(12) PATENT **(11)** Application No. **AU 200138831 B2**
(19) AUSTRALIAN PATENT OFFICE **(10)** Patent No. **781462**

(54) Title
Subunit optimized fusion proteins

(51)[7] International Patent Classification(s)

C07H 021/02	**C12N 005/00**
A01K 067/00	**C12N 005/02**
A01K 067/033	**C12N 015/00**
C07H 021/04	**C12N 015/09**
C07K 001/00	**C12N 015/63**
C07K 014/00	**C12N 015/70**
C07K 016/00	**C12N 015/74**
C07K 017/00	**C12P 021/00**
C08H 001/00	

(21) Application No: **200138831** (22) Application Date: **2000.09.18**

(87) WIPO No: **WO01/19842**

(30) Priority Data

(31) Number	(32) Date	(33) Country
09/399079	**1999.09.17**	**US**

(43) Publication Date : **2001.04.17**
(43) Publication Journal Date : **2001.06.14**
(44) Accepted Journal Date : **2005.05.26**

(71) Applicant(s)
GTC Biotherapeutics, Inc.

(72) Inventor(s)
Dan Pollock; Harry M. Meade; Klaus Bosslet

(74) Agent/Attorney
Davies Collison Cave,Level 15,1 Nicholson Street,MELBOURNE VIC 3000

(56) Related Art
US 5665357

AU 200138831

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 March 2001 (22.03.2001)

PCT

(10) International Publication Number
WO 01/19842 A1

(51) International Patent Classification[7]: C07H 21/02, 21/04, C12N 5/00, 5/02, 15/00, 15/09, 15/63, 15/70, 15/74, C07K 1/00, 14/00, 16/00, 17/00, C08H 1/00, C12P 21/00, A01K 67/00, 67/033

(21) International Application Number: PCT/US00/25558

(22) International Filing Date:
18 September 2000 (18.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/399,079 17 September 1999 (17.09.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/399,079 (CIP)
Filed on 17 September 1999 (17.09.1999)

GTC Biotherapeutics, Inc.

(71) Applicant (for all designated States except US): ~~GENZYME TRANSGENICS CORPORATION~~ [US/US]; 175 Crossing Boulevard, Framingham, MA 01702-~~9322~~ (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): POLLOCK, Dan [US/US]; 4 Redgate Drive, Medway, MA 02053 (US). MEADE, Harry, M. [US/US]; 62 Grasmere Street, Newton, MA 02458 (US). BOSSLET, Klaus [DE/DE]; Schering AG, 13342 Berlin (DE).

(74) Agent: MYERS, Louis; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— With international search report.

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.

WO 01/19842 A1

(54) Title: SUBUNIT OPTIMIZED FUSION PROTEINS

(57) Abstract: A method of making a fusion protein having: a first member, fused to a second member wherein the first and second members are chosen such that the fusion protein assembles into a complex having a number of subunits which optimizes activity of the multimeric form of the second member.

SUBUNIT OPTIMIZED FUSION PROTEINS

Related Applications

This application claims the benefit of a previously filed Provisional Application No. 60/101,083 filed September 18, 1998, which is hereby incorporated by reference.

Field of the Invention

The invention relates to a fusion protein having a first and a second member, wherein the second member of the fusion protein assembles into a multimer and the other member is chosen, or modified, such that it promotes assembly of the second member into a preselected or an optimal number of subunits.

Background of the Invention

Fusion proteins can combine useful properties of distinct proteins. E.g., a fusion protein can combine the targeting property of an antibody molecule with the cytotoxic effect of a toxin.

Summary of the Invention

In general, the invention features, a method of making a fusion protein having: a first member, e.g., a targeting moiety, e.g., an immunoglobulin subunit (e.g., an immunoglobulin heavy chain or light chain, or a fragment of either) fused to a second member, e.g., an enzyme, e.g., a toxin (e.g., an enzyme or toxin subunit). The first and second members are chosen such that the fusion protein assembles into a complex having a number of subunits which optimizes activity of the multimeric form of the second member. In preferred embodiments the first member, or the fusion protein, assembles into a form having the same number of subunits as are present in an active, e.g., native, form of the second member. In preferred embodiments the first member, or the fusion protein, assembles into a form having fewer subunits than are present in an active, e.g., native, form of the second member.

In preferred embodiments, the fusion protein assembles into a complex, e.g., a di-, tri-, tetra-, or higher multi-meric complex. Preferably, the fusion protein assembles into a dimer or a tetramer.

In preferred embodiments, the fusion protein assembles into a complex having enzymatic activity.

In a preferred embodiment, the first member is a monomer. E.g., it is a species which is normally monomeric, or which has been modified, e.g., by mutation of a site which modulates formation or maintenance of a multimer of subunits. In some embodiments the monomeric form is useful because it does not prevent formation of a multimer by the second member.

In another preferred embodiment, the first member is a forms a dimer, e.g., a heterodimer or homodimer. For example, it is a species which is normally dimeric, or which has been modified, e.g., by mutation of a site which modulates formation or maintenance of a multimer of subunits, to be dimeric. In some embodiments the dimeric form is useful because it does not prevent formation of a multimer by the second member.

In preferred embodiments, the fusion protein has the formula: R1-L-R2; R2-L-R1; R2-R1; or R1-R2, wherein R1 is a first member, e.g., an immunoglobulin subunit, L is a peptide linker and R2 is a second member, e.g., an enzyme subunit. Preferably, R1 and R2 are covalently linked, e.g., directly fused or linked via a peptide linker.

In preferred embodiments, the first or the second member of the fusion protein, or both are modified by, e.g., substituting or deleting, a portion of the amino acid sequence. In a particularly preferred embodiment the fusion protein includes a first member which is an Ig superfamily member, preferably an Ig subunit, which has been modified to inhibit formation of a multimeric form, e.g., a tetrameric form. Preferably the modification, which can be a change, insertion, or deletion of one or more amino acid residues, results in a subunit which does not form a multimer or which forms a lower order multimer that it normally would form, e.g., it forms a dimer rather than a tetramer.

Preferably, a region which mediates formation or maintenance of a multimeric structure is modified and thereby wholly or partly inactivated. E.g., a portion of an immunoglobulin

subunit , e.g., a heavy chain, e.g., the hinge region, is modified, e.g., deleted. In those embodiments where the hinge region of the immunoglobulin is modified, e.g., removed, the modified immunoglobulin is monovalent.

In preferred embodiments, the modification of the first member inhibits the assembly of the first member, or the fusion protein into a multimer, e.g., results in the production of a monomer, or, e.g., of a dimer, where a higher order multimer would otherwise be formed.

In preferred embodiments, the first member is a targeting agent, e.g., a polypeptide having a high affinity for a target, e.g., an antibody, a ligand, or an enzyme.

In preferred embodiments, the first member is an immunoglobulin or a fragment thereof, e.g., an antigen binding fragment thereof. Preferably, the immunoglobulin is a monoclonal antibody, e.g., a human, murine (e.g., mouse) monoclonal antibody; or a recombinant monoclonal antibody. Preferably, the monoclonal antibody is a human antibody. In other embodiments, the monoclonal antibody is a recombinant antibody, e.g., a chimeric or a humanized antibody (e.g., it has a variable region, or at least a complementarity determining region (CDR), derived from a non-human antibody (e.g., murine) with the remaining portion(s) are human in origin); or a transgenically produced human antibody (e.g., an antibody produced by a hybridoma which includes a B cell obtained from a transgenic non-human animal, e.g., a transgenic mouse, having a genome comprising a human heavy chain transgene and a light chain transgene fused to an immortalized cell).

In preferred embodiments, the first member is a full-length antibody (*e.g.*, an IgG1 or IgG4 antibody) or includes only an antigen-binding portion (*e.g.*, a Fab, $F(ab')_2$, Fv or a single chain Fv fragment).

In preferred embodiments, the first member is an immunoglobulin subunit selected from the group consisting of a subunit of : IgG (e.g., IgG1, IgG2, IgG3, IgG4), IgM, IgA1, IgA2, IgA.sub.sec, IgD, of IgE. Preferably, the immunoglobulin subunit is an IgG isotype, e.g., IgG3.

In preferred embodiments, the first member is a monomer, e.g., a single chain antibody; or forms a dimer, e.g., a dimer of an immunoglobulin heavy chain and a light chain.

In preferred embodiments, the first member is a monovalent antibody (e.g., it includes one pair of heavy and light chains, or antigen binding portions thereof). In other embodiments, the first member is divalent antibody (e.g., it includes two pairs of heavy and light chains, or antigen binding portions thereof).

In preferred embodiments, the first member includes an immunoglobulin heavy chain or a fragment thereof, e.g., an antigen binding fragment thereof. Preferably, the immunoglobulin heavy chain or fragment thereof (e.g., an antigen binding fragment thereof) is linked, e.g., linked via a peptide linker or is directly fused, to an enzyme. Preferably, the immunoglobulin heavy chain-enzyme fusion protein is capable of assembling into a functional complex, e.g., a di-, tri-, tetra-, or multi-meric complex having enzymatic activity. The most preferred form is dimeric

In preferred embodiments, the first member includes an immunoglobulin heavy chain or fragment thereof (e.g., an antigen binding fragment thereof), and a light chain or a fragment thereof (e.g., an antigen binding fragment thereof). Preferably, the immunoglobulin heavy chain is linked, e.g., linked via a peptide linker or directly fused, to an enzyme. Preferably, the fused immunoglobulin heavy chain -enzyme fusion protein assembles with a light chain, e.g., to produce a functional complex, e.g., a di-, tri-, tetra-, or multi-meric complex having enzymatic activity. The most preferred form is dimeric.

In preferred embodiments, the first member is an immunoglobulin that interacts with (e.g., binds to) a cell surface antigen on a target cell, e.g., a cancer cell. For example, the immunoglobulin binds to a tumor cell antigen, e.g., carcinoembryonic antigen (CEA), TAG-72, her-2/neu, epidermal growth factor receptor, transferrin receptor, among others.

In preferred embodiments, the first member localizes, e.g., increases the concentration of, a fusion protein in proximity to a target cell, e.g., a cancer cell.

In preferred embodiments, the second member is a subunit of an enzyme, e.g., an enzyme having one or more subunits (e.g., catalytic subunits). Preferably, the enzyme include one, preferably two, more preferably three, most preferably four subunits. A

preferred enzyme is beta-glucuronidase, e.g., a human beta-glucuronidase. The enzyme can be a homo-, or a hetero-multimer. If the enzyme is a heteromultimer, two (or more) fusion proteins are needed to form the active product.

In preferred embodiments, the second member is capable of converting a precursor drug, e.g., a prodrug, to a toxic drug.

In preferred embodiments, the first member is an immunoglobulin G (IgG) heavy and light chains, and the second member is human beta-glucuronidase fusion protein.

In preferred embodiments, the light chain of the first member has an amino acid sequence as shown in Figure 1B (SEQ ID NO:2); the light chain of the first member has an amino acid sequence at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with an amino acid sequence from Figure 1B (SEQ ID NO:2).

In preferred embodiments, the light chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence as shown in Figure 1B (SEQ ID NO:1), or Figure 2 (SEQ ID NO:37); the light chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with a nucleotide sequence shown in Figure 1B (SEQ ID NOs:2, 3, or 4), or Figure 2 (SEQ ID NO:37); the light chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence that is capable of hybridizing under stringent conditions to the nucleotide sequence shown in Figure 1B.

In preferred embodiments, the heavy chain of the first member has an amino acid sequence as shown in Figure 4B (SEQ ID NO:6, 7, 8, 9, 10 and/or 11), or Figure 5 (SEQ ID NOs:13, 14, 15 and/or 16); the heavy chain of the first member has an amino acid sequence at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with an amino acid sequence from Figure 4B (SEQ ID NO: 6, 7, 8, 9, 10 and/or 11), or Figure 5 (SEQ ID Nos:13, 14, 15 and/or 16).

In preferred embodiments, the heavy chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence as shown in Figure 4B (SEQ ID NO: 5), or Figure 5 (SEQ ID NO:12); the heavy chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with a nucleotide sequence shown in Figure 4B (SEQ ID NO:5), or Figure 5 (SEQ ID NO:12); the heavy chain of the first member has an amino acid sequence that is encoded by a nucleotide sequence that is capable of hybridizing under stringent conditions to the nucleotide sequence shown in Figure 4B, or 5.

In a preferred embodiment, the fusion protein includes a peptide linker and the peptide linker has one or more of the following characteristics: a) it allows for the rotation of the first and the second member relative to each other; b) it is resistant to digestion by proteases; c) it does not interact with the first or the second; d) it allows the fusion protein to form a complex (e.g., a di-, tri-, tetra-, or multi-meric complex) that retains enzymatic activity; and e) it promotes folding and/or assembly of the fusion protein into an active complex.

In a preferred embodiment: the fusion protein includes a peptide linker and the peptide linker is 5 to 60, more preferably, 10 to 30, amino acids in length; the peptide linker is 20 amino acids in length; the peptide linker is 17 amino acids in length; each of the amino acids in the peptide linker is selected from the group consisting of Gly, Ser, Asn, Thr and Ala; the peptide linker includes a Gly-Ser element.

In a preferred embodiment, the fusion protein includes a peptide linker and the peptide linker includes a sequence having the formula $(Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_y$ wherein y is 1, 2, 3, 4, 5, 6, 7, or 8. Preferably, the peptide linker includes a sequence having the formula $(Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_3$. Preferably, the peptide linker includes a sequence having the formula $((Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_3\text{-}Ser\text{-}Pro)$.

In preferred embodiments, the fusion protein is produced recombinantly, e.g., produced in a host cell (e.g., a cultured cell), or in a transgenic animal, e.g., a transgenic mammal (e.g., a goat, a cow, or a rodent (e.g., a mouse).

In preferred embodiments, the fusion protein is produced in a transgenic mammal (e.g., a goat, a cow, or a rodent (e.g., a mouse). Thus, the method further includes: providing a transgenic animal, which includes a transgene which provides for the expression of a fusion protein described herein; allowing the transgene to be expressed; and, preferably, recovering fusion protein, from the milk of the transgenic mammal.

For embodiments where the fusion protein is produced transgenically, the fusion protein can further include:

a signal sequence which directs the secretion of the fusion protein, e.g., a signal from a secreted protein (e.g., a signal from a protein secreted into milk; or an immunoglobulin secretory signal); and

(optionally) a sequence which encodes a sufficient portion of the amino terminal coding region of a secreted protein, e.g., a protein secreted into milk, or an immunoglobulin, to promote secretion, e.g., in the milk of a transgenic mammal, of the fusion protein.

In preferred embodiments, the fusion protein is made in a mammary gland of the transgenic mammal, e.g., a ruminant, e.g., a goat or a cow.

In preferred embodiments, the fusion protein is secreted into the milk of the transgenic mammal, e.g., a ruminant, e.g., a dairy animal, e.g., a goat or a cow.

In preferred embodiments, the fusion protein is secreted into the milk of a transgenic mammal at concentrations of at least about 0.1 mg/ml, 0.5 mg/ml, 1.0 mg/ml, 1.5 mg/ml, 2 mg/ml, 3 mg/ml, 5 mg/ml or higher.

In preferred embodiments, the fusion protein is made under the control of a mammary gland specific promoter, e.g., a milk specific promoter, e.g., a milk serum protein or casein promoter. The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the transgene encoding the fusion protein is a nucleic acid construct which includes:

(a) optionally, an insulator sequence;

(b) a promoter, e.g., a mammary epithelial specific promoter, e.g., a milk protein promoter;

(c) a nucleotide sequence which encodes a signal sequence which can direct the secretion of the fusion protein, e.g. a signal from a milk specific protein, or an immunoglobulin;

(d) optionally, a nucleotide sequence which encodes a sufficient portion of the amino terminal coding region of a secreted protein, e.g. a protein secreted into milk, or an immunoglobulin, to allow secretion, e.g., in the milk of a transgenic mammal, of the fusion protein;

(e) one or more nucleotide sequences which encode a fusion protein, e.g., an immunoglobulin-enzyme fusion protein as described herein; and

(f) (optionally) a 3' untranslated region from a mammalian gene, e.g., a mammary epithelial specific gene, (e.g., a milk protein gene).

In preferred embodiments, elements a (if present), b, c, d (if present), and f of the transgene are from the same gene; the elements a (if present), b, c, d (if present), and f of the transgene are from two or more genes. For example, the signal sequence, the promoter sequence and the 3' untranslated sequence can be from a mammary epithelial specific gene, e.g., a milk serum protein or casein gene (e.g., a β casein gene). Preferably, the signal sequence, the promoter sequence and the 3' untranslated sequence are from a goat β casein gene.

In preferred embodiments, the promoter of the transgene is a mammary epithelial specific promoter, e.g., a milk serum protein or casein promoter (e.g., a β casein promoter). The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the signal sequence encoded by the transgene is an amino terminal sequence which directs the expression of the protein to the exterior of a cell, or into the cell membrane. For example, the signal sequence can be obtained from an immunoglobulin protein. Preferably, the signal sequence is from a protein which is secreted into the milk, e.g., the milk of the transgenic animal.

In preferred embodiments, the one or more nucleotide sequences encoding a fusion protein include one or more of: a nucleotide sequence encoding a first member, e.g., an immunoglobulin heavy chain (or an antigen binding portion thereof) operably linked to a second member, e.g., an enzyme; (optionally) a nucleotide sequence encoding an immunoglobulin light chain (or an antigen binding portion thereof), or both. In one embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are operatively linked in a single construct, e.g., a single cosmid. In another embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are introduced into a transgenic animal in separate constructs. Preferably, when linked, the nucleotide sequences are arranged in the following order:
5'-N1-3' linked to 5'-N2-3'; or 5'-N2-3' linked to 5'-N1-3' wherein N1 is a first member, e.g., an immunoglobulin heavy chain (or an antigen binding portion thereof) operably linked to a second member, e.g., an enzyme; and N2 is an immunoglobulin light chain (or an antigen binding portion thereof). The nucleotide sequences can be in any orientation with respect to each other, e.g., sense/sense; reverse/reverse; sense/reverse; or reverse/sense.

In preferred embodiments, the 3' untranslated region of the transgene includes a polyadenylation site, and is obtained from a mammalian gene, e.g., a mammary epithelial specific gene, e.g., a milk serum protein gene or casein gene. The 3' untranslated region can be obtained from a casein gene (e.g., a β casein gene), a beta lactoglobulin gene, whey acid protein gene, or lactalbumin gene. Preferably, the 3' untranslated region is from a goat β casein gene.

In preferred embodiments, the transgene, e.g., the transgene as described herein, integrates into a germ cell and/or a somatic cell of the transgenic animal.

In another aspect, the invention features, a method for providing a transgenically produced fusion protein, e.g., a fusion protein as described herein, in the milk, of a transgenic mammal. The method includes obtaining milk from a transgenic mammal, which includes a fusion protein encoding transgene, e.g., one which has been introduced into its germline, e.g., a nucleic acid construct as described herein, that result in the

expression of the protein-coding sequence of fusion protein in mammary gland epithelial cells, thereby secreting the fusion protein in the milk of the mammal.

In preferred embodiments the transgenic mammal is selected from the group consisting of sheep, mice, pigs, cows and goats. The preferred transgenic mammal is a goat.

In preferred embodiments, the fusion protein is secreted into the milk of a transgenic mammal at concentrations of at least about 0.1 mg/ml, 0.5 mg/ml, 1.0 mg/ml, 1.5 mg/ml, 2 mg/ml, 3 mg/ml, 5 mg/ml or higher.

In preferred embodiments, the transgene encoding the immunoglobulin-enzyme fusion protein is a nucleic acid construct which includes:

(a) optionally, an insulator sequence;

(b) a promoter, e.g., a mammary epithelial specific promoter, e.g., a milk protein promoter;

(c) a nucleotide sequence which encodes a signal sequence which can direct the secretion of the fusion protein, e.g., a signal from a milk specific protein, or an immunoglobulin;

(d) optionally, a nucleotide sequence which encodes a sufficient portion of the amino terminal coding region of a secreted protein, e.g., a protein secreted into milk, or an immunoglobulin, to allow secretion, e.g., in the milk of a transgenic mammal, of the non-secreted protein;

(e) one or more nucleotide sequences which encode a fusion protein, e.g., a fusion protein as described herein; and

(f) optionally, a 3' untranslated region from a mammalian gene, e.g., a mammary epithelial specific gene, (e.g., a milk protein gene).

In preferred embodiments, elements a (if present), b, c, d (if present), and f of the transgene are from the same gene; the elements a (if present), b, c, d (if present), and f of the transgene are from two or more genes. For example, the signal sequence, the promoter sequence and the 3' untranslated sequence can be from a mammalian gene, e.g., a mammary epithelial specific gene, e.g., a milk serum protein or casein gene (e.g., a β casein gene).

Preferably, the signal sequence, the promoter sequence and the 3' untranslated sequence are from a goat β casein gene.

In preferred embodiments, the promoter of the transgene is a mammary epithelial specific promoter, e.g., a milk serum protein or casein promoter (e.g., a β casein promoter). The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the signal sequence encoded by the transgene is an amino terminal sequence which directs the expression of the protein to the exterior of a cell, or into the cell membrane. Preferably, the signal sequence is from a protein which is secreted into the milk, e.g., the milk of the transgenic animal.

In preferred embodiments, the one or more nucleotide sequences encoding an fusion protein include one or more of: a nucleotide sequence encoding an immunoglobulin heavy chain (or an antigen binding portion thereof) fused to an enzyme; a nucleotide sequence encoding an immunoglobulin light chain (or an antigen binding portion thereof), or both. In one embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are operatively linked in a single construct, e.g., a single cosmid. In another embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are introduced into a transgenic animal in separate constructs. Preferably, when linked, the nucleotide sequences are arranged in the following order:
5'-N1-3' linked to 5'-N2-3'; or 5'-N2-3' linked to 5'-N1-3' wherein N1 is an immunoglobulin heavy chain (or an antigen binding portion thereof) linked to an enzyme; and N2 is an immunoglobulin light chain (or an antigen binding portion thereof). The nucleotide sequences can be in any orientation with respect to each other, e.g., sense/sense; reverse/reverse; sense/reverse; or reverse/sense.

In preferred embodiments, the 3' untranslated region of the transgene includes a polyadenylation site, and is obtained from a mammalian gene, e.g., a mammary epithelial specific gene, (e.g., a milk serum protein gene or casein gene). The 3' untranslated region can be obtained from a casein gene (e.g., a β casein gene), a beta lactoglobulin gene, whey

acid protein gene, or lactalbumin gene. Preferably, the 3' untranslated region is from a goat β casein gene.

In preferred embodiments, the transgene, e.g., the transgene as described herein, integrates into a germ cell and/or a somatic cell of the transgenic animal.

In another aspect, the invention features, a transgene, e.g., a nucleic acid construct, preferably, an isolated nucleic acid construct, which includes:

(a) optionally, an insulator sequence;

(b) a promoter, e.g., a mammary epithelial specific promoter, e.g., a milk protein promoter;

(c) a nucleotide sequence which encodes a signal sequence which can direct the secretion of the fusion protein, e.g. a signal sequence from a milk specific protein, or an immunoglobulin;

(d) optionally, a nucleotide sequence which encodes a sufficient portion of the amino terminal coding region of a secreted protein, e.g. a protein secreted into milk, or an immunoglobulin, to allow secretion, e.g., in the milk of a transgenic mammal, of the fusion protein protein;

(e) one or more nucleotide sequences which encode a fusion protein, e.g., a fusion protein as described herein; and

(f) optionally, a 3' untranslated region from a mammalian gene, e.g., a mammary epithelial specific gene, (e.g., a milk protein gene).

In preferred embodiments, elements a (if present), b, c, d (if present), and f of the transgene are from the same gene; the elements a (if present), b, c, d (if present), and f of the transgene are from two or more genes. For example, the signal sequence, the promoter sequence and the 3' untranslated sequence can be from a mammalian gene, e.g., a mammary epithelial specific gene, e.g., a milk serum protein or casein gene (e.g., a β casein gene). Preferably, the signal sequence, the promoter sequence and the 3' untranslated sequence are from a goat β casein gene.

In preferred embodiments, the promoter of the transgene is a mammary epithelial specific promoter, e.g., a milk serum protein or casein promoter (e.g., a β casein promoter).

The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the signal sequence encoded by the transgene is an amino terminal sequence which directs the expression of the protein to the exterior of a cell, or into the cell membrane. Preferably, the signal sequence is from a milk specific protein, or an immunoglobulin. Preferably, the signal sequence directs secretion of the encoded fusion protein into the milk of a transgenic animal, e.g., a transgenic mammal.

In preferred embodiments, the one or more nucleotide sequences encoding a fusion protein include one or more of: a nucleotide sequence encoding an immunoglobulin heavy chain (or an antigen binding portion thereof) fused to an enzyme; a nucleotide sequence encoding an immunoglobulin light chain (or an antigen binding portion thereof), or both. In one embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are operatively linked in a single construct, e.g., a single cosmid. In another embodiment, the nucleotide sequences encoding the heavy chain fusion and the light chain are introduced into a transgenic animal in separate constructs. Preferably, when linked, the nucleotide sequences are arranged in the following order:
5'-N1-3' linked to 5'-N2-3'; or 5'-N2-3' linked to 5'-N1-3' wherein N1 is an immunoglobulin heavy chain (or an antigen binding portion thereof) linked to an enzyme; and N2 is an immunoglobulin light chain (or an antigen binding portion thereof). The nucleotide sequences can be in any orientation with respect to each other, e.g., sense/sense; reverse/reverse; sense/reverse; or reverse/sense.

In preferred embodiments, the 3' untranslated region of the transgene includes a polyadenylation site, and is obtained from a mammalian gene, e.g., a mammary epithelial specific gene, (e.g., a milk serum protein gene or casein gene). The 3' untranslated region can be obtained from a casein gene (e.g., a β casein gene), a beta lactoglobulin gene, whey acid protein gene, or lactalbumin gene. Preferably, the 3' untranslated region is from a goat β casein gene.

In another aspect, the invention features a nucleic acid molecule encoding a fusion protein, e.g., a fusion protein as described herein.

In preferred embodiments, nucleic acid has a nucleotide sequence as shown in Figure 1B (SEQ ID NO:1), Figure 2 (SEQ ID NO:37), Figure 4B (SEQ ID NO:5), or Figure 5 (SEQ ID NO:12); the nucleic acid has a nucleotide sequence at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with a nucleotide sequence shown in Figure 1B (SEQ ID NO:1), Figure 2 (SEQ ID NO:37), Figure 4B (SEQ ID NO:5), or Figure 5 (SEQ ID NO:12); the nucleic acid has a nucleotide sequence that is capable of hybridizing under stringent conditions to the nucleotide sequence shown in Figure 1B, Figure 2, Figure 4B, or Figure 5.

In a preferred embodiment, the nucleic acid has a nucleotide sequence which encodes an amino acid sequence as shown in Figure 1A (SEQ ID NOs:2, 3, 4), Figure 4B (SEQ ID NO:6, 7, 8, 9, 10, 11), or Figure 5 (SEQ ID NO:13, 14, 15, 16); the nucleic acid has a nucleotide sequence which encodes an amino acid sequence which is at least 60%, 70%, 75%, more preferably at least 85%, more preferably at least 90%, more preferably at least 95%, most preferably at least 98%, 99% sequence identity or homology with an amino acid sequence from Figure 1A (SEQ ID NO:2, 3, 4), Figure 4B (SEQ ID NO:6, 7, 8, 9, 10, 11), or Figure 5 (SEQ ID NO:13, 14, 15, 16).

In another aspect, the invention features a host cell, e.g., an isolated host cell (e.g., a cultured cell), which includes a nucleic acid of the invention (e.g., a nucleic acid, or a transgene, e.g., a nucleic acid construct, as described herein).

In another aspect, the invention features, a fusion protein described herein, or a purified preparation thereof

In another aspect, the invention features, a pharmaceutical or nutraceutical composition having a therapeutically effective amount of a fusion protein, e.g., a fusion protein as described herein, and a pharmaceutically acceptable carrier.

In a preferred embodiment, the composition includes milk.

In another aspect, the invention features, a transgenic animal which includes a transgene that encodes a fusion protein, e.g., a transgene which encodes a fusion protein described herein.

Preferred transgenic animals include: mammals; birds; reptiles; marsupials; and amphibians. Suitable mammals include: ruminants; ungulates; domesticated mammals; and dairy animals. Particularly preferred animals include: mice, goats, sheep, camels, rabbits, cows, pigs, horses, oxen, and llamas. Suitable birds include chickens, geese, and turkeys. Where the transgenic protein is secreted into the milk of a transgenic animal, the animal should be able to produce at least 1, and more preferably at least 10, or 100, liters of milk per year. Preferably, the transgenic animal is a ruminant, e.g., a goat, cow or sheep. Most preferably, the transgenic animal is a goat.

In preferred embodiments, the transgenic mammals has germ cells and somatic cells containing a transgene that encodes a fusion protein, e.g., a transgene which encodes a fusion protein described herein.

In preferred embodiments, the fusion protein expressed in the transgenic animal is under the control of a mammary gland specific promoter, e.g., a milk specific promoter, e.g., a milk serum protein or casein promoter. The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the transgenic animal is a mammal, and the fusion protein is secreted into the milk of the transgenic animal at concentrations of at least about 0.1 mg/ml, 0.5 mg/ml, 1.0 mg/ml, 1.5 mg/ml, 2 mg/ml, 3 mg/ml, 5 mg/ml or higher.

In another aspect, the invention features, a method of making a transgenic organism which has a fusion protein transgene. The method includes providing or forming in a cell of an organism, a fusion protein, e.g., a transgene which encodes a fusion protein described herein; and allowing the cell, or a descendent of the cell, to give rise to a transgenic organism.

In a preferred embodiment, the transgenic organism is a transgenic plant or animal. Preferred transgenic animals include: mammals; birds; reptiles; marsupials; and amphibians. Suitable mammals include: ruminants; ungulates; domesticated mammals; and dairy animals. Particularly preferred animals include: mice, goats, sheep, camels, rabbits, cows, pigs, horses, oxen, and llamas. Suitable birds include chickens, geese, and turkeys. Where the transgenic protein is secreted into the milk of a transgenic animal, the animal should be able to produce at least 1, and more preferably at least 10, or 100, liters of milk per year.

In preferred embodiments, the fusion protein is under the control of a mammary gland specific promoter, e.g., a milk specific promoter, e.g., a milk serum protein or casein promoter. The milk specific promoter can be a casein promoter, beta lactoglobulin promoter, whey acid protein promoter, or lactalbumin promoter. Preferably, the promoter is a goat β casein promoter.

In preferred embodiments, the organism is a mammal, and the fusion protein is secreted into the milk of the transgenic animal at concentrations of at least about 0.1 mg/ml, 0.5 mg/ml, 1.0 mg/ml, 1.5 mg/ml, 2 mg/ml, 3 mg/ml, 5 mg/ml or higher.

In another aspect, the invention features, a method of selectively killing an aberrant or diseased cell which expresses on its surface a target antigen, e.g., a cancer cell expressing a cell suface antigen. The method includes:

contacting said aberrant or diseased cell with an effective amount of a fusion protein, e.g., a fusion protein described herein, wherein either the first or the second member of the fusion protein recognizes said target antigen, such that selective killing of the cell occurs.

The subject method can be used on cells in culture, e.g. *in vitro* or *ex vivo* (e.g., cultures comprising cancer cells). For example, cells can be cultured *in vitro* in culture medium and the contacting step can be effected by adding the fusion protein of the invention to the culture medium. Alternatively, the method can be performed on cells (e.g., cancer cells) present in a subject, e.g., as part of an *in vivo* (e.g., therapeutic or prophylactic) protocol.

In another aspect, the invention features, a method of selectively killing an aberrant or diseased cell which expresses on its surface a target antigen, e.g., a cancer cell expressing a cell suface antigen. The method includes:

introducing into said aberrant or diseased cell a nucleic acid encoding a fusion protein, e.g., a fusion protein described herein, wherein either the first or the second member of the fusion protein recognizes said target antigen, such that selective killing of the cell occurs.

The subject method can be used on cells in culture, e.g. *in vitro* or *ex vivo* (e.g., cultures comprising cancer cells). For example, cells can be cultured *in vitro* in culture medium and the nucleic acids of the invention can be introduced to the culture medium. Alternatively, the method can be performed on cells (e.g., cancer cells) present in a subject, e.g., as part of an *in vivo* (e.g., therapeutic or prophylactic) gene therapy protocol.

In another aspect, the invention provides, a method of treating in a subject, a disorder characterized by aberrant growth or activity of a cell which expresses on its surface a target antigen, e.g., a cancer cell expressing a target antigen. The method includes administering to the subject an effective amount of a fusion protein, or a nucleic acid encoding a fusion protein (e.g., a fusion protein described herein), wherein either the first or the second member of the fusion protein recognizes said target antigen.

In a preferred embodiment, the disease is characterized by aberrant growth or activity of a cell, e.g., cancer cell, an immune cell.

In yet another aspect, the present invention provides a method for detecting *in vitro* or *in vivo* the presence of target antigen in a sample, e.g., for diagnosing a disease. The method comprises (i) contacting a sample or a control sample under conditions that allow interaction of a labelled fusion protein, e.g.. a fusion protein as described herein, and (ii) detecting formation of a complex. A statistically significant change in the formation of the complex between the fusion protein antibody and the target antigen with respect to a control sample is indicative the presence of target antigen in the sample.

In preferred embodiments, the second member is an enzyme, e.g., horseradish peroxidase.

The invention features fusion proteins in which the ability of a first member of the fusion to from a multimer is chosen so as to optimize a characteristic, e.g., activity or solubility, of the second member.

The terms peptides, proteins, and polypeptides are used interchangeably herein.

A purified preparation, substantially pure preparation of a polypeptide, or an isolated polypeptide as used herein, means a polypeptide that has been separated from at least one other protein, lipid, or nucleic acid with which it occurs in the cell or organism which expresses it, e.g., from a protein, lipid, or nucleic acid in a transgenic animal or in a fluid, e.g., milk, or other substance, e.g., an egg, produced by a transgenic animal. The polypeptide is preferably separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. The polypeptide preferably constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 μg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

A substantially pure nucleic acid, is a nucleic acid which is one or both of: not immediately contiguous with either one or both of the sequences, e.g., coding sequences, with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid sequence with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant

DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional fusion protein sequence.

Homology, or sequence identity, as used herein, refers to the sequence similarity between two polypeptide molecules or between two nucleic acid molecules. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total # of positions x 100).

For example, if 6 of 10, of the positions in two sequences are matched or homologous then the two sequences are 60% homologous or have 60% sequence identity. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology or sequence identity. Generally, a comparison is made when two sequences are aligned to give maximum homology or sequence identity.

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci.* USA 87:2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to ITALY nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to ITALY protein molecules of the invention. To obtain

gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

As used herein, the term transgene means a nucleic acid sequence (encoding, e.g., one or more fusion protein polypeptides), which is introduced into the genome of a transgenic organism. A transgene can include one or more transcriptional regulatory sequences and other nucleic acid, such as introns, that may be necessary for optimal expression and secretion of a nucleic acid encoding the fusion protein. A transgene can include an enhancer sequence. A fusion protein sequence can be operatively linked to a tissue specific promoter, e.g., mammary gland specific promoter sequence that results in the secretion of the protein in the milk of a transgenic mammal, a urine specific promoter, or an egg specific promoter.

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

A transgenic organism, as used herein, refers to a transgenic animal or plant.

As used herein, a "transgenic animal" is a non-human animal in which one or more, and preferably essentially all, of the cells of the animal contain a transgene introduced by way of human intervention, such as by transgenic techniques known in the art. The transgene can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus.

Mammals are defined herein as all animals, excluding humans, that have mammary glands and produce milk.

As used herein, a "dairy animal" refers to a milk producing non-human animal which is larger than a rodent. In preferred embodiments, the dairy animal produce large volumes of milk and have long lactating periods, e.g., cows or goats.

As used herein, the language "subject" includes human and non-human animals. The term "non-human animals" of the invention includes vertebrates, e.g., mammals and non-mammals, such as non-human primates, ruminants, birds, amphibians, reptiles and rodents, e.g., mice and rats. The term also includes rabbits.

As used herein, a "transgenic plant" is a plant, preferably a multi-celled or higher plant, in which one or more, and preferably essentially all, of the cells of the plant contain a transgene introduced by way of human intervention, such as by transgenic techniques known in the art.

As used herein, the term "plant" refers to either a whole plant, a plant part, a plant cell, or a group of plant cells. The class of plants which can be used in methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. It includes plants of a variety of ploidy levels, including polyploid, diploid and haploid.

As used herein, the terms "immunoglobulin" and "antibody" refer to a glycoprotein comprising at least two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds. Each heavy chain is comprised of a heavy chain variable region (abbreviated herein as HCVR or VH) and a heavy chain constant region. The heavy chain constant region is comprised of three domains, CH1, CH2 and CH3. Each light chain is comprised of a light chain variable region (abbreviated herein as LCVR or VL) and a light chain constant region. The light chain constant region is comprised of one domain, CL. The VH and VL regions can be further subdivided into regions of hypervariability, termed complementarity determining regions (CDR), interspersed with regions that are more conserved, termed framework regions (FR). Each VH and VL is composed of three CDRs and four FRs, arranged from amino-terminus to carboxy-terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3, FR4. The variable regions of the heavy and light chains contain a binding domain that interacts with an antigen. The constant regions of the antibodies may mediate the binding of the immunoglobulin to host tissues or factors,

including various cells of the immune system (e.g., effector cells) and the first component (Clq) of the classical complement system.

The term "antigen-binding portion" of an antibody (or simply "antibody portion"), as used herein, refers to one or more fragments of an antibody that retain the ability to specifically bind to an antigen (*e.g.* a target antigen). It has been shown that the antigen-binding function of an antibody can be performed by fragments of a full-length antibody. Examples of binding fragments encompassed within the term "antigen-binding portion" of an antibody include (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a $F(ab')_2$ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) a Fd fragment consisting of the VH and CH1 domains; (iv) a Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward *et al.*, (1989) *Nature* 341:544-546), which consists of a VH domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be joined, using recombinant methods, by a synthetic linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules (known as single chain Fv (scFv); see *e.g.*, Bird *et al.* (1988) *Science* 242:423-426; and Huston *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:5879-5883). Such single chain antibodies are also intended to be encompassed within the term "antigen-binding portion" of an antibody. These antibody fragments are obtained using conventional techniques known to those with skill in the art, and the fragments are screened for utility in the same manner as are intact antibodies.

The term "monoclonal antibody" as used herein refers to an antibody molecule of single molecular composition. A monoclonal antibody composition displays a single binding specificity and affinity for a particular epitope. Accordingly, the term "human monoclonal antibody" refers to antibodies displaying a single binding specificity which have variable and constant regions derived from human germline immunoglobulin sequences. In one embodiment, the human monoclonal antibodies are produced by a hybridoma which includes a B cell obtained from a transgenic non-human animal, e.g., a

transgenic mouse, having a genome comprising a human heavy chain transgene and a light chain transgene fused to an immortalized cell.

The term "recombinant human antibody", as used herein, is intended to include all human antibodies that are prepared, expressed, created or isolated by recombinant means, such as antibodies isolated from an animal (*e.g.*, a mouse) that is transgenic for human immunoglobulin genes; antibodies expressed using a recombinant expression vector transfected into a host cell, antibodies isolated from a recombinant, combinatorial human antibody library, or antibodies prepared, expressed, created or isolated by any other means that involves splicing of human immunoglobulin gene sequences to other DNA sequences. Such recombinant human antibodies have variable and constant regions derived from human germline immunoglobulin sequences. In certain embodiments, however, such recombinant human antibodies are subjected to *in vitro* mutagenesis (or, when an animal transgenic for human Ig sequences is used, *in vivo* somatic mutagenesis) and thus the amino acid sequences of the VH and VL regions of the recombinant antibodies are sequences that, while derived from and related to human germline VH and VL sequences, may not naturally exist within the human antibody germline repertoire *in vivo*.

A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence. With respect to transcription regulatory sequences, operably linked means that the DNA sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame.

The terms "vector" or "construct", as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) can be integrated into the genome of

a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated vectors.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Detailed Description

The drawings are first described.

Figure 1A is a schematic diagram of a construct containing the genomic sequence of the light chain (LC) of humanized anti-carcinoembryonic antigen antibody 431. The location of the signal peptide sequence (s) and the light chain variable (Vk) and the Ck regions are also indicated. The location of the restriction enzyme sites is also indicated.

Figure 1B depicts the nucleotide and amino acid sequence for the light chain of humanized anti-carcinoembryonic antigen antibody 431. The location of the restriction enzyme sites is indicated.

Figure 2 depicts the nucleotide sequence for a Sal I insert containing the coding sequences for light chain of humanized anti-carcinoembryonic antigen antibody 431.

Figure 3 is a schematic diagram of a construct (Bc 458) which includes the Sal I insert containing the coding sequences for light chain of humanized anti-carcinoembryonic antigen antibody 431. Also indicated is the location of the silencer, 5' β-casein untranslated region, the light chain coding region, and the 3' β-casein untranslated region.

Figure 4A is a schematic diagram of a construct containing the genomic sequence of the heavy chain (HC) of humanized anti-carcinoembryonic antigen antibody 431 linked to the β-glucuronidase sequence. The location of the signal peptide sequence (s) and the heavy chain variable (Vh) and CH1 are also indicated. The location of the restriction enzyme sites is also indicated.

Figure 4B depicts the nucleotide and amino acid sequence for the heavy chain of humanized anti-carcinoembryonic antigen antibody 431. The location of the restriction enzyme sites is indicated.

Figure 5 depicts the nucleotide and amino acid sequence for the mutant heavy chain of humanized anti-carcinoembryonic antigen antibody 431. The mutant heavy chain lacks the hinge region. The location of the restriction enzyme sites is indicated.

Figure 6 is a schematic diagram of a construct (Bc 454) containing the mutant heavy chain of humanized anti-carcinoembryonic antigen antibody 431 linked to the β-glucuronidase sequence. The location of the silencer, 5' β-casein untranslated region, the heavy chain mutant/β-glucuronidase fusion coding region, and the 3' β-casein untranslated region. The location of the restriction enzyme sites is also indicated.

Figure 7 is an overview of the construction of the heavy chain mutants.

Figure 8 is an enlarged view of the mutations to β-glucuronidase

The present invention provides, at least in part, transgenically produced fusion proteins wherein one member of the fusion protein assembles into a multimer and the other member is chosen, or modified, to promote assembly into the optimal number of subunits. In one embodiment, the fusion protein includes an immunoglobulin subunit (e.g., an immunoglobulin heavy or light chain) fused to a toxin (e.g., a subunit of an enzyme). The immunogloblulin-enzyme fusion proteins described herein serve to target a cytotoxic agent (e.g. the enzyme) to an undesirable cell, e.g., a tumor cell. For example, the fusion proteins

described in the Examples below, (i.e., an antibody against carcinoembryonic antigen (CEA) fused to an enzyme, e.g., glucuronidase) can be used to target, to a tumor cell. After allowing sufficient time for the immunoglobulin-enzyme fusion to localize at the tumor site, a non-toxic prodrug can be administered. This prodrug is converted to a highly cytotoxic drug by the action of the targeted enzyme localized at the tumor site, permitting to achieve therapeutic levels of the drug without unacceptable toxicity for the patients.

Production of Immunoglobulins

A monoclonal antibody against a target antigen, e.g., a cell surface protein (e.g., receptor) on a cell can be produced by a variety of techniques, including conventional monoclonal antibody methodology e.g., the standard somatic cell hybridization technique of Kohler and Milstein, Nature 256: 495 (1975). Although somatic cell hybridization procedures are preferred, in principle, other techniques for producing monoclonal antibody can be employed e.g., viral or oncogenic transformation of B lymphocytes.

The preferred animal system for preparing hybridomas is the murine system. Hybridoma production in the mouse is a very well-established procedure. Immunization protocols and techniques for isolation of immunized splenocytes for fusion are known in the art. Fusion partners (e.g., murine myeloma cells) and fusion procedures are also known.

Human monoclonal antibodies (mAbs) directed against human proteins can be generated using transgenic mice carrying the complete human immune system rather than the mouse system. Splenocytes from these transgenic mice immunized with the antigen of interest are used to produce hybridomas that secrete human mAbs with specific affinities for epitopes from a human protein (see, e.g., Wood et al. International Application WO 91/00906, Kucherlapati et al. PCT publication WO 91/10741; Lonberg et al. International Application WO 92/03918; Kay et al. International Application 92/03917; Lonberg, N. et al. 1994 *Nature* 368:856-859; Green, L.L. et al. 1994 *Nature Genet.* 7:13-21; Morrison, S.L. et al. 1994 *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Bruggeman et al. 1993 *Year Immunol* 7:33-40; Tuaillon et al. 1993 *PNAS* 90:3720-3724; Bruggeman et al. 1991 *Eur J Immunol* 21:1323-1326).

Monoclonal antibodies can also be generated by other methods known to those skilled in the art of recombinant DNA technology. An alternative method, referred to as the "combinatorial antibody display" method, has been developed to identify and isolate antibody fragments having a particular antigen specificity, and can be utilized to produce monoclonal antibodies (for descriptions of combinatorial antibody display see e.g., Sastry et al. 1989 *PNAS* 86:5728; Huse et al. 1989 *Science* 246:1275; and Orlandi et al. 1989 *PNAS* 86:3833). After immunizing an animal with an immunogen as described above, the antibody repertoire of the resulting B-cell pool is cloned. Methods are generally known for obtaining the DNA sequence of the variable regions of a diverse population of immunoglobulin molecules by using a mixture of oligomer primers and PCR. For instance, mixed oligonucleotide primers corresponding to the 5' leader (signal peptide) sequences and/or framework 1 (FR1) sequences, as well as primer to a conserved 3' constant region primer can be used for PCR amplification of the heavy and light chain variable regions from a number of murine antibodies (Larrick et al.,1991, *Biotechniques* 11:152-156). A similar strategy can also been used to amplify human heavy and light chain variable regions from human antibodies (Larrick et al., 1991, *Methods: Companion to Methods in Enzymology* 2:106-110).

In an illustrative embodiment, RNA is isolated from B lymphocytes, for example, peripheral blood cells, bone marrow, or spleen preparations, using standard protocols (e.g., U.S. Patent No. 4,683,202; Orlandi, et al. *PNAS* (1989) 86:3833-3837; Sastry et al., *PNAS* (1989) 86:5728-5732; and Huse et al. (1989) *Science* 246:1275-1281.) First-strand cDNA is synthesized using primers specific for the constant region of the heavy chain(s) and each of the κ and λ light chains, as well as primers for the signal sequence. Using variable region PCR primers, the variable regions of both heavy and light chains are amplified, each alone or in combinantion, and ligated into appropriate vectors for further manipulation in generating the display packages. Oligonucleotide primers useful in amplification protocols may be unique or degenerate or incorporate inosine at degenerate positions. Restriction endonuclease recognition sequences may also be incorporated into the primers to allow for the cloning of the amplified fragment into a vector in a predetermined reading frame for expression.

The V-gene library cloned from the immunization-derived antibody repertoire can be expressed by a population of display packages, preferably derived from filamentous phage, to form an antibody display library. Ideally, the display package comprises a system that allows the sampling of very large variegated antibody display libraries, rapid sorting after each affinity separation round, and easy isolation of the antibody gene from purified display packages. In addition to commercially available kits for generating phage display libraries (e.g., the Pharmacia *Recombinant Phage Antibody System*, catalog no. 27-9400-01; and the Stratagene *SurfZAP*^{TM} phage display kit, catalog no. 240612), examples of methods and reagents particularly amenable for use in generating a variegated antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. International Publication No. WO 92/18619; Dower et al. International Publication No. WO 91/17271; Winter et al. International Publication WO 92/20791; Markland et al. International Publication No. WO 92/15679; Breitling et al. International Publication WO 93/01288; McCafferty et al. International Publication No. WO 92/01047; Garrard et al. International Publication No. WO 92/09690; Ladner et al. International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum Antibod Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J Mol Biol* 226:889-896; Clackson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc Acid Res* 19:4133-4137; and Barbas et al. (1991) *PNAS* 88:7978-7982.

In certain embodiments, the V region domains of heavy and light chains can be expressed on the same polypeptide, joined by a flexible linker to form a single-chain Fv fragment, and the scFV gene subsequently cloned into the desired expression vector or phage genome. As generally described in McCafferty et al., *Nature* (1990) 348:552-554, complete V_H and V_L domains of an antibody, joined by a flexible $(Gly_4\text{-}Ser)_3$ linker can be used to produce a single chain antibody which can render the display package separable based on antigen affinity. Isolated scFV antibodies immunoreactive with the antigen can subsequently be formulated into a pharmaceutical preparation for use in the subject method.

Once displayed on the surface of a display package (e.g., filamentous phage), the antibody library is screened with the target antigen, or peptide fragment thereof, to identify and isolate packages that express an antibody having specificity for the target antigen. Nucleic acid encoding the selected antibody can be recovered from the display package (e.g., from the phage genome) and subcloned into other expression vectors by standard recombinant DNA techniques.

Specific antibody molecules with high affinities for a surface protein can be made according to methods known to those in the art, e.g, methods involving screening of libraries (Ladner, R.C., *et al.*, U.S. Patent 5,233,409; Ladner, R.C., *et al.*, U.S. Patent 5,403,484). Further, the methods of these libraries can be used in screens to obtain binding determinants that are mimetics of the structural determinants of antibodies.

In particular, the Fv binding surface of a particular antibody molecule interacts with its target ligand according to principles of protein-protein interactions, hence sequence data for V_H and V_L (the latter of which may be of the κ or λ chain type) is the basis for protein engineering techniques known to those with skill in the art. Details of the protein surface that comprises the binding determinants can be obtained from antibody sequence information, by a modeling procedure using previously determined three-dimensional structures from other antibodies obtained from NMR studies or crytallographic data. See for example Bajorath, J. and S. Sheriff, 1996, *Proteins: Struct., Funct., and Genet. 24 (2)*, 152-157; Webster, D.M. and A. R. Rees, 1995, "Molecular modeling of antibody-combining sites,"in S. Paul, Ed., *Methods in Molecular Biol. 51*, Antibody Engineering Protocols, Humana Press, Totowa, NJ, pp 17-49; and Johnson, G., Wu, T.T. and E.A. Kabat, 1995, "Seqhunt: A program to screen aligned nucleotide and amino acid sequences," in *Methods in Molecular Biol.51, op. cit.*, pp 1-15.

In one embodiment, a variegated peptide library is expressed by a population of display packages to form a peptide display library. Ideally, the display package comprises a system that allows the sampling of very large variegated peptide display libraries, rapid sorting after each affinity separation round, and easy isolation of the peptide-encoding gene from purified display packages. Peptide display libraries can be in, e.g., prokaryotic organisms and viruses, which can be amplified quickly, are relatively easy to manipulate,

and which allows the creation of large number of clones. Preferred display packages include, for example, vegetative bacterial cells, bacterial spores, and most preferably, bacterial viruses (especially DNA viruses). However, the present invention also contemplates the use of eukaryotic cells, including yeast and their spores, as potential display packages. Phage display libraries are described above.

Other techniques include affinity chromatography with an appropriate "receptor", e.g., a target antigen, followed by identification of the isolated binding agents or ligands by conventional techniques (e.g., mass spectrometry and NMR). Preferably, the soluble receptor is conjugated to a label (e.g., fluorophores, colorimetric enzymes, radioisotopes, or luminescent compounds) that can be detected to indicate ligand binding. Alternatively, immobilized compounds can be selectively released and allowed to diffuse through a membrane to interact with a receptor.

Combinatorial libraries of compounds can also be synthesized with "tags" to encode the identity of each member of the library (see, e.g., W.C. Still *et al.*, International Application WO 94/08051). In general, this method features the use of inert but readily detectable tags, that are attached to the solid support or to the compounds. When an active compound is detected, the identity of the compound is determined by identification of the unique accompanying tag. This tagging method permits the synthesis of large libraries of compounds which can be identified at very low levels among to total set of all compounds in the library.

The term modified antibody is also intended to include antibodies, such as monoclonal antibodies, chimeric antibodies, and humanized antibodies which have been modified by, e.g., deleting, adding, or substituting portions of the antibody. For example, an antibody can be modified by deleting the hinge region, thus generating a monovalent antibody. Any modification is within the scope of the invention so long as the antibody has at least one antigen binding region specific.

Chimeric mouse-human monoclonal antibodies (i.e., chimeric antibodies) can be produced by recombinant DNA techniques known in the art. For example, a gene encoding the Fc constant region of a murine (or other species) monoclonal antibody molecule is digested with restriction enzymes to remove the region encoding the murine Fc, and the

equivalent portion of a gene encoding a human Fc constant region is substituted. (see Robinson et al., International Patent Publication PCT/US86/02269; Akira, et al., European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., International Application WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al., European Patent Application 125,023; Better et al. (1988 *Science* 240:1041-1043); Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al., 1987, *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al., 1988, *J. Natl Cancer Inst.* 80:1553-1559).

The chimeric antibody can be further humanized by replacing sequences of the Fv variable region which are not directly involved in antigen binding with equivalent sequences from human Fv variable regions. General reviews of humanized chimeric antibodies are provided by Morrison, S. L., 1985, *Science* 229:1202-1207 and by Oi et al., 1986, *BioTechniques* 4:214. Those methods include isolating, manipulating, and expressing the nucleic acid sequences that encode all or part of immunoglobulin Fv variable regions from at least one of a heavy or light chain. Sources of such nucleic acid are well known to those skilled in the art and, for example, may be obtained from 7E3, an anti-$GPIIb_bIII_a$ antibody producing hybridoma. The recombinant DNA encoding the chimeric antibody, or fragment thereof, can then be cloned into an appropriate expression vector. Suitable humanized antibodies can alternatively be produced by CDR substitution U.S. Patent 5,225,539; Jones et al. 1986 *Nature* 321:552-525; Verhoeyan et al. 1988 *Science* 239:1534; and Beidler et al. 1988 *J. Immunol.* 141:4053-4060.

All of the CDRs of a particular human antibody may be replaced with at least a portion of a non-human CDR or only some of the CDRs may be replaced with non-human CDRs. It is only necessary to replace the number of CDRs required for binding of the humanized antibody to the Fc receptor.

An antibody can be humanized by any method, which is capable of replacing at least a portion of a CDR of a human antibody with a CDR derived from a non-human antibody. Winter describes a method which may be used to prepare the humanized antibodies of the present invention (UK Patent Application GB 2188638A, filed on March 26, 1987), the

contents of which is expressly incorporated by reference. The human CDRs may be replaced with non-human CDRs using oligonucleotide site-directed mutagenesis.

Also within the scope of the invention are chimeric and humanized antibodies in which specific amino acids have been substituted, deleted or added. In particular, preferred humanized antibodies have amino acid substitutions in the framework region, such as to improve binding to the antigen. For example, in a humanized antibody having mouse CDRs, amino acids located in the human framework region can be replaced with the amino acids located at the corresponding positions in the mouse antibody. Such substitutions are known to improve binding of humanized antibodies to the antigen in some instances. Antibodies in which amino acids have been added, deleted, or subsituted are referred to herein as modified antibodies or altered antibodies.

Target Antigens

In preferred embodiments, the first member of the fusion proteins of the present invention is a targeting agent, e.g., a polypeptide having a high affinity for a target, e.g., an antibody, a ligand, or an enzyme. Accordingly, the fusion proteins of the invention can be used to selectively direct (e.g., localize) the second member of the fusion protein to the vicinity of an undesirable cell.

For example, the first member can be an immunoglobulin that interacts with (e.g., binds to a target antigen). In certain embodiments, the target antigen is present on the surface of a cell, e.g., an aberrant cell such a hyperproliferative cell (e.g., a cancer cell). Exemplary target antigens include carcinoembryonic antigen (CEA), TAG-72, her-2/neu, epidermal growth factor receptor, transferrin receptor, among others.

As used herein, "target cell" shall mean any undesirable cell in a subject (e.g., a human or animal) that can be targeted by a fusion protein of the invention. Exemplary target cells include tumor cells, such as carcinoma or adenocarcinoma-derived cells (e.g., colon, breast, prostate, ovarian and endometrial cancer cells) (Thor, A. *et al.* (1997) *Cancer Res* 46: 3118; Soisson A. P. *et al.* (1989) *Am. J. Obstet. Gynecol.*:1258-63). The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary

system carcinomas, testicular carcinomas, breast carcinomas, ovarian carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

Production of Fusion Proteins

The first and second members of the fusion protein can be linked to each other, preferably via a linker sequence. The linker sequence should separate the first and second members of the fusion protein by a distance sufficient to ensure that each member properly folds into its secondary and tertiary structures. Preferred linker sequences (1) should adopt a flexible extended conformation, (2) should not exhibit a propensity for developing an ordered secondary structure which could interact with the functional first and second members, and (3) should have minimal hydrophobic or charged character, which could promote interaction with the functional protein domains. Typical surface amino acids in flexible protein regions include Gly, Asn and Ser. Permutations of amino acid sequences containing Gly, Asn and Ser would be expected to satisfy the above criteria for a linker sequence. Other near neutral amino acids, such as Thr and Ala, can also be used in the linker sequence.

A linker sequence length of 20 amino acids can be used to provide a suitable separation of functional protein domains, although longer or shorter linker sequences may also be used. The length of the linker sequence separating the first and second members can be from 5 to 500 amino acids in length, or more preferably from 5 to 100 amino acids in length. Preferably, the linker sequence is from about 5-30 amino acids in length. In preferred embodiments, the linker sequence is from about 5 to about 20 amino acids, and is advantageously from about 10 to about 20 amino acids. Amino acid sequences useful as

linkers of the first and second member include, but are not limited to, $(SerGly_4)_y$ wherein y is greater than or equal to 8, or $Gly_4SerGly_5Ser$. A preferred linker sequence has the formula $(SerGly_4)_4$. Another preferred linker has the sequence ((Ser-Ser-Ser-Ser-Gly)$_3$-Ser-Pro).

The first and second members can be directly fused without a linker sequence. Linker sequences are unnecessary where the proteins being fused have non-essential N-or C-terminal amino acid regions which can be used to separate the functional domains and prevent steric interference. In preferred embodiments, the C-terminus of first member can be directly fused to the N-terminus of second, or viceversa.

Recombinant Production

A fusion protein of the invention can be prepared with standard recombinant DNA techniques using a nucleic acid molecule encoding the fusion protein. A nucleotide sequence encoding a fusion protein can be synthesized by standard DNA synthesis methods.

A nucleic acid encoding a fusion protein can be introduced into a host cell, e.g., a cell of a primary or immortalized cell line. The recombinant cells can be used to produce the fusion protein. A nucleic acid encoding a fusion protein can be introduced into a host cell, e.g., by homologous recombination. In most cases, a nucleic acid encoding the fusion protein is incorporated into a recombinant expression vector.

The nucleotide sequence encoding a fusion protein can be operatively linked to one or more regulatory sequences, selected on the basis of the host cells to be used for expression. The term "operably linked" means that the sequences encoding the fusion protein compound are linked to the regulatory sequence(s) in a manner that allows for expression of the fusion protein. The term "regulatory sequence" refers to promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990), the content of which are incorporated herein by reference. Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cells, those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences) and those that direct expression in a regulatable manner (*e.g.*, only in

the presence of an inducing agent). It will be appreciated by those skilled in the art that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression of fusion protein desired, and the like. The fusion protein expression vectors can be introduced into host cells to thereby produce fusion proteins encoded by nucleic acids.

Recombinant expression vectors can be designed for expression of fusion proteins in prokaryotic or eukaryotic cells. For example, fusion proteins can be expressed in bacterial cells such as *E. coli*, insect cells (e.g., in the baculovirus expression system), yeast cells or mammalian cells. Some suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.*, (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of fusion proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, (1983) *Mol. Cell. Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39).

Examples of mammalian expression vectors include pCDM8 (Seed, B., (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987), *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

In addition to the regulatory control sequences discussed above, the recombinant expression vector can contain additional nucleotide sequences. For example, the recombinant expression vector may encode a selectable marker gene to identify host cells that have incorporated the vector. Moreover, to facilitate secretion of the fusion protein from a host cell, in particular mammalian host cells, the recombinant expression vector can encode a signal sequence operatively linked to sequences encoding the amino-terminus of the fusion protein such that upon expression, the fusion protein is synthesized with the signal sequence fused to its amino terminus. This signal sequence

directs the fusion protein into the secretory pathway of the cell and is then cleaved, allowing for release of the mature fusion protein (*i.e.*, the fusion protein without the signal sequence) from the host cell. Use of a signal sequence to facilitate secretion of proteins or peptides from mammalian host cells is known in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells *via* conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, electroporation, microinjection and viral-mediated transfection. Suitable methods for transforming or transfecting host cells can be found in Sambrook *et al.* (*Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory manuals.

Often only a small fraction of mammalian cells integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) can be introduced into the host cells along with the gene encoding the fusion protein. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding the fusion protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Transgenic Mammals

Methods for generating non-human transgenic animals are described herein. DNA constructs can be introduced into the germ line of a mammal to make a transgenic mammal. For example, one or several copies of the construct can be incorporated into the genome of a mammalian embryo by standard transgenic techniques.

It is often desirable to express the transgenic protein in the milk of a transgenic mammal. Mammals that produce large volumes of milk and have long lactating periods are preferred. Preferred mammals are ruminants, e.g., cows, sheep, camels or goats, e.g., goats of Swiss origin, e.g., the Alpine, Saanen and Toggenburg breed goats. Other preferred animals include oxen, rabbits and pigs.

In an exemplary embodiment, a transgenic non-human animal is produced by introducing a transgene into the germline of the non-human animal. Transgenes can be introduced into embryonal target cells at various developmental stages. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used should, if possible, be selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness.

Introduction of the fusion protein transgene into the embryo can be accomplished by any of a variety of means known in the art such as microinjection, electroporation, or lipofection. For example, a fusion protein transgene can be introduced into a mammal by microinjection of the construct into the pronuclei of the fertilized mammalian egg(s) to cause one or more copies of the construct to be retained in the cells of the developing mammal(s). Following introduction of the transgene construct into the fertilized egg, the egg can be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. One common method is to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

The progeny of the transgenically manipulated embryos can be tested for the presence of the construct by Southern blot analysis of a segment of tissue. An embryo having one or more copies of the exogenous cloned construct stably integrated into the genome can be used to establish a permanent transgenic mammal line carrying the transgenically added construct.

Litters of transgenically altered mammals can be assayed after birth for the incorporation of the construct into the genome of the offspring. This can be done by hybridizing a probe corresponding to the DNA sequence coding for the fusion protein or a segment thereof onto chromosomal material from the progeny. Those mammalian progeny found to contain at least one copy of the construct in their genome are grown to maturity.

The female species of these progeny will produce the desired protein in or along with their milk. The transgenic mammals can be bred to produce other transgenic progeny useful in producing the desired proteins in their milk.

Transgenic females may be tested for protein secretion into milk, using an art-known assay technique, e.g., a Western blot or enzymatic assay.

Other Transgenic Animals

Fusion protein can be expressed from a variety of transgenic animals. A protocol for the production of a transgenic pig can be found in White and Yannoutsos, *Current Topics in Complement Research: 64th Forum in Immunology*, pp. 88-94; US Patent No. 5,523,226; US Patent No. 5,573,933; PCT Application WO93/25071; and PCT Application WO95/04744. A protocol for the production of a transgenic mouse can be found in US Patent No. 5,530,177. A protocol for the production of a transgenic rat can be found in Bader and Ganten, *Clinical and Experimental Pharmacology and Physiology*, Supp. 3:S81-S87, 1996. A protocol for the production of a transgenic cow can be found in *Transgenic Animal Technology, A Handbook*, 1994, ed., Carl A. Pinkert, Academic Press, Inc. A protocol for the production of a transgenic sheep can be found in *Transgenic Animal Technology, A Handbook*, 1994, ed., Carl A. Pinkert, Academic Press, Inc. A protocol for the production of a transgenic rabbit can be found in Hammer et al., *Nature* 315:680-683, 1985 and Taylor and Fan, *Frontiers in Bioscience* 2:d298-308, 1997.

Production of Transgenic Protein in the Milk of a Transgenic Animal

Milk Specific Promoters

Useful transcriptional promoters are those promoters that are preferentially activated in mammary epithelial cells, including promoters that control the genes encoding milk proteins such as caseins, beta lactoglobulin (Clark et al., (1989) *Bio/Technology* 7: 487-492), whey acid protein (Gorton et al. (1987) *Bio/Technology* 5: 1183-1187), and lactalbumin (Soulier et al., (1992) *FEBS Letts.* 297: 13). The alpha, beta, gamma or kappa casein gene promoter of any mammalian species can be used to provide mammary

expression; a preferred promoter is the goat beta casein gene promoter (DiTullio, (1992) *Bio/Technology* 10:74-77). Milk-specific protein promoter or the promoters that are specifically activated in mammary tissue can be isolated from cDNA or genomic sequences. Preferably, they are genomic in origin.

DNA sequence information is available for mammary gland specific genes listed above, in at least one, and often in several organisms. See, e.g., Richards et al., *J. Biol. Chem.* 256, 526-532 (1981) (α-lactalbumin rat); Campbell et al., *Nucleic Acids Res.* 12, 8685-8697 (1984) (rat WAP); Jones et al., *J. Biol. Chem.* 260, 7042-7050 (1985) (rat β-casein); Yu-Lee & Rosen, *J. Biol. Chem.* 258, 10794-10804 (1983) (rat γ-casein); Hall, *Biochem. J.* 242, 735-742 (1987) (α-lactalbumin human); Stewart, *Nucleic Acids Res.* 12, 389 (1984) (bovine αs1 and κ casein cDNAs); Gorodetsky et al., *Gene* 66, 87-96 (1988) (bovine β casein); Alexander et al., *Eur. J. Biochem.* 178, 395-401 (1988) (bovine κ casein); Brignon et al., *FEBS Lett.* 188, 48-55 (1977) (bovine αS2 casein); Jamieson et al., *Gene* 61, 85-90 (1987), Ivanov et al., *Biol. Chem.* Hoppe-Seyler 369, 425-429 (1988), Alexander et al., *Nucleic Acids Res.* 17, 6739 (1989) (bovine β lactoglobulin); Vilotte et al., *Biochimie* 69, 609-620 (1987) (bovine α-lactalbumin). The structure and function of the various milk protein genes are reviewed by Mercier & Vilotte, *J. Dairy Sci.* 76, 3079-3098 (1993) (incorporated by reference in its entirety for all purposes). If additional flanking sequence are useful in optimizing expression, such sequences can be cloned using the existing sequences as probes. Mammary-gland specific regulatory sequences from different organisms can be obtained by screening libraries from such organisms using known cognate nucleotide sequences, or antibodies to cognate proteins as probes.

Signal Sequences

Useful signal sequences are milk-specific signal sequences or other signal sequences which result in the secretion of eukaryotic or prokaryotic proteins. Preferably, the signal sequence is selected from milk-specific signal sequences, i.e., it is from a gene which encodes a product secreted into milk. Most preferably, the milk-specific signal sequence is related to the milk-specific promoter used in the expression system of this invention. The size of the signal sequence is not critical for this invention. All that is required is that the

sequence be of a sufficient size to effect secretion of the desired recombinant protein, e.g., in the mammary tissue. For example, signal sequences from genes coding for caseins, e.g., alpha, beta, gamma or kappa caseins, beta lactoglobulin, whey acid protein, and lactalbumin are useful in the present invention. A preferred signal sequence is the goat β-casein signal sequence.

Signal sequences from other secreted proteins, e.g., immunoglobulins, or proteins secreted by liver cells, kidney cell, or pancreatic cells can also be used.

Insulator Sequences

The DNA constructs of the invention further comprise at least one insulator sequence. The terms "insulator", "insulator sequence" and "insulator element" are used interchangeably herein. An insulator element is a control element which insulates the transcription of genes placed within its range of action but which does not perturb gene expression, either negatively or positively. Preferably, an insulator sequence is inserted on either side of the DNA sequence to be transcribed. For example, the insulator can be positioned about 200 bp to about 1 kb, 5' from the promoter, and at least about 1 kb to 5 kb from the promoter, at the 3' end of the gene of interest. The distance of the insulator sequence from the promoter and the 3' end of the gene of interest can be determined by those skilled in the art, depending on the relative sizes of the gene of interest, the promoter and the enhancer used in the construct. In addition, more than one insulator sequence can be positioned 5' from the promoter or at the 3' end of the transgene. For example, two or more insulator sequences can be positioned 5' from the promoter. The insulator or insulators at the 3' end of the transgene can be positioned at the 3' end of the gene of interest, or at the 3'end of a 3' regulatory sequence, e.g., a 3' untranslated region (UTR) or a 3' flanking sequence.

A preferred insulator is a DNA segment which encompasses the 5' end of the chicken β-globin locus and corresponds to the chicken 5' constitutive hypersensitive site as described in PCT Publication 94/23046, the contents of which is incorporated herein by reference.

DNA Constructs

A fusion protein can be expressed from a construct which includes a promoter specific for mammary epithelial cells, e.g., a casein promoter, e.g., a goat beta casein promoter, a milk-specific signal sequence, e.g., a casein signal sequence, e.g., a β-casein signal sequence, and a DNA encoding a fusion protein.

A construct can also include a 3' untranslated region downstream of the DNA sequence coding for the non-secreted protein. Such regions can stabilize the RNA transcript of the expression system and thus increases the yield of desired protein from the expression system. Among the 3' untranslated regions useful in the constructs of this invention are sequences that provide a poly A signal. Such sequences may be derived, e.g., from the SV40 small t antigen, the casein 3' untranslated region or other 3' untranslated sequences well known in the art. Preferably, the 3' untranslated region is derived from a milk specific protein. The length of the 3' untranslated region is not critical but the stabilizing effect of its poly A transcript appears important in stabilizing the RNA of the expression sequence.

A construct can include a 5' untranslated region between the promoter and the DNA sequence encoding the signal sequence. Such untranslated regions can be from the same control region from which promoter is taken or can be from a different gene, e.g., they may be derived from other synthetic, semi-synthetic or natural sources. Again their specific length is not critical, however, they appear to be useful in improving the level of expression.

A construct can also include about 10%, 20%, 30%, or more of the N-terminal coding region of a gene preferentially expressed in mammary epithelial cells. For example, the N-terminal coding region can correspond to the promoter used, e.g., a goat β-casein N-terminal coding region.

Prior art methods can include making a construct and testing it for the ability to produce a product in cultured cells prior to placing the construct in a transgenic animal. Surprisingly, the inventors have found that such a protocol may not be of predictive value in determining if a normally non-secreted protein can be secreted, e.g., in the milk of a transgenic animal. Therefore, it may be desirable to test constructs directly in transgenic animals, e.g., transgenic mice, as some constructs which fail to be secreted in CHO cells are secreted into the milk of transgenic animals.

Purification from milk

The transgenic fusion protein can be produced in milk at relatively high concentrations and in large volumes, providing continuous high level output of normally processed peptide that is easily harvested from a renewable resource. There are several different methods known in the art for isolation of proteins from milk.

Milk proteins usually are isolated by a combination of processes. Raw milk first is fractionated to remove fats, for example, by skimming, centrifugation, sedimentation (H.E. Swaisgood, Developments in Dairy Chemistry, I: Chemistry of Milk Protein, Applied Science Publishers, NY, 1982), acid precipitation (U.S. Patent No. 4,644,056) or enzymatic coagulation with rennin or chymotrypsin (Swaisgood, *ibid.)*. Next, the major milk proteins may be fractionated into either a clear solution or a bulk precipitate from which the specific protein of interest may be readily purified.

USSN 08/648,235 discloses a method for isolating a soluble milk component, such as a peptide, in its biologically active form from whole milk or a milk fraction by tangential flow filtration. Unlike previous isolation methods, this eliminates the need for a first fractionation of whole milk to remove fat and casein micelles, thereby simplifying the process and avoiding losses of recovery and bioactivity. This method may be used in combination with additional purification steps to further remove contaminants and purify the component of interest.

Production of Transgenic Protein in the Eggs of a Transgenic Animal

A fusion protein can be produced in tissues, secretions, or other products, e.g., an egg, of a transgenic animal. For example, fusion proteins can be produced in the eggs of a transgenic animal, preferably a transgenic turkey, duck, goose, ostrich, guinea fowl, peacock, partridge, pheasant, pigeon, and more preferably a transgenic chicken, using methods known in the art (Sang et al., Trends Biotechnology, 12:415-20, 1994). Genes encoding proteins specifically expressed in the egg, such as yolk-protein genes and albumin-protein genes, can be modified to direct expression of fusion protein.

Egg Specific Promoters

Useful transcriptional promoters are those promoters that are preferentially activated in the egg, including promoters that control the genes encoding egg proteins, e.g., ovalbumin, lysozyme and avidin. Promoters from the chicken ovalbumin, lysozyme or avidin genes are preferred. Egg-specific protein promoters or the promoters that are specifically activated in egg tissue can be from cDNA or genomic sequences. Preferably, the egg-specific promoters are genomic in origin.

DNA sequences of egg specific genes are known in the art (see, e.g., Burley et al., "The Avian Egg", John Wiley and Sons, p. 472, 1989, the contents of which are incorporated herein by reference). If additional flanking sequence are useful in optimizing expression, such sequences can be cloned using the existing sequences as probes. Egg specific regulatory sequences from different organisms can be obtained by screening libraries from such organisms using known cognate nucleotide sequences, or antibodies to cognate proteins as probes.

Transgenic Plants

A fusion protein can be expressed in a transgenic organism, e.g., a transgenic plant, e.g., a transgenic plant in which the DNA transgene is inserted into the nuclear or plastidic genome. Plant transformation is known as the art. See, in general, *Methods in Enzymology* Vol. 153 ("Recombinant DNA Part D") 1987, Wu and Grossman Eds., Academic Press and European Patent Application EP 693554.

Foreign nucleic acid can be introduced into plant cells or protoplasts by several methods. For example, nucleic acid can be mechanically transferred by microinjection directly into plant cells by use of micropipettes. Foreign nucleic acid can also be transferred into a plant cell by using polyethylene glycol which forms a precipitation complex with the genetic material that is taken up by the cell (Paszkowski et al. (1984) *EMBO J.* 3:2712-22). Foreign nucleic acid can be introduced into a plant cell by electroporation (Fromm et al. (1985) *Proc. Natl. Acad. Sci. USA* 82:5824). In this technique, plant protoplasts are electroporated in the presence of plasmids or nucleic acids containing the relevant genetic construct. Electrical impulses of high field strength reversibly permeabilize biomembranes

allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form a plant callus. Selection of the transformed plant cells with the transformed gene can be accomplished using phenotypic markers.

Cauliflower mosaic virus (CaMV) can be used as a vector for introducing foreign nucleic acid into plant cells (Hohn et al. (1982) "Molecular Biology of Plant Tumors," Academic Press, New York, pp. 549-560; Howell, U.S. Pat. No. 4,407,956). CaMV viral DNA genome is inserted into a parent bacterial plasmid creating a recombinant DNA molecule which can be propagated in bacteria. The recombinant plasmid can be further modified by introduction of the desired DNA sequence. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the plant cells or plants.

High velocity ballistic penetration by small particles can be used to introduce foreign nucleic acid into plant cells. Nucleic acid is disposed within the matrix of small beads or particles, or on the surface (Klein et al. (1987) *Nature* 327:70-73). Although typically only a single introduction of a new nucleic acid segment is required, this method also provides for multiple introductions.

A nucleic acid can be introduced into a plant cell by infection of a plant cell, an explant, a meristem or a seed with *Agrobacterium tumefaciens* transformed with the nucleic acid. Under appropriate conditions, the transformed plant cells are grown to form shoots, roots, and develop further into plants. The nucleic acids can be introduced into plant cells, for example, by means of the Ti plasmid of *Agrobacterium tumefaciens*. The Ti plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and is stably integrated into the plant genome (Horsch et al. (1984) "Inheritance of Functional Foreign Genes in Plants," *Science* 233:496-498; Fraley et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:4803).

Plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed so that whole plants are recovered which contain the transferred foreign gene. Some suitable plants include, for example, species from the genera Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum,

Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciohorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Hererocallis, Nemesia, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Lolium, Zea, Triticum, Sorghum, and Datura.

Plant regeneration from cultured protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," *Handbook of Plant Cell Cultures* 1:124-176 (MacMillan Publishing Co. New York 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts* (1983)-Lecture Proceedings, pp. 12-29, (Birkhauser, Basal 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," *Protoplasts* (1983)-Lecture Proceedings, pp. 31-41, (Birkhauser, Basel 1983); and H. Binding, "Regeneration of Plants," *Plant Protoplasts*, pp. 21-73, (CRC Press, Boca Raton 1985).

Regeneration from protoplasts varies from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the exogenous sequence is first generated. In certain species, embryo formation can then be induced from the protoplast suspension, to the stage of ripening and germination as natural embryos. The culture media can contain various amino acids and hormones, such as auxin and cytokinins. It can also be advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In vegetatively propagated crops, the mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants for trialling, such as testing for production characteristics. Selection of a desirable transgenic plant is made and new varieties are obtained thereby, and propagated vegetatively for commercial sale. In seed propagated crops, the mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the gene for the newly introduced foreign gene activity level. These seeds can be grown to produce plants that have the selected phenotype. The inbreds according to this invention can be used

to develop new hybrids. In this method a selected inbred line is crossed with another inbred line to produce the hybrid.

Parts obtained from a transgenic plant, such as flowers, seeds, leaves, branches, fruit, and the like are covered by the invention, provided that these parts include cells which have been so transformed. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention, provided that these parts comprise the introduced DNA sequences. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention.

Selection of transgenic plants or plant cells can be based upon a visual assay, such as observing color changes (e.g., a white flower, variable pigment production, and uniform color pattern on flowers or irregular patterns), but can also involve biochemical assays of either enzyme activity or product quantitation. Transgenic plants or plant cells are grown into plants bearing the plant part of interest and the gene activities are monitored, such as by visual appearance (for flavonoid genes) or biochemical assays (Northern blots); Western blots; enzyme assays and flavonoid compound assays, including spectroscopy, see, Harborne et al. (Eds.), (1975) *The Flavonoids*, Vols. 1 and 2, [Acad. Press]). Appropriate plants are selected and further evaluated. Methods for generation of genetically engineered plants are further described in US Patent No. 5,283,184, US Patent No. 5, 482,852, and European Patent Application EP 693 554, all of which are hereby incorporated by reference.

Embodiments of the invention are further illustrated by the following examples which should not be construed as being limiting. The contents of all cited references (including literature references, issued patents, published patent applications, and co-pending patent applications) cited throughout this application are hereby expressly incorporated by reference.

Examples 1 and 2 below describe the generation of two constructs: a light chain construct and a heavy chain/β-glucuronidase fusion constructs. Two plasmids, one containing a clone of an antibody heavy chain/ human (β-glucuronidase fusion protein and

the other containing kappa light chain sequence were received obtained from Behringwerke AG.

EXAMPLE 1: Construction of Light Chain (LC) Construct

The Example describes the generation of a light chain nucleic acid construct using the light chain nucleotide sequence from a humanized monoclonal antibody against carcinoembryonic antigen (431) subcloned into a mammary specific expression vector (Bc163) and a commercial mammalian expression vector (pcDNA3).

Briefly, a Hind III -Eco RI fragment containing the light chain sequence was subcloned into pGEM3z to facilitate further manipulation. Two mutations were made:

a) To create a Sal I, Xho I, and Kozak consensus sequence at the beginning of the coding region; and

b) To creation a Sal I site immediately after the termination codon.

The original construct contained approximately 1300 bases of unknown sequence. To remove the unknown sequences, the Gapped Heteroduplex method was used to create a Sal I site just after the termination codon. Sac I sites just before the termination codon and near the Eco RI site were used to make the gap, which was filled using Klenow fragment, deoxynucleotides, T4 DNA ligase, and the following oligonucleotide:

TGT TAG AGG TCG ACG CCC CAC (SEQ ID NO:21)
term Sal I

The gapped region (through the termination codon and new Sal I site) was then sequenced to confirm that no changes were made in sequence.

A second Nco I site was found in the unknown sequence that was removed for a subsequent step described below. To remove this site, the construct containing the new Sal I site was digested with Eco RI, ends filled with Klenow fragment and deoxynucleotides, and ligated to a Sal I linker, purchased from New England Biolabs following routine experimental procedures. This construct containing two Sal I sites was then digested with Sal I-and religated, removing the unknown sequence containing the second Nco I site.

A Sal I site and Kozak consensus sequence were then inserted immediately before the initial methionine codon (instead of simply changing the Hind III site) because there were several ATG sequences prior to the correct starting codon that could possibly have been used as alternative start sites. While these ATG sequences do not seem to be a problem in tissue culture, the safest route was to remove this region. These ATG sequences were removed by excising the Hind III Nco I site and replacing it with a Hind III --Nco I adapter containing Sal I and Xho I and a Kozak consensus sequence. The replaced region was also confirmed by sequencing.

The sequence changes were as follows:

The original 5' prime region had the nucleotide sequence (ATG sequences are capitalized; ATG corresponding to initial methionine is indicated in bold):

aagctt ATG aat ATG caaatcctgctc ATG aat ATG caaatcctctga atctac ATG gtaaatataggtttgtctataccacaaacagaaaaac ATG agat cacagttctctctacagttactgagcacacaggacctcacc **ATG**
(SEQ ID NO:22)

The original sequence was replace with the following replacement sequence:

Hind III	Sal I	XhoI
AAGCTT	GTCGAC	CTCGAG CCACC**ATG**

Kozak (consensus sequence) (SEQ ID NO:23)

The Sal I fragment containing the entire coding region of the light chain was then subcloned into the Xho I site of Bc163, a mammary specific expression vector and pcDNA3, a commercial mammalian expression vector. Orientation was determined by restriction enzyme analysis and/or sequencing. Figure 1A is a schematic diagram of the light chain construct (431A). The nucleotide and amino acid sequences are shown in Figure 1B. Figure 2 depicts the nucleotide sequence for a Sal I insert containing the coding sequences

for light chain of humanized anti-carcinoembryonic antigen antibody 431. Shown as Figure 3 is a schematic diagram of a construct (Bc 458) which includes the Sal I insert containing the coding sequences for light chain of humanized anti-carcinoembryonic antigen antibody 431. Also indicated is the location of the silencer, 5' β-casein untranslated region, the light chain coding region, and the 3' β-casein untranslated region.

EXAMPLE 2: Construction of Heavy Chain/β-Glucuronidase Fusion Construct

The Example describes the generation of a heavy chain/β-glucuronidase fusion construct using the heavy chain nucleotide sequence from a humanized monoclonal antibody against carcinoembryonic antigen (431) subcloned into a mammary specific expression vector (Bc163) and a commercial mammalian expression vector (pcDNA3).

The Hind III -Xba I fragment containing the heavy chain/β-glucuronidase fusion sequence was subcloned into pGEM3z to facilitate further manipulation. Three mutations were made to the coding region of the heavy chain/β-glucuronidase fusion construct:

a) To create a Sal I, Xho I, and Kozak consensus sequence at the beginning of the coding region;

b) to change the sequence at the internal Sal I site while retaining thecorrect amino acid sequence; and

c) to create a Sal I site immediately after the termination codon.

The signal sequence that was used for the light chain was also used for the heavy chain. Again, the region between the Hind III and Nco I sites was removed and replaced with the same set of oligonucleotides used in the light chain to create a Sal I site and Kozak consensus sequence immediately before the initial methionine codon. (see above).

The internal Sal I site had to be changed for the purpose of subcloning the fragment into a beta casein expression vector.

	Asn Gly Val Asp Thr Leu (SEQ ID NO:24)
original sequence	AAT GGG GTC GAC ACG CTA (SEQ ID NO:25)
new sequence	GTG GAT (SEQ ID NO:26)

Val Asp

The 3-prime flanking sequence contained two polyadenylation signal sites and a string of 16 adenine residues between the translational stop codon and the Xba I site. To remove these sequences, a Sal I site was inserted just after the stop codon.

	Phe Thr * * *
original sequence	TTT ACT TGA GCA AGA CTG (SEQ ID NO:27)
new sequence	TTT ACT TGA GGT CGA CTG (SEQ ID NO:28)
	Sal I

The Gapped Heteroduplex method was used to make the changes above. The original plan was to gap the DNA between the Not I and Xba I sites and change the internal Sal I site and add the 3-prime Sal I site at the same time. This proved difficult to accomplish so the 3-prime Sal I site was added first and a new gap was made between the two Bgl II sites to change the internal Sal I site. The gapped regions were then sequenced in entirety to confirm that no changes were made to the sequence. The only difference found was in the fourth intron, 1673 bases from the initial ATG. A cytosine was found in both the mutated and the original plasmid instead of adenine, as shown in the printed sequence above. The Sal I fragment containing the entire coding region of the heavy chain -- glucuronidase fusion protein was then subcloned into the Xho I site of Bc163, a mammary specific expression vector and pcDNA3, a commercial mammalian expression vector. Orientation was determined by restriction enzyme analysis and/or sequencing.
Figure 4A is a schematic diagram of the light chain construct (431A). The nucleotide and amino acid sequences are shown in Figure 4B.

EXAMPLE 3: Generation of Linked Construct

This Example described the generation of a construct which includes the light chain and the heavy chain/β-glucuronidase fusion, along with their corresponding upstream and downstream beta casein sequences ligated together into a single cosmid.

In order to eliminate the possibility of integrating only one chain of a two chain protein, such as an antibody, that has been co-injected into mice or other species, both chains along with their own corresponding upstream and downstream beta casein sequences were ligated together into a single cosmid.

To achieve this, supercos1 (Stratagene) was modified by inserting the following oligonucleotides into the Bam HI site:

Pvu I Pvu I

T3... GAT CAC CGA TCG TCG ACC CCC TCG AGC GAT CGA... T7 (SEQ ID N0: 29)

TG GCT AGC AGC TGG GCG AGC TCG CTA GCT ACT AG (SEQ ID NO: 30)

Sal I XhoI

These modifications create a new supercos plamid, designated supercos 334, with unique Sal I and Xho I sites. Pvu I, Not I, and Eco RI sites flank these sites and the Bam HI site is destroyed.

The Sal I fragments from BC174 or BC175, containing the modified light chain and heavy chain/ β-glucuronidase coding regions within the beta casein 5-prime and 3-prime flanking regions respectively, were inserted into the Xho I site of supercos 334. Three clones were isolated and prepared. The orientation was determined by restriction enzyme analysis.

clone #	name	insert	orientation
1	LC14	LC	reverse
2	LC13	LC	sense
11	HC9	HC	reverse

The complementary Sal I fragments from BC174 and BC175 (used above) were then ligated into the Sal I site of the above constructions. (Heavy chain fragment into LC13 and LC14, light chain fragment into HC9). The resulting ligations were then large enough to package *in-vitro* into lambda phage particles (Amersham kit N. 334) and were used to infect *E. coli* XL1 Blue. Three versions were generated and one of each of these clones was isolated and prepared:

clone #	name	insert	orientation
1	Bc180	HC/LC	reverse/reverse
9	Bc181	HC/LC	sense /sense
20	Bc182	LC/HC	reverse/reverse

Although made through two different pathways, Bc181 and Bc182 are essentially the same insert when cut away from the vector. When viewed in the sense direction, they both contain the heavy chain/β-glucuronidase Sal I cassette followed by and linked to the light chain Sal I cassette. Each Sal I cassette contains the 5-prime beta casein promoter region, the antibody coding region, and the 3-prime beta casein flanking sequence.

In essence, two species were made: the light chain cassette followed by the heavy chain cassette, or the heavy chain cassette followed by the light chain cassette.

EXAMPLE 4: Characterization of the Light Chain and Heavy Chain/b-Glucuronidase Constructs

The manipulated DNA fragments were tested in tissue culture using the pcDNA3 constructs described above transfected into cos 7 cells using the standard protocol for Lipofectamine using Opti-MEM (Gibco-BRL). Conditioned media (DMEM +10%FBS) was removed after 48 hours and run on a 10 -20% SDS-PAGE gels for Western blotting.

Western Blots were conducted following standard procedures, Briefly, for the heavy chain/beta-glucuronidase, samples were run in triplicate under reducing conditions and electroblotted onto nitrocellulose. The nitrocellulose was then cut into three sections and incubated overnight with each of three monoclonal antibodies: Mab 2149/80, Mab 2156/94, and Mab 2156/215. The secondary antibody used for detection was from Cappel (cat. no. 55570), affinity purified horse radish peroxidase conjugated goat anti-mouse IgG. Detection was with the ECL kit from Amersham. Mab 2149/80 was the only antibody that showed a signal on the western blot.

For the light chain, samples were again run under reducing conditions and electroblotted onto nitrocellulose. The nitrocellulose was then incubated overnight with

horse radish peroxidase conjugated goat anti-human Kappa chain antibody (Cappel no. 55233). Detection was with the ECL kit from Amersham.

EXAMPLE 5: Production of Transgenic Animals

Microinjection fragments were prepared by cutting the beta casein constructs BC174 (light chain) and BC175 (heavy chain) with Sal I to release the bacterial sequences. Fragments were gel purified then buffer exchanged and concentrated using the Wizard system by Promega.

Microinjections of the original nucleotide sequences were tested in the mouse model system using an expression vector containing the goat beta casein upstream and coding sequences. Two separate constructions were made and co-injected into mouse embryos, from which founder lines were identified and tested further. The original DNA sequences were also co-injected with an "insulator" sequence which allows us to produce a higher percentage of high producing animal lines. For example, without the insulator generally one in three lines would be a relatively high producer. With the insulator, in many cases, almost all of the lines produced are high expressing lines.

Two sets of injections were carried out as follows:

For the first set of injections, 1249 embryos were injected of which 838 survived, and 737 were transferred to pseudopregnant females. From these females 80 live pups were born, of which 8 were transgenic, 7 of which carried both chains.

For the second set of injections, 508 embryos were injected of which 435 survived, and 426 were transferred to pseudopregnant females. From these females 44 live pups were born of which 2 were transgenic, both of which carried both chains.

BC181 was injected over three days. In this set, 840 embryos were injected of which 641 survived, and 618 were transferred to pseudopregnant females. From these females 39 live pups were born, of which 5 were transgenic, 3 of which carried both chains. Due to the repetition of the flanking beta casein sequences, it appears that in some cases recombination occurs deleting one chain or the other.

BC181 was co injected with the silencer fragment over four days. In this case, 1495 embryos were injected of which 1183 survived, and 1073 were transferred to pseudopregnant females. From these, 111 live pups were born and 10 of these were

transgenic, six carrying both chains. Two of the pups carry both the silencer fragment and both antibody chains.

EXAMPLE 6: Generation of Mutants of the Heavy chain/β-glucuronidase fusion Protein

In an attempt to increase expression of active molecules, two mutations to the heavy chain fusion protein were carried out. The first mutation was to remove the hinge region of the construct. The second mutation removes the hinge and linker sequence (ala-ala-ala-ala-val) (SEQ ID NO: 31) at the beginning of the β-glucuronidase coding sequence, fusing the CH2 portion to β glucuronidase.

To achieve this, gapped heteroduplex mutagenesis was again used. The construct Behring HC5 (which contains the fusion protein in pGem3Z with both ends modified and an internal Sal I site removed) was linearized with (Xba I). A second aliquot was cut with BstE2 plus Not I. When boiled together and cooled some of each strand anneal forming the heteroduplex containing a single stranded gap, in this case between the BstE2 and Not I sites. Two new constructs were then made, sequencing over the gapped portion to make sure no other mutations were made inadvertently.

GTC #403: using the oligonucleotide "Behr hinge-alternate" (in bold below) removes the hinge region and part of the introns immediately preceding and after it.

ccaaactctctactc**ACTCAGCTCA CGCATCCACC**tccatcccagatccccgt (SEQ ID NO: 32)
intron intron

GTC #406: using the oligonucleotide "Behr hinge/linker" (in bold below) removes the hinge region and ala-ala-ala-ala-val linker, fusing the CH2 and β-glucuronidase coding regions.

agcaacaccaaggtg**GACAAGAGAGTT CAGGGCGGGATG**ctgtacccccaggag
CH2 coding sequence β-glucuronidase (SEQ ID NO: 33)

The mutated fusion protein coding sequence can then be excised using Sal I and subcloned into an appropriate expression vector.

High levels of expression of the encoded proteins was obtained with a vector consisting of a silencer (or insulator) fragment followed by the goat beta casein promoter, insert DNA, and goat beta casein 3 prime untranslated regions. Both mutant heavy chains and the light chain have been subcloned into such a vector, Bc450, which is flanked by Sal I sites which release the entire injection fragment.

Bc454: Bc450 with heavy chain mutant 403 (minus hinge)

Bc456: Bc450 with heavy chain mutant 406 (minus hinge/linker)

Bc458: Bc450 with the light chain

Figure 5 depicts the nucleotide and amino acid sequence for the mutant heavy chain of humanized anti-carcinoembryonic antigen antibody 431 lacking the hinge region.

Figure 6 is a schematic diagram of a construct (BC 454) containing the mutant heavy chain of humanized anti-carcinoembryonic antigen antibody 431 linked to the β-glucuronidase sequence. The location of the silencer, 5' β-casein untranslated region, the heavy chain mutant/β-glucuronidase fusion coding region, and the 3' β-casein untranslated region.

EXAMPLE 7: Characterization of Transgenic Animals

The previous examples describe the testing of the original fusion protein and two heavy chain mutants in the milk expression system. The original fusion proteins were tested both without the insulator and also co injected with a separate insulator fragments. The heavy chain mutants, on the other hand, were tested with the insulator integrated into the construct.

Initially, the concentration of the fusion protein produced in milk was estimated by comparing the signal of a sample to that of a standard on a Western blot. Later, experiments measured activity rather than concentration based on Western blots. The activity measurements were more accurate.

Except for the first set of constructs, Bc174 + Bc175, estimates of protein concentration by Western blot are rough estimates. Generally, lines that express well appear to be in the 1-2 mg/ml range.

Expression data is summarized below, with more detained data sets for each construct attached.

Constructs	**DNA**	**Insulator**	**Western (HC) ug/ml estimated**	**Maximum activity ug/ml**
Bc174/Bc175	Original	no	~ 800	20
Bc181	original, linked	No	1000-2000	Na
Bc181 + insulator	original, linked	Co injected fragment	~ 1000	100
Bc456/Bc458	Minus hinge/linker	yes	1000-2000	8
Bc454/Bc458	Minus hinge	yes	1000-2000	800

Essentially the results shown herein indicate that while high levels of protein can be made in milk, most of this protein is not active. Such inactivity may be due to a folding problem or a problem in the assembly of the tetramer. Removal of the hinge and linker also produced a protein with low activity. In contast, substantial amount of enzymatic activity was achieved upon the removal of the hinge alone.

Approximately, 8 mg of this protein have been produced in mouse milk. The isolated protein is currently being tested in *in vivo* studies ("human CEA positive colon cancer metastasis model").

A summary of the data regarding the mice produced and analysis done follows, in table form.

A. Bc174/175 founders

Original DNA without the insulator

Line	2nd Gen.	Sex	PCR		Copy #		Western ug/ml	Activity ug/ml*
			LC	HC	LC	HC	HC	
2		F	+	+	10	10	---	0.14
4		F	+	+	1	1	---	<0.1
10		F	+	+	10	10	~800	18
	142	F	+	+			~800	
22		F	+	+	10	10	---	<0.1
	154	F	+	+			~800	
23		F	+	+	50	50	~400	4
	200	F	+	+			0.0	
40		M	+	+	100	100	---	<0.1
62		F	+	+	+	+	---	<0.1
81		M	+	---	---	---	n.a.	
85		F	+	+	25	25	---	.3
116		M	+	+	5	5		
	216	F	+	+			~800	
	221	F	+	+			~800	

n.a. = not analyzed (line carries only one chain)

B. Bc181 founders

Original DNA without the insulator; a fusion of the Bc 174 and Bc175 injection fragments

			PCR		Copy #		Western ug/ml (approx.)	Activity ug/ml*
Line	F1	Sex	LC	HC	LC	HC	HC	
6		M	+	+	12	12		
	49	F					0.0	
	50	F					0.0	
	52	F					>1000	39
	60	F					>1000	41
25		F	+	+	15	15	0.0	
29		F	---	+	n.a.	n.a.	0.0	
33		M	---	+	n.a.	n.a.	n.a.	
36		F	+	+	100	100	0.0	

n.a. = not analyzed

C. Bc181 + insulator founders

Original DNA (fusion of the Bc 174 and Bc 175 injection fragments) co-injected with the insulator

			PCR			Copy #			Western ug/ml (approx.)	Activity ug/ml*
Line	2nd gen.	Sex	LC	HC	Sil	LC	HC	sil	HC	
9		M	---	+	n.a.	0	1	---	n.a.	
13		M	+	---	n.a.	2	0	---	n.a.	
15		F	+	---	n.a.	3	0	---	n.a.	
33		F	+	+	n.a.	3	3	---	~800	
40		M	---	+	n.a.	0	2	---	n.a.	
58		M	+	+	n.a.	2	10	---		
	2-139	F	+	+	---				~800	43
	2-140	F	+	+	---				~800	31
66		F	+	+	n.a.	1	1	+		
78		F	+	+	n.a.	1	1	---	0.0	
81		M	+	+	n.a.	1	1	---	Not pass	
90		M	+	+	n.a.	20	20	+		
	2-123	F	+	+	+				~1000	~100
	2-124	F	+	+	+				~1000	60
	2-126	F	+	+	---				Low	15

n.a. = not analyzed

D. Bc456 + Bc458 founders

Mutation removing the hinge and linker:

1st Generation	2nd Generation	Sex	Western HC	Activity ug/ml.
6		F	none	0
8		F		0
13		F	none	0
18		F		0
24		F	none	0
57		F	good	2
65		F	good	8
66		F	low	0
138		M		
	175	F		
152		F	none	0

E. Bc454 + Bc458

Mutation removing the hinge only

1st Generation	2nd Generation	3rd Generation	Sex	Western HC	Activity ug/ml.
162 died	c-section		F	---	
	4		F	High	66
180			F	Good	177
	11		F		133
	12		F		185
182			M	---	
	15		F	Good	83
187			M	Not passing gene	
193			M	---	
	27		F	High	838
		57	F		829
		58	F		742
		59	F		944
		60	F		574
		61	F		752
		62	F		534
201			F	Good	416
215			F		(died)
219			M	---	
			F		
220			M	---	
			F		

EXAMPLE 8: Generation and Characterization of Transgenic Goats

The sections outlined below briefly describe the major steps in the production of transgenic goats.

Goat Species and breeds:

Swiss-origin goats, e.g., the Alpine, Saanen, and Toggenburg breeds, are preferred in the production of transgenic goats.

Goat superovulation:

The timing of estrus in the donors is synchronized on Day 0 by 6 mg subcutaneous norgestomet ear implants (Syncromate-B, CEVA Laboratories, Inc., Overland Park, KS). Prostaglandin is administered after the first seven to nine days to shut down the endogenous synthesis of progesterone. Starting on Day 13 after insertion of the implant, a total of 18 mg of follicle-stimulating hormone (FSH-Schering Corp., Kenilworth, NJ) is given intramuscularly over three days in twice-daily injections. The implant is removed on Day 14. Twenty-four hours following implant removal the donor animals are mated several times to fertile males over a two-day period (Selgrath, et al., Theriogenology, 1990. pp. 1195-1205).

Embryo collection:

Surgery for embryo collection occurs on the second day following breeding (or 72 hours following implant removal). Superovulated does are removed from food and water 36 hours prior to surgery. Does are administered 0.8 mg/kg Diazepam (VALIUM®) IV, followed immediately by 5.0 mg/kg Ketamine (Keteset), IV. Halothane (2.5%) is administered during surgery in 2 L/min oxygen via an endotracheal tube. The reproductive tract is exteriorized through a midline laparotomy incision. Corpora lutea, unruptured follicles greater than 6 mm in diameter, and ovarian cysts are counted to evaluate superovulation results and to predict the number of embryos that should be collected by oviductal flushing. A cannula is placed in the ostium of the oviduct and held in place with a

single temporary ligature of 3.0 Prolene. A 20 gauge needle is placed in the uterus approximately 0.5 cm from the uterotubal junction. Ten to twenty ml of sterile phosphate buffered saline (PBS) is flushed through the cannulated oviduct and collected in a Petri dish. This procedure is repeated on the opposite side and then the reproductive tract is replaced in the abdomen. Before closure, 10-20 ml of a sterile saline glycerol solution is poured into the abdominal cavity to prevent adhesions. The linea alba is closed with simple interrupted sutures of 2.0 Polydioxanone or Supramid and the skin closed with sterile wound clips.

Fertilized goat eggs are collected from the PBS oviductal flushings on a stereomicroscope, and are then washed in Ham's F12 medium (Sigma, St. Louis, MO) containing 10% fetal bovine serum (FBS) purchased from Sigma. In cases where the pronuclei are visible, the embryos is immediately microinjected. If pronuclei are not visible, the embryos can be placed in Ham's F12 containing 10% FBS for short term culture at 37°C in a humidified gas chamber containing 5% CO2 in air until the pronuclei become visible (Selgrath, et al., Theriogenology, 1990. pp. 1195-1205).

Microinjection procedure:

One-cell goat embryos are placed in a microdrop of medium under oil on a glass depression slide. Fertilized eggs having two visible pronuclei are immobilized on a flame-polished holding micropipet on a Zeiss upright microscope with a fixed stage using Normarski optics. A pronucleus is microinjected with the DNA construct of interest, e.g., a BC355 vector containing the fusion protein gene operably linked to the regulatory elements of the goat beta-casein gene, in injection buffer (Tris-EDTA) using a fine glass microneedle (Selgrath, et al., Theriogenology, 1990. pp. 1195-1205).

Embryo development:

After microinjection, the surviving embryos are placed in a culture of Ham's F12 containing 10% FBS and then incubated in a humidified gas chamber containing 5% CO2 in air at 37°C until the recipient animals are prepared for embryo transfer (Selgrath, et al., Theriogenology, 1990. p. 1195-1205).

Preparation of recipients:

Estrus synchronization in recipient animals is induced by 6 mg norgestomet ear implants (Syncromate-B). On Day 13 after insertion of the implant, the animals are given a single non-superovulatory injection (400 I.U.) of pregnant mares serum gonadotropin (PMSG) obtained from Sigma. Recipient females are mated to vasectomized males to ensure estrus synchrony (Selgrath, et al., Theriogenology, 1990. pp. 1195-1205).

Embryo Transfer:

All embryos from one donor female are kept together and transferred to a single recipient when possible. The surgical procedure is identical to that outlined for embryo collection outlined above, except that the oviduct is not cannulated, and the embryos are transferred in a minimal volume of Ham's F12 containing 10% FBS into the oviductal lumen via the fimbria using a glass micropipet. Animals having more than six to eight ovulation points on the ovary are deemed unsuitable as recipients. Incision closure and post-operative care are the same as for donor animals (see, e.g., Selgrath, et al., Theriogenology, 1990. pp. 1195-1205).

Monitoring of pregnancy and parturition:

Pregnancy is determined by ultrasonography 45 days after the first day of standing estrus. At Day 110 a second ultrasound exam is conducted to confirm pregnancy and assess fetal stress. At Day 130 the pregnant recipient doe is vaccinated with tetanus toxoid and Clostridium C&D. Selenium and vitamin E (Bo-Se) are given IM and Ivermectin was given SC. The does are moved to a clean stall on Day 145 and allowed to acclimatize to this environment prior to inducing labor on about Day 147. Parturition is induced at Day 147 with 40 mg of PGF2a (Lutalyse®, Upjohn Company, Kalamazoo Michigan). This injection is given IM in two doses, one 20 mg dose followed by a 20 mg dose four hours later. The doe is under periodic observation during the day and evening following the first injection of Lutalyse® on Day 147. Observations are increased to every 30 minutes beginning on the morning of the second day. Parturition occurred between 30 and 40 hours after the first

injection. Following delivery the doe is milked to collect the colostrum and passage of the placenta is confirmed.

Verification of the transgenic nature of F_0 animals:

To screen for transgenic F_0 animals, genomic DNA is isolated from two different cell lines to avoid missing any mosaic transgenics. A mosaic animal is defined as any goat that does not have at least one copy of the transgene in every cell. Therefore, an ear tissue sample (mesoderm) and blood sample are taken from a two day old F_0 animal for the isolation of genomic DNA (Lacy, et al., A Laboratory Manual, 1986, Cold Springs Harbor, NY; and Herrmann and Frischauf, Methods Enzymology, 1987. 152: pp. 180-183). The DNA samples are analyzed by the polymerase chain reaction (Gould, et al., Proc. Natl. Acad. Sci, 1989. 86:pp. 1934-1938) using primers specific for the fusion protein gene and by Southern blot analysis (Thomas, Proc Natl. Acad. Sci., 1980. 77:5201-5205) using a random primed first member or second member cDNA probe (Feinberg and Vogelstein, *Anal. Bioc.*, 1983. 132: pp. 6-13). Assay sensitivity is estimated to be the detection of one copy of the transgene in 10% of the somatic cells.

Generation and Selection of production herd

The procedures described above can be used for production of transgenic founder (F_0) goats, as well as other transgenic goats. The transgenic F_0 founder goats, for example, are bred to produce milk, if female, or to produce a transgenic female offspring if it is a male founder. This transgenic founder male, can be bred to non-transgenic females, to produce transgenic female offspring.

Transmission of transgene and pertinent characteristics

Transmission of the transgene of interest, in the goat line is analyzed in ear tissue and blood by PCR and Southern blot analysis. For example, Southern blot analysis of the founder male and the three transgenic offspring shows no rearrangement or change in the copy number between generations. The Southern blots are probed with immunoglobulin-enzyme fusion protein cDNA probe. The blots are analyzed on a Betascope 603 and copy

number determined by comparison of the transgene to the goat beta casein endogenous gene.

Evaluation of expression levels

The expression levels of the transgenic protein, in the milk of transgenic animals, is determined using enzymatic assays or Western blots.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that that prior art forms part of the common general knowledge in Australia.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

Other embodiments are within the following claims.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method of making a biologically active fusion protein having a first member, fused to a second member comprising:

choosing a first member and a second member such that the first member of the fusion protein causes said fusion protein to assemble(s) into a complex with at least one additional expressed copy of said fusion protein so that said complex optimizes the biological activity of the multimeric form of said second member in said complex;

developing said biologically active fusion protein by:

(a) introducing an expression system into a fertilized egg of a non-human mammal, wherein said expression system comprises: an in frame DNA construct encoding said fusion protein that is to be expressed in the mammary gland of an adult mammal harboring said fusion protein in its genome; a mammary tissue-specific promoter operatively linked to the DNA constructs of said fusion protein which enables the fusion protein to be expressed in the milk of a transgenic non-human placental mammal during lactation; and, a DNA sequence encoding a signal sequence to provide for secretion of said fusion protein from mammary epithelial cells;

(b) introducing said fertilized egg containing said expression system into a host non-human mammal of the same species as said fertilized egg;

(c) allowing the host non-human mammal to produce progeny; and

(d) selecting a progeny of said non-human mammal that produces recoverable amounts of said fusion protein in its milk,

producing the fusion protein; and,

allowing the fusion protein to assemble into said complex such that said complex is comprised of at least two of said fusion proteins each containing a first member and a

second member wherein the biological activity of this multimeric form of the expressed fusion protein and its second member is optimized.

2. The method of claim 1, wherein the fusion protein assembles into a form having the same number of subunits as are present in an active form of the second member.

3. The method of claim 1, wherein the first member includes an Ig subunit.

4. The method of claim 1, wherein the second member is other than an Ig subunit.

5. The method of claim 1, wherein the first member has been modified at a site which modulates formation or maintenance of a multimer of subunits.

6. The method of claim 1, wherein the first member is a dimer.

7. The method of claim 1, wherein the first member includes an Ig subunit which has been modified to inhibit formation of a multimeric form.

8. The method of claim 7, wherein the modification of the first member is a substitution, insertion, or deletion of one or more amino acid residues, that results in a subunit which does not form a multimer or which forms a lower order multimer than it normally would form.

9. The method of claim 7, wherein the hinge region of the immunoglobulin is modified.

10. The method of claim 7, wherein the modification results in a dimeric Ig structure.

11. The method of claim 10, wherein the dimer includes a heavy chain fusion and a light chain fusion.

12. The method of claim 1, wherein the second member includes beta-glucuronidase.

13. The method of claim 1, wherein the first member is an immunoglobulin (Ig) heavy or light chain, and the second member is a human beta-glucuronidase fusion protein.

14. The method of claim 1, wherein the fusion protein is produced in a non-human transgenic mammal.

15. A method for providing a transgenically produced fusion protein of claim 1, comprising obtaining milk from a transgenic mammal, which includes a fusion protein encoding transgene that result in the expression of the protein-coding sequence of fusion protein in mammary gland epithelial cells, thereby secreting the fusion protein in the milk of the mammal.

16. A nucleic acid construct, encoding a fusion protein according to claim 1, further comprising:

(a) optionally, an insulator sequence;

(b) a promoter, e.g., a mammary epithelial specific promoter, e.g., a milk protein promoter;

(c) a nucleotide sequence which encodes a signal sequence which can direct the secretion of the fusion protein, e.g. a signal sequence from a milk specific protein, or an immunoglobulin;

(d) optionally, a nucleotide sequence which encodes a sufficient portion of the amino terminal coding region of a secreted protein, e.g. a protein secreted into milk, or an immunoglobulin, to allow secretion, e.g., in the milk of a transgenic mammal, of the fusion protein;

(e) one or more nucleotide sequences which encode a fusion protein, e.g., a fusion protein as described herein: and,

(f) optionally, a 3' untranslated region from a mammary epithelial specific gene, e.g., a milk protein gene.

17. A nucleic acid construct, which includes a nucleic acid molecule encoding a fusion protein of claim 1.

18. A fusion protein described in claim 1.

19. A non-human transgenic animal which includes a transgene that encodes a fusion protein of claim 1.

20. The method of claim 14, wherein the fusion protein is produced in the milk of the non-human transgenic mammal.

21. A method of making a biologically active fusion protein having a first member, fused to a second member comprising:

choosing a first member and a second member such that the first member of the fusion protein causes said fusion protein to assemble(s) into a complex with at least one additional expressed copy of said fusion protein so that said complex optimizes the biological activity of the multimeric form of said second member in said complex;

developing said biologically active fusion protein by:

(a) introducing an expression system into a fertilized egg of a non-human mamma, wherein said expression system comprises: an in frame DNA construct encoding said fusion protein that is to be expressed in the mammary gland of an adult mammal harboring said fusion protein in its genome; a mammary tissue-specific promoter operatively linked to the DNA constructs of said fusion protein which enables the fusion protein to

be expressed in the milk of a transgenic non-human placental mammal during lactation; and, a DNA sequence encoding a signal sequence to provide for secretion of said fusion protein from mammary epithelial cells;

(b) introducing said fertilized egg containing said expression system into a host non-human mammal of the same species as said fertilized egg;

(c) allowing the host non-human mammal to produce progeny; and

(d) selecting a progeny of said non-human mammal that produces recoverable amounts of said fusion protein in its milk,

providing a nucleic acid molecule that encodes the fusion protein;

producing said fusion protein; and

allowing the fusion protein to assemble into said complex such that said complex is comprised of at least two of said fusion proteins each containing a first member and a second member wherein the biological activity of this multimeric form of the expressed fusion protein and its second member is optimized; and,

wherein a region of said first member is modified such that this modification inhibits the assembly or maintenance of the multimeric structure of said fusion protein such that said complex results in at most a tetramer of said fusion protein;

wherein said second member is a toxin.

22. The method of claim 21, wherein the fusion protein assembles into a form having the same number of subunits as are present in an active form of the second member.

23. The method of claim 21, wherein the fusion protein is produced in a transgenic non-human mammal.

24. The method of claim 23, wherein the fusion protein is produced in the milk of the transgenic non-human mammal.

25. The method of claim 21, wherein the fusion protein is produced in cell culture.

26. A method of making a biologically active fusion protein having a first member, fused to a second member comprising:

choosing a first member and a second member wherein the first member includes an Ig subunit which has been modified at a site which modulates the formation or maintenance of a multimer of Ig subunits such that the first member of the fusion protein causes said fusion protein to assemble(s) into a complex with at least one additional expressed copy of said fusion protein so that said complex optimizes the biological activity of the multimeric form of said second member in said complex:

developing said biologically active fusion protein by:

(a) introducing an expression system into a fertilized egg of a non-human mammal, wherein said expression system comprises: an in frame DNA construct encoding said fusion protein that is to be expressed in the mammary gland of an adult mammal harboring said fusion protein in its genome; a mammary tissue-specific promoter operatively linked to the DNA constructs of said fusion protein which enables the fusion protein to be expressed in the milk of a transgenic non-human placental mammal during lactation; and, a DNA sequence encoding a signal sequence to provide for secretion of said fusion protein from mammary epithelial cells;

(b) introducing said fertilized egg containing said expression system into a host non-human mammal of the same species as said fertilized egg;

(c) allowing the host non-human mammal to produce progeny; and

(d) selecting a progeny of said non-human mammal that produces recoverable amounts of said fusion protein in its milk.

providing a nucleic acid molecule that encodes the fusion protein;

producing the fusion protein; and

allowing the fusion protein to assemble into said complex such that said complex is comprised of at least two of said fusion proteins each containing a first member and a second member wherein the biological activity of this multimeric form of the expressed fusion protein and its second member is optimized;

wherein said fusion protein is used to treat or diagnosis a disease condition; and

wherein the mammary gland-specific promoter is the β-casein promoter.

27. The method of claim 21, wherein the fusion protein assembles into a form having the same number of subunits as are present in an active form of the second member.

28. The method of claim 21, wherein the first member includes an Ig subunit.

29. The method of claim 21, wherein the first member has been modified at a site which modulates formation or maintenance of a multimer of subunits.

30. The method of claim 21, wherein the first member is a dimer.

31. The method of claim 30, wherein the modification of the first member is a substitution, insertion, or deletion of one or more amino acid residues, that results in a subunit which does not form a multimer or which forms a lower order multimer than it normally would form.

32. The method of claim 30, wherein the hinge region of the immunoglobulin is deleted.

33. The method of claim 30, wherein the modification results in a dimeric Ig structure.

34. The method of claim 33, wherein said dimeric Ig structure includes a heavy chain fusion and a light chain fusion.

35. The method of claim 21, wherein the first member is an immunoglobulin (Ig) heavy or light chain, and the second member is a human beta- glucuronidase fusion protein.

36. The method of claim 28, wherein said first member is an immunoglobulin subunit selected from the group consisting of a subunit of:

a) IgG;
b) IgM,
c) IgA1;
d) IgA2;
e) IgA;
f) IgD;
g) IgE.

37. The method of claim 21, wherein said second member is capable of converting a separately administered prodrug into a physiologically active compound.

38. The method of claim 37, wherein said prodrug, when physiologically active, is capable of converting a separately administered prodrug into a physiologically active compound.

39. The method of claim 21, wherein said fusion protein is used to treat or diagnosis a disease condition.

40. The method of claim 21, wherein said disease condition is characterized by aberrant growth or activity of a cell such that a target antigen is expressed on the surface of a target cell.

41. The method of claim 21, wherein said fusion protein used for diagnosis can recognize a target antigen with either its first or the second member.

42. The method of claim 21, wherein said non-human transgenic mammal is selected from a group of mammals consisting of:

a) caprine;

b) bovine;

c) porcine;

d) rodent; and,

e) ovine.

43. The method of claim 21, wherein said fusion protein is expressed in the milk of said non-human transgenic mammal milk at a level of at least 1.0 mg/ml.

44. The method of claim 21, wherein the mammary gland-specific promoter is selected from the group consisting of:

a) a casein promoter;

b) a beta lactogiobulin promoter;

c) a whey acid protein promoter; and

d) a lactalbumin promoter.

45. The method of claim 21, wherein said in frame DNA construct encoding said fusion protein that is to be expressed in the mammary gland of an adult mammal also contains an insulator sequence.

46. The method of claim 1, wherein said second member is capable of converting a separately administered prodrug into a physiologically active compound.

47. The method of claim 46, wherein said prodrug, when physiologically active, is capable of converting a separately administered prodrug into a physiologically active compound.

48. The method of claim 1, wherein said fusion protein is used to treat or diagnosis a disease condition.

49. The method of claim 1, wherein said disease condition is characterized by aberrant growth or activity of a cell such that a target antigen is expressed on the surface of a target cell.

50. The method of claim 1, wherein said fusion protein used for diagnosis can recognize a target antigen with either its first or the second member.

51. The method of claim 1, wherein said non-human transgenic mammal is selected from a group of mammals consisting of:

a) caprine;
b) bovine;
c) porcine;
d) rodent; and,
e) ovine.

52. The method of claim 1, wherein said second member is selected from a group of mammals consisting of:

a) a toxin;
b) an enzyme;
c) a physiologically active enzyme fragment;
d) an angiogenin;
e) beta-glucuronidase;
f) a peroxidase; and,
g) a phosphatase.

53. The method of claim 1, wherein said fusion protein is expressed in the milk of said non-human transgenic mammal milk at a level of at least 1.0 mg/ml.

54. The method of claim 1, wherein the mammary gland-specific promoter is selected from the group consisting of:

a) a casein promoter;
b) a beta lactoglobulin promoter;
c) a whey acid protein promoter; and,
d) a lactalbumin promoter.

55. The method of claim 1, wherein said in frame DNA construct encoding said fusion protein that is to be expressed in the mammary gland of an adult mammal also contains an insulator sequence.

56. A method according to any one of claims 1 to 15 or 21 to 55, a nucleic acid according to any one of claims 16 to 17, a fusion protein according to claim 18, a transgenic animal according to claim 19 substantially as hereinbefore described with reference to the Figures and/or Examples.

DATED this 23rd day of February, 2005
GTC BIOTHERAPEUTICS, INC.
by DAVIES COLLISON CAVE
Patent Attorneys for the Applicant(s)

Figure 1A

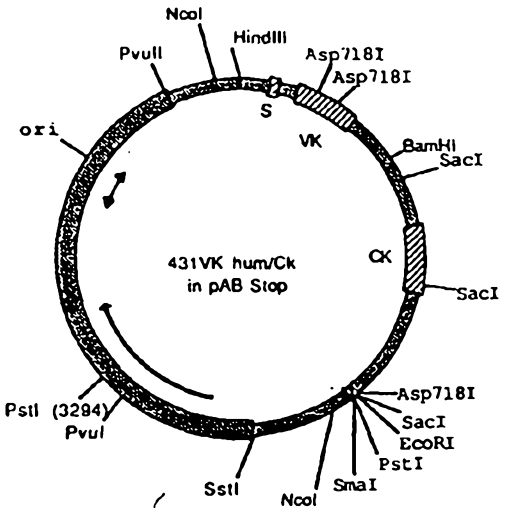


Figure 1B

```
     HindIII
  1  AAGCTTATGA ATATGCAAAT CCTGCTCATG AATATGCAAA TCCTCTGAAT CTACATGGTA AATATAGGTT

 71  TGTCTATACC ACAAACAGAA AAACATGAGA TCACAGTTCT CTCTACAGTT ACTGAGCACA CAGGACCTCA
                                Signal
     NcoI
141  CC ATG GGA TGG AGC TGT ATC ATC CTC TTC TTG GTA GCA ACA GCT ACA GGTAAGGGGC
      1▶Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr

198  TCACAGTAGC AGGCTTGAGG TCTGGACATA TATATGGGTG ACAATGACAT CCACTTTGCC TTTCTCTCCA

268  CA GAC ATC CAG ATG ACC CAG AGC CCA AGC AGC CTG AGC GCC AGC GTG GGT GAC AGA
      1▶Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser Val Gly Asp Arg
                                                                              VK
                                                                  Asp718I
324  GTG ACC ATC ACC TGT AGT ACC AGC TCG AGT GTA AGT TAC ATG CAC TGG TAC CAG CAG
 19▶Val Thr Ile Thr Cys Ser Thr Ser Ser Ser Val Ser Tyr Met His Trp Tyr Gln Gln

381  AAG CCA GGT AAG GCT CCA AAG CTG CTG ATC TAC AGC ACA TCC AAC CTG GCT TCT GGT
 38▶Lys Pro Gly Lys Ala Pro Lys Leu Leu Ile Tyr Ser Thr Ser Asn Leu Ala Ser Gly

                                                  Asp718I
438  GTG CCA AGC AGA TTC AGC GGT AGC GGT AGC GGT ACC GAC TTC ACC TTC ACC ATC AGC
 57▶Val Pro Ser Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Phe Thr Ile Ser

495  AGC CTC CAG CCA GAG GAC ATC GCC ACC TAC TAC TGC CAT CAG TGG AGT AGT TAT CCC
 76▶Ser Leu Gln Pro Glu Asp Ile Ala Thr Tyr Tyr Cys His Gln Trp Ser Ser Tyr Pro

552  ACG TTC GGC CAA GGG ACC AAG GTG GGTGAGTCCT TACAACCTCT CTCTTAGTCT CCTCAGGTGA
 95▶Thr Phe Gly Gln Gly Thr Lys Val

616  GTCCTTACAA CCTCTCTCTT CTATTCAGCT TAAATAGATT TTACTGCATT TGTTGGGGGG GAAATGTGTG

686  TATCTGAATT TCAGGTCATG AAGGACTAGG GACACCTTGG GAGTCAGAAA GGGTCATTGG GAGCCGTGGC
```

Figure 1B (continued)

```
 756 TGATGCAGAC AGACATCCTC AGCTCCCAGA CCTCATGGCC AGAGATTTAT AGGATCCTTC TAAACTCTGA

 826 GGGGGTCGGA TGACGTGGCC ATTCTTTGCC TAAAGCATTG AGTTTACTGC AAGGTCAGAA AAGCATGCAA

                                  SacI
 896 AGCCCTCAGA ATGGCTGCAA AGAGCTCCAA CAAAACAATT TAGAACTTTA TTAAGGAATA GGGGGAAGCT

 966 AGGAAGAAAC TCAAAACATC AAGATTTTAA ATACGCTTCT TGGTCTCCTT GCTATAATTA TCTGGGATAA

1036 GCATGCTGTT TTCTGTCTGT CCCTAACATG CCCTGTGATT ATCCGCAAAC AACACACCCA AGGGCAGAAC

1106 TTTGTTACTT AAACACCATC CTGTTTGCTT CTTTCCTCA GGA ACT GTG GCT GCA CCA TCT GTC
                                              1▶Gly Thr Val Ala Ala Pro Ser Val

1169 TTC ATC TTC CCG CCA TCT GAT GAG CAG TTG AAA TCT GGA ACT GCC TCT GTT GTG TGC
   9▶Phe Ile Phe Pro Pro Ser Asp Glu Gln Leu Lys Ser Gly Thr Ala Ser Val Val Cys

1226 CTG CTG AAT AAC TTC TAT CCC AGA GAG GCC AAA GTA CAG TGG AAG GTG GAT AAC GCC
  28▶Leu Leu Asn Asn Phe Tyr Pro Arg Glu Ala Lys Val Gln Trp Lys Val Asp Asn Ala

1283 CTC CAA TCG GGT AAC TCC CAG GAG AGT GTC ACA GAG CAG GAC AGC AAG GAC AGC ACC
  47▶Leu Gln Ser Gly Asn Ser Gln Glu Ser Val Thr Glu Gln Asp Ser Lys Asp Ser Thr

1340 TAC AGC CTC AGC AGC ACC CTG ACG CTG AGC AAA GCA GAC TAC GAG AAA CAC AAA GTC
  66▶Tyr Ser Leu Ser Ser Thr Leu Thr Leu Ser Lys Ala Asp Tyr Glu Lys His Lys Val

                                                 SacI
1397 TAC GCC TGC GAA GTC ACC CAT CAG GGC CTG AGC TCG CCC GTC ACA AAG AGC TTC AAC
  85▶Tyr Ala Cys Glu Val Thr His Gln Gly Leu Ser Ser Pro Val Thr Lys Ser Phe Asn

1454 AGG GGA GAG TGT TAG AGGGAGAAGT GCCCCCACCT GCTCCTCAGT TCCAGCCTGA CCCCCTCCCA
 104▶Arg Gly Glu Cys ···

1519 TCCTTTGGCC TCTGACCCTT TTTCCACAGG  NNNN<~1300~>NNNN GGACCTACCC CTATTGCGGT

1586 CCTCCAGCTC ATCTTTCACC TCACCCCCCT CCTCCTCCTT GGCTTTAATT ATGCTAATGT TGGAGGAGAA

1656 TGAATAAATA AAGTGAATCT TTGCACCTGT GGTTTCTCTC TTTCCTCAAT TTAATAATTA TTATCTGTTG

1726 TTTACCAACT ACTCAATTTC TCTTATAAGG GACTAAATAT GTAGTCATCC TAAGGCGCAT AACCATTTAT

1796 AAAAATCATC CTTCATTCTA TTTTACCCTA TCATCCTCTG CAAGACAGTC CTCCCTCAAA CCCACAAGCC

1866 TTCTGTCCTC ACAGTCCCCT GGGCCGTGGT AGGAGAGACT TGCTTCCTTG TTTTCCCCTC CTCAGCAAGC

1936 CCTATAGTCC TTTTTAAGGG TGACAGGTCT TACGGTCATA TATCCTTTGA TTCAATTCCC TGGGAATCAA

                                          SacI
                                    Asp718I      EcoRI   PstI  SmaI
2006 CCAAGGCAAA TTTTTCAAAA GAAGAAACCT GCGGGTACCG AGCTCGAATT CCTGCAGCCC GGGGGATCGA

2076 TCC
```

Sal I insert containing 431 light chain coding sequences

```
    SalI (11)
 11 GTCGACCTCGAGCCA CC ATG GGA TGG AGC TGT ATC ATC CTC TTC TTG GTA
 61 GCA ACA GCT ACA GGTAAGGGGC TCACAGTAGC AGGCTTGAGG TCTGGACATA
113 TATATGGGTG ACAATGACAT CCACTTTGCC TTTCTCTCCA CA GGTGTCCACTCC GAC
170 ATC CAG ATG ACC CAG AGC CCA AGC AGC CTG AGC GCC AGC GTG GGT GAC
218 AGA GTG ACC ATC ACC TGT AGT ACC AGC TCG AGT GTA AGT TAC ATG CAC
266 TGG TAC CAG CAG AAG CCA GGT AAG GCT CCA AAG CTG CTG ATC TAC AGC
314 ACA TCC AAC CTG GCT TCT GGT GTG CCA AGC AGA TTC AGC GGT AGC GGT
362 AGC GGT ACC GAC TTC ACC TTC ACC ATC AGC AGC CTC CAG CCA GAG GAC
410 ATC GCC ACC TAC TAC TGC CAT CAG TGG AGT AGT TAT CCC ACG TTC GGC
458 CAA GGG ACC AAG GTG G AAATCAAACGTGAGTAGAATTTAAACTTTGCTTCCTCAGTTG
516 GATCCCCAATTCTAAACTCTGAGGGGGTCGGATGACGTGGCCATTCTTTGCCTAAAGCATTGAG
580 TTTACTGCAAGGTCAGAAAAGCATGCAAAGCCCTCAGAATGGCTGCAAAGAGCTCCAACAAAAC
644 AATTTAGAACTTTATTAAGGAATAGGGGGAAGCTAGGAAGAAACTCAAAACATCAAGATTTTAA
708 ATACGCTTCTTGGTCTCCTTGCTATAATTATCTGGGATAAGCATGCTGTTTTCTGTCTGTCCCT
772 AACATGCCCTGTGATTATCCGCAAACAACACCCAAGGGCAGAACTTTGTTACTTAAACACCA
836 TCCTGTTTGCTTCTTTCCTCAGGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTG
900 ATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCTGCTGAATAACTTCTATCCCAGAGA
964 GGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCACA
1028 GAGCAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACT
1092 ACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCTCGCCCGTCACAAA
                                    SalI (1183)
1156 GAGCTTCAACAGGGGAGAGTGTTAGAGGTCGAC
```

Figure 2

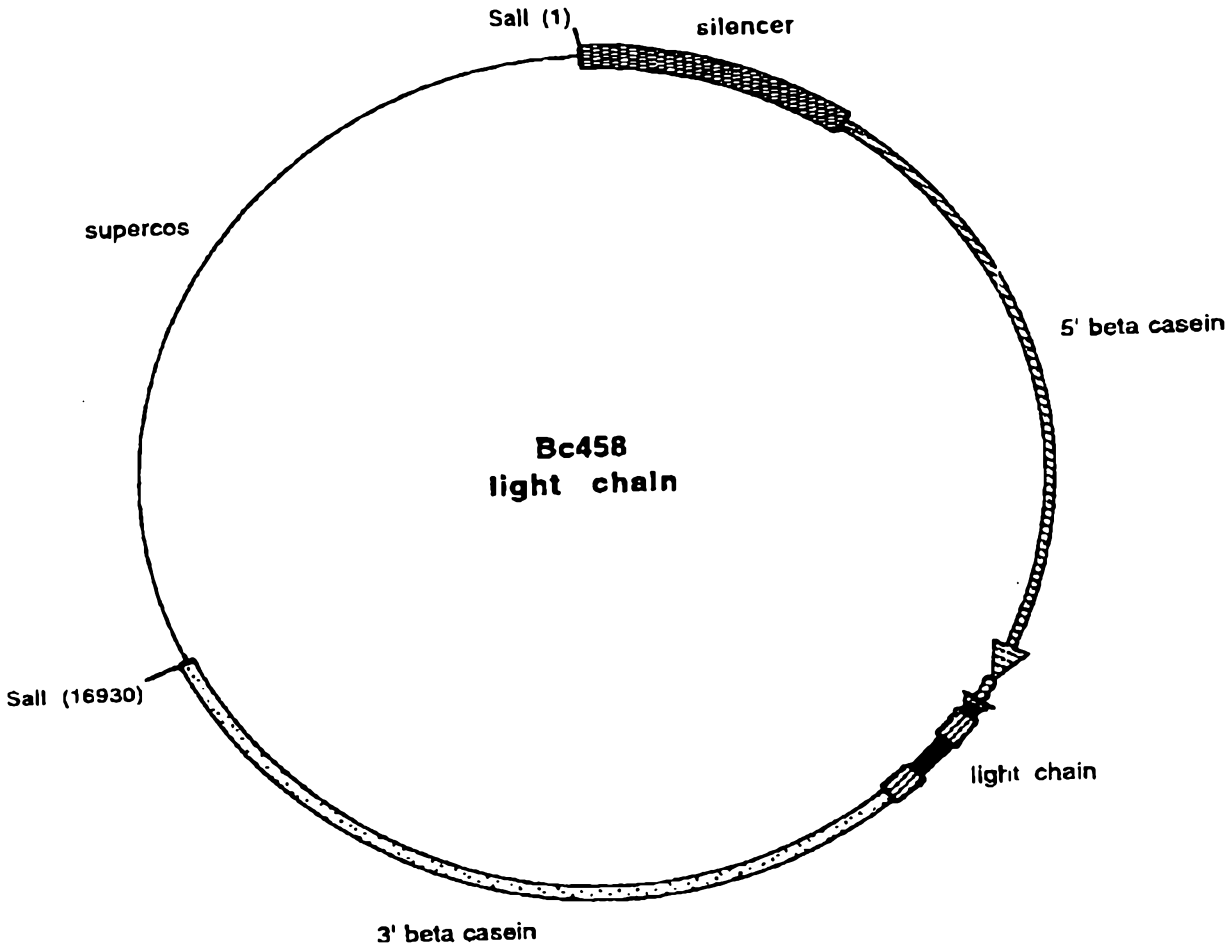

SalI (1)
silencer
supercos
5' beta casein
Bc458
light chain
SalI (16930)
light chain
3' beta casein

figure 3

FIGURE

genomic construct of 431 HC link hum-β-Gluc in pAB Stop

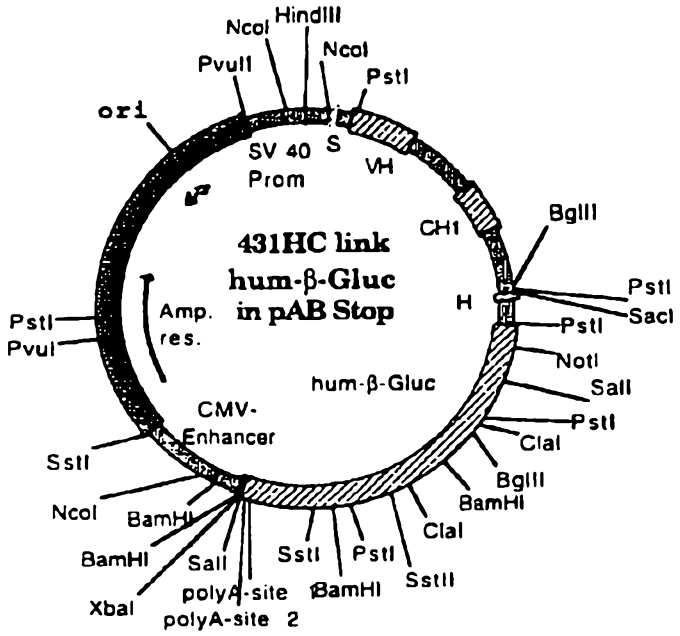
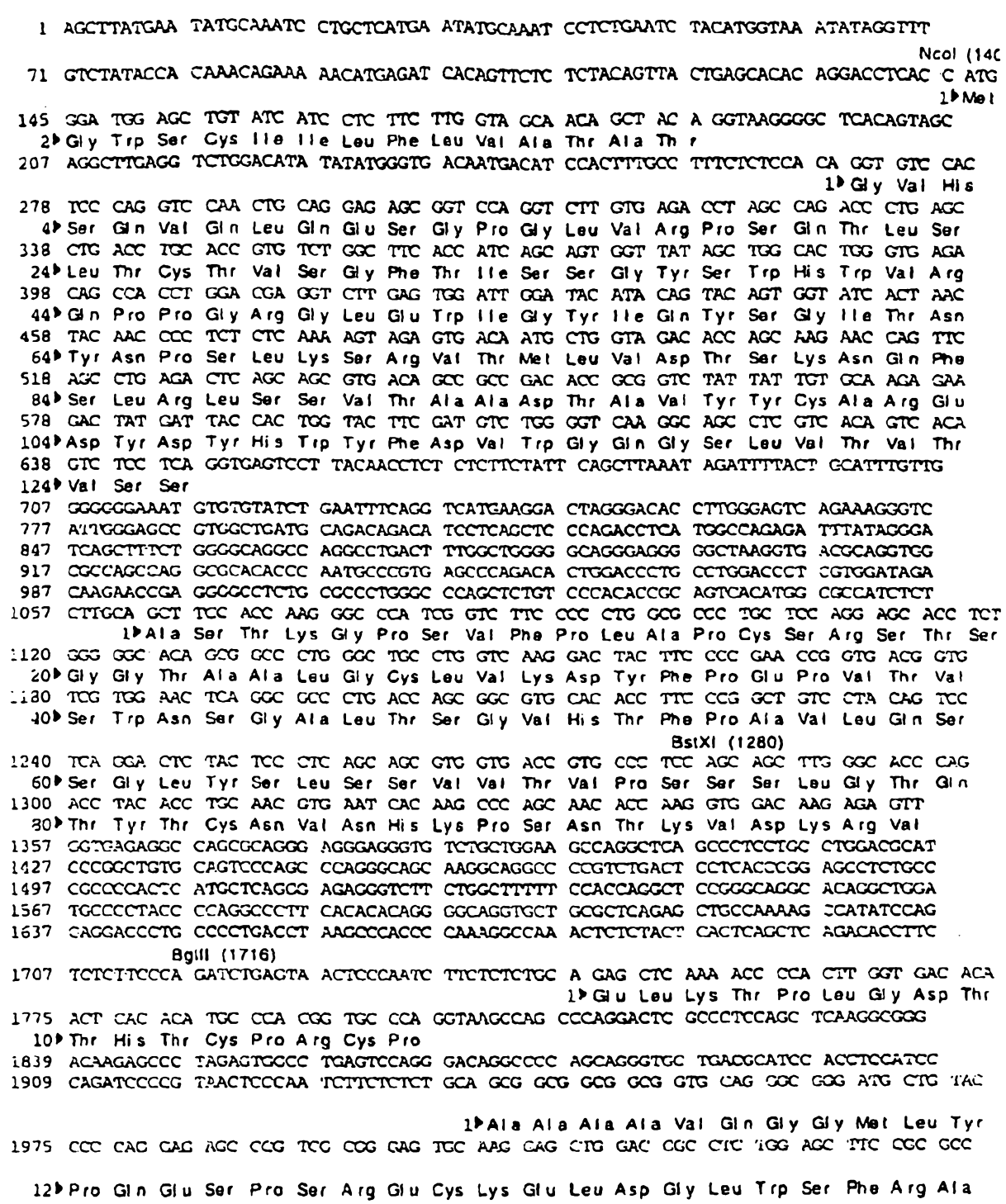


Figure 4A

FIGURE 4B




```
1 AGCTTATGAA TATGCAAATC CTGCTCATGA ATATGCAAAT CCTCTGAATC TACATGGTAA ATATAGGTTT
71 GTCTATACCA CAAACAGAAA AACATGAGAT CACAGTTCTC TCTACAGTTA CTGAGCACAC AGGACCTCAC C ATG
145 GGA TGG AGC TGT ATC ATC CTC TTC TTG GTA GCA ACA GCT AC A GGTAAGGGGC TCACAGTAGC
207 AGGCTTGAGG TCTGGACATA TATATGGGTG ACAATGACAT CCACTTTGCC TTTCTCTCCA CA GGT GTC CAC
278 TCC CAG GTC CAA CTG CAG GAG AGC GGT CCA GGT CTT GTG AGA CCT AGC CAG ACC CTG AGC
338 CTG ACC TGC ACC GTG TCT GGC TTC ACC ATC AGC AGT GGT TAT AGC TGG CAC TGG GTG AGA
398 CAG CCA CCT GGA CGA GGT CTT GAG TGG ATT GGA TAC ATA CAG TAC AGT GGT ATC ACT AAC
458 TAC AAC CCC TCT CTC AAA AGT AGA GTG ACA ATG CTG GTA GAC ACC AGC AAG AAC CAG TTC
518 AGC CTG AGA CTC AGC AGC GTG ACA GCC GCC GAC ACC GCG GTC TAT TAT TGT GCA AGA GAA
578 GAC TAT GAT TAC CAC TGG TAC TTC GAT GTC TGG GGT CAA GGC AGC CTC GTC ACA GTC ACA
638 GTC TCC TCA GGTGAGTCCT TACAACCTCT CTCTTCTATT CAGCTTAAAT AGATTTTACT GCATTTGTTG
707 GGGCGGAAAT GTGTGTATCT GAATTTCAGG TCATGAAGGA CTAGGGACAC CTTGGGAGTC AGAAAGGGTC
777 ATTGGGAGCC GTGGCTGATG CAGACAGACA TCCTCAGCTC CCAGACCTCA TGGCCAGAGA TTTATAGGGA
847 TCAGCTTTCT GGGGCAGGCC AGGCCTGACT TTGGCTGGGG GCAGGGAGGG GGCTAAGGTG ACGCAGGTGG
917 CGCCAGCCAG GCGCACACCC AATGCCCGTG AGCCCAGACA CTGGACCCTG CCTGGACCCT CGTGGATAGA
987 CAAGAACCGA GGGGCCTCTG CGCCCTGGGC CCAGCTCTGT CCCACACCGC AGTCACATGG CGCCATCTCT
1057 CTTGCA GCT TCC ACC AAG GGC CCA TCG GTC TTC CCC CTG GCG CCC TGC TCC AGG AGC ACC TCT
1120 GGG GGC ACA GCG GCC CTG GGC TGC CTG GTC AAG GAC TAC TTC CCC GAA CCG GTG ACG GTG
1180 TCG TGG AAC TCA GGC GCC CTG ACC AGC GGC GTG CAC ACC TTC CCG GCT GTC CTA CAG TCC
1240 TCA GGA CTC TAC TCC CTC AGC AGC GTG GTG ACC GTG CCC TCC AGC AGC TTG GGC ACC CAG
1300 ACC TAC ACC TGC AAC GTG AAT CAC AAG CCC AGC AAC ACC AAG GTG GAC AAG AGA GTT
1357 GGTGAGAGGC CAGCGCAGGG AGGGAGGGTG TCTGCTGGAA GCCAGGCTCA GCCCTCCTGC CTGGACGCAT
1427 CCCGGCTGTG CAGTCCCAGC CCAGGGCAGC AAGGCAGGCC CCGTCTGACT CCTCACCCGG AGCCTCTGCC
1497 CGCCCCACTC ATGCTCAGCG AGACGGTCTT CTGGCTTTTT CCACCAGGCT CCGGGCAGGC ACAGGCTGGA
1567 TGCCCCTACC CCAGGCCCTT CACACACAGG GGCAGGTGCT GCGCTCAGAG CTGCCAAAAG TCATATCCAG
1637 CAGGACCCTG CCCCTGACCT AAGCCCACCC CAAAGGCCAA ACTCTCTACT CACTCAGCTC AGACACCTTC
1707 TCTCTTCCCA GATCTGAGTA ACTCCCAATC TTCTCTCTGC A GAG CTC AAA ACC CCA CTT GGT GAC ACA
1775 ACT CAC ACA TGC CCA CGG TGC CCA GGTAAGCCAG CCCAGGACTC GCCCTCCAGC TCAAGGCGGG
1839 ACAAGAGCCC TAGAGTGGCC TGAGTCCAGG GACAGGCCCC AGCAGGGTGC TGACGCATCC ACCTCCATCC
1909 CAGATCCCCG TAACTCCCAA TCTTCTCTCT GCA GCG GCG GCG GCG GTG CAG GGC GGG ATG CTG TAC
1975 CCC CAG GAG AGC CCG TCC CGG GAG TGC AAG GAG CTG GAC GGC CTC TGG AGC TTC CGC GCC
```

FIGURE 4B (Continued)

NotI (2082)

2035 GAC TTC TCT GAC AAC CGA CGC CGG GGC TTC GAG GAG CAG TGG TAC CGG CGG CCG CTG TGG

32▶Asp Phe Ser Asp Asn Arg Arg Arg Gly Phe Glu Glu Gln Trp Tyr Arg Arg Pro Leu Trp
2095 GAG TCA GGC CCC ACC GTG GAC ATG CCA GTT CCC TCC AGC TTC AAT GAC ATC AGC CAG GAC

52▶Glu Ser Gly Pro Thr Val Asp Met Pro Val Pro Ser Ser Phe Asn Asp Ile Ser Gln Asp
2155 TGG CGT CTG CGG CAT TTT GTC GGC TGG GTG TGG TAC GAA CGG GAG GTG ATC CTG CCG GAG

72▶Trp Arg Leu Arg His Phe Val Gly Trp Val Trp Tyr Glu Arg Glu Val Ile Leu Pro Glu
2215 CGA TGG ACC CAG GAC CTG CGC ACA AGA GTG GTG CTG AGG ATT GGC AGT GCC CAT TCC TAT

92▶Arg Trp Thr Gln Asp Leu Arg Thr Arg Val Val Leu Arg Ile Gly Ser Ala His Ser Tyr
SalI (2296)
2275 GCC ATC GTG TGG GTG AAT GGG GTC GAC ACG CTA GAG CAT GAG GGG GGC TAC CTC CCC TTC

112▶Ala Ile Val Trp Val Asn Gly Val Asp Thr Leu Glu His Glu Gly Gly Tyr Leu Pro Phe
2335 GAG GCC GAC ATC AGC AAC CTG GTC CAG GTG GGG CCC CTG CCC TCC CGG CTC CGA ATC ACT

132▶Glu Ala Asp Ile Ser Asn Leu Val Gln Val Gly Pro Leu Pro Ser Arg Leu Arg Ile Thr
2395 ATC GCC ATC AAC AAC ACA CTC ACC CCC ACC ACC CTG CCA CCA GGG ACC ATC CAA TAC CTG

152▶Ile Ala Ile Asn Asn Thr Leu Thr Pro Thr Thr Leu Pro Pro Gly Thr Ile Gln Tyr Leu
2455 ACT GAC ACC TCC AAG TAT CCC AAG GGT TAC TTT GTC CAG AAC ACA TAT TTT GAC TTT TTC

172▶Thr Asp Thr Ser Lys Tyr Pro Lys Gly Tyr Phe Val Gln Asn Thr Tyr Phe Asp Phe Phe
2515 AAC TAC GCT GGA CTG CAG CGG TCT GTA CTT CTG TAC ACG ACA CCC ACC ACC TAC ATC GAT

192▶Asn Tyr Ala Gly Leu Gln Arg Ser Val Leu Leu Tyr Thr Thr Pro Thr Thr Tyr Ile Asp
BstXI (2588) BglII (2627)
2575 GAC ATC ACC GTC ACC ACC AGC GTG CAG CAA GAC AGT GGG CTG GTG AAT TAC CAG ACC TCT

212▶Asp Ile Thr Val Thr Thr Ser Val Glu Gln Asp Ser Gly Leu Val Asn Tyr Gln Ile Ser
2635 GTC AAG GGC AGT AAC CTG TTC AAG TTG GAA GTG CGT CTT TTG GAT GCA GAA AAC AAA GTC

232▶Val Lys Gly Ser Asn Leu Phe Lys Leu Glu Val Arg Leu Leu Asp Ala Glu Asn Lys Val
2695 GTG GCG AAT GGG ACT GAG ACC CAG GGC CAA CTT AAG GTG CCA GGT GTC AGC CTC TGG TGG

252▶Val Ala Asn Gly Thr Gly Thr Gln Gly Gln Leu Lys Val Pro Gly Val Ser Leu Trp Trp
2755 GCG TAC CTG ATG CAC GAA CGC CCT GCC TAT CTG TAT TCA TTG GAG GTG CAG CTG ACT GCA

272▶Pro Tyr Leu Met His Glu Arg Pro Ala Tyr Leu Tyr Ser Leu Glu Val Gln Leu Thr Ala
BamHI (2861)
2815 CAG AGC TCA CTG GGG CCT GTG TCT GAC TTC TAC ACA CTC CCT GTG GGG ATC CGC ACT GTG

292▶Gln Thr Ser Leu Gly Pro Val Ser Asp Phe Tyr Thr Leu Pro Val Gly Ile Arg Thr Val
2875 GCT GTC ACC AAG AGC CAG TTC CTC ATC AAT GGG AAA CCT TTC TAT TTC CAC GGT GTC AAC

312▶Ala Val Thr Lys Ser Gln Phe Leu Ile Asn Gly Lys Pro Phe Tyr Phe His Gly Val Asn
2935 AAG CAT GAG GAT GCG GAC ATC CGA GGG AAG GGC TTC GAC TGG CCG CTG CTG GTG AAG GAC

332▶Lys His Glu Asp Ala Asp Ile Arg Gly Lys Gly Phe Asp Trp Pro Leu Leu Val Lys Asp
2995 TTC AAC CTG CTT CGC TGG CTT GGT GCC AAC GCT TTC CGT ACC AGC CAC TAC CCC TAT GCA

352▶Phe Asn Leu Leu Arg Trp Leu Gly Ala Asn Ala Phe Arg Thr Ser His Tyr Pro Tyr Ala
3055 GAG GAA GTG ATG CAG ATG TGT GAC CGC TAT GGG ATT GTG GTC ATC GAT GAG TGT CCC GGC

372▶Glu Glu Val Met Gln Met Cys Asp Arg Tyr Gly Ile Val Val Ile Asp Glu Cys Pro Gly

FIGURE 4B (Continued)

```
3115 GTG GGC TTG GCG CTG CCG CAG TTC TTC AAC AAC GTT TCT CTG CAT CAC CAC ATG CAG GTG

 392▶Val Gly Leu Ala Leu Pro Gln Phe Phe Asn Asn Val Ser Leu His His His Met Gln Val
3175 ATG GAA GAA GTG GTG CGT AGG GAC AAG AAC CAC CCC GCG GTC GTG ATG TGG TCT GTG GCC

 412▶Met Glu Glu Val Val Arg Arg Asp Lys Asn His Pro Ala Val Val Met Trp Ser Val Ala
3235 AAC GAG CCT GCG TCC CAC CTA GAA TCT GCT GGC TAC TAC TTG AAG ATG GTG ATC GCT CAC

 432▶Asn Glu Pro Ala Ser His Leu Glu Ser Ala Gly Tyr Tyr Leu Lys Met Val Ile Ala His
      BstXI (3296)
3295 ACC AAA TCC TTG GAC CCC TCC CGG CCT GTG ACC TTT GTG AGC AAC TCT AAC TAT GCA GCA

 452▶Thr Lys Ser Leu Asp Pro Ser Arg Pro Val Thr Phe Val Ser Asn Ser Asn Tyr Ala Ala
3355 GAC AAG GGG GCT CCG TAT GTG GAT GTG ATC TGT TTG AAC AGC TAC TAC TCT TGG TAT CAC

 472▶Asp Lys Gly Ala Pro Tyr Val Asp Val Ile Cys Leu Asn Ser Tyr Tyr Ser Trp Tyr His
3415 GAC TAC GGG CAC CTG GAG TTG ATT CAG CTG CAG CTG GCC ACC CAG TTT GAG AAC TGG TAT

 492▶Asp Tyr Gly His Leu Glu Leu Ile Gln Leu Gln Leu Ala Thr Gln Phe Glu Asn Trp Tyr
3475 AAG AAG TAT CAG AAG CCC ATT ATT CAG AGC GAG TAT GGA GCA GAA ACG ATT GCA GGG TTT

 512▶Lys Lys Tyr Gln Lys Pro Ile Ile Gln Ser Glu Tyr Gly Ala Glu Thr Ile Ala Gly Phe
          BamHI (3540)
3535 CAC CAG GAT CCA CCT CTG ATG TTC ACT GAA GAG TAC CAG AAA AGT CTG CTA GAG CAG TAC

 532▶His Gln Asp Pro Pro Leu Met Phe Thr Glu Glu Tyr Gln Lys Ser Leu Leu Glu Gln Tyr
3595 CAT CTG GGT CTG GAT CAA AAA CGC AGA AAA TAT GTG GTT GGA GAG CTC ATT TGG AAT TTT

 552▶His Leu Gly Leu Asp Gln Lys Arg Arg Lys Tyr Val Val Gly Glu Leu Ile Trp Asn Phe
3655 GCC GAT TTC ATG ACT GAA CAG TCA CCG ACG AGA GTG CTG GGG AAT AAA AAG GGG ATC TTC

 572▶Ala Asp Phe Met Thr Glu Gln Ser Pro Thr Arg Val Leu Gly Asn Lys Lys Gly Ile Phe
3715 ACT CGG CAG AGA CAA CCA AAA AGT GCA GCG TTC CTT TTG CGA GAG AGA TAC TGG AAG ATT

 592▶Thr Arg Gln Arg Gln Pro Lys Ser Ala Ala Phe Leu Leu Arg Glu Arg Tyr Trp Lys Ile
3775 GCC AAT GAA ACC AGG TAT CCC CAC TCA GTA GCC AAG TCA CAA TGT TTG GAA AAC AGC CCG

 612▶Ala Asn Glu Thr Arg Tyr Pro His Ser Val Ala Lys Ser Gln Cys Leu Glu Asn Ser Pro
3835 TTT ACT TGA GCAAGACTGA TACCACCTGC GTGTCCCTTC CTCCCCGAGT CAGGGCGACT TCCAGAGCAG
                ➔
 632▶Phe Thr •••
3904 CAGAACAAGT GCCTCCTGGA CTGTTCACGG CAGACCAGAA CGTTTCTGGC CTGGGTTTTG TGGTCATCTA
3974 TTCTAGCAGG GAACACTAAA GGTGGAAATA AAAGATTTTC TATTATGGAA ATAAAGAGTT GGCATGAAAG
                          XbaI (4063)
                    SalI (4057)     BamHI (4069)
4044 TCGCTACTGN NNNGTCGACT CTAGAGGATC CCCGCTTAAT TAAGTTGTTT ATTGCAGCTT ATAATGGTTA
4114 CAAATAAAGC AATAGCATCA CAAATTTCAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTG
                                          BamHI
4184 TCCAAACTCA TCAATGTATC TTATCATGTC TGGATCCGAA TTGATCCCCT GGAGACTTGG AAATCCCCGT
                                                NcoI
4254 GAGTCAAACC GCTATCCACG CCCATTGATG TACTGCCAAA ACCGCATCAC CATGGTAATA GCGATGACTA
4324 ATACGTAGAT GTACTGCCAA GTAGGAAAGT CCCATAAGGT CATGTACTGG GCATAATGCC AGGCGGGCCA
4394 TTTACCGTCA TTGACGTCAA TAGGGGGCGT ACTTGGCATA TGATACACTT GATGTACTGC CAAGTGGGCA
4464 GTTTACCGTA AATACTCCAC CCATTGACGT CAATGGAAAG TCCCTATTGG CGTTACTATG GGAACATACG
4534 TCATTATTGA CGTCAATGGG CGGGGGTCGT TGGGCGGTCA GCCAGGCGGG CCATTTACCG TAAGTTATGT
4604 AACGCGGAAC TCCATATATG GGCTATGAAC TAATGACCCC GTAATTGATT ACTATTAATA ACTAGTCAAT
4674 AATCAATGTC CGAGCTCGAA ATTCTTGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT ATAGGTTAAT
4744 GTCATGATAA TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATTT
4814 GTTTATTTTT CTAAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCAATA
```

FIGURE 4B (Continued)

```
4884 ATATTGAAAA AGGAAGAGTA TGAGTATT CAA CAT TTC CGT GTC GCC CTT ATT CCC TTT TTT GCG GCA
                                  1▶Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
4951 TTT TGC CTT CCT GTT TTT GCT CAC CCA GAA ACG CTG GTG AAA GTA AAA GAT GCT GAA GAT
  14▶Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp
5011 CAG TTG GGT GCA CGA GTG GGT TAC ATC GAA CTG GAT CTC AAC AGC GGT AAG ATC CTT GAG
  34▶Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys Ile Leu Glu
5071 AGT TTT CGC CCC GAA GAA CGT TTT CCA ATG ATG AGC ACT TTT AAA GTT CTG CTA TGT GGC
  54▶Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly
5131 GCG GTA TTA TCC CGT GTT GAC GCC GGG CAA GAG CAA CTC GGT CGC CGC ATA CAC TAT TCT
  74▶Ala Val Leu Ser Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
                        ScaI (5207)
5191 CAG AAT GAC TTG GTT GAG TAC TCA CCA GTC ACA GAA AAG CAT CTT ACG GAT GGC ATG ACA
  94▶Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr
5251 GTA AGA GAA TTA TGC AGT GCT GCC ATA ACC ATG AGT GAT AAC ACT GCG GCC AAC TTA CTT
 114▶Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala Asn Leu Leu
5311 CTG ACA ACG ATC GGA GGA CCG AAG GAG CTA ACC GCT TTT TTG CAC AAC ATG GGG GAT CAT
 134▶Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His
5371 GTA ACT CGC CTT GAT CGT TGG GAA CCG GAG CTG AAT GAA GCC ATA CCA AAC GAC GAG CGT
 154▶Val Thr Arg Leu Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
5431 GAC ACC ACG ATG CCT GCA GCA ATG GCA ACA ACG TTG CGC AAA CTA TTA ACT GGC GAA CTA
 174▶Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu
5491 CTT ACT CTA GCT TCC CGG CAA CAA TTA ATA GAC TGG ATG GAG GCG GAT AAA GTT GCA GGA
 194▶Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys Val Ala Gly
5551 CCA CTT CTG CGC TCG GCC CTT CCG GCT GGC TGG TTT ATT GCT GAT AAA TCT GGA GCC GGT
 214▶Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly
5611 GAG CGT GGG TCT CGC GGT ATC ATT GCA GCA CTG GGG CCA GAT GGT AAG CCC TCC CGT ATC
 234▶Glu Arg Gly Ser Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
5671 GTA GTT ATC TAC ACG ACG GGG AGT CAG GCA ACT ATG GAT GAA CGA AAT AGA CAG ATC GCT
 254▶Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala
5731 GAG ATA GGT GCC TCA CTG ATT AAG CAT TGG TAA CTGTCAGACC AAGTTTACTC ATATATACTT
 274▶Glu Ile Gly Ala Ser Leu Ile Lys His Trp •••
5794 TAGATTGATT TAAAACTTCA TTTTTAATTT AAAAGGATCT AGGTGAAGAT CCTTTTTGAT AATCTCATGA
5864 CCAAAATCCC TTAACGTGAG TTTTCGTTCC ACTGAGCGTC AGACCCCGTA GAAAAGATCA AAGGATCTTC
5934 TTGAGATCCT TTTTTTCTGC GCGTAATCTG CTGCTTGCAA ACAAAAAAAC CACCGCTACC AGCGGTGGTT
6004 TGTTTGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG TAACTGGCTT CAGCAGAGCG CAGATACCAA
6074 ATACTGTCCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT CAAGAACTCT GTAGCACCGC CTACATACCT
6144 CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC GATAAGTCGT GTCTTACCGG GTTGGACTCA
6214 AGACGATAGT TACCGGATAA GGCGCAGCGG TCGGGCTGAA CGGGGGGTTC GTGCACACAG CCCAGCTTGG
6284 AGCGAACGAC CTACACCGAA CTGAGATACC TACAGCGTGA GCATTGAGAA AGCGCCACGC TTCCCGAAGG
                                                                    ←
6354 GAGAAAGGCG GACAGGTATC CGGTAAGCGG CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG
6424 GGAAACGCCT GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTTGTGAT
                                  ori
6494 GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG
5564 CTGGCCTTTT GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT
     ──────►
6634 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG
6704 AGCGCCTGAT GCGGTATTTT CTCCTTACGC ATCTGTGCGG TATTTCACAC CGCATATGGT GCACTCTCAG
6774 TACAATCTGC TCTGATGCCG CATAGTTAAG CCAGTATACA CTCCGCTATC GCTACGTGAC TGGGTCATGG
6844 CTGCGCCCCG ACACCCGCCA ACACCCGCTG ACGCGCCCTG ACGGGCTTGT CTGCTCCCGG CATCCGCTTA
6914 CAGACAAGCT GTGACCGTCT CCGGGAGCTG CATGTGTCAG AGGTTTTCAC CGTCATCACC GAAACGCGCG
6984 AGGCAGCTGT GGAATGTGTG TCAGTTAGGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC AGAAGTATGC
7054 AAAGCATGCA TCTCAATTAG TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT
7124 GCAAAGCATG CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT
                                  NcoI
7194 CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT TTATGCAGAG GCCGAGGCCG
```

FIGURE 4B (Continued)

HindIII (7328)
7264 CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC CTAGGCTTTT GCAAA

Figure 5

```
      XhoI (1)
   1  CTCGAGCCAC C ATG GGA TGG AGC TGT ATC ATC CTC TTC TTG GTA GCA ACA GCT AC A
                 1▶Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Th r
  57  GGTAAGGGGC TCACAGTAGC AGGCTTGAGG TCTGGACATA TATATGGGTG ACAATGACAT CCACTTTGCC
 127  TTTCTCTCCA CA GGT GTC CAC TCC CAG GTC CAA CTG CAG GAG AGC GGT CCA GGT CTT GTG
                   1▶Gly Val His Ser Gln Val Gln Leu Gln Glu Ser Gly Pro Gly Leu Val
 187  AGA CCT AGC CAG ACC CTG AGC CTG ACC TGC ACC GTG TCT GGC TTC ACC ATC AGC AGT
  17▶Arg Pro Ser Gln Thr Leu Ser Leu Thr Cys Thr Val Ser Gly Phe Thr Ile Ser Ser
 244  GGT TAT AGC TGG CAC TGG GTG AGA CAG CCA CCT GGA CGA GGT CTT GAG TGG ATT GGA
  36▶Gly Tyr Ser Trp His Trp Val Arg Gln Pro Pro Gly Arg Gly Leu Glu Trp Ile Gly
 301  TAC ATA CAG TAC AGT GGT ATC ACT AAC TAC AAC CCC TCT CTC AAA AGT AGA GTG ACA
  55▶Tyr Ile Gln Tyr Ser Gly Ile Thr Asn Tyr Asn Pro Ser Leu Lys Ser Arg Val Thr
 358  ATG CTG GTA GAC ACC AGC AAG AAC CAG TTC AGC CTG AGA CTC AGC AGC GTG ACA GCC
  74▶Met Leu Val Asp Thr Ser Lys Asn Gln Phe Ser Leu Arg Leu Ser Ser Val Thr Ala
 415  GCC GAC ACC GCG GTC TAT TAT TGT GCA AGA GAA GAC TAT GAT TAC CAC TGG TAC TTC
  93▶Ala Asp Thr Ala Val Tyr Tyr Cys Ala Arg Glu Asp Tyr Asp Tyr His Trp Tyr Phe
 472  GAT GTC TGG GGT CAA GGC AGC CTC GTC ACA GTC ACA GTC TCC TCA GGTGAGTCCT
 112▶Asp Val Trp Gly Gln Gly Ser Leu Val Thr Val Thr Val Ser Ser
 527  TACAACCTCT CTCTTCTATT CAGCTTAAAT AGATTTTACT GCATTTGTTG GGGGGGAAAT GTGTGTATCT
 597  GAATTTCAGG TCATGAAGGA CTAGGGACAC CTTGGGAGTC AGAAAGGGTC ATTGGGAGCC GTGGCTGATG
 667  CAGACAGACA TCCTCAGCTC CCAGACCTCA TGGCCAGAGA TTTATAGGGA TCAGCTTTCT GGGGCAGGCC
 737  AGGCCTGACT TTGGCTGGGG GCAGGGAGGG GGCTAAGGTG ACCCAGGTGG CGCCAGCCAG GCGCACACCC
 807  AATGCCCGTG AGCCCAGACA CTGGACCCTG CCTGGACCCT CGTGGATAGA CAAGAACCGA GGGGCCTCTG
 877  CGCCCTGGGC CCAGCTCTGT CCCACACCGC AGTCACATGG CGCCATCTCT CTTGCA GCT TCC ACC AAG
                                                          1▶Ala Ser Thr Lys
 945  GGC CCA TCG GTC TTC CCC CTG GCG CCC TGC TCC AGG AGC ACC TCT GGG GGC ACA GCG
   5▶Gly Pro Ser Val Phe Pro Leu Ala Pro Cys Ser Arg Ser Thr Ser Gly Gly Thr Ala
1002  GCC CTG GGC TGC CTG GTC AAG GAC TAC TTC CCC GAA CCG GTG ACG GTG TCG TGG AAC
  24▶Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr Val Ser Trp Asn
1059  TCA GGC GCC CTG ACC AGC GGC GTG CAC ACC TTC CCG GCT GTC CTA CAG TCC TCA GGA
  43▶Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro Ala Val Leu Gln Ser Ser Gly
                                  BstEII (1138)          BstXI (1150)
1116  CTC TAC TCC CTC AGC AGC GTG GTG ACC GTG CCC TCC AGC AGC TTG GGC ACC CAG ACC
  62▶Leu Tyr Ser Leu Ser Ser Val Val Thr Val Pro Ser Ser Ser Leu Gly Thr Gln Thr
1173  TAC ACC TGC AAC GTG AAT CAC AAG CCC AGC AAC ACC AAG GTG GAC AAG AGA GTT
  81▶Tyr Thr Cys Asn Val Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys Arg Val
1227  GGTGAGAGGC CAGCGCAGGG AGGGAGGGTG TCTGCTGGAA GCCAGGCTCA GCCCTCCTGC CTGGACGCAT
1297  CCCGGCTGTG CAGTCCCAGC CCAGGGCAGC AAGGCAGGCC CCGTCTGAGT CCTCACCCCGG AGCCTCTGCC
1367  CGCCCCACTC ATGCTCAGGG AGAGGGTCTT CTGGCTTTTT CCACCAGGCT CCGGGCAGGC ACAGGCTGGA
1437  TGCCCCTACC CCAGGCCCTT CACACACAGG GGCAGGTGCT GCGCTCAGAG CTGCCAAAAG CCATATCCAG
                                                                    J

1507  GAGGACCCTG CCCCTGACCT AAGCCCACCC CAAAGGCCAA ACTCTCTACT CACTCAGCTC ACGCATCCAC
1577  CTCCATCCCA GATCCCCGTA ACTCCCAATC TTCTCTCTGC A GCG GCG GCG GCG GTG CAG GGC GGG

                                                  1▶Ala Ala Ala Ala Val Gln Gly Gly
1642  ATG CTG TAC CCC CAG GAG AGC CCG TCG CGG GAG TGC AAG GAG CTG GAC GGC CTC TGG

   9▶Met Leu Tyr Pro Gln Glu Ser Pro Ser Arg Glu Cys Lys Glu Leu Asp Gly Leu Trp
1699  AGC TTC CGC GCC GAC TTC TCT GAC AAC CGA CGC CGG GGC TTC GAG GAG CAG TGG TAC

  28▶Ser Phe Arg Ala Asp Phe Ser Asp Asn Arg Arg Arg Gly Phe Glu Glu Gln Trp Tyr
      NotI (1758)
1756  CGG CGG CCG CTG TGG GAG TCA GGC CCC ACC GTG GAC ATG CCA GTT CCC TCC AGC TTC

  47▶Arg Arg Pro Leu Trp Glu Ser Gly Pro Thr Val Asp Met Pro Val Pro Ser Ser Phe
1813  AAT GAC ATC AGC CAG GAC TGG CGT CTG CGG CAT TTT GTC GGC TGG GTG TGG TAC GAA

  66▶Asn Asp Ile Ser Gln Asp Trp Arg Leu Arg His Phe Val Gly Trp Val Trp Tyr Glu
```

Figure 5 Continued

```
1870 CGG GAG GTG ATC CTG CCG GAG CGA TGG ACC CAG GAC CTG CGC ACA AGA GTG GTG CTG

  85▶Arg Glu Val Ile Leu Pro Glu Arg Trp Thr Gln Asp Leu Arg Thr Arg Val Val Leu
1927 AGG ATT GGC AGT GCC CAT TCC TAT GCC ATC GTG TGG GTG AAT GGG GTGGAT ACG CTA GA
                                                          24◀HisIle
 104▶Arg Ile Gly Ser Ala His Ser Tyr Ala Ile Val Trp Val Asn Gly ValAsp Thr Leu Gl
1987 CAT GAG GGG GGC TAC CTC CCC TTC GAG GCC GAC ATC AGC AAC CTG GTC CAG GTG GGG

 124▶His Glu Gly Gly Tyr Leu Pro Phe Glu Ala Asp Ile Ser Asn Leu Val Gln Val Gly
2044 CCC CTG CCC TCC CGG CTC CGA ATC ACT ATC GCC ATC AAC AAC ACA CTC ACC CCC ACC

 143▶Pro Leu Pro Ser Arg Leu Arg Ile Thr Ile Ala Ile Asn Asn Thr Leu Thr Pro Thr
2101 ACC CTG CCA CCA GGG ACC ATC CAA TAC CTG ACT GAC ACC TCC AAG TAT CCC AAG GGT

 162▶Thr Leu Pro Pro Gly Thr Ile Gln Tyr Leu Thr Asp Thr Ser Lys Tyr Pro Lys Gly
2158 TAC TTT GTC CAG AAC ACA TAT TTT GAC TTT TTC AAC TAC GCT GGA CTG CAG CGG TCT

 181▶Tyr Phe Val Gln Asn Thr Tyr Phe Asp Phe Phe Asn Tyr Ala Gly Leu Gln Arg Ser
                                                          BstXI (2264)
2215 GTA CTT CTG TAC ACG ACA CCC ACC ACC TAC ATC GAT GAC ATC ACC GTC ACC ACC AGC

 200▶Val Leu Leu Tyr Thr Thr Pro Thr Thr Tyr Ile Asp Asp Ile Thr Val Thr Thr Ser
                                        BglII (2303)
2272 GTG GAG CAA GAC AGT GGG CTG GTG AAT TAC CAG ATC TCT GTC AAG GGC AGT AAC CTG

 219▶Val Glu Gln Asp Ser Gly Leu Val Asn Tyr Gln Ile Ser Val Lys Gly Ser Asn Leu
2329 TTC AAG TTG GAA GTG CGT CTT TTG GAT GCA GAA AAC AAA GTC GTG GCG AAT GGG ACT

 238▶Phe Lys Leu Glu Val Arg Leu Leu Asp Ala Glu Asn Lys Val Val Ala Asn Gly Thr
2386 GGG ACC CAG GGC CAA CTT AAG GTG CCA GGT GTC AGC CTC TGG TGG CCG TAC CTG ATG

 257▶Gly Thr Gln Gly Gln Leu Lys Val Pro Gly Val Ser Leu Trp Trp Pro Tyr Leu Met
2443 CAC GAA CGC CCT GCC TAT CTG TAT TCA TTG GAG GTG CAG CTG ACT GCA CAG ACG TCA

 276▶His Glu Arg Pro Ala Tyr Leu Tyr Ser Leu Glu Val Gln Leu Thr Ala Gln Thr Ser
                                              BamHI (2537)
2500 CTG GGG CCT GTG TCT GAC TTC TAC ACA CTC CCT GTG GGG ATC CGC ACT GTG GCT GTC

 295▶Leu Gly Pro Val Ser Asp Phe Tyr Thr Leu Pro Val Gly Ile Arg Thr Val Ala Val
2557 ACC AAG AGC CAG TTC CTC ATC AAT GGG AAA CCT TTC TAT TTC CAC GGT GTC AAC AAG

 314▶Thr Lys Ser Gln Phe Leu Ile Asn Gly Lys Pro Phe Tyr Phe His Gly Val Asn Lys
2614 CAT GAG GAT GCG GAC ATC CGA GGG AAG GGC TTC GAC TGG CCG CTG CTG GTG AAG GAC

 333▶His Glu Asp Ala Asp Ile Arg Gly Lys Gly Phe Asp Trp Pro Leu Leu Val Lys Asp
2671 TTC AAC CTG CTT CGC TGG CTT GGT GCC AAC GCT TTC CGT ACC AGC CAC TAC CCC TAT

 352▶Phe Asn Leu Leu Arg Trp Leu Gly Ala Asn Ala Phe Arg Thr Ser His Tyr Pro Tyr
2728 GCA GAG GAA GTG ATG CAG ATG TGT GAC CGC TAT GGG ATT GTG GTC ATC GAT GAG TGT

 371▶Ala Glu Glu Val Met Gln Met Cys Asp Arg Tyr Gly Ile Val Val Ile Asp Glu Cys
2785 CCC GGC GTG GGC CTG GCG CTG CCG CAG TTC TTC AAC AAC GTT TCT CTG CAT CAC CAC

 390▶Pro Gly Val Gly Leu Ala Leu Pro Gln Phe Phe Asn Asn Val Ser Leu His His His
2842 ATG CAG GTG ATG GAA GAA GTG GTG CGT AGG GAC AAG AAC CAC CCC GCG GTC GTG ATG

 409▶Met Gln Val Met Glu Glu Val Val Arg Arg Asp Lys Asn His Pro Ala Val Val Met
2899 TGG TCT GTG GCC AAC GAG CCT GCG TCC CAC CTA GAA TCT GCT GGC TAC TAC TTG AAG

 428▶Trp Ser Val Ala Asn Glu Pro Ala Ser His Leu Glu Ser Ala Gly Tyr Tyr Leu Lys
```

Figure 5 Continued

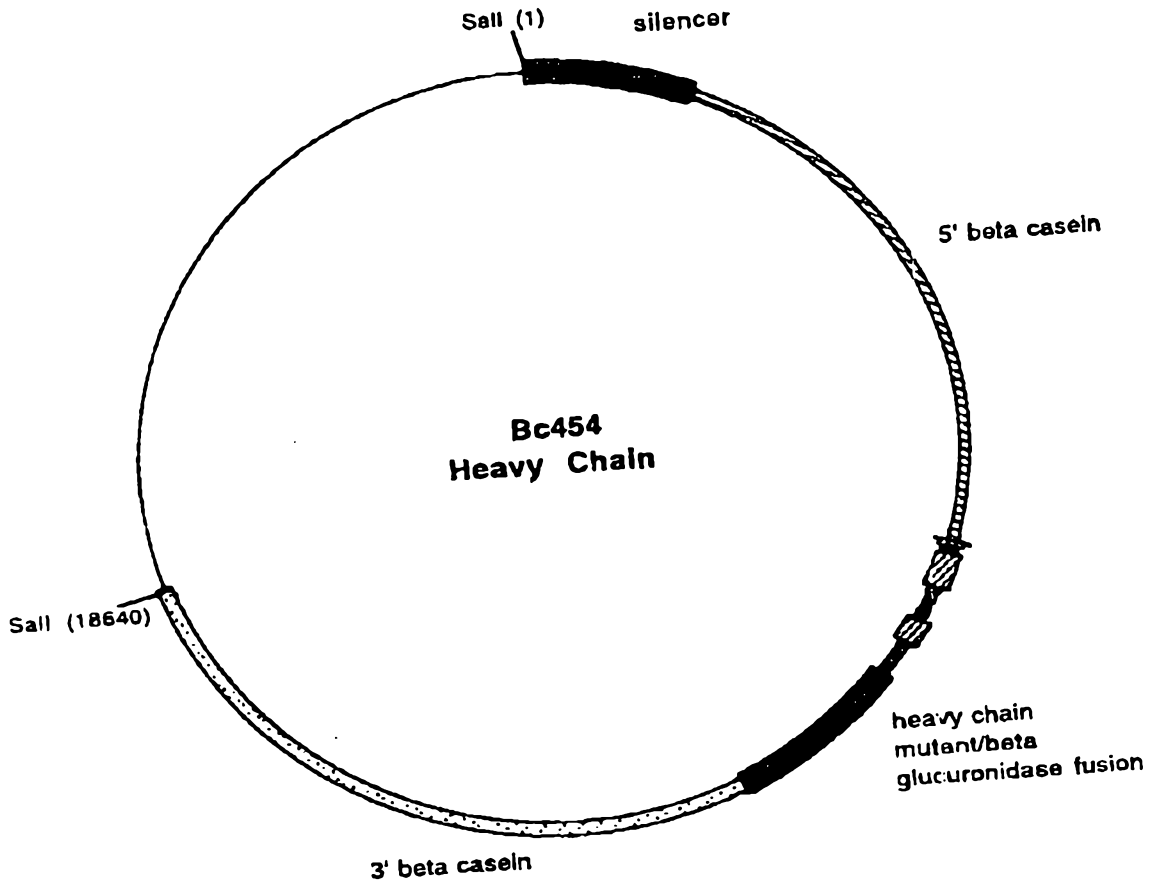
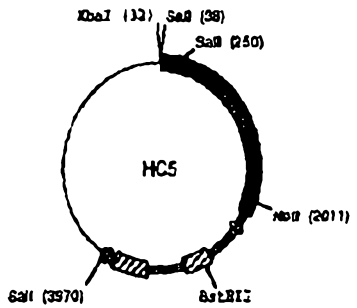

SalI (1)
silencer
5' beta casein
Bc454
Heavy Chain
SalI (18640)
heavy chain
mutant/beta
glucuronidase fusion
3' beta casein

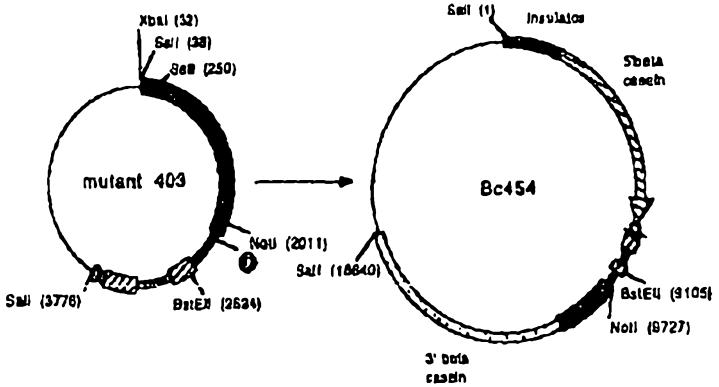
Figure 6



The construct above contains the Behring 431 heavy chain/B-glucuronidase fusion protein with a Kozak consensus sequence, flanked by Sal I sites.

The gapped heteroduplex method was used to remove (a) the hinge region and (b) the hinge and ala-ala-ala-ala-val linker, fusing the CH1 region to B-glucuronidase. Plasmid HC5 was gapped between the BstEII and Not I sites, and the following oligonucleotides used to make mutants #403 and #406, as shown.

oligo used:	ACTCAGCTCA	CGCATCCACC	
for #403		intron	intron
oligo used:	GACAAGAGAGTT	CAGGGCGGGATG	
for #406		CH1	B-glucuronidase



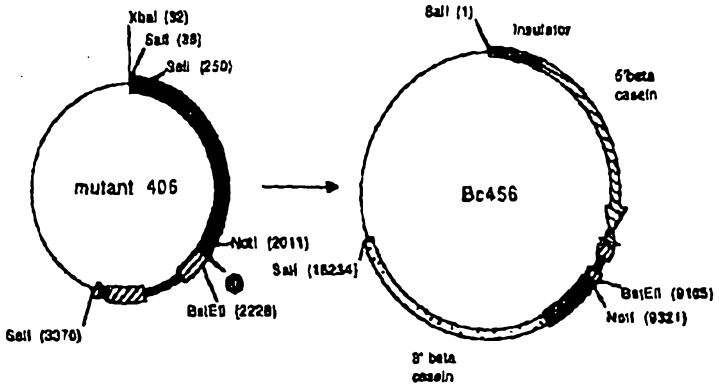
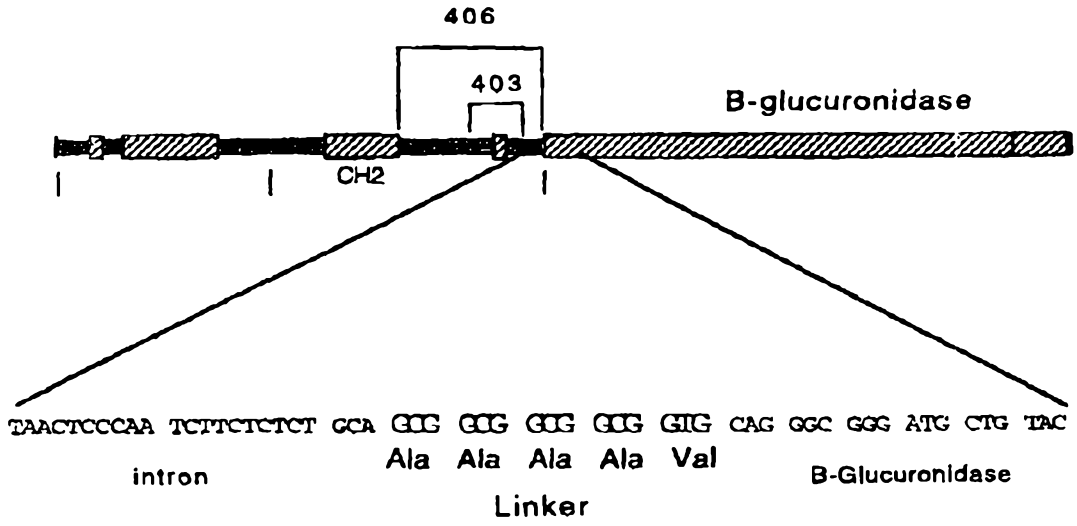


Figure 7

MUTATIONS TO B-GLUCURONIDASE



The gapped heteroduplex method was used to remove:
(a) the hinge region
(b) the hinge and ala-ala-ala-ala-val linker, fusing the CH1 region to B-glucuronidase.

oligo used: ACTCAGCTCA CGCATCCACC
for #403 intron intron

oligo used: GACAAGAGAGTT CAGGGCGGGATG
for #406 CH2 B-glucuronidase

Figure 8