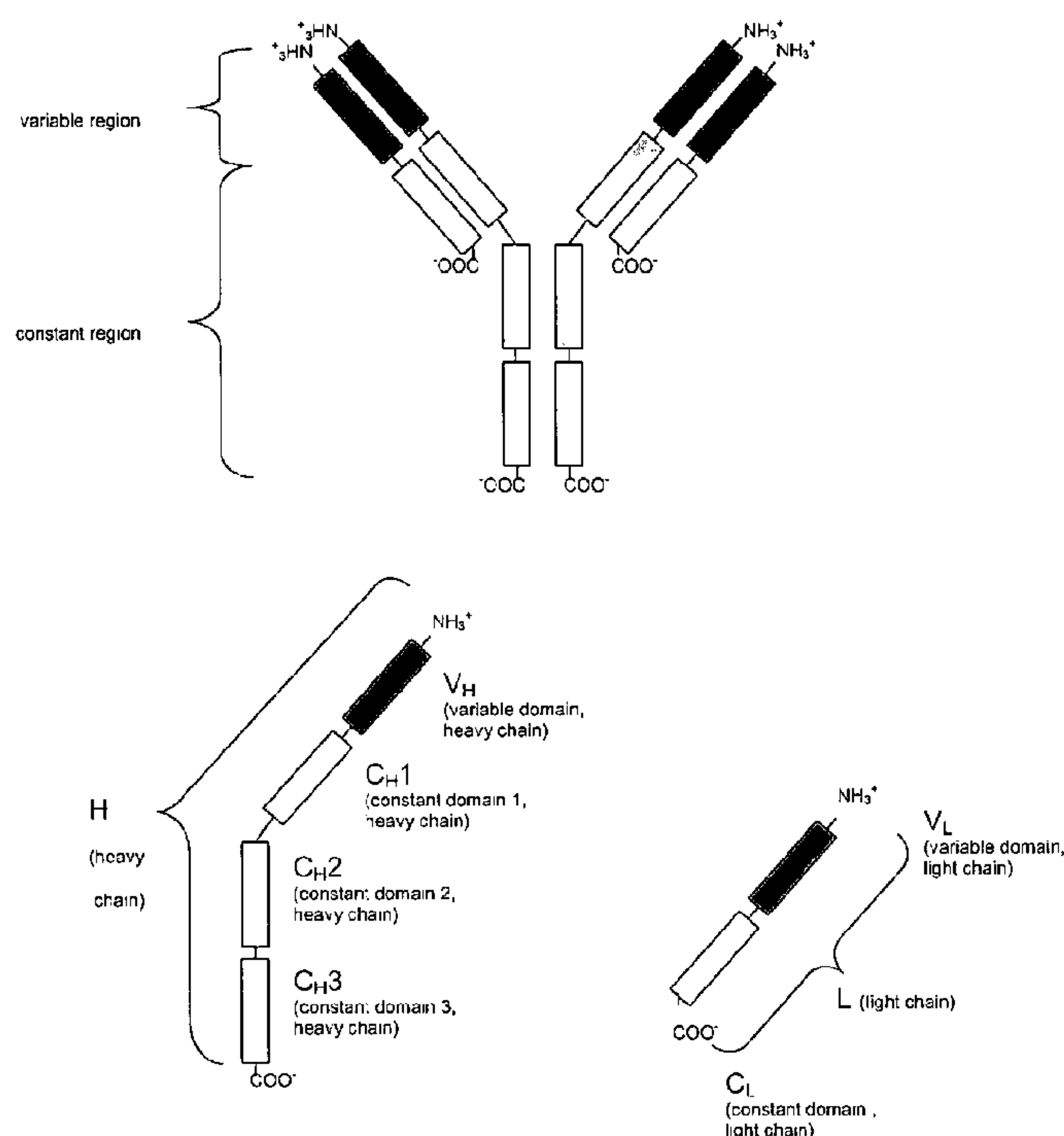




(86) Date de dépôt PCT/PCT Filing Date: 2006/10/19
(87) Date publication PCT/PCT Publication Date: 2007/04/26
(45) Date de délivrance/Issue Date: 2015/03/17
(85) Entrée phase nationale/National Entry: 2008/04/03
(86) N° demande PCT/PCT Application No.: EP 2006/010067
(87) N° publication PCT/PCT Publication No.: 2007/045465
(30) Priorités/Priorities: 2005/10/21 (EP05023003.6);
2006/05/24 (EP06010665.5)

(51) Cl.Int./Int.Cl. *C12N 15/62* (2006.01),
C07K 19/00 (2006.01)
(72) Inventeur/Inventor:
KOPETZKI, ERHARD, DE
(73) Propriétaire/Owner:
F. HOFFMANN-LA ROCHE AG, US
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : PROCEDE D'EXPRESSION PAR RECOMBINAISON D'UN POLYPEPTIDE
(54) Title: METHOD FOR THE RECOMBINANT EXPRESSION OF A POLYPEPTIDE



(57) Abrégé/Abstract:

A method for the recombinant production of a heterologous polypeptide in a eukaryotic host cell is described. The host cell comprises an expression plasmid, whereby the expression plasmid comprises in a 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide, whose amino acid sequence is selected from Table 1 depending on the first two amino acids of the second polypeptide, c) a nucleic acid encoding a second polypeptide comprising a nucleic acid encoding a heterologous polypeptide, a nucleic acid encoding a linker, and a nucleic acid encoding an immunoglobulin fragment, and d) a 3' untranslated region comprising a polyadenylation signal. Further a plasmid and a kit are described.



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

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
26 April 2007 (26.04.2007)

PCT

(10) International Publication Number
WO 2007/045465 A1

(51) International Patent Classification:

C12N 15/62 (2006.01) C07K 19/00 (2006.01)

(21) International Application Number:

PCT/EP2006/010067

(22) International Filing Date: 19 October 2006 (19.10.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

05023003.6 21 October 2005 (21.10.2005) EP

06010665.5 24 May 2006 (24.05.2006) EP

(71) Applicant (for all designated States except US): **F. HOFF-MANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **KOPETZKI, Erhard** [DE/DE]; Kastnerhofstrasse 21, 82377 Penzberg (DE).(74) Agent: **SCHREINER, Siegfried**; Roche Diagnostics GmbH, Patent Department (TR-E), Postfach 11 52, 82372 Penzberg (DE).

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

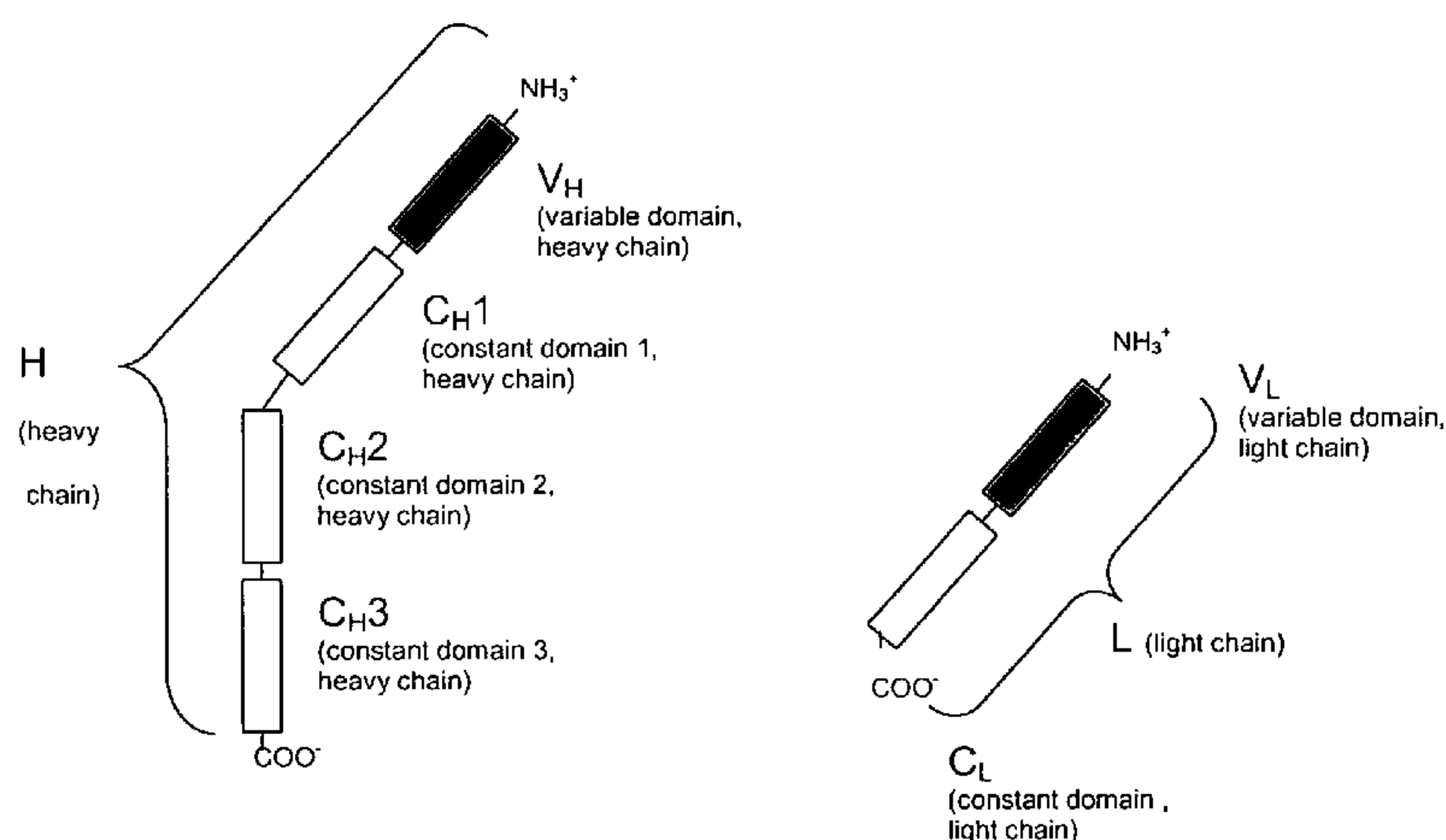
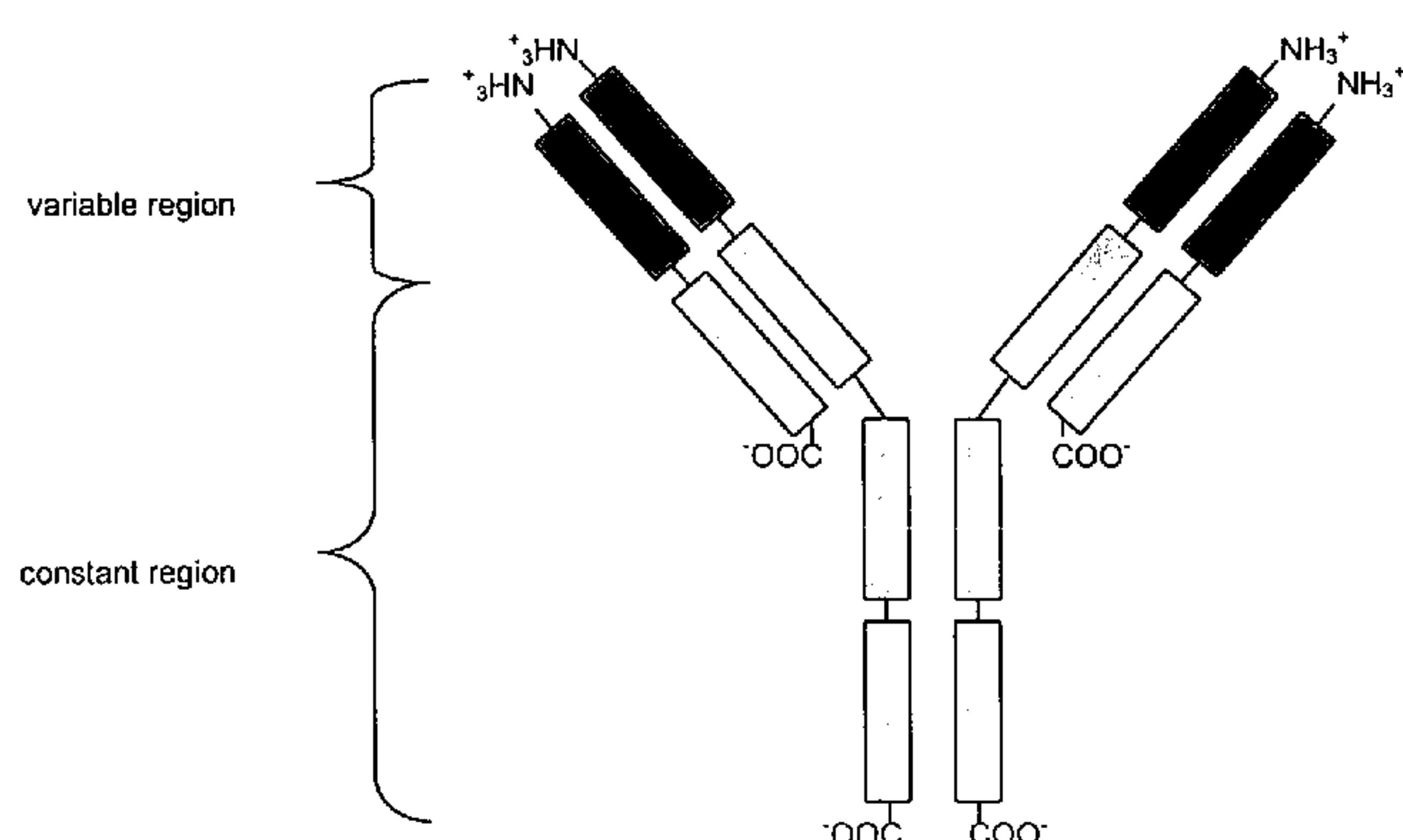
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv))

[Continued on next page]

(54) Title: METHOD FOR THE RECOMBINANT EXPRESSION OF A POLYPEPTIDE



(57) Abstract: A method for the recombinant production of a heterologous polypeptide in a eukaryotic host cell is described. The host cell comprises an expression plasmid, whereby the expression plasmid comprises in a 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide, whose amino acid sequence is selected from Table 1 depending on the first two amino acids of the second polypeptide, c) a nucleic acid encoding a second polypeptide comprising a nucleic acid encoding a heterologous polypeptide, a nucleic acid encoding a linker, and a nucleic acid encoding an immunoglobulin fragment, and d) a 3' untranslated region comprising a polyadenylation signal. Further a plasmid and a kit are described.

WO 2007/045465 A1

WO 2007/045465 A1



Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

Method for the recombinant expression of a polypeptide

The present invention relates to a method for the recombinant expression of a polypeptide in eukaryotic cells.

Background of the Invention

Expression systems for the production of recombinant polypeptides are well-known in the state of the art and are described by, e.g., Marino, M.H., Biopharm. 2 (1989) 18-33; Goeddel, D.V., et al., Methods Enzymol. 185 (1990) 3-7; Wurm, F., and Bernard, A., Curr. Opin. Biotechnol. 10 (1999) 156-159. Polypeptides for use in pharmaceutical applications are preferably produced in mammalian cells such as CHO cells, NS0 cells, Sp2/0 cells, COS cells, HEK cells, BHK cells and the like. The essential elements of an expression plasmid are a prokaryotic plasmid propagation unit for example for E.coli comprising an origin of replication and a selection marker, an eukaryotic selection marker, and one or more expression cassettes for the expression of the structural gene(s) of interest each comprising a promoter, a structural gene, and a transcription terminator including a polyadenylation signal. For transient expression in mammalian cells a mammalian origin of replication, such as the SV40 Ori or OriP, can be included. As promoter a constitutive or inducible promoter can be selected. For optimized transcription a Kozak sequence may be included in the 5' untranslated region. For mRNA processing, in particular mRNA splicing and transcription termination, mRNA splicing signals, depending on the organization of the structural gene (exon/intron organization), may be included as well as a polyadenylation signal.

Expression of a gene is performed either as transient or as permanent expression. The polypeptide(s) of interest are in general secreted polypeptides and therefore contain an N-terminal extension (also known as the signal sequence) which is necessary for the transport/secretion of the polypeptide through the cell into the extracellular medium.

In general, the signal sequence can be derived from any gene encoding a secreted polypeptide. If a heterologous signal sequence is used, it preferably is one that is recognized and processed (i.e. cleaved by a signal peptidase) by the host cell. For secretion in yeast for example the native signal sequence of a heterologous gene to be expressed may be substituted by a homologous yeast signal sequence derived from a secreted gene, such as the yeast invertase signal sequence, alpha-factor leader

- 2 -

(including *Saccharomyces*, *Kluyveromyces*, *Pichia*, and *Hansenula* α -factor leaders, the second described in US 5,010,182), acid phosphatase signal sequence, or the *C. albicans* glucoamylase signal sequence (EP 0 362 179). In mammalian cell expression the native signal sequence of the protein of interest is satisfactory, although other mammalian signal sequences may be suitable, such as signal sequences from secreted polypeptides of the same or related species, e.g. for immunoglobulins from human or murine origin, as well as viral secretory signal sequences, for example, the herpes simplex glycoprotein D signal sequence. The DNA fragment encoding for such a presegment is ligated in frame to the DNA fragment encoding a polypeptide of interest.

In WO 98/28427 a genetically or chemically prepared fusion protein comprising the Fc immunoglobulin region, a derivative or analog fused to the N-terminal portion of the OB protein is reported. A chimeric molecule, i.e. antibody fusion or fusion protein, comprising a carboxy terminal protein import sequence and an amino terminal cargo region is presented in WO 03/035892.

In US 2003/0049227 a method for the induction of a cytotoxic immune response against a tumor in a mammal by administering an immunocytokine, which is a fusion protein comprising an amino-terminal immunoglobulin part and a carboxy-terminal cytokine part, is reported.

WO 91/16437 reports a soluble recombinant fused protein which is stable in the mammalian circulatory system comprising a polypeptide which contains a recognition site for a target molecule, such as a complement receptor site, and is joined to the N-terminal end of an immunoglobulin chain. A fusion protein made up of an antibody and a peptide having a biological activity is reported in US 2003/0103984.

In US 2004/0033511 an antibody-cytokine fusion protein and in US 2004/0180035 an antibody-cytokine immunoconjugate are reported. An immunotoxin comprising Gelonin and an antibody is reported in WO 94/26910.

Summary of the Invention

The current invention comprises a method for the recombinant production of a heterologous polypeptide in a eukaryotic host cell comprising an expression plasmid, whereby the expression plasmid comprises in a 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide, whose amino acid

- 3 -

sequence is selected from Table 1 depending on the first two amino acids of the second polypeptide, c) a nucleic acid encoding a second polypeptide comprising a nucleic acid encoding the heterologous polypeptide, a nucleic acid encoding a linker, and a nucleic acid encoding an immunoglobulin fragment, and d) a 3' untranslated region comprising a polyadenylation signal. The method further comprises the introduction of the expression plasmid into a eukaryotic host cell which is cultivated under conditions suitable for the expression of the second polypeptide and the second polypeptide is recovered from the culture medium.

In one embodiment of the invention the nucleic acid encoding the second polypeptide contains in 5' position to the nucleic acid encoding the heterologous polypeptide an additional nucleic acid encoding either a single amino acid or a dipeptide or the peptide of the amino acid sequence QIWNN (SEQ ID NO: 472) or a fragment thereof.

In another embodiment the immunoglobulin fragment is obtained either from an IgG or from an IgE.

In a further embodiment the eukaryotic cell is a mammalian cell, especially a CHO cell, NS0 cell, Sp2/0 cell, COS cell, K562 cell, BHK cell, PER.C6 cell or HEK cell.

In still another embodiment the linker is a peptide or polypeptide selected from the group consisting of SEQ ID NOs: 06, 07, 08, 09, 10, 139, 140, 554, 555, 556, and 557.

In another embodiment the immunoglobulin fragment comprises the carboxy-terminal constant domain of a heavy or light chain of a naturally occurring or synthetic immunoglobulin, i.e. either the C_H1-, the hinge region, the C_H2-, the C_H3-domain of a heavy chain or the C_L-domain of a light chain. Additionally the immunoglobulin fragment comprises a variable domain fragment.

In another embodiment the variable domain fragment is a variable domain of an immunoglobulin heavy or light chain in which of from one to six amino acids of the variable domain are deleted.

In a further embodiment of from one to six regions (FR1, FR2, FR3, CDR1, CDR2, CDR3) of the variable domain are deleted.

In a further embodiment the variable domain is deleted.

- 4 -

In another embodiment the immunoglobulin fragment is derived from a naturally occurring immunoglobulin or a variant thereof.

In a further embodiment the immunoglobulin fragment is derived from an at least partially synthetic immunoglobulin.

5 In still another embodiment of the invention the amino acid sequence of the heterologous polypeptide is of from 5 to 500 amino acid residues, more preferred of from 10 to 350 amino acid residues, most preferred of from 15 to 150 amino acid residues.

10 The invention further comprises a plasmid comprising in a 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide, whose amino acid sequence is selected from Table 1 depending on the first two amino acids of the second polypeptide, c) a nucleic acid encoding a second polypeptide comprising a nucleic acid encoding a heterologous polypeptide, a nucleic acid encoding a linker, and a nucleic acid encoding an immunoglobulin fragment, and d) a 3'-untranslated
15 region comprising a polyadenylation signal.

The invention still further comprises a kit for the preparation of a plasmid for the expression of a heterologous polypeptide in a eukaryotic cell comprising a plasmid which comprises in a 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide, whose amino acid sequence is selected from the group consisting
20 of SEQ ID NO: 36, 37, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, and 329, c) a nucleic acid encoding a second polypeptide comprising i) a nucleic acid encoding a peptide of the amino acid sequence QIWNN (SEQ ID NO: 472) or a N-terminal fraction thereof, ii) a cloning site comprising at least one restriction cleavage site suitable for the insertion of a nucleic acid encoding a heterologous polypeptide, iii)
25 a nucleic acid encoding a linker selected from the group consisting of SEQ ID NOs 06, 07, 08, 09, 10, 139, 140, 554, 555, 556, and 557, and iv) a nucleic acid encoding an immunoglobulin fragment, and d) a 3'-untranslated region comprising a polyadenylation signal.

Detailed Description of the Invention

30 The current invention comprises a method for the recombinant expression of a heterologous polypeptide of interest in an eukaryotic host cell comprising an expression vector, which contains a suitable promoter, a transcription terminator, a selectable marker, a nucleic acid sequence encoding a polypeptide, and a nucleic

- 5 -

acid sequence encoding a signal sequence, whereby the nucleic acid sequence encoding the signal sequence is selected from Table 1 depending on the first two amino acids of the following polypeptide. The nucleic acid sequence encoding the heterologous polypeptide starts within fifteen nucleotides after the end of the nucleic acid sequence encoding the signal sequence. The nucleic acid sequence encoding the heterologous polypeptide can either be inserted within a FR1-region of an immunoglobulin, within a V_L-region of an immunoglobulin or within the first constant domain of an immunoglobulin or it can replace all or a fraction of a FR1-region of an immunoglobulin, a V_L-region of an immunoglobulin, or the first constant domain of an immunoglobulin.

Within the scope of the present invention some of the terms used are defined as follows:

A "nucleic acid molecule" as used herein, refers to a naturally occurring, or partially or fully non-naturally occurring nucleic acid encoding a polypeptide which can be produced recombinantly. The nucleic acid molecule can be build up of DNA-fragments which are either isolated or synthesized by chemical means. The nucleic acid molecule can be integrated into another nucleic acid, e.g. in an expression plasmid or the genome/chromosome of a eukaryotic host cell. Plasmid includes shuttle and expression vectors. Typically, the plasmid will also comprise a prokaryotic propagation unit comprising an origin of replication (e.g. the ColE1 origin of replication) and a selectable marker (e.g. ampicillin or tetracycline resistance gene) for replication and selection, respectively, of the vector in bacteria.

An "expression cassette" refers to a nucleic acid sequence that contains the elements necessary for expression and secretion of at least the contained structural gene in a cell.

A nucleic acid molecule is likewise characterized by its nucleic acid sequence consisting of individual nucleotides or/and by an amino acid sequence encoded by the nucleic acid molecule.

A "gene" denotes a segment e.g. on a chromosome or on a plasmid which is necessary for the expression of a peptide, polypeptide or protein. Beside the coding region the gene comprises other functional elements including a promoter, introns, and terminators.

A "structural gene" denotes the coding region of a gene without a signal sequence.

- 6 -

A "resistance gene" or a "selectable marker", which is used interchangeably within this application, is a gene that allows cells carrying the gene to be specifically selected for or against, in the presence of a corresponding selection agent. A useful positive selectable marker is an antibiotic resistance gene. This selectable marker allows the host cell transformed with the gene to be positively selected for in the presence of the corresponding antibiotic; a non-transformed host cell would not be capable to grow or survive under the selective culture conditions. Selectable markers can be positive, negative or bifunctional. Positive selectable markers allow selection for cells carrying the marker, whereas negative selectable markers allow cells carrying the marker to be selectively eliminated. Typically, a selectable marker will confer resistance to a drug or compensate for a metabolic or catabolic defect in the host cell. Resistance genes useful with eukaryotic cells include, e.g., the genes for aminoglycoside phosphotransferase (APH), such as the hygromycin phosphotransferase (hyg), neomycin and G418 APH, dihydrofolate reductase (DHFR), thymidine kinase (tk), glutamine synthetase (GS), asparagine synthetase, tryptophan synthetase (indole), histidinol dehydrogenase (histidinol D), and genes encoding resistance to puromycin, bleomycin, phleomycin, chloramphenicol, Zeocin, and mycophenolic acid. Further marker genes are described in WO 92/08796 and WO 94/28143.

"Regulatory elements" as used herein, refer to nucleotide sequences present in cis, necessary for transcription and/or translation of the gene comprising the nucleic acid sequence encoding a polypeptide of interest. The transcriptional regulatory elements normally comprise a promoter upstream of the structural gene sequence to be expressed, transcriptional initiation and termination sites, and a polyadenylation signal sequence. The term "transcriptional initiation site" refers to the nucleic acid base in the gene corresponding to the first nucleic acid incorporated into the primary transcript, i.e. the mRNA precursor; the transcriptional initiation site may overlap with the promoter sequence. The term "transcriptional termination site" refers to a nucleotide sequence normally represented at the 3' end of a gene of interest to be transcribed, that causes RNA polymerase to terminate transcription. The polyadenylation signal sequence, or poly-A addition signal provides the signal for the cleavage at a specific site at the 3' end of eukaryotic mRNA and the post-transcriptional addition in the nucleus of a sequence of about 100-200 adenine nucleotides (polyA tail) to the cleaved 3' end. The polyadenylation signal sequence may include the consensus sequence AATAAA located at about 10-30 nucleotides upstream from the site of cleavage.

- 7 -

To produce a secreted polypeptide, the structural gene of interest includes a DNA segment that encodes a signal sequence/leader peptide. The signal sequence directs the newly synthesized polypeptide to and through the ER membrane where the polypeptide can be routed for secretion. The signal sequence is cleaved off by a signal peptidases during the protein crosses the ER membrane. As for the function of the signal sequence the recognition by the host cell's secretion machinery is essential. Therefore the used signal sequence has to be recognized by the host cell's proteins and enzymes of the secretion machinery.

Translational regulatory elements include a translational initiation (AUG) and stop codon (TAA, TAG, or TGA). An internal ribosome entry site (IRES) can be included in some constructs.

A "promoter" refers to a polynucleotide sequence that controls transcription of a gene/structural gene or nucleic acid sequence to which it is operably linked. A promoter includes signals for RNA polymerase binding and transcription initiation. The promoters used will be functional in the cell type of the host cell in which expression of the selected sequence is contemplated. A large number of promoters including constitutive, inducible and repressible promoters from a variety of different sources, are well known in the art (and identified in databases such as GenBank) and are available as or within cloned polynucleotides (from, e.g., depositories such as ATCC as well as other commercial or individual sources). A "promoter" comprises a nucleotide sequence that directs the transcription of a structural gene. Typically, a promoter is located in the 5' non-coding or untranslated region of a gene, proximal to the transcriptional start site of a structural gene. Sequence elements within promoters that function in the initiation of transcription are often characterized by consensus nucleotide sequences. These promoter elements include RNA polymerase binding sites, TATA sequences, CAAT sequences, differentiation-specific elements (DSEs; McGehee, R.E., et al., Mol. Endocrinol. 7 (1993) 551), cyclic AMP response elements (CREs), serum response elements (SREs; Treisman, R., Seminars in Cancer Biol. 1 (1990) 47), glucocorticoid response elements (GREs), and binding sites for other transcription factors, such as CRE/ATF (O'Reilly, M.A., et al., J. Biol. Chem. 267 (1992) 19938), AP2 (Ye, J., et al., J. Biol. Chem. 269 (1994) 25728), SP1, cAMP response element binding protein (CREB; Loeken, M.R., Gene Expr. 3 (1993) 253) and octamer factors (see, in general, Watson et al., eds., Molecular Biology of the Gene, 4th ed. (The Benjamin/Cummings Publishing Company, Inc. 1987), and Lemaigre, F.P. and Rousseau, G.G., Biochem. J. 303 (1994) 1-14). If a promoter is an inducible

- 8 -

promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter. Repressible promoters are also known. For example, the c-fos promoter is specifically activated upon binding of growth hormone to its receptor on the cell surface. Tetracycline (tet) regulated expression can be achieved by artificial hybrid promoters that consist e.g. of a CMV promoter followed by two Tet-operator sites. The Tet-repressor binds to the two Tet-operator sites and blocks transcription. Upon addition of the inducer tetracycline, Tet-repressor is released from the Tet-operator sites and transcription proceeds (Gossen, M. and Bujard, H., PNAS 89 (1992) 5547-5551). For other inducible promoters including metallothionein and heat shock promoters, see, e.g., Sambrook et al. (supra) and Gossen et al., Curr. Opin. Biotech. 5 (1994) 516-520. Among the eukaryotic promoters that have been identified as strong promoters for high-level expression are the SV40 early promoter, adenovirus major late promoter, mouse metallothionein-I promoter, Rous sarcoma virus long terminal repeat, Chinese hamster elongation factor 1 alpha (CHEF-1, see e.g. US 5,888,809), human EF-1 alpha, ubiquitin, and human cytomegalovirus immediate early promoter (CMV IE).

The "promoter" can be constitutive or inducible. An enhancer (i.e., a cis-acting DNA element that acts on a promoter to increase transcription) may be necessary to function in conjunction with the promoter to increase the level of expression obtained with a promoter alone, and may be included as a transcriptional regulatory element. Often, the polynucleotide segment containing the promoter will include enhancer sequences as well (e.g., CMV or SV40).

An "enhancer", as used herein, refers to a polynucleotide sequence that enhances transcription of a gene or coding sequence to which it is operably linked. Unlike promoters, enhancers are relatively orientation and position independent and have been found 5' or 3' (Lusky, M., et al., Mol. Cell Bio. 3 (1983) 1108) to the transcription unit, within an intron (Banerji, J., et al., Cell 33 (1983) 729) as well as within the coding sequence itself (Osborne, T.F., et al., Mol. Cell Bio. 4 (1984) 1293). Therefore, enhancers may be placed upstream or downstream from the transcription initiation site or at considerable distances from the promoter, although in practice enhancers may overlap physically and functionally with promoters. A large number of enhancers, from a variety of different sources are well known in the art (and identified in databases such as GenBank) and available as or within cloned polynucleotide sequences (from, e.g., depositories such as the ATCC

- 9 -

as well as other commercial or individual sources). A number of polynucleotides comprising promoter sequences (such as the commonly-used CMV promoter) also comprise enhancer sequences. For example, all of the strong promoters listed above may also contain strong enhancers (see e.g. Bendig, M.M., Genetic Engineering 7 (1988) 91-127).

An "internal ribosome entry site" or "IRES" describes a sequence which functionally promotes translation initiation independent from the gene 5' of the IRES and allows two cistrons (open reading frames) to be translated from a single transcript in an animal cell. The IRES provides an independent ribosome entry site for translation of the open reading frame immediately downstream (downstream is used interchangeably herein with 3') of it. Unlike bacterial mRNA which can be polycistronic, i.e. encode several different polypeptides that are translated sequentially from the mRNAs, most mRNAs of animal cells are monocistronic and code for the synthesis of only one protein. With a polycistronic transcript in a eukaryotic cell, translation would initiate from the most 5' translation initiation site, terminate at the first stop codon, and the transcript would be released from the ribosome, resulting in the translation of only the first encoded polypeptide in the mRNA. In a eukaryotic cell, a polycistronic transcript having an IRES operably linked to the second or subsequent open reading frame in the transcript allows the sequential translation of that downstream open reading frame to produce the two or more polypeptides encoded by the same transcript. The use of IRES elements in vector construction has been previously described, see, e.g., Pelletier, J., et al., Nature 334 (1988) 320-325; Jang, S.K., et al., J. Virol. 63 (1989) 1651- 1660; Davies, M.V., et al., J. Virol. 66 (1992) 1924-1932; Adam, M.A., et al., J. Virol. 65 (1991) 4985-4990; Morgan, R.A., et al., Nucl. Acids Res. 20 (1992) 1293-1299; Sugimoto, Y., et al., Biotechnology 12 (1994) 694-698; Ramesh, N., et al., Nucl. Acids Res. 24 (1996) 2697-2700; and Mosser, D.D., et al., Biotechniques 22 (1997) 150-152).

"Operably linked" refers to a juxtaposition of two or more components, wherein the components so described are in a relationship permitting them to function in their intended manner. For example, a promoter and/or enhancer are operably linked to a coding sequence, if it acts in cis to control or modulate the transcription of the linked sequence. Generally, but not necessarily, the DNA sequences that are "operably linked" are contiguous and, where necessary to join two protein encoding regions such as a secretory leader/signal sequence and a polypeptide, contiguous and in reading frame. However, although an operably linked promoter is generally located upstream of the coding sequence, it is not necessarily contiguous with it.

- 10 -

Enhancers do not have to be contiguous. An enhancer is operably linked to a coding sequence if the enhancer increases transcription of the coding sequence. Operably linked enhancers can be located upstream, within, or downstream of coding sequences and at considerable distance from the promoter. A
5 polyadenylation site is operably linked to a coding sequence if it is located at the downstream end of the coding sequence such that transcription proceeds through the coding sequence into the polyadenylation sequence. Linking is accomplished by recombinant methods known in the art, e.g., using PCR methodology and/or by ligation at convenient restriction sites. If convenient restriction sites do not exist,
10 then synthetic oligonucleotide adaptors or linkers are used in accord with conventional practice.

The term "expression" as used herein refers to transcription and/or translation occurring within a host cell. The level of transcription of a desired product in a host cell can be determined on the basis of the amount of corresponding mRNA that is
15 present in the cell. For example, mRNA transcribed from a selected sequence can be quantitated by PCR or by Northern hybridization (see Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989)). Protein encoded by a selected sequence can be quantitated by various methods, e.g. by ELISA, by assaying for the biological activity of the protein, or by employing
20 assays that are independent of such activity, such as Western blotting or radioimmunoassay, using antibodies that recognize and bind to the protein (see Sambrook et al., 1989, supra).

A "host cell" refers to a cell into which the gene encoding the polypeptide of the invention is introduced. Host cell includes both prokaryotic cells used for
25 propagation of the plasmids/vectors, and eukaryotic cells for expression of the structural gene. Typically, the eukaryotic cells are mammalian cells.

A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, whether produced naturally or synthetically. Polypeptides of less than about 20 amino acid residues may be referred to as "peptides." Polypeptides comprising one
30 or more polypeptide chains or comprising an amino acid chain of a length of 100 amino acids or more may be referred to as "proteins".

A "protein" is a macromolecule comprising one or more polypeptide chains whereby at least one chain has an amino acid length of 100 amino acids or more. A protein may also comprise non-peptidic components, such as carbohydrate groups.

- 11 -

Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and may vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures; additions such as carbohydrate groups are generally not specified, but may be present nonetheless.

"Heterologous DNA" or „heterologous polypeptide“ refers to a DNA molecule or a polypeptide, or a population of DNA molecules or a population of polypeptides, that do not exist naturally within a given host cell. DNA molecules heterologous to a particular host cell may contain DNA derived from the host cell species (i.e. endogenous DNA) so long as that host DNA is combined with non-host DNA (i.e. exogenous DNA). For example, a DNA molecule containing a non-host DNA segment encoding a polypeptide operably linked to a host DNA segment comprising a promoter is considered to be a heterologous DNA molecule. Conversely, a heterologous DNA molecule can comprise an endogenous structural gene operably linked with an exogenous promoter.

A peptide or polypeptide encoded by a non-host DNA molecule is a "heterologous" peptide or polypeptide.

A "cloning vector" is a nucleic acid molecule, such as a plasmid, cosmid, phageimide or bacterial artificial chromosome (BAC), which has the capability of replicating autonomously in a host cell. Cloning vectors typically contain one or a small number of restriction endonuclease recognition sites that allow insertion of a nucleic acid molecule in a determinable fashion without loss of an essential biological function of the vector, as well as nucleotide sequences encoding a resistance gene that is suitable for use in the identification and selection of cells transformed with the cloning vector. Resistance genes typically include genes that provide tetracycline resistance or ampicillin resistance.

An "expression plasmid" is a nucleic acid molecule encoding a protein to be expressed in a host cell. Typically, an expression plasmid comprises a prokaryotic plasmid propagation unit, e.g. for E.coli, comprising an origin of replication, and a selection marker, an eukaryotic selection marker, and one or more expression cassettes for the expression of the structural gene(s) of interest comprising a promoter, a structural gene, and a transcription terminator including a polyadenylation signal. Gene expression is usually placed under the control of a promoter, and such a structural gene is said to be "operably linked to" the

- 12 -

promoter. Similarly, a regulatory element and a core promoter are operably linked if the regulatory element modulates the activity of the core promoter.

A "polycistronic transcription unit" is a transcription unit in which more than one structural gene is under the control of the same promoter.

5 An "isolated polypeptide" is a polypeptide that is essentially free from contaminating cellular components, such as carbohydrate, lipid, or other proteinaceous impurities associated with the polypeptide in nature. Typically, a preparation of an isolated polypeptide contains the polypeptide in a highly purified form, i.e. at least about 80% pure, at least about 90% pure, at least about 95% pure,
10 greater than 95% pure, or greater than 99% pure. One way to show that a particular protein preparation contains an isolated polypeptide is by the appearance of a single band following sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of the protein preparation and Coomassie Brilliant Blue staining of the gel. However, the term "isolated" does not exclude the presence of the same
15 polypeptide in alternative physical forms, such as dimers or alternatively glycosylated or derivatized forms.

The term "immunoglobulin" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes. The recognized immunoglobulin genes include the different constant domain (region) genes as well
20 as the myriad immunoglobulin variable region genes. Immunoglobulins may exist in a variety of formats, including, for example, Fv, Fab, and F(ab)₂ as well as single chains (scFv) (e.g. Huston, J.S., et al., PNAS USA 85 (1988) 5879-5883; Bird, R.E., et al., Science 242 (1988) 423-426; in general, Hood et al., Immunology, Benjamin N.Y., 2nd edition (1984); and Hunkapiller, T. and Hood, L., Nature 323 (1986) 15-
25 16).

An immunoglobulin in general comprises at least two light chain polypeptides and two heavy chain polypeptides. Each of the heavy and light polypeptide chains may contain a variable domain (region) (generally the amino terminal portion of the polypeptide chain) which contains a binding region (domain) that is able to
30 interact with an antigen. Each of the heavy and light polypeptide chains comprises a constant region (generally the carboxyl terminal portion). The constant region of the heavy chain mediates the binding of the antibody i) to cells bearing a Fc gamma receptor (FcγR), such as phagocytic cells, or ii) to cells bearing the neonatal Fc receptor (FcRn) also known as Brambell receptor. It also mediates the binding to

- 13 -

some factors including factors of the classical complement system such as component (C1q).

5 The variable domain of an immunoglobulin's light or heavy chain in turn comprises different segments, i.e. four framework regions (FR) and three hypervariable regions (CDR).

10 An "immunoglobulin fragment" denotes a polypeptide comprising at least the constant domains of a chain of an immunoglobulin, i.e. C_H1, hinge-region, C_H2, and C_H3 and optionally C_H4 of a heavy chain of an immunoglobulin or C_L of a light chain of an immunoglobulin. Also comprised are derivatives and variants thereof. Additionally a variable domain, in which one or more amino acids or amino acid regions are deleted, may be present. In a preferred embodiment the variable domain is deleted in the immunoglobulin fragment.

15 „Transcription terminator“ as denoted within this application is a DNA sequence of 50-750 base pairs in length which gives the RNA polymerase the signal for termination of the mRNA synthesis. Very efficient (strong) terminators at the 3' end of an expression cassette are advisable to prevent the RNA polymerase from reading through particularly when using strong promoters. Inefficient transcription terminators can lead to the formation of an operon-like mRNA which can be the reason for an undesired, e.g. plasmid-coded, gene expression.

20 The term "linker" as used within this application denotes peptide linkers of natural or synthetic origin. They are building up a linear amino acid chain. The chain has a length of from 1 to 50 amino acids, preferred between 3 and 25 amino acids. The linker may contain repetitive amino acid sequences or parts of naturally occurring polypeptides, such as polypeptides with a hinge-function.

25 "Synthetic linkers" are designated to be rich in glycine, glutamine, and serine residues. These residues are arranged in a small peptide unit of up to five amino acids, such as GGGGS, QQQQG, or SSSSG. The small peptide unit is repeated for two to five times to form a multimeric unit. At each of the amino- and/or carboxy-terminal end of the multimeric unit up to six additional amino acids can be added.

30 The term „biologically active molecule“ as used herein refers to an organic molecule, e.g. a biological macromolecule such as a peptide, protein, glycoprotein, nucleoprotein, mucoprotein, lipoprotein, synthetic polypeptide or protein, that causes a biological effect when administered in or to artificial biological systems,

- 14 -

such as bioassays using cell lines and viruses, or in vivo to an animal, including but not limited to birds and mammals, including humans. This biological effect can be but is not limited to enzyme inhibition or activation, binding to a receptor or a ligand, either at the binding site or circumferential, signal triggering or signal modulation.

Biologically active molecules are without limitation for example hormones, cytokines, growth factors, receptor ligands, agonists or antagonists, cytotoxic agents, antiviral agents, imaging agents, enzyme inhibitors, enzyme activators or enzyme activity modulators such as allosteric substances.

The term "amino acid" as used within this application comprises alanine (three letter code: ala, one letter code: A), arginine (arg, R), asparagine (asn, N), aspartic acid (asp, D), cysteine (cys, C), glutamine (gln, Q), glutamic acid (glu, E), glycine (gly, G), histidine (his, H), isoleucine (ile, I), leucine (leu, L), lysine (lys, K), methionine (met, M), phenylalanine (phe, F), proline (pro, P), serine (ser, S), threonine (thr, T), tryptophan (trp, W), tyrosine (tyr, Y), and valine (val, V).

Methods and techniques known to a person skilled in the art, which are useful for carrying out the current invention, are described e.g. in Ausubel, F.M., ed., Current Protocols in Molecular Biology, Volumes I to III (1997), Wiley and Sons; Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

The invention comprises a method for the recombinant production of a heterologous polypeptide in a eukaryotic host cell. The host cell comprises an expression plasmid, which comprises in 5' to 3' direction a) a promoter, b) a nucleic acid encoding a first polypeptide whose amino acid sequence is selected from Table 1 depending of the first two amino acids of the second polypeptide, c) a nucleic acid encoding a second polypeptide comprising a nucleic acid encoding a heterologous polypeptide having a biological activity, a nucleic acid encoding a peptide or polypeptide selected from the group consisting of SEQ ID NO: 06-10, 139, 140, and 554-557, a nucleic acid encoding an immunoglobulin fragment, and d) a 3' untranslated region. This expression plasmid is introduced into a host cell which is cultivated under conditions suitable for the expression of the second polypeptide. The secreted second polypeptide is recovered from the culture medium.

- 15 -

The first polypeptide is a so-called signal sequence. The signal sequence is responsible for the secretion of the attached/succeeding/operably linked polypeptide. To be effective the signal sequence has to be recognized and processed by the proteins and enzymes within the cell expressing the polypeptide. In case of a eukaryotic host cell, the signal sequence is preferably a eukaryotic one. To assure that the second polypeptide according to the current invention is secreted correctly, the signal sequence is selected from human and murine immunoglobulin signal sequences. A compilation of signal sequences is shown in table 1.

Which signal sequence is selected depends on the succeeding amino acids. It has to be assured that the signal peptidase, which cleaves the signal sequence after the secretion process, recognizes the signal sequence of the secreted polypeptide and removes it. To provide a "natural" transition from the signal sequence to the heterologous polypeptide the signal sequence should be chosen that way, that the first two amino acids of the heterologous polypeptide are identical to the first two amino acids of the amino acid sequences of the naturally following immunoglobulin FR1-region.

Table 1: Set of the first two amino acids (given in one letter code) of the second polypeptide assigned to signal peptides (first polypeptide given in one letter code).

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
AC	MEFQTQVLMSLLCMS	163
	MESQTQVLMFLLWVS	164
	MVSTPQFLVFLFWIP	165
AD	MESQTLVFISILLWLY	166
AG	MSVPTQLLGLLLWLT	167
AH	MKSQTQVFIFLLCVS	168
	MKSQTQVFVFLLCVS	169
AI	MDMRVPAQLLGLLLWLRGARC	68
	MDMRVPAQLLGLLLWLRGARC	69
	MDMRVPAQLLGLLQLWLSGARC	70
	MDMRVPAQLLGLLLWLSGARC	71
	MDMRVPAQLLGLLLWLPDTRC	72
	MDMRVPAQLLGLLLWFPGARC	73
	MDMRVPAQLLGLLLWFPGARC	74
	MDMRVLAQLLGLLLCFPGARC	75
	MDMRVLAQLLGLLLCFPGARC	76
	MDMRVPAQLLGLLLWLPGARC	77
	MDMRVPAQLLGLLLWLPGARC	78
	MDMRVPAQLLGLLLWFPGSRC	79
	MDMRVPAQLLGLLLWFPGSRC	80
	MDMRVPAQLLGLLLWLPGARC	81

- 16 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
AI (continued)	MDMRVPAQRLGLLLLWFPGARC	82
	MRVPAQLLGLLLWLPGARC	83
	MDMRVPAQLLGLLLWLPGARC	84
	MDMRVPAQLLGLLLWLPGARC	85
	MDMRVPAQLLGLLLWLPGAKC	86
	MAWISLILSLLALSS	170
	MAWTSILISLLALCS	171
	MRCLAEFLGLLVLWIP	172
AL	MGWNWIFILISVTT	173
AQ	MRFQVQVLGLLLLWIS	174
	MRPSIQFLGLLFWLH	175
AR	MDIRAPAQFLGILLWFP	176
	MDMMVLAQFLAFLLLWFP	177
	MDMRAPAQFLGILLWFP	178
	MDMRAPAQFLGILLWFP	179
	MDMRAPAQVFGFLLWFP	180
	MDMRASAQFHGILLWFP	181
	MDMRASAQFHGILLWFP	182
	MDMWTSAQFLGILLWFL	183
	MNTRAPAEFLGFLLWFL	184
	MRAPAPFLGLLFCFL	185
	M RTPAPFLGLLFCFS	186
	MSISTQLLGLLLWLT	187
	MSLPTQLQGILLWLT	188
	MSVLTQVLALLLWLT	189
	MSVPTQLLALLLWLT	190
	MSVPTQVLGILLWLT	191
	TDFHMQIFSFMLISFT	192
AS	MAWTSILISLLALCS	193
AT	MRCLAEFLRLLVLWIP	194
CQ	MPWALLLTLLTHSAVSVV	138
DA	MKLPVRLLVLMFWIPSSS	195
DI	MDMRVPAQLLGLLLWLRGARC	68
	MDMRVPAQLLGLLLWLRGARC	69
	MDMRVPAQLLGLLQLWLSGARC	70
	MDMRVPAQLLGLLLWLSGARC	71
	MDMRVPAQLLGLLLWLDPTRC	72
	MDMRVPAQLLGLLLWFPGARC	73
	MDMRVPAQLLGLLLWFPGARC	74
	MDMRVLAQLLGLLLCFPGARC	75
	MDMRVLAQLLGLLLCFPGARC	76
	MDMRVPAQLLGLLLWLPGARC	77
	MDMRVPAQLLGLLLWLPGARC	78
	MDMRVPAQLLGLLLWFPGSRC	79
	MDMRVPAQLLGLLLWFPGSRC	80
	MDMRVPAQLLGLLLWLPGARC	81
	MDMRVPAQRLGILLWFPGARC	82
	MRVPAQLLGLLLWLPGARC	83
	MDMRVPAQLLGLLLWLPGARC	84

- 17 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
DI (continued)	MDMRVPAQLLGLLLWLPGARC	85
	MDMRVPAQLLGLLLWLPGAKC	86
	MRLPAQLLGLLMLWVPGSSE	87
	MRLPAQLLGLLMLWVPGSSE	88
	MRLPAQLLGLLMLWVPGSSG	89
	MRLPAQLLGLLMLWVPGSSG	90
	MRLPAQLLGLLMLWIPGSSA	91
	MRLPAQLLGLLMLWIPGSSA	92
	MRLPAQLLGLLMLWVSGSSG	93
	MRLPAQLLGLLMLWVSGSSG	94
	MRLLAQLLGLLMLWVPGSSG	95
	MVLQTQVFISLLWISGAYG	103
	MDIRAPAQFLGILLWFPARC	196
	MDMMVLAQFLAFLLLWFPARC	197
	MDMRAPAQFFGILLWFPIRC	198
	MDMRAPAQFLGILLWFPARC	199
	MDMRAPAQIFGFLLLFQTRC	200
	MDMRAPAQVFGFLLLWFPARC	201
	MDMRASAQFLGFLLLWFP	202
	MDMRDPPQFLAFLLLWIP	203
	MDMRTPAQFLGILLWFPIKC	204
	MDMRVPAHVFGFLLLWFPTRC	205
	MDSQAQVLILLLLWVSTCG	206
	MDSQAQVLMILLLSVSTCG	207
	MDSQAQVLMILLLLWVSTCG	208
	MDSQARVLMILLLLWVSTCG	209
	MEFQTQVFVFLWLSVDG	210
	MEFQTQVLMISLLCMSACA	211
	MEKDTLLLWVLLWVPSTG	212
	MESDTLLLWVLLWVPSTG	213
	MESQIQAFVFLWLSVDG	214
	MESQIQVFVFLWLSVDG	215
	MESQNHVLMFLLWVSTCG	473
	MESQTHVLMFLLWVSTCG	474
	MESQTQVFVYMLLWLSVDG	475
	MESQTQVLISLLFWVSTCG	476
	MESQTQVLMFLLWVSACA	477
	MESQTQVLMISLLFWVSTCG	478
	METDPLLWVLLWVPSTG	479
	METDTILLWVLLWVPSTG	480
	METDTLLLWVLLWVPSTG	481
	METDTLLLWVLLWVPSTG	482
	METHSQVFVYMLLWLSVEG	483
	MGFKMESHTQAFVFAFLWLSVDG	484
	MGVPTQLLLWLTVRC	485
	MIASAQFLGLLLLCFQTRC	486
	MKFPSQLLLFLLFRITHC	487
	MKFPSQLLLLLFGIPMIC	488
	MMSSAQFLGLLLLCFQTRC	489

- 18 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
DI (continued)	MMSSAQFLGLLLLCFQTRY	490
	MNMLTQLLGLLLLWFAGKC	491
	MRCLAEFLGLLVLWIPAIG	492
	MRCLAEFLRLLVLWIPATG	493
	MRCSLQFLGVLMFWISVSG	494
	MRFSAQLLGLLVLWIPSTA	495
	MRPSIQFLGLLLFWLHAQC	496
	MRVLAELLGLLLFCFLVRC	497
	MRVLPEFLGLLLWISVRC	498
	MSISTQLLGLLLWLTARC	499
	MSVLTQVLALLLWLTARC	500
	MSVPTQLLALLLWLTARC	501
	MSVPTQLLGLLLWLTAGC	502
	MSVPTQVLGLLLWLTARC	503
	MVFTPQILGLMLFWISSTG	504
	MVFTPQILGLMLFWISSRG	505
	MVLGLKWVFFVVFYQSRG	506
	MVSTSQLLGLLLFWTSSRG	507
	PAQFLLLVLWIQSRC	508
DR	MSLLTQLQGLLLLWLT	216
DV	MRLPAQLLGLLMLWVPGSSE	87
	MRLPAQLLGLLMLWVPGSSE	88
	MRLPAQLLGLLMLWVPGSSG	89
	MRLPAQLLGLLMLWVPGSSG	90
	MRLPAQLLGLLMLWIPGSSA	91
	MRLPAQLLGLLMLWIPGSSA	92
	MRLPAQLLGLLMLWVSGSSG	93
	MRLPAQLLGLLMLWVSGSSG	94
	MRLLAQLLGLLMLWVPGSSG	95
	MLPSQLIGFLLLWVPASRG	105
	MLPSQLIGFLLLWVPASRG	106
	MVSPLQFLRLLLWVPASRG	107
	LILKVQC	217
	LVLKVLC	218
	MDMRASAQFHGILLWFPARC	219
	MKLPVLLVLLLFTSPSSS	220
	MKLPVRLLVLMFWIPSSS	221
	MMSPAQFLFLLVLWIQTNG	222
	MMSPAQFLFLLVLWIRTNG	223
	MMSPVHSIFILLWIVISG	224
	MMSPVQFLFLLMLWIQTNG	225
	MNFGRLRIFLVLTLKVQC	226
	MNLPVHLLVLLFWIPSRG	227
	MNTRAPAEFLGFLLWFLARC	228
	MRFQVQVLGLLLWISAQC	229
	MRVLSLLYLLTAIPGILS	230
EI	METPAQLLFLLLWLPD TTG	96
	METPAQLLFLLLWLPD TTG	97
	MEAPAQLLFLLLWLPD TTG	98

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
EI (continued)	MEAPAQLLFLLLWLPDTTG	99
	MEAPAQLLFLLLWLPDTTG	100
	MEAPAQLLFLLLWLTDTTG	101
	MEPWKPQHSHFFLLLWLPDTTG	102
	MLPSQLIGFLLLWVPASRG	105
	MLPSQLIGFLLLWVPASRG	106
	MVSPLQFLRLLLWVPASRG	107
	MDFHVQIFSFM LISVTILSSG	231
	MDFQMQUIISLLISVTIVSNG	232
	MDFQVQIFSFLISVTILTNG	233
	MDMRAPAQFLGILLWF PARC	234
	MNFHVQIFSFM LISVTIGSSG	235
	MTMLSLVLLSFLLCSRA	236
	MVSTPQFLVFLFWIPACG	237
	TELICVFLFLSVTAILSSG	238
EI	MDCGISLVFLVLILKVC	239
EM	MDMWVQIFSLLICVTSKG	240
EN	LLISVTIMSRG	241
	MDFQVQIFSFLISASIMSRG	242
	MDFQVQIFSFLISISVMSRG	243
	MDFQVQIFSFLISVSIMSRG	244
	MDLQVQIISFLLIIVTIMSRG	245
ET	MGSQVHLLSFLLLWISDTRA	104
	MGEQRIRSCHATSGAESAR	246
	MGSQVHLLSFLLLWISDTRA	247
	MTMFSLALLSLLLLCVSSRA	248
	MTMLSLAPLLSLLLSRA	249
	MXTMDEHESGAVTPHQVLKSRA	250
EV	MDWTWRILFLVAAATGAHS	28
	MDWTWRILFLVAAATGAHS	29
	MDWTWRILFLVAAATSAHS	30
	MDWTWSILFLVAAPTGAHS	31
	MDCTWRILFLVAAATGTHA	32
	MDWTWRILFLVAAATDAYS	33
	MDWTWRVFCLLAVAPGAHS	34
	MDWIWRILFLVGAATGAHS	35
	MELGLSWVFLVAILEGVQC	38
	MELGLSWIFLLAILKGVQC	39
	MEFGLSWVFLVAIIKGVQC	40
	MELGLSWVFLVAILEGVQC	41
	MEFGLSWIFLAAILKGVQC	42
	MEFGLSWVFLVAILKGVQC	43
	MELGLRWVFLVAILEGVQC	44
	MEFGLSWLFLVAILKGVQC	45
	MEFGLSWVFLVALLRGVQC	46
	MEFGLSWVFLVALLRGVQC	47
	MEFGLSWVFLVAILKGVQC	48
	MELGLCWVFLVAILEGVQC	49
	MEFGLSWVFLVAILKGVQC	50

- 20 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
EV (continued)	MEFWLSWVFLVAILKGVQC	51
	MTEFGLSWVFLVAIFKGVQC	52
	MEFGLSWVFLVAILKGVQC	53
	MEFGLSWVFLVVILQGVQC	54
	MEFGLSWVFLVAILKGVQC	55
	MGSTAILALLAVLQGVCS	64
	MGSTAILGLLAVLQGVCA	65
	IKWSWIFLFLLSGTAVHS	251
	IKWSWISLFLLSGTAVHS	252
	LILKVQC	253
	LVLKVQC	254
	MAVVTGKGLPSPKLEVNS	255
	MDFGLIFFIVALLKVQC	256
	MDFGLSLVFLVLILKVQC	257
	MDMRASAQFHGILLWFPARC	258
	MEWELSLIFIFALLKDVQC	259
	MEWSCIFLFLLSVTAVHS	260
	MEWSCIFLFLLSVTAIHS	261
	MEWSWIFLFLLSGTAVLS	262
	MGWNWIFILILSVTTALS	263
	MGWSCIILFLVATATVHS	264
	MGWNWIFILILSVTTVHS	265
	MGWSCIMLFLAATATVHS	266
	MGWSWIFFLFLSGTAVLS	267
	MGWSWIFLFLFLSGTAVLS	268
	MGWSWIFLFLLSGSVLS	269
	MGWSWIFLFLLSGSVHS	270
	MGWSWIFLFLLSGTAVHS	271
	MGWSWIFLFLLSGTAVLS	509
	MGWSWIFLFLLSGTAVLS	510
	MGWSWIFLLFLSGTAVLS	511
	MGWSWIFLLFLSGTAVHS	512
	MGWSWIFLLFLSGTAVLS	513
	MGWSWVFLSFLSGTAVLS	514
	MKCSWVIFFLMAVVIINS	515
	MKLWLNWILLVALLNIQC	516
	MLLGLKWVFFVVFYQVHC	517
	MLLGLKWVFFVVFYQGVHC	518
	MMVLSLLYLLTALPGILS	519
	MNFGLSLIFLVLILKVQC	520
	MQLGHLLPDGSVNS	521
	MVSETHVLIFLLWVSVHC	522
	RSVPTQLLGLLLWLTVNS	523
GH	MGWSYIILFLVATAT	272
GI	IDINVQIFRFLISVTSSG	273
GK	MNMLTQLLGLLLWFA	274
GR	MRTPAHFLGLLLCFL	275
HV	MRWSCIIILFLVATATVHS	276
IG	MNFHVQIFSFM LISVT	277

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
IH	MEWSCIFLFLLSVTA	278
II	MDFQVQIFQIPVKQCL	279
	MDFQVQIFSFLISAS	280
	MKFPSQLLLFLLFRIT	281
IK	MDMRTPAQFLGILLWFP	282
IL	MAVLALLFCLVTFPS	283
	MDFHVQIFSFMILSVT	284
	MDFQVQIFSFLISAS	285
	MDFQVQIFSFLISR	286
	MDFQVQIFSFLISVT	287
	TELICVFLFLSVTA	288
IM	LLISVT	289
	MDFQVQIFSFLISAS	290
	MDFQVQIFSFLISVS	291
	MDFQVQIFSFLISVS	292
	MDFQVQIFSFLMSAS	293
	MDLQVQIISFLIIVT	294
	MHFQVQIFSFLISAS	295
IN	MKCSWVIFFLMAVVI	296
IQ	MKLWLNWILLVALLN	297
IR	MDMRAPAQFFGILLWFP	298
IS	MIYSLQLLRMLVLWIP	299
	MMSPVHSIFILLWIV	300
	MSYSLQLLRMLVLWIP	301
IT	MSYSLQLLRMLVLWIP	302
IV	MDFQMQUIISLLISVT	303
KN	MDFQVQIFQIPVKQCLIISRG	304
LM	MDFQVQIFSFLISAS	305
LP	MAWVSFYLLPFIFSTGLCA	127
	MAWTQLLLFPLLHWTGSL	128
	MAWTPLLFLTLLHCTGSL	129
LR	MRPTLSFLGSCCSSLI	306
MI	MKFPSQLLLLLLFGIP	307
NF	MAW AP LLLTLLAHCTGSWA	133
NI	MDMRVPAQLLGLLLWLWGARC	68
	MDMRVPAQLLGLLLWLWGARC	69
	MDMRVPAQLLGLLQLWLSGARC	70
	MDMRVPAQLLGLLLWLWGARC	71
	MDMRVPAQLLGLLLWLPGDTRC	72
	MDMRVPAQLLGLLLWFPGARC	73
	MDMRVPAQLLGLLLWFPGARC	74
	MDMRVLAQLLGLLLCFPGARC	75
	MDMRVLAQLLGLLLCFPGARC	76
	MDMRVPAQLLGLLLWLPGARC	77
	MDMRVPAQLLGLLLWLPGARC	78
	MDMRVPAQLLGLLLWFPGSRC	79
	MDMRVPAQLLGLLLWFPGSRC	80
	MDMRVPAQLLGLLLWLPGARC	81

- 22 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
NI (continued)	MDMRVPAQRLGLLLLWFPGARC	82
	MRVPAQLLGLLLLWLPGARC	83
	MDMRVPAQLLGLLLLWLPGARC	84
	MDMRVPAQLLGLLLLWLPGARC	85
	MDMRVPAQLLGLLLLWLPGAKC	86
	MESDTLLLWVLLLWVPSTS	308
	MESQTLVFISILLWLYADG	309
	MESQTQVFLSLLLWVSTCG	310
	METDTLLLWVLLLWVPSTG	311
PV	MGWSCIMLFLAATATVHS	312
	MGWSCIMLFLAATATGVHS	313
QA	MD M RVP AQ LLGLLLLWLRGARC	68
	MD M RVP AQ LLGLLLLWLRGARC	69
	MD M RVP AQ LLGLLQLWLSGARC	70
	MD M RVP AQ LLGLLLLWLSGARC	71
	MD M RVP AQ LLGLLLLWLPDTRC	72
	MD M RVP AQ LLGLLLLWFPGARC	73
	MD M RVP AQ LLGLLLLWFPGARC	74
	MD M RVL AQ LLGLLLLCFPGARC	75
	MD M RVL AQ LLGLLLLCFPGARC	76
	MD M RVP AQ LLGLLLLWLPGARC	77
	MD M RVP AQ LLGLLLLWLPGARC	78
	MD M RVP AQ LLGLLLLWFPGSRC	79
	MD M RVP AQ LLGLLLLWFPGSRC	80
	MD M RVP AQ LLGLLLLWLPGARC	81
	MD M RVP AQ RLGLLLLWFPGARC	82
	M RVP AQ LLGLLLLWLPGARC	83
	MD M RVP AQ LLGLLLLWLPGARC	84
	MD M RVP AQ LLGLLLLWLPGARC	85
	MD M RVP AQ LLGLLLLWLPGAKC	86
	MAW TP LLLLLSHCTGSLS	130
	MAW TP LLLFLSHCTGSLS	131
	MAW TL LLLVLLSHCTGSLS	132
	MAW TP LFLFLTCCPGSNS	134
	MAW TP LFLFLTCCPGSNS	135
	MAWISLILSLLALSSAIS	314
	IGWSYIILLVATATVHS	315
	MAWTSILSLLALCSASS	316
	MAWTSILSLLALCSAIS	317
	MGWSCVLLFLVSGTAVLC	318
QI	MDT LC STLLLLTIPSWVLS	36
	MDT LC YTLLLLTTPSWVLS	37
	MDFQVQIFSFLISASIISRG	319
	MDFQVQIFSFLISASILFRG	320
	MDFQVQIFSFLISASILSRG	321
	MDFQVQIFSFLISASIMSRG	322
	MDFQVQIFSFLISASLMSRG	323
	MDFQVQIFSFLISRILSRG	324
	MDFQVQIFSFLISVSIMSRG	325

- 23 -

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
QI (continued)	MDFQVQIFSFLMSASIMSRG	326
	MDTLCSTLLLLTIPSWVLS	327
	MGWSWIFLFLLSGTAVHC	328
	MHFQVQIFSFLISASIMSRG	329
QL	MKH LW FLLLVAAPRWVLS	56
	MKH LW FLLLVAAPRWVLS	57
	MKH LW FLLLVAAPRWVLP	58
	MKH LW FLLLVAAPRWVLS	59
	MKH LW FLLLVAAPRWVLS	60
	MKH LW FLLLVAAPRWVLS	61
	MKH LW FLLLVAAPRWVLS	62
	MKH LW FLLLVAAPRWVLS	63
	MAW VS FYLLPFIFSTGLCA	127
	M A WTQ LL LFPLLLHWTGSLS	128
	M AWT PL LFTLLLHCTGSLS	129
	MAWTPLFFFFVLHCSSFS	330
QM	MDWTWRILFLVAAATGAHS	28
	MDWTWRILFLVAAATGAHS	29
	MDWTWRILFLVAAATSAHS	30
	MDWTWSILFLVAAPTGAHS	31
	MDCTWRILFLVAAATGTHA	32
	MDWTWRILFLVAAATDAYS	33
	MDWTWRVFCLLAVAPGAHS	34
	MDWIWRILFLVGAATGAHS	35
	MRVLGFLCLVTVLPGSLS	331
QP	MAWVSFYLLPFIFSTGLCA	127
	MAWTQLLLLFPLLHWTGSLS	128
	MAWTPLLFTLLLHCTGSLS	129
	MAWTPLLLLLSHCTGSLS	130
	MAWTPLLLLFLSHCTGSLS	131
	MAWTLLLLVLLSHCTGSLS	132
	MAWAPLLLTLLSLLTGSLS	137
	MAWTPLFFFFLLHCSSFS	332
QS	MAWSPLFLTLLTHCAGSWA	108
	MAWSPLLLTLLAHCTGSWA	109
	MASFPLLLTLLTHCAGSWA	110
	MAGFPLLLTLLTHCAGSWA	111
	MTCSPLLLTLIHCTGSWA	112
	MAWALLLTLLTQGTGSWA	113
	MAWALLLSLLTQGTGSWA	114
	MAWALLLTLLTQGTGSWA	115
	MAWALLLTLLTQGTGSWA	116
	MAWALLLTLLTQDTGSWA	117
QT	MAWTPLFLFLTCCPGSNS	134
	MAWTPLFLFLTCCPGSNS	135
	MAWMMLLLGLLAYGSGVDS	136
QV	MDWTWRILFLVAAATGAHS	28
	MDWTWRILFLVAAATGAHS	29
	MDWTWRILFLVAAATSAHS	30

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
QV (continued)	MDWTWSILFLVAAPTGAHS	31
	MDCTWRILFLVAAATGTHA	32
	MDWTWRILFLVAAATDAYS	33
	MDWTWRVFCLLAVAPGAHS	34
	MDWIWRILFLVGAATGAHS	35
	MDTLCSTLLLLTIPSWVLS	36
	MDTLCYTLLLLTTPSWVLS	37
	MELGLSWVFLVAILEGVQC	38
	MELGLSWIFLLAILKGVQC	39
	MEFGLSWVFLVAIIKGVQC	40
	MELGLSWVFLVAILEGVQC	41
	MEFGLSWIFLAAILKGVQC	42
	MEFGLSWVFLVAILKGVQC	43
	MELGLRWVFLVAILEGVQC	44
	MEFGLSWLFLVAILKGVQC	45
	MEFGLSWVFLVALLRGVQC	46
	MEFGLSWVFLVALLRGVQC	47
	MEFGLSWVFLVAILKGVQC	48
	MELGLCWVFLVAILEGVQC	49
	MEFGLSWVFLVAILKGVQC	50
	MEFWLSWVFLVAILKGVQC	51
	MTEFGLSWVFLVAIFKGVQC	52
	MEFGLSWVFLVAILKGVQC	53
	MEFGLSWVFLVVILQGVQC	54
	MEFGLSWVFLVAILKGVQC	55
	MKHLWFFLLLVAAPRWVLS	56
	MKHLWFFLLLVAAPRWVLS	57
	MKHLWFFLLLVAAPRWVLP	58
	MKHLWFFLLLVAAPRWVLS	59
	MKHLWFFLLLVAAPRWVLS	60
	MKHLWFFLLLVAAPRWVLS	61
	MKHLWFFLLLVAAPRWVLS	62
	MKHLWFFLLLVAAPRWVLS	63
	MSVSFLIFLPVLGLPWGVLS	66
	MDWTWRILFLVAAATGAHS	67
	IFLFLLSITAVHC	333
	KGGSCVSLFLVATANVHF	334
	MAVLALLFCLVTFPSILS	335
	MAVLGLLFCLVTFPSVLS	336
	MAVLGLLLCLVTFPSVLS	337
	MAWSWVFLFFLSVTTVHS	338
	MDWIWIMLHLLAATGIQS	339
	MECSWVFLFLLSLTAVHC	340
	MEFGLSWVFLVALLRGVQC	341
	MEWLXXFLLFLSLTAVHC	342
	MEWSGVFIFLLSVTAVHS	343
	MEWSGVFIFLLSVTAVYS	524
	MEWSRVFIFLLSVTAVHS	525
	MEWSWVFLFFLSVTTVHS	526

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
QV (continued)	MEWSWVFLFLLSLTSVHS	527
	MGRLTFSFLLLLPVPAVLS	528
	MGWSCIIFFLVATATVHF	529
	MGWSCIIILFLVAAANVHS	530
	MGWSCIIILFLVAAATVHS	531
	MGWSCIIILFLVATATVHS	532
	MGWSCIIILFLVATATVHS	533
	MGWSCIIILFLVSTATVHS	534
	MGWSCIIILVAAATVHS	535
	MGWSCIIILVAAATVHS	536
	MGWSCIIILVAAATVQF	537
	MGWSCIMLFLAARATVHS	538
	MGWSCIMLFLAATATVHF	539
	MGWSCIMLFLAATATVHF	540
	MGWSCIMLFLAATATVHS	541
	MGWSCIMLFLAATATVHS	542
	MGWSCIMLFLAATATVHS	543
	MGWSFLPLFLAATATGVHS	544
	MGWSRIFLFLLSITAVHC	545
	MGWSSIILFLVATATVHS	546
	MGWSWIFPFLLSGTAVHC	547
	MGWSYIIFFLVATATVHF	548
	MGWSYIIFFLVATATVHS	549
	MGWSYIILFLVATATGHS	550
	MGWSYIILFLVATATVHS	551
	MGWSYIILFLVATATVNS	552
	MRWSCIIILFLVATATVHS	553
SA	METPASFLCLLLLWTT	344
SF	MAWTPLFFFFLLHCS	345
	MAWTPLFFFFVLHCS	346
SI	MKSQTQVFIFLLCVSAHG	347
	MKSQTQVFVFLLCVSAHG	348
SK	MDMWVQIFSLLLICVT	349
SR	MXTMDEHESGAVTPHQVLK	350
	MGEQRIRCHATSGAE	351
	MNLPVHLLVLLFWIP	352
	MTMFSLALLSLLLCVS	353
	MTMLSLAPLLSLLL	354
	MTMLSLVLLSFLLC	355
	MVFTPQILGLMLFWIS	356
	MVLGLKWVFFVVFYQ	357
	MVSTSQLLGLLFWTS	358
	PAQFLFLLVLWIQ	359
SS	MAWIPLFLGVLAYCTGSVA	118
	MAWTALLSLLAHFTGSVA	119
	MAWTPLLPLLTFTCTVSEA	120
	MAWIPLLPLLTCTGSEA	121
	MAWTPLWLTLTLCIGSVV	122
	MAWTVLLGLLSHCTGSVT	123

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
SS (continued)	MAWATLLLPLLNLYTGSIA	124
	MAWIPLLLPLLTCTGSEA	125
	MAWIPLLLPLLILCTVSVA	126
	IDINVQIFRLLISVT	360
	MKLPVLLVLLLFTSP	361
	MKLPVRLLVLMFWIP	362
	MKLPVRLLVLMFWIP	363
ST	MEKDTLLLWVLLWVP	364
	MESDTLLLWVLLWVP	365
	MESDTLLLWVLLWVP	366
	METDPLLWVLLWVP	367
	METDTILLWVLLWVP	368
	METDTLLLWVLLWVP	369
	METDTLLLWVLLWVP	370
	METDTLLLWVLLWVP	371
	MRFSAQLLGLLVWIP	372
	MVFTPQILGLMLFWIS	373
SY	MAWIPLFLGVLAYCTGSVA	118
	MAWTALLSLLAHFTGSVA	119
	MAWTPLLPLLTCTVSEA	120
	MAWIPLLLPLLTCTGSEA	121
	MAWTPLWLTLLTLCIGSVV	122
	MAWTVLLGLLSHCTGSVT	123
	MAWATLLLPLLNLYTGSIA	124
	MAWIPLLLPLLTCTGSEA	125
	MAWIPLLLPLLILCTVSVA	126
TC	MDSQAQVLILLWVS	374
	MDSQAQVLMLLLSVS	375
	MDSQAQVLMLLWVS	376
	MDSQARVLMLLWVS	377
	MESQNHVLMFLLWVS	378
	MESQTHVLMFLLWVS	379
	MESQTQVFLSLLWVS	380
	MESQTQVLISLLFWVS	381
	MESQTQVLMSELLFWVS	382
TG	METPASFLCLLLWTTSAV	383
TN	QHGHEGLCSVSWVPVA	384
	MMSPAQFLFLLVLWIQ	385
	MMSPAQFLFLLVLWIR	386
	MMSPVQFLFLLMLWIQ	387
TR	MIASAQFLGLLLCFQ	388
	MMSSAQFLGLLLCFQ	389
	MMSSAQFLGLLLCFQ	390
	MDMRAPAQIFGFLLLFQ	391
	MDMRVPAHVFGFLLLWFP	392
VC	MDCGISLVFLVLILK	393
VD	MEFQTQVFVFLWLS	394
	MESQIQAFVFVFLWLS	395
	MESQIQVFVFVFLWLS	396

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
VD (continued)	MESQTQVFVYMLLWLS	397
	MGFKMESHTQAFVFAFLWLS	398
VE	METHSQVFVYMLLWLS	399
VH	MEWLXXFLLFLSLTA	400
	MEWSCIFLFLLSVTA	401
	MEWSGVFIFLLSVTA	402
	MEWSRVFIFLLSVTA	403
	MEWSWVFLFFLSVTT	404
	MEWSWVFLFLLSLTS	405
	MGWNWIFILISVTT	406
	MGWSCIIFFLVATAT	407
	MGWSCIIFLVAAAN	408
	MGWSCIIFLVAAAT	409
	MGWSCIIFLVATAT	410
	MGWSCIIFLVATAT	411
	MGWSCIIFLVATAT	412
	MGWSCIIFLVSTAT	413
	MGWSCIIILVAAAT	414
	MGWSCIIILVAAAT	415
	MGWSCIMLFLAARAT	416
	MGWSCIMLFLAATAT	417
	MGWSCIMLFLAATAT	418
	MGWSCIMLFLAATAT	419
	MGWSCIMLFLAATAT	420
	MGWSCIMLFLAATAT	421
	MGWSCIMLFLAATAT	422
	MGWSCIMLFLAATAT	423
	MGWSRIFLFLLSITA	424
	MGWSSIILFLVATAT	425
	MGWSWIFLFLLSGSA	426
	MGWSWIFLFLLSGTA	427
	MGWSWIFLFLLSGTA	428
	MGWSWIFLFLLSGTA	429
	MGWSWIFPFLLSGTA	430
	MGWSYIIFFLVATAT	431
	MGWSYIIFFLVATAT	432
	MGWSYIILFLVATAT	433
	MGWSYIILFLVATAT	434
	MGWSYIILFLVATAT	435
	MLLGLKWVFFVVFYQ	436
	MRWSCIIILVATAT	437
	MRWSCIIILVATAT	438
	MVSETHVLIFLLWVS	439
	IFLFLLSITA	440
	IGWSYIILLVATAT	441
	IKWSWIFLFLLSGTA	442
	IKWSWISLFLLSGTA	443
	KGGSCVSLFLVATAN	444
	MAWSWVFLFFLSVTT	445

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
VH (continued)	MECSWVFLFLLSLTA	446
VI	MDMRVPAQLLGLLLWLRGARC	68
	MDMRVPAQLLGLLLWLRGARC	69
	MDMRVPAQLLGLLQLWLSGARC	70
	MDMRVPAQLLGLLLWLSGARC	71
	MDMRVPAQLLGLLLWLPDTRC	72
	MDMRVPAQLLGLLLWFPGARC	73
	MDMRVPAQLLGLLLWFPGARC	74
	MDMRVLAQLLGLLLCFPGARC	75
	MDMRVLAQLLGLLLCFPGARC	76
	MDMRVPAQLLGLLLWLPGARC	77
	MDMRVPAQLLGLLLWLPGARC	78
	MDMRVPAQLLGLLLWFPGSRC	79
	MDMRVPAQLLGLLLWFPGSRC	80
	MDMRVPAQLLGLLLWLPGARC	81
	MDMRVPAQRLGLLLWFPGARC	82
	MRVPAQLLGLLLWLPGARC	83
	MDMRVPAQLLGLLLWLPGARC	84
	MDMRVPAQLLGLLLWLPGARC	85
	MDMRVPAQLLGLLLWLPGAKC	86
	MIYSLQLLRMLVLWIPISK	447
	MSYSLQLLRMLVLWIPISK	448
	MSYSLQLLRMLVLWIPITK	449
VL	LVLK	450
	MAVLGLLFCLVTFPS	451
	MAVLGLLLCLVTFPS	452
	MDRLTSSFLLIIVPA	453
	MEWSWIFLFLLSGTA	454
	MGRLTFSFLLLPVPA	455
	MGWSCVLLFLVSGTA	456
	MGWSWIFFFLLSGTA	457
	MGWSWIFLFFLSGTA	458
	MGWSWIFLFLLSGSA	459
	MGWSWIFLFLLSGTA	460
	MGWSWIFLFLLSGTA	461
	MGWSWIFLLFLSGTA	462
	MGWSWIFLLFLSGTA	463
	MGWSWVFLSFLSGTA	464
VM	MDFQVQIFSFLISIS	465
VN	MAVVTGKGLPSPKLE	466
	MGWSYIILFLVATAT	467
	MLGHLLPDGS	468
	RSVPTQLLGLLLWLT	469
VQ	LILK	470
	LVLK	471
	MDFGLIFFIVALLK	141
	MDFGLSLVFLVLILK	02
	MGWSCHILILVAAAT	05
	MNFGRLIFLVLTLK	20

second polypeptide starts with amino acids	first peptide (signal sequence) amino acid sequence	SEQ ID NO:
VQ (continued)	MNFGLSLIFLVLILK	21
VR	MGVPTQLLLLWLT	22
	MRVLAELLGLLLFCFL	23
	MRVLPEFLGLLLWIS	24
VS	MRCSLQFLGVLMFWIS	25
	REWSWNFLFLLSGTT	26
VY	MEWSGVFIFLLSVTA	27
if the combination of the first two amino acids of the second polypeptide is not listed in this table these first polypeptides may be used	ELWVLMVWVP	142
	ELWVLMVWVPSTS	143
	HDHALTSSSPQPSSPLCL	144
	LAVITSNIWFPMVCMS	145
	MDMWTSAQFLGILLWFLARC	146
	MDRLTSSFLLIIVPAVLS	147
	MLRAIKAAPFSRFGCS	148
	MRAPAPFLGLLLFCFLARC	149
	MRCSPHFLELLVFWIL	150
	MRPTLSFLGSCSSLILRC	151
	MRTPAHFLGLLLCFLGRC	152
	MRTPAFLGLLLFCFSARC	153
	MSLLTQLQGLLLWLTDRC	154
	MSLPTQLQGLLLWLTARC	155
	MTMLSLAPLLSLLLCVS	156
	MTSLSQLLGMLMLQSL	157
	MVFAPQILGFLLWIS	158
	MVFTPHILGLLFWIS	159
	QHGHEGLCSVSWVPVATNS	160
	REWSWNFLFLLSGTTVSS	161
	TDFHMQIFSFMLISFTARC	162

5 If the dipeptide of the first two amino acids of the second polypeptide is not explicitly listed in Table 1, and no sequence as listed in the last row of Table 1 is intended to be used, it is beneficial not to string the first polypeptide and the second polypeptide directly together. In such a case it is favorable to insert a short sequence of up to five amino acids to resemble the beginning of the immunoglobulin FR1 region sequence which would/could naturally follow the first polypeptide. This sequence can be a single amino acid or a dipeptide, the peptide QIWNN (SEQ ID NO: 472) or a fragment thereof to resemble the first two amino acids of the naturally following immunoglobulin FR1 region. In one embodiment, this sequence is the peptide QIWNN (SEQ ID NO:472), or an N-terminal fraction thereof comprising at least the dipeptide QI.

10

After the first polypeptide or optionally after the inserted short sequence the second polypeptide comprises a heterologous polypeptide. This heterologous polypeptide

- 30 -

has an amino acid sequence of from 5 to 500 amino acid residues. In a preferred embodiment of the invention the amino acid sequence is of from 10 to 350 amino acid residues and in a more preferred embodiment of from 15 to 150 amino acid residues. The heterologous polypeptide conjugated to the immunoglobulin is selected from the group comprising biologically active molecules. These molecules exhibit a biological effect when administered to an artificial biological system or a living cell, such as in assay-systems, or to a living organism, such as birds or mammals, including humans. These biologically active compounds comprise, but are not limited to, agonists as well as antagonists of receptors, inhibitors as well as activators of enzymes, and the like, and also peptides, polypeptides, and proteins exhibiting cytotoxic, antiviral, antibacterial, or anti-cancer activity, as well as antigens. The biological effect can be, but is not limited, to enzyme inhibition, binding to a receptor, either at the binding site or circumferential, and signal triggering. These biologically active compounds are, for example, useful for pharmaceutical, therapeutic, or diagnostic applications.

The second polypeptide further comprises after the heterologous polypeptide a linker. Linkers that can preferably be used with the current invention are listed in Table 2.

Table 2: Possible linkers.

linker No.	linker amino acid sequence	SEQ ID NO:
1	[Ser(Gly) ₄] ₃	06
2	[Ser(Gly) ₄] ₅	07
3	[Gly(Gln) ₄] ₃	08
4	Gly(Ser) ₁₅ Gly	09
5	GST	10
6	[(Gly) ₄ Ser] ₃ -Gly-Ala-Ser	139
7	Gly(Ser) ₁₅ Gly-Ala-Ser	140
8	[(Gly) ₄ Ser] ₃ -Gly	554
9	[(Gly) ₄ Ser] ₅ -Gly	555
10	[(Gly) ₄ Ser] ₃ -Gly ₂	556
11	[(Gly) ₄ Ser] ₅ -Gly ₂	557

After the linker an immunoglobulin fragment follows as the carboxy-terminal part of the second polypeptide.

The second polypeptide comprises a heterologous polypeptide followed by a linker and followed by an immunoglobulin fragment as carboxy-terminal part, i.e. a nucleic acid encoding the second polypeptide comprises in a 5' to 3' direction

nucleic acids encoding a heterologous polypeptide, a linker, and an immunoglobulin fragment.

5 Immunoglobulin molecules are assigned to five different classes: IgA (Immunoglobulin A), IgD, IgE, IgG, and IgM. Of these IgG and IgE are more frequently used in pharmaceutic and diagnostic applications. Within these classes the immunoglobulins differ in their over-all structure but the building blocks are similar. All immunoglobulins are built up of two different polypeptide chains, a light chain and a heavy chain.

10 An immunoglobulin fragment comprises the carboxy-terminal constant domain(s) of an immunoglobulin light or heavy chain, e.g. it comprises either at least the C_{H1} -, C_{H2} -, C_{H3} -domain and the hinge-region of an immunoglobulin heavy chain and optionally a C_{H4} -domain, or the C_L -domain of an immunoglobulin light chain. The immunoglobulin from which the fragment is derived can be a naturally occurring or a synthetic immunoglobulin. In one embodiment of the invention the
15 immunoglobulin fragment additionally contains a fragment of a heavy or light chain variable domain or of a variant thereof. In the variable domain fragment amino acid(s) or region(s) are deleted. In one embodiment of from one to six amino acids of the variable domain are deleted. In another embodiment of from one to six regions of the variable domain are deleted. In a further embodiment the
20 variable domain is deleted. The presence of a functional, i.e. antigen recognizing, variable domain in the immunoglobulin fragment is not essential for the current invention. A not functionable immunoglobulin according to the invention is an immunoglobulin not possessing an antigen recognizing variable domain. In one
25 embodiment, the immunoglobulin fragment is a not functionable immunoglobulin. In one embodiment, the variable and constant domains comprised in the immunoglobulin fragment are of/derived from the same antibody, i.e. belong to the same antibody.

30 The different nucleic acid sequences are operably linked on an expression plasmid. For expression the plasmid is introduced into a host cell. Proteins are preferably produced in mammalian cells such as CHO cells, NS0 cells, Sp2/0 cells, COS cells, HEK cells, K562 cells, BHK cells, PER.C6 cells, and the like.

Description of the Figures:

- 5 Figure 1 Common structure of immunoglobulins of the IgG class.
- Figure 2 Plasmid map of the anti-IGF-1R γ 1-heavy chain expression vector
 4818.
- Figure 3 Plasmid map of the anti-IGF-1R κ -light chain expression vector
 4802.
- 10 Figure 4 Plasmid map of the γ 1-heavy chain constant region gene vector
 4962.
- Figure 5 Plasmid map of the modified anti-IGF-1R κ -light chain
 expression vector 4964.
- Figure 6 Plasmid map of the modified anti-IGF-1R light chain expression
15 vector 4963.
- Figure 7 Coomassie Blue stained SDS-PAGE-gels of affinity purified
 immunoglobulin conjugates; sample arrangement according to
 Table 6.
- Figure 8 Immunodetection of the light chain in cell culture supernatants
20 after transient expression in HEK293 EBNA cells; sample
 arrangement according to Table 6.
- Figure 9 Immunodetection of the heavy chain in cell culture supernatants
 after transient expression in HEK293 EBNA cells; sample
 arrangement according to Table 6.

25 ExamplesMaterial & Methods

30 General information regarding the nucleotide sequences of human
 immunoglobulins light and heavy chains is given in: Kabat, E.A., et al., (1991)
 Sequences of Proteins of Immunological Interest, fifth ed., NIH Publication No 91-
 3242.

Amino acids of antibody chains are numbered according to EU numbering
(Edelman, G.M., et al., PNAS 63 (1969) 78-85; Kabat, E.A., et al., (1991) Sequences
of Proteins of Immunological Interest, Fifth Ed., NIH Publication No 91-3242).

- 33 -

Recombinant DNA techniques

Standard methods were used to manipulate DNA as described in Sambrook, J., et al., Molecular cloning: A laboratory manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. The molecular biological reagents were used according to the manufacturer's instructions.

Protein determination

The protein concentration was determined by determining the optical density (OD) at 280 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence.

DNA sequence determination

DNA sequences were determined by double strand sequencing performed at MediGenomix GmbH (Martinsried, Germany).

DNA and protein sequence analysis and sequence data management

The GCG's (Genetics Computer Group, Madison, Wisconsin) software package version 10.2 and Infomax's Vector NTI Advance suite version 8.0 was used for sequence creation, mapping, analysis, annotation, and illustration.

Gene synthesis

Desired gene segments were prepared by Medigenomix GmbH (Martinsried, Germany) from oligonucleotides made by chemical synthesis. The 100 - 600 bp long gene segments which are flanked by singular restriction endonuclease cleavage sites were assembled by annealing and ligation of oligonucleotides including PCR amplification and subsequently cloned into the pCR2.1-TOPO-TA cloning vector (Invitrogen Corp., USA) via A-overhangs. The DNA sequence of the subcloned gene fragments were confirmed by DNA sequencing.

Affinity purification of immunoglobulin conjugates

The expressed and secreted immunoglobulin conjugates were purified by affinity chromatography using Protein A-SepharoseTM CL-4B (GE Healthcare former Amersham Bioscience, Sweden) according to known methods. Briefly, after centrifugation (10,000 g for 10 minutes) and filtration through a 0.45 µm filter the immunoglobulin conjugate containing clarified culture supernatants were applied on a Protein A-SepharoseTM CL-4B column equilibrated with PBS buffer (10 mM Na₂HPO₄, 1 mM KH₂PO₄, 137 mM NaCl and 2.7 mM KCl, pH 7.4). Unbound proteins were washed out with PBS equilibration buffer and 0.1 M citrate buffer, pH 5.5. The immunoglobulin conjugates were eluted with 0.1 M citrate buffer, pH

- 34 -

3.0, and the immunoglobulin conjugate containing fractions were neutralized with 1 M Tris-Base. Then, the immunoglobulin conjugates were extensively dialyzed against PBS buffer at 4°C, concentrated with an ultrafree centrifugal filter device equipped with a Biomax-SK membrane (Millipore Corp., USA) and stored in an ice-water bath at 0°C.

Example 1

Making of the expression plasmids

The gene segments encoding an anti-insulin-like growth factor I receptor (IGF-1R) antibody light chain variable domain (region) (V_L) and the human kappa-light chain constant domain (region) (C_L) were joined as were gene segments for the anti-IGF-1R heavy chain variable domain (region) (V_H) and the human gamma1-heavy chain constant region (C_{H1} -Hinge- C_{H2} - C_{H3}).

a) Vector 4818

Vector 4818 is the expression plasmid for the transient expression of an anti-IGF-1R antibody (also denoted as anti-IGF-1R in the following) heavy chain (genomically organized expression cassette; exon-intron organization) in HEK293 EBNA cells (for sequences see US 2005/0008642). It comprises the following functional elements:

Beside the anti-IGF-1R $\gamma 1$ -heavy chain expression cassette this vector contains:

- a hygromycin resistance gene as a selectable marker,
- an origin of replication, oriP, of Epstein-Barr virus (EBV),
- an origin of replication from the vector pUC18 which allows replication of this plasmid in *E. coli*, and
- a beta-lactamase gene which confers ampicillin resistance in *E. coli*.

The transcription unit of the anti-IGF-1R gamma1-heavy gene is composed of the following elements:

- the immediate early enhancer and promoter from the human cytomegalovirus (HCMV),
- a synthetic 5'-untranslated region (UT),
- a murine immunoglobulin heavy chain signal sequence including a signal sequence intron (signal sequence 1, intron, signal sequence 2 [L1-intron-L2]),
- the cloned anti-IGF-1R variable heavy chain encoding segment arranged

- 35 -

with a unique BsmI restriction site at the 5'-end (L2 signal sequence) and a splice donor site and a unique NotI restriction site at the 3'-end,

- a mouse/human heavy chain hybrid intron 2 including the mouse heavy chain enhancer element (part JH₃, JH₄) (Neuberger, M.S., EMBO J. 2 (1983) 1373-1378),
- the genomic human γ 1-heavy gene constant region,
- the human γ 1-immunoglobulin polyadenylation ("poly A") signal sequence, and
- the unique restriction sites AscI and SgrAI at the 5'- and 3'-end, respectively.

The plasmid map of the anti-IGF-1R γ 1-heavy chain expression vector 4818 is shown in Figure 2.

b) Vector 4802

Vector 4802 is the expression plasmid for the transient expression of an anti-IGF-1R antibody light chain (cDNA) in HEK293 EBNA cells. It comprises the following functional elements.

Beside the anti-IGF-1R kappa-light chain expression cassette this vector contains:

- a hygromycin resistance gene as a selectable marker,
- an origin of replication, oriP, of Epstein-Barr virus (EBV),
- an origin of replication from the vector pUC18 which allows replication of this plasmid in E. coli, and
- a β -lactamase gene which confers ampicillin resistance in E. coli.

The transcription unit of the anti-IGF-1R κ -light chain gene is composed of the following elements:

- the immediate early enhancer and promoter from the human cytomegalovirus (HCMV),
- the cloned anti-IGF-1R variable light chain cDNA including
- the native 5'-UT and
- the native light chain signal sequence of the human immunoglobulin germline gene arranged with a unique BglII restriction site at the 5'-end,
- the human κ -light chain gene constant region,
- the human immunoglobulin κ -polyadenylation ("poly A") signal sequence, and

- 36 -

- the unique restriction sites *AscI* and *FseI* at the 5'- and 3'-end, respectively.

5 The plasmid map of the anti-IGF-1R κ -light chain expression vector 4802 is shown in Figure 3.

c) Plasmid 4962

10 Vector 4962 served as basic structure for the assembling of expression plasmids 4965, 4966, and 4967. These plasmids enabled the transient expression of modified antibody heavy chains (N-terminal conjugation without variable domain, cDNA organization) in HEK 293 EBNA cells. Plasmid 4962 comprises the following functional elements.

Beside the expression cassette for the γ 1-heavy chain constant region this vector contains:

- a hygromycin resistance gene as a selectable marker,
- 15 - an origin of replication, *oriP*, of Epstein-Barr virus (EBV),
- an origin of replication from the vector pUC18 which allows replication of this plasmid in *E. coli*, and
- a beta-lactamase gene which confers ampicillin resistance in *E. coli*.

20 The transcription unit of the γ 1-heavy chain constant region gene (C_H1 -Hinge- C_H2 - C_H3) is composed of the following elements:

- the immediate early enhancer and promoter from the human cytomegalovirus (HCMV),
- a synthetic linker (SEQ ID NO: 01) comprising a single *BglII* restriction site at the 5'-end and a single *NheI* restriction site at the 3'-end (*NheI* site within the C_H1 N-terminus)

HCMV-promoter AlaSer (CH1)

...agatcttttgccaccgctagc...

30 *BglII* *NheI*

- the human γ 1-heavy chain gene constant domains (region) (C_H1 -Hinge- C_H2 - C_H3 , cDNA organization),
 - the human γ 1-immunoglobulin polyadenylation ("poly A") signal
- 35 sequence, and

- 37 -

- the unique restriction sites *AscI* and *FseI* at the 5'- and 3'-end, respectively.

5 The plasmid map of the $\gamma 1$ -heavy chain constant domains/region gene vector 4962 is shown in Figure 4.

d) Plasmid 4964

Vector 4964 served as basic structure for the assembling of expression plasmids 4976 and 4977. These plasmids enabled the transient expression of modified anti-IGF-1R antibody light chains (N-terminal conjugation) in HEK 293 EBNA cells.

10 The plasmid 4964 is a variant of expression plasmid 4802.

The transcription unit of the anti-IGF-1R κ -light gene was modified as indicated below:

15 The native light chain signal sequence is replaced by a synthetic linker arranged with a unique *BglII* restriction site at the 5'- and a unique *NheI* restriction site at the 3'-end directly joined to the V_L -IGF-1R variable domain (region) (SEQ ID NO: 03).

		- V_L -1R18
	... <u>agatct</u> atatatatatatg	ctagcgaaattgtgttgaca...
		AlaSerGluIleValLeuThr...
20	<i>BglII</i>	<i>NheI</i>

The plasmid map of the modified anti-IGF-1R κ -light chain expression vector 4964 is shown in Figure 5.

e) Plasmid 4969

25 The expression plasmid 4969 is derived from plasmid 4802 which is an expression plasmid for the anti-IGF-1R antibody light chain. The plasmid encodes a modified antibody light chain fragment (N-terminal conjugation without variable domain; polypeptide-linker-constant region of kappa chain).

30 For the construction of plasmid 4969 a unique *BglII* restriction site was introduced at the 3'-end of the CMV-promoter and a unique *BbsI* restriction site was introduced inside of the constant region of the anti-IGF-1R antibody light chain (SEQ ID NO: 04).

- 38 -

|-- C-kappa BbsI
 cgaactgtggctgcaccatctgtcttcatcttc...
 ArgThrValAlaAlaProSerValPheIlePhe...

5 f) Plasmid 4963

This plasmid enabled the transient expression of anti-IGF-1R antibody light chains in HEK 293 EBNA cells.

The plasmid 4963 is a variant of expression plasmid 4802.

10 The transcription unit of the anti-IGF-1R κ -light gene was modified as indicated below:

- the human κ -light chain constant gene region was slightly modified at the C-kappa-Ig-kappa pA joining region (insertion of a unique HindIII and KasI restriction site, SEQ ID NO: 558).

15 ...C-kappa Ig-kappa-pA
 ...Aaaagcttcaacaggggagagtgt**TGA**agggagaggcgcccca
 ...LysSerPheAsnArgGlyGluCys
 HindIII KasI

20 The plasmid map of the modified anti-IGF-1R light chain expression vector 4963 is shown in Figure 6.

Example 2

Making of the final expression plasmids

25 The immunoglobulin fusion genes (heavy and light chain) comprising the immunoglobulin gene segment, linker gene segment, and polypeptide gene segment have been assembled with known recombinant methods and techniques by connection of the according gene segments (nucleic acids).

The nucleic acid sequences encoding the peptide linkers and polypeptides were each synthesized by chemical synthesis and then ligated into an E.coli plasmid. The subcloned nucleic acid sequences were verified by DNA sequencing.

30 The employed immunoglobulin polypeptide chains, the immunoglobulin fragment, the location of the polypeptide conjugation (N-terminal), the employed linker and the employed polypeptide are listed in Table 2 (page 30), Table 3, and Table 3a.

- 39 -

Table 3: Employed proteins and polypeptides; the amino acid sequence and the numbering of the positions is as in the BH8 reference strain (Locus HIVH3BH8; HIV-1 isolate LAI/IIIB clone BH8 from France; Ratner, L., et al., Nature 313 (1985) 277-384).

proteins and polypeptides	SEQ ID NO:
HIV-1 gp41 (position 507-851 of BH8 gp 160)	11
T-651 (see e.g. US 6,656,906)	12
HIV-1 gp41 ectodomain variant single mutant: I568P	13
HIV-1 gp41 ectodomain variant quadruple mutant: I568P, L550E, L566E, I580E	14

5

Table 3a: Chemically prepared gene segments used for immunoglobulin conjugate gene construction.

Insert	SEQ ID NO:
Insert 4964 (introduction of unique restriction sites)	15
Insert 4965 (with T-651) comprising signal sequence (MDTLCSTLLLLTIPSWVLS), inserted short sequence (QIWNN), heterologous polypeptide (MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL), linker (GGGGSGGGGSGGGGSG)	16
Insert 4966 (with T-651) comprising signal sequence (MDTLCSTLLLLTIPSWVLS), inserted short sequence (QIWNN), heterologous polypeptide (MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL), linker (GGGGSGGGGSGGGGSGGGGSGGGGSG)	17
Insert 4967 (with T-651) comprising signal sequence (MDTLCSTLLLLTIPSWVLS), inserted short sequence (QIWNN), heterologous polypeptide (MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL), linker (GSSSSSSSSSSSSSSSG)	18

- 40 -

Insert	SEQ ID NO:
Insert 4969 (gp41 single mutant) comprising signal peptide (MEFGLSWVFLVALLRGVQC), inserted short sequence (Q), heterologous polypeptide (VQARQLLSGIVQQQNNLLRAIEGQQHLLQLTVWGPQLQARIL AVERYLKDQQLLGIWGC SGKLICTTAVPWNASWSNKSLEQIWNN MTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELL), linker (GGGGSGGGGSGGGGSG)	19

5 The components used for the construction of the final expression plasmids for transient expression of the modified immunoglobulin polypeptide light and heavy chains (the expression cassettes) are listed in Table 4 with respect to the used basis plasmid, cloning site, and inserted nucleic acid sequence encoding the conjugated immunoglobulin polypeptides.

Table 4: Components employed in the construction of the used expression plasmids.

Expression plasmid	Basis vector	Inserted DNA gene segment	Cloning sites
N-terminal conjugation: Heavy chain (without variable domain)			
4965	4962	Insert 4965 (249 Bp)	BglIII / NheI
4966	4962	Insert 4966 (279 Bp)	BglIII / NheI
4967	4962	Insert 4967 (252 Bp)	BglIII / NheI
N-terminal conjugation: Light chain (without variable domain)			
4969	4802	Insert 4969 (589 Bp)	BglIII / BbsI
N-terminal conjugation: Light chain (including the variable domain)			
4976	4964	Insert 4965 (249 Bp)	HindIII / KasI
4977	4964	Insert 4967 (252 Bp)	HindIII / KasI

10 In Table 5 is listed: the used polypeptides with HIV-1 inhibitory properties (T-651 and HIV-1 gp41 ectodomain variants), the used linkers to join the immunoglobulin light or heavy chain with the polypeptide and the deduced molecular weight of the modified antibody chains as deduced from the encoded amino acid sequences.

Table 5: Summary of the employed polypeptides and the deduced molecular weight of the modified immunoglobulin polypeptide chains.

expression plasmid	polypeptide	molecular weight [Da]	linker SEQ ID NO:
Reference plasmids			
4818	anti-IGF-1R heavy chain	49263.5	no linker
4802	anti-IGF-1R light chain	23572.2	no linker
N-terminal fusions: Heavy chain (without variable domain)			
4965	T-651	42227.3	554
4966	T-651	42857.9	555
4967	T-651	42644.7	09
N-terminal fusions: Light chain (without variable domain)			
4969	Gp41 single mutant	27247.3	554
N-terminal fusions: Light chain (including the variable domain)			
4976	T-651	29851.9	139
4977	T-651	30269.2	140

Example 3

5 Transient expression of immunoglobulin variants in HEK293 EBNA cells

Recombinant immunoglobulin variants were generated by transient transfection of adherent growing HEK293-EBNA cells (human embryonic kidney cell line 293 expressing Epstein-Barr-Virus nuclear antigen; American type culture collection deposit number ATCC # CRL-10852) cultivated in DMEM (Dulbecco's modified Eagle's medium, Gibco, Invitrogen Corp., USA) supplemented with 10% ultra-low IgG FCS (fetal calf serum, Gibco, Invitrogen Corp., USA), 2 mM Glutamine (Gibco, Invitrogen Corp., USA), 1% volume by volume (v/v) nonessential amino acids (Gibco, Invitrogen Corp., USA) and 250 µg/ml G418 (Roche Molecular Biochemicals, Roche Diagnostics GmbH, Germany). For transfection Eugene™ 6 Transfection Reagent (Roche Molecular Biochemicals, Roche Diagnostics GmbH, Germany) was used in a ratio of reagent (µl) to DNA (µg) ranging from 3:1 to 6:1.

- 42 -

Immunoglobulin polypeptide light and heavy chains were expressed from two different plasmids using a molar ratio of light chain to heavy chain encoding plasmid from 1:2 to 2:1. Immunoglobulin variants containing cell culture supernatants were harvested at day 4 to 11 after transfection. Supernatants were
5 stored at 0° C in an ice-water bath until purification.

General information regarding the recombinant expression of human immunoglobulins in e.g. HEK293 cells is given in: Meissner, P., et al., Biotechnol. Bioeng. 75 (2001) 197-203.

Example 4

10 Expression analysis using SDS PAGE, Western blotting transfer and detection with immunoglobulin specific antibody conjugates

The expressed and secreted polypeptides were processed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (SDS-PAGE), and the separated polypeptides were transferred to a membrane from the gel and subsequently
15 detected by an immunological method.

SDS-PAGE

LDS sample buffer, fourfold concentrate (4x): 4 g glycerol, 0.682 g Tris-Base, 0.666 g Tris-hydrochloride, 0.8 g LDS (lithium dodecyl sulfate), 0.006 g EDTA (ethylene diamin tetra acid), 0.75 ml of a 1% by weight (w/v) solution of Serva Blue
20 G250 in water, 0.75 ml of a 1% by weight (w/v) solution of phenol red, add water to make a total volume of 10 ml.

The culture broth containing the secreted polypeptide was centrifuged to remove cells and cell debris. An aliquot of the clarified supernatant was admixed with 1/4 volumes (v/v) of 4xLDS sample buffer and 1/10 volume (v/v) of 0.5 M 1,4-dithiotreitol (DTT). Then the samples were incubated for 10 min. at 70°C and
25 protein separated by SDS-PAGE. The NuPAGE® Pre-Cast gel system (Invitrogen Corp., USA) was used according to the manufacturer's instruction. In particular, 10% NuPAGE® Novex® Bis-Tris Pre-Cast gels (pH 6.4) and a NuPAGE® MOPS running buffer was used.

30 **Western blot**

Transfer buffer: 39 mM glycine, 48 mM Tris-hydrochloride, 0.04% by weight (w/v) SDS, and 20% by volume methanol (v/v)

- 43 -

After SDS-PAGE the separated immunoglobulin conjugate polypeptide chains were transferred electrophoretically to a nitrocellulose filter membrane (pore size: 0.45 μ m) according to the „Semidry-Blotting-Method“ of Burnette (Burnette, W.N., Anal. Biochem. 112 (1981) 195-203).

5 **Immunological detection**

TBS-buffer: 50 mM Tris-hydrochloride, 150 mM NaCl, adjusted to pH 7.5

Blocking solution: 1% (w/v) Western Blocking Reagent (Roche Molecular Biochemicals, Roche Diagnostics GmbH, Germany) in TBS-buffer

TBST-Buffer: 1x TBS-buffer with 0.05% by volume (v/v) Tween-20

10 For immunological detection the western blotting membranes were incubated with shaking at room temperature two times for 5 minutes in TBS-buffer and once for 90 minutes in blocking solution.

Detection of the immunoglobulin conjugate polypeptide chains

15 Heavy chain: For detection of the heavy chain or heavy chain fragment containing polypeptides a purified rabbit anti-human IgG antibody conjugated to a peroxidase was used (Code No. P 0214, DAKO, Denmark).

Light chain: Polypeptides containing light chain or light chain fragments were detected with a purified peroxidase conjugated rabbit anti-human kappa light chain antibody (DAKO, Denmark, Code No. P 0129).

20 For visualization of the antibody light and heavy chains or fragments thereof washed and blocked Western blot membranes were first incubated in case of a heavy chain with a purified rabbit anti-human IgG antibody conjugated to a peroxidase or in case of a light chain with a purified peroxidase conjugated rabbit anti-human kappa light chain antibody in a 1:10,000 dilution in 10 ml blocking
25 solution at 4°C with shaking over night. After washing the membranes three times with TBTS-buffer and once with TBS buffer for 10 min. at room temperature, the Western-blot membranes were developed with a Luminol/peroxid-solution generating chemiluminescence (Lumi-Light^{PLUS} Western Blotting Substrate, Roche Molecular Biochemicals, Roche Diagnostics GmbH, Germany). Therefore the
30 membranes were incubated in 10 ml Luminol/peroxid-solution for 10 seconds to 5 minutes and the emitted light was detected afterwards with a Lumi-Imager F1 Analysator (Roche Molecular Biochemicals, Roche Diagnostics GmbH, Germany) and/or was recorded with an x-ray-film.

The intensity of the spots was quantified with the LumiAnalyst Software (Version 3.1).

Multiple-staining of immunoblots

5 The secondary peroxidase-labeled antibody conjugate used for the detection can be removed from the stained blot by incubating the membrane for one hour at 70°C in 1 M Tris-hydrochloride-buffer (pH 6.7) containing 100 mM beta-mercaptoethanol and 20% (w/v) SDS. After this treatment the blot can be stained with a different secondary antibody a second time. Prior to the second detection the blot is washed three times at room temperature with shaking in TBS-buffer for 10 minutes each.

10 The sample arrangement is listed in Table 6.

Table 6: Sample arrangement of SDS PAGE gels / Western blots

sample	expression plasmids		note
	light chain	heavy chain	
MW marker			
anti-IGF-1R (reference Ab), 50 ng			
anti-IGF-1R (reference Ab), 150 ng			
anti-IGF-1R (reference Ab), 500 ng			
HEK293 culture medium			
3	4802 (wt)	4818 (wt)	anti-IGF-1R (reference Ab) control
4	4802 (wt)	4961 (wt)	anti-IGF-1R (reference Ab) control
5	4963 (wt)	4818 (wt)	anti-IGF-1R (reference Ab) control
6	4802 (wt)	4965	N-term; heavy; without V _H
7	4802 (wt)	4966	N-term; heavy; without V _H
8	4802 (wt)	4967	N-term; heavy; without V _H
9	4969	4818 (wt)	N-term; light; without V _L
10	4976	4818 (wt)	N-term; light
11	4977	4918 (wt)	N-term; light

- 45 -

sample	expression plasmids		note
	light chain	heavy chain	
12	4969	4966	N-term; light; without V _L N-term; heavy; without V _H
13	4976	4966	N-term; light; N-term; heavy; without V _H
14	4977	4967	N-term; light N-term; heavy; without V _H

Example 5**Detection of assembled immunoglobulin polypeptides**

5 **Purification and concentration of immunoglobulin polypeptides by affinity binding to Protein A SepharoseTM CL-4B**

10 HEK 293 EBNA cells containing one or more plasmids were cultivated under conditions suitable for the transient expression of the polypeptide gene(s) located on the plasmid(s) for 6 to 10 days. To 1 ml clarified culture supernatant in a 1.8 ml Eppendorf cup 0.1 ml of a Protein A SepharoseTM CL-4B (GE Healthcare former
15 Amersham Biosciences, Sweden) suspension (1:1 (v/v) suspension of Protein A Sepharose in PBS buffer (10 mM Na₂HPO₄, 1 mM KH₂PO₄, 137 mM NaCl and 2.7 mM KCl, pH 7.4)) was added. The suspension was incubated for a time of between one and sixteen hours at room temperature with shaking. Thereafter the Sepharose beads were sedimented by centrifugation (30 s, 5000 rpm) and the supernatant was
20 discarded. The Sepharose pellet was washed subsequently each with 1.6 ml PBS buffer, 1.6 ml 0.1 M citrate buffer pH 5.0, and 1.6 ml distilled water. The protein A bound immunoglobulin was extracted from the Sepharose beads with 0.1 ml 1xLDS-PAGE sample buffer at 70°C for 5 to 10 min. The analysis was done by SDS-PAGE separation and staining with Coomassie brilliant blue as described in example 4.

Results:

Expression/Secretion-analysis of heavy and/or light chain fragment containing polypeptides after transient expression:

25 **Figure 7a-c:** Coomassie Blue stained SDS-PAGE-gels of affinity purified polypeptides; sample arrangement according to Table 6.

Immunodetection of immunoglobulin containing polypeptides:

Figure 8a-c: Immunodetection of light chain fragment containing polypeptides in cell culture supernatants after transient expression in HEK293 EBNA cells.

5 Figure 9a-c: Immunodetection of the heavy chain fragment containing polypeptides in cell culture supernatants after transient expression in HEK293 EBNA cells.

10 From Figure 7 a-c, 8 a-c and 9 a-c it can be deduced that the polypeptides are transiently expressed and secreted into the culture medium. In the case that the immunoglobulin containing polypeptide possesses one or several glycosylation sites the final polypeptides have no exactly defined molecular weight but a molecular weight distribution depending on the extent of glycosylation. This causes in SDS-PAGE that the species all representing one polypeptide do not migrate homogeneously and thus the bands are broadened.

Example 6

15 Quantitation of the expressed heavy chain containing polypeptides with human IgG ELISA

20 The immunoglobulin heavy chain fragment containing polypeptide concentration in cell culture supernatants was determined by a sandwich ELISA which used a biotinylated anti-human IgG F(ab')₂ fragment as the capture reagent and for detection a peroxidase-conjugated anti-human IgG F(ab')₂ antibody fragment.

25 Streptavidin coated 96-well plates (Pierce Reacti-Bind™ Streptavidin Coated Polystyrene Strip Plates, Code No. 15121, Pierce Chemical Company, USA) were coated with 0.5 µg/ml biotinylated goat polyclonal anti-human IgG F(ab')₂ antibody fragment ((F(ab')₂<h-Fcγ>Bi; Dianova, Germany, Code No. 109-066-098) capture antibody (0.1 ml/well) in diluent buffer (diluent buffer: PBS buffer containing 0.5% weight by volume (w/v) bovine serum albumin) by incubation for one hour at room temperature (RT) under shaking. Thereafter, the plates were washed three times with more than 0.3 ml wash buffer (wash buffer: PBS containing 1% weight by volume (w/v) Tween™₂₀). IgG immunoglobulin conjugate
30 containing cell culture supernatants (samples) were diluted serially (twofold) up to a concentration of 0.5–20 ng/ml in diluent buffer, added to plates and incubated for one hour at RT with shaking. Purified anti-IGF-1R standard antibody (0.5 – 20 ng/ml) in diluent buffer was used for the generation of an IgG protein standard

- 47 -

curve. After washing the plates three times with 0.3 ml/well wash buffer, bound complexes to human Fcγ were detected with a peroxidase-conjugated F(ab')₂ fragment of goat polyclonal anti-human F(ab')₂-specific IgG [F(ab')₂<h-Fcγ>POD; Dianova, Code No. 109-036-098]. After washing the plates thrice with 0.3 ml/well wash buffer the plates were developed with ABTS[®] (2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) peroxidase substrate solution (Roche Molecular Biochemicals, Code No. 1684302, Roche Diagnostics GmbH, Germany). After 10-30 minutes the absorbance was measured at 405 nm and 490 nm against a reagent blank (incubation buffer + ABTS[®] solution) on a Tecan Spectrafluorplus plate reader (Tecan Deutschland GmbH, Germany). For background correction the absorbance at 490 nm was subtracted from the absorbance at 405 nm according to formula I. All samples were assayed at least as duplicates, and the values from double or triple absorbance measurements were averaged. The IgG content of the samples were calculated from a standard curve.

Formula I:
$$\Delta A = (A_{sample}^{405} - A_{sample}^{490}) - (A_{blank}^{405} - A_{blank}^{490})$$

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

What is claimed is:

1. Method for the recombinant production of a heterologous polypeptide as a fusion polypeptide comprising the heterologous polypeptide, a linker and an immunoglobulin fragment in a eukaryotic host cell comprising an expression plasmid, wherein
 - a) the expression plasmid comprises in a 5' to 3' direction
 - aa) a promoter,
 - ab) a nucleic acid encoding a first polypeptide which is a signal sequence, whose amino acid sequence is selected from the group consisting of SEQ ID NO: 163, 164, 165, 166, 167, 168, 169, 68, 70, 71, 72, 73, 75, 77, 79, 82, 83, 86, 170, 171, 172, 173, 174, 175, 176, 177, 178, 180, 181, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 194, 138, 195, 87, 89, 91, 93, 95, 103, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 473, 474, 475, 476, 477, 478, 479, 480, 481, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 216, 105, 107, 217, 218, 219, 220, 222, 223, 224, 225, 226, 227, 228, 229, 230, 96, 97, 98, 101, 102, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 104, 246, 247, 248, 249, 250, 28, 30, 31, 32, 33, 34, 35, 38, 39, 40, 42, 43, 44, 45, 46, 49, 51, 52, 54, 64, 65, 251, 252, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 286, 287, 288, 289, 291, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 304, 127, 128, 129, 306, 307, 133, 308, 309, 310, 313, 130, 131, 132, 134, 314, 315, 316, 317, 318, 36, 37, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 56, 58, 330, 331, 137, 332, 108, 109, 110, 111, 112, 113, 114, 117, 136, 66, 67, 333, 334, 335, 336, 337, 338, 339, 340, 342, 343, 524, 525, 526, 527, 528, 529, 530, 531, 534, 535, 537, 538, 539, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 118, 119, 120, 121, 122, 123, 124, 126, 360, 361, 362, 364, 365, 367, 368, 369,

- 49 -

372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 402, 403, 404, 405, 407, 408, 409, 410, 413, 414, 416, 417, 424, 425, 426, 427, 429, 430, 431, 436, 437, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 464, 465, 466, 468, 469, 470, 141, 02, 20, 21, 22, 23, 24, 25, 26, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, and 162, depending on the first two amino acids of a second polypeptide chosen that way, that the first two amino acids of the second polypeptide are identical to the first two amino acids of the amino acid sequences of the naturally following immunoglobulin FR1-region,

- ac) a nucleic acid encoding the second polypeptide comprising
 - i) a nucleic acid encoding said heterologous polypeptide,
 - ii) a nucleic acid encoding a linker,
 - iii) a nucleic acid encoding an immunoglobulin fragment comprising at least the constant domains of a chain of an immunoglobulin,
 - ad) a 3' untranslated region comprising a polyadenylation signal,
 - b) the expression plasmid is introduced into a eukaryotic host cell,
 - c) the host cell is cultivated under conditions suitable for the expression of the second polypeptide, and
 - d) the second polypeptide comprising the heterologous polypeptide, a linker and an immunoglobulin fragment is recovered from the culture medium.
2. Method according to claim 1, wherein the nucleic acid encoding said second polypeptide comprises in 5' to 3' direction a nucleic acid encoding the heterologous polypeptide, a nucleic acid encoding the linker and a nucleic acid encoding the immunoglobulin fragment as the carboxy-terminal part of the second polypeptide.
 3. Method according to claim 1 or 2, wherein the second polypeptide contains in 5' position to the nucleic acid encoding said heterologous polypeptide an additional nucleic

- 50 -

acid encoding either a single amino acid or a dipeptide or the peptide QIWNN (SEQ ID NO: 472) or a fragment thereof.

4. Method according to any one of claims 1 to 3, wherein the immunoglobulin fragment is obtained either from an IgG or from an IgE.
5. Method according to any one of the claims 1 to 4, wherein the eukaryotic cell is a mammalian cell.
6. Method according to claim 5, wherein the mammalian cell is a CHO cell, NS0 cell, Sp2/0 cell, COS cell, K562 cell, BHK cell, PER.C6 cell or HEK cell.
7. Method according to any one of claims 1 to 6, wherein the linker is a peptide or polypeptide selected from the group consisting of SEQ ID NO: 06, 07, 08, 09, 10, 139, 140, 554, 555, 556, and 557.
8. Method according to any one of claims 1 to 7, wherein the immunoglobulin fragment comprises
 - a) either the C_H1-, C_H2-, C_H3-domain and the hinge region of an immunoglobulin heavy chain or the C_L-domain of an immunoglobulin light chain,
 - and
 - b) a fragment of a variable immunoglobulin heavy or light chain domain.
9. Method according to any one of claims 1 to 7, wherein the immunoglobulin fragment comprises only constant domains.
10. A plasmid comprising in a 5' to 3' direction
 - a) a promoter,
 - b) a nucleic acid encoding a first polypeptide which is a signal sequence, whose amino acid sequence is selected from the group consisting of SEQ ID NO: 163, 164, 165, 166, 167, 168, 169, 68, 70, 71, 72, 73, 75, 77, 79, 82, 83, 86, 170, 171, 172, 173, 174, 175, 176, 177, 178, 180, 181, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 194, 138, 195, 87, 89, 91, 93, 95, 103, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 473, 474,

- 51 -

475, 476, 477, 478, 479, 480, 481, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 216, 105, 107, 217, 218, 219, 220, 222, 223, 224, 225, 226, 227, 228, 229, 230, 96, 97, 98, 101, 102, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 104, 246, 247, 248, 249, 250, 28, 30, 31, 32, 33, 34, 35, 38, 39, 40, 42, 43, 44, 45, 46, 49, 51, 52, 54, 64, 65, 251, 252, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 286, 287, 288, 289, 291, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 304, 127, 128, 129, 306, 307, 133, 308, 309, 310, 313, 130, 131, 132, 134, 314, 315, 316, 317, 318, 36, 37, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 56, 58, 330, 331, 137, 332, 108, 109, 110, 111, 112, 113, 114, 117, 136, 66, 67, 333, 334, 335, 336, 337, 338, 339, 340, 342, 343, 524, 525, 526, 527, 528, 529, 530, 531, 534, 535, 537, 538, 539, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 118, 119, 120, 121, 122, 123, 124, 126, 360, 361, 362, 364, 365, 367, 368, 369, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 402, 403, 404, 405, 407, 408, 409, 410, 413, 414, 416, 417, 424, 425, 426, 427, 429, 430, 431, 436, 437, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 464, 465, 466, 468, 469, 470, 141, 02, 20, 21, 22, 23, 24, 25, 26, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, and 162, depending on the first two amino acids of a second polypeptide chosen that way, that the first two amino acids of the second polypeptide are identical to the first two amino acids of the amino acid sequences of the naturally following immunoglobulin FR1-region;

- c) a nucleic acid encoding the second polypeptide comprising
 - i) a nucleic acid encoding a heterologous polypeptide,
 - ii) a nucleic acid encoding a linker,

- 52 -

- iii) a nucleic acid encoding an immunoglobulin fragment comprising at least the constant domains of a chain of an immunoglobulin; and
 - d) a 3' untranslated region comprising a polyadenylation signal.
11. Kit for the preparation of a plasmid for the expression of a heterologous polypeptide in a eukaryotic cell, comprising:
- a plasmid which comprises in a 5' to 3' direction
- a) a promoter,
 - b) a nucleic acid encoding a first polypeptide which is a signal sequence, whose amino acid sequence is selected from the group consisting of SEQ ID NO: 36, 37, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, and 329,
 - c) a nucleic acid encoding a second polypeptide comprising
 - i) a nucleic acid encoding a peptide QIWNN (SEQ ID NO: 472) or an N-terminal fraction thereof comprising at least the dipeptide QI,
 - ii) a cloning site comprising at least one restriction cleavage site for inserting the nucleic acid encoding the heterologous polypeptide,
 - iii) a nucleic acid encoding a linker selected from the group consisting of SEQ ID NO: 06, 07, 08, 09, 10, 139, 140, 554, 555, 556, and 557, and
 - iv) a nucleic acid encoding an immunoglobulin fragment comprising at least the constant domains of a chain of an immunoglobulin; and
 - d) a 3' untranslated region comprising a polyadenylation signal, and
- one or more buffers for use in preparation of the plasmid.

Figure 1:

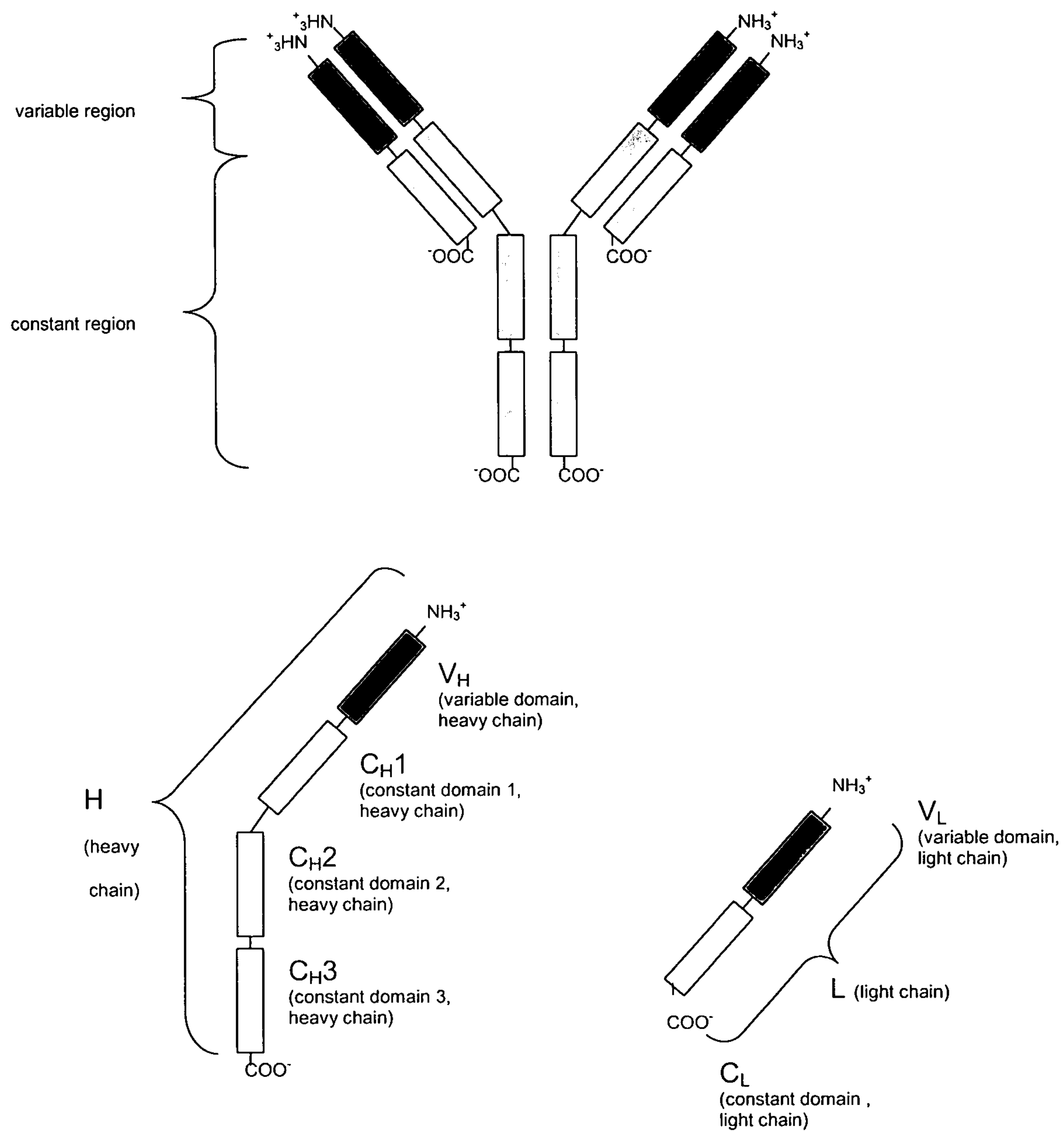


Figure 2

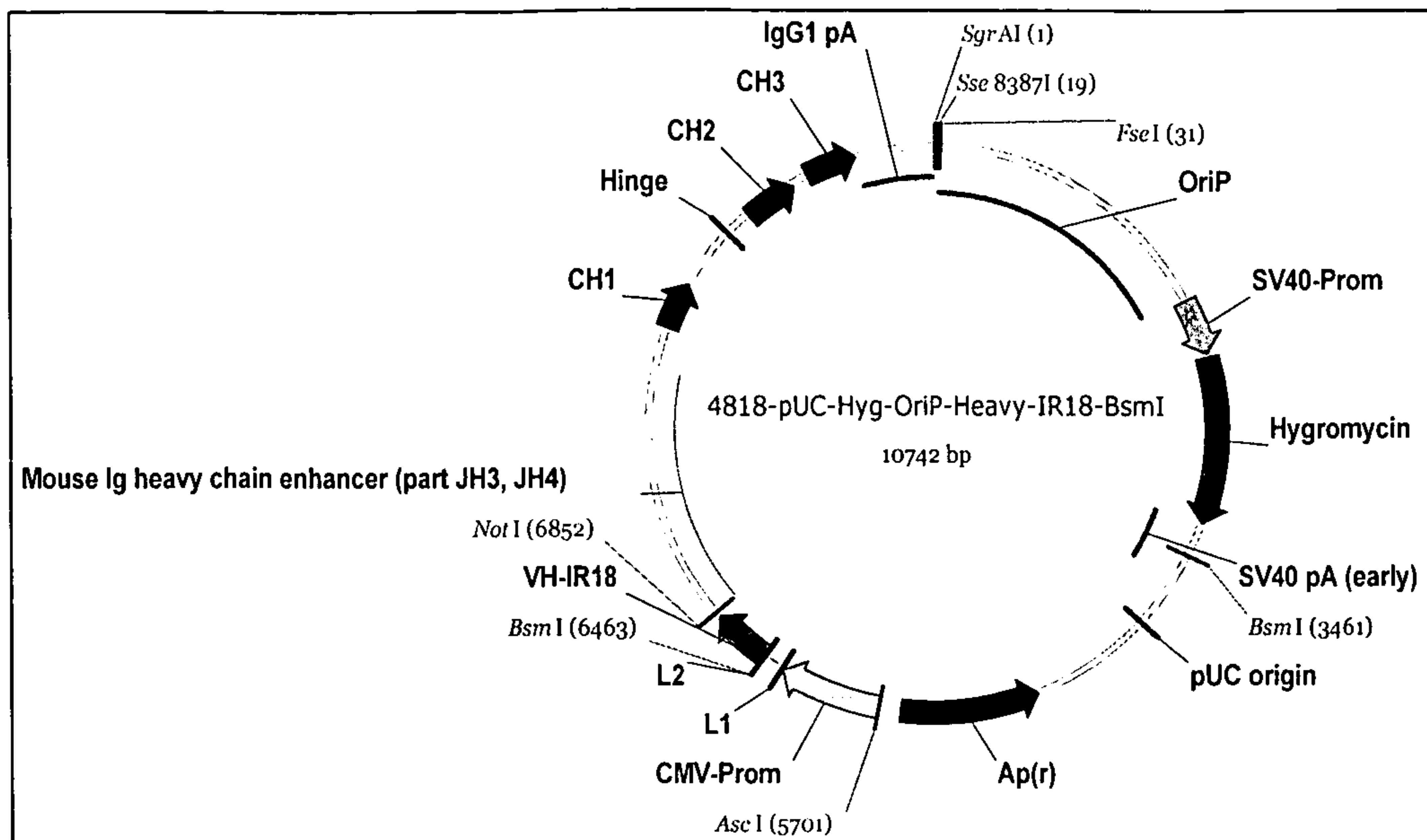


Figure 3

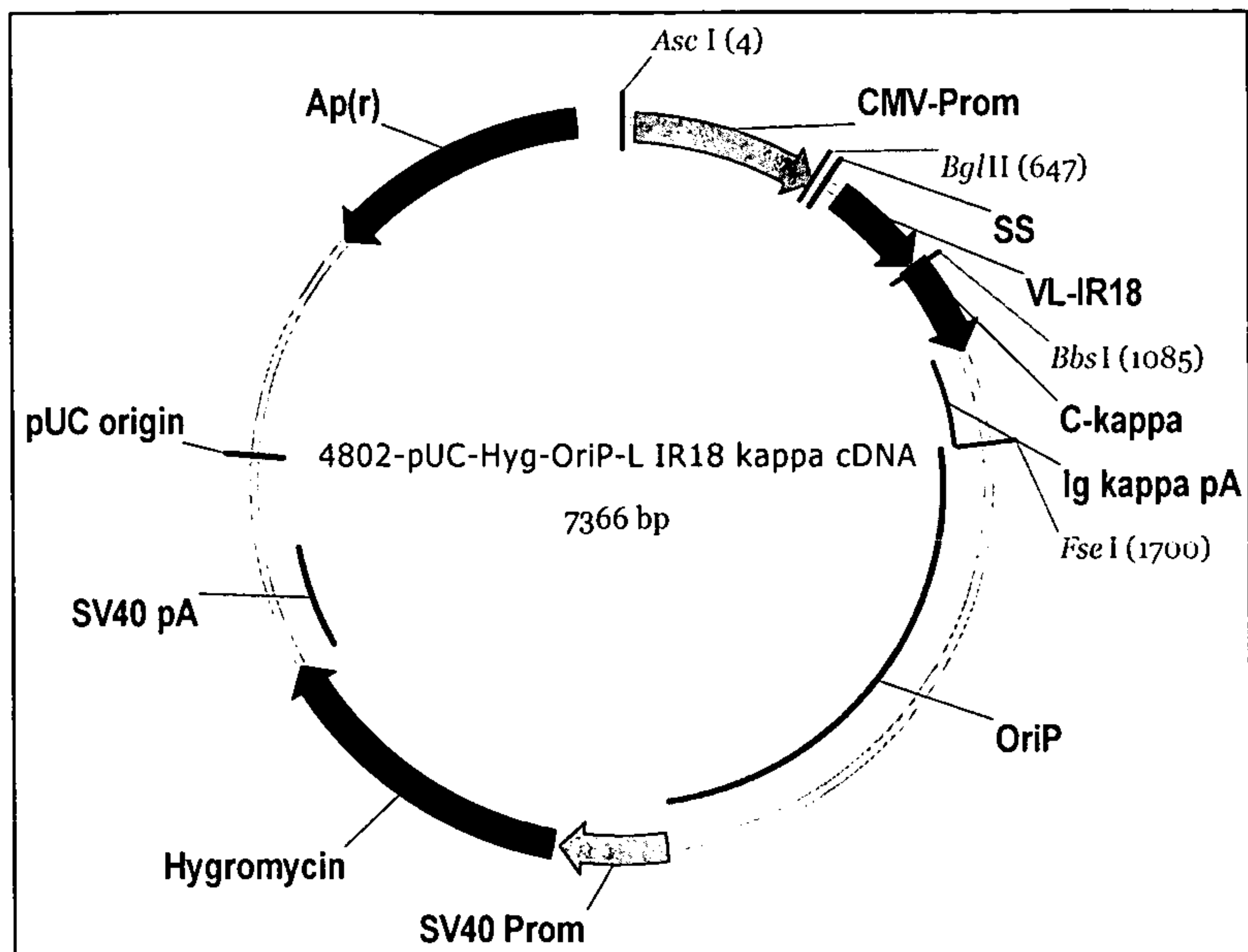


Figure 4

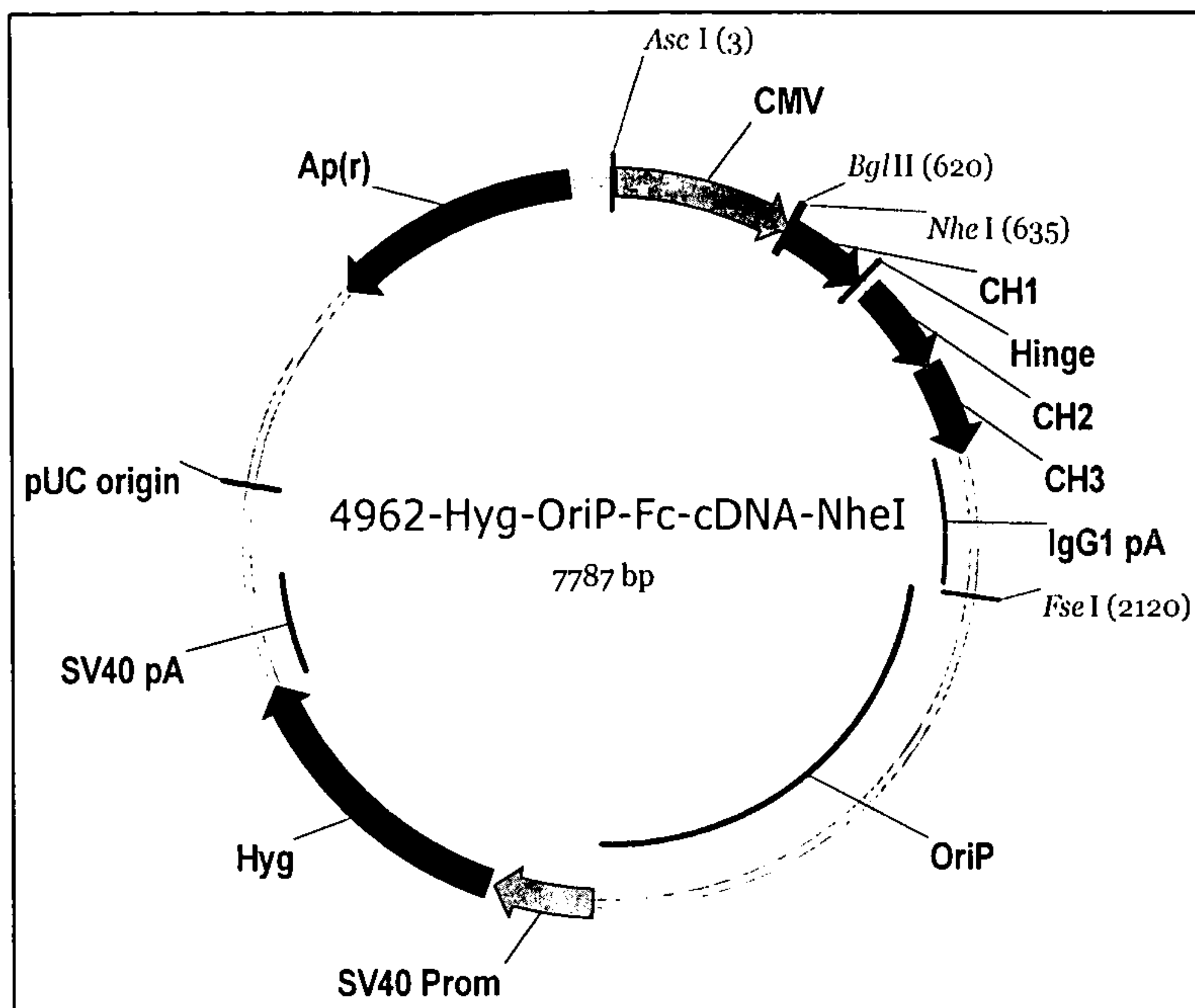


Figure 5

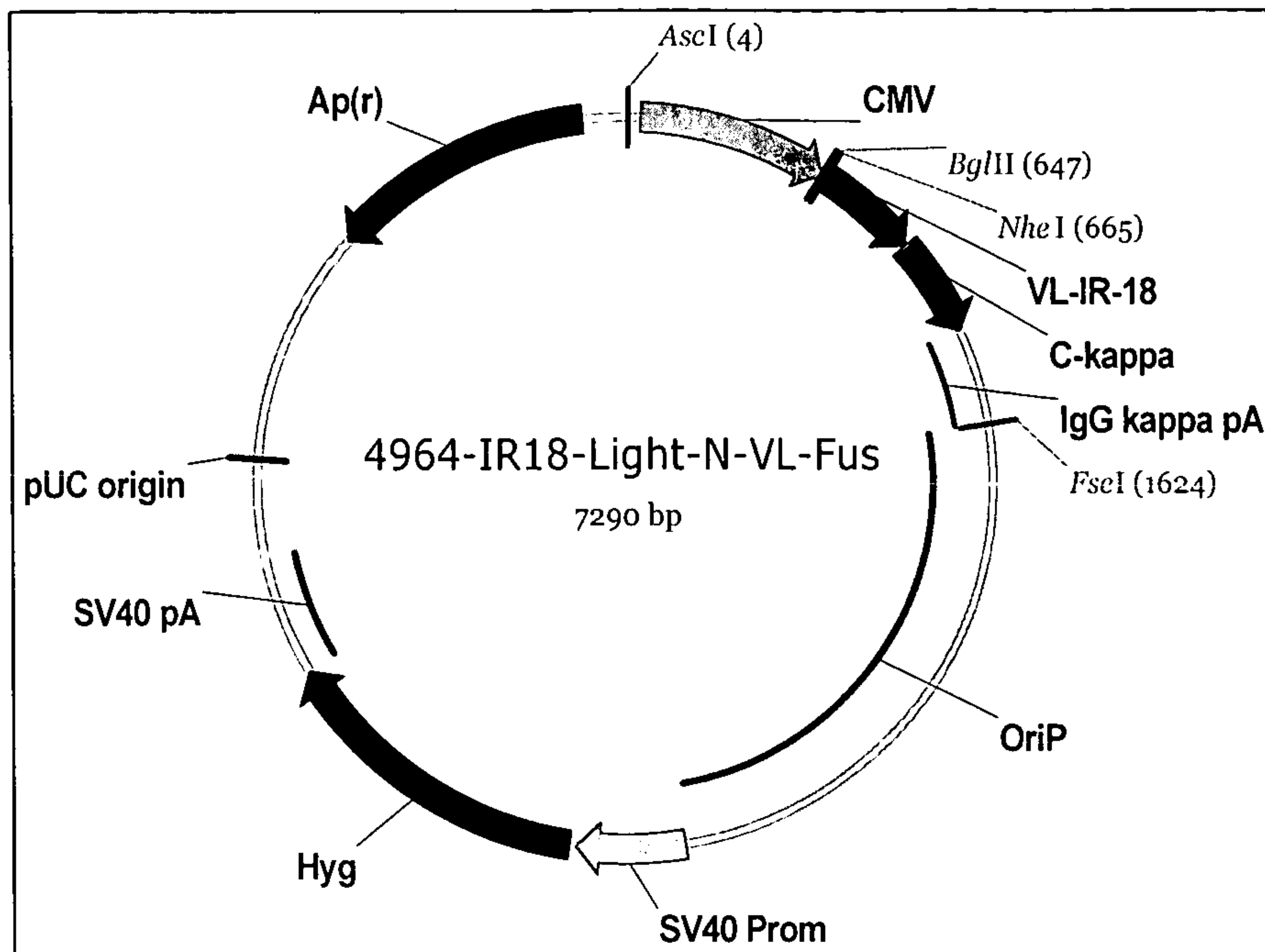


Figure 6

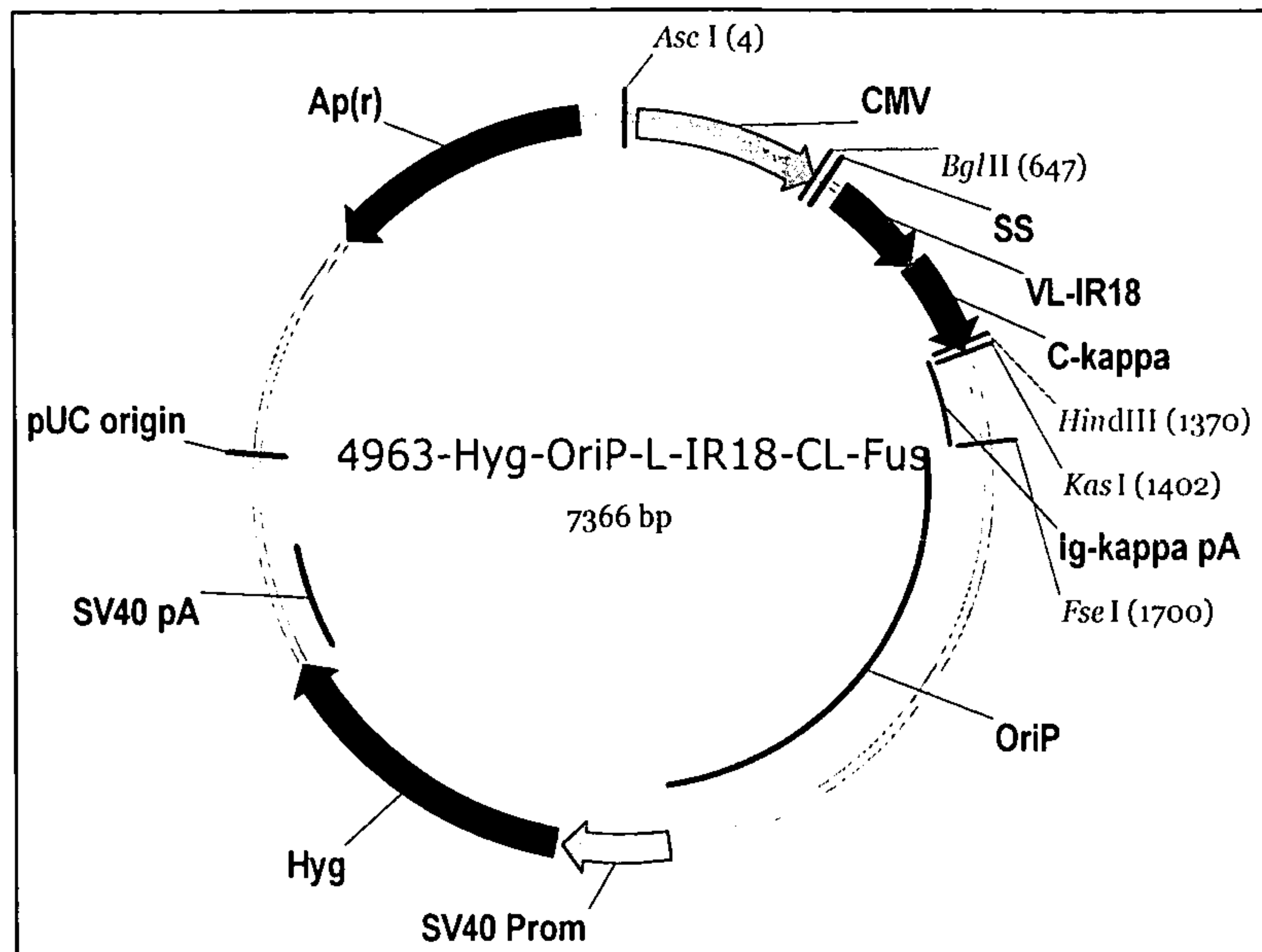


Figure 7a

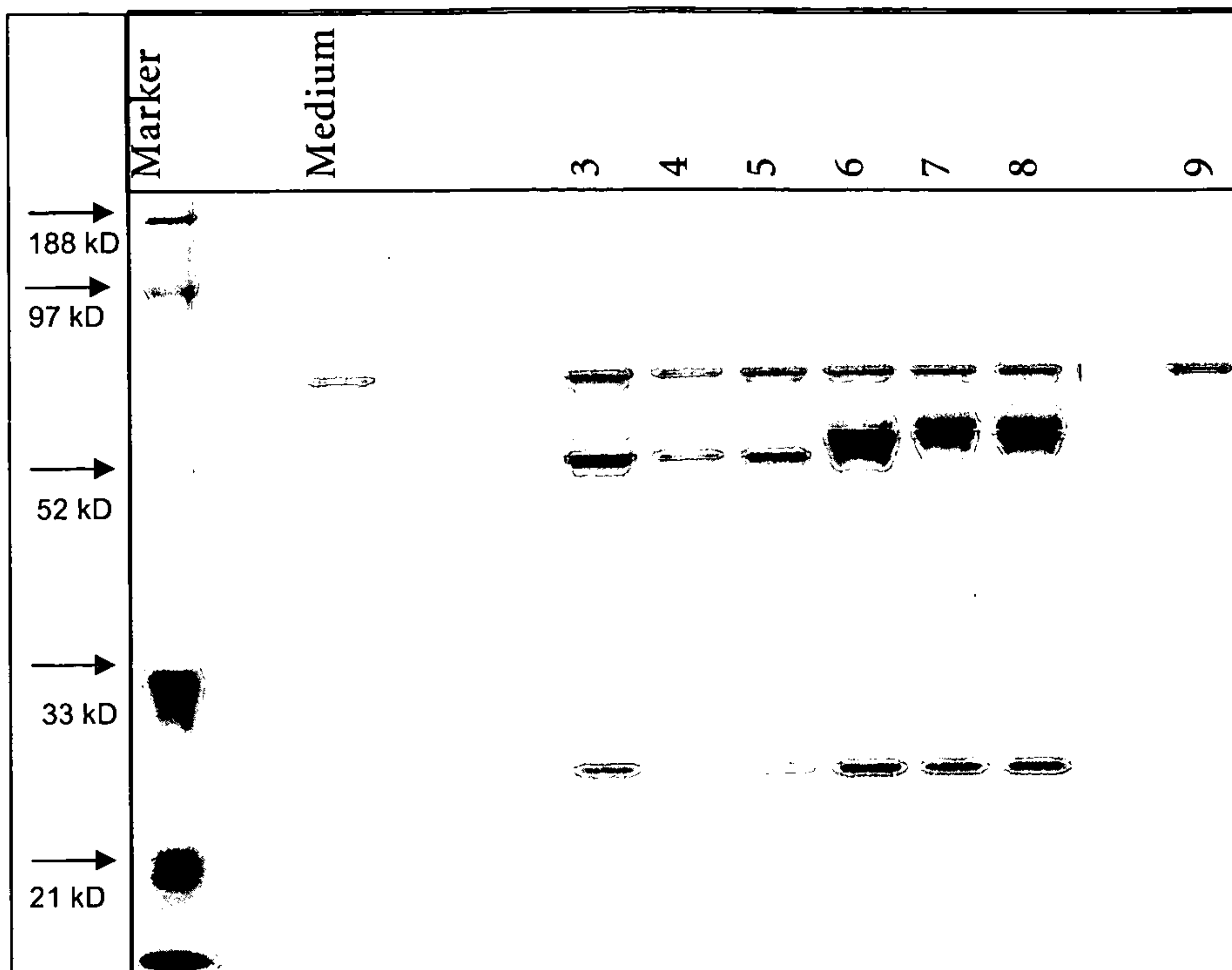


Figure 7b

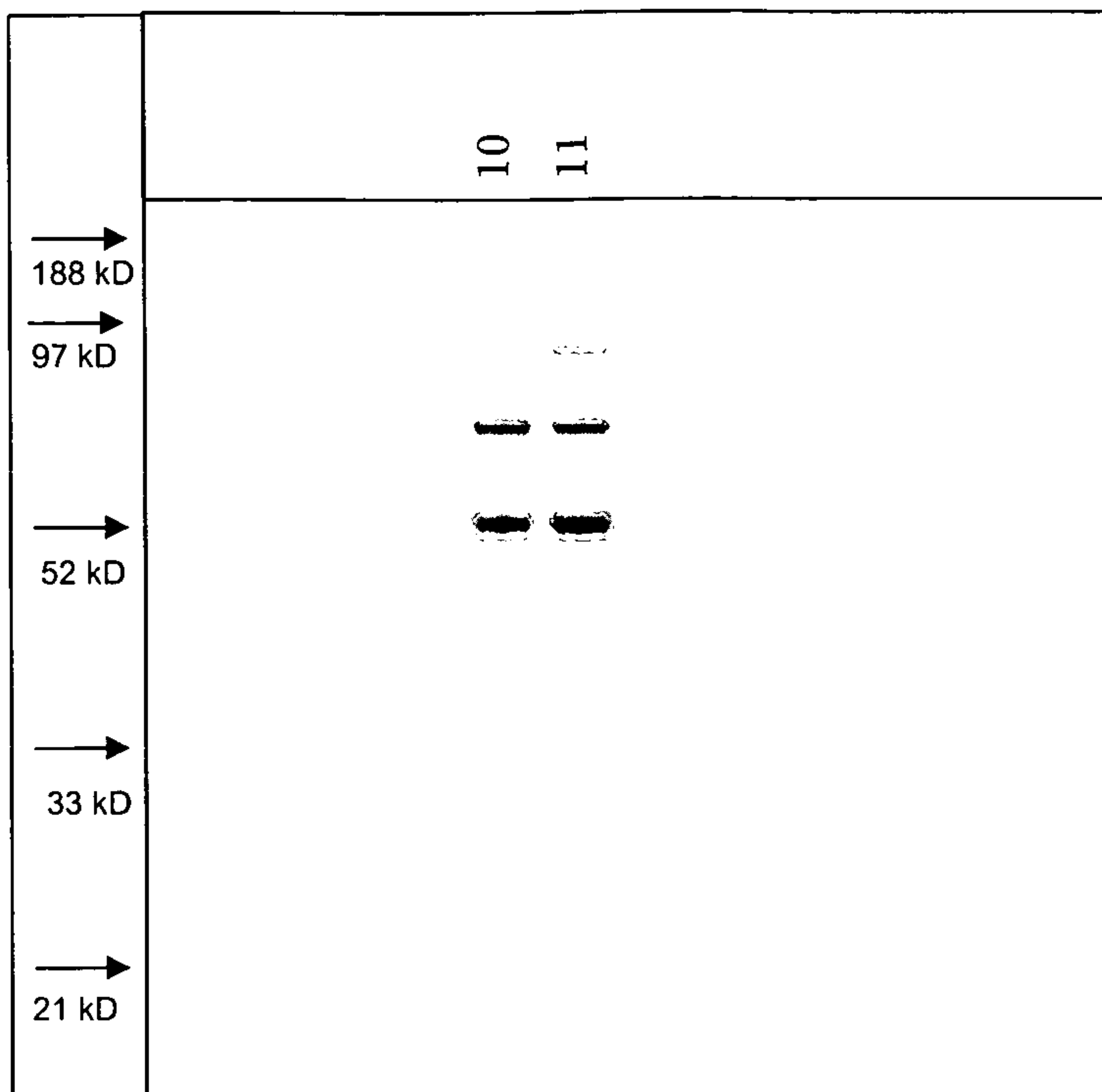


Figure 7c

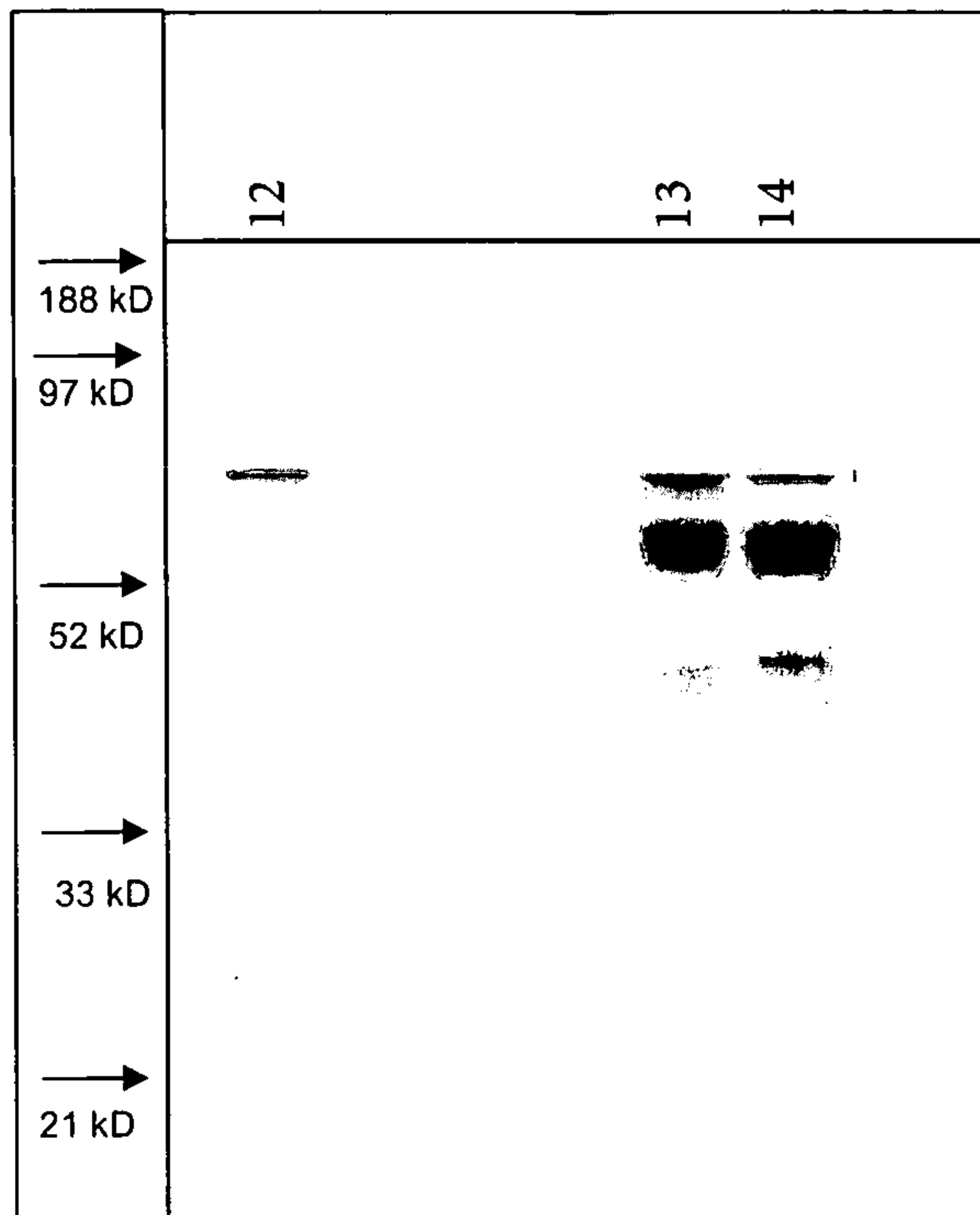


Figure 8a

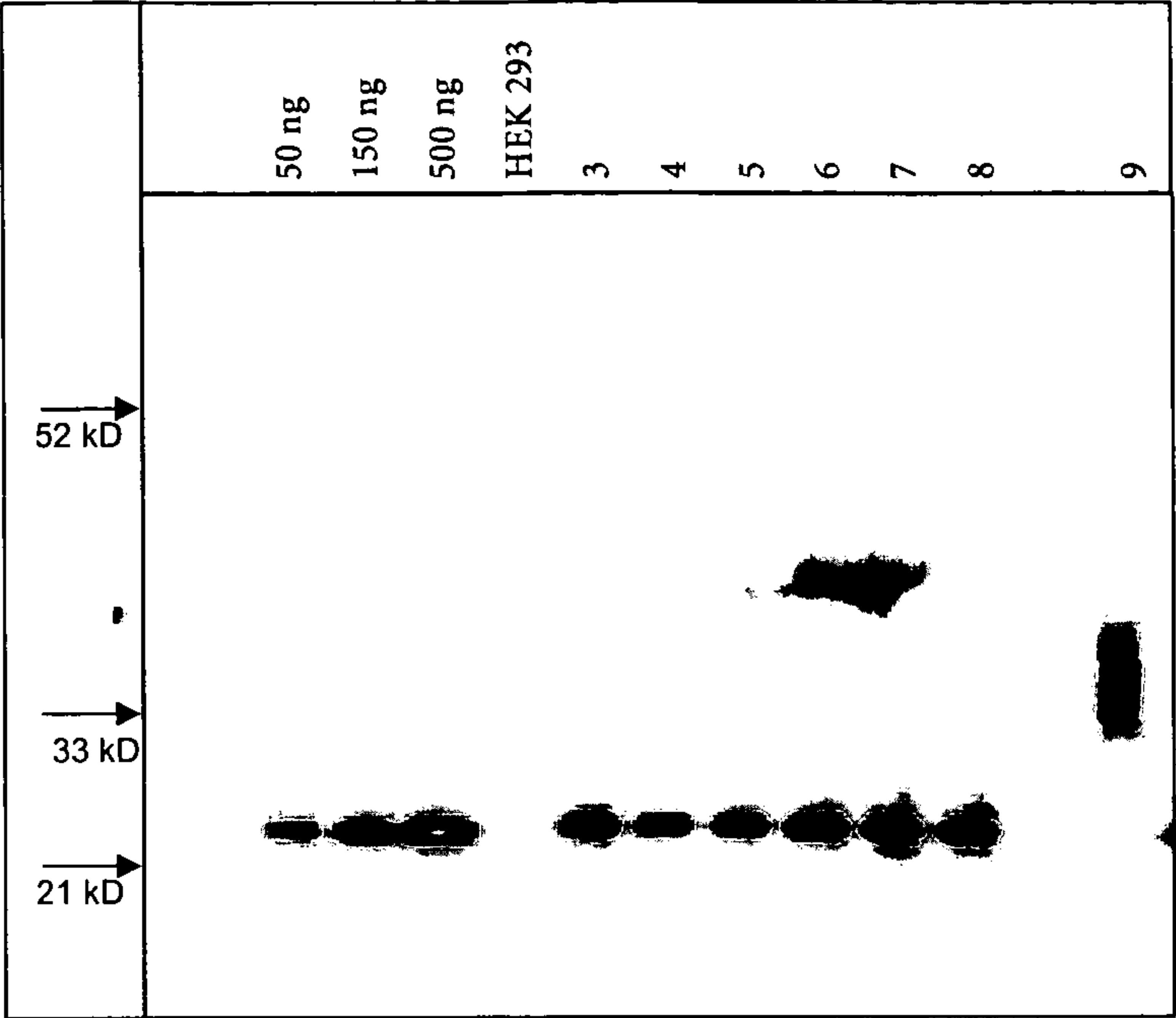


Figure 8b

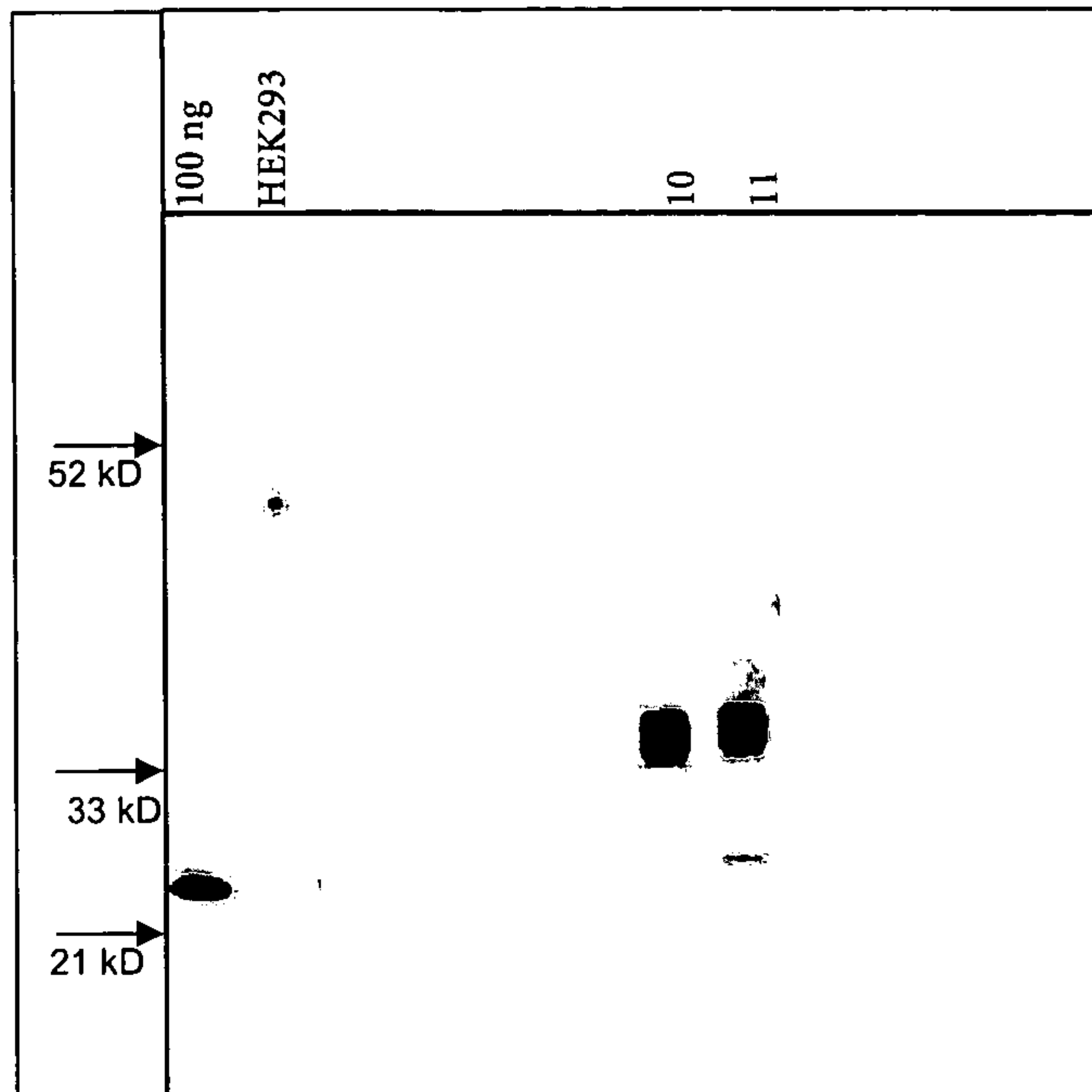


Figure 8c

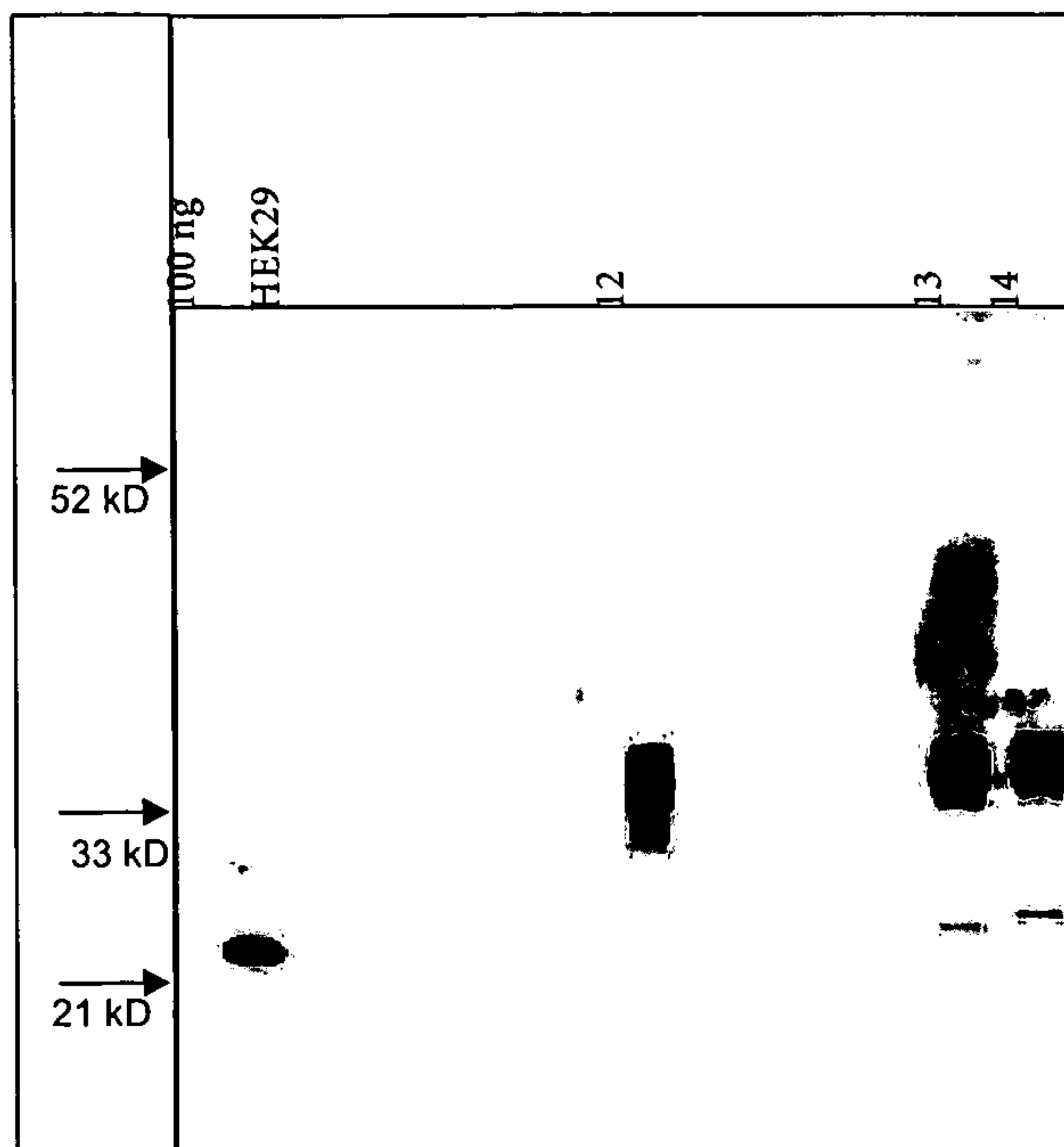


Figure 9a

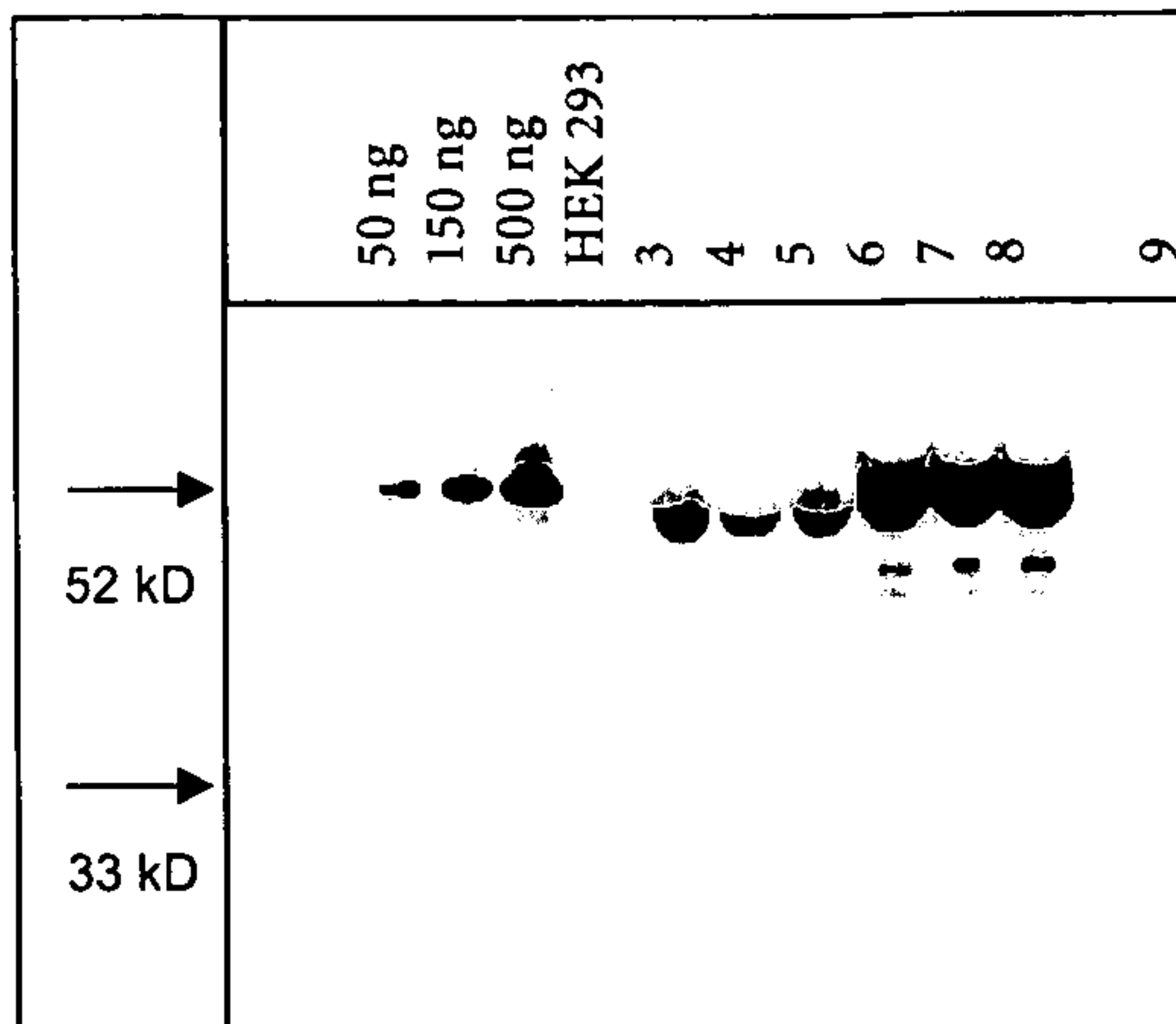


Figure 9b

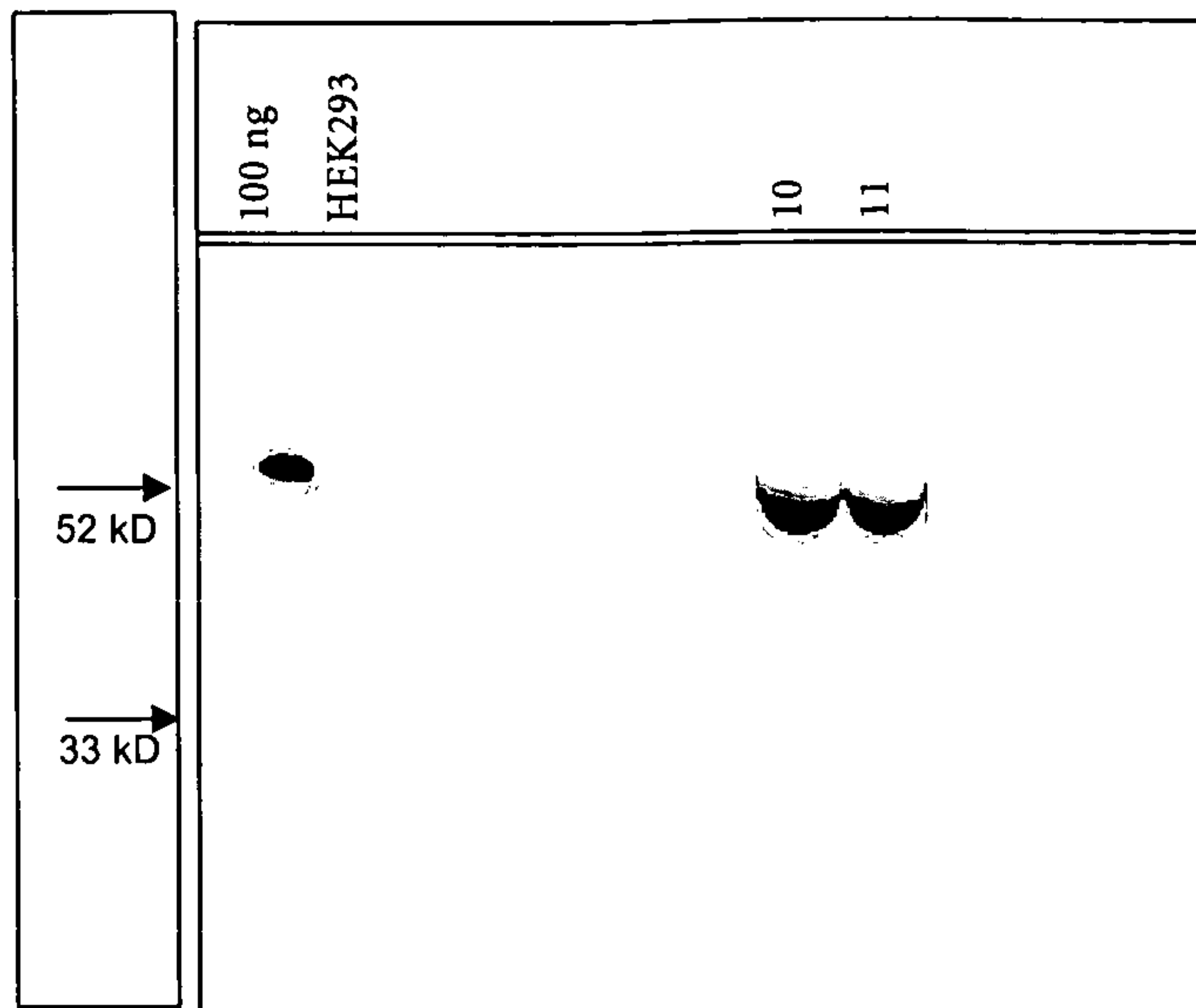


Figure 9c:

