(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 646 650 B9**

(12) **CORRECTED EUROPEAN PATENT SPECIFICATION**

(15) Correction information:
**Corrected version no 1 (W1 B1)**
**Corrections, see**
**Claims DE 3**

(48) Corrigendum issued on:
**09.09.2009 Bulletin 2009/37**

(45) Date of publication and mention of the grant of the patent:
**22.04.2009 Bulletin 2009/17**

(21) Application number: **04755315.1**

(22) Date of filing: **14.06.2004**

(51) Int Cl.:
***C07K 14/65*** *(2006.01)* ***C07K 14/00*** *(2006.01)*

(86) International application number:
**PCT/US2004/019070**

(87) International publication number:
**WO 2004/113382 (29.12.2004 Gazette 2004/53)**

# (54) RECOMBINANT IGF EXPRESSION SYSTEMS

RECOMBINANTE IGF EXPRESSIONSSYSTEME

SYSTEMES D'EXPRESSION D'IGF RECOMBINE

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PL PT RO SE SI SK TR**

(30) Priority: **13.06.2003 US 477941 P**

(43) Date of publication of application:
**19.04.2006 Bulletin 2006/16**

(73) Proprietor: **Insmed, Inc.**
**Richmond, VA 23235 (US)**

(72) Inventors:
- **SUMNER, Emma, Jayne**
  **Washington,**
  **Tyne & Wear NE38 7TB (GB)**
- **HODGSON, Ian**
  **North Yorkshire, TS14 6DJ (GB)**

(74) Representative: **Viering, Jentschura & Partner**
**Postfach 22 14 43**
**80504 München (DE)**

(56) References cited:
**WO-A-96/40722** **US-A- 6 068 994**

- **OLSON P ET AL: "High-Level Expression of Eukaryotic Polypeptides from Bacterial Chromosomes" PROTEIN EXPRESSION AND PURIFICATION, ACADEMIC PRESS, SAN DIEGO, CA, US, vol. 14, no. 2, November 1998 (1998-11), pages 160-166, XP004445187 ISSN: 1046-5928**
- **BAKER R T: "Protein expression using ubiquitin fusion and cleavage." CURRENT OPINION IN BIOTECHNOLOGY. OCT 1996, vol. 7, no. 5, October 1996 (1996-10), pages 541-546, XP002303484 ISSN: 0958-1669**



EP 1 646 650 B9

## Description

**[0001]** This application claims the benefit of U.S. Provisional Application No. 60/477,941, filed June 13, 2003.

## BACKGROUND OF THE INVENTION

Field of the Invention

**[0002]** The present invention provides recombinant insulin-like growth factor (IGF) expression systems that yield high levels of product which are easily processed to provide IGF with a proper N-terminus for human use.

Description of the Related Art

**[0003]** Recombinant production allows the large scale manufacturing of therapeutic proteins while avoiding many of the difficulties and hazards of protein purification from natural sources. In the case of human proteins, recombinant manufacturing is frequently the only practical method for producing the amounts of protein required for commercial sales of therapeutic products. Recombinant production also eliminates worker exposure to human fluid and tissues, avoiding potential exposure to infectious agents such as viruses.

**[0004]** Recombinant manufacturing involves the expression of a DNA construct encoding for the desired protein in a recombinant host cell. The host cell can be either prokaryotic (*e.g.*, bacteria such as *Escherichia coli)* or eukaryotic (*e.g.*, yeast or mammalian cell-line). For large scale recombinant manufacturing, bacterial or yeast host cells are most commonly used, due to the ease of manipulation and growth of these organisms and also because these organisms require relatively simple growth media.

**[0005]** Recombinant manufacturing, however, does have its difficulties. Expression constructs must be optimized for a particular protein and for a particular host cell. Expressing a recombinant protein in a host cell exposes the recombinant protein to a new set of host cell enzymes, such as proteases, which can modify or even degrade the recombinant protein. Modification and degradation of the recombinant protein is undesirable, as it decreases yields and can complicate the purification of the recombinant protein. Polypeptides over-expressed in the bacterial cytoplasm often accumulate as insoluble "inclusion bodies" (Williams et al., Science 215:687-688 (1982); Schoner et al., Biotechnology 3:151-154 (1985)). Polypeptides accumulated in the form of inclusion bodies are relatively useless. Conversion of this insoluble material into active, soluble polypeptide requires slow and difficult solubilization and refolding protocols that often greatly reduce the net yield of biologically active polypeptide. This problem has particularly impacted the production of IGF, resulting in numerous attempts to solve the "refolding problem." Even when polypeptides are expressed in the cytoplasm of bacteria in soluble form, they often accumulate poorly as a result of degradation by host proteases. Furthermore, the accumulated polypeptides often lack the desired amino terminus. This problem is commonly addressed by the expression of a fusion protein, in which the N-teminus of the desired polypeptide is fused to a carrier protein.

**[0006]** However, the use of fusion polypeptides has drawbacks. It is often necessary to cleave the desired polypeptide away from the fusion partner by enzymatic or chemical means. This can be accomplished by placing an appropriate target sequence for cleavage between the fusion partner and the desired polypeptide. Unfortunately, the enzymes most widely used for polypeptide cleavage are expensive, inefficient, or imprecise in their cleavage and cannot be successfully applied to a majority of fusion constructs. For example, enterokinase and Factor Xa are mammalian enzymes that are expensive to produce and exhibit highly variable cleavage efficiency. Meanwhile, enzymes like subtilisin are relatively inexpensive to produce, but their precision is unacceptable for commercial-scale processes under current "Good Manufacturing Practices" (GMP). The human rhinovirus 14 protease, termed 3C protease, is a robust, precise, and inexpensive enzyme that cleaves the amino acid sequence E-(V or T)-L-F-Q-G-P immediately N-terminal to the glycine residue. 3C protease has been used to cleave IGF from dsbA-IGF fusion constructs (Olson et al., Protein Expr Purif. 14:160-166 (1998) WO 96/40722A). As a substantial portion of the dsbA-IGF is refractory to cleavage, however, 3C protease may not be universally applicable to 3C site containing IGF fusion constructs.

**[0007]** The patent literature describes numerous fusion systems for the production of IGF, each of which has certain advantages and disadvantages. For example, constructs encoding native mature IGF-1 (with an additional methionine) yield low expression levels. A variety of approaches were developed to enhance expression, including the use of fusion partners (Schulz M.F. et al., J. Bact. 169:585-53921(1987)), and the use of multiple protease-deficient hosts (Buell et al., Nucleic Acids Res. 13:1923-1938 (1985)). However, the growth characteristics of these protease deficient hosts are not ideal for high intensity fermentation. It also has been found that good expression can be obtained simply by adding Met-Arg-Lys to the N-terminus, but this does not produce authentic IGF-1 (Belagaje et al. Protein Sci. 6:1953-19623 (1997)).

**[0008]** An additional problem that is often overlooked when designing IGF fusion constructs is that the fusion partner is essentially a waste product. Prokaryotes can express only a finite amount of recombinant protein - typically up to 40%

of the dry mass of the cell. If the fusion partner is a large protein relative to the IGF, then the effective amount of IGF is reduced. For example, dsbA-IGF fusion proteins are approximately 31,500 daltons, of which the IGF is approximately 7,500 daltons, or approximately 24 percent of the total fusion protein. See U.S. Patent No. 5,629,172. Thus, the efficiency of IGF protein production can be enhanced by using a proportionally smaller fusion partner.

**[0009]** Identifying an effective fusion partner is a difficult task that typically involves significant trial and error. Therefore, improved constructs with more effective fusion partners are required to enhance the production of recombinant IGF.

## SUMMARY OF THE INVENTION

**[0010]** According to certain aspects of the invention, therefore, there are provided nucleic acid sequences, expression systems, host cells, polypeptides, and methods of producing recombinant insulin-like growth factor (IGF).

**[0011]** In one embodiment, the present invention provides for a nucleic acid sequence encoding the polypeptide sequence:

$$MQIFVKTLTGK[X^1]_{0\text{-}30}L\ E[X^2]LFQ\ [IGF]$$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I. In other embodiments, the present invention provides a nucleic acid encoding the above polypeptide, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (*i.e*. zero) amino acid.

**[0012]** In another embodiment, the present invention provides for a protein expression vector encoding the polypeptide sequence:

$$MQIFVKTLTGK[X^1]_{0\text{-}30}L\ E[X^2]LFQ\ [IGF]$$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I. In other embodiments, the present invention provides a protein expression vector encoding the above polypeptide, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (*i.e*. zero) amino acid.

**[0013]** In another embodiment, the present invention provides for a protein of the sequence:

$$MQIFVKTLTGK[X^1]_{0\text{-}30}L\ E[X^2]LFQ\ [IGF]$$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I. In other embodiments, the present invention provides a protein of the above sequence, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (*i.e*. zero) amino acid.

**[0014]** In yet another embodiment, the present invention provides recombinant host cells expressing a protein of the sequence:

$$MQIFVKTLTGK[X^1]_{0\text{-}30}L\ E[X^2]LFQ\ [IGF]$$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I. In other embodiments, the present invention provides a protein of the above sequence, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (i. e. zero) amino acid. In one embodiment, the host cells are prokaryotic. In another embodiment, the host cells are eukaryotic.

**[0015]** In another embodiment, the present invention provides for a method of expressing IGF, comprising transfecting a host cell with the expression vector described above, culturing those cells under conditions which permit protein expression and isolating the protein from the cells, supernatant or broth. In one embodiment, the expression vector encodes a protein of the above sequence, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (*i.e*. zero) amino acid. In one embodiment, the host cells are prokaryotic, while in another, the host cells are eukaryotic.

**[0016]** The present invention also provides a method of preparing N-terminal IGF, wherein a cleavage mixture is prepared containing a chaotropic agent, 3C protease and a polypeptide of the sequence:

$$MQIFVKTLTGK[X^1]_{0\text{-}30}L\ E[X^2]LFQ\ [IGF]$$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I. In one aspect, the method employs a cleavage mixture containing a protein of the above sequence, wherein $X^1$ is TITLEV, TITLEVESSDTIDNVKSKIQDKEGIPPDQQ, or represents no (*i.e*. zero) amino acid. The cleavage mixture further contains a reducing agent, such as dithiothreitol, and can be in solution/liquid form.

**[0017]** Other objects, features and advantages of the present invention will become apparent from the following detailed description. Further, the examples demonstrate the principle of the invention and cannot be expected to specifically illustrate the application of this invention to all the examples where it will be obviously useful to those skilled in the prior art.

## DESCRIPTION OF THE DRAWINGS

**[0018]** Figure 1 provides the nucleic acid sequence (SEQ ID NO: 13) and the amino acid sequence (SEQ ID NO: 21) of human IGF.

**[0019]** Figure 2 provides the amino acid sequences of seven IGF-I fusion constructs described herein (SEQ ID NO: 14 through SEQ ID NO: 20).

**[0020]** Figure 3 sets forth a comparison of mass, Expression disposition, expression yield, and protease cleavage of the seven IGF-I fusion constructs described herein.

**[0021]** Figure 4 depicts the soluble and insoluble properties of the full length Ubiquitin-3C-IGF and TR41 Ubi-3C-IGF proteins. A SDS-PAGE gel was loaded with the following: (M) molecular weight standards; (1) Ubiquitin-IGF standard; (2) empty lane; (3) full length ubiquitin-3C-IGF total cell extract; (4) full length ubiquitin-3C-IGF soluble cell extract; (5) full length ubiquitin-3C-IGF insoluble cell extract; (6) empty lane; (7) TR41 Ubi-3C-IGF total cell extract; (8) TR41 Ubi-3C-IGF soluble cell extract; and (9) TR41 Ubi-3C-IGF insoluble cell extract.

**[0022]** Figure 5 depicts the soluble and insoluble properties of the TR11 Ubi-3C-IGF and TR17 Ubi-3C-IGF proteins. A SDS-PAGE gel was loaded with the following: (M) molecular weight standards; (1) empty lane; (2) T411 Ubi-3C-IGF total cell extract; (3) T411 Ubi-3C-IGF soluble cell extract; (4) TR11 Ubi-3C-IGF insoluble cell extract; (5) empty lane; (6) empty lane; (7) TR17 Ubi-3C-IGF total cell extract; (8) TR17 Ubi-3C-IGF soluble cell extract ; and (9) TR11 Ubi-3C-IGF insoluble cell extract.

**[0023]** Figure 6 depicts the relative molecular weights of dsba-3C-IGF, full length Ubiquitin-3C-IGF, and TR41 Ubi-3C-IGF, and the cleavage products, such as mature IGF of these proteins. A SDS-PAGE gel was loaded with the following: (M) molecular weight standards; (1) IGF Standard; (2) IGF Standard; (3) dsbA-3C-IGF cell extract; (4) dsbA-3C-IGF cell extract- 3C cleaved; (5) dsbA-3C-IGF cell extract 3C cleaved; (6) dsbA-3C-IGF cell extract - 3C cleaved; (7) full length ubiuitin-3C-IGF cell extract; (8) full length ubiuitin-3C-IGF cell extract - 3C cleaved; (9) full length ubiuitin-3C-IGF cell extract - 3C cleaved; (10) full length ubiuitin-3C-IGF cell extract - 3C cleaved; (11) TR41 Ubi-3C-IGF cell extract; (12) TR41 Ubi-3C-IGF cell extract - 3C cleaved; (13) TR41 Ubi-3C-IGF cell extract -3C cleaved; and (14) TR41 Ubi-3C-IGF cell extract- 3C cleaved.

## DETAILED DESCRIPTION

**[0024]** DNA constructs for making IGF are provided. The inventive constructs enhance production of IGF by dramatically improving the yield of expression systems employing fusion constructs comprising ubiquitin-derived peptides, a 3C protease cleavage site, and a N-terminal IGF.

**[0025]** In one embodiment, a short peptide sequence, composed of approximately 11 consecutive amino acids from the protein ubiquitin is used to provide high-level expression of IGF in *E. coli.* In another embodiment, a protease cleavage site from the 3C protein can be incorporated into IGF fusion constructs. In other embodiments, small proteins having a C-terminus with the sequence E-V/T-L-F-Q can be fused in-frame with IGF. As the N-terminus of IGF begins with the amino acids G-P, the resulting fusion protein contains the protease cleavage site E-V/T-L-F-Q//G-P. The 3C protease recognizes this cleavage site (denoted here by "//") and releases IGF from the construct. Thus, in one aspect of the invention, there is provided a prokaryotic expression system for efficiently producing human IGF. Moreover, the inventive constructs and methods dramatically improve the yield of recombinant IGF.

### Definitions

**[0026]** Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. Likewise, the terminology used herein describes particular embodiments only, and is not intended to limit the scope of the invention.

**[0027]** Nucleic acid code and amino acid code may be expressed in single-letter designations as described by IUB Nomenclature Committee, Eur. J. Biochem. 150:1-5 (1985). Similarly, genes and proteins, such as restriction endonucleases, may be expressed in abbreviated forms that are well recognized in the art (Kotyk A., ed., Quantities, Symbols, Units, and Abbreviations in the Life Sciences: A Guide for Authors and Editors, Totowa, N.J., Humana Press (1999)).

**[0028]** As used herein, "N-terminal IGF" refers to a polypeptide encoding IGF-I, particularly human IGF-I, in which the first N-terminal two amino acids are Glycine (G) and Proline (P) respectively. N-terminal IGF includes, but is not limited to, full length IGF, C-terminal truncated forms of IGF, chimeras of N-terminal IGF and point mutants of IGF which do

not modify the N-terminal amino acids G and P. Particular examples of N-terminal IGF are shown in the Figures alone or as members of complete fusion constructs. By way of further example, "N-terminal IGF" does not embrace a naturally occurring truncated form of IGF-I lacking the first three N-terminal amino acids (hereinafter "des(1-3)-IGF-I"). See Heding et al., J. Biol. Chem., 271:13948-13952 (1996).

**[0029]** As used herein, "GP-IGF" refers to an IGF that has been cleaved from a fusion construct to yield a polypeptide with the proper amino terminal sequence of human IGF.

**[0030]** As used herein, "mature IGF" refers to a GP-IGF that has been refolded and optionally isolated such that it exhibits a useful biological property. The art recognizes utility in IGF molecules that stimulate cell or anabolic growth, (via agonist activity), IGF molecules that antagonize cell or anabolic growth (via antagonist activity or by displacing agonist IGF [so called "null IGF"]), and IGF molecules that bind to one or more Insulin Like Growth Factor Binding Proteins.

**[0031]** As used herein, "expression systems" refers to a single nucleic acid or a plurality of nucleic acids that are used to direct expression of a target nucleic acid or the production and/ or assembly of a polypeptide. Expression systems of the present invention are not particularly limiting in their form and may be either plasmid based or may direct chromosomal integration of the target sequence. See Sambrook et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory Press (1989); United States Patent 5,470,727.

**[0032]** As used herein, "host cells," in one embodiment, refer to bacteria, particularly *E. coli* host strains including K strains and B strains and their derivatives including protease deficient strains. In another embodiment, the term refers to eukaryotic cells, such as yeast or tissue culture cell line which have been immortalized for propagation in the laboratory. In yet another embodiment, the term refers to cells derived from animal, plant, fungal, or insect tissue.

**[0033]** As used herein, "3C protease" refers to a protease that recognizes the amino acid sequence E-(V or T)-L-F-Q-G-P. The protease cleaves the sequence N-terminal to the glycine. The designation "3C protease" is derived from the human rhinovirus 14 protease and known to those skilled in art as 3C. See Stanway et al., Nucleic Acids Res. 12: 7859-75 (1984).

**[0034]** The following examples serve to illustrate various embodiments of the present invention and should not be construed, in any way, to limit the scope of the invention.

## Examples

### Example 1. Production of Ubiq-3C-IGF-1 expression vectors

**[0035]** The human IGF-1 gene containing a 5' 3C protease cleavage site (hereinafter 3C-IGF-1) has been cloned into three expression vectors.

#### A. Cloning and verification of expression vectors

**[0036]** Human IGF-1 (Celtrix Pharmaceuticals, Inc., Glen Allen, VA), whose nucleotide and amino acids sequences are set forth in Figure 1, was cloned with an oligonucleotide encoding the six amino acids of the 3C protease cleavage site plus methionine into vectors pET29a and pPop. This fusion strategy represents the smallest possible fusion partner for IGF-1 and was used to investigate protein accumulation and degradation.

**[0037]** 3C-IGF-1 was cloned into pUC19 downstream of the LacZ gene (23 amino acids) to create a second small fusion gene.

**[0038]** In each instance the IGF gene was amplified from the *E. coli* codon-optimised IGF sequence present in plasmid p10723 (Celtrix Pharmaceuticals, Inc., Glen Allen, VA) using PCR. pPop and pET29a - PCR primers were designed to incorporate NdeI and BamHI sites at the 5' and 3' termini respectively to facilitate cloning into vector. The 3C protease cleavage site was incorporated immediately downstream of the 5' NdeI site and in-frame with the IGF sequence. pUC19 - PCR primers were designed to incorporate SacI and EcoRI sites at the 5' and 3' termini respectively to facilitate cloning into vector. The 3C protease cleavage site was incorporated downstream of the SacI site. The PCR primers are described in the following table:

Table 1. PCR primers used to clone 3C-IGF into each of the expression vectors.

| Target | SEQ ID | Primer |
|---|---|---|
| pPop and | 1 | CATATGCTGGAAGTTCTGTTCCAGGGTCCGGAAACCCTG |
| pET29a | 2 | CCGGGATCCTTAAGCGGATTTAGCCG |
| PUC19 | 3 | GAGCTCGCTGGAAGTTCTGTTCCAA |
| | 4 | GAATTCTTAAGCGGATTTAGC |

**[0039]** For each 3C-IGF construct, the IGF gene was amplified using the Expand High Fidelity PCR system (Roche Diagnostics Corporation, Indianapolis, IN) according to the manufacturers' instructions. In a typical PCR reaction, 2μls of plasmid DNA were mixed with 5μls of 10X Expand Buffer containing $MgCl_2$, 1μl of each primer at 1μg/μl, 2μls of 10mM nucleotide mix and sterile deionised water to a final volume of 50μls.

**[0040]** The PCR cycle was performed in a thermocycler (Eppendorf Mastercycler, Brinkmann Instruments, Inc.) and included denaturation, annealing and polymerization. The same cycle was used for each PCR. For the first stage of PCR, the template DNA was denatured by heating at 94˚C for 2 minutes. Expand High Fidelity Taq polymerase was added to the reaction mix. In the second stage, the reaction was denatured at 94˚C for 1 minute, cooled to 60˚C for 30 seconds for primer annealing, and heated to 72˚C for 1 minute to enable DNA elongation. Thirty-five subsequent cycles were performed. Following the 35 cycles, a final 10 minute heating step at 72˚C was performed, before cooling the samples to 4˚C.

**[0041]** The resulting 3C-IGF PCR product was analyzed by agarose gel electrophoresis. Once satisfied that the PCR product was of the correct size of approximately 230bp, it was sub-cloned into vector pCR2.1 (Invitrogen Corporation, Carlsbad, CA) using the TOPOTA cloning kit (Invitrogen Corporation, Carlsbad, CA) according to the manufacturers' instructions. Following ligation, pCR2.1/3C-IGF-1 was transformer into *E. coli* Top10F' cells (Invitrogen Corporation, Carlsbad, CA). Transformants were selected on LB agar containing kanamycin at 50μg/ml.

**[0042]** Putative 3C-IGF-1 PCR clones were analyzed initially using whole cell lysis preparations to detect the presence of plasmid DNA and its approximate size. Restriction enzyme digestion was performed on the clones to excise the 3C-IGF-1 gene to confirm its presence. NdeI and BamHI were used to excise 3C-IGF-1 from sub-clones intended for ligation into pPop. Typically, 18μls of plasmid DNA was digested with 1μl of each enzyme in the presence of 2μls of 10X enzyme specific buffer. The digests were left at 37˚C in a waterbath for 4 hours and electrophoresed on an agarose gel.

**[0043]** The 3C-IGF-1 harboring clones identified above were designated Top10F'::pCR2.1/3C-IGF-1, and Top10F'::pCR2.1/3C-IGF-1(LacZ) and were used to source the 3C-IGF-1 for ligation into pPop, pET29a and pUC19.

B. Ligation of 3C-IGF-1 into pPop and pET29a

**[0044]** Plasmid DNA was isolated from 5 ml overnight cultures of Top10F'::pCR2.1/3C-IGF-1, MSD3363 pPop and XL1-Blue MR::pET29a using the QIAGEN plasmid mini-prep kit (QIAGEN Inc., Valencia, CA) according to the manufacturers' instructions. Vectors pPop and pET29a were prepared for ligation to 3C-IGF by digestion with BamHI and NdeI. Similarly, the 3C-IGF gene was excised directly from pCR2.1/3C-IGF-1 by digestion with BamHI and NdeI. The 3C-IGF-1 gene was subsequently ligated overnight to phosphatased pPop and pET29a at 17C using T4 DNA ligase. The resulting constructs, pPop/3C-IGF-1 and pET29a/3C-IGF-1, were transformed into *E. coli* XL1-Blue MR strain (Stratagene, La Jolla, CA) according to the manufacturers' instructions. Transformants were selected using tetracycline resistance for pPop/3C-IGF-1(10μg/ml), and kanamycin (50μg/ml) for pET29a/3C-IGF-1.

**[0045]** Transformants were grid plated onto LB agar containing tetracycline (10μg/ml) for pPop/3C-IGF-1 and kanamycin (50μg/ml) for pET29a/3C-IGF, and incubated at 37˚C overnight. Clones were analyzed using whole cell lysis preparations and by restriction enzyme digestion with NdeI and BamHI. The resulting constructs which were shown to possess 3C-IGF-1 were designated XL1-Blue MR::pPop/3C-IGF-1 and XL1-Blue MR::pET29a/3C-IGF-1.

C. Ligation of 3C-IGF-1 into pUC19

**[0046]** Plasmid DNA was isolated from a 5 ml overnight cultures of Top10F'::pCR2.1/3C-IGF-1 (LacZ), using the QIAGEN plasmid mini-prep kit (QIAGEN Inc., Valencia, CA) according to the manufacturers' instructions. Vector pUC19 was prepared for ligation to 3C-IGF-1 by digestion with EcoRI and SacI. Similarly, the 3C-IGF-1 gene was excised directly from pCR2.1/3C-IGF-1 (LacZ) by digestion with EcoRI and SacI. The 3C-IGF-1 gene was subsequently ligated overnight to phosphatased pUC19 at 17˚C using T4 DNA ligase. The resulting construct, pUC19/3C-IGF-1, was transformed into *E. coli* XL1-Blue MR strain (Stratagene) according to the manufacturers' instructions. Transformants were selected using ampicillin resistance (100μg/ml).

**[0047]** Transformants were grid plated onto LB agar containing ampicillin (100μg/ml) and incubated at 37˚C overnight. Clones were analysed using whole cell lysis preparations and by restriction enzyme digestion with SacI and EcoRI. The resulting construct which was shown to possess 3C-IGF-1 was designated XL1-Blue MR::pUC19/3C-IGF-1.

D. Electroporation of pPop/3C-IGF-1 and pET29a/3C-IGF-1 into *E. coli* host strains

**[0048]** 1μl each of pPop/3C-IGF-1 and pET29a/3C-IGF-1 from each of the XL1-Blue MR strains harbouring the expression vectors, were electroporated into 5 *E. coli* DE3 electrocompetent host strains - BL21, MSD68, MSD2252, UT5600 and MSD2254pLysS.

**[0049]** The hosts were electroporated using a Bio-Rad Gene Pulser, according to the manufacturers' instructions

(parameters included: 25μF, 200Ω, 2.5kV). Transformants harbouring pPop/3C-IGF-1 were selected on LB agar containing 10μg/ml tetracycline. Transformants harbouring pET29a/3C-IGF-1, were selected on LB agar containing 50μg/ml kanamycin. The transformants were grid plated out onto LB agar containing appropriate antibiotic and stored at 4˚C. Glycerol stocks were made of each strain and stored at -70˚C.

E. Expression of 3C-IGF-1 from pPop

**[0050]** Initially, lab-scale shake flasks were used to investigate if 3C-TGF-1 could be expressed from each of the *E. coli* host strains electroporated with pPop/3C-IGF-1, without being degraded by intracellular proteases (due to the small size of the molecule).

**[0051]** 5mls of LB broth containing 10μg/ml tetracycline was inoculated with the *E. coli* host strains harbouring pPop/3C-IGF-1. Cells were grown overnight at 37˚C in an orbital incubator with good aeration (200 rpm). 0.5mls of each overnight culture was then used to inoculate 50mls of LB broth containing the 10μg/ml tetracycline. Cultures were incubated at 37˚C until an OD600 of approximately 0.4 - 0.6 was obtained. The cultures were induced at this point with 1mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 2.5 hours.

**[0052]** Following induction, both pre- and post-induction samples were electrophoresed to obtain equivalent loadings, on a 4-12% Bis-Tris NuPAGE gel under reduced denatured conditions, according to the manufacturers' instructions

**[0053]** No expression of 3C-IGF-1 from any host cell (BL21, MSD68, MSD2252, UT5600 and MSD2254pLysS) was visible on the gel. The results suggest that (a) there was no expression of the protein or (b) the fusion protein was expressed but was degraded by intracellular proteases due to the small size of the fusion. The strength of the promoter/induction makes it highly unlikely that the protein was not expressed, suggesting that the protein was degraded.

F. Ubiquitin-3C-IGF fusion gene construction

**[0054]** The SOE-PCR technique (Warrens et al., Gene 186:29-35 (1997)) was used to clone an existing ubiquitin IGF-I construct (pER10088 - Celtrix Pharmaceuticals, Inc., Glen Allen, VA) into the pPOP expression vector and simultaneously introduce the 3C cleavage sequence between the ubiquitin and the IGF genes. The amino acid sequences of the fusion proteins is shown in Figure 2

**[0055]** Four plasmids were constructed. Full length ubiquitin (FL) had exactly the same ubiquitin fusion partner as found in pER10088, but with the 6 amino acid 3C recognition site inserted between ubiquitin and the N- terminal of mature IGF-1.

**[0056]** TR41 (truncated at 41) consisted of the first 41 amino acids of ubiquitin followed by the 3C recognition site and IGF-1

**[0057]** T411 (truncated at 41) consisted of the first 11 amino acids of ubiquitin followed by the 3C recognition site and IGF-1; and

**[0058]** TR17 (truncated at 41) consisted of the first 11 amino acids of ubiquitin followed by the 3C recognition site and IGF-1.

**[0059]** Primers were designed to enable cloning into Nde I/Bam HI cut pPop or pET vectors, and purchased from a commercial synthesizer.

**[0060]** Using the primers identified in Table 2, UQ1 and UQ2 produce a full length ubiquitin with a 3C IGF overhang. UQ1 and UQ5 produce a truncated (41 aa) ubiquitin with a 3C IGF overhang. UQ3 and UQ4 produce a 3C IGF with a full length ubiquitin overhang. UQ6 and UQ4 produce a 3C IGF with a truncated ubiquitin overhang. UQ13 and UQ4 produce a truncated (11 aa) ubiquitin with a 3C-IGF overhang. UQ16 and UQ4 produce a truncated (11 aa) ubiquitin with a 3C-IGF overhang.

Table 2. SOE-PCR primers used to clone 3C-IGF into each of the expression vectors.

| Primer | SEQ ID | Sequence |
|---|---|---|
| UQ1 | 5 | CATATGCAGATTTTCGTCAAGACTTTGACC |
| UQ4 | 6 | GATCCTTAAGCGGATTTAGCCGGTTTCAG |
| UQ2FL | 7 | GGTTTCCGGACCCTGGAACAGAACTTCCAGACCACCGCGGAG |
| UQ3FL | 8 | CTCCGCGGTGGTCTGGAAGTTCTGTTCCAGGGTCCGGAAACC |
| UQ5TR | 9 | GGTTTCCGGACCCTGGAACAGAACTTCCAGTTGTTGATCTGG |
| UQ6TR | 10 | CCAGATCAACAACTGGAAGTTCTGTTCCAGGGTCCGGAAACC |
| UQ13 | 11 | CATATGCAGATTTTCGTCAAGACTTTGACCGGTAAACTGGAAGTTCTGTTC |

(continued)

| Primer | SEQ ID | Sequence |
|---|---|---|
| UQ16 | 12 | CATATGCAGATTTTCGTCAAGACTTTGACCGGTAAAACCATA ACATTGGAAGTTCTGGAAGTTCTGTTC |

**[0061]** Using the above primer pairs, PCR was carried out using the Expand High Fidelity PCR system according to the manufacturers' instructions. In each PCR reaction, 2 μls of plasmid DNA (pER10088 or p10724) was mixed with 5μls of 10X Expand Buffer containing $MgCl_2$, 1μl of each primer at 1μg/μl, 2μls of 10mM nucleotide mix and sterile deionised water to a final volume of 50μls.

**[0062]** The PCR cycle was performed in a thermocycler and included denaturation, annealing and polymerization. The same cycle was used for each PCR. For the first stage of PCR, the template DNA was denatured by heating at 94˚C for 2 minutes. Expand High Fidelity Taq polymerase was then added. In the second stage, the reaction was denatured at 94˚C for 1 minute, cooled to 68˚C for 45 seconds for primer annealing, and then heated to 72˚C for 1 minute to enable DNA elongation. 33 subsequent cycles were performed. Next, a final 10 minute heating step at 72˚C was performed, then the samples were cooled to 4˚C (temperature hold).

**[0063]** The PCR products were analyzed by agarose gel electrophoresis. The bands were excised from the gel and the DNA extracted using a commercial extraction kit (QIAGEN Inc., Valencia, CA), and the DNA eluted in 30 μls of buffer.

**[0064]** A second round of PCR was carried out using these PCR products as templates and UQ1 and UQ 4 as primers. The PCR reactions were identical to those used previously except that 1μl of DNA was used and the annealing temperature was 65˚C.

**[0065]** The PCR products were analysed by agarose gel electrophoresis. The bands were excised from the gel and the DNA extracted using a commercial extraction kit, and the DNA eluted in 30 μls of buffer.

**[0066]** 3 μls of purified PCR product was sub-cloned into vector pCR2.1 using the TOPOTA cloning kit according to the manufacturers' instructions. Following ligation, the plasmids were transformed into *E. coli* Top 10F' cells. Transformants were selected for on LB agar containing kanamycin at 50μg/ml.

**[0067]** PCR clones were analyzed initially using whole cell lysis preps to detect the presence of plasmid DNA and its approximate size. Restriction enzyme digestion was performed on the clones to excise the 3C-IGF-1 gene to confirm its presence.

**[0068]** NdeI and BamHI were used to excise 3C-IGF-1 from sub-clones intended for ligation into pPop. Typically, 18 μls of plasmid DNA was digested with 1 μl of each enzyme in the presence of 2 μls of 10X enzyme specific buffer. The digests were left at 37˚C in a waterbath for 4 hours and electrophoresed on an agarose gel.

**[0069]** Vector pPop was prepared for ligation by digestion with BamHI and NdeI. The 3C-IGF1 fusions genes were subsequently ligated overnight to phosphatased pPop at 17˚C using T4 DNA ligase. The resulting constructs were transformed into *E. coli* XL1-Blue MR strain according to the manufacturers' instructions. Transformants were selected using tetracycline resistance (10 μg/ml).

**[0070]** Transformants were grid plated onto LB agar containing tetracycline (10μg/ml) and incubated at 37˚C overnight. Clones were analysed using whole cell lysis preparations and by restriction enzyme digestion with NdeI and BamHI.

Example 2. Creation of *E. coli* stock by electroporation

**[0071]** In this example, *E. coli* stock strains harbouring the expression vectors were created by electroporation.

**[0072]** The electrocompetent versions of *E. coli* W3110 and BL21 were electroporated with 1μl of plasmid DNA isolated from each of the XL1-Blue MR strains harbouring the expression vectors. The hosts were electroporated using a Bio-Rad Gene Pulser, according to the manufacturers' instructions (parameters included: 25μF, 200Ω, 2.5kV). Transformants were selected on LB agar containing 10μg/ml tetracycline. The transformants were grid plated out onto LB agar containing appropriate antibiotic and stored at 4˚C. Glycerol stocks were made of each strain and stored at -70˚C.

Example 3. Expression of IGF1

**[0073]** Lab-scale shake flasks were used to investigate expression at 37˚C in various host strains.

A. Growth of *E. coli* stock strains harbouring the expression vectors

**[0074]** 5mls of LB broth containing the appropriate antibiotic were inoculated with the host strains containing the expression vectors. Cells were grown overnight at 37˚C in an orbital incubator with good aeration. 0.5mls of each

overnight was used to inoculate 50mls of LB broth containing the appropriate antibiotic. Cultures were incubated at 37˚C until an $OD_{600}$ of approximately 0.4 - 0.6 was obtained. The cultures were induced at this point with 1mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 2.5 hours.

**[0075]** Following induction, both pre- and post-induction samples were electrophoresed on a 4-12% Bis-Tris NuPAGE gel under reduced denatured conditions, according to the manufacturers' instructions. 1.5 mls of each pre- and post-induction sample were harvested in a benchtop microcentrifuge at 13,000 rpm for 2 minutes. The supernatants were decanted and the pellets re-suspended in deionised water. The amount of water in which each pellet was re-suspended varied and was calculated so as to give suspensions of equivalent optical density so that comparisons between tracks on the gels were meaningful. The loadings on the gels are indicated in brackets after the sample lane. Each gel was quantified using densitometry.

B. Fractionation of soluble and insoluble protein

**[0076]** Cell pellets were harvested by centrifugation at approximately 2,500 g for 10 minutes and re-suspended in one forth of the original volume of 10 mM PBS pH 7.4. The cell suspension was sonicated using a Heat Systems XL2020 sonicator for 7 x 30 seconds, with 30 seconds cooling between.

**[0077]** The sonicate was harvested by centrifugation at 12,000 rpm (4˚C) for 15 minutes in a Sorvall RC-5B centrifuge. The supernatant (soluble fraction) was decanted and stored on ice. The pellet (insoluble fraction) was re-suspended in 10 mls of sterile deionised water. Samples of each were analyzed on SDS-PAGE gels.

C. Cleavage of the ubiguitin-IGF polypeptides with 3C protease

**[0078]** The different Ubi-3C-IGF-1 fusions constructed as discussed above and Dsb-3C-IGF fusion (p10723 - Celtrix Pharmaceuticals, Inc., Glen Allen, VA), were treated with three different sources of 3C protease.

**[0079]** Shake-flasks preparations were made from two different Dsb3C protease-containing strains. One strain contained a chromosomally integrated 3C protease gene and the other strain carried the 3C protease on a plasmid (Zhang et al., Protein Expr Purif. 12:159-165(1998)). Cell free lysates were prepared using sonication followed by centrifugation. A third source of 3C protease was obtained commercially from Amersham Biosciences under the tradename PRECISSION PROTEASE.

**[0080]** The following were mixed in a tube at incubated at room temperature overnight: 2 mls of cell free extract containing IGF fusions, 0.4 mls of cell free extract containing 3C proteases or 5 μls of Prescission protease, 4.8 μls of 0.5 M DTT, and 2.4 μls of 1 M EDTA. Samples were analyzed by SDS-PAGE.

D. Expression of IGF1

1. The amino terminus of ubiquitin confers high-levels expression to IGF fusion proteins

**[0081]** The constructs designed herein were intended to find minimal IGF fusion constructs incorporating a 3C protease cleavage site that yielded high-levels expression in *E. coli* strains. As shown in Figure 3, Met-3C-IGF and LacZ-3C-IGF did not accumulate to any significant degree upon expression, whereas all ubiquitin-3C-IGF constructs were abundantly expressed. Firgure 4 also demonstrated the relative expression levels of Ubiqutin-IGF, Ubiquitin-3C-IGF, and TR41 Ubi-3C-IGF). Moreover, truncated ubiquitin-3C constructs expressed as-well-as or better than the full length ubiquitin-3C construct. Thus the N-terminal 11 amino acids are sufficient to confer high-level expression of human IGF in *E. coli.*

2. Truncated ubiquitin constructs of IGF increase the yield of IGF protein

**[0082]** As shown in Figure 3, the truncated ubiquitin constructs provided higher yields of IGF. This is because (1) the relative percentage of IGF is increased as compared to the total fusion protein and (2) because the clones are more efficient at protein expression than the parent ubiquitin-IGF construct. Moreover, because they are exclusively isolated as insoluble proteins, these constructs effectively increase the yield of IGF by reducing the amount of down-stream processing. As is well known in the art, each processing step has a loss associated with it and a corresponding cost associated with the processing. Having the protein exclusively located in one form eliminates the need to process the soluble protein from the cell extract through extra precipitation steps.

3. Truncated ubiquitin constructs do not possess ubiquitin-solubilizing properties

**[0083]** Figures 4 demonstrates that full-length ubiquitin-3C-IGF is relatively soluble with approximately 50% of the protein in the soluble cell extract and approximately 50% of the protein isolated as insoluble form. Conversely, Figure

5 shows that TR41 Ubi-3C-IGF and TR11 Ubi-3C-IGF were isolated exclusively in the insoluble form. This allows advantageous isolation of the protein, since the disposition of the protein is exclusively in one form in the cell extract. All three truncated proteins are relatively insoluble and require high concentrations (>4 M Urea) of chaotrope in order to solublize them. In particular, TR11 Ubi-3C-IGF and TR17 Ubi-3C-IGF were more insoluble than TR41 Ubi-3C-IGF, both requiring very high levels (>8M urea) of chaotrope to substantially solubilize the proteins.

4. Truncated ubiquitin constructs of IGF are cleaved by 3C protease to yield mature IGF with the correct N-terminus

**[0084]** Figure 6 demonstrates dsbA-3C-IGF, full length ubiquitin-3C-IGF, and TR41 Ubi-3C-IGF cleavage by three different sources of 3C protease (as described in Example 3C, above). In each case, the IGF is liberated as seen by comparison of the band corresponding with the same molecular weight of the IGF standard. Note that the cleavage of dsbA-3C-IGF and TR41 Ubi-3C-IGF is essentially quantitative, while the full length ubiquitin-3C-IGF has some uncleaved product. The experiment was not optimized for cleavage conditions. Thus, the lack of quantitative cleavage does not negatively reflect on the quality of the construct. In a single experiment of TR11 Ubi-3C-IGF and TR17 Ubi-3C-IGF, these constructs were not cleaved by 3C protease at all. The inventors determined, however, that the amount of chaotrope in the reaction inactivated the enzyme. It is likely that these constructs would be cleaved by the 3C protease if the basal solution was sufficiently free of interfering chaotrope.

**[0085]** N-terminal amino acid sequencing was conducted on the IGF isolated from 3C protease cleaved TR 41 Ubi-3C-IGF and the mature IGF demonstrated the correct N-terminus sequence for human IGF.

**[0086]** The invention has been disclosed broadly and illustrated in reference to representative embodiments described above.

## Claims

1. An isolated nucleic acid sequence encoding a polypeptide comprising the following sequence:

   $MQIFVKTLTGK[X^1]_{0\text{-}30}LE[X^2]LFQ[IGF]$

   wherein $X^1$ is a peptide having a sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I.

2. The nucleic acid of claim 1, wherein $X^1$ is selected from the group consisting of (A) zero amino acids, (B) six amino acids of the sequence TITLEV and (C) 30 amino acids of the sequence TTTLEVESSDTIDNVKSKIQDKEGIPPDQQ.

3. A protein expression system, comprising a protein expression vector encoding a protein:

   $MQIFVKTLTGK[X^1]_{0\text{-}30}LE[X^2]LFQ[IGF]$

   wherein $X^1$ is a peptide having from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I.

4. The protein expression system of claim 3, wherein $X^1$ is selected from the group consisting of (A) zero amino acids, (B) six amino acids of the sequence TITLEV and (C) 30 amino acids of the sequence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

5. A protein expression system, comprising
   a host cell and
   a protein expression vector encoding a polypeptide:

   $MQIFVKTLTGK[X^1]_{0\text{-}30}LE[X^2]LFQ[IGF]$

   wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I.

6. The protein expression system of claim 5, wherein $X^1$ is selected from the group consisting of (A) zero amino acids, (B) six amino acids of the sequence TITLEV and (C) 30 amino acids of the sequence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

7. The protein expression system of claim 5, wherein the host cell is prokaryotic.

**8.** The protein expression system of claim 5, wherein the host cell is eukaryotic.

**9.** A method of expressing N-terminal IGF comprising:

(A) transfecting a host cell with an expression vector encoding a polypeptide:

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

wherein $X^1$ is a peptide having from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I;
(B) culturing said host cell under conditions that permit expression of said polypeptide; and
(C) isolating the polypeptide.

**10.** The method of claim 9, wherein $X^1$ is selected from the group consisting of (A) zero amino acids, (B) six amino acids of the sequence TITLEV and (C) 30 amino acids of the sequence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**11.** The method of claim 9, wherein the host cell is prokaryotic.

**12.** The method of claim 9, wherein the host cell is eukaryotic.

**13.** A method of preparing recombinant N-terminal IGF, comprising

(A) preparing a cleavage mixture of

(1) about 1 M to about 3 M of a chaotropic agent, wherein said chaotropic agent is either urea or guanidine hydrochloride;
(2) 3C protease; and
(3) a polypeptide comprising the sequence:

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

wherein $X^1$ is a peptide having from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I;
(B) incubating the cleavage mixture at a temperature from about 4 degrees Celsius to about 37 degrees Celsius for a period of time sufficient to allow the protease to cleave the polypeptide; and
(C) isolating the N-terminal IGF.

**14.** The method of claim 13, wherein $X^1$ is selected from the group consisting of (A) zero amino acids, (B) six amino acids of the sequence TITLEV and (C) 30 amino acids of the sequence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**15.** A polypeptide comprising the polypeptide sequence:

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

wherein $X^1$ is a peptide sequence from 0 to 30 amino acids; $X^2$ is either V or T; and IGF is N-terminal IGF-I.

**16.** The polypeptide of claim 15, wherein $X^1$ is selected from the group consisting of (A) 0 amino acids; (B) 6 amino acids of the sequence "TITLEV"; and (C) 30 amino acids of the sequence "TITLEVESSDTIDNVKSKIQDKEGIPPDQQ."

## Patentansprüche

**1.** Eine isolierte Nukleinsäuresequenz, die ein Polypeptid kodiert, das die folgende Sequenz umfasst:

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

wobei $X^1$ ein Peptid ist, das eine Sequenz von 0 bis 30 Aminosäuren besitzt; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist.

**2.** Die Nukleinsäure nach Anspruch 1, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**3.** Ein Proteinexpressionssystem, umfassend ein Proteinexpressionsvektor, der ein Protein kodiert:

$$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$$

wobei $X^1$ ein Peptid ist, das 0 bis 30 Aminosäuren besitzt; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist.

**4.** Das Proteinexpressionssystem nach Anspruch 3, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**5.** Ein Proteinexpressionssystem, umfassend
eine Wirtszelle und
ein Proteinexpressionsvektor, der ein Polypeptid kodiert:

$$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$$

wobei $X^1$ eine Peptidsequenz mit 0 bis 30 Aminosäuren ist; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist.

**6.** Das Proteinexpressionssystem nach Anspruch 5, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**7.** Das Proteinexpressionssystem nach Anspruch 5, wobei die Wirtszelle prokaryotisch ist.

**8.** Das Proteinexpressionssystem nach Anspruch 5, wobei die Wirtszelle eukaryotisch ist.

**9.** Ein Verfahren zur Expression von N-terminalem IGF umfassend:

(A) Transfizieren einer Wirtszelle mit einem Expressionsvektor, der ein Polypeptid kodiert:

$$MQ\ IFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$$

wobei $X^1$ ein Peptid ist, das 0 bis 30 Aminosäuren besitzt; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist;
(B) Kultivieren der Wirtszelle unter Bedingungen, die die Expression des Polypeptids erlauben; und
(C) Isolieren des Polypeptids.

**10.** Das Verfahren nach Anspruch 9, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**11.** Verfahren nach Anspruch 9, wobei die Wirtszelle prokaryotisch ist.

**12.** Verfahren nach Anspruch 9, wobei die Wirtszelle eukaryotisch ist.

**13.** Ein Verfahren zum Herstellen rekombinanten N-terminalen IGFs, umfassend

(A) Herstellung einer Spaltungsmischung aus

(1) ungefähr 1 M bis ungefähr 3 M eines chaotropen Mittels, wobei das chaotrope Mittel entweder Harnstoff oder Guanidinhydrochlorid ist;
(2) 3C Protease; und
(3) ein Polypeptid umfassend die Sequenz:

$$MQIFVKTLTGK[X^1]_{0-3}LE[X^2]LFQ[IGF]$$

wobei $X^1$ ein Peptid ist, das 0 bis 30 Aminosäuren besitzt; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist;
(B) Inkubieren der Spaltungsmischung bei einer Temperatur von ungefähr 4˚C bis ungefähr 37˚C für einen Zeitraum, der ausreichend ist, um der Protease zu erlauben, das Polypeptid zu spalten; und
(C) Isolieren des N-terminalen IGFs.

**14.** Das Verfahren nach Anspruch 13, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**15.** Ein Polypeptid umfassend die Polypeptidsequenz:

MQIFVKTLTGK$[X^1]_{0\text{-}30}$LE$[X^2]$LFQ[IGF]

wobei $X^1$ ein Peptid ist, das 0 bis 30 Aminosäuren besitzt; $X^2$ entweder V oder T ist; und IGF N-terminales IGF-I ist.

**16.** Das Polypeptid nach Anspruch 15, wobei $X^1$ ausgewählt wird aus der Gruppe bestehend aus (A) null Aminosäuren, (B) 6 Aminosäuren der Sequenz TITLEV und (C) 30 Aminosäuren der Sequenz TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**Revendications**

**1.** Séquence d'acide nucléique isolée codant pour un polypeptide, comprenant la séquence suivante :

MQIFVKTLTGK$[X^1]_{0\text{-}30}$LE$[X^2]$LFQ[IGF]

dans laquelle $X^1$ est un peptide ayant une séquence de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal.

**2.** L'acide nucléique de la revendication 1, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence TITLEV et (C) 30 acides aminés correspondant à la séquence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**3.** Système d'expression de protéine, comprenant un vecteur d'expression de protéine codant pour une protéine :

MQIFVKTLTGK$[X^1]_{0\text{-}30}$LE$[X^2]$LFQ[IGF]

dans laquelle $X^1$ est un peptide comportant de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal.

**4.** Le système d'expression de protéine de la revendication 3, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence TITLEV et (C) 30 acides aminés correspondant à la séquence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**5.** Système d'expression de protéine, comprenant :

une cellule hôte et
un vecteur d'expression de protéine codant pour un polypeptide :

MQIFVKTL TGK$[X^1]_{0\text{-}30}$LE$[X^2]$LFQ[IGF]

dans lequel $X^1$ est un peptide ayant une séquence de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal.

**6.** Le système d'expression de protéine de la revendication 5, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence TITLEV et (C) 30 acides aminés correspondant à la séquence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**7.** Le système d'expression de protéine de la revendication 5, dans lequel la cellule hôte est procaryote.

**8.** Le système d'expression de protéine de la revendication 5, dans lequel la cellule hôte est eucaryote.

**9.** Procédé d'expression d'IGF N-terminal, comprenant :

(A) la transfection d'une cellule hôte avec un vecteur d'expression codant pour un polypeptide :

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

dans lequel $X^1$ est un peptide ayant de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal ;
(B) la culture de ladite cellule hôte dans des conditions qui permettent l'expression dudit polypeptide ; et
(C) l'isolement du polypeptide.

**10.** Le procédé de la revendication 9, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence TITLEV et (C) 30 acides aminés correspondant à la séquence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**11.** Le système d'expression de protéine de la revendication 9, dans lequel la cellule hôte est procaryote.

**12.** Le système d'expression de protéine de la revendication 9, dans lequel la cellule hôte est eucaryote.

**13.** Procédé de préparation d'IGF N-terminal recombinant, comprenant

(A) la préparation d'un mélange pour clivage, de :

(1) environ 1 M à environ 3 M d'un agent chaotrope, dans lequel ledit agent chaotrope est soit l'urée soit le chlorhydrate de guanidine ;
(2) la protéase 3C ; et
(3) un polypeptide comprenant la séquence :

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

dans laquelle $X^1$ est un peptide comportant de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal ;
(B) l'incubation du mélange pour clivage, à une température d'environ 4 °C à environ 37 °C pendant une période de temps suffisante pour permettre à la protéase de cliver le polypeptide ;
(C) l'isolement de l'IGF N-terminal.

**14.** Le procédé de la revendication 13, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence TITLEV et (C) 30 acides aminés correspondant à la séquence TITLEVESSDTIDNVKSKIQDKEGIPPDQQ.

**15.** Polypeptide comprenant la séquence polypeptidique :

$MQIFVKTLTGK[X^1]_{0-30}LE[X^2]LFQ[IGF]$

dans laquelle $X^1$ est un peptide ayant de 0 à 30 acides aminés ; $X^2$ représente V ou T ; et IGF représente l'IGF-I N-terminal.

**16.** Polypeptide de la revendication 15, dans lequel $X^1$ est choisi parmi le groupe consistant en (A) zéro acide aminé, (B) six acides aminés correspondant à la séquence "TITLEV" et (C) 30 acides aminés correspondant à la séquence "TITLEVESSDTIDNVKSKIQDKEGIPPDQQ".

**Figure 1**

Nucleic Acid Sequence of IGF-1

GGTCCGGAAACCCTGTGCGGTGCTGAACTGGTTGACGCTCTGCAGTTCGTTTGCGGTGACCGTGGTTTCTACTTCAACAAACCGA
CCGGTTACGGTTCCTCCTCCCGTCGTGCTCCGCAGACCGGTATCGTTGACGAATGCTGCTTCCGGTCCTGCGACCTGCGTCGTCTG
GAAATGTACTGCGCTCCGCTGAAACCGGCTAAATCCGCTTAA

Amino Acid Sequence Of Human IGF-1

GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA*

## Figure 2

**Amino Acid sequence of fusion proteins constructed and tested**

| Clone | SEQ.ID.NO. | Sequence |
|---|---|---|
| Met-3C-IGF | 14 | MLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| LacZ-3C-IGF | 15 | MITPSLHACRFSTLEDPRVPSSNSLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| Ubi-IGF | 16 | MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| UBI-3C-IGF | 17 | MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGGLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| Tr41 Ubi-3C-IGF | 18 | MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| Tr17 Ubi-3C-IGF | 19 | MQIFVKTLTGKTITLEVLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |
| Tr11 Ubi-3C-IGF | 20 | MQIFVKTLTGKLEVLFQ^GPETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRARQTGIVDECCFRSCDLRRLEMYCAPLKPAKSA |

^ = cleavage juncture between fusion protein and IGF

**Figure 3**

| Clone | Total Mass | Fusion Mass | IGF Mass | Disposition | Shake Flask Total expression (g/L) | Shake Flask IGF expression (g/L) | % increase IGF expression | 3C clevage |
|---|---|---|---|---|---|---|---|---|
| Met-3C-IGF | 8534 | 879 | 7655 | 100% insoluble | < 0.005 | nd | | nd |
| LacZ-3C-IGF | 11031 | 3376 | 7655 | 100% insoluble | < 0.005 | nd | | nd |
| Ubi-IGF | 16212 | 8557 | 7655 | 50:50 soluble/insoluble | 0.08 | N/A | | N/A |
| UBI-3C-IGF | 16942 | 9287 | 7655 | 50:50 soluble/insoluble | 0.07 | 0.036 | -4 % | Yes |
| Tr41 Ubi-3C-IGF | 12961 | 5306 | 7655 | 100% insoluble | 0.11 | 0.065 | 72 % | Yes |
| Tr17 Ubi-3C-IGF | 10307 | 2652 | 7655 | 100% insoluble | 0.12 | 0.089 | 136 % | No* |
| Tr11 Ubi-3C-IGF | 9650 | 1995 | 7655 | 100% insoluble | 0.11 | 0.087 | 131 % | No* |

nd = Not determined

n/a = Not applicable

* = Extreme insolubility requires high concentration of chaotrope. The samples were not sufficiently diluted and the 3C protease was inactivated by the chaotrope.

**Figure 4**

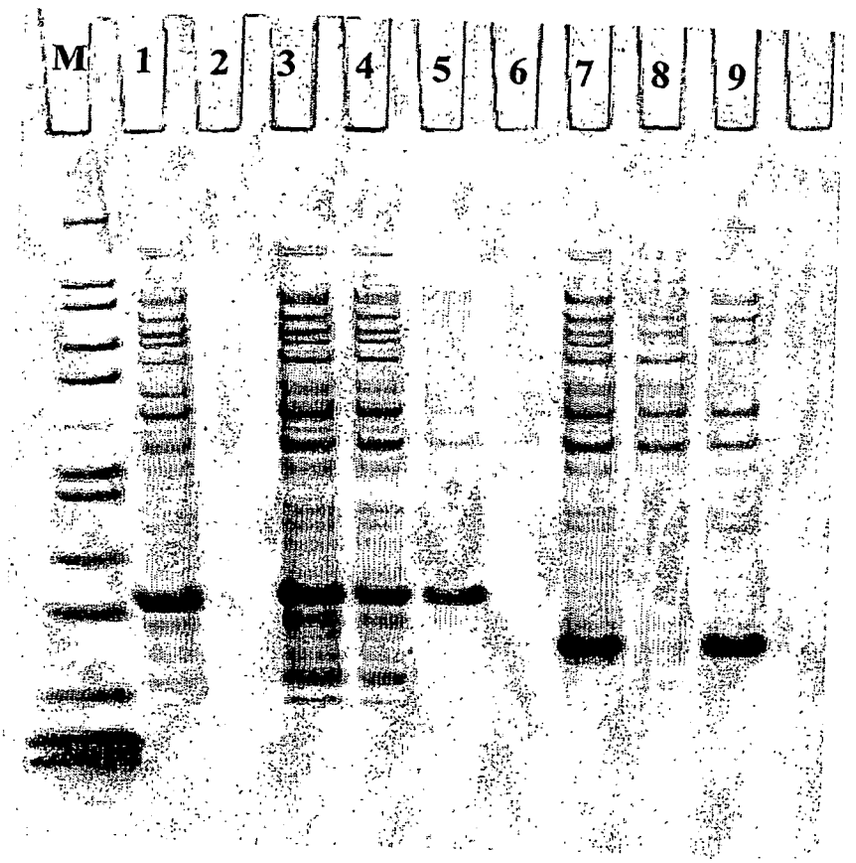


**Figure 5**

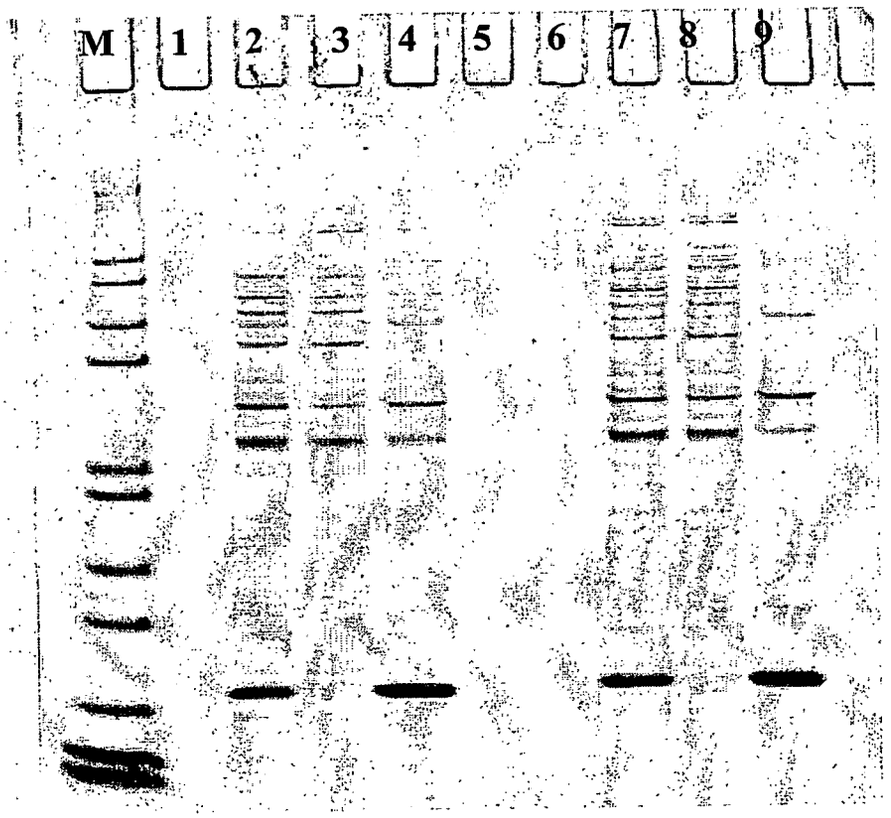
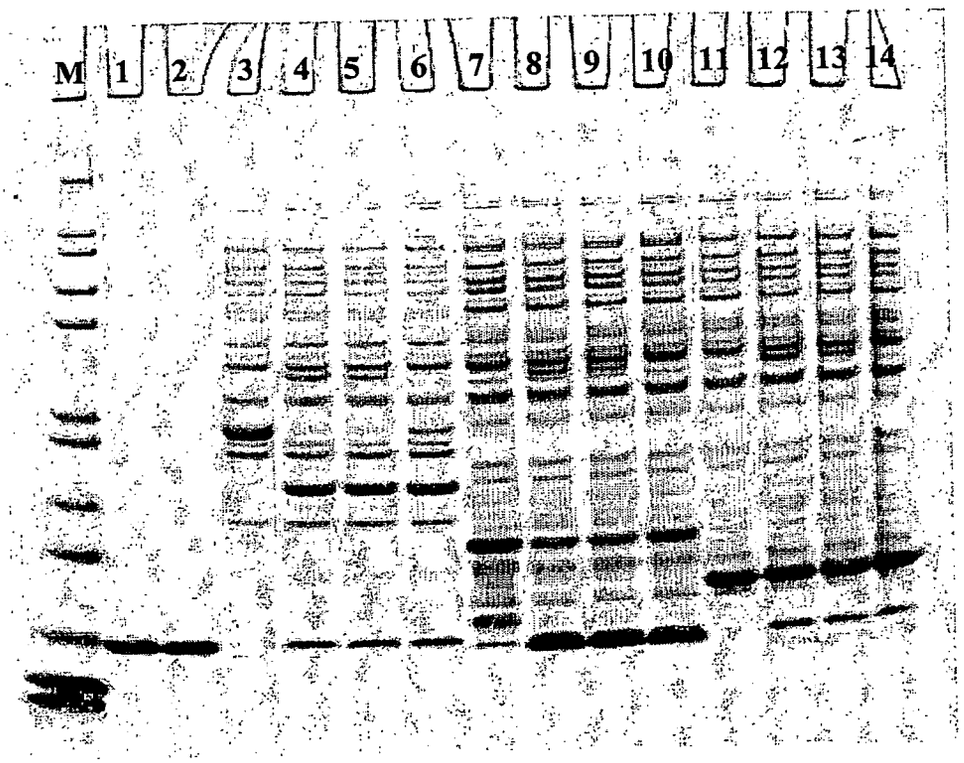


**Figure 6**



## REFERENCES CITED IN THE DESCRIPTION

*This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.*

### Patent documents cited in the description


- US 47794103 P **[0001]**
- WO 9640722 A **[0006]**
- US 5629172 A **[0008]**
- US 5470727 A **[0031]**


### Non-patent literature cited in the description


- **WILLIAMS et al.** *Science,* 1982, vol. 215, 687-688 **[0005]**
- **SCHONER et al.** *Biotechnology,* 1985, vol. 3, 151-154 **[0005]**
- **OLSON et al.** *Protein Expr Purif.,* 1998, vol. 14, 160-166 **[0006]**
- **SCHULZ M.F. et al.** *J. Bact.,* 1987, vol. 169, 585-53921 **[0007]**
- **BUELL et al.** *Nucleic Acids Res.,* 1985, vol. 13, 1923-1938 **[0007]**
- **BELAGAJE et al.** *Protein Sci.,* 1997, vol. 6, 1953-19623 **[0007]**
- *Eur. J. Biochem.,* 1985, vol. 150, 1-5 **[0027]**
- Quantities, Symbols, Units, and Abbreviations in the Life Sciences: A Guide for Authors and Editors. Humana Press, 1999 **[0027]**
- **HEDING et al.** *J. Biol. Chem.,* 1996, vol. 271, 13948-13952 **[0028]**
- **SAMBROOK et al.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1989 **[0031]**
- **STANWAY et al.** *Nucleic Acids Res.,* 1984, vol. 12, 7859-75 **[0033]**
- **WARRENS et al.** *Gene,* 1997, vol. 186, 29-35 **[0054]**
- **ZHANG et al.** *Protein Expr Purif.,* 1998, vol. 12, 159-165 **[0079]**