

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

21 March 2024 (21.03.2024)



(10) International Publication Number

WO 2024/056874 A1

(51) International Patent Classification:

C07K 14/00 (2006.01) C12P 21/00 (2006.01)

C07K 14/435 (2006.01) C07K 4/00 (2006.01)

C07K 14/195 (2006.01)

Published:

— with international search report (Art. 21(3))

— with sequence listing part of description (Rule 5.2(a))

(21) International Application Number:

PCT/EP2023/075465

(22) International Filing Date:

15 September 2023 (15.09.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

22196120.4 16 September 2022 (16.09.2022) EP

(71) Applicant: NUMAFERM GMBH [DE/DE]; Merowingerplatz 1a, 40225 Düsseldorf (DE).

(72) Inventors: SCHWARZ, Christian; Nixhütter Weg 78b, 41468 Neuss (DE). STOCK, Janpeter; c/o NUMAFERM GmbH, Merowingerplatz 1a, 40225 Düsseldorf (DE). MONTOYA, José; c/o NUMAFERM GmbH, Merowingerplatz 1a, 40225 Düsseldorf (DE).

(74) Agent: MAIWALD GMBH; Grünstraße 25, 40212 Düsseldorf (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: FUSION PROTEINS COMPRISING GG REPEAT SEQUENCES III

(57) Abstract: The present invention relates to polypeptides comprising a first amino acid sequence comprising one or more GG repeat sequences and a peptide or polypeptide of interest in form of a fusion protein that exhibits increased renaturation efficiency and optionally also improved expression. Also encompassed are nucleic acids encoding these polypeptides, host cells that comprise said nucleic acids, and methods for protein expression and renaturation using said nucleic acids, host cells and polypeptides.



WO 2024/056874 A1

## FUSION PROTEINS COMPRISING GG REPEAT SEQUENCES III

### FIELD OF THE INVENTION

The present invention lies in the field of molecular biology, recombinant peptide and protein expression and relates to amino acid sequences that support the refolding and renaturation of a peptide or protein of interest recombinantly expressed as a fusion protein.

### BACKGROUND OF THE INVENTION

To date, recombinant protein/enzyme production for use in industrial processes is widely established. It is expected that in the future more and more industrial processes that are currently based on traditional chemistry will be adapted to involve recombinant technologies.

The type 1 secretion system (T1SS) is a relatively simple yet highly conserved secretion strategy used throughout Gram-negative bacteria to translocate small substrates as well as extremely large proteins to the extracellular environment. Much of our current knowledge of T1SS has been gathered over the last several decades from studies of the T1SS toxins HlyA and CyaA of *Escherichia coli* and *Bordetella pertussis*, respectively. These toxins, and the majority of type 1 secreted proteins, belong to the repeats-in-toxin (RTX) family (Welch (1991), *Mol Microbiol* **5**:512–528).

While the technologies based on the T1SS, in particular fusion proteins of peptides and proteins to be expressed with HlyA are known in the art, for example from the international patent publications WO 2013/057312 A1 and WO 2014/170430 A1, and are commercially available, it has been found that in various instances high yields are prevented by suboptimal solubility or susceptibility to proteolysis of the fusion constructs. Thus, there still exists need for further optimized methods that allow more efficient production of peptides and proteins that overcome some of the drawbacks of existing methods.

### SUMMARY OF THE INVENTION

The present invention is based on the inventor's surprising finding that two or more so-called GG repeats, short amino acid sequences of the general consensus sequence GGxGxDxUx derived from a specific class of naturally occurring proteins can be artificially combined and optionally inserted into a scaffold to form a novel, non-natural amino acid sequence (herein referred to as first amino acid sequence) that, if fused to another peptide or polypeptide of interest that is to be recombinantly produced by expression of a fusion protein in a host cell, provides for increased renaturation efficiency, while at the same time providing for good stability and/or solubility. Interestingly, it was found that efficiency of renaturation of the peptide or polypeptide of interest by the GG-repeat sequence containing part is dominated by the specific GG repeat sequence in the C-terminal position (based on N- to C-terminal orientation), e.g. the second position in case there are two GG repeats present, and may correlate to a specific consensus sequence that represents a narrower subgroup of the broad genus of all possible GG repeat sequences that satisfy the above more general consensus sequence. In contrast to the specific importance of the most C-terminal GG repeat sequence on renaturation efficiency, it was found

that the exact sequence of the first GG repeat (in N- to C-terminal orientation) has only little influence on renaturation efficiency. Based on this finding, artificial GG repeat combinations could be developed that performed significantly better than naturally occurring GG repeat combinations. Interestingly, it was found that by choosing such an approach replacements of natural GG repeat sequence elements could be designed that provide for improved renaturation efficiency. It needs to be noted that the possibility to shorten the renaturation tag while maintaining renaturation efficiency also provides for an improvement, as the relative yields of the peptide/polypeptide of interest will still be higher.

The newly developed amino acid sequences allow for the production of the fusion protein in form of high titer inclusion bodies with high initial purities after extraction while at the same time protecting fused peptides/proteins from degradation, e.g. proteolysis, desamidation, truncation, oxidation and the like and, most importantly, allowing their renaturation from a denatured into a properly folded functional state (by means of using the novel sequences as a fusion tag).

In a first aspect, the present invention is therefore directed to an isolated polypeptide comprising a first and a second amino acid sequence, wherein

(a) the first amino acid sequence

(a1) comprises in N- to C-terminus orientation the general amino acid sequence element



wherein GGR1 and each GGRN independently are a GG repeat sequence of the consensus sequence GGxGxDxUx, wherein each x can independently be any amino acid and U is independently a hydrophobic, large amino acid selected from F, V, A, Y, I, L and M;

wherein GGRC is a GG repeat sequence of the consensus sequence GGX<sup>11</sup>GX<sup>12</sup>DX<sup>13</sup>X<sup>14</sup>X<sup>15</sup>, wherein X<sup>11</sup> is selected from A and D; X<sup>12</sup> is selected from N, A, and D; X<sup>13</sup> is selected from T and S; X<sup>14</sup> is L or F; and X<sup>15</sup> is V or I;

wherein each "Linker" is independently either a peptide bond or an amino acid sequence of 1 to 25 amino acids not comprising a GG repeat sequence as defined above, preferably a peptide bond;

wherein n is 0 or an integer from 1 to 10, preferably 0 or 1 to 8, more preferably 0 or 1 to 4;

(a2) wherein the first amino acid sequence is 18 to 300, preferably 30 to 200 amino acids in length; and

(a3) wherein the first amino acid sequence and/or the amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is a non-natural sequence element, wherein, optionally, the amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%, and/or wherein, optionally, the complete first amino acid sequence is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%;

(b) the second amino acid sequence is at least one peptide or polypeptide of interest, and

(c) wherein the first and second amino acid sequence are heterologous to each other, preferably in that they do not occur in combination in a single naturally occurring polypeptide or fragment thereof,

wherein, preferably, the first amino acid sequence increases the solubility, stability and/or renaturation efficiency of the second amino acid sequence relative to a polypeptide not comprising a GG repeat sequence or not comprising the first amino acid sequence.

In various embodiments, the amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%, and/or wherein the complete first amino acid sequence is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%.

In various embodiments, the first and second amino acid are heterologous to each other in that they do not naturally occur in combination on the same natural polypeptide. This may mean that they originate or are derived from different naturally occurring polypeptides or that at least one is artificial in that it is not identical to any naturally occurring polypeptide or fragment thereof.

In various embodiments, GGRC is selected from any one of the amino acid sequences set forth in SEQ ID Nos. 1-48.

In various embodiments, the first amino acid sequence increases the solubility, stability and/or renaturation efficiency of the second amino acid sequence (for example determined by the test methods disclosed herein), in particular relative to a polypeptide comprising (only) the second amino acid sequence but not any GG repeat or not comprising the first amino acid sequence.

In various embodiments, the first amino acid sequence comprises only 2 GG repeats, namely those in the sequence element GGR1-Linker-GGRC.

In various other embodiments, the first amino acid sequence comprises at least 3, at least 4, at least 5, at least 6, at least 7 or at least 8 GG repeat sequences of the general consensus sequence GGxGxDxUx, such that the first amino acid sequence optionally comprises the general structure, from N- to C-terminus,

- (1) GGR1-Linker-GGRC; or
- (2) GGR1-Linker-GGR2-Linker-GGRC;
- (3) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGRC;
- (4) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGRC;
- (5) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGRC; or
- (6) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGRC; or
- (7) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGR7-Linker-GGRC;

wherein GGR2, GGR3, GGR4, GGR5, GGR6, and GGR7 are defined as GGRN above. In said embodiments, "Linker" and "GGRC" are defined as above.

In various embodiments, the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) disclosed above is flanked by additional N- and/or C-terminal amino acid sequences, optionally each being up to 150 amino acids in length. The N-terminal flanking sequence may be 1 to 100 amino acids in length, preferably 1 to 80, 1 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 8 to 20, 10 to 15 or 10 amino acids in length.

In various embodiments, the N-terminal flanking sequence may be derived from an RTX protein, i.e. may have an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein, for example any one of HlyA, CyaA, EhxA, LktA, PILkA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29, RTX30, RTX31, RTX32, and RTX33, with the sequences as defined herein below. In various embodiments, the sequence from which the flanking sequence is derived is also located N-terminal to any one of or all the natural GG repeats in said RTX protein. In various embodiments, the N-terminal flanking sequence may comprise, consist essentially of (i.e. at least 80% of its total length is made up by) or consist of the amino acid sequence set forth in SEQ ID NO:212, which is derived from HlyA. In various other embodiments, the N-terminal flanking sequence is derived from RTX12, in particular comprises or consists of the 10 to 46 amino acids directly N-terminal to the 6 GG repeats in the C-terminal part of RTX, said C-terminal part of RTX12 comprising the C-terminal 190 amino acids. Said N-terminal flanking sequence may thus comprise or consist of the amino acids starting from any one of positions 1 to 37 and extending up to position 46 of said 190 aa long C-terminal RTX12 fragment.

The C-terminal flanking sequence may be 10 to 150 amino acids in length, preferably 20 to 120, 30 to 120, 40 to 120, 50 to 120, 60 to 120, 70 to 120, 80 to 120, 90 to 120, 100 to 120 or 110 to 120 amino acids in length.

In various embodiments, the C-terminal flanking sequence may be derived from an RTX protein, i.e. may have an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein, for example any one of HlyA, CyaA, EhxA, LktA, PILkA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29,

RTX30, RTX31, RTX32, and RTX33, with the sequences as defined herein below. In various embodiments, the sequence from which the flanking sequence is derived is also located C-terminal to any one of or all the natural GG repeats in said RTX protein. In various embodiments, the C-terminal flanking sequence may comprise, consist essentially of (i.e. at least 80% of its total length is made up by) or consist of the amino acid sequence set forth in SEQ ID NO:213, which is derived from HlyA. In various other embodiments, the C-terminal flanking sequence is derived from RTX12, in particular comprises or consists of the 10 to 90 amino acids directly C-terminal to the 6 GG repeats in the C-terminal part of RTX, said C-terminal part of RTX12 comprising the C-terminal 190 amino acids. Said C-terminal flanking sequence may thus comprise or consist of the amino acids starting from position 101 and extending up to any one of positions 110 to 190 of said 190 aa long C-terminal RTX12 fragment.

Accordingly, the N- and/or C-terminal flanking sequences are scaffold sequences that may be derived from a natural occurring RTX protein, in which any one or more of the naturally-occurring GG repeat sequences have been replaced by the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) referred to above. The scaffold sequence may be derived from any RTX protein, in particular any one of HlyA, CyaA, EhxA, LktA, PILktA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29, RTX30, RTX31, RTX32, and RTX33, with the sequences as defined herein below. In some embodiments, they are derived from hemolysin A protein, preferably hemolysin A1 having the amino acid sequence set forth in SEQ ID NO:214 or a variant or fragment thereof, more preferably the amino acid sequence as set forth in SEQ ID NO:215 or a variant or fragment thereof, wherein the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) as disclosed above are inserted between positions 10 and 11 of SEQ ID NO:215 using the positional numbering of SEQ ID NO:215.

In various embodiments, the first amino acid sequence consists of the GG repeat element as disclosed herein, in particular the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) as disclosed above, and optionally the N- and/or C-terminal flanking sequences, as defined herein above, in particular the N- and/or C-terminal flanking sequences derived from RTX proteins. In various embodiments, the first amino acid sequence does not comprise a protease recognition site, such as a TEV protease recognition site.

In various embodiments, U in the consensus sequence GGxGxDxUx is selected from F, V, A, I, L, M and Y, for example F, V, I, L, M and Y. The consensus sequence of the GG repeat sequences may thus be GGX<sup>1</sup>GX<sup>2</sup>DX<sup>3</sup>UX<sup>4</sup>, wherein X<sup>1</sup>, X<sup>2</sup>, X<sup>3</sup> and X<sup>4</sup> may be any amino acid with the exception of P, preferably X<sup>1</sup> is selected from G, A, L, V, I, M, F, S, T, Q; Y, K, R, D, E, N, and Y, preferably from G, A, E, S, T, Q, L, R and D, more preferably from G, D, A, E, S and T; and/or X<sup>2</sup> is selected from N, D, A, G,

S, H, T, E, M, R and H, preferably from N, D, A, S, and H, more preferably from N, D, A, and S; and/or X<sup>3</sup> is selected from A, K, V, L, I, F, M, R, Y, S, L, T, V, Q, N, D, E, and H, preferably from T, R, V, L, S, I, A, Y, Q, D and H, more preferably from T, R, V, L, S and I, even more preferably from S, T, R and V; and/or X<sup>4</sup> is selected from W, K, R, Y, V, L, T, H, D, A, M, E, F, I, S, N and Q, preferably from V, L, I, F, S, R, N, Y and T.

In various embodiments, the sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is selected from the amino acid sequences set forth in SEQ ID Nos. 90-122, 124-131, 165, 170-185, 194-197, 200, 203, and 204.

The isolated polypeptide typically is a fusion protein, with the second amino acid sequence optionally being 10 to 1000 amino acids in length, preferably 15 to 500 amino acids in length. In various embodiments, the second amino acid sequence is N-terminal or C-terminal to the first amino acid sequence, preferably C-terminal. The second amino acid sequence may be linked directly or via a linker sequence to the N- or C-terminal end of the first amino acid sequence, the linker sequence optionally being 1 to 30 amino acids in length and preferably comprising a protease recognition and cleavage site. The protease recognition and cleavage site may be a TEV protease recognition and cleavage site.

The isolated polypeptide may have a total length of up to 1000 amino acids, preferably up to 800, up to 700, up to 600 or up to 500 amino acids, for example up to 350 amino acids.

In another aspect, the invention also relates to a nucleic acid, nucleic acid molecule or isolated nucleic acid molecule encoding the isolated polypeptide as described herein. In one aspect, said nucleic acid is part of a vector. One aspect thus features a (nucleic acid) vector comprising a nucleic acid molecule according to the invention. The vector may be an expression vector and may comprise additional nucleic acid sequences necessary to facilitate its function in a host cell. In various embodiments, the vector may be a plasmid.

One further aspect of the invention relates to a host cell comprising a nucleic acid molecule according to the invention or a vector according to the invention. The host cell may be a prokaryotic host cell, for example an *E.coli* cell.

In a still further aspect, the invention is directed to a method for the production of a polypeptide (isolated polypeptide) as described herein, comprising

- (1) cultivating the host cell described herein under conditions that allow the expression of the polypeptide; and
- (2) isolating the expressed polypeptide from the host cell; and, optionally,
- (3) renaturing the isolated polypeptide.

The method may in various embodiments comprise recovering the expressed peptide or protein from the host cell in form of insoluble protein aggregates, in particular inclusion bodies (IBs). Such embodiments, in which the fusion protein is accumulated within the host cells, optionally in insoluble/denatured form, may be particularly advantageous alternatives of such methods.

In various embodiments, the methods may comprise a step of re-solubilizing the peptide/protein and/or reconstituting/refolding it under suitable conditions. Said step is also referred to herein as “renaturing” step. Such a step is described in more detail below in relation to methods for renaturing the isolated polypeptide.

In still another aspect, the present invention is therefore also directed to methods for renaturing the isolated polypeptide of the invention, wherein the method may comprise contacting, for example in form of re-solubilizing, the isolated polypeptide with a suitable medium for renaturation, typically an aqueous medium, optionally comprising earth alkaline metal ions, including but not limited to calcium ions. It has been found that such a medium may trigger conformational changes in the GG repeats (comprised in the first amino acid sequence) that in turn then facilitate refolding/renaturation of the fused heterologous peptide or polypeptide (second amino acid sequence). This effect is independent of the type of fused peptide or polypeptide in that the conformational change of the first amino acid sequence effects the renaturation of the second amino acid sequence.

In various embodiments of the methods of the invention, the host cell is a prokaryotic cell, for example an *E.coli* cell. In various embodiments, the expression is performed in minimal culture medium. In various embodiments, the recombinant peptide or protein is purified using a method selected from affinity chromatography, ion exchange chromatography, reverse phase chromatography, size exclusion chromatography, and combinations thereof; and/or the method comprises treatment of the recombinant peptide or protein with a protease suitable for cleavage of a protease cleavage site within the recombinant peptide or protein and, optionally, said treatment with a protease is followed by purification of the recombinant peptide or protein.

In still another aspect, the present invention also relates to the use of an isolated polypeptide of the invention for facilitating the production of a recombinant peptide or protein. Said recombinant peptide or protein may be part of the isolated polypeptide, which is expressed as a fusion protein. In such embodiments, the first amino acid sequences disclosed herein may be used to facilitate renaturation/refolding of said recombinant peptide or polypeptide. Said recombinant peptide or protein is identical with the peptide or polypeptide of interest as referred to herein and may thus consist of or be comprised in the second amino acid sequence disclosed herein. As stated above, said use is preferably directed to recombinant polypeptides expressed as fusion proteins in a suitable expression system. The present invention thus features the use of the isolated polypeptide of the present invention for expression and renaturation of the second amino acid sequence in form of a fusion protein, wherein the expression is in a recombinant host cell in form of inclusion bodies.



It is understood that all combinations of the above disclosed embodiments are also intended to fall within the scope of the present invention.

#### **DETAILED DESCRIPTION OF THE INVENTION**

The terms used herein have, unless explicitly stated otherwise, the meanings as commonly understood in the art.

"At least one", as used herein, relates to one or more, in particular 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. "At least" if used in relation to a numerical value and in particular a list of numerical values relates to each of said separate numerical values and defines said numerical value as the minimum value. "At least 81, 82, 83, etc." thus means at least 81, at least 82, at least 83 and so on.

"Isolated" as used herein in relation to a molecule means that said molecule has been at least partially separated from other molecules it naturally associates with or other cellular components. "Isolated" may mean that the molecule has been purified to separate it from other molecules and components, such as other proteins and nucleic acids and cellular debris, in particular those that accompany it due to its recombinant production in host cells.

"Nucleic acid" as used herein includes all natural forms of nucleic acids, such as DNA and RNA. Preferably, the nucleic acid molecules of the invention are DNA.

The term "peptide" is used throughout the specification to designate a polymer of amino acid residues connected to each other by peptide bonds. A peptide according to the present invention may have 2-100 amino acid residues. The terms "protein" and "polypeptide" are used interchangeably throughout the specification to designate a polymer of amino acid residues connected to each other by peptide bonds. A protein or polypeptide according to the present invention has preferably more than 100 amino acid residues.

The term "an N-terminal fragment" relates to a peptide or protein sequence which is in comparison to a reference peptide or protein sequence C-terminally truncated, such that a contiguous amino acid polymer starting from the N-terminus of the peptide or protein remains. In some embodiments, such fragments may have a length of at least 30 amino acids, at least 50 amino acids or at least 70 amino acids.

The term "a C-terminal fragment" relates to a peptide or protein sequence which is in comparison to a reference peptide or protein sequence N-terminally truncated, such that a contiguous amino acid polymer starting from the C-terminus of the peptide or protein remains. In some embodiments, such fragments may have a length of at least 30 amino acids, at least 50 amino acids or at least 70 amino acids.

The term "fusion protein" as used herein concerns two or more peptides and proteins which are N- or C-terminally connected to each other, typically by (a) peptide bond(s), including via an amino acid/peptide linker sequence. Such fusion proteins may be encoded by two or more nucleic acid sequences which are operably fused to each other. In certain embodiments, a fusion protein refers to at least one peptide or protein of interest C-terminally or N-terminally fused to a first amino acid sequence according to the invention. In various embodiments, the peptide or protein of interest is fused to the C-terminus of the first amino acid sequence, optionally via a linker sequence.

"Stability", as used herein in relation to the polypeptides of the invention, primarily relates to resistance to proteolytic degradation, which is a commonly encountered issue, in particular in conditions of high cell density and/or high protein yields.

"Solubility", as used herein in relation to the polypeptides of the invention, primarily relates to solubility in the cytoplasm and/or culture medium so that the polypeptide can be successfully secreted and purified from the cultivation medium.

"Renaturation" and "renaturation efficiency", as used herein in relation to the polypeptides of the invention, primarily relates to the possibility to isolate the polypeptide in denatured/unfolded form, for example from insoluble protein aggregates, such as inclusion bodies, and induce its proper folding to the active, three-dimensional conformation. The efficiency may be given as the mass of successfully renatured protein relative to the total mass of all expressed protein, including insoluble protein, in percent or the solubilized protein mass in percent. The refolding may be effected in special refolding buffers including earth alkaline metal ions, in particular calcium ions, to induce refolding of the first amino acid sequence, which in turn facilitates renaturation of the complete fusion construct into its functional conformation. "Renaturing" and "refolding" are used interchangeably herein and generally relate to the process of folding a given protein that is unfolded or denatured such that it adopts its functional three-dimensional conformation.

Generally, the skilled person understands that for putting the present invention into practice any nucleotide sequence described herein may comprise an additional start and/or stop codon or that a start and/or stop codon included in any of the sequences described herein may be deleted, depending on the nucleic acid construct used. The skilled person will base this decision, e.g., on whether a nucleic acid sequence comprised in the nucleic acid molecule of the present invention is to be translated and/or is to be translated as a fusion protein. In various embodiments, the isolated polypeptides of the invention additionally comprise the amino acid M on the N-terminus.

Determination of the sequence identity of nucleic acid or amino acid sequences can be done by a sequence alignment based on well-established and commonly used BLAST algorithms (See, e.g. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search

tool." J. Mol. Biol. 215:403-410, and Altschul, Stephan F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Hheng Zhang, Webb Miller, and David J. Lipman (1997): "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs"; Nucleic Acids Res., 25, S.3389-3402). Such an alignment is based on aligning similar nucleotide or amino acid sequences stretches with each other. Another algorithm known in the art fo said purpose is the FASTA algorithm. Alignments, in particular multiple sequence comparisons, are typically done by using computer programs. Commonly used are the Clustal series (See, e.g., Chenna et al. (2003): Multiple sequence alignment with the Clustal series of programs. Nucleic Acid Research 31, 3497-3500), T-Coffee (See, e.g., Notredame et al. (2000): T-Coffee: A novel method for multiple sequence alignments. J. Mol. Biol. 302, 205-217) or programs based on these known programs or algorithms.

Also possible are sequence alignments using the computer program Vector NTI® Suite 10.3 (Invitrogen Corporation, 1600 Faraday Avenue, Carlsbad, CA, USA) with the set standard parameters, with the AlignX module for sequence comparisons being based on the ClustalW. If not indicated otherwise, the sequence identity is determined using the BLAST algorithm.

Such a comparison also allows determination of the similarity of the compared sequences. Said similarity is typically expressed in percent identify, i.e. the portion of identical nucleotides/amino acids at the same or corresponding (in an alignment) sequence positions relative to the total number of the aligned nucleotides/amino acids. For example, if in an alignment 90 amino acids of a 100 aa long query sequence are identical to the amino acids in corresponding positions of a template sequence, the sequence identity is 90%. The broader term "homology" additionally considers conserved amino acid substitutions, i.e. amino acids that are similar in regard to their chemical properties, since those typically have similar chemical properties in a protein. Accordingly, such homology can be expressed in percent homology. If not indicated otherwise, sequence identity and sequence homology relate to the entire length of the aligned sequence.

In the context of the present invention, the feature that an amino acid position corresponds to a numerically defined position in a reference sequence means that the respective position correlates to the numerically defined position in said reference sequence in an alignment obtained as described above.

The hemolysin (Hly) secretion system is a protein secretion system which mostly occurs in gram-negative bacteria. This secretion system belongs to the family of type I secretion systems which transport their substrates in an ATP driven manner in a single step from the cytosol to the extracellular space without an intermediate station in the periplasm. The Hly secretion system comprises hemolysin B (HlyB) which represents an ATP-binding cassette (ABC) transporter, the membrane fusion protein hemolysin D (HlyD), and the universal outer membrane protein TolC. The 110 kDa hemolytic toxin hemolysin A (HlyA) is a transport substrate of the Hly secretion system. On genetic level, the components necessary for hemolysin A-specific secretion are organized in an operon structure. The

nucleic acid sequence encoding for hemolysin C (HlyC) also forms part of this operon but is not required for HlyA secretion through the Hly secretion system. HlyC catalyzes acylation of HlyA which renders HlyA hemolytic. HlyA is a protein which consists of 1024 amino acid residues and requires for its export via the Hly secretion system its C-terminus, comprising about 40-60 amino acids. Furthermore, HlyA is characterized in that it comprises N-terminally to the 40-60 C-terminal amino acids a domain comprising several glycine rich (GG) repeats. The Hly secretion system and its T1 SS allocrit HlyA are used as a scaffold herein for the newly combined GG repeat sequences to demonstrate the functionality of the inventive concept. However, it is to be understood that the invention is not limited to these constructs derived from HlyA, since similar results have also been obtained with the GG repeat-containing sequences of other RTX proteins.

“GG repeat”, “GG repeat sequence” and “glycine rich repeat”, as used herein interchangeably, are the characteristic of the repeats in toxin (RTX) toxin family and have the consensus sequence disclosed herein. The glycine rich repeats bind  $\text{Ca}^{2+}$  which induces their folding. Hence, in absence of earth alkaline metal ions, in particular  $\text{Ca}^{2+}$ , the domain comprising the glycine rich repeats is typically unstructured.

“Flanking sequence”, as used herein, relates to an amino acid sequence that is in the immediate vicinity of a given sequence element, i.e. is connected to the N- or C-terminus of said given sequence element by a peptide bond. The flanking sequence may have a length as defined herein, with the positional numbering of such a flanking sequence starting such that position 1 is the amino acid directly neighboring the sequence element and then counting away from the sequence element that is flanked either upstream, i.e. in the direction of the N-terminus if it flanks the N-terminus of the sequence element, or in direction of the C-terminus if it flanks the C-terminus of the sequence element, unless explicitly indicated otherwise

The present invention is based on the inventor's surprising finding that two or more naturally occurring GG repeat sequences may be newly combined and, optionally, inserted into a scaffold structure to yield non-naturally occurring amino acid sequences which when fused to a peptide/polypeptide of interest allow to efficiently renature the complete fusion protein including the peptide/polypeptide of interest due to the presence of the GG repeats, with the renaturation efficiency being at least equal to or higher than that obtained with a natural RTX protein, in particular if the C-terminal GG repeat has a specific consensus sequence, identified by the inventors.

“Non-natural”, as used herein in relation to the first amino acid sequences and general sequence elements, means that said sequence does not naturally occur in an organism, but has been obtained by protein/peptide engineering. The first amino acid sequence is thus generally a recombinant sequence either by virtue of the GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC element already being non-natural or by one or more of the GGR repeats and the remainder of the first amino acid sequence being heterologous to each other, i.e. being in their totality not identical to a naturally occurring protein or fragment thereof.

In various embodiments, the GG repeat sequences and the additional flanking sequences of the scaffold, which are not GG repeat sequences, are heterologous to each other, i.e. do not occur in combination in any naturally occurring protein or peptides. The term "non-natural" also means that the first amino acid sequence is not a fragment of a naturally occurring polypeptide, but rather differs from any such fragment of a naturally occurring polypeptide in sequence. The sequence identity of the first amino acid sequence over its entire length is thus not 100% relative to any naturally occurring protein, polypeptide or fragment thereof, but always differs in a least 1 amino acid position. In some embodiments, this may also be true for the GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC element, i.e. it is over its entire length not 100 % identical to naturally occurring polypeptide or fragment thereof. In other embodiments, the GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC element can be identical to a fragment of a naturally occurring polypeptide, but then the remainder of the first amino acid sequence is different from said naturally occurring polypeptide or fragment thereof.

"Heterologous", as used herein in relation to amino acid sequences, means that the respective amino acid sequences do not occur in combination in a single natural or naturally occurring protein or polypeptide or fragment thereof. This means that these two sequence elements do not occur in the same linear polypeptide sequence of a known natural polypeptide. In various embodiments, the second amino acid sequence may be a mammalian sequence, including *Homo sapiens* sequence, while the first amino acid sequence is no mammalian sequence, but rather is an artificial, non-natural sequence. As the first amino acid sequence is always non-natural, i.e. is excluded from being identical to any naturally occurring polypeptide or fragment thereof, thus being an artificial sequence, the feature that the first and second amino acid sequence are heterologous to each other is essentially understood.

The present invention is directed to such fusion proteins that comprise an artificial, non-natural "expression and renaturation tag" that comprises at least two GG repeats as defined herein that imparts the desired properties of increased renaturation efficiency to the fusion protein. This is achieved by using said tag sequence as the first amino acid sequence, while the peptide(s) and/or polypeptide(s) of interest, i.e. the cargo that is to be produced, is the second amino acid sequence. These fusion proteins are also more generally referred to herein as polypeptides that are isolated after production. These fusion proteins and isolated polypeptides are artificial constructs, since the first and second amino acid sequences are necessarily heterologous to each other and the fusion construct has been genetically engineered.

As it is desirable to have maximum yields of the peptide or protein of interest, the first amino acid sequences of the invention are preferably as short as possible, generally with a maximum length of about 250 amino acids, typically about 200 amino acids or less. Herein constructs are used and exemplified wherein the first amino acid sequence has a length of about 150 amino acids. Together with optional linker and protease cleavage sites, the total length of the part of the fusion protein not including the peptide or protein of interest is typically limited to about 250 amino acids or less. In various embodiments, it may be desirable to use first amino acid sequences significantly short than the upper

limit of 200 amino acids, such as up to 190, up to 180, up to 170, up to 160, or up to 150 amino acids in length. In various embodiments, the first amino acid sequence is long enough to accommodate 4, 5, or 6 GG repeats, i.e. is at least 36, at least 45 or at least 54 amino acids in length.

In the polypeptides described herein, the first amino acid sequence functions by increasing the solubility, stability and/or renaturation efficiency, in particular the renaturation efficiency, of the second amino acid sequence relative to a polypeptide not comprising any GG repeat sequence or not comprising the first amino acid sequence as defined herein or a reference sequence that comprises the scaffold of SEQ ID NO:215 and the GG repeat cassette of SEQ ID NO:218 inserted between positions 10 and 11 of SEQ ID NO:215 as the first amino acid sequence. In various other embodiments, the solubility, stability and/or renaturation efficiency, in particular the renaturation efficiency, of the second amino acid sequence is at least equal to, preferably higher than that obtained with a reference sequence that comprises the scaffold of SEQ ID NO:215 and the GG repeat cassette of SEQ ID NO:218 inserted between positions 10 and 11 of SEQ ID NO:215 as the first amino acid sequence. The renaturation efficiency may be determined by expressing the polypeptides of the invention as inclusion bodies, for example in *E.coli* host cells, isolating the inclusion bodies, for example by known kits (such via BugBuster Kit) and then dissolving in 6 M GuHCl (guanidium hydrochloride) (1:4 w/v). After solubilization (overnight), renaturation can be performed in a suitable renaturation buffer with or without calcium ions, with the renaturation efficiency being determined in the cleared supernatant by UV/Vis spectroscopy (renaturation is typically done at room temperature for 20 min).

The first amino acid sequence comprises a first GG repeat sequence, herein referred to as GGR1, and a second GG repeat sequence, herein referred to as GGRC. The terms "first" and "second" as used in this context, as well as any further numbering terms, such as "third" etc., may relate to the position of the respective GG repeat in the N- to C-terminal direction. The GG repeat closest to the N-terminus is thus always referred to as first GG repeat, the one following in C-terminal direction is the second GG repeat and so on. However, this does not apply to the "first amino acid sequence" and "second amino acid sequence" and any further amino acid sequence, which may be arranged in any order.

The first GG repeat, i.e. GGR1, as well as any non-C-terminal GG repeat GGRN has the general consensus sequence GGxGxDxUx, wherein x can be any amino acid and U is a hydrophobic, large amino acid selected from F, V, A, I, L, M and Y. In various embodiments, the U in the consensus sequence is selected from F, V, I, L, M and Y or F, I, L, M and Y. In various embodiments, x is typically selected from the 20 proteinogenic amino acids G, A, V, L, I, F, C, M, P, R, K, H, N, Q, D, E, S, T, W, and Y. If not indicated otherwise, it is generally preferred that any amino acid not specifically defined is selected from these 20 proteinogenic amino acids. In various embodiments, x may be any one of the above amino acids with the exception of proline. In even more specific embodiments, the consensus sequence may be GGX<sup>1</sup>GX<sup>2</sup>DX<sup>3</sup>UX<sup>4</sup>, wherein X<sup>1</sup>, X<sup>2</sup>, X<sup>3</sup> and X<sup>4</sup> may be any amino acid with the exception of P. In various embodiments, X<sup>1</sup> is selected from G, A, L, V, I, M, F, S, T, Q; Y, K, R, D, E, N, and Y, for example from G, A, E, S, T, Q, L, R and D, or from G, D, A, E, S and T. In various

embodiments,  $X^2$  is selected from N, D, A, G, S, H, T, E, M, R and H, for example from N, D, A, S, and H, or from N, D, A, and S. In various embodiments,  $X^3$  is selected from A, K, V, L, I, F, M, R, Y, S, L, T, V, Q, N, D, E, and H, for example from T, R, V, L, S, I, A, Y, Q, D and H, or from T, R, V, L, S and I, or from S, T, R and V. In various embodiments,  $X^4$  is selected from W, K, R, Y, V, L, T, H, D, A, M, E, F, I, S, N and Q, for example from V, L, I, F, S, R, N, Y and T. These embodiments of  $X^1$ - $X^4$  may be combined with each other.

In contrast to GGR1 and all other non-C-terminal GG repeats, the C-terminal GG repeat sequence GGRC has a specific consensus sequence, namely  $GGX^{11}GX^{12}DX^{13}X^{14}X^{15}$ , wherein  $X^{11}$  is selected from A and D;  $X^{12}$  is selected from N, A, and D;  $X^{13}$  is selected from T and S;  $X^{14}$  is L or F; and  $X^{15}$  is V or I, for example any one of the amino acid sequences set forth in SEQ ID Nos. 1-48, such as, for example SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, or 48.

It has been surprisingly found by the inventors that the C-terminal GG repeat sequence in a given sequence element, such as the first amino acid sequence, dominates the renaturation efficiency of the first amino acid sequence and also its effect on any second amino acid sequence fused thereto in a sequence specific matter. This means that certain GG repeat sequences if used as C-terminal GG repeats, GGRC, can surprisingly influence the renaturation of the fusion protein to a greater extent than the same sequence in another position, such as GGR1.

With respect to the specific sequences useful as GGRC, it has been found that efficiency in increasing renaturation roughly correlates with matching the above defined specific consensus sequence  $GGX^{11}GX^{12}DX^{13}X^{14}X^{15}$  that only allows a limited number of amino acids in the variable positions of the GG repeat.

In contrast, while the first GG repeat is a necessary structural element for effecting renaturation of the fusion protein, its sequence is much less relevant. This means that the presence of any GG repeat sequence as GGR1 or GGRN is sufficient for good renaturation, while GGRC needs to be selected from those GG repeat sequences that match the given consensus sequence.

In various embodiments, the first amino acid sequence may comprise more than 2 GG repeat sequences, such as at least 3, at least 4, at least 5, at least 6, at least 7 or at least 8 GG repeat sequences of the general consensus sequence  $GGxGxDxUx$ . However, while it has been found that in some few instances a third and fourth GG repeat may further increase renaturation efficiency, although not to the same extent as the step from one to two GG repeats, adding a fifth and further GG repeat sequence appears to only slightly influence renaturation efficiency in most cases. As the use of only two GG repeat sequences as reported herein has been found to perform significantly better than a natural sequence of 3 GG repeats and one GG repeat-like sequence and it is generally desirable to keep the first amino acid sequence as short as possible, in various embodiments, it is preferred to use 2, 3 or 4

GG repeats. As stated above, if more than 2 GG repeat sequences are present, it has been found that it is the most C-terminal GG repeat sequence that has the most influence on renaturation efficiency.

Accordingly, in various embodiments where the first amino acid sequence comprises more than two GG repeat sequences, at least the most C-terminal GG repeat is defined as GGRC herein above.

In the first amino acid sequences of the invention, the GG repeats may be arranged directly adjacent to each other, i.e. directly connected by peptide bonds, or they may be separated by linkers. If there are more than 2 GG repeats both options can also be mixed in that some GG repeats are directly adjacent, i.e. connected via peptide bonds, while others are connected via linkers ("Linker"). These linker sequences may be amino acid sequences of 1 to 25 amino acids, preferably 1 to 20 or 1 to 15 or 1 to 10 or 1 to 9 or 1 to 8 or 1 to 7 or 1 to 6 or 1 to 5 or 1 to 4 or 1 to 3 or 1-2 amino acids in length. These are not limited with respect to their sequences (aside from not containing a GG repeat sequence), but, in various embodiments, it may be preferred to keep these linking sequences short, for example of only 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid in length. If two or more GG repeat sequences are contained in the first amino acid sequence it may be preferred that at least two, or depending on the number present, at least three, at least 4, at least 5, at least 6 or all of them are directly linked to each other by a peptide bond. It has been found, for example, that in some embodiments constructs where the first amino acid sequence comprises GG repeat sequences, with these sequences directly linked to each other, perform particularly well so that the presence of such linker sequences between the individual GG repeat sequences may be unnecessary in most cases. In various embodiments, all "Linker" elements in the GG repeat cassette are a peptide bond.

In various embodiments, the first amino acid sequence therefore comprises the general structure, from N- to C-terminus:

GGR1-Linker-GGRC.

This sequence element may be further expanded to be:

- (2) GGR1-Linker-GGR2-Linker-GGRC;
- (3) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGRC;
- (4) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGRC;
- (5) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGRC;
- (6) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGRC; or
- (7) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGR7-Linker-GGRC.

In these embodiments, GGR1, GGR2, GGR3, GGR4, GGR5, GGR6, and GGR7 are GG repeats of the consensus sequence GGxGxDxUx, as defined above, and each "Linker" is independently either a peptide bond or an amino acid sequence of 1 to 25 amino acids, preferably 1 to 20 or 1 to 15 amino



acids in length. The "Linker" does not comprise a GG repeat sequence. As disclosed above, it may be preferred that at least one "Linker" is a peptide bond. It is understood that also constructs with more than 8 GG repeats are covered, with the ninth and further GG repeat also linked via a "Linker" as herein defined. It is further understood that the first amino acid sequence does not contain any further GG repeat sequences beyond those indicated in the general GG repeat sequence element and/or structures (1) to (7) above.

While the GG repeat sequences disclosed may be combined in any possible manner covered by the present disclosure, in various embodiments, the GG repeat sequence element in the first amino acid sequence in the isolated polypeptide has the amino acid sequence set forth in any one of SEQ ID Nos. 90-122, 124-131, 165, 170-185, 194-197, 200, 203, and 204.

In all the afore-mentioned embodiments, the linker may be as defined herein, but is preferably a peptide bond so that all GG repeats are directly adjacent to each other.

The GG repeat sequence elements of the first amino acid sequence disclosed herein and more particularly the general structures given above, i.e. GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC and (1) to (7) as well as the specific embodiments disclosed above, may be flanked by additional N- and/or C-terminal amino acid sequences. These do not contain any GG repeat sequences and are each typically up to 200, preferably up to 150 amino acids in length. In various embodiments, these flanking sequences are of a length that the preferred length of the first amino acid sequence disclosed above is not exceeded.

In various embodiments, the N-terminal flanking sequence is 1 to 100 amino acids in length, preferably 1 to 80, 1 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 8 to 20, 10 to 15 or 10 amino acids in length. It can be preferred that the N-terminal flanking sequence is as short as possible.

In various embodiments, the N-terminal flanking sequence may be derived from an RTX protein, for example may have an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein, for example any one of HlyA, CyaA, EhxA, LktA, PILktA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29, RTX30, RTX31, RTX32, and RTX33, with the sequences as defined herein below. As defined above, the flanking sequence does not comprise any further GG repeat sequences but are rather derived from the scaffold sequences flanking the GG repeat elements in these naturally occurring RTX proteins. It is thus preferred that they comprise a sequence that starts with the amino acid directly neighboring the GG repeat sequence in the RTX protein and extending into the direction of the N-

terminus up to the desired length. While it may be preferred that these flanking sequences have a sequence identity to the naturally occurring flanking sequence as high as possible, they may comprise some sequence deviations. Accordingly, in various embodiments, it has an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein.

In various embodiments, the sequence from which the flanking sequence is derived is also located N-terminal to any one of or all the natural GG repeats in said RTX protein.

In various embodiments, the N-terminal flanking sequence comprises, consists essentially of (i.e. at least 80% of its total length is made up by the given sequence or it does not contain more than 1-10 additional amino acids) or consists of the amino acid sequence GNSLAKNVLF (SEQ ID NO:212). This sequence is derived from HlyA and artificially combined with the GG repeat elements disclosed herein.

In various other embodiments, the N-terminal flanking sequence is derived from RTX12, in particular comprises or consists of the 10 to 46 amino acids directly N-terminal to the 6 GG repeats in the C-terminal part of RTX, said C-terminal part of RTX12 comprising the C-terminal 190 amino acids. Said N-terminal flanking sequence may thus comprise or consist of the amino acids starting from any one of positions 1 to 37 and extending up to position 46 of said 190 aa long C-terminal RTX12 fragment. The scaffold may thus be the sequence set forth in SEQ ID NO:189 with the GG repeat element of SEQ ID NO:194 comprised therein replaced by another GG repeat sequence element as defined herein.

In various embodiments, the C-terminal flanking sequence is 10 to 150 amino acids in length, preferably 20 to 120, 30 to 120, 40 to 120, 50 to 120, 60 to 120, 70 to 120, 80 to 120, 90 to 120, 100 to 120 or 110 to 120 amino acids in length. In various embodiments, the C-terminal flanking sequence is longer than the N-terminal flanking sequence.

In various embodiments, the C-terminal flanking sequence may be derived from an RTX protein, i.e. may have an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein, for example any one of HlyA, CyaA, EhxA, LktA, PILktA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29, RTX30, RTX31, RTX32, and RTX33, with the sequences as defined herein below. As defined above, the flanking sequence does not comprise any further GG repeat sequences but are rather derived from the scaffold sequences flanking the GG repeat elements in these naturally occurring RTX proteins. It is

thus preferred that they comprise a sequence that starts with the amino acid directly neighboring the GG repeat sequence in the RTX protein and extending into the direction of the C-terminus up to the desired length. While it may be preferred that these flanking sequences have a sequence identity to the naturally occurring flanking sequence as high as possible, they may comprise some sequence deviations. Accordingly, in various embodiments, it has an amino acid sequence that has at least 80% sequence identity, preferably at least 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 % sequence identity, over its entire length to the amino acid sequence of any RTX protein.

In various embodiments, the sequence from which the flanking sequence is derived is also located C-terminal to any one of or all the natural GG repeats in said RTX protein.

In various embodiments, the C-terminal flanking sequence comprises, consists essentially of (i.e. does not contain more than 1 to 50 additional amino acids or at least 80% thereof are made up by the given sequence) or consist of the amino acid sequence set forth in SEQ ID NO:213. This sequence is derived from HlyA and artificially combined with the GG repeat elements disclosed herein.

In various other embodiments, the C-terminal flanking sequence is derived from RTX12, in particular comprises or consists of the 10 to 90 amino acids directly C-terminal to the 6 GG repeats in the C-terminal part of RTX, said C-terminal part of RTX12 comprising the C-terminal 190 amino acids. Said C-terminal flanking sequence may thus comprise or consist of the amino acids starting from position 101 and extending up to any one of positions 110 to 190 of said 190 aa long C-terminal RTX12 fragment. The scaffold may thus be the sequence set forth in SEQ ID NO:189 with the GG repeat element of SEQ ID NO:194 comprised therein replaced by another GG repeat sequence element as defined herein.

In various embodiments, the N- and/or C-terminal flanking sequences are scaffold sequences derived from a naturally occurring RTX protein, in which any one or more of the naturally occurring GG repeat sequences have been replaced by the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7). It may be that the complete GG repeat cassette in the respective RTX protein is replaced by the GG repeat elements disclosed herein and that the sequences N- and C-terminal to this cassette in the natural protein are N- and/or C-terminally truncated such that only the directly adjacent flanking sequences remain. These N- and C-terminal flanking sequences in combination form a scaffold in which the GG repeat sequence elements disclosed herein can be inserted.

In various embodiments, the flanking sequences, for example in combination, may be fragments of a naturally occurring RTX protein, such as those disclosed herein below. Such fragments are typically at least 30 amino acids in length with a maximum length of 200 amino acids and do not comprise GG repeat sequences. In various embodiments, said fragments, in particular those parts used as C-terminal flanking sequences, do not retain a functional, typical C-terminal, secretion signal, e.g. at least the C-terminal 40, preferably 50 or 60 amino acids (i.e. typically the 40 to 60 C-terminal amino acids of the

native, naturally occurring protein) are not part of the fragment. The fragments used as a scaffold structure, i.e. the combined flanking sequences, or the C-terminal flanking sequences are thus typically N-terminal fragments that may additionally lack most of the N-terminus and retain only the amino acid sequences in direct vicinity of the replaced GG repeats. Accordingly, the fragments used as a scaffold and in particular the part used as an N-terminal flanking sequence are thus typically also C-terminal fragments that lack most of the N-terminus and retain only the amino acid sequences in direct vicinity of the replaced GG repeats. These retained sequences adjacent to the core GG repeat sequence and derived from RTX proteins fragments are referred to herein as “flanking sequences”.

These RTX proteins from which the flanking sequences are derived may be allocrits of a T1 SS and may be selected from the group consisting of, without limitation, HlyA, CyaA, EhxA, LktA, PILkTA, PasA, PvxA, MmxA, LtxA, ApxIA, ApxIIA, ApxIIIA, ApxIVA, ApxI, Apx11, AqxA, VcRtxA, VvRtxA, MbxA, RTX cytotoxin, RtxL1, RtxL2, FrhA, LipA, TliA, PrtA, PrtSM, PrtG, PrtB, PrtC, AprA, AprX, ZapA, ZapE, Sap, HasA, colicin V, LapA, ORF, RzcA, RtxA, XF2407, XF2759, RzcA, RsaA, Crs, CsxA, CsxB, SlaA, SwmA, S111951, NodO, PlyA, PlyB, FrpA, FrpC, RTX1, RTX3, RTX7, RTX8, RTX9, RTX12, RTX14, RTX16, RTX18, RTX19, RTX20, RTX21, RTX22, RTX24, RTX26, RTX28, RTX29, RTX30, RTX31, RTX32, and RTX33. The amino acid sequences of the recited RTX proteins are available in the Uni Prot database under the following accession numbers:

RTX1 (SCP-domain containing protein, *Aromatoleum aromaticum strain EbN1*; NCBI Reference WP\_011239843.1; UniProt Accession No. Q5NXL6),

RTX3 (Serralyisin (*Caulobacter sp.* strain K31); UniProt Accession No. B0T558),

RTX7 (uncharacterized protein of *Pelobacter propionicus*; UniProt Accession No. A1ARK7; NCBI Reference WP\_011736233.1),

RTX8 (Metallopeptidase AprX (*Pseudomonas fluorescens strain ATCC BAA-477/ NRRL B-23932 / Pf-5*); UniProt Accession No. Q4KDU3; NCBI Reference WP\_011060780.1),,

RTX9 (glycosyl hydrolase family 16, hemolysin-type calcium-binding repeat protein (*Rhodobacter sphaeroides strain ATCC 17023 / DSM 158 / JCM 6121*); UniProt Accession No. Q3J0J9; NCBI Reference WP\_011338282.1),

RTX12 (endonuclease/exonuclease/phosphatase (*Caulobacter sp.* strain K31); UniProt Accession No. B0T829; NCBI Reference WP\_012287684.1),

RTX14 (Glycoside hydrolase family 16 (*Caulobacter sp.* strain K31); UniProt Accession No. B0T438),

RTX16 (lipase class 3 (*Chlorobium phaeobacteroides strain DSM266*); UniProt Accession No. A1BIU7; NCBI Reference WP\_011746110.1),

RTX18 (animal haem peroxidase (*Leptotrix cholodnii strain ATCC 51168 / LMG 8142 / SP-6*); UniProt Accession No. B1Y442; NCBI Reference WP\_012347322.1),

RTX19 (putative outer membrane adhesin like protein (*Magnetococcus marinus strain ATCC BAA-1437 / JCM 17883 / MC-1*); UniProt Accession No. A0L6L9; ENA Reference ABK43612.1),

RTX20 (putative outer membrane adhesin like protein (*Magnetococcus marinus strain ATCC BAA-1437 / JCM 17883 / MC-1*); UniProt Accession No. A0LDY6; NCBI Reference WP\_011715232.1),

RTX21 (Glycerophosphoryl diester phosphodiesterase (*Methylobacterium nodulans* (strain LMG 21967 / CNCM I-2342 / ORS 2060); UniProt Accession No. B8ICD8; EMBL Reference ACL55526.1),  
RTX22 (5' nucleotidase domain protein (*Methylobacterium* sp. strain 4-64); UniProt Accession No. B0UQ63; NCBI Reference WP\_012331605.1),  
RTX24 (polyurethanase A (*Pseudomonas fluorescens* strain ATCC BAA-477 / NRRL B 23932 / Pf-5); UniProt Accession No. Q4KBS6; NCBI Reference WP\_011061486.1),  
RTX26 (extracellular alkaline metalloprotease AprA (*Pseudomonas fluorescens* strain ATCC BAA-477 / NRRL B-23932 / Pf-5); UniProt Accession No. Q4KBR8; NCBI Reference WP\_011061494.1),  
RTX28 (hemolysin type calcium-binding protein (*Rhodobacter sphaeroides* strain ATCC 17023 / DSM 158 / JCM 6121); UniProt Accession No. Q3IVG4; NCBI Reference WP\_011331290.1),  
RTX29 (neutral zinc metallopeptidase (*Rhodobacter sphaeroides* strain ATCC 17023 / DSM 158 / JCM 6121); UniProt Accession No. Q3J1D3; NCBI Reference WP\_011338088.1),  
RTX30 (glycoside hydrolase family 16 (*Rhodobacter sphaeroides* strain ATCC 17029 / ATH 2.4.9); UniProt Accession No. A3PLQ2; NCBI Reference WP\_011841485.1),  
RTX31 (peptidase domain protein (*Sinorhizobium medicae* strain WSM419); UniProt Accession No. A6UK74; NCBI Reference WP\_011970163.1),  
RTX32 (Serralysin (*Sinorhizobium medicae* strain WSM419); UniProt Accession No. A6UK83; NCBI Reference WP\_011970172.1), and  
RTX33 (hemolysin type calcium-binding region (*Verminephrobacter eiseniae* strain EF01-2); UniProt Accession No. A1WL15; NCBI Reference WP\_011810323.1).

As disclosed herein, neither are these proteins used as such as the first amino acid sequence, nor are fragments thereof that contain the natural GG repeat sequences used as the first amino acid sequence. Rather only the flanking sequences of these proteins are used to create a scaffold sequence for the newly developed GG repeat elements disclosed herein and combined such as to form a non-natural first amino acid sequence that does not share 100% sequence identity with any natural protein sequence or fragment thereof over its entire length (meaning the GG repeat element plus optional flanking sequences). In various embodiments, the sequence identity of such a first amino acid sequence to a natural protein or fragment thereof is 99% or less, 98% or less, 97% or less, 96% or less, 95% or less, 94% or less, 93% or less, 92% or less, 91% or less, 90% or less, 89% or less, 88% or less, 87% or less, 86% or less, or 85% or less.

The scaffold sequence may be derived from hemolysin A protein, preferably hemolysin A1 having the amino acid sequence set forth in SEQ ID NO:214 or a variant or fragment thereof, more preferably the amino acid sequence as set forth in SEQ ID NO:215 or a variant or fragment thereof, wherein the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any one of the general structures (1)-(7) disclosed herein is inserted between positions 10 and 11 using the positional numbering of SEQ ID NO:215.

In alternative embodiments, the scaffold sequence formed by the flanking sequences can be derived from any other RTX protein, such as RTX1, RTX9, RTX12, RTX18, RTX21, RTX22, RTX30, RTX33, RTX28, RTX24, RTX19, RTX16, RTX26, RTX3, RTX20, RTX31, RTX14, RTX32 and RTX29.

In various exemplary embodiments, the flanking sequences may be derived from a known C-terminal fragment of HlyA, namely HlyA1 (SEQ ID NO:214). Said known fragment comprises 3 GG repeats and one GG repeat like sequence. These GG repeats are deleted in such scaffolds and replaced by a GG repeat element that differs from the natural, deleted GG repeat element. This may mean that the GG repeat element used, in particular if it is GGR1-Linker-GGR2-Linker-GGRC, is not that corresponding to amino acids 11-46 of SEQ ID NO:214 using the positional numbering of this sequence. In various embodiments, the scaffold is a further shortened version of HlyA1 (with deleted GG repeat cassette, i.e. positions 11 to 46 of SEQ ID NO:214 being deleted) obtained by any one or more of an N-terminal truncation, a C-terminal truncation or a deletion of one or more amino acids, in particular a C-terminal truncation. In such embodiments, the remaining length of the scaffold may be at least 50 amino acids in length, preferably at least 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, or 180 amino acids long. In various embodiments, the scaffold fragment of SEQ ID NO:214 has a C-terminal truncation, for example up to 30, up to 35, up to 40, up to 45, up to 50, up to 55 or up to 60 amino acids. In various embodiments, the scaffold sequence lacks the GG repeat cassette of HlyA1, i.e. amino acids 11-46, and the 54 C-terminal amino acids, i.e. positions 165-218, using the positional number of SEQ ID NO:214. This yields the scaffold sequence set forth in SEQ ID NO:215. In this scaffold sequence, which is devoid of GG repeats, the GG repeat element of choice is then inserted between amino acids 10 and 11, using the positional numbering of SEQ ID NO:215.

The invention further comprises, as flanking sequences, variants of all the afore-mentioned fragments of RTX proteins. These variants may have over their entire length, at least 80 %, at least 81 %, at least 82 %, at least 83 %, at least 84 %, at least 85 %, at least 86 %, at least 87 %, at least 88 %, at least 89 %, at least 90 %, at least 91 %, at least 92 %, at least 93 %, at least 94 %, at least 95%, at least 96 %, at least 97 %, at least 97.5%, at least 98 %, at least 98.5 %, at least 99% or 99.5% sequence identity with the respective amino acid sequence. In various embodiments, the variant is of the same length as the respective reference sequence. In other embodiments, it is a further shortened fragment thereof that may be obtainable by deletions/truncations. As noted, this variability is only given for the so-called flanking sequences that are N-terminal to the most N-terminal GG repeat and C-terminal to the most C-terminal GG repeat.

The first amino acid sequence of the present invention is, in various embodiments, 18 to 300 amino acids in length, for example 30 to 200. The lower limit may be 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 amino acids in length. As disclosed above, it may be preferred that the first amino acid sequence is as short as possible, as long as its beneficial influence on expression levels and/or yields is not impaired and it retains sufficiently high solubility and/or proteolytic stability. Generally, the skilled person will know how to find a balance between a sequence that is as short as

possible while still providing for the desired expression levels/yields within the scope defined by the present invention, in particular in view of the methods and techniques described in the examples herein below.

In various embodiments, if the first amino acid sequence has a scaffold derived from SEQ ID NO:214, the first amino acid sequence does not comprise any one of the full amino acid sequences of SEQ ID Nos. 46-50 as set forth in WO 2013/057312 A1. In various other embodiments, the first amino acid sequence does not comprise any one of the full amino acid sequences of SEQ ID Nos. 33, 34, 36 of EP 2 583 975 A1, as well as SEQ ID Nos. 4 and 7 of WO 2006/036406 A2 and EPOP:A12703 (200 aa long C-terminal fragment of HlyA) of WO 8706953 A1. In various embodiments, the isolated polypeptide is not that of SEQ ID NO: 28, 30, 32, 34 or 36 of EP 2792686 A1. In various embodiments, the first amino acid sequence does not comprise the first amino acid sequence as disclosed in PCT/EP2021/056948 or EP20163961 or EP20168779, such as any one of SEQ ID Nos. 1, 8-26, 44-63, and 65-71 of PCT/EP2021/056948. Further, in various embodiments, the first amino acid sequence is not that of databank accession number WP\_199357260. The first amino acid sequence is preferably also neither any of those used in WO 2021/185968 A1 nor is the polypeptide of the invention any one of those disclosed therein. In various embodiments, the first amino acid sequence does not comprise the GG repeat sequence element of the HlyA fragment with the sequence set forth in SEQ ID NO: 214, namely does not comprise amino acids 11-46 of said sequence using the positional numbering of SEQ ID NO:214. The isolated polypeptide is also not the one described in Example 4 of WO 2022/194403 A1, nor does the first amino acid sequence comprise the GG repeat cassette with the amino acid sequence of SEQ ID NO:235 of WO 2022/194403 A1. These known sequences are excluded as being the first amino acid sequence already by the definitions given herein, as they are fragments of the natural proteins, such as the natural HlyA protein and thus excluded by the definition that the first amino acid sequence is a "non-natural" sequence (that differs from any natural protein or fragment thereof by at least one amino acid substitution, preferably more).

The first amino acid sequence, as disclosed herein, may in various embodiments be of the general structure

"N-terminal flanking sequence" – "GGR repeat element" – "C-terminal flanking sequence"

with the GG repeat element being GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any one of (1) to (7) disclosed above. The first amino acid sequence may consist of such a structure.

In various embodiments, the first amino acid sequence does not comprise a secretion signal, i.e. an amino acid sequence that facilitates secretion from the host cell into the medium. This may mean that the first amino acid sequence does not facilitate secretion of the fusion protein to a significant degree, which may mean that at least 50 % or more of the expressed fusion protein remain inside of the cell, preferably as inclusion bodies, and are not secreted to the surrounding medium. This percentage may

be higher, such as 60%, 70%, 80%, 90% or more. In various embodiments, essentially all of the expressed polypeptide remains inside in the cell. In various embodiments, it is expressed as insoluble inclusion bodies, as further described herein below.

In various embodiments, the first amino acid sequence does also not comprise a protease recognition site, such as a TEV protease recognition site. While these may form part of the polypeptides of the invention, they are preferably not considered part of the first amino acid sequence.

The first amino acid sequence and second amino acid sequence as well as any third and further amino acid sequence are all individually continuous amino acid sequences, i.e. none of these sequences may be inserted into another, but all of them need to be linearly connected (in any possible orientation). In various embodiments, the polypeptide of the invention consists of the first and second amino acid sequences and any linking sequence between those two, as defined herein below.

The isolated polypeptide comprises a second amino acid sequence N-terminal or C-terminal to the first amino acid sequence, wherein the second amino acid sequence is at least one peptide or polypeptide of interest. The second amino acid sequence may be 5 to 1500 amino acids in length, preferably 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or 25 to 1000, to 800, to 700, to 600, to 500, to 400, to 300, to 200, to 180, to 150, to 120 or to 100 amino acids in length. The terms "second amino acid sequence" and the terms "peptide of interest" and "polypeptide/protein of interest" are used interchangeably herein. "Peptides of interest" are typically oligopeptides with a length of up to 100 amino acids, typically 3 to 80, 4 to 75, 5 to 70, 6 to 70, 7 to 70, 8 to 70, 9 to 70, or 10 to 70 amino acids. "Polypeptides/proteins of interest" typically comprise more than 100 amino acids, with the upper limit as defined above. The (poly)peptides of interest are preferably not the (poly)peptides from which either the GG repeats and/or the flanking sequences of the first amino acid sequence are derived. They are also heterologous to the first amino acid sequence, meaning that they do not naturally occur in the same natural polypeptide together.

In various embodiments, the second amino acid is a naturally occurring peptide or polypeptide sequence or fragment thereof, typically not containing any sequence deviation over its entire sequence relative to its natural template. This may mean that the second amino acid sequence shares 100% sequence identity over its entire length to a naturally occurring peptide or polypeptide or fragment thereof.

In various embodiments, the peptide or protein of interest may comprise two or more naturally occurring peptides or proteins, the two or more peptides or proteins may be separated by protease cleavage sites. This also includes embodiments, where the same peptide or protein is included multiple times in said second amino acid sequence. This then allows production of higher amounts of the respective peptide or protein of interest. In other embodiments, the peptide or protein of interest is only a single peptide or protein.



Generally, any peptide or protein may be chosen as protein of interest. In certain embodiments, the protein of interest is a protein, which does not form a homo-dimer or homo-multimer. The protein of interest may also be a peptide or protein which is a subunit of a larger peptide or protein complex. Such a peptide or protein may be isolated after expression and be suitable for an *in vitro* reconstitution of the multi peptide or protein complex. In certain embodiments, the protein or peptide of interest is a protein or peptide having less than 1000, less than 800, less than 700, less than 600, less than 500, less than 400, less than 300 amino acid residues, for example less than 200 amino acids or less than 150 amino acids. If these peptides comprise pre- and/or pro- sequences in their native state after translation the nucleic acid sequence encoding for the peptide of interest may be engineered to be limited to the sequence encoding the mature peptide. One exemplary peptide is insulin, e.g., human insulin. The expression of over-expressed peptides and proteins as inclusion bodies is especially advantageous where the peptide or protein is harmful to the host cell. For this reason, the present invention is particularly advantageous for expression of lipases and proteases which are known to be toxic to the host cell and thus the expression of these proteins by the inventive systems and methods represents a specific embodiment of the present invention.

In various embodiments, the peptide or protein of interest is an enzyme. The International Union of Biochemistry and Molecular Biology has developed a nomenclature for enzymes, the EC numbers; each enzyme is described by a sequence of four numbers preceded by "EC". The first number broadly classifies the enzyme based on its mechanism. The complete nomenclature can be browsed at <http://www.chem.qmul.ac.uk/iubmb/enzyme/>.

Accordingly, a peptide or protein of interest according to the present invention may be chosen from any of the classes EC 1 (Oxidoreductases), EC 2 (Transferases), EC 3 (Hydrolases), EC 4 (Lyases), EC 5 (Isomerases), and EC 6 (Ligases), and the subclasses thereof.

In certain embodiments, the peptide or protein of interest is cofactor dependent or harbors a prosthetic group. For expression of such peptides or proteins, in some embodiments, the corresponding cofactor or prosthetic group may be added to the culture medium during expression.

In certain cases, the peptide or protein of interest is a dehydrogenase or an oxidase. In case the peptide or protein of interest is a dehydrogenase, in some embodiments, the peptide or protein of interest is chosen from the group consisting of alcohol dehydrogenases, glutamate dehydrogenases, lactate dehydrogenases, cellobiose dehydrogenases, formate dehydrogenases, and aldehydes dehydrogenases. In case the peptide or protein of interest is an oxidase, in some embodiments, the peptide or protein of interest is chosen from the group consisting of cytochrome P450 oxidoreductases, in particular P450 BM3 and mutants thereof, peroxidases, monooxygenases, hydrogenases, monoamine oxidases, aldehydes oxidases, xanthin oxidases, amino acid oxidases, and NADH oxidases.

In further embodiments, the peptide or protein of interest is a transaminase or a kinase. In case the peptide or protein of interest is a transaminase, in some embodiments, the peptide or protein of interest is chosen from the group consisting of alanine aminotransferases, aspartate aminotransferases, glutamate-oxaloacetic transaminases, histidinol-phosphate transaminases, and histidinol-pyruvate transaminases. In various embodiments, if the peptide or protein of interest is a kinase, the peptide or protein of interest is chosen from the group consisting of nucleoside diphosphate kinases, nucleoside monophosphate kinases, pyruvate kinase, and glucokinases. In some embodiments, if the peptide or protein of interest is a hydrolase, the peptide or protein of interest is chosen from the group consisting of lipases, amylases, proteases, cellulases, nitrile hydrolases, halogenases, phospholipases, and esterases.

In certain embodiments, if the peptide or protein of interest is a lyase, the peptide or protein of interest is chosen from the group consisting of aldolases, e.g., hydroxynitrile lyases, thiamine-dependent enzymes, e.g., benzaldehyde lyases, and pyruvate decarboxylases. In various embodiments, if the peptide or protein of interest is an isomerase, the peptide or protein of interest is chosen from the group consisting of isomerases and mutases.

In some embodiments, if the peptide or protein of interest is a ligase, the peptide or protein of interest may be a DNA ligase.

In certain embodiments, the peptide or protein of interest may be an antibody. This may include a complete immunoglobulin or fragment thereof, which immunoglobulins include the various classes and isotypes, such as IgA, IgD, IgE, IgG1, IgG2a, IgG2b and IgG3, IgM, etc. Fragments thereof may include Fab, Fv and F(ab')<sub>2</sub>, Fab', the variable domain of the light chain (VL) or the variable domain of the heavy chain (VH) and related fragments, such as nanobodies, and the like.

Also contemplated herein are therapeutically active peptides and proteins of interest, e.g., cytokines.

Thus, in certain embodiments the peptide or protein of interest is selected from the group consisting of cytokines, in particular human or murine interferons, interleukins (IL-1, IL-2, IL-3, IL-4; IL-5; IL-6; IL-7; IL-8; IL-9; IL-10; IL-11; IL-12; IL-13; IL-14; IL-15; IL-16; and IL-17), colony-stimulating factors, necrosis factors, e.g., tumor necrosis factor, such as TNF alpha, and growth factors, such as transforming growth factor beta family members, such as TGF-beta1; TGF-beta2 and TGF-beta3.

In some embodiments, if the peptide or protein of interest is an interferon, the peptide or protein of interest may be selected from the group consisting of interferon alpha, e.g., alpha-1, alpha-2, alpha-2a, and alpha-2b, alpha-2, alpha-16, alpha 21, beta, e.g., beta-1, beta-1a, and beta-1b, or gamma.

In further embodiments, the peptide or protein of interest is an antimicrobial peptide, in particular a peptide selected from the group consisting of bacteriocines and lantibiotics, e.g., nisin, cathelicidins, defensins, and saposins.

Also disclosed herein are peptides or proteins of interest which are therapeutically active peptides or proteins. In certain embodiments, the peptide or protein of interest is a therapeutically active peptide. In some embodiments, a therapeutically active peptide may be selected from the group consisting of Fuzeon/T20, human calcitonin, salmon calcitonin, human corticotropin release factor, Mab40, Mab42, peptides associated with Alzheimer's disease, exenatide, Tesamorelin, Teriparatide, BMP-2, Corticorelin ovine triflutate, Linaclotide, Nesiritide, Lucinactant, Bivalirudin, Lepirudin, Thymalfasin, Glatiramer, Glucagon, Aviptadil, Secretin, Thymosin-b4, Teduglutide, GLP-1, GLP-2 and analoga, Plecanatide, Ecallantide, Anakinra, Disiteritide, Lixisenatide, Liraglutide, Semaglutide, Abaloparatide, Goserelin, Lanreotide, Carfilzomib, Enfuvirtide, T-20, Terlipressin, Elcatonin, Afamelanotide, Oxodotretotide, Caspofungin, Colistin, Polymyxin E, Cyclosporine, Dactinomycin, Lyovac-Cosmegen, Degarelix, Vancomycin, Secretin, Ziconotide, Gonadorelin, Somastatin, Sincalide, Eptifibatid, Vapreotide, Triptorelin, Desmopressin, Lypressin, Atosiban, Pramintide, Pasireotide, Sandostatin, and Icatibant. The afore-mentioned peptides may be of mammalian or human origin. Also encompassed are analogues of the afore-mentioned peptides that originate from other species, for example homologues from other animals, microorganisms, virus and others.

In certain embodiments, the peptide or protein of interest is a type I secretion substrate. More than 1000 proteins are annotated or have been described as type I secretion substrates in the literature. Many of them have interesting characteristics for the biotechnological usage, in particular proteases and lipases. Suitable proteases and lipases have been described by Baumann et al. (1993) EMBO J 12, 3357-3364; and Meier et al. (2007) J. BIOL. CHEM.: 282(43), pp. 31477-31483. The content of each of these documents is incorporated by reference herein in its entirety.

In certain embodiments, the second amino acid sequence is a peptide or protein of interest which is chosen from the group consisting of MBP, lipase CalB, protease SprP, hydrolase PlaB, hydrolase PlaK, hydrolase PlbF, lipase TesA, Vif, human interferon alpha- 1, alpha-2, alpha-8, alpha- 16, alpha-21, human interferon beta, human interferon gamma, murine interferon alpha, murine interferon gamma, IFABP, Cas2, affibody protein ZA3, nisin, corticotropin release factor, amyloid- beta peptide, exenatide, Fuzeon/T20, salmon calcitonin, Mab40, Mab42, lipase LipA, SprP, the HIV- 1 protein Vif, human calcitonin, Tesamorelin, Teriparatide, BMP-2, Corticorelin ovine triflutate, Linaclotide, Nesiritide, Lucinactant, Bivalirudin, Lepirudin, Thymalfasin, Glatiramer, Glucagon, Aviptadil, Secretin, Thymosin-b4, Teduglutide, GLP-1, GLP-2 and analoga, Plecanatide, Ecallantide, Anakinra, Disiteritide, Lixisenatide, Liraglutide, Semaglutide, Abaloparatide, Goserelin, Lanreotide, Carfilzomib, Enfuvirtide, T-20, Terlipressin, Elcatonin, Afamelanotide, Oxodotretotide, Caspofungin, Colistin, Polymyxin E, Cyclosporine, Dactinomycin, Lyovac-Cosmegen, Degarelix, Vancomycin, Secretin, Ziconotide,

Gonadorelin, Somastatin, Sincalide, Eptifibatid, Vapreotide, Triptorelin, Desmopressin, Lypressin, Atosiban, Pramintide, Pasireotide, Sandostatin, and Icatibant.

The second amino acid sequence may be directly or via a linker sequence linked to the first amino acid sequence. It may be located N- or C-terminally relative to the first amino acid sequence, typically, for example but without limitation, C-terminally. This means that the N-terminus of the second amino acid sequence is linked, optionally via a linker sequence, to the C-terminus of the first amino acid sequence. However, in other embodiments, the C-terminus of the second amino acid sequence may be fused, optionally via a linker sequence, to the N-terminus of the first amino acid sequence.

In various embodiments, the linker sequence that connects the first and second amino acid sequences is also a peptide sequence and connected to the respective ends of the first and second amino acid sequence via a peptide bond. In various embodiments, the linker sequence may be 1 to 50 amino acids in length, for example 1 to 30 amino acids or 5 to 20 amino acids. As disclosed above, the first amino acid sequence together with said linker sequence may have a maximum length of 250 amino acids. In various embodiments, the linker sequence is also heterologous to the first and the second amino acid sequence and/or does not naturally occur as part of the first and/or second amino acid sequence.

The linker sequence may be functional in that it may provide for easy cleavage and separation of the first and second amino acid. To facilitate such a purpose, it can comprise or consist of a protease recognition and cleavage site. The linker may also comprise both, a linker sequence that serves only as a link and a protease cleavage site. The linker sequence may be a G-rich sequence, for example 4 or 5 consecutive G residues, optionally followed by an S residue. Such a linker does not comprise any GG repeat sequence. The G-rich sequence may be the one set forth in SEQ ID NO:216.

The term "protease (recognition and) cleavage site" refers to a peptide sequence which can be cleaved by a selected protease thus allowing the separation of peptide or protein sequences which are interconnected by a protease cleavage site. In certain embodiments the protease cleavage site is selected from the group consisting of a Factor Xa, a tobacco etch virus (TEV) protease, a enterokinase, a SUMO Express protease, an Arg-C proteinase, an Asp-N endopeptidases, an Asp-N endopeptidase + N-terminal Glu, a caspase 1, a caspase 2, a caspase 3, a caspase 4, a caspase 5, a caspase 6, a caspase 7, a caspase 8, a caspase 9, a caspase 10, a chymotrypsin-high specificity, a chymotrypsin-low specificity, a clostripain (Clostridiopeptidase B), a glutamyl endopeptidase, a granzyme B, a pepsin, a proline-endopeptidase, a proteinase K, Welcut protease, Clean Cut protease, a staphylococcal peptidase I, a Thrombin, a Trypsin, inteins, SprB or SplA-E from *Staphylococcus aureus*, and a Thermolysin cleavage site. In various embodiments, it may be a Factor Xa, SprB or TEV protease cleavage site. It can be preferred, in some embodiments, to design the protease recognition site such that as few amino acids as possible of the recognition and cleavage site remain attached to the peptide or protein of interest. In various embodiments, a protease recognition and cleavage site is included, the site being for example a TEV protease recognition and cleavage site. The TEV protease cleavage site

typically comprises the amino acid sequence ENLYFQX (SEQ ID NO:217), with X being G or S and cleaves between the Gln (Q) and X residues. In various embodiments, the P1' amino acids (X) may be A, M or C instead of G or S.

In various embodiments, the isolated polypeptide may further comprise at least one third amino acid sequence, for example an affinity tag.

The term "affinity tag" as used herein relates to entities which are coupled to a molecule of interest and allow enrichment of the complex between the molecule of interest and the affinity tag using an affinity tag receptor. In certain embodiments affinity tags may be selected from the group consisting of the Strep-tag® or Strep-tag® II, the myc-tag, the FLAG-tag, the His-tag, the small ubiquitin-like modifier (SUMO) tag, the covalent yet dissociable NorpD peptide (CYD) tag, the heavy chain of protein C (HPC) tag, the calmodulin binding peptide (CBP) tag, or the HA-tag or proteins such as Streptavidin binding protein (SBP), maltose binding protein (MBP), and glutathione-S-transferase.

In various embodiments, the isolated polypeptide has relative to a polypeptide comprising a first amino acid sequence not covered by the present definition, for example a first amino acid sequence comprising the scaffold of SEQ ID NO:215 and the GG repeat cassette of SEQ ID NO:218 inserted between positions 10 and 11 of SEQ ID NO:215, a comparable or increased renaturation efficiency. "Increased", as used in this connection, refers to yields of renaturated polypeptides that are at least 10%, preferably at least 20% higher than those achieved with the reference sequence, in particular that mentioned above. Renaturation efficiency may be determined by comparing the levels of (i) renaturation of isolated polypeptides comprising as the first amino acid sequence a reference sequence, for example the scaffold of SEQ ID NO:215 and the GG repeat cassette of SEQ ID NO:218 inserted between positions 10 and 11 of SEQ ID NO:215, and (ii) non-hydrolyzed isolated polypeptides comprising as the first amino acid sequence an amino acid sequence as described herein, after expression in form of inclusion bodies, purification and renaturation under otherwise identical conditions.

In various embodiments, the first amino acid sequence comprises as the first, N-terminal amino acid the residue M. If this is not present within the specific sequences disclosed herein, it may be artificially added, if desired, in particular to facilitate expression in a host cell.

In various embodiments, the polypeptides of the present invention consist of the elements

- (1) First amino acid sequence, as defined herein;
- (2) Second amino acid sequence, as defined herein;
- (3) Optional linker between first and second amino acid sequence, as defined herein;
- (4) Optional third amino acid sequence, as defined herein, with optional short linker sequences (1-20 amino acids) between the third amino acid sequence and any adjacent element(s).

The polypeptides of the invention may thus comprise or consist of the structure

“First amino acid sequence”-linker-“Second amino acid sequence”

or

“Second amino acid sequence”-linker-“First amino acid sequence”.

Said linker sequence is defined as described above for the linker between the first and second amino acid sequence. Its definition is thus different from that of the “Linker” between the individual GG repeats.

If a third amino acid sequence is present this may be attached to the N- or C-terminus of any one of these structures.

The invention further relates to the nucleic acid, in particular the isolated nucleic acid molecule, encoding the polypeptide as described above. The polypeptide comprises the first amino acid sequence and the second amino acid sequence and optionally at least one third amino acid sequence. All of these amino acid sequences are typically linked by peptide bonds and expressed as a single fusion protein. To facilitate said expression, the nucleic acid molecule comprises a first nucleotide sequence encoding the first amino acid sequence and optionally a second, third and further nucleotide sequence encoding the second, third and further amino acid sequence, with said nucleotide sequences being operably linked to allow expression of the single fusion protein comprising all afore-mentioned amino acid sequences.

The term “operably linked” in the context of nucleic acid sequences means that a first nucleic acid sequence is linked to a second nucleic acid sequence such that the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter sequence is operably linked to a coding sequence of a heterologous gene if the promoter can initiate the transcription of the coding sequence. In a further context, a sequence encoding for the first amino acid sequence is linked such to a second amino acid sequence encoding for a peptide or protein of interest, that if the two sequences are translated a single peptide/protein chain is obtained.

In certain embodiments, the above defined nucleic acid molecules may be comprised in a vector, for example a cloning or expression vector. Generally, the nucleic acid molecules of the invention can also be part of a vector or any other kind of cloning vehicle, including, but not limited to a plasmid, a phagemid, a phage, a baculovirus, a cosmid, or an artificial chromosome. Generally, a nucleic acid molecule disclosed in this application may be “operably linked” to a regulatory sequence (or regulatory sequences) to allow expression of this nucleic acid molecule.

Such cloning vehicles can include, besides the regulatory sequences described above and a nucleic acid sequence of the present invention, replication and control sequences derived from a species compatible with the host cell that is used for expression as well as selection markers conferring a selectable phenotype on transformed or transfected cells. Large numbers of suitable cloning vectors are known in the art, and are commercially available.

In certain embodiments the nucleic acid molecules disclosed herein are comprised in a cloning vector. In some embodiments the nucleic acid molecules disclosed herein are comprised in an expression vector. The vectors may comprise regulatory elements for replication and selection markers. In certain embodiments, the selection marker may be selected from the group consisting of genes conferring ampicillin, kanamycin, chloramphenicol, tetracycline, blasticidin, spectinomycin, gentamicin, hygromycin, and zeocin resistance. In various other embodiments, the selection may be carried out using antibiotic-free systems, for example by using toxin/antitoxin systems, cer sequence, triclosan, auxotrophies or the like. Suitable methods are known to those skilled in the art.

The above-described nucleic acid molecule of the present invention, comprising a nucleic acid sequence encoding for a protein of interest, if integrated in a vector, must be integrated such that the peptide or protein of interest can be expressed. Therefore, a vector of the present invention comprises sequence elements which contain information regarding to transcriptional and/or translational regulation, and such sequences are "operably linked" to the nucleotide sequence encoding the polypeptide. An operable linkage in this context is a linkage in which the regulatory sequence elements and the sequence to be expressed are connected in a way that enables gene expression. The precise nature of the regulatory regions necessary for gene expression may vary among species, but in general these regions comprise a promoter which, in prokaryotes, contains both the promoter per se, i.e. DNA elements directing the initiation of transcription, as well as DNA elements which, when transcribed into RNA, will signal the initiation of translation. Such promoter regions normally include 5' non-coding sequences involved in initiation of transcription and translation, such as the -35/- 10 boxes and the Shine-Dalgarno element in prokaryotes or the TATA box, CAAT sequences, and 5'-capping elements in eukaryotes. These regions can also include enhancer or repressor elements as well as translated signal and leader sequences for targeting the native polypeptide to a specific compartment of a host cell.

In addition, the 3' non-coding sequences may contain regulatory elements involved in transcriptional termination, polyadenylation or the like. If, however, these termination sequences are not satisfactory functional in a particular host cell, then they may be substituted with signals functional in that cell.

In various embodiments, a vector comprising a nucleic acid molecule of the invention can therefore comprise a regulatory sequence, preferably a promoter sequence. In certain embodiments, the promoter is identical or homologous to promoter sequences of the host genome. In such cases endogenous polymerases may be capable to transcribe the nucleic acid molecule sequence comprised in the vector. In various embodiments, the promoter is selected from the group of weak, intermediate and strong promoters, preferably from weak to intermediate promoters.

In another preferred embodiment, a vector comprising a nucleic acid molecule of the present invention comprises a promoter sequence and a transcriptional termination sequence. Suitable promoters for prokaryotic expression are, for example, the araBAD promoter, the tet-promoter, the lacUV5 promoter,

the CMV promoter, the EF1 alpha promoter, the AOX1 promoter, the tac promoter, the T7 promoter, or the lac promoter. Examples of promoters useful for expression in eukaryotic cells are the SV40 promoter or the CMV promoter. Furthermore, a nucleic acid molecule of the invention can comprise transcriptional regulatory elements, e.g., repressor elements, which allow regulated transcription and translation of coding sequences comprised in the nucleic acid molecule. Repressor element may be selected from the group consisting of the Lac-, AraC-, or MalR-repressor.

The vector may be effective for prokaryotic or eukaryotic protein expression. In particular, the nucleic acid molecules of the present invention may be comprised in a vector for prokaryotic protein expression. Such vector sequences are constructed such that a sequence of interest can easily be inserted using techniques well known to those skilled in the art. In certain embodiments, the vector is selected from the group consisting of a pET-vector, a pBAD-vector, a pK184-vector, a pMONO-vector, a pSELECT-vector, pSELECT-Tag-vector, a pVITRO-vector, a pVIVO-vector, a pORF-vector, a pBLAST-vector, a pUO-vector, a pDUO-vector, a pZERO-vector, a pDeNy-vector, a pDRIVE-vector, a pDRIVE-SEAP-vector, a HaloTag@Fusion-vector, a pTARGET<sup>TM</sup>-vector, a Flexi@-vector, a pDEST-vector, a pHIL-vector, a pPIC-vector, a pMET-vector, a pPink-vector, a pLP -vector, a pTOPO-vector, a pBud-vector, a pCEP-vector, a pCMV-vector, a pDisplay-vector, a pEF-vector, a pFL-vector, a pFRT-vector, a pFastBac-vector, a pGAPZ-vector, a pIZ/V5-vector, a p3S-vector, a pIAR-vector, pSEC, pMS, a pSU2726-vector, a pLenti6 -vector, a pMIB-vector, a pOG-vector, a pOpti-vector, a pREP4-vector, a pRSET-vector, a pSCREEN- vector, a pSecTag-vector, a pTEFI -vector, a pTracer-vector, a pTrc-vector, a pUB6-vector, a pVAXI-vector, a pYC2-vector, a pYES2-vector, a pZeo-vector, a pcDNA-vector, a pFLAG-vector, a pTAC-vector, a pT7-vector, a gateway@-vector, a pQE-vector, a pLEXY-vector, a pRNA-vector, a pPK-vector, a pUMVC-vector, a pLIVE-vector, a pCRUZ-vector, a Duet-vector, and other vectors or derivatives thereof.

The vectors of the present invention may be chosen from the group consisting of high, medium and low copy vectors.

The above described vectors of the present invention may be used for the transformation or transfection of a host cell in order to achieve expression of a peptide or protein which is encoded by an above described nucleic acid molecule and comprised in the vector DNA. Thus, in a further aspect, the present invention also relates to a host cell comprising a vector or nucleic acid molecule as disclosed herein.

Also contemplated herein are host cells, which comprise a nucleic acid molecule as described herein integrated into their genomes. The skilled person is aware of suitable methods for achieving the nucleic acid molecule integration. For example, the molecule may be delivered into the host cells by means of liposome transfer or viral infection and afterwards the nucleic acid molecule may be integrated into the host genome by means of homologous recombination. In certain embodiments, the nucleic acid molecule is integrated at a site in the host genome, which mediates transcription of the peptide or protein of the invention encoded by the nucleic acid molecule. In various embodiments, the nucleic acid



molecule further comprises elements which mediate transcription of the nucleic acid molecule once the molecule is integrated into the host genome and/or which serve as selection markers.

In certain embodiments, the nucleic acid molecule of the present invention is transcribed by a polymerase natively encoded in the host genome. In various embodiments, the nucleic acid molecule is transcribed by a RNA-polymerase which is non-native to the host genome. In such embodiments, the nucleic acid molecule of the present invention may further comprise a sequence encoding for a polymerase and/or the host genome may be engineered or the host cell may be infected to comprise a nucleic acid sequence encoding for an exogenous polymerase. The host cell may be specifically chosen as a host cell capable of expressing the gene. In addition or otherwise, in order to produce the isolated polypeptide of the invention, the nucleic acid coding for it can be genetically engineered for expression in a suitable system. Transformation can be performed using standard techniques (Sambrook, J. et al. (2001), *Molecular Cloning: A Laboratory Manual*, 3<sup>rd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Prokaryotic or eukaryotic host organisms comprising such a vector for recombinant expression of the polypeptide as described herein form also part of the present invention. Suitable host cells can be prokaryotic cells. In certain embodiments the host cells are selected from the group consisting of gram positive and gram negative bacteria. In some embodiments, the host cell is a gram negative bacterium, such as *E.coli*. In certain embodiments, the host cell is *E. coli*, in particular *E. coli* BL21 (DE3) or other *E. coli* K12 or *E. coli* B834 derivatives. In further embodiments, the host cell is selected from the group consisting of *Escherichia coli* (*E. coli*), *Pseudomonas*, *Serratia marcescens*, *Salmonella*, *Shigella* (and other enterobacteriaceae), *Neisseria*, *Haemophilus*, *Klebsiella*, *Proteus*, *Enterobacter*, *Helicobacter*, *Acinetobacter*, *Moraxella*, *Helicobacter*, *Stenotrophomonas*, *Bdellovibrio*, *Legionella*, acetic acid bacteria, *Bacillus*, *Bacilli*, *Carynebacterium*, *Clostridium*, *Listeria*, *Streptococcus*, *Staphylococcus*, and Archaea cells. Suitable eukaryotic host cells are among others CHO cells, insect cells, fungi, yeast cells, e.g., *Saccharomyces cerevisiae*, *S. pombe*, *Pichia pastoris*.

In certain embodiments, the host cell is a prokaryotic cell, such as *E.coli*, in particular *E.coli* BL21 (DE3), *E. coli* BL21, *E. coli* K12, *E. coli* BLR, *E. coli* BL21 AI, *E. coli* BL21 pLysS, *E. coli* XL1 and *E. coli* DH5a. Further suitable *E.coli* strains include, but are not limited to DH1, DH5a, DM1, HB101, JmlOI-110, Rosetta(DE3)pLysS, SURE, TOP10, XLI-Blue, XL2-Blue and XLIO-Blue strains.

The transformed host cells are cultured under conditions suitable for expression of the nucleotide sequence encoding the polypeptide of the invention. In certain embodiments, the cells are cultured under conditions suitable for expression of the nucleotide sequence encoding a polypeptide of the invention.

For producing the recombinant peptide or protein of interest in form of the fusion proteins described herein, a vector of the invention can be introduced into a suitable prokaryotic or eukaryotic host organism

by means of recombinant DNA technology (as already outlined above). For this purpose, the host cell is first transformed with a vector comprising a nucleic acid molecule according to the present invention using established standard methods (Sambrook, J. et al. (2001), supra). The host cell is then cultured under conditions, which allow expression of the heterologous DNA and thus the synthesis of the corresponding polypeptide. Subsequently, the polypeptide is recovered ("isolated") either from the cell or from the cultivation medium.

For expression of the peptides and proteins of the present invention several suitable protocols are known to the skilled person. The expression of a recombinant polypeptide of the present invention may be achieved by the following method comprising: (a) introducing a nucleic acid molecule or vector of the invention into a host cell, wherein the nucleic acid molecule or vector encodes the recombinant polypeptide; and (b) cultivating the host cell in a culture medium under conditions that allow expression of the recombinant polypeptide.

Step (a) may be carried out by using suitable transformation and transfection techniques known to those skilled in the art. These techniques are usually selected based on the type of host cell into which the nucleic acid is to be introduced. In some embodiments, the transformation may be achieved using electroporation or heat shock treatment of the host cell.

Step (b) may include a cultivation step that allows growth of the host cells. Alternatively, such step allowing growth of the host cells and a step that allows expression of the polypeptide may be performed separately in that the cells are first cultivated such that they grow to a desired density and then they are cultivated under conditions that allow expression of the polypeptide. The expression step can however still allow growth of the cells.

The method may further include a step of recovering the expressed polypeptide. The polypeptide may be recovered from the growth medium, if it is secreted, or from the cells or both. Preferably the polypeptide is recovered from the cells. The recovery of the polypeptide may include various purification steps.

Generally, any known culture medium suitable for growth of the selected host may be employed in this method. In various embodiments, the medium is a rich medium or a minimal medium. Also contemplated herein is a method, wherein the steps of growing the cells and expressing the peptide or protein comprise the use of different media. For example, the growth step may be performed using a rich medium which is replaced by a minimal medium in the expression step. In certain cases, the medium is selected from the group consisting of LB medium, TB medium, 2YT medium, synthetic medium and minimal medium.

In various preferred embodiments of the invention, the polypeptide of the invention is not secreted. In such embodiments, it may preferably be expressed in form of inclusion bodies (IBs). In many cases, it

may be useful to express the polypeptide of the invention in such an insoluble form, for example in cases where the peptide of interest is rather short, normally soluble and/or subject to proteolytic degradation within the host cell. Production of the peptide in insoluble form both facilitates simple recovery and additionally protects the peptide from the undesirable proteolytic degradation. In such embodiments, the first amino acid sequence may serve as a solubility tag, i.e., an inclusion body (IB) tag, that induces IB formation. Calcium ion (or earth alkaline metal ion) binding to the GG repeat(s) or any other change of environment may later catalyze the GG-repeat-facilitated folding of the fusion polypeptide into the native, active conformation, optionally with the calcium/earth alkaline metal ions acting as a folding helper/chaperone. Renaturation/refolding can, in certain instances, surprisingly also occur in the absence of calcium and/or earth alkaline metal ions, but is still critically dependent on the presence of the GG repeat elements disclosed herein.

It has been described above, that the inventor has found a variety of first amino acid sequences that increase renaturation efficiency in such methods compared to the respective full-length proteins they are derived from or natural fragments thereof. It has further surprisingly found that even comparably short first amino acid sequences can successfully induce proper renaturation of a comparably long second amino acid sequence, with said second amino acid sequence not comprising any such refolding sequence motifs, such as GG repeats.

The terms "inclusion body" or "IB", as interchangeably used herein, relate to nuclear or cytoplasmic aggregates of substances, for instance proteins. In the method of the present invention, the IBs mainly consist of the fusion protein comprising at least one peptide/protein of interest and the first amino acid sequence, as defined herein.

In various embodiments, in particular where expression of the polypeptide of the invention in form of IBs is desired, the expression of the endogenous ABC transporter gene, the endogenous MFP gene and/or the endogenous OMP gene of the T1 SS or the activity of the corresponding gene products in the host cell is inhibited. In various embodiments, the host cell does not express endogenous ABC transporter, endogenous MFP and/or endogenous OMP of the T1 SS. The host cell may be engineered accordingly.

Methods to inhibit the expression of genes such as their deletion or insertion of nucleotide sequences destroying the integrity of the promoter sequence or the gene itself are known in the art. A preferred gene expression activity after deletion or disruption may be less than 35 %, 30 %, 25 %, 20 %, 15 %, 10 % or 5 % of the activity measured in untreated cells. In other various embodiments of the invention, the endogenous ABC transporter, the endogenous MFP and/or the endogenous OMP of the type 1 secretion system are inhibited by antibodies or small molecule inhibitors. In preferred embodiments of the invention, the ABC transporter activity is inhibited by orthovanadate or an ATP homologous inhibitor such as 8-azido-ATP. Such ATP mimetics are known in the art. The preferred protein activity after inhibitor treatment may be less than 35 %, 30 %, 25 %, 20 %, 15 %, 10 % or 5 % of the activity measured

in untreated cells. In other embodiments of the invention, the transport is inhibited or blocked by the polypeptide of the invention itself, for example by over-expressing it.

In various embodiments of the afore-mentioned methods, the polypeptide of the invention is recovered from the host cells in form of insoluble inclusion bodies, for example by any of the purification methods disclosed herein.

Renaturation may then be effected by exposing the isolated/purified polypeptide to renaturation conditions. In some instances, this may include buffer conditions wherein the so-called refolding buffer comprises at least 0.01, more preferably 0.01-40 mM of alkaline earth metal ions, in particular  $\text{Ca}^{2+}$ . By virtue of the presence of GG repeats in the first amino acid sequence, such treatment triggers refolding of the polypeptide into its functional (native) conformation. Interestingly, the refolding of the first amino acid sequence also induces proper folding of the peptide/polypeptide of interest fused thereto. It has surprisingly been found by the inventors that such refolding can, in various embodiments, also occur in the absence of calcium or other earth alkaline metal ions. These are therefore not mandatory components of the refolding buffer. This may be particularly advantageous for peptides and proteins that are susceptible to adverse effects by the presence of calcium or other earth alkaline metal ions. In various embodiments, the renaturation/refolding thus occurs in the absence of calcium and/or earth alkaline metal ions in general. It may be preferred that the refolding occurs under reducing conditions, for example by using a reduction agent, such as TCEP ((tris(2-carboxyethyl)phosphine)).

The general advantages of expression of a fusion protein in form of inclusion bodies are set forth in greater detail in WO 2014/170430 A1, which is herewith included by reference in its entirety.

In various embodiments, the method also encompasses the purification the recombinant polypeptide, wherein the recombinant polypeptide is purified using a method selected from affinity chromatography, ion exchange chromatography, reverse phase chromatography, size exclusion chromatography, and combinations thereof.

In several embodiments, the method may comprise the treatment of the recombinant polypeptide with a protease suitable for cleavage of a protease cleavage site within the recombinant polypeptide. In preferred embodiments, the recombinant polypeptide is purified and, optionally, renatured prior to proteolytic cleavage using one or more methods disclosed above. Also after cleavage of the recombinant peptide or protein, the method may comprise a further purification step as defined above. Thus, in some embodiments the recombinant polypeptide is purified, subjected to proteolytic cleavage and the peptide or protein of interest is further purified. In other embodiments, the protease may be co-expressed or added to the cultivation medium or expressed by co-cultivated microorganisms, such that cleavage occurs before purification. However, since renaturation of the peptide or polypeptide of interest is mediated by the first amino acid sequence, such strategy may not be preferred.

In a further aspect, the present invention relates to the use of a vector or nucleic acid molecule as disclosed herein for the expression of a recombinant polypeptide.

A method for expression of a recombinant peptide or protein using the above-described nucleic acid molecules may comprise the steps of:

- (a) introducing a nucleic acid molecule or a vector as described above into a suitable host cell, wherein the nucleic acid molecule or vector encodes the recombinant polypeptide; and
- (b) cultivating the host cell in a culture medium under conditions that allow expression of the recombinant polypeptide.

The expression of the recombinant polypeptide in step (b) may be in form of inclusion bodies.

The method can further be defined as the other methods of the invention described above. Specifically, the method may further comprise recovering the expressed peptide or protein from the host cell and/or the culture medium. In addition, the host cell may be a prokaryotic cell; and/or the expression may be performed in minimal culture medium; and/or the recombinant polypeptide may be purified using a method selected from affinity chromatography, ion exchange chromatography, reverse phase chromatography, size exclusion chromatography, and combinations thereof; and/or the method may comprise treatment of the recombinant polypeptide with a protease suitable for cleavage of a protease cleavage site within the recombinant polypeptide; and/or the method may comprise a cleavage step followed by purification of the recombinant polypeptide.

In various embodiments, the polypeptide is recovered in form of inclusion bodies and, after purification, exposed to a refolding buffer. The buffer may comprise at least 0.01, more preferably 0.01-40 mM of earth alkaline metal ions, such as  $\text{Ca}^{2+}$ . In other instances, the refolding buffer is free of calcium and other earth alkaline metal ions. In various embodiments, the methods of the invention thus may also comprise a step of re-solubilizing the peptide/protein and/or reconstituting/refolding it under suitable conditions that may be conditions with or without calcium. Said step is also referred to herein as "renaturing" step. "Without calcium", as used in this context, means that the concentration of calcium is so low that it has virtually no impact on the folding. Such concentration may be less than 0.01 mM, less than 0.001 mM or less than 0.0001 mM.

In still another aspect, the present invention is therefore also directed to methods for renaturing the isolated polypeptide of the invention, wherein the method may comprise contacting, for example in form of re-solubilizing, the isolated polypeptide with a suitable medium, typically an aqueous medium, optionally comprising earth alkaline metal ions, for example the refolding buffer disclosed above. Such a medium allows the GG repeats (comprised in the first amino acid sequence) to induce conformational changes that in turn then facilitate refolding/renaturation of the fused heterologous peptide or polypeptide (second amino acid sequence).

In still another aspect, the present invention also relates to the use of an isolated polypeptide of the invention for facilitating the production of a recombinant peptide or protein. Said recombinant peptide or protein is part of the isolated polypeptide (in form of the second amino acid sequence), which is expressed as a fusion protein. In such embodiments, the first amino acid sequences disclosed herein may be used to facilitate renaturation/refolding of said recombinant peptide or polypeptide. Said recombinant peptide or protein is identical with the peptide or polypeptide of interest as referred to herein and may thus consist of the second amino acid sequence disclosed herein. As stated above, said use is preferably directed to recombinant polypeptides expressed as fusion proteins in a suitable expression system.

## EXAMPLES

**Table 1:** Overview of SEQ ID Nos. and corresponding amino acid sequences. The amino acid sequence matching the consensus of GGRC as defined herein are given in **bold**.

SEQ ID NO:	Amino acid sequence
1	<b>GGAGNDTLV</b>
2	<b>GGAGNDTLI</b>
3	<b>GGAGNDFV</b>
4	<b>GGAGNDFI</b>
5	<b>GGAGNDSL</b>
6	<b>GGAGNDSLI</b>
7	<b>GGAGNDSFV</b>
8	<b>GGAGNDSFI</b>
9	<b>GGAGADTLV</b>
10	<b>GGAGADTLI</b>
11	<b>GGAGADTFV</b>
12	<b>GGAGADTFI</b>
13	<b>GGAGADSLV</b>
14	<b>GGAGADSLI</b>
15	<b>GGAGADSFV</b>
16	<b>GGAGADSFV</b>
17	<b>GGAGADSLV</b>
18	<b>GGAGADSLI</b>
19	<b>GGAGADSFV</b>
20	<b>GGAGADSLV</b>
21	<b>GGAGADSLI</b>
22	<b>GGAGADSFV</b>
23	<b>GGAGADSLV</b>
24	<b>GGAGADSLI</b>
25	<b>GGAGADSFV</b>
26	<b>GGAGADSLV</b>
27	<b>GGAGADSLI</b>
28	<b>GGAGADSFV</b>
29	<b>GGAGADSLV</b>
30	<b>GGAGADSLI</b>
31	<b>GGAGADSFV</b>
32	<b>GGAGADSLV</b>
33	<b>GGAGADSLI</b>
34	<b>GGAGADSFV</b>
35	<b>GGAGADSLV</b>
36	<b>GGAGADSLI</b>
37	<b>GGAGADSLV</b>

38	<b>GGDGADSLI</b>
39	<b>GGDGADSFV</b>
40	<b>GGDGADSF</b>
41	<b>GGDGDDTLV</b>
42	<b>GGDGDDTLI</b>
43	<b>GGDGDDTFV</b>
44	<b>GGDGDDTFI</b>
45	<b>GGDGDDSLV</b>
46	<b>GGDGDDSLI</b>
47	<b>GGDGDDSFV</b>
48	<b>GGDGDDSF</b>
49	GGAGADRLD
50	GGAGNDRID
51	GGAGADRLN
52	GGSGSQMN
53	GGSGVDVVE
54	GGLGNDILT
55	GGAGFDTLV
56	GGAGNDVLN
57	GKGNLDF
58	GKGNLTLN
59	GGFGDRLN
60	GGSGQDLLR
61	GGIGYDDIY
62	GGSGDDRDL
63	GGAGDDRDL
64	GGGKDSL
65	GGAGNDLLD
66	GGAGNDTLD
67	GGQQLI
68	GGDGDDTLT
69	GGSGVDQLF
70	GGDGGDRLE
71	GGDGNDTVA
72	GGAGKDTLA
73	GGAGSDRLL
74	GGQQLLE
75	GGTGDLLL
76	GGLGDELS
77	GGNGDDQLY
78	GGAGNDTLS
79	GGFGNDVYI
80	GGDGNDKLI
81	GGDGNDKLT
82	GGAGDDTIF
83	GGGGDTLW
84	GGDGNDVLL
85	GKGDII
86	GGLGSDVMV
87	GGYNDIYR
88	GGEGDLLK
89	GGGGHDMQ
90	<b>GGAGNDTLVGGDGNSFV</b>
91	<b>GGAGDDRDLGGAGNDTLV</b>
92	<b>GGAGNDTLVGGAGNSFV</b>
93	<b>GGAGNDTLVGGDGNSFI</b>
94	<b>GGAGNDTLVGGDGNSLV</b>

95	<b>GGAGNDTLVGGDGADSFV</b>
96	<b>GGAGNDTFVGGAGNDSFV</b>
97	<b>GGAGNDTLVGGAGNDSFI</b>
98	<b>GGAGNDTLVGGDGADTFI</b>
99	<b>GGAGNDTLVGGAGDDTFV</b>
100	<b>GGAGKDTLAGGAGNDTLV</b>
101	<b>GGDGNDTVAGGAGNDTLV</b>
102	<b>GGAGNDTLVGGDGDDFSFI</b>
103	<b>GGAGNDTLVGGAGNDTLV</b>
104	<b>GGDGDDTLTGGAGNDTLV</b>
105	<b>GGAGNDTLVGGDGNDTFI</b>
106	<b>GGAGNDTLDGGAGNDTLV</b>
107	<b>GGAGNDTLVGGAGADTFV</b>
108	<b>GGAGNDTLVGGAGNDTFV</b>
109	<b>GGLGDDELSGGAGNDTLV</b>
110	<b>GGQQQDELIGGAGNDTLV</b>
111	<b>GGAGNDTLVGGAGNDSLII</b>
112	<b>GGAGNDSLIVGGAGNDTLV</b>
113	<b>GGAGADTLVGGAGNDTFV</b>
114	<b>GGAGNDTLVGGDGDDTLV</b>
115	<b>GGAGNDTLVGGAGADTFI</b>
116	<b>GGDGGDRLEGGAGNDTLV</b>
117	<b>GGAGNDTLVGGDGNDTLI</b>
118	<b>GGTGDDLLLGGAGNDTLV</b>
119	<b>GGSGDDRDLGGAGNDTLV</b>
120	<b>GGGGKDSLIVGGAGNDTLV</b>
121	<b>GGQQQDLLEGGAGNDTLV</b>
122	<b>GGAGNDLLDGGAGNDTLV</b>
123	<b>GGAGNDTLVGGYGNDIYR</b>
124	<b>GGAGNDTLVGGDGDDSLII</b>
125	<b>GGAGNDTLVGGAGADTLV</b>
126	<b>GGAGADTLVGGAGNDTLV</b>
127	<b>GGAGNDTLVGGAGNDSLIV</b>
128	<b>GGAGNDTLVGGAGNDTLI</b>
129	<b>GGAGNDTLVGGDGADTLI</b>
130	<b>GGAGSDRLLGGAGNDTLV</b>
131	<b>GGAGNDTLVGGDGADSLIV</b>
132	<b>GGSGVDQLFVGGAGNDTLV</b>
133	<b>GGAGNDTLVGGDGNDVLL</b>
134	<b>GGAGNDTLVGGSGVDSYR</b>
135	<b>GGAGNDTLVGGAGADSLII</b>
136	<b>GGAGNDTLVGGDGNDKLI</b>
137	<b>GGAGNDTLVGGAGNDTLS</b>
138	<b>GGAGNDTLVGGAGNDRID</b>
139	<b>GGAGNDTLVGGAGDDTIF</b>
140	<b>GGAGNDTLVGGAGNDTLD</b>
141	<b>GGAGNDTLVGGSGSDQMN</b>
142	<b>GGAGNDTLVGGKGNDTLN</b>
143	<b>GGAGNDTLVGGNGDDQLY</b>
144	<b>GGAGNDTLVGGAGDDRDL</b>
145	<b>GGAGNDTLVGGVGSVDMV</b>
146	<b>GGAGNDTLVGGKGDIIID</b>
147	<b>GGAGNDTLVGGIGSDRAD</b>
148	<b>GGAGNDTLVGGSGDDRDL</b>
149	<b>GGAGNDTLVGGAGADRLN</b>
150	<b>GGAGNDTLVGGAGADRLD</b>
151	<b>GGAGNDTLVGGAGNDVLN</b>
152	<b>GGAGNDTLVGGKGNDDLFI</b>





205	GGGGHDRMQGGAGNDFV
206	GGGGHDRMQGGAGNDSFV
207	GGGGHDRMQGGDGNDSEFI
208	GGGGHDRMQGGIGYDDIY
209	GGGGHDRMQGGSGQDLLR
210	GGGGHDRMQGGAGFDTLV
211	GGGGHDRMQGGFGFDRLN
212	GNSLAKNVLF
213	YLSGYGHIIIDDDGGKEDKLSLADIDFRDVAFKREGNDLIMYKAEGNVLSIGHKNGITFRN WFEKESGDISHQIEQIFDKSGRIITPDSLKKALEYQQRNNKASYVYGNDALAYGSQ
214	GNSLAKNVLFGGKGNKLYGSEGADLLDGGEGDILLKGGYGNDIYRYLSGYGHIIIDDD GGKEDKLSLADIDFRDVAFKREGNDLIMYKAEGNVLSIGHKNGITFRNWFEKESGDISH QIEQIFDKSGRIITPDSLKKALEYQQRNNKASYVYGNDALAYGSQGDNLPLINEISKIISAAG SFDVKEERTAASLLQLSGNASDFSYGRNSITLTTSA
215	GNSLAKNVLFYLSGYGHIIIDDDGGKEDKLSLADIDFRDVAFKREGNDLIMYKAEGNVLSI GHKNGITFRNWFEKESGDISHQIEQIFDKSGRIITPDSLKKALEYQQRNNKASYVYGND ALAYGSQ
216	GGGGS
217	ENLYFQX
218	GGKGNKLYGSEGADLLDGGEGDILLKGGYGNDIYR
219	NSDSECLSHDGYCLHDGVCMIYIYALDKYACNCVVGYIGERCQYRDLKWWELR
220	GGEGDILLKGGYGNDIYR

### *Materials and methods*

Expression host: Escherichia coli BL21 DaraB (Novagen).

All oligonucleotides were purchased from Microsynth Seqlab GmbH.

Codon optimized nucleotide sequences encoding peptides were purchased from Thermo Fisher Scientific. All enzymes were purchased from NEB, Clontech, Invitrogen or Fermentas.

### *Expression protocol*

1. Transformation of chemically competent cells with an expression vector encoding for a first amino acid sequence that is a scaffold derived from HlyA1 with the amino acid sequence set forth in SEQ ID NO:215 with the GG repeat elements, as disclosed below, being inserted between positions 10 and 11 of SEQ ID NO:215 and the peptide/polypeptide of interest (second amino acid sequence; as a fusion protein) and plating of the transformed cells on LB agar plates comprising suitable antibiotic(s) for selection of the transformed cells.
2. Incubation of the agar plates over night at 37°C.
3. Inoculation of 2YT medium comprising antibiotics with a single colony from the agar plate for an overnight culture.
4. Incubation at 37°C and shaking of the culture over night.
5. Inoculation of the main culture comprising 2x YT medium (16 g tryptone/peptone from casein (Roth, #8952.2), 10 g yeast extract (Roth, #2363.2), 5 g NaCl (Roth, #3957.1), ad 1 L demineralized water with the overnight culture resulting in an OD600 of 0.01 – 0.2 (flasks with baffles)
6. Incubation of the culture at 37°C at different rpm
7. Induction of the expression the peptide or protein of interest with 2g/L arabinose at an OD600 of 0.4-1.0.2
8. Incubation of the cultures for 4 hrs.

9. Culture samples were taken at 0 hrs and 3 hrs post induction and centrifuged for 10 min., 13,000 x g, RT.

10. Cell samples were resuspended in water to adjust an OD of 5.0, mixed 4:1 with 5x SDS loading dye and heated (95°C, 10 min).

11. 20 µL samples were loaded on 15% SDS-PAGE gels and SDS-PAGE analysis was performed at 160 V for about 45 min.

*Example 1: Optimized GG repeat sequences for renaturation*

Fusion constructs of the modified HlyA1 scaffold (SEQ ID NO:215) with different GG repeat sequences being inserted between positions 10 and 11 (cf. Table 1) and an N-terminal M as first amino acid sequence with a polypeptide of interest (hEGF; SEQ ID NO:219) C-terminally fused thereto via a linker comprising a 5 amino acid long G rich sequence (SEQ ID NO:216) and a TEV protease site (SEQ ID NO:217) were designed.

hEGF was used as test peptide, because renaturation efficiencies with HlyA1-D166-219 are moderate, so that renaturation data fall in a middle range and allow comparison with constructs significantly increasing or decreasing renaturation efficiencies.

Fusion constructs were expressed as inclusion bodies (lbs) in *E.coli* BL21 (DE3), lbs were prepared via BugBuster Kit and the lbs dissolved in 6 M GuHCl (1:4 w/v). After solubilization overnight, 20 mg/mL stock solution of each construct was set in 6 M GuHCl and 5 mM TCEP (Tris(2-carboxyethyl)phosphine) added. Renaturation was performed at room temperature for 20 min setting a final concentration of 1 mg/mL in a total volume of 500 µL renaturation buffer with calcium ions (10 mM Tris-HCl, 120 mM NaCl, pH 7.3 supplemented with 10 mM CaCl<sub>2</sub>) or without calcium ions (HEPES + EDTA). The renaturation efficiency was determined in the cleared supernatant by UV/Vis spectroscopy.

The wildtype control (wt) exhibits a renaturation value of nearly 50 % in renaturation buffer, with high calcium dependence. When removing its first two GG repeats, the resulting natural dimer (SEQ ID NO:220) leads to a renaturation value slightly above 60 %.

When replacing the wt GG repeat cassette of the control by a dimer made of an inactivated version of GGAGNDTLV (SEQ ID NO:1) called D6A (GGAGNATLV; SEQ ID NO:169) N-terminally plus the unaltered GGAGNDTLV (SEQ ID NO:1) at the C-terminal position, the renaturation value is slightly reduced, but stays over 40 %. When inverting such a dimer, leaving the inactivated monomer D6A in the C-terminal position, the renaturation value goes down to nearly 20 %. This demonstrates that the C-terminal position is decisive for the resulting renaturation.

Moreover, when replacing the GG repeats by a dimer made of twice the inactivated monomer D6A, the renaturation value goes further down to around 10 %, confirming the cumulative detrimental effect of the inactivated monomers in the resulting renaturation.

Table 2 shows the results. GGR1, GGR2, GGR3, GGR4, GGR5, GGR6 and GGR7 are given as the respective SEQ ID Nos. It is understood that in said table the C-terminal GG repeat, which may be GGR2, GGR3, GGR4, GGR5, GGR6 or GGR7, corresponds to GGRC thus, for the sake of representation in a single table, deviating from the nomenclature used in the rest of this application. All GG repeats were directly linked to each other by a peptide bond. SEQ ID NO:169 is an inactivated control GG repeat. SEQ ID NO:218 is the natural GG repeat element of the scaffold of SEQ ID NO:215. All GG repeat sequences matching the consensus for GGRC are given in bold. Also given is the SEQ ID NO: for the complete GG repeat element comprising all GG repeats present (GG repeat cassette).

Table 2

Nr.	GGR1	GGR2	GGR3/GGR4/GGR5/GGR6/GGR7	Renaturation efficiency (%) NanoDrop-based (+Ca)	GG repeat cassette (SEQ NO ID)
1	<b>1</b>	<b>31</b>	-	71.3	90
2	63	<b>1</b>	-	68.1	91
3	<b>1</b>	<b>7</b>	-	68.1	92
4	<b>1</b>	<b>32</b>	-	67.9	93
5	<b>1</b>	<b>29</b>	-	67.2	94
6	<b>1</b>	<b>39</b>	-	66.1	95
7	<b>3</b>	<b>7</b>	-	64.8	96
8	<b>1</b>	<b>8</b>	-	64.6	97
9	<b>1</b>	<b>36</b>	-	64.1	98
10	<b>1</b>	<b>19</b>	-	63.7	99
11	72	<b>1</b>	-	62.9	100
12	71	<b>1</b>	-	62.6	101
13	<b>1</b>	<b>48</b>	-	62.2	102
14	<b>1</b>	<b>1</b>	-	60.9	103
15	68	<b>1</b>	-	60.9	104
16	<b>1</b>	<b>28</b>	-	60.0	105
17	66	<b>1</b>	-	59.6	106
18	<b>1</b>	<b>11</b>	-	57.9	107
19	<b>1</b>	<b>3</b>	-	57.7	108
20	76	<b>1</b>	-	56.4	109
21	67	<b>1</b>	-	56.3	110
22	<b>1</b>	<b>6</b>	-	56.3	111
23	<b>5</b>	<b>1</b>	-	56.0	112

24	9	3	-	56.0	113
25	1	41	-	55.9	114
26	1	12	-	54.6	115
27	70	1	-	54.3	116
28	1	26	-	52.5	117
29	75	1	-	52.1	118
30	62	1	-	51.6	119
31	64	1	-	50.6	120
32	74	1	-	49.7	121
33	65	1	-	49.6	122
34	1	87	-	48.6	123
35	1	46	-	48.0	124
36	1	9	-	47.6	125
37	9	1	-	45.5	126
38	1	5	-	45.1	127
39	1	2	-	43.6	128
40	1	34	-	41.9	129
41	73	1	-	41.1	130
42	1	37	-	40.3	131
43	1	84	-	32.9	133
44	1	166	-	30.9	134
45	1	80	-	28.2	136
46	1	78	-	27.9	137
47	1	50	-	27.8	138
48	1	82	-	25.9	139
49	1	66	-	25.7	140
50	1	52	-	25.0	141
51	1	58	-	23.5	142
52	1	77	-	22.9	143
53	1	63	-	22.5	144
54	1	86	-	22.4	145
55	1	85	-	22.3	146
56	1	167	-	21.3	147
57	1	62	-	20.4	148
58	1	51	-	20.1	149
59	1	49	-	19.4	150
60	1	56	-	19.0	151
61	1	57	-	18.0	152

62	1	168	-	18.0	153
63	1	83	-	16.0	154
64	1	54	-	14.5	155
65	1	65	-	14.2	156
66	1	55	-	13.7	157
67	1	61	-	12.8	158
68	1	53	-	8.8	159
69	1	59	-	8.7	160
70	1	60	-	6.5	161
71 (wt)	186	187	88/87	47.6	162
72 (control)	1	169	-	23.5	163
73 (control)	169	169	-	10.1	164
74 (control)	169	1	-	42.9	165
75	88	87	-	62.6	220
76	1	3	5/7	73.6	170
77	1	1	1/5	69.4	171
78	1	1	1/7	66.5	172
79	5	5	7/3	66.2	173
80	1	7	1/3	63.9	174
81	1	3	5/5	63.5	175
82	5	7	3/3	59.6	176
83	7	7	1/1	59.2	177
84	3	3	1/1	57.2	178
85	1	5	5/1	54.6	179
86	3	5	3/7	49.6	180
87	9	9	3	84.6	181
88	1	9	3/9/9/9/1	76.5	182
89	9	9	9/3	66.4	183
90	1	9	9/3	63.1	184
91	1	9	3	61.1	185

It can be seen from the data in Table 2 that the inventive fusion constructs provided for similar or better renaturation efficiency compared to a longer control construct (wt; or controls with 1 or 2 inactivated GG

repeats) or those constructs that do not have a C-terminal GG repeat sequence that meets the consensus sequence defined herein. It needs to be noted that for first amino acid sequences (with two GG repeat sequences) even renaturation efficiencies slightly lower than the wt control (with four GG repeats/repeat-like sequences) may still be advantageous, as the proportion of the peptide/polypeptide of interest in the construct is significantly higher.

*Example 2: Optimized GG repeat sequences for renaturation*

The experiments carried out in Example 1 were repeated with a different scaffold, namely the C-terminal part of RTX12 (Endonuclease from *Caulobacter* sp. K31). The RTX12 fragment used consisted of a start methionine and the last 190 amino acids of the native endonuclease from *Caulobacter* sp. K31 (SEQ ID NO:189), containing a stretch of five canonical GG repeats and one consensus GG repeat (SEQ ID NO:194).

It was known from previous experiments that renaturation of the RTX12 scaffold fused to a polypeptide of interest (hEGF; SEQ ID NO:219) C-terminally via a linker comprising a 5 amino acid long G rich sequence (SEQ ID NO:216) and a TEV protease site (SEQ ID NO:217) in renaturation buffer is extremely high, probably because its C-terminal native GG repeat is a consensus sequence.

Renaturation of this construct was almost 100 %, with very low calcium dependency. The more GG repeats were deleted, the lower the resulting renaturation efficiency and the higher the calcium dependence. With the natural GG repeat dimer corresponding to the two most C-terminal of the six native GG repeat sequences in the C-terminal part of RTX12, a renaturation efficiency of 60 % was achieved in renaturation buffer. It was decided to take this RTX12 scaffold with the dimer of SEQ ID NO:197 (which already meets the criteria set out herein for the GG repeat element, as its terminal GG repeat matches the given consensus sequence) as a reference for analyzing the influence of constructs with different monomers on the resulting renaturation.

Baseline data for constructs based on the dimer of SEQ ID NO:197 were then collected. When replacing the N-terminal GG repeat by an inactivated version of GG71 called 71D6A (GGGGHARMQ) (SEQ ID NO:198), the renaturation value goes down to nearly 50 %. In contrast, when replacing the C-terminal GG repeat by an inactivated version of GG50 called 50D6A (GGAGDATFV) (SEQ ID NO:199), the renaturation efficiency goes down to nearly 35 %. Finally, when replacing both GG repeats by a dimer made of the inactivated monomers 71D6A and 50D6A, the renaturation value goes further down to 32 %, confirming also here the importance of the C-terminal position in the GG repeat stretch and the cumulative detrimental effect of inactivated monomers in the resulting renaturation:

With these data as foundation, an experiment with artificial dimers replacing the GG repeat stretch of RTX12\_G\_TEV\_hEGF was conducted, leaving the N-terminal position fixed as GG71 and setting different consensus or non-consensus monomers in the C-terminal position. The results show that out of 5 tested consensus monomers, only 2 lead to a higher renaturation value than the dimer GG71- GG50

(SEQ ID NO:197). Nevertheless, the 4 tested non-consensus monomers reduce the resulting renaturation of the dimeric construct of SEQ ID NO:197 when set at the C-terminal position of the dimer instead of last GG repeat: Again, all GG repeats matching the consensus sequence for GGRC are given in bold. Further, the nomenclature deviates from the rest of the description in that the GG repeats are consecutively numbered. This means the highest numbered GG repeat is GGRC in this setup. As different lengths of GG repeat cassettes are used, GGRC may be GGR2, GGR3, GGR4 or GGR6.

Table 3.

Nr.	GGR1	GGR2	GGR3/GGR4/ GGR5/GGR6	Renaturation efficiency (%) NanoDrop-based (+Ca)	GG repeat cassette (SEQ NO ID)
94 (wt)	190	191	192/193/89/19	99.8	194
95	192	193	89/19	95.6	195
96	193	89	19	69.9	196
97	89	19	-	60.0	197
98 (control)	198	19	-	49.6	200
99 (control)	89	199	-	34.8	201
100 (control)	198	199	-	32.0	202
101	89	29	-	98.3	203
102	89	31	-	64.8	204
103	89	3	-	39.2	205
104	89	7	-	37.4	206
105	89	32	-	35.9	207
106	89	61	-	27.8	208
107	89	60	-	22.2	209
108	89	55	-	15.4	210
109	89	59	-	15.4	211

With these results, it can be concluded that the positive effect observed for our consensus monomers at the C-terminal position of a GG repeat dimer in in the HlyA1 scaffold described in Example 1 can also be transferred to the truncated RTX12 scaffold, when replacing its C-terminal GG repeat GG50 (SEQ ID NO: 19) by certain consensus monomers.



In summary, our data clearly demonstrate that the artificial consensus sequence of GG repeats described herein and oligomers of this sequence feature a very good calcium-dependent renaturation of fusion proteins based on different scaffolds, such as the HlyA1 scaffold, leading to a renaturation efficiency significantly higher than with the native GG repeat stretch

All documents cited herein, are hereby incorporated by reference in their entirety. The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention. The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group. Further embodiments of the invention will become apparent from the following claims.

**CLAIMS**

1. Isolated polypeptide comprising a first and a second amino acid sequence, wherein

(a) the first amino acid sequence

(a1) comprises in N- to C-terminus orientation the general amino acid sequence element



wherein GGR1 and each GGRN independently are a GG repeat sequence of the consensus sequence GGxGxDxUx, wherein each x can independently be any amino acid and U is independently a hydrophobic, large amino acid selected from F, V, A, Y, I, L and M;

wherein GGRC is a GG repeat sequence of the consensus sequence GGX<sup>11</sup>GX<sup>12</sup>DX<sup>13</sup>X<sup>14</sup>X<sup>15</sup>, wherein X<sup>11</sup> is selected from A and D; X<sup>12</sup> is selected from N, A, and D; X<sup>13</sup> is selected from T and S; X<sup>14</sup> is L or F; and X<sup>15</sup> is V or I;

wherein each "Linker" is independently either a peptide bond or an amino acid sequence of 1 to 25 amino acids not comprising a GG repeat sequence as defined above, preferably a peptide bond;

wherein n is 0 or an integer from 1 to 10, preferably 0 or 1 to 8, more preferably 0 or 1 to 4;

(a2) wherein the first amino acid sequence is 18 to 300, preferably 30 to 200 amino acids in length; and

(a3) wherein the first amino acid sequence and/or the amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is a non-natural sequence element, wherein, optionally, the amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%, and/or wherein, optionally, the complete first amino acid sequence is a non-natural sequence in that its sequence identity over its entire length relative to any naturally occurring polypeptide or fragment thereof is less than 100%;

(b) the second amino acid sequence is at least one peptide or polypeptide of interest, and

(c) wherein the first and second amino acid sequence are heterologous to each other in that they do not occur in combination in a single naturally occurring polypeptide or fragment thereof,

wherein, preferably, the first amino acid sequence increases the solubility, stability and/or renaturation efficiency of the second amino acid sequence relative to a polypeptide not comprising a GG repeat sequence or the first amino acid sequence.

2. The isolated polypeptide of claim 1, wherein GGRC is selected from any one of the amino acid sequences set forth in SEQ ID Nos. 1-48.

3. The isolated polypeptide of claim 1 or 2, wherein the general amino acid sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is selected from the following:

(1) GGR1-Linker-GGRC; or

(2) GGR1-Linker-GGR2-Linker-GGRC;

(3) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGRC;

(4) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGRC;

(5) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGRC; or

- (6) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGRC; or  
(7) GGR1-Linker-GGR2-Linker-GGR3-Linker-GGR4-Linker-GGR5-Linker-GGR6-Linker-GGR7-Linker-GGRC;

wherein GGR2, GGR3, GGR4, GGR5, GGR6, and GGR7 are defined as GGRN above.

4. The isolated polypeptide of any one of claims 1 to 3, wherein the sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC is selected from the amino acid sequences set forth in SEQ ID Nos. 90-122, 124-131, 165, 170-185, 194-197, 200, 203 and 204.

5. The isolated polypeptide of any one of claims 1 to 4, wherein the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) according to claim 3 is flanked by additional N- and/or C-terminal amino acid sequences, optionally each being up to 150 amino acids in length.

6. The isolated polypeptide of claim 5, wherein

(1a) the N-terminal flanking sequence is 1 to 100 amino acids in length, preferably 1 to 80, 1 to 70, 5 to 60, 5 to 50, 5 to 40, 5 to 30, 8 to 20, 10 to 15 or 10 amino acids in length; and/or

(1b) the N-terminal flanking sequence comprises or consist of the amino acid sequence set forth in SEQ ID NO:212; and/or

(2a) the C-terminal flanking sequence is 10 to 150 amino acids in length, preferably 20 to 120, 30 to 120, 40 to 120, 50 to 120, 60 to 120, 70 to 120, 80 to 120, 90 to 120, 100 to 120 or 110 to 120 amino acids in length; and/or

(2b) the C-terminal flanking sequence comprises or consist of the amino acid sequence set forth in SEQ ID NO:213; and/or

(3) the N- and/or C-terminal flanking sequences are scaffold sequences derived from a natural occurring RTX protein, in which any one or more of the naturally occurring GG repeat sequences have been replaced by the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) according to claim 3; and/or

(4) the scaffold sequence is derived from hemolysin A protein, preferably hemolysin A1 having the amino acid sequence set forth in SEQ ID NO:214 or a variant or fragment thereof, more preferably the amino acid sequence as set forth in SEQ ID NO:215 or a variant or fragment thereof, wherein the general sequence element GGR1-Linker-(GGRN-Linker)<sub>n</sub>-GGRC or any of the general structures (1)-(7) according to claim 3 are inserted between positions 10 and 11 using the positional numbering of SEQ ID NO:215.

7. The isolated polypeptide of any one of claims 1 to 6, wherein

(1) U in the consensus sequence GGxGxDxUx is selected from F, V, I, L, M and Y or F, I, L, M and Y; and/or

(2) the consensus sequence of the GG repeat sequences GGR1, GGR2, GGR3, GGR4, GGR5, GGR6, GGR7 and GGRN is GGX<sup>1</sup>GX<sup>2</sup>DX<sup>3</sup>UX<sup>4</sup>, wherein X<sup>1</sup>, X<sup>2</sup>, X<sup>3</sup> and X<sup>4</sup> may be any amino acid with the

exception of P, preferably  $X^1$  is selected from G, A, L, V, I, M, F, S, T, Q; Y, K, R, D, E, N, and Y, preferably from G, A, E, S, T, Q, L, R and D, more preferably from G, D, A, E, S and T; and/or  $X^2$  is selected from N, D, A, G, S, H, T, E, M, R and H, preferably from N, D, A, S, and H, more preferably from N, D, A, and S; and/or  $X^3$  is selected from A, K, V, L, I, F, M, R, Y, S, L, T, V, Q, N, D, E, and H, preferably from T, R, V, L, S, I, A, Y, Q, D and H, more preferably from T, R, V, L, S and I, even more preferably from S, T, R and V; and/or  $X^4$  is selected from W, K, R, Y, V, L, T, H, D, A, M, E, F, I, S, N and Q, preferably from V, L, I, F, S, R, N, Y and T.

8. The isolated polypeptide of any one of claims 1 to 7, wherein "Linker" between GGR1 and GGR2 is a peptide bond.

9. The isolated polypeptide of any one of the preceding claims, wherein

(1) the isolated polypeptide is a fusion protein; and/or

(2) the second amino acid sequence is 10 to 1000 amino acids in length, preferably 15 to 500 amino acids in length; and/or

(3) the second amino acid sequence is N-terminal or C-terminal to the first amino acid sequence, preferably C-terminal, wherein the second amino acid sequence is optionally linked directly or via a linker sequence to the N- or C-terminal end of the first amino acid sequence, the linker sequence optionally being 1 to 30 amino acids in length and preferably comprising a protease recognition and cleavage site, the protease recognition and cleavage site preferably being a TEV protease recognition and cleavage site; and/or

(4) the isolated polypeptide has a total length of up to 700 amino acids, preferably up to 500 amino acids, more preferably up to 350 amino acids.

10. Nucleic acid encoding the polypeptide of any one of claims 1 to 9.

11. Vector comprising a nucleic acid molecule according to claim 10, wherein the vector preferably is a plasmid.

12. Host cell comprising a nucleic acid molecule according to claim 10 or a vector according to claim 11, the host cell preferably being a prokaryotic host cell.

13. Method for the production of a polypeptide of any one of claims 1 to 9, comprising

(1) cultivating the host cell of claim 12 under conditions that allow the expression of the polypeptide;

(2) isolating the expressed polypeptide from the host cell, wherein the expression of the polypeptide in the host cell optionally is in insoluble form, preferably in form of inclusion bodies; and

(3) renaturing the isolated polypeptide.

14. Method for renaturing the isolated polypeptide of any one of claims 1 to 9 or obtained according to the method of claim 13, wherein the method comprises dissolving or dispersing the isolated polypeptide in an aqueous medium suitable for renaturation, optionally an aqueous medium comprising earth alkaline metal ions, preferably calcium ions.

15. Use of the isolated polypeptide of any one of claims 1-9 for expression and renaturation of the second amino acid sequence in form of a fusion protein, wherein the expression is in a recombinant host cell in form of inclusion bodies.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/075465

## Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
  - a.  forming part of the international application as filed.
  - b.  furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)).  
 accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2.  With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

**INTERNATIONAL SEARCH REPORT**

International application No  
**PCT/EP2023/075465**

**A. CLASSIFICATION OF SUBJECT MATTER**  
**INV. C07K14/00 C07K14/435 C07K14/195 C12P21/00 C07K4/00**  
**ADD.**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**  
 Minimum documentation searched (classification system followed by classification symbols)  
**C07K C12P**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
**EPO-Internal, WPI Data, BIOSIS, EMBL, Sequence Search**

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<b>X</b>	<p><b>DATABASE UniProt [Online]</b></p> <p><b>22 November 2017 (2017-11-22),</b>  <b>"SubName: Full-Hemolysin-type</b>  <b>calcium-binding repeat-containing protein</b>  <b>{ECO:0000313 EMBL:SFL17339.1}; Flags:</b>  <b>Fragment;"</b>,  <b>XP093104518,</b>  <b>retrieved from EBI accession no.</b>  <b>UNIPROT:A0A1I4FIR5</b>  <b>Database accession no. A0A1I4FIR5</b>  <b>sequence</b></p> <p align="center">-----</p>	<b>1</b>
<b>X</b>	<p><b>WO 2021/185968 A1 (NUMAFERM GMBH [DE])</b>  <b>23 September 2021 (2021-09-23)</b>  <b>Whole document, especially page 11, 3rd</b>  <b>paragraph and claims.</b></p> <p align="center">-----</p> <p align="center">-/--</p>	<b>1-15</b>

Further documents are listed in the continuation of Box C.       See patent family annex.

\* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p>
---	---

Date of the actual completion of the international search <b>28 November 2023</b>	Date of mailing of the international search report <b>07/12/2023</b>
--	---

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer <b>Kools, Patrick</b>
--	---

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/075465

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 2014/170430 A1 (HEINRICH HEINE UNIVERSITÄT DÜSSELDORF [DE]) 23 October 2014 (2014-10-23) cited in the application Whole document, especially the claims -----</p>	1-15
X	<p>EP 3 892 290 A1 (NUMAFERM GMBH [DE]) 13 October 2021 (2021-10-13) cited in the application Whole document, especially the claims -----</p>	1-15
X,P	<p>WO 2022/194403 A1 (NUMAFERM GMBH [DE]) 22 September 2022 (2022-09-22) the whole document -----</p>	1-15



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

**PCT/EP2023/075465**

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
<b>WO 2021185968</b>	<b>A1</b>	<b>23-09-2021</b>	<b>EP 4121086 A1</b>	<b>25-01-2023</b>
			<b>EP 4121444 A1</b>	<b>25-01-2023</b>
			<b>US 2023142538 A1</b>	<b>11-05-2023</b>
			<b>US 2023144204 A1</b>	<b>11-05-2023</b>
			<b>WO 2021185968 A1</b>	<b>23-09-2021</b>
			<b>WO 2021185969 A1</b>	<b>23-09-2021</b>
			<b>WO 2022194375 A1</b>	<b>22-09-2022</b>
-----				
<b>WO 2014170430</b>	<b>A1</b>	<b>23-10-2014</b>	<b>AU 2014255697 A1</b>	<b>03-12-2015</b>
			<b>BR 112015026384 A2</b>	<b>10-10-2017</b>
			<b>CA 2912300 A1</b>	<b>23-10-2014</b>
			<b>CN 105473609 A</b>	<b>06-04-2016</b>
			<b>DK 2986633 T3</b>	<b>12-03-2018</b>
			<b>EP 2792686 A1</b>	<b>22-10-2014</b>
			<b>EP 2986633 A1</b>	<b>24-02-2016</b>
			<b>ES 2661404 T3</b>	<b>28-03-2018</b>
			<b>JP 6714902 B2</b>	<b>01-07-2020</b>
			<b>JP 2016515396 A</b>	<b>30-05-2016</b>
			<b>NO 2986633 T3</b>	<b>30-06-2018</b>
			<b>PL 2986633 T3</b>	<b>29-06-2018</b>
			<b>PT 2986633 T</b>	<b>08-03-2018</b>
			<b>TR 201802551 T4</b>	<b>21-03-2018</b>
			<b>US 2016138066 A1</b>	<b>19-05-2016</b>
			<b>WO 2014170430 A1</b>	<b>23-10-2014</b>
-----				
<b>EP 3892290</b>	<b>A1</b>	<b>13-10-2021</b>	<b>NONE</b>	
-----				
<b>WO 2022194403</b>	<b>A1</b>	<b>22-09-2022</b>	<b>NONE</b>	
-----				