic evaluation of the release from a fusion polypeptide according to SEQ ID No.: 41 at pH 7.0 at 421 nm and 397 nm. The upper curve shows the absorption at 421 nm and the lower curve the absorption at 397 nm. GFP has an absorption maximum at 397 nm in its linear form and at 421 nm in its cyclic form. It is shown that most of the product released from the fusion peptide is cyclic GFP, wherein the amount of linear GFP released is low. **Figure 4 B** is a non-purified sample of a fusion

protein according to SEQ ID No.: 41 carrying GFP as target peptide to be cyclized after activation. The product cyclic GFP has a size of 28.5 kDa.

5 **Figure 5 A** shows the results of the kinetic evaluation of the release from a fusion polypeptide according to SEQ ID No.: 42 at pH 7.0 at 421 nm and 397 nm. The upper curve shows the absorption at 421 nm and the lower curve the absorption at 397 nm. GFP has an absorption maximum at 397 nm in its linear form and at 421 nm in its cyclic form. It is shown that most of the product released from the fusion peptide is cyclic GFP, wherein the amount of linear GFP released is low. **Figure 5 B** is a non-purified sample of a fusion protein according to SEQ ID No.: 42 carrying GFP as target peptide to be cyclized after activation. The product cyclic GFP has a size of 28.5 kDa.

15 **Figure 6 A** shows the results of the kinetic evaluation of the release from a fusion polypeptide according to SEQ ID No.: 44 at pH 7.0 at 421 nm and 397 nm. The upper curve shows the absorption at 421 nm and the lower curve the absorption at 397 nm. GFP has an absorption maximum at 397 nm in its linear form and at 421 nm in its cyclic form. It is shown that most of the product released from the fusion peptide is cyclic GFP, wherein the amount of linear GFP released is low. **Figure 6 B** is a non-purified sample of a fusion protein according to SEQ ID No.: 44 carrying GFP as target peptide to be cyclized after activation. The product cyclic GFP has a size of 28.5 kDa. The second lane shows the marker having bands at 25 and 75 kDa.

25 **Figure 7** shows a tricine gel loaded with samples of *E.coli* cultures expressing constructs of fusion polypeptide according to SEQ ID NO.: 39 carrying different target peptides: GFP according to SEQ ID No.: 45 (SEQ ID No.: 39-PcGFP), MCoTI-II (trypsin inhibitor) snake venom (SEQ ID No.: 39-46) and Cycloviolacin O14 (SEQ ID No.: 39-47). The bands are the following: Std. indicates the marker, a) is the uninduced control, b) culture after induction, c) the supernatant of the culture after separation of lysate supernatant and inclusion body pellet, d) inclusion body pellet after denaturation and e) sample d) mixed with starch and buffer according to table 10. The size of the fusion polypeptide according to SEQ ID No.: 39 without any product shows as 56.62 kDa on the gel and can be also detected if the cyclization reaction has happened and thus, no product is included in the fusion protein anymore. Its size including cyclic GFP is 85.2 kDa. The fusion polypeptide with MCoTI-II has a size of 59.8 kDa and with Cycloviolacin O14 a size of 59.8 kDa. Thus, all three target peptides can be produced using a polypeptide according to the invention.

30

35

Figure 8 shows a tricine gel loaded with samples of *E.coli* cultures expressing constructs of fusion polypeptide according to SEQ ID NO.: 40 carrying different target peptides: GFP according to SEQ ID No.: 45 (SEQ ID No.: 40-PcGFP), MCoTI-II trypsin inhibitor snake venom (SEQ ID No.: 40-46) and Cycloviolacin O14 (SEQ ID No.: 40-47). The bands are the following: Std. indicates the marker, a) is the uninduced control, b) culture after induction, c) the supernatant of the culture after separation of lysate supernatant and inclusion body pellet, d) inclusion body pellet after denaturation and e) sample d) mixed with starch and buffer according to table 10. The size of the fusion polypeptide according to SEQ ID No.: 40 without any product shows as 56.63 kDa on the gel. Its size including cyclic GFP is 85.2 kDa. The fusion polypeptide with MCoTI-II has a size of 59.8 kDa and with Cycloviolacin O14 a size of 59.8 kDa. Thus, all three target peptides can be produced using a polypeptide according to the invention.

Figure 9 shows a tricine gel loaded with samples of *E.coli* cultures expressing constructs of fusion polypeptide according to SEQ ID NO.: 42 carrying different target peptides: GFP according to SEQ ID No.: 45 (SEQ ID No.: 42-PcGFP), MCoTI-II trypsin inhibitor snake venom (SEQ ID No.: 42-46) and Cycloviolacin O14 (SEQ ID No.: 42-47) and sunflower trypsin inhibitor 1 (SEQ ID No.: 42-48). The bands are the following: Std. indicates the marker, a) is the uninduced control, b) culture after induction, c) the supernatant of the culture after separation of lysate supernatant and inclusion body pellet, d) inclusion body pellet after denaturation and e) sample d) mixed with starch and buffer according to table 10. The size of the fusion polypeptide according to SEQ ID No.: 42 without any product shows as 56.87 kDa on the gel. Its size including cyclic GFP is 85.42 kDa. The fusion polypeptide with MCoTI-II has a size of 60.01 kDa, with Cycloviolacin O14 a size of 60.03 kDa and with sunflower trypsin inhibitor 1 a size of 58.63 kDa. Thus, all four target peptides can be produced using a polypeptide according to the invention.

30

Examples

Example 1: Production of cyclic GFP

Construct design

The feasibility of the method according to the invention using a fusion polypeptide according to the invention was tested with three different peptides having an N-terminal glycine and different numbers of intramolecular cysteine bridges and which cannot be cyclized using methods available in the state of the art. GFP (SEQ ID No.: 45) was tested as a proof of principle. The peptides are listed in Table 1.

Table 1: Tested target sequences

SEQ ID No.	Product	Mass k[Da]
45	GFP	28.56
46	MCoTI-II (trypsin inhibitor) snake venom	32.07
47	Cycloviolacin O14	31.68
48	Sunflower Trypsin inhibitor 1	14.98

During the cyclization reaction, different fragments are produced to show

- a) the existence of the fusion polypeptide and its different parts
- 5 b) the different mechanisms of product release und different chemical environments.

The size of these fragments is listed in Table 2. The list of fragments shows the differences in the masses between the different constructs. The constructs can be distinguished by single mutations or exchanges, additions or deletions of motives. The alterations or
10 distinguishing features between the different constructs can be found in the catalytic domains of both the N-terminal and the C-terminal domain of the fusion protein.

Table 2: Components of the fusion polypeptide, N-terminally and C-terminally to the product

SEQ ID No. of the polypeptide construct	Sequence name	Complete fusion polypeptide [kDa]	N-terminal Fragment [kDa]	C-terminal Fragment [kDa]
39	SC3-UK	56.62	22.49	34.13
40	SC4-UK	56.63	22.49	34.14
41	SC5-UK	56.82	22.46	34.36
42	SC6-UK	56.87	22.53	34.33

The following table 3 shows the composition and masses of different fusion polypeptides carrying the target peptides GFP (SEQ ID No.: 45), MCo-TI-II (SEQ ID No.: 46), Cycloviolacin O14 (SEQ ID No.: 47), Sun Flower Trypsin Inhibitor I (SFT-I) (SEQ ID No.: 48) to be cyclized:

SEQ ID No.	Sequence name	Sequence with cyclic GFP [kDa]	Sequence with MCo-TI-II [kDa]	Sequence with Cycloviolacin O14 [kDa]	Sequence with Sun Flower Trypsin Inhibitor I [kDa]
39	SC3-UK	85.18	59.827	59.788	58.118
40	SC4-UK	85.19	59.837	59.798	58.128
41	SC5-UK	85.38	60.027	59.988	58.318
42	SC6-UK	85.43	60.077	60.038	58.368

5

Table 4: Composition of different fusion polypeptides carrying a product peptide to be cyclized.

SEQ ID No.	Sequence name	C-split [SEQ ID No.:]	Inclusion body tag [SEQ ID No.:]	Auto-protease [SEQ ID No.:]	N-split [SEQ ID No.:]	Purification tag [SEQ ID No.:]
39	SC3-UK	19	13	8	20	3
40	SC4-UK	21	13	8	22	3
41	SC5-UK	23	13	8	24	3
42	SC6-UK	19	13	8	28	3

Explanation of the reaction

The fusion polypeptide that is tested to produce cyclized polypeptide sequences consists of a combination of an N-terminal and a C-terminal split intein, an autocatalytic protease domain, a cyclization product sequence and a purification domain. The product sequence is N-terminally flanked by the split C-intein and the catalytic domain. The C-terminus is composed of the split N-intein domain and a purification domain that binds starches. The purification domain enhances purification and yield of the cyclized product. When the fusion polypeptide is activated, the N-terminus of the C-split intein and the C-terminus of the N-split intein form a peptide bond producing a circular polypeptide. The product sequence is cyclized when it is released by the autoprotease domain and the split N-intein domain. The cyclic polypeptide release leads to two different products, the unloaded and linearized fusion polypeptide that remains in the purification matrix and the cyclized product polypeptide that is also released from the column. Replacing the catalytic domain of the split C-intein with another autoproteolytic domain will enhance both the controllability of the product release and the product spectrum, as more amino acids will be allowed at the N-terminus of the cyclization product than cysteine, threonine and serine. The size of the fusion polypeptide in relation to the cyclization product has an influence on the reaction, especially on the pH value for the cyclization reaction. The cyclization reaction can be adapted to different target pH values or pH ranges for a better controllability of the product release.

The cyclization reaction is depicted in Figure 1 and schematically shows the cyclization reaction of a product. The given pH values are examples and may vary dependent on the product and used fusion polypeptide construct.

Example 2: Production of fusion polypeptides

The loaded fusion polypeptides were produced in *E.coli*. The gene coding for the fusion polypeptide for cyclic peptide release is contained on a pET vector allowing for controlled and enhanced expression of said fusion polypeptide from the plasmid in an appropriate cell strain.

The vector containing the fusion polypeptide for cyclic peptide release is used as a vector to transform an appropriate host cell (e.g. *E. coli*) and to store the transformed host cells on select agar plates. The colonies will be cultivated for expression of recombinant fusion polypeptide.

A colony is picked into 10 mL Luria Bertolli medium containing 30mg/L kanamycin and 30mg/L chloramphenicol. The culture is agitated over night at 37°C. The overnight culture is transferred in 180 mL of Luria Bertolli medium and 20 mL potassium phosphate solution (Table 6). The bacteria were cultivated in 200 mL of the culture medium for between three and six hours at a temperature of between 23°C and 37°C at 170 rpm. The cells are collected and resuspended in 900 ml of Luria Bertolli medium containing 30 mg/L kanamycin and 30 mg/L chloramphenicol and 100 mL phosphate solution (Table 6). The pH-value of the culture medium before the introduction of cells is slightly higher than pH 7.2, which is close to the activation point of the autoprotease domain. For this reason, a pH stabilizing measure is taken. To one litre of culture medium, 25 mL of alkaline solution (Table 5) are added after 2 h hours of incubation. After two further hours of incubation, again 25 mL of alkaline solution (Table 5) are added before induction. The culture is allowed to grow for another set of three to six hours depending on the desired optical density of the culture. The expression culture is then induced by applying 25 mL of a feeding salt solution (Table 7) over a period of one hour. The feed contains Isopropyl β -d-1-thiogalactopyranoside (IPTG) or the IPTG is applied independently. The final concentration is 2 mM. The expression culture is cultivated at temperatures between 23°C and 37°C for between three and twelve hours at 170 rpm.

Table 5: Alkaline solution

(V=0.05l, pH=11.2 at 22°C)	
Sodium hydroxide (NaOH)	4 M
glycerol	20% (v/v)

20

Table 6: Potassium phosphate solution

(V=0.5l)	
Dibasic potassium phosphate (K_2HPO_4)	0.72 M
Monobasic potassium phosphate (KH_2PO_4)	0.17 M
Thiamine sulfate	4 g/L
Glucose	50 g/L
Glycerine	200 g/L

Table 7: Salt solution for the feed

(V=0.5l)	
Citrate	34 g/L
Cobalt (II) Chloride	0.16 g/L
Manganese (II) Chloride	0.047 g/L

Copper (II) Chloride	0.03 g/L
Boric Acid	0.2 g/L
Disodium molybdate	0.16 g/L
Zink acetate	0.64 g/L
2,2',2'',2'''-(Ethane-1,2-diylidinitrilo)tetraacetic acid [EDTA]	0.52 g/L
Iron (III) Citrate	1.6 g/L
Magnesium sulfate	24 g/L
Glycerin	200 g/L

Downstream processing

The cells are harvested by centrifugation at 8,000 rpm and 4°C for 10 minutes. The medium is discarded and the cells are stored at -80°C over night. The pH-value is kept higher than

5 7.4.

The cells are thawed in lysis and washing buffer (Table 8) at a ratio of 35 mL per 8 g cells which is a ratio of approximately 1:4 (w/v). The cells are resuspended in the buffer and immediately homogenized in a disperser for 5 minutes and 30 seconds without further cooling at 2,800 rpm. The lysate is then collected and stored at 4°C for 20 minutes. The
10 lysate is centrifuged at 7,000 rpm for five to eight minutes at 4°C. The supernatant is discarded and the pellet is resuspended in the same volume of 4°C cold lysis and wash buffer (Table 8) by vortexing for 60 seconds. The suspension is then stored at 4°C again and centrifuged at 7,000 rpm for five to eight minutes at 4°C once more. The supernatant is discarded and replaced with 35 mL of desalted water. The lysate is resuspended by
15 vortexing for 60 seconds again and stored at 4°C for at least twenty minutes. The lysate in water is centrifuged a third time for five to eight minutes at 4°C at 6,000 rpm. The supernatant is discarded including DNA-rich contaminations. The washed inclusion bodies can be stored at -80°C or be immediately used for the next step.

2 g of inclusion bodies are immersed in 8 mL to 10 mL of denaturing buffer (Table 7). Ratios
20 between 1:4 and 1:10 (w/v) are suitable. The suspension is vortexed for one minute and 40 seconds and then incubated at room temperature for 8 minutes. This sequence is repeated six times. After one hour the suspension should be clear. If this is not the case, the suspension is incubated at room temperature for further 60 minutes. When clear, the suspension is stored at 4°C for two to twelve hours and centrifuged at 8,000 to 10,000 rpm
25 for six to eight minutes. The supernatant is collected and stored at 4°C for 30 minutes to two hours. If more SDS polypeptide aggregates precipitate, the suspension is centrifuged once more. The detergent free sample can be stored at 4°C or at room temperature for several weeks. The pH will be above 11.0.

The suspension is now brought in contact with the starch mixture consisting of a 2:3 mixture of wheat to potato starch washed in starch wash medium (Table 10). The starch was sieved prior to use. The starch has grain sizes between 25 and 32 μm . The starches are mixed in the buffer according to Table 6 and washed in it twice. The starch is centrifuged at 23°C for 5 minutes at 5,000 rpm. The starch is loaded to the column or a centrifuge beaker as a slurry. The buffer is removed by elution or centrifugation, after loading the fusion polypeptide sample. Other starches such as corn starch, rice starch, and starches from other plant, fungal or animal sources can be used.

After loading the sample, the buffer pH on the column is gradually lowered by adding activating buffer (Table 11) at different pH values at different rates and eluting them at different rates. The eluent or centrifugation supernatant that contain the desired cyclized product are freeze-dried. Supernatants are collected by centrifugation at between 6000 and 8000 rpm at 4°C for 6 to 8 minutes. The elution is performed at flow rates of between 0.1 ml/min and 1 ml/min at atmospheric pressure.

Table 8: Lysis and wash buffer

(V=1l, pH = 9.0 at T=23°C)	
Sodium acetate	0.05 M
Potassium hydroxide	0.005 M
Magnesium chloride	0.002 M
HEPES	0.2 M (optional)
NaCl	0.5 M (optional)
Triton X-100	1 %

Table 9: Denaturing buffer

(V=1l, pH = 9.0 at T=23°C)	
Potassium hydroxide	0.005 M
HEPES	0.2 M (optional)
SDS	2 %

Table 10: Starch washing buffer

(V=1l, pH = 7.5 at T=23°C)	
NaCl	0.05 M
Tris	0.05 M
Acetic acid	Up to 0.01 M

Table 11: Activating buffer

(V=1l, pH as needed for release of product, at T=23°C)	
Sucrose	0.02 to 0.02 M (optional)
Potassium hydroxide	Up to 0.005 M
Arginine	0.005 to 0.05 M (optional)
HEPES	Up to 0.2 M (optional)
Acetic acid	Up to 0.001 M (optional)

Example 3: Photometric and gel electrophoretic investigation of the cyclization process

The fusion protein performance was tested at different pH values, ranging between pH 4.0 and 8.0. In this case, „performance“ refers to the successful production of cyclic product peptide or polypeptide.

The tested samples are mixtures of one part fusion protein in detergent free denaturation buffer (Table 9) with nine parts of an aqueous solution having a pH value of between 5.0 to 9.0. The best results regarding the release of cyclic product and purification efficacy, was found in a 1:10 (one part protein solution and 9 parts buffer) mixture of fusion polypeptide in detergent free denaturation buffer with activating buffer (Table 11) having a pH value of 7.2.

The analytics and results are discussed in more detail below.

Investigation of kinetics

The kinetic investigation of the fusion protein that has been activated by dissolution in an activating buffer at a ratio of 1:10 is performed photometrically using an Implen n120 nanophotometer and GFP as product to be cyclized.

(1) 10 µl of clear fusion polypeptide (according to SEQ ID No.: 39, 40, 41, 42 and 44) solution in denaturing buffer (Table 7) without detergent were transferred to 90 µl of an activation buffer at pH 7.0 that allowed for the activation of the catalytic domains of the fusion polypeptide and release of the cyclic GFP.

(2) The two solutions were carefully mixed by slowly pipetting and slowly agitating the mixture for 30 seconds. The solution was incubated at room temperature for another 30 seconds and then spun down.

- (3) 2 µl of the solution were transferred to the sample field of the nanophotometer. The sample was referenced against the activation buffer. The sample was then measured over a time range of 20 minutes and at wavelengths between 250 nm and 600 nm.
- 5 (4) The results of the same fusion polypeptide carrying cyclic GFP as a product were compared to the same fusion polypeptide with other cyclic products at 280nm, 397nm and 421nm.

Figures 2 to 6 show the release of cyclic GFP from the fusion polypeptide. Cyclic GFP is released if an adsorption is detected at 421 nm, linear GFP has an absorption maximum at 397 nm. The figures show that GFP is released in its cyclic form and only in a small amount in its linear form.

10

Investigation of productivity and purification

The product cyclization was tested using fusion polypeptides according to SEQ ID No.: 39, SEQ ID No.: 40 and SEQ ID No.: 42 with GFP, MCoTI-II, Cycloviolacin O14 or Sunflower Trypsin Inhibitor-I in small scale cultures that were expressed and processed as described in Example 2. The construction of these sequences is described in Example 1.

15

The following pH values and product codes were used for releasing of the cyclic product:

Product	SEQ ID	Product Code	pH value for release [-]	Size of product on gel [kDa]
GFP	45	PcGFP	7 to 8	28
MCoTI-II	46	Pc1	7 to 8	3.207
Cycloviolacin O14	47	Pc2	7 to 8	3.168
Sunflower Trypsin Inhibitor-I	48	Pc3	7 to 8	1.498

Samples of the fractions of the whole production process were collected and analyzed via tricine gel electrophoresis as shown in Figures 7 to 9. Samples of all intermediate supernatants were investigated via gel electrophoresis. The samples a) to e) displayed on the gel are the following fractions:

- 5 a) Sample of uninduced *E. coli* culture (No inclusion body production)
- b) Sample of induced *E. coli* culture (Inclusion body production)
- c) Supernatant of *E. coli* culture after separation of lysate supernatant and inclusion body pellet.
- d) Inclusion body pellet in denaturing buffer
- 10 e) Protein solution after refolding from inclusion bodies and binding to purification domain

Figure 7 shows the products PcGFP, Pc-1 and Pc2 in the fusion polypeptide according to SEQ ID No.: 39.

- Figure 8 shows the products PcGFP, Pc-1 und Pc2 in the fusion polypeptide according to
- 15 SEQ ID No.: 40.

Figure 9 shows the products PcGFP, Pc-1, Pc2 and Pc3 in the fusion polypeptide according to SEQ ID No.: 42.

- It can be seen from the sizes of the proteins detected in the tricine gels that all fusion polypeptides carrying the desired target proteins can be produced as inclusion bodies and
- 20 are thus capable to be used in a purification method according to the invention, wherein the products are released in its cyclic form.

Example 3 - Binding experiments

- The samples for the experiments in presence of starch are performed at pH 6.5 to pH 9.0 for samples containing a fusion polypeptide with cyclic GFP as target protein according to
- 25 SEQ ID No.: 29 and 36. Similar experiments have been conducted using fusion polypeptides according to SEQ ID NO.: 39 to 43 carrying cyclic GFP (SEQ ID NO.: 45) as product.

The starch is taken from a stock mixture of 40% corn starch and 60% potato starch. The starch stock was resuspended in a buffer of 50 mM Tris, 100 mM NaCl and 1 mM EDTA. The starch mix was vortexed and 200 μ L of the mixture pipetted in 2 mL reaction vessels. The vessels were centrifuged in a microfuge at 13,000 rpm and room temperature for 2
5 minutes.

The supernatant was removed and replaced with 100 μ L of the respective fusion polypeptide sample at pH 11.5 without detergent. The sample vessels were vortexed and agitated for ten minutes at room temperature ($T=24.5^{\circ}\text{C}$). Then, 900 μ L of buffer (Table 11, pH 7.2) was added to the corresponding reaction vessel with starch and fusion polypeptide.
10 This mixture was now agitated as described above. Then, all samples were centrifuged in a microfuge at 13,000 rpm and room temperature for 2 minutes. The supernatants were precipitated in 1 mL of ethanol per sample, vortexed and the precipitate was pelleted by centrifugation. The pellet was resuspended in 50 μ L of Laemmli SDS sample loading buffer and incubated at 66°C for 10 minutes. The starch pellet was extracted using 50 μ L of
15 Laemmli SDS sample loading buffer and incubated at 66°C for 10 minutes. The samples were then loaded on TGX stain free polyacrylamide gels with an acrylamide gradient of 4-20% and compared to 1 μ L of a Dual Xtra polypeptide standard (Marker) from Bio-Rad.

Claims

1. Fusion polypeptide comprising or consisting of, in direction from the N-terminus to the C-terminus:
 - (i) a split C-intein domain,
 - (ii) an autoprotease domain,
 - (iii) a target polypeptide sequence,
 - (iv) a split N-intein domain, and
 - (v) a purification domain embedded into the split N-intein domain after aspartate at position 71 of the split N-intein domain,wherein the purification domain (v) binds to a carbohydrate matrix.
2. Fusion polypeptide according to claim 1, wherein the purification domain comprises or consists of an amino acid sequence according to SEQ ID No.: 1 to SEQ ID No.: 3 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 1 to SEQ ID No.: 3.
3. Fusion polypeptide according to claim 1 or 2, wherein the autoprotease domain (ii) comprises or consists of an amino acid sequence according to SEQ ID No.: 8 to SEQ ID No.: 12 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 8 to SEQ ID No.: 12.
4. Fusion polypeptide according to any one of the preceding claims, wherein the fusion polypeptide additionally comprises an inclusion body sequence selected from SEQ ID No.: 13 to SEQ ID No.: 15 or an amino acid sequence having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 13 to SEQ ID No.: 15.
5. Fusion polypeptide according to any one of the preceding claims, wherein the split C-intein domain (i) and the split N-intein domain (ii) are derived from a natural

occurring DnaE split intein sequence, preferably derived from the organism *Senyochocystes sp* PCC6803 or *Nostoc punctiformes*.

6. Fusion polypeptide according to any one of the preceding claims, wherein the split C-intein domain (i) and the split N-intein domain (ii) comprise or consist of amino acid sequence pairs selected from the group consisting of SEQ ID No.: 17 and SEQ ID No.: 18, SEQ ID No.: 19 and SEQ ID No.: 20, SEQ ID No.: 21 and SEQ ID No.: 22, SEQ ID No.: 23 and SEQ ID No.: 24, SEQ ID No.: 25 and SEQ ID No.: 26, SEQ ID No.: 27 and SEQ ID No.: 28, , or amino acid sequence pairs having having a sequence identity of 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % or more to SEQ ID No.: 17 and SEQ ID No.: 18, SEQ ID No.: 19 and SEQ ID No.: 20, SEQ ID No.: 21 and SEQ ID No.: 22, SEQ ID No.: 23 and SEQ ID No.: 24, SEQ ID No.: 25 and SEQ ID No.: 26, SEQ ID No.: 27 and SEQ ID No.: 28,

wherein the first SEQ ID No. relates to the C-intein domain (i) and the second SEQ ID No. relates to the N-intein domain (ii).

7. Recombinant nucleic acid molecule encoding a fusion polypeptide according to any one of the preceding claims.
8. Genetically modified cell, including a recombinant nucleic acid molecule according to claim 7, wherein the cell is capable of expressing a fusion polypeptide according to any one of claims 1 to 6.
9. Genetically modified cell according to claim 8, wherein the cell is selected from the group consisting of *Escherichia coli*, *Vibrio natrigens*, *Saccheromyces cerevisiae*, *Aspergillus niger*, green algae, microalgae, HEK T293 and Chinese hamster ovary cells (CHO).
10. Method of preparing a target peptide comprising the steps of:
- (a) providing a genetically modified cell according to claim 8 or 9,
 - (b) culturing the cell under conditions suitable for expression of a fusion polypeptide according to any one of claims 1 to 6,
 - (c) obtaining the fusion polypeptide and optionally, unfolding of the obtained fusion polypeptide and directed refolding of said fusion polypeptide,

- (d) contacting the fusion polypeptide obtained in step (c) with a carbohydrate matrix,
 - (e) circulizing the fusion polypeptide via the split C-intein domain (i) and the split N-intein domain (ii),
 - (f) cleaving the fusion polypeptide by activating the autoprotease domain of the fusion polypeptide, thereby obtaining a cyclic target peptide,
 - (g) collecting a mixture comprising the cyclic target peptide.
11. Method according to claim 10, wherein the carbohydrate matrix in step (d) consists of or comprises a substance selected from the group consisting of starch, lignin, carbohydrate polymers, copolymers with alpha-1,4- and alpha-1,6 glycosidic bonds of glucose or other sugars and mixtures thereof and is preferably present in a packed column, as a packed substrate or as starch grains consisting of amylose and amylopectin.
12. Method according to claim 10 or 11, wherein the circulizing of the fusion polypeptide via the split C-intein domain (i) and the split N-intein domain (ii) in step (e) is performed at a pH of above 7.5, preferably of above 9.0.
13. Method according to any one of claims 10 to 12, wherein the activation of the autoprotease domain in step (e) is performed at pH 6 to pH 8, preferably at pH 6.5 to 7.5, especially preferably at pH 7 to pH 7.4.
14. Recombinant nucleic acid molecule, encoding a fusion polypeptide according to any one of claims 1 to 6 and a cloning site for incorporation of a recombinant nucleic acid molecule according to claim 7, optionally operatively linked to an expression control sequence.
15. Mixture comprising or consisting of a cyclic target peptide, preferably of a synthetic cyclic target peptide and a total amount of 0.001 to 1 wt.-% sodium and/or potassium, based on the total weight of the sum of sodium (if present), potassium (if present) and target peptide, wherein the mixture is obtained or obtainable by a method according to claims 10 to 13.