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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2010 (29.12.2010)

PCT

(10) International Publication Number
WO 2010/151808 A1

(51) **International Patent Classification**:
C12N 15/10 (2006.01)

(21) **International Application Number**:
PCT/US2010/040052

(22) **International Filing Date**:
25 June 2010 (25.06.2010)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/220,878 26 June 2009 (26.06.2009) US

(71) **Applicant** *(for all designated States except US)*: **SEA LANE BIOTECHNOLOGIES, LLC** [US/US]; 356 Selby Lane, Atherton, CA 94027 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **HOROWITZ, Lawrence** [US/US]; 362 Selby Lane, Atherton, CA 94027 (US). **BHATT, Ramesh** [US/US]; 2729 Waltham Cross, Belmont, CA 94002 (US).

(74) **Agents**: **BERNHARDT, Jeffery, P.** et al.; Arnold & Porter, LLP, 1400 Page Mill Road, Palo Alto, CA 94303-1124 (US).

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

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

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

WO 2010/151808 A1

(54) **Title**: EXPRESSION OF SURROGATE LIGHT CHAINS

(57) **Abstract**: The present invention concerns surrogate light chain (SURROBODY™) constructs comprising surrogate light chain sequences with heterologous signal sequences.

EXPRESSION OF SURROGATE LIGHT CHAINS

Field of the Invention

The present invention concerns surrogate light chain (SURROBODY™) constructs comprising surrogate light chain sequences with heterologous signal sequences.

Background of the Invention

Antibody (Ig) molecules produced by B-lymphocytes are built of heavy (H) and light (L) chains. The amino acid sequences of the amino terminal domains of the H and L chains are variable (V_H and V_L), especially at the three hypervariable regions (CDR1, CDR2, CDR3) that form the antigen combining site. The assembly of the H and L chains is stabilized by a disulfide bond between the constant region of the L chain (C_L) and the first constant region of the heavy chain (C_{H1}) and by non-covalent interactions between the V_H and V_L domains.

In humans and many animals, such as mice, the genes encoding the antibody H and L chains are assembled by stepwise somatic rearrangements of gene fragments encoding parts of the V regions. Various stages of B lymphocyte development are characterized by the rearrangement status of the Ig gene loci (see, e.g. Melchers, F. & Rolink, A., B-Lymphocyte Development and Biology, Paul, W.E., ed., 1999, Lippincott, Philadephia).

Precursors of B cells (pre-B cells) have been identified in the bone marrow by their production of a set of genes called VpreB(1-3) and λ5, instead of the fully developed light chains, and coexpression of μ heavy chains.

The main isoform of human VpreB1 (CAG30495) is a 145 aa-long polypeptide (SEQ ID NO: 1). It has an Ig V domain-like structure, but lacks the last β-strand (β7) of a typical V domain, and has a carboxyl terminal end that shows no sequence homologies to any other proteins. VpreB2 has several isoforms, including a 142-amino acid mouse VpreB2 polypeptide (P13373; SEQ ID NO: 2), and a 171 amino acids long splice variant of the mouse VpreB2 sequence (CAA01964l SEQ ID NO: 3). VpreB1 and VpreB2 sequences have been disclosed in EP 0 269 127 and U.S. Patent No. 5,182,205; Collins et al., *Genome Biol.* 5(10):R84 (2004); and Hollins et al., *Proc. Natl. Acad. Sci. USA* 86(14):5552-5556 (1989). The main isoform of human VpreB3 (SEQ ID NO: 4) is a 123 aa-long protein (CAG30496), disclosed in Collins et al., *Genome Biol.* 5(10):R84 (2004).

VpreB(1-3) are non-covalently associated with another protein, λ5. The human λ5 is a 209-amino acid polypeptide (CAA01962; SEQ ID NO: 5), that carries an Ig C domain-like structure with strong homologies to antibody light chains and, towards its amino terminal end, two functionally distinct regions, one of which shows strong homology to the β7 strand of the Vλ domains. A human λ5-like protein has 213 amino acids (NP_064455; SEQ ID NO: 6) and shows about 84% sequence identity to the antibody λ light chain constant region.

For further details, see the following review papers: Karasuyama et al., *Adv. Immunol.* 63:1-41 (1996); Melchers et al., *Immunology Today* 14:60-68 (1993); and Melchers, *Proc. Natl. Acad. Sci. USA* 96:2571-2573 (1999).

The VpreB and λ5 polypeptides together form a non-covalently associated, Ig light chain-like structure, which is called the surrogate light chain or pseudo light chain. On the surface of early preB cells, the surrogate light chain is disulfide-linked to membrane-bound Ig μ heavy chain in association with a signal transducer CD79a/CD79b heterodimer to form a B cell receptor-like structure, the so-called preB cell receptor (pre-BCR).

Surrobodies are based on the pre-B cell receptor (pre-BCR), which is produced during normal development of antibody repertoire. Unlike antibodies, pre-BCR is a trimer, composed of an antibody heavy chain paired with two surrogate light chain components, VpreB and λ5. Both VpreB and λ5 are encoded by genes that do not undergo gene rearrangement and are expressed in early pre-B cells before V(D)J recombination begins. The pre-BCR is structurally different from a mature immunoglobulin in that it is composed of a heavy chain and two non-covalently associated proteins: VpreB and λ5, i.e., they have three components as opposed to two in antibodies. Furthermore, although VpreB is homologous to the Vλ Ig domain, and λ5 is homologous to the Cλ domain of antibodies, each has noncanonical peptide extensions: VpreB1 has additional 21 residues on its C terminus; λ5 has a 50 amino acid extension at its N terminus.

A κ-like B cell receptor (κ-like BCR) has been identified, utilizing a κ-like surrogate light chain (κ-like SLC) (Frances et al., *EMBO J* 13:5937-43 (1994); Thompson et al., *Immunogenetics* 48:305-11 (1998); Rangel et al., *J Biol Chem* 280:17807-14 (2005)).

Rangel et al., *J Biol Chem* 280(18):17807-17814 (2005) report the identification and molecular characterization of a Vκ-like protein that is the product of an unrearranged *V*κ gene, which turned out to the be identical to the cDNA sequence previously reported by Thompson et al., *Immunogenetics* 48:305-311 (1998). Whereas, Frances et al., *EMBO J* 13:5937-43 (1994)

reported the identification and characterization of a rearranged germline JCk that has the capacity to associate with μ heavy chains at the surface of B cell precursors, thereby providing an alternative to the λ5 pathway for B cell development.

It has been proposed that κ-like and λ-like pre-BCRs work in concert to promote light chain rearrangement and ensure the maturation of B cell progenitors. For a review, see McKeller and Martinez-Valdez *Seminars in Immunology* 18:4043 (2006).

Further details of the design and production of Surrobodies are provided in Xu et al., *Proc. Natl. Acad. Sci. USA* 2008, 105(31):10756-61, in PCT Publication WO 2008/118970 published on October 2, 2008, in U.S. Provisional Application No.61/134,929 filed July 11, 2008, and in Xu et al., *J. Mol. Biol.* 2010, 397, 352-360, the entire disclosures of which are expressly incorporated by reference herein.

Surrogate light chains have leader sequences to enable their protein production and extracellular display on pre-B cells. However, it has been found that typically the recombinant expression of engineered surrogate light chain constructs is lower than antibodies using identical heavy chains. Therefore, there is a need for improving the efficiency of recombinant expression of surrogate light chain constructs.

Summary of the Invention

The present invention is based, at least in part, on the experimental finding that the efficiency of recombinant expression of surrogate light chain constructs can be significantly improved by using heterologous leader sequences.

In one aspect, the present invention provides isolated nucleic acid molecules encoding a surrogate light chain (SLC) polypeptide or SLC construct containing an SLC polypeptide, wherein the native secretory leader sequence of the polypeptide is replaced by a heterologous secretory leader sequence. In one embodiment, the SLC polypeptide includes a VpreB polypeptide, a λ5 polypeptide, or fragments or variants thereof. In another embodiment, the VpreB polypeptide is selected from the group consisting of a native VpreB1 sequence, a native VpreB2 sequence, a native VpreB3 sequence, and fragments and variants thereof. In some embodiments, the native VpreB sequence is selected from the group consisting of human VpreB1 of SEQ ID NO: 1, mouse VpreB2 of SEQ ID NOS: 2 and 3, human VpreB3 of SEQ ID NO: 4, human VpreB-like polypeptide of SEQ ID NO:5, human VpreB dTail polypeptide of SEQ ID NO:6 and fragments and variants thereof. In one other embodiment, the λ5 polypeptide

is selected from the group consisting of a human λ5-like of SEQ ID NO: 7; a human λ5 polypeptide of SEQ ID NO: 8, a human λ5 dTail polypeptide of SEQ ID NO:9, and fragments and variants thereof. In another embodiment, the SLC polypeptide includes a Vκ -like polypeptide, a JCκ polypeptide, or fragments or variants thereof. In one other embodiment, the Vκ -like polypeptide sequence is selected from the group consisting of SEQ ID NOS: 12-24, and fragments and variants thereof. In some embodiments, the JCκ polypeptide sequence is selected from the group consisting of SEQ ID NOS:26-39, and fragments and variants thereof.

In another aspect, the present invention provides isolated nucleic acid molecules encoding a surrogate light chain (SLC) polypeptide, wherein the native secretory leader sequence of the polypeptide is replaced by a heterologous secretory leader sequence and the SLC polypeptide includes an SLC polypeptide fusion, or fragments or variants thereof. In one embodiment, the SLC fusion includes a VpreB-λ5 polypeptide fusion, or fragments or variants thereof. In another embodiment, the fusion of the VpreB polypeptide sequence and λ5 polypeptide sequence takes place at or around the CDR3 analogous regions of the VpreB sequence and the λ5 sequence respectively. In one other embodiment, the VpreB polypeptide sequence is linked at its carboxy terminus to the amino terminus of the λ5 polypeptide sequence. In one embodiment, the SLC fusion includes a Vκ-like-JCκ polypeptide fusion, or fragments or variants thereof. In another embodiment, the fusion of the Vκ-like polypeptide sequence and JCκ polypeptide sequence takes place at or around the CDR3 analogous regions of the Vκ-like sequence and the JCκ sequence respectively. In one other embodiment, the Vκ-like polypeptide sequence is fused at its carboxy terminus to the amino terminus of the JCκ polypeptide sequence.

In one other aspect, the present invention provides SLC fusions that contain a non-SLC molecule. In one embodiment, the SLC fusion contains a non-SLC molecule and at least one of a VpreB, a λ5, a Vκ-like, and a JCκ sequence. In another embodiment, the non-SLC molecule may be a non-SLC polypeptide. In one embodiment, the fusion comprises a λ5 sequence or a VpreB sequence fused to a non-SLC polypeptide. In one other embodiment, the fusion takes place at or around the CDR3 analogous regions of the VpreB sequence or the λ5 sequence. In some embodiments, the N-terminus of a λ5 sequence is fused to the C-terminus of a non-SLC polypeptide, or the C-terminus of a VpreB sequence is fused to the N-terminus of a non-SLC polypeptide. In another embodiment, the fusion comprises a Vκ-like or a JCκ sequence fused to a non-SLC polypeptide. In one other embodiment, the fusion takes place at or around the CDR3 analogous regions of the Vκ-like sequence or the JCκ sequence. In some embodiments, the N-

terminus of a JCκ sequence is fused to the C-terminus of a non-SLC polypeptide, or the C-terminus of a Vκ-like sequence is fused to the N-terminus of a non-SLC polypeptide. In one embodiment, the present invention provides isolated nucleic acid molecules encoding an SLC polypeptide, wherein the SLC polypeptide comprises an SLC fusion polypeptide containing a non-SLC molecule.

In all embodiments, the heterologous secretory leader sequence may be a leader sequence of a secreted polypeptide selected from the group consisting of antibodies, cytokines, lymphokines, monokines, chemokines, polypeptide hormones, digestive enzymes, and components of the extracellular matrix. In one embodiment, the cytokine may be selected from the group consisting of growth hormone, such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor; fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor-α and -β (TNF-α and -β); mullerian-inhibiting substance; mouse gonadotropin-associated peptide; inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF-β; platelet-growth factor; transforming growth factors (TGFs) such as TGF-α and TGF-β; insulin-like growth factor-I and -II; erythropoietin (EPO); osteoinductive factors; interferons such as interferon-α, -β and -γ; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1a, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12; a tumor necrosis factor such as TNF-α or TNF-β; MIP-1α; MIP-1β; and other polypeptide factors including LIF and kit ligand (KL).

In all embodiments, the secretory leader sequence may be selected from the group consisting of leader sequences of human and non-human mammalian albumin, transferrin, CD36, growth hormone, tissue plasminogen activator (t-PA), erythropoietin (EPO), and neublastin.

In all embodiments, the secretory leader sequence may be a synthetic sequence.

In all embodiments, the secretory leader sequence may be a consensus sequence of native secretory leader sequences.

In all embodiments, the heterologous signal sequence may be SEQ ID NO:36 (METDTLLLWVLLLWVPGSTG).

In all embodiments, the present invention provides an isolated nucleic acid molecule encoding a surrogate light chain (SLC) construct.

In one aspect, the present invention provides vectors and recombinant host cells. In all embodiments, the vectors may contain a nucleic acid molecule described herein. In all embodiments, the recombinant host cells may be transformed with a nucleic acid described herein.

In another aspect, the present invention provides methods for the expression of a surrogate light chain (SLC) polypeptide or SLC construct in a recombinant host cell. In one embodiment, the method includes the step of transforming the recombinant host cell with a nucleic acid molecule encoding an SLC polypeptide or SLC construct, wherein the native secretory leader sequence of the polypeptide is replaced by a heterologous secretory leader sequence. In another embodiment, the recombinant host cell is an eukaryotic cell. In one other embodiment, the recombinant host cell is a Chinese Hamster Ovary (CHO) cell or a human embryonic kidney (HEK) 293 cell. In some embodiments, the SLC polypeptide or SLC construct is selected from the group consisting of an SLC polypeptide comprising one or more of a VpreB polypeptide, a λ5 polypeptide, a VpreB-λ5 polypeptide fusion, a Vκ -like polypeptide, a JCκ polypeptide, and a Vκ-like-JCκ polypeptide fusion.

Brief Description of the Drawings

Figure 1 shows the human VpreB1 amino acid sequence of SEQ ID NO: 1 with a native leader sequence; the mouse VpreB2 sequences of SEQ ID NOS: 2 and 3; the human VpreB3-like sequence of SEQ ID NO: 4, the sequence of the truncated VpreB1 sequence in the "trimer" designated in Figure 11 as "VpreB dTail" (SEQ ID NO: 5); and the human VpreB1 amino acid sequence of SEQ ID NO:6 with a murine Ig κ leader sequence. Underlining indicates the leader sequences within the VpreB amino acid sequences.

Figure 2 shows the human λ5-like sequence of SEQ ID NO: 7; the human λ5 sequence of SEQ ID NO: 8; the sequence of the truncated λ5 sequence in the "trimer" designated in Figure 11 as "λ5 dTail" (SEQ ID NO: 9); and the human λ5 dTail sequence of SEQ ID NO: 10 with a murine Ig κ leader sequence. Underlining indicates the leader sequences within the λ5 amino acid sequences.

Figure 3 shows the human VpreB1-λ5 chimeric amino acid sequence as SEQ ID NO:35 (murine Ig κ leader sequence underlined).

Figures 4A and 4B show (A) the human Vκ-like nucleotide sequence of SEQ ID NO:11 and the amino acid sequence of the encoded protein (AJ004956; SEQ ID NO:12) (native leader sequence underlined), and (B) the predicted mature amino acid sequences of Vκ-like proteins possible from all Vκ families, each bearing different lengths of extensions (SEQ ID NOS: 13-24) aligned with AJ004956 Vκ-like prototype sequence (SEQ ID NO:12).

Figures 5A-C shows (A) the human JCκ nucleotide sequence of SEQ ID NO:25 and the amino acid sequence of the encoded protein (SEQ ID NO:26) (unique sequence compared to predicted mature JCk proteins is doubly underlined and potential leader cleavage sequence singly underlined), (B) the predicted JCκ-like amino acid sequences from the remaining kappa J-constant region rearrangements (J1-J5Cκ) (SEQ ID NOS:27-31), and (C) the JCk engineered secretion optimized variants, including JCκ with an appended murine Ig κ leader sequence underlined (SEQ ID NO:32), a recombined JCκ only with an appended murine Ig κ leader sequence underlined (SEQ ID NO:33), and a predicted processed JCκ with an appended murine Ig κ leader sequence underlined (SEQ ID NO:34).

Figure 6 is a schematic illustration of a surrogate light chain formed by VpreB and λ5 sequences, illustrative fusion polypeptides comprising surrogate light chain sequences, and an antibody light chain structure derived from V-J joining.

Figure 7 is a schematic illustration of various surrogate light chain deletion and single chain constructs.

Figure 8 schematically illustrates the incorporation of combinatorial functional diversity into surrogate light chain constructs.

Figure 9 shows the gene and protein structures of various illustrative surrogate light chain constructs.

Figure 10 illustrates various representative ways of adding functionality to surrogate light chain (SLC) components.

Figure 11 illustrates various trimeric and dimeric surrogate light chain (SLC) constructs.

Figure 12 is a schematic illustration of various heterodimeric surrogate κ light chain deletion variants. In the "full length" construct, both the Vκ-like and JCκ sequence retains the C- and N-terminal extensions (tails), respectively. In the dJ variant, the N-terminal extension of JCκ has been deleted. In the dVκ tail variants, the C-terminal extension of the Vκ-like sequence had been removed but the N-terminal extension of JCκ is retained. In the "short kappa" variant, both the C-terminal tail of the Vκ-like sequence and the N-terminal extension of the JCκ sequence are retained.

Figure 13: κ-like light chain deletion and single chain constructs, which can be used individually or with another protein, such as an antibody heavy chain or a fragment thereof.

Figure 14: Incorporating combinatorial functional diversity into κ-like surrogate light chain constructs. Red lines indicate appended diversity, such as a peptide library.

Figure 15: Light chains are products of gene rearrangement and RNA processing.

Figure 16A illustrates that Vκ-like protein is derived from unrearranged VκIV-gene transcription and translation. VκIV is one of seventy-one VL germline genes. Since there are an additional 70 VL germline genes capable of creating Vκ-like proteins, there are 39 more κ V genes and 31 more λ V genes.

Figure 16B illustrates that JCκ is a product of processed RNA from unrearranged J and C germlines. JCκ is one of forty-five JC germline combinations. There are an additional 44 VL germline genes capable of creating JCκ-like proteins 4 more Jκ genes to combine with Cκ and 4 Jλ genes to combine with 10 Cλ genes (40 total).

Figure 17 shows a schematic illustration of adding functionality to κ-like surrogate light chain components. Bifunctional and trifunctional structures are illustrated. A: scFv constrained fusion; B: Vκ-like scFv fusion; C: JCκ scFv fusion; D: SLC dual fusion.

Figure 18 illustrates the types of surrogate light chain functional tail extensions.

Figure 19 illustrates κ-like and λ-like surrogate light chain functional chimeras.

Figure 20A-C illustrates (A) a Surrobody format, (B) a bifunctional and bispecific Surrobody formats, and (C) cloning strategies for the molecules depicted in (A) and (B).

Detailed Description of the Invention

A. Definitions

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* 2nd ed., J. Wiley & Sons (New York, NY 1994), provides one skilled in the art with a general guide to many of the terms used in the present application.

One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Indeed, the present invention is in no way limited to the methods and materials described. For purposes of the present invention, the following terms are defined below.

Throughout this application, the use of singular includes the plural unless expressly stated otherwise.

In this application, the use of "or" includes "and/or", unless expressly stated otherwise.

Furthermore, the terms, "include," "including," and "included," are not limiting.

In the context of the present invention, the term "antibody" (Ab) is used to refer to a native antibody from a classically recombined heavy chain derived from V(D)J gene recombination and a classically recombined light chain also derived from VJ gene recombination, or a fragment thereof.

A "native antibody" is heterotetrameric glycoprotein of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains. Each light chain is linked to a heavy chain by covalent disulfide bond(s), while the number of disulfide linkages varies between the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has, at one end, a variable domain (V_H) followed by a number of constant domains. Each light chain has a variable domain at one end (V_L) and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light- and heavy-chain variable domains,

Chothia *et al., J. Mol. Biol.* 186:651 (1985); Novotny and Haber, *Proc. Natl. Acad. Sci. U.S.A.* 82:4592 (1985).

The term "variable" with reference to antibody chains is used to refer to portions of the antibody chains which differ extensively in sequence among antibodies and participate in the binding and specificity of each particular antibody for its particular antigen. Such variability is concentrated in three segments called hypervariable regions both in the light chain and the heavy chain variable domains. The more highly conserved portions of variable domains are called the framework region (FR). The variable domains of native heavy and light chains each comprise four FRs (FR1, FR2, FR3 and FR4, respectively), largely adopting a β-sheet configuration, connected by three hypervariable regions, which form loops connecting, and in some cases forming part of, the β-sheet structure. The hypervariable regions in each chain are held together in close proximity by the FRs and, with the hypervariable regions from the other chain, contribute to the formation of the antigen-binding site of antibodies (see Kabat *et al., Sequences of Proteins of Immunological Interest,* 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md. (1991), pages 647-669). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the antibody in antibody-dependent cellular toxicity.

The term "hypervariable region" when used herein refers to the amino acid residues of an antibody which are responsible for antigen-binding. The hypervariable region comprises amino acid residues from a "complementarity determining region" or "CDR" (*i.e.,* residues 30-36 (L1), 46-55 (L2) and 86-96 (L3) in the light chain variable domain and 30-35 (H1), 47-58 (H2) and 93-101 (H3) in the heavy chain variable domain; MacCallum *et al,. J Mol Biol.* 262(5):732-45 (1996).

The term “framework region” refers to the art recognized portions of an antibody variable region that exist between the more divergent CDR regions. Such framework regions are typically referred to as frameworks 1 through 4 (FR1, FR2, FR3, and FR4) and provide a scaffold for holding, in three-dimensional space, the three CDRs found in a heavy or light chain antibody variable region, such that the CDRs can form an antigen-binding surface.

Depending on the amino acid sequence of the constant domain of their heavy chains, antibodies can be assigned to different classes. There are five major classes of antibodies IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA, and IgA2. In a preferred embodiment, the immunoglobulin

sequences used in the construction of the immunoadhesins of the present invention are from an IgG immunoglobulin heavy chain domain. For human immunoadhesins, the use of human IgG1 and IgG3 immunoglobulin sequences is preferred. A major advantage of using the IgG1 is that IgG1 immunoadhesins can be purified efficiently on immobilized protein A. However, other structural and functional properties should be taken into account when choosing the Ig fusion partner for a particular immunoadhesin construction. For example, the IgG3 hinge is longer and more flexible, so that it can accommodate larger "adhesin" domains that may not fold or function properly when fused to IgG1. Another consideration may be valency; IgG immunoadhesins are bivalent homodimers, whereas Ig subtypes like IgA and IgM may give rise to dimeric or pentameric structures, respectively, of the basic Ig homodimer unit. For VEGF receptor Ig-like domain/immunoglobulin chimeras designed for in vivo applications, the pharmacokinetic properties and the effector functions specified by the Fc region are important as well. Although IgG1, IgG2 and IgG4 all have in vivo half-lives of 21 days, their relative potencies at activating the complement system are different. Moreover, various immunoglobulins possess varying numbers of allotypic isotypes.

The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called α, δ, ε, γ, and μ, respectively.

The "light chains" of antibodies from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains. Any reference to an antibody light chain herein includes both κ and λ light chains.

"Antibody fragments" comprise a portion of a full length antibody, generally the antigen binding or a variable domain thereof. Examples of antibody fragments include, but are not limited to, Fab, Fab′, $F(ab')_2$, scFv, and $(scFv)_2$ fragments.

As used herein the term “antibody binding region” refers to one or more portions of an immunoglobulin or antibody variable region capable of binding an antigen(s). Typically, the antibody binding region is, for example, an antibody light chain (VL) (or variable region thereof), an antibody heavy chain (VH) (or variable region thereof), a heavy chain Fd region, a combined antibody light and heavy chain (or variable region thereof) such as a Fab, $F(ab')_2$, single domain, or single chain antibody (scFv), or a full length antibody, for example, an IgG (*e.g.*, an IgG1, IgG2, IgG3, or IgG4 subtype), IgA1, IgA2, IgD, IgE, or IgM antibody.

The term "epitope" as used herein, refers to a sequence of at least about 3 to 5, preferably at least about 5 to 10, or at least about 5 to 15 amino acids, and typically not more than about 500, or about 1,000 amino acids, which define a sequence that by itself, or as part of a larger sequence, binds to an antibody generated in response to such sequence. An epitope is not limited to a polypeptide having a sequence identical to the portion of the parent protein from which it is derived. Indeed, viral genomes are in a state of constant change and exhibit relatively high degrees of variability between isolates. Thus the term "epitope" encompasses sequences identical to the native sequence, as well as modifications, such as deletions, substitutions and/or insertions to the native sequence. Generally, such modifications are conservative in nature but non-conservative modifications are also contemplated. The term specifically includes "mimotopes," i.e. sequences that do not identify a continuous linear native sequence or do not necessarily occur in a native protein, but functionally mimic an epitope on a native protein. The term "epitope" specifically includes linear and conformational epitopes.

The term "surrogate light chain polypeptide" or "SLC polypeptide" is used herein to refer to a VpreB polypeptide, a λ5 polypeptide, a Vκ-like polypeptide, a JCκ polypeptide, or variants thereof.

The term "non-surrogate light chain molecule" or "non-SLC molecule" is used herein to refer to a molecule that is not an SLC polypeptide. The non-SLC molecule may be a polypeptide, such as a cytokine or antibody fragment.

The term "VpreB" is used herein in the broadest sense and refers to any native sequence or variant VpreB polypeptide, specifically including, without limitation, human VpreB1 of SEQ ID NO: 1, mouse VpreB2 of SEQ ID NOS: 2 and 3, human VpreB3-like sequence of SEQ ID NO: 4, human VpreB dT of SEQ ID NO:5 and isoforms, including splice variants and variants formed by posttranslational modifications, other mammalian homologues thereof, as well as variants of such native sequence polypeptides.

The term "λ5" is used herein in the broadest sense and refers to any native sequence or variant λ5 polypeptide, specifically including, without limitation, human λ5 of SEQ ID NO: 6, human λ5-like protein of SEQ ID NO: 7, the human λ5 dT shown as SEQ ID NO: 9, the human VpreB1 amino acid sequence of SEQ ID NO:10 and their isoforms, including splice variants and variants formed by posttranslational modifications, other mammalian homologous thereof, as well a variants of such native sequence polypeptides.

The terms "variant VpreB polypeptide" and "a variant of a VpreB polypeptide" are used interchangeably, and are defined herein as a polypeptide differing from a native sequence VpreB polypeptide at one or more amino acid positions as a result of an amino acid modification. The "variant VpreB polypeptide," as defined herein, will be different from a native antibody λ or κ light chain sequence, or a fragment thereof. The "variant VpreB polypeptide" will preferably retain at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity with a native sequence VpreB polypeptide. In another preferred embodiment, the "variant VpreB polypeptide" will be less than 95%, or less than 90%, or less than 85%, or less than 80%, or less than 75%, or less than 70%, or less than 65%, or less than 60% identical in its amino acid sequence to a native antibody λ or κ light chain sequence. Variant VpreB polypeptides specifically include, without limitation, VpreB polypeptides in which the non-Ig-like unique tail at the C-terminus of the VpreB sequence is partially or completely removed.

The terms "variant λ5 polypeptide" and "a variant of a λ5 polypeptide" are used interchangeably, and are defined herein as a polypeptide differing from a native sequence λ5 polypeptide at one or more amino acid positions as a result of an amino acid modification. The "variant λ5 polypeptide," as defined herein, will be different from a native antibody λ or κ light chain sequence, or a fragment thereof. The "variant λ5 polypeptide" will preferably retain at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity with a native sequence λ5 polypeptide. In another preferred embodiment, the "variant λ5 polypeptide" will be less than 95%, or less than 90%, or less than 85%, or less than 80%, or less than 75%, or less than 70%, or less than 65%, or less than 60% identical in its amino acid sequence to a native antibody λ or κ light chain sequence. Variant λ5 polypeptides specifically include, without limitation, λ5 polypeptides in which the unique tail at the N-terminus of the λ5 sequence is partially or completely removed.

The terms "variant Vκ-like polypeptide" and "a variant of a Vκ-like polypeptide" are used interchangeably, and are defined herein as a polypeptide differing from a native sequence Vκ-like polypeptide at one or more amino acid positions as a result of an amino acid modification. The "variant Vκ-like polypeptide," as defined herein, will be different from a native antibody λ or κ light chain sequence, or a fragment thereof. The "variant Vκ-like polypeptide" will preferably retain at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%,

or at least about 98% sequence identity with a native sequence Vκ-like polypeptide. In another preferred embodiment, the "variant Vκ-like polypeptide" will be less than 95%, or less than 90%, or less than 85%, or less than 80%, or less than 75%, or less than 70%, or less than 65%, or less than 60% identical in its amino acid sequence to a native antibody λ or κ light chain sequence. Variant Vκ-like polypeptides specifically include, without limitation, Vκ-like polypeptides in which the non-Ig-like unique tail at the C-terminus of the Vκ-like sequence is partially or completely removed.

The terms "variant JCκ polypeptide" and "a variant of a JCκ polypeptide" are used interchangeably, and are defined herein as a polypeptide differing from a native sequence JCκ polypeptide at one or more amino acid positions as a result of an amino acid modification. The "variant JCκ polypeptide," as defined herein, will be different from a native antibody λ or κ light chain sequence, or a fragment thereof. The "variant JCκ polypeptide" will preferably retain at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity with a native sequence JCκ polypeptide. In another preferred embodiment, the "variant JCκ polypeptide" will be less than 95%, or less than 90%, or less than 85%, or less than 80%, or less than 75%, or less than 70%, or less than 65%, or less than 60% identical in its amino acid sequence to a native antibody λ or κ light chain sequence. Variant JCκ polypeptides specifically include, without limitation, JCκ polypeptides in which the unique tail at the N-terminus of the JCκ sequence is partially or completely removed.

Percent amino acid sequence identity may be determined using the sequence comparison program NCBI-BLAST2 (Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997)). The NCBI-BLAST2 sequence comparison program may be downloaded from http://www.ncbi.nlm.nih.gov or otherwise obtained from the National Institute of Health, Bethesda, MD. NCBI-BLAST2 uses several search parameters, wherein all of those search parameters are set to default values including, for example, unmask = yes, strand = all, expected occurrences = 10, minimum low complexity length = 15/5, multi-pass e-value = 0.01, constant for multi-pass = 25, dropoff for final gapped alignment = 25 and scoring matrix = BLOSUM62.

The term "VpreB sequence" is used herein to refer to the sequence of "VpreB," as hereinabove defined, or a fragment thereof.

The term "λ5 sequence" is used herein to refers to the sequence of "λ5," as hereinabove defined, or a fragment thereof.

The term "Vκ-like sequence" is used herein to refer to the sequence of "Vκ-like," as hereinabove defined, or a fragment thereof.

The term "JCκ sequence" is used herein to refer to the sequence of "JCκ," as hereinabove defined, or a fragment thereof.

The term "λ-like surrogate light chain," as used herein, refers to a dimer formed by the non-covalent association of a VpreB and a λ5 protein.

The term "κ-like surrogate light chain," as used herein, refers to a dimer formed by the non-covalent association of a Vκ-like and a JCκ protein.

The term "λ-like surrogate light chain sequence," as defined herein, means any polypeptide sequence that comprises a "VpreB sequence" and/or a "λ5 sequence," as hereinabove defined. The "λ-like surrogate light chain sequence," as defined herein, specifically includes, without limitation, the human VpreB1 sequence of SEQ ID NO 1, the mouse VpreB2 sequences of SEQ ID NOS: 2 and 3, and the human VpreB3 sequence of SEQ ID NO: 4, the human VpreB dT shown as SEQ ID NO: 5; and the human VpreB1 amino acid sequence of SEQ ID NO:6 and their various isoforms, including splice variants and variants formed by posttranslational modifications, homologues thereof in other mammalian species, as well as fragments and variants thereof. The term "λ-like surrogate light chain sequence" additionally includes, without limitation, the human λ5 sequence of SEQ ID NO: 7, the human λ5-like sequence of SEQ ID NO: 8, the human λ5 dTail shown as SEQ ID NO: 9, the human λ5 dTail sequence of SEQ D NO: 10 and their isoforms, including splice variants and variants formed by posttranslational modifications, homologues thereof in other mammalian species, as well as fragments and variants thereof. The term "λ-like surrogate light chain sequence" additionally includes a sequence comprising both VpreB and λ5 sequences as hereinabove defined.

The term "κ-like surrogate light chain sequence," as defined herein, means any polypeptide sequence that comprises a "Vκ-like sequence" and/or a "JCκ," as hereinabove defined. The "κ-like surrogate light chain sequence," as defined herein, specifically includes, without limitation, the human Vκ-like sequence of any of SEQ ID NOS:12-24, and their various isoforms, including splice variants and variants formed by posttranslational modifications, homologues thereof in other mammalian species, as well as fragments and variants thereof. The term "κ-like surrogate light chain sequence" additionally includes, without limitation, the human Vκ-like sequence of any of SEQ ID NOS:12-24, the human JCκ sequence of any of SEQ ID

NO:25-35, and their isoforms, including splice variants and variants formed by posttranslational modifications, homologues thereof in other mammalian species, as well as fragments and variants thereof. The term "κ-like surrogate light chain sequence" additionally includes a sequence comprising both Vκ-like and JCκ sequences as hereinabove defined.

The term, "surrogate light chain construct" is used in the broadest sense and includes any and all additional heterogeneous components, including a heterogeneous amino acid sequence, nucleic acid, and other molecules conjugated to a surrogate light chain sequence, wherein "conjugation" is defined below.

A "surrogate light chain construct" is also referred herein as a "Surrobody™," or "Surrobody" and the two terms are used interchangeably. Certain Surrobody™ λ-like surrogate light chain constructs are disclosed in Xu et al., *Proc. Natl. Acad. Sci. USA* 2008, 105(31):10756-61 and in PCT Publication WO 2008/118970 published on October 2, 2008. Also contemplated are κ-like surrogate light chain constructs as described in U.S. Patent Publication No. 2010-0062950, and Xu et al., *J. Mol. Biol.* 2010, 397, 352-360, the entire disclosures of which are expressly incorporated by reference herein.

In the context of the polypeptides of the present invention, the term "heterogeneous amino acid sequence," relative to a first amino acid sequence, is used to refer to an amino acid sequence not naturally associated with the first amino acid sequence, at least not in the form it is present in the surrogate light chain constructs herein. Thus, a "heterogeneous amino acid sequence" relative to a VpreB, λ5, Vκ-like, or JCκ is any amino acid sequence not associated with native VpreB, λ5, Vκ-like, or JCκ in its native environment. These include, without limitation, i) λ5 sequences that are different from those λ5 sequences that, together with VpreB, form the surrogate light chain on developing B cells, such as amino acid sequence variants, e.g. truncated and/or derivatized λ5 sequences; ii) VpreB sequences that are different from those VpreB sequences that, together with λ5, form the surrogate light chain on developing B cells, such as amino acid sequence variants, e.g. truncated and/or derivatized VpreB sequences, iii) Vκ-like sequences that are different from those Vκ-like sequences that, together with JCκ, form the κ-like surrogate light chain on developing B cells, such as amino acid sequence variants, e.g. truncated and/or derivatized Vκ-like sequences; and iv) JCκ sequences that are different from those JCκ sequences that, together with Vκ-like, form the κ-like surrogate light chain on developing B cells, such as amino acid sequence variants, e.g. truncated and/or derivatized JCκ sequences.

A "heterogeneous amino acid sequence" relative to a VpreB or λ5 also includes VpreB or λ5 sequences covalently associated with, e.g. fused to, a corresponding VpreB or λ5, including native sequence VpreB or λ5, since in their native environment, the VpreB and λ5 sequences are not covalently associated, e.g. fused, to each other. Similarly, a "heterogeneous amino acid sequence" relative to a Vκ-like or JCκ also includes Vκ-like or JCκ sequences covalently associated with, e.g. fused to, a corresponding Vκ-like or JCκ, including native sequence Vκ-like or JCκ, since in their native environment, the Vκ-like or JCκ sequences are not covalently associated, e.g. fused, to each other. Heterogeneous amino acid sequences also include, without limitation, antibody sequences, including antibody and heavy chain sequences and fragments thereof, such as, for example, antibody light and heavy chain variable region sequences, and antibody light and heavy chain constant region sequences.

The terms "conjugate," "conjugated," and "conjugation" refer to any and all forms of covalent or non-covalent linkage, and include, without limitation, direct genetic or chemical fusion, coupling through a linker or a cross-linking agent, and non-covalent association, for example through Van der Waals forces, or by using a leucine zipper.

The term "flexible linker" is used herein to refer to any linker that is not predicted, based on its chemical structure, to be fixed in three-dimensional space in its intended context and environment.

The term "fusion" is used herein to refer to the combination of amino acid sequences of different origin in one polypeptide chain by in-frame combination of their coding nucleotide sequences. The term "fusion" explicitly encompasses internal fusions, i.e., insertion of sequences of different origin within a polypeptide chain, in addition to fusion to one of its termini.

As used herein, the terms "peptide," "polypeptide" and "protein" all refer to a primary sequence of amino acids that are joined by covalent "peptide linkages." In general, a peptide consists of a few amino acids, typically from about 2 to about 50 amino acids, and is shorter than a protein. The term "polypeptide," as defined herein, encompasses peptides and proteins.

The term "amino acid" or "amino acid residue" typically refers to an amino acid having its art recognized definition such as an amino acid selected from the group consisting of: alanine (Ala); arginine (Arg); asparagine (Asn); aspartic acid (Asp); cysteine (Cys); glutamine (Gln); glutamic acid (Glu); glycine (Gly); histidine (His); isoleucine (Ile): leucine (Leu); lysine (Lys); methionine (Met); phenylalanine (Phe); proline (Pro); serine (Ser); threonine (Thr); tryptophan

(Trp); tyrosine (Tyr); and valine (Val) although modified, synthetic, or rare amino acids may be used as desired. Thus, modified and unusual amino acids listed in 37 CFR 1.822(b)(4) are specifically included within this definition and expressly incorporated herein by reference. Amino acids can be subdivided into various sub-groups. Thus, amino acids can be grouped as having a nonpolar side chain (*e.g.*, Ala, Cys, Ile, Leu, Met, Phe, Pro, Val); a negatively charged side chain (*e.g.*, Asp, Glu); a positively charged side chain (*e.g.*, Arg, His, Lys); or an uncharged polar side chain (*e.g.*, Asn, Cys, Gln, Gly, His, Met, Phe, Ser, Thr, Trp, and Tyr). Amino acids can also be grouped as small amino acids (Gly, Ala), nucleophilic amino acids (Ser, His, Thr, Cys), hydrophobic amino acids (Val, Leu, Ile, Met, Pro), aromatic amino acids (Phe, Tyr, Trp, Asp, Glu), amides (Asp, Glu), and basic amino acids (Lys, Arg).

The term "polynucleotide(s)" refers to nucleic acids such as DNA molecules and RNA molecules and analogs thereof (*e.g.*, DNA or RNA generated using nucleotide analogs or using nucleic acid chemistry). As desired, the polynucleotides may be made synthetically, *e.g.*, using art-recognized nucleic acid chemistry or enzymatically using, *e.g.*, a polymerase, and, if desired, be modified. Typical modifications include methylation, biotinylation, and other art-known modifications. In addition, the nucleic acid molecule can be single-stranded or double-stranded and, where desired, linked to a detectable moiety.

The term "variant" with respect to a reference polypeptide refers to a polypeptide that possesses at least one amino acid mutation or modification (i.e., alteration) as compared to a native polypeptide. Variants generated by "amino acid modifications" can be produced, for example, by substituting, deleting, inserting and/or chemically modifying at least one amino acid in the native amino acid sequence.

An "amino acid modification" refers to a change in the amino acid sequence of a predetermined amino acid sequence. Exemplary modifications include an amino acid substitution, insertion and/or deletion.

An "amino acid modification at" a specified position, refers to the substitution or deletion of the specified residue, or the insertion of at least one amino acid residue adjacent the specified residue. By insertion "adjacent" a specified residue is meant insertion within one to two residues thereof. The insertion may be N-terminal or C-terminal to the specified residue.

An "amino acid substitution" refers to the replacement of at least one existing amino acid residue in a predetermined amino acid sequence with another different "replacement" amino acid residue. The replacement residue or residues may be "naturally occurring amino acid residues"

(i.e. encoded by the genetic code) and selected from the group consisting of: alanine (Ala); arginine (Arg); asparagine (Asn); aspartic acid (Asp); cysteine (Cys); glutamine (Gln); glutamic acid (Glu); glycine (Gly); histidine (His); isoleucine (Ile): leucine (Leu); lysine (Lys); methionine (Met); phenylalanine (Phe); proline (Pro); serine (Ser); threonine (Thr); tryptophan (Trp); tyrosine (Tyr); and valine (Val). Substitution with one or more non-naturally occurring amino acid residues is also encompassed by the definition of an amino acid substitution herein.

A "non-naturally occurring amino acid residue" refers to a residue, other than those naturally occurring amino acid residues listed above, which is able to covalently bind adjacent amino acid residues(s) in a polypeptide chain. Examples of non-naturally occurring amino acid residues include norleucine, ornithine, norvaline, homoserine and other amino acid residue analogues such as those described in Ellman et al. *Meth. Enzym.* 202:301 336 (1991). To generate such non-naturally occurring amino acid residues, the procedures of Noren et al. Science 244:182 (1989) and Ellman et al., supra, can be used. Briefly, these procedures involve chemically activating a suppressor tRNA with a non-naturally occurring amino acid residue followed by in vitro transcription and translation of the RNA.

An "amino acid insertion" refers to the incorporation of at least one amino acid into a predetermined amino acid sequence. While the insertion will usually consist of the insertion of one or two amino acid residues, the present application contemplates larger "peptide insertions", e.g. insertion of about three to about five or even up to about ten amino acid residues. The inserted residue(s) may be naturally occurring or non-naturally occurring as disclosed above.

An "amino acid deletion" refers to the removal of at least one amino acid residue from a predetermined amino acid sequence.

The term “mutagenesis” refers to, unless otherwise specified, any art recognized technique for altering a polynucleotide or polypeptide sequence. Preferred types of mutagenesis include error prone PCR mutagenesis, saturation mutagenesis, or other site directed mutagenesis.

"Site-directed mutagenesis" is a technique standard in the art, and is conducted using a synthetic oligonucleotide primer complementary to a single-stranded phage DNA to be mutagenized except for limited mismatching, representing the desired mutation. Briefly, the synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the single-stranded phage DNA, and the resulting double-stranded DNA is transformed into a phage-supporting host bacterium. Cultures of the transformed bacteria are plated in top agar, permitting plaque formation from single cells that harbor the phage. Theoretically, 50% of the

new plaques will contain the phage having, as a single strand, the mutated form; 50% will have the original sequence. Plaques of interest are selected by hybridizing with kinased synthetic primer at a temperature that permits hybridization of an exact match, but at which the mismatches with the original strand are sufficient to prevent hybridization. Plaques that hybridize with the probe are then selected, sequenced and cultured, and the DNA is recovered.

The term “vector” is used to refer to a rDNA molecule capable of autonomous replication in a cell and to which a DNA segment, e.g., gene or polynucleotide, can be operatively linked so as to bring about replication of the attached segment. Vectors capable of directing the expression of genes encoding for one or more polypeptides are referred to herein as "expression vectors. "The term “control sequences” refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers. A vector may be a “plasmid” referring to a circular double-stranded DNA loop into which additional DNA segments may be ligated. A vector may be a phage vector or a viral vector, in which additional DNA segements may be ligated into the viral genome. Suitable vectors are capable of autonomous replication in a host cell into which they are introduced, e.g., bacterial vector with a bacterial origin or replication and episomal mammalian vectors. A vector may be integrated into the host cell genome, e.g., a non-episomal mammalian vector, upon introduction into the host cell, and replicated along with the host genome.

Nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, “operably linked” means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

A "phage display library" is a protein expression library that expresses a collection of cloned protein sequences as fusions with a phage coat protein. Thus, the phrase "phage display

library" refers herein to a collection of phage (e.g., filamentous phage) wherein the phage express an external (typically heterologous) protein. The external protein is free to interact with (bind to) other moieties with which the phage are contacted. Each phage displaying an external protein is a "member" of the phage display library.

The term "filamentous phage" refers to a viral particle capable of displaying a heterogeneous polypeptide on its surface, and includes, without limitation, f1, fd, Pf1, and M13. The filamentous phage may contain a selectable marker such as tetracycline (e.g., "fd-tet"). Various filamentous phage display systems are well known to those of skill in the art (see, e.g., Zacher et al. Gene 9: 127-140 (1980), Smith et al. Science 228: 1315-1317 (1985); and Parmley and Smith Gene 73: 305-318 (1988)).

The term “panning” is used to refer to the multiple rounds of screening process in identification and isolation of phages carrying compounds, such as antibodies, with high affinity and specificity to a target.

A "leader sequence," “signal peptide,” or a "secretory leader," which terms are used interchangeably, contains a sequence comprising amino acid residues that directs the intracellular trafficking of the polypeptide to which it is a part. Polypeptides contain secretory leaders, signal peptides or leader sequences, typically at their N-terminus. These polypeptides may also contain cleavage sites where the leader sequences may be cleaved from the rest of the polypeptides by signal endopeptidases. Such cleavage results in the generation of mature polypeptides. Cleavage typically takes place during secretion or after the intact polypeptide has been directed to the appropriate cellular compartment.

A "host cell" includes an individual cell or cell culture which can be or has been a recipient for transformation of nucleic acid(s) and/or vector(s) containing nucleic acids encoding the molecules described herein. In methods of the present invention, a host cell can be a eukaryotic cell, such as a Chinese Hamster Ovary (CHO) cell, or a human embryonic kidney (HEK) 293 cell. Other suitable host cells are known to those skilled in the art.

B. Detailed Description

Techniques for performing the methods of the present invention are well known in the art and described in standard laboratory textbooks, including, for example, Ausubel *et al.*, Current Protocols of Molecular Biology, John Wiley and Sons (1997); Molecular Cloning: A Laboratory Manual, Third Edition, J. Sambrook and D. W. Russell, eds., Cold Spring Harbor, New York,

USA, Cold Spring Harbor Laboratory Press, 2001; O'Brian et al., Analytical Chemistry of Bacillus Thuringiensis, Hickle and Fitch, eds., Am. Chem. Soc., 1990; Bacillus thuringiensis: biology, ecology and safety, T.R. Glare and M. O'Callaghan, eds., John Wiley, 2000; Antibody Phage Display, Methods and Protocols, Humana Press, 2001; and Antibodies, G. Subramanian, ed., Kluwer Academic, 2004. Mutagenesis can, for example, be performed using site-directed mutagenesis (Kunkel et al., Proc. Natl. Acad. Sci USA 82:488-492 (1985)). PCR amplification methods are described in U.S. Pat. Nos. 4,683,192, 4,683,202, 4,800,159, and 4,965,188, and in several textbooks including "PCR Technology: Principles and Applications for DNA Amplification", H. Erlich, ed., Stockton Press, New York (1989); and PCR Protocols: A Guide to Methods and Applications, Innis et al., eds., Academic Press, San Diego, Calif. (1990).

Heterologous Leader Sequences

The main isoform of human VpreB1 (CAG30495) is a 145 amino acid long polypeptide (SEQ ID NO: 1 in Figure 1), including a 19 amino acid leader sequence. Similar leader sequences are present in other VpreB polypeptides. The human truncated VpreB1 sequence (lacking the characteristic "tail" at the C-terminus of native VpreB1), is also referred to as the "VpreB1 dTail sequence" and shown as SEQ ID NO:5.

The main isoform of human λ5 (CAA10962) is a 209-amino acid polypeptide (SEQ ID NO:7), including a 30 amino acid leader sequence. Similar leader sequences are present in other λ5 polypeptides. The human truncated λ5 sequence (lacking the characteristic "tail" at the N-terminus of native λ5), is also referred to as the "λ5 dTail sequence" and shown as SEQ ID NO:9.

Native human Vκ-like polypeptide sequences specifically include, without limitation, the human κ-like polypeptide (SEQ ID NO:12), encoded by the polynucleotide of AJ004956 shown as SEQ ID NO:11, including a 20 amino acid leader sequence. Similar leader sequences are present in other Vκ-like polypeptides.

Native sequence JCκ-like polypeptides include, without limitation, the AAB32987 human JCκ polyepeptide shown in Figure 5A that lacks a prototypical leader sequence (SEQ ID NO: 26), including a potential 22 amino acid leader sequence, of which 15 amino acids are uniquely appended to classically recombined JCκ sequence. Similar recombined leader sequences are present in other JCκ polypeptides.

The present invention provides nucleic acid and polypeptide constructs for producing surrogate light chain constructs in higher yields than when such constructs are produced from sequences that comprise an endogenous leader VpreB leader sequence and/or λ5 leader sequence, or an endogenous Vκ-like leader sequence and/or JCκ leader sequence. The present invention also provides vectors, host cells and methods for producing surrogate light chain constructs in higher yields than when such constructs are produced from DNA sequences that include the coding sequence of the endogenous leader of VpreB and/or λ5, or the endogenous leader of Vκ-like and/or JCκ, or without an endogenous leader sequence. The higher yields are achieved by replacing at least one endogenous secretory leader sequence with a heterologous leader sequence of the invention. Accordingly, the present invention provides surrogate light chains and surrogate light chain constructs comprising heterologous leader sequences.

Preferably, the expression level achieved by a heterologous leader peptide is at least about 5% higher, at least about 10% higher, at least about 20% higher, at least about 30% higher, at least about 40% higher, or at least about 50% higher than the expression level achieved by using a homologous leader sequence, when expression is conducted under essentially the same conditions.

In the present invention, a heterologous leader sequence is fused to the amino terminus of a surrogate light chain polypeptide, in place of the native VpreB leader sequence and/or the native λ5 leader sequence, or a κ-like surrogate light chain polypeptide, in place of the native Vκ-like leader sequence and/or the native JCκ leader sequence. The inventors have discovered that certain heterologous leader sequences function surprisingly well, in contrast to the native leader sequence of the surrogate light chain during the production of surrogate light chain constructs, comprising a surrogate light chain sequence (VpreB/λ5 or Vκ-like/JCκ sequences either fused together or non-covalently associated) and an antibody heavy chain sequence.

According to the present invention, the heterologous leader sequence can be any leader sequence from a highly translated protein, including leader sequences of antibody light chains and human and non-human mammalian secreted proteins. Secreted proteins are included and their sequences are available from public databases, such as Swiss-Prot, UniProt, TrEMBL, RefSeq, Ensembl and CBI-Gene. In addition, SPD, a web based secreted protein database is a resource for such sequences, available at http://spd.cbi.pku.edu.cn. (See, Chen et al., *Nucleic Acids Res.*, 2005, 33:D169-D173). Such secreted proteins include, without limitation, antibodies, cytokines, lymphokines, monokines, chemokines, polypeptide hormones, digestive enzymes, and components of the extracellular matrix.

Included among the cytokines are growth hormone, such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor; fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor-α and -β (TNF-α and -β); mullerian-inhibiting substance; mouse gonadotropin-associated peptide; inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF-β; platelet-growth factor; transforming growth factors (TGFs) such as TGF-α and TGF-β; insulin-like growth factor-I and -II; erythropoietin (EPO); osteoinductive factors; interferons such as interferon-α, -β and -γ; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1a, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12; a tumor necrosis factor such as TNF-α or TNF-β; MIP-1α; MIP-1β; and other polypeptide factors including LIF and kit ligand (KL).

Further leader sequences suitable for use in the constructs of the present invention are included in publicly available signal peptide databases, such as, the SPdb signal peptide database, accessible at http://proline.bis.nus.edu.sg/spdb (See, Choo et al., *BMC Bioinformatics* 2005, 6:249).

Specific examples of suitable heterologous leader sequences include, without limitation, leader sequences of human and non-human mammalian albumin, transferrin, CD36, growth hormone, tissue plasminogen activator (t-PA), erythropoietin (EPO), neublastin leader sequences and leader peptides from other secreted human and non-human proteins.

The murine Ig kappa leader sequence may be used (METDTLLLWVLLLWVPGSTG - SEQ ID NO:36) as a heterologous leader sequence.

When heterologous leader sequences are present in i) both a VpreB and a λ5 surrogate light chain construct, or ii) both a Vκ-like and a JCκ surrogate light chain construct, each heterologous leader sequence in i) or ii) may be identical to the other or may be different from the other.

In addition to signal peptides from native proteins, the heterologous leader sequences of the present invention include synthetic and consensus leader sequences, which can be designed to further improve the performance of leader sequences occurring in nature, and specifically

adapted for best performance in the host organism used for the expression of the surrogate light chain constructs of the present invention.

Surrogate light chain constructs

The surrogate light chain (SLC) constructs herein are based on the pre-B cell receptor (pre-BCR), which is produced during normal development of an antibody repertoire. Unlike antibodies, pre-BCR is a trimer, that is composed of an antibody heavy chain paired with two surrogate light chain components, VpreB and λ5. Both VpreB and λ5 are encoded by genes that do not undergo gene rearrangement and are expressed in early pro-B cells before V(D)J recombination begins. The pre-BCR is structurally different from a mature immunoglobulin in that it is composed of a heavy chain and two non-covalently associated proteins: VpreB and λ5, i.e., they have three components as opposed to two in antibodies. Furthermore, although VpreB is homologous to the Vλ Ig domain, and λ5 is homologous to the Cλ domain of antibodies, each has noncanonical peptide extensions: VpreB1 has additional 21 residues on its C terminus; λ5 has a 50 amino acid extension at its N terminus. Further details of the design and production of Surrobodies are provided in Xu et al., *Proc. Natl. Acad. Sci. USA* 2008, 105(31):10756-61 and in PCT Publication WO 2008/118970 published on October 2, 2008.

Similarly, the κ-like surrogate light chain constructs described herein are based on the pre-B cell receptor (pre-BCR). The κ-like light chain is the germline VκIV gene partnered with a JCκ fusion gene. In each of these genes a peptidic extension exists in the vicinity surrounding a site analogous for CDR3. As these two proteins do not appear to recombine at the genomic level it is likely their association to a heavy chain are mutually exclusive of each other and analogous to the associations described for the λ-like surrogate light chain. Further details of the design and production of κ-like surrogate light chain constructs can be found in U.S. Patent Publication No. 2010-0062950, and Xu et al., *J. Mol. Biol.* 2010, 397, 352-360, the entire disclosures of which are expressly incorporated herein by reference.

The present invention contemplates surrogate light chain (SLC) polypeptides and SLC constructs containing an SLC polypeptide having surrogate light chain sequences with heterologous signal sequences. In one embodiment, the SLC construct may comprise a VpreB sequence conjugated to a λ5 sequence, wherein the native secretory leader sequence of said VpreB sequence and/or said λ5 sequence is replaced by a heterologous secretory leader sequence. In another embodiment, the VpreB sequence is selected from the group consisting of a native VpreB1 sequence, a native VpreB2 sequence, a native VpreB3 sequence and fragments

and variants thereof. In one other embodiment, the native VpreB sequence is selected from the group consisting of human VpreB1 of SEQ ID NO: 1, mouse VpreB2 of SEQ ID NOS: 2 and 3, human VpreB3 of SEQ ID NO: 4, human VpreB-like polypeptide of SEQ ID NO:5, human VpreB dTail polypeptide of SEQ ID NO:6 and fragments and variants thereof. In other embodiments, the λ5 sequence comprises all or part of a human λ5-like of SEQ ID NO: 7; a human λ5 polypeptide of SEQ ID NO: 8, or a human λ5 dTail polypeptide of SEQ ID NO:9.

The present invention also contemplates SLC constructs wherein a λ5 sequence and a VpreB sequence are connected by a covalent linker. In one embodiment, the invention provides an SLC construct wherein the λ5 sequence is non-covalently associated with the VpreB sequence. In one other embodiment, the invention contemplates an SLC construct wherein the conjugate of said VpreB sequence and λ5 sequence is non-covalently associated with an antibody heavy chain sequence.

As described herein, the present invention concerns isolated nucleic acid molecules encoding SLC polypeptides and the SLC constructs comprising the SLC polypeptides. In one embodiment, the invention provides a nucleic acid encoding a surrogate light chain comprising a VpreB sequence fused to a λ5 sequence, wherein the native secretory leader sequence of said VpreB sequence and/or said λ5 sequence is replaced by a heterologous secretory leader sequence. In another embodiment, the invention provides a nucleic acid encoding a surrogate light chain comprising a VpreB sequence connected to a λ5 sequence by a peptide or polypeptide linker, wherein the native secretory leader sequence of said VpreB sequence and/or said λ5 sequence is replaced by a heterologous secretory leader sequence. In one other embodiment, the invention provides a vector comprising the nucleic acid. In another embodiment, the invention provides a recombinant host cell transformed with the nucleic acid.

In another aspect, the invention provides a library of surrogate light chain constructs. In another embodiment, the library comprises a nucleic acid encoding an SLC. In one other embodiment, the library may be in the form of a display.

In one other aspect, the present invention contemplates κ-like surrogate light chain polypeptides and SLC constructs comprising κ-like SLC polypeptides. In one embodiment, the invention relates to a κ-like SLC construct comprising a Vκ -like sequence conjugated to JCκ sequence, wherein the native secretory leader sequence of said Vκ -like sequence and/or said JCκ sequence is replaced by a heterologous secretory leader sequence. In another embodiment, the Vκ -like sequence is selected from the group consisting of SEQ ID NOS: 12-24, and

fragments and variants thereof. In one other embodiment, the JCκ sequence is selected from the group consisting of SEQ ID NOS:26-39, and fragments and variants thereof.

In one embodiment, the invention contemplates a κ-like SLC construct wherein the Vκ-like sequence is fused to said JCκ sequence. In another embodiment, the fusion takes place at or around the CDR3 analogous regions of said Vκ-like sequence and said JCκ sequence respectively. In one embodiment, the invention contemplates a κ-like SLC construct, wherein said Vκ-like sequence and said JCκ sequence are connected by a covalent linker.

In one embodiment, the invention provides a κ-like SLC construct, wherein said Vκ-like sequence is non-covalently associated with said JCκ sequence. In one embodiment, the invention provides a κ-like SLC construct wherein the conjugate of said Vκ-like sequence and JCκ sequence is non-covalently associated with an antibody heavy chain sequence.

In one embodiment, the invention provides a κ-like SLC construct, wherein said secretory leader sequence may be a synthetic sequence. In one embodiment, the invention provides a κ-like SLC construct, wherein said secretory leader sequence may be a consensus sequence of native secretory leader sequences.

In another aspect, the invention provides isolated nucleic acids encoding a κ-like SLC construct. In one embodiment, the invention provides a nucleic acid encoding a κ-like surrogate light chain comprising a Vκ-like sequence fused to a JCκ sequence, wherein the native secretory leader sequence of said Vκ-like sequence and/or said JCκ sequence is replaced by a heterologous secretory leader sequence. In another embodiment, the invention provides a nucleic acid encoding a κ-like surrogate light chain comprising a Vκ-like sequence connected to a JCκ sequence by a peptide or polypeptide linker, wherein the native secretory leader sequence of said Vκ-like sequence and/or said JCκ sequence is replaced by a heterologous secretory leader sequence. In one other embodiment, the invention provides a vector comprising the nucleic acid. In another embodiment, the invention provides a recombinant host cell transformed with the nucleic acid.

In one embodiment, the invention provides a library a κ-like surrogate light chain construct. In another embodiment, the library comprises a nucleic acid encoding a κ-like SLC. In one other embodiment, the library may be in the form of a display.

In one other aspect, the invention provides a method for the expression of a κ-like SLC. In one embodiment, the invention provides a method for the expression of a κ-like surrogate

light chain in a recombinant host cell comprising transforming said recombinant host cell with nucleic acid encoding a chimeric molecule comprising a Vκ-like sequence covalently connected to a JCκ sequence, wherein the native secretory leader sequence of said Vκ-like sequence and/or said JCκ sequence is replaced by a heterologous secretory leader sequence. In one other embodiment, the Vκ-like sequence is fused to the JCκ sequence. In another embodiment, the Vκ-like sequence is connected to the JCκ sequence through a peptide or polypeptide linker. In another embodiment, the recombinant host cell is an eukaryotic cell. In one embodiment, the recombinant host cell is a Chinese Hamster Ovary (CHO) cell, or a human embryonic kidney (HEK) 293 cell.

In one other embodiment, the invention provides an SLC construct comprising a VpreB sequence shown as SEQ ID NO:6. In another embodiment, the invention provides an SLC construct comprising a λ5 sequence shown as SEQ ID NO:10. In one embodiment, the invention provides an SLC construct comprising a polypeptide shown as SEQ ID NO:35.

Specific examples of λ-like Surrobodies include polypeptides in which a VpreB sequence, such as a VpreB1, VpreB2, or VpreB3 sequence, including fragments and variants of the native sequences, is conjugated to a λ5 sequence, including fragments and variants of the native sequence. Representative fusions of this type are provided in PCT Publication WO 2008/118970 published on October 2, 2008, the entire disclosures of which are expressly incorporated by reference herein. An example of a fusion with a heterogeneous leader sequence is illustrated in Figure 3 (SEQ ID NO:35).

In a direct fusion, typically the C-terminus of a VpreB sequence (e.g. a VpreB1, VpreB2 or VpreB3 sequence) is fused to the N-terminus of a λ5 sequence. While it is possible to fuse the entire length of a native VpreB sequence to a full-length λ5 sequence (see, e.g., the first diagram in Figure 7), typically the fusion takes place at or around a CDR3 analogous site in each of the two polypeptides. A representative fusion construct based on the analogous CDR3 sites for VpreB1 and λ5 is illustrated in Figure 6. In this embodiment, the fusion may take place within, or at a location within about 10 amino acid residues at either side of the CDR3 analogous region. In a preferred embodiment, the fusion takes place between about amino acid residues 116-126 of the native human VpreB1 sequence (SEQ ID NO: 1) and between about amino acid residues 82 and 93 of the native human λ5 sequence (SEQ ID NO: 7).

It is also possible to fuse the VpreB sequence or the λ5 sequence to the CDR3 region of an antibody λ light chain or the variable region of the antibody light chain respectively. Further

constructs, in which only one of VpreB and λ5 is truncated are also shown in Figure 7. Similar constructs can be prepared using antibody κ light chain sequences. Illustrations of κ-like surrogate light chain constructs can be found in Figures 12-19.

Further direct fusion structures are illustrated on the right side of Figure 11. The structure designated "SLC fusion 1" is a tetramer, composed of two dimers, in which the fusion of a truncated V-preB1 sequence (lacking the characteristic "tail" at the C-terminus of native VpreB1) to a similarly truncated λ5 sequence is non-covalently associated with an antibody heavy chain. The structure designated "SLC fusion 2" is a tetramer, composed of two dimers, in which the fusion of a truncated VpreB1 sequence (lacking the characteristic "tail" at the C-terminus of native VpreB1) to an antibody light chain constant region is non-covalently associated with an antibody heavy chain. The structure designated "SLC fusion 3" is a tetramer, composed of two dimers, in which the fusion of an antibody light chain variable region to a truncated λ5 sequence (lacking the characteristic "tail" at the N-terminus of native λ5) is non-covalently associated with an antibody heavy chain.

As noted above, in addition to direct fusions, the polypeptide constructs of the present invention include non-covalent associations of a VpreB sequence (including fragments and variants of a native sequence) with a heterogeneous sequence, such as a λ5 sequence (including fragments and variants of the native sequence), and/or an antibody sequence. Thus, for example, a full-length VpreB sequence may be non-covalently associated with a truncated λ5 sequence. Alternatively, a truncated VpreB sequence may be non-covalently associated with a full-length λ5 sequence.

Surrogate light chain constructs comprising non-covalently associated VpreB1 and λ5 sequences, in non-covalent association with an antibody heavy chain, are shown on the left side of Figure 11. As the various illustrations show, the structures may include, for example, full-length VpreB1 and λ5 sequences, a full-length VpreB1 sequence associated with a truncated λ5 sequence ("Lambda 5dT"), a truncated VpreB1 sequence associated with a full-length λ5 sequence (VpreB dT") and a truncated VpreB1 sequence associated with a truncated λ5 sequence ("Short").

Although Figure 11 illustrates certain specific constructs, one of ordinary skill will appreciate that a variety of other constructs can be made and used in a similar fashion. For example, the structures can be asymmetrical, comprising different surrogate light chain

sequences in each arm, and/or having trimeric or pentameric structures, as opposed to the structures illustrated in Figure 11.

All surrogate light chain constructs (Surrobodies) herein may be associated with antibody sequences. For example, as shown in Figure 9, a VpreB-λ5 fusion can be linked to an antibody heavy chain variable region sequence by a peptide linker. In another embodiment, a VpreB-λ5 fusion is non-covalently associated with an antibody heavy chain, or a fragment thereof including a variable region sequence to form a dimeric complex. In yet another embodiment, the VpreB and λ5 sequences are non-covalently associated with each other and an antibody heavy chain, or a fragment thereof including a variable region sequence, thereby forming a trimeric complex. Exemplary constructs comprising an antibody heavy chain are illustrated in Figure 11.

Specific examples of κ-like Surrobodies include polypeptides in which a Vκ-like sequence, including fragments and variants of the native sequences, is conjugated to a JCκ sequence, including fragments and variants of the native sequence. Representative fusions of this type are illustrated in U.S. Patent Publication No. 2001-0062950, and Xu et al., *J. Mol. Biol.* 2010, 397, 352-360, the entire disclosures of which are expressly incorporated by reference herein.

Specific examples of the polypeptide constructs herein include polypeptides in which a Vκ-like and/or JCκ sequence is associated with an antibody heavy chain, or a fragment thereof. Specific heterodimeric constructs, comprising both Vκ-like and JCκ sequences, are illustrated in Figure 12. As shown in Figure 12, in the κ-like surrogate light chain constructs of the present invention, the Vκ-like polypeptide and/or the JCκ polypeptide may contain the C- and N-terminal extensions, respectively, that are not present in similar antibody sequences. Alternatively, part or whole of the extension(s) can be removed from the κ-like surrogate light chain constructs herein.

Other κ-like surrogate light chain constructs, which can be used individually or can be further derivatized and/or associated with additional heterogeneous sequences, such as antibody heavy chain sequences, such as a full-length antibody heavy chain or a fragment thereof.

While the C- and N-terminal extensions of the Vκ-like polypeptide and/or the JCκ polypeptide do not need to be present in the constructs of the present invention, it is advantageous to retain at least a part of at least one of such appendages, because they provide a unique opportunity to create combinatorial functional diversity, either by linear extensions or,

for example, in the form of constrained diversity, as a result of screening loop libraries, as shown in Figure 14. In addition, the "tail" portions of the Vκ-like polypeptide and/or the JCκ polypeptide can be fused to other peptides and/or polypeptides, to provide for various desired properties, such as, for example, enhanced binding, additional binding specificities, enhanced pK, improved half-life, reduced half-life, cell surface anchoring, enhancement of cellular translocation, dominant negative activities, etc. Specific functional tail extensions are listed in Figure 18.

If desired, the constructs of the present invention can be engineered, for example, by incorporating or appending known sequences or sequence motifs from the CDR1, CDR2 and/or CDR3 regions of antibodies, including known therapeutic antibodies into the CDR1, CDR2 and/or CDR3 analogous regions of the κ-like surrogate light chain sequences. This allows the creation of molecules that are not antibodies, but will exhibit binding specificities and affinities similar to or superior over those of a known therapeutic antibody.

As Vκ-like and the JCκ genes encode polypeptides that can function as independent proteins and function as surrogate light chains, surrogate-like light chains can be engineered from true light chains and be used in every previous application proposed for engineered true surrogate light chains. This can be accomplished by expressing the variable light region to contain a peptidic extension analogous to either the VpreB or Vκ-like gene. Similarly the constant region can be engineered to resemble either the λ5 or JCκ genes and their peptidic extensions. Furthermore any chimeras or heterodimeric partnered combinations are within the scope herein.

In some embodiments, the SLC constructs comprise heterogenous amino acid sequences or non-SLC polypeptides. In certain embodiments, the heterogeneous amino acid sequence can add one or more additional functionalities to the construct of the present invention. SLC constructs may be designed to include non-SLC polypeptides. In one embodiment, the non-SLC polypeptide fused to a first SLC component and/or a second SLC component. The amino terminus of the non-SLC polypeptide may be fused to the carboxy terminus of the first SLC component and/or the carboxy terminus of the non-SLC polypeptide may be fused to the amino terminus of the second SLC component. In another embodiment, the first SLC component is a VpreB polypeptide or a Vκ-like polypeptide. In one other embodiment, the second SLC component is a λ5 polypeptide or a JCκ polypeptide.

λ-like SLC constructs with additional functionalities including antibody variable region sequences with desired binding specificities are illustrated in Figure 10. In particular, Figure 10 illustrates the insertion of an anti-VEGF single chain Fv (scFv) to create a fusion protein linking VpreB and λ5 (Figure 10A). This resulting engineered SLC-constrained scFv is paired with the heavy chain of an anti-TNF-α antibody. Figure 10B depicts the fusion of the anti-VEGF scFv to the C-terminus of VpreB. Figure 10C depicts the fusion of an anti-ovalbumin scFv to the amino terminus of λ5. A tripartite protein complex having the potential to bind to both TNF-α and ovalbumin can be formed. Figure 10D depicts the combination of two fusion constructs (VpreB-anti-VEGF scFv and the λ5-anti-ovalbumin) with the heavy chain of the anti-TNF-α antibody to create a trispecific molecule. A variety of bifunctional and trifunctional constructs, including VpreB and λ5 polypeptide sequences may be constructed using such a strategy. In addition, as depicted in Figure 8, combinatorial functional diversity may be incorporated into λ-like SLC constructs.

κ-like SLC constructs with additional functionalities including antibody variable region sequences with desired binding specificities are illustrated in Figure 17. In particular, Figure 17 illustrates a variety of bifunctional and trifunctional constructs, including Vκ-like and JCκ polypeptide sequences as hereinabove described.

The surrogate light chain (SLC) constructs of the present invention may be provided in dimeric or 2-piece format. Examples of this format are provided in Figure 9, which show a protein structure of a VpreB-λ5 fusion and an antibody heavy chain (right side, 2nd depiction from the bottom) corresponding to a 2-piece format. The SLC constructs may also be provided in a trimeric or 3-piece format. Figure 9 shows a protein structure of VpreB, λ5, and an antibody heavy chain (right side, depiction at the bottom) corresponding to a 3-piece format.

The surrogate light chain (SLC) constructs of the present invention may be provided in bifunctional or bispecific formats. Figure 20 shows examples of this: (A) depicts a Surrobody format, while (B) depicts bifunctional and bispecific Surrobody formats. As shown in Figure 20 (B), an SLC construct may include a SLC fusion polypeptide having an SLC polypeptide component (e.g., VpreB, λ5, Vκ-like, JCκ polypeptides, or fragments or variants thereof), and a non-SLC molecule. In one embodiment, the non-SLC molecule may be any polypeptide having a certain function. In another embodiment, the polypetpide may be a cytokine, which can provide additional functionality. In another embodiment, the non-SLC polypeptide may be an antibody fragment, which can provide additional specificity. Figure 20 (C) depicts an exemplary

SLC fusion cloning strategies and the respective amino acids. The diagonal hatched areas represent non-immunoglobulin tail regions of VpreB (amino acids 120-145) and λ5 (amino acids 38-92). An "L" indicates an endogenous leader sequence while an "mL" indicates a synthetic Ig κ leader sequence. A "Fusion" indicates fusion sites for genes of interest, such as a non-SLC molecule.

Preparation of surrogate light chain constructs

Nucleic acids encoding surrogate light chain, e.g. VpreB and λ5 polypeptides or Vκ-like or JCκ polypeptides, can be isolated from natural sources, e.g. developing B cells and/or obtained by synthetic or semi-synthetic methods. Once this DNA has been identified and isolated or otherwise produced, it can be ligated into a replicable vector for further cloning or for expression.

Cloning and expression vectors that can be used for expressing the coding sequences of the polypeptides herein are well known in the art and are commercially available. The vector components generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Suitable host cells for cloning or expressing the DNA encoding the surrogate light chain constructs in the vectors herein are prokaryote, yeast, or higher eukaryote (mammalian) cells, mammalian cells are being preferred.

Examples of suitable mammalian host cell lines include, without limitation, monkey kidney CV1 line transformed bySV40 (COS-7, ATCC CRL 1651); human embryonic kidney (HEK) line 293 (HEK 293 cells) subcloned for growth in suspension culture, Graham et al, *J. Gen Virol.* 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO, Urlaub et al., *Proc. Natl. Acad. Sci. USA* 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod. 23:243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., *Annals N.Y. Acad. Sci.* 383:44-68 (1982)); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2).

For use in mammalian cells, the control functions on the expression vectors are often provided by viral material. Thus, commonly used promoters can be derived from the genomes of polyoma, Adenovirus2, retroviruses, cytomegalovirus, and Simian Virus 40 (SV40). Other promoters, such as the β-actin protomer, originate from heterologous sources. Examples of suitable promoters include, without limitation, the early and late promoters of SV40 virus (Fiers *et al.*, *Nature*, 273: 113 (1978)), the immediate early promoter of the human cytomegalovirus (Greenaway *et al.*, *Gene*, 18: 355-360 (1982)), and promoter and/or control sequences normally associated with the desired gene sequence, provided such control sequences are compatible with the host cell system.

Transcription of a DNA encoding a desired heterologous polypeptide by higher eukaryotes is increased by inserting an enhancer sequence into the vector. The enhancer is a cis-acting element of DNA, usually about from 10 to 300 bp, that acts on a promoter to enhance its transcription-initiation activity. Enhancers are relatively orientation and position independent, but preferably are located upstream of the promoter sequence present in the expression vector. The enhancer might originate from the same source as the promoter, such as, for example, from a eukaryotic cell virus, e.g. the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Expression vectors used in mammalian host cells also contain polyadenylation sites, such as those derived from viruses such as, e.g., the SV40 (early and late) or HBV.

An origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral (e.g., Polyoma, Adeno, VSV, BPV) source, or may be provided by the host cell.

The expression vectors usually contain a selectable marker that encodes a protein necessary for the survival or growth of a host cell transformed with the vector. Examples of suitable selectable markers for mammalian cells include dihydrofolate reductase (DHFR), thymidine kinase (TK), and neomycin.

Suitable mammalian expression vectors are well known in the art and commercially available. Thus, for example, the surrogate light chain constructs of the present invention can be produced in mammalian host cells using a pCI expression vector (Promega), carrying the human cytomegalovirus (CMV) immediate-early enhancer/promoter region to promote constitutive expression of a DNA insert. The vector may also be the pTT5 expression vector (National

Research Council, Canada). The vector can contain a neomycin phosphotransferase gene as a selectable marker.

The surrogate light chain constructs of the present invention can also be produced in bacterial host cells. Control elements for use in bacterial systems include promoters, optionally containing operator sequences, and ribosome binding sites. Suitable promoters include, without limitation, galactose (gal), lactose (lac), maltose, tryptophan (trp), β-lactamase promoters, bacteriophage λ and T7 promoters. In addition, synthetic promoters can be used, such as the tac promoter. Promoters for use in bacterial systems also generally contain a Shine-Dalgarno (SD) sequence operably linked to the DNA encoding the Fab molecule. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria.

The coding sequences of the individual chains within a multi-chain construct comprising antibody surrogate light chain sequences can be present in the same expression vector, under control of separate regulatory sequences, or in separate expression vectors, used to co-transfect a desired host cells, including eukaryotic and prokaryotic hosts. Thus, multiple genes can be coexpressed using the Duet™ vectors commercially available from Novagen.

The transformed host cells may be cultured in a variety of media. Commercially available media for culturing mammalian host cells include Ham's F10 (Sigma), Minimal Essential Medium ((MEM), (Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium ((DMEM), Sigma). In addition, any of the media described in Ham *et al.*, *Meth. Enz.* 58:44 (1979) and Barnes *et al.*, *Anal. Biochem.* 102:255 (1980) may be used as culture media for the host cells. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression, and are included in the manufacturer's instructions or will otherwise be apparent to the ordinarily skilled artisan.

Further suitable media for culturing mammalian, bacterial (e.g. *E. coli*) or other host cells are also described in standard textbooks, such as, for example, Sambrook et al., *supra*, or Ausubel et al., *supra*.

In one aspect, the present invention provides a method for the expression of a surrogate light chain in a recombinant host cell. In one embodiment, the method includes the step of providing a nucleic acid encoding an SLC polypeptide or an SLC fusion polypeptide. In another embodiment, the method includes the step of transforming or transfecting the recombinant host cell with a nucleic acid encoding an SLC polypeptide or SLC fusion polypeptide. In one

embodiment, the nucleic acid encoding an SLC fusion polypeptide is a chimeric molecule comprising a first SLC sequence covalently connected to a second SLC sequence, wherein the native secretory leader sequence of the first SLC sequence and/or the second SLC sequence is replaced by a heterologous secretory leader sequence. The first SLC sequence may be a VpreB sequence, a Vκ-like sequence, or a fusion polypeptide thereof. The second SLC sequence may be a λ5 sequence, a JCκ sequence, or a fusion polypeptide thereof.

In one embodiment, a VpreB sequence is covalently connected to a λ5 sequence, wherein the native secretory leader sequence of said VpreB sequence and/or said λ5 sequence is replaced by a heterologous secretory leader sequence. In another embodiment, the VpreB sequence is fused to the λ5 sequence. In one other embodiment, the VpreB sequence is connected to the λ5 sequence through a peptide or polypeptide linker. In one other embodiment, a Vκ-like sequence is covalently connected to a JCκ sequence, wherein the native secretory leader sequence of said Vκ-like sequence and/or said JCκ sequence is replaced by a heterologous secretory leader sequence. In one other embodiment, the Vκ-like sequence is fused to the JCκ sequence. In another embodiment, the Vκ-like sequence is connected to the JCκ sequence through a peptide or polypeptide linker.

In all embodiments, the methods of expression may comprise the step of transforming or transfecting a host cell with more than one nucleic acid encoding a surrogate light chain polypeptide, including surrogate light chain polypeptides and/or surrogate light chain fusion polypeptides.

In all embodiments, the methods may further comprise the step of transforming or transfecting a host cell with a nucleic acid encoding an antibody heavy chain.

In one aspect, the present invention provides methods for the expression of surrogate light chain polypeptides and/or surrogate light chain fusion polypeptides having improved yields. In one embodiment, the methods of the present invention utilizing heterologous leader sequences in place of native leader sequences are characterized greater polypeptide expression and yield than methods which do not replace native leader sequences with heterologous leader sequences.

In one embodiment, the recombinant host cell is bacterial cell. In another embodiment, the host cell is a eukaryotic cell. In one embodiment, the recombinant host cell is a Chinese Hamster Ovary (CHO) cell, or a human embryonic kidney (HEK) 293 cell.

In one aspect, the present invention provides host cells containing the nucleic acids described herein. In one embodiment, the invention provides a recombinant host cell transformed with at least one nucleic acid described herein. In one other embodiment, the host cell is transformed with a nucleic acid encoding an SLC fusion, which may or may not include a non-SLC molecule.

In all embodiments, the host cell is further transformed with a nucleic acid encoding an antibody heavy chain.

In all embodiments, the present invention provides vectors that contain the nucleic acids described herein. In all embodiments, the host cell is transformed with at least one vector containing a nucleic acid described herein.

Purification can be performed by methods known in the art. In a preferred embodiment, the surrogate light chain constructs are purified in a 6xHis-tagged form, using the Ni-NTA purification system (Invitrogen).

κ-like SLC molecules can be engineered from existing light chain V genes and light chain constant genes. As shown in Figure 15, light chains are products of gene rearrangement and RNA processing. As the components of the κ-like SLC molecules provide alternative function from unrearranged light chain V genes and rearranged light chain JC genes, it is feasible to engineer similar translated proteins from all remaining kappa and lambda light chain V genes to make Vκ -like molecules (Figure 16A) and all combinations of the remaining kappa JC rearrangements (4 JCκ -like) (Figure 16B) and lambda JC rearrangements (4 "J" x 10 "constant" = 40 JCλ-like) (Figure 16B). Each one of these engineered molecules can serve purposes similar to those using Vκ-like and JCκ, as well as those contained in PCT Publication WO 2008/118970 published on October 2, 2008, with VpreB and λ5, and combinations and chimeras thereof.

The surrogate light chains of the present invention can be used to construct molecules for the prevention and/or treatment of disease. For such applications, molecules containing a surrogate light chain are usually used in the form of pharmaceutical compositions. Techniques and formulations generally may be found in Remington's Pharmaceutical Sciences, 18th Edition, Mack Publishing Co. (Easton, Pa. 1990). See also, Wang and Hanson "Parenteral Formulations of Proteins and Peptides: Stability and Stabilizers," Journal of Parenteral Science and Technology, Technical Report No. 10, Supp. 42-2S (1988).

Polypeptide-based pharmaceutical compositions are typically formulated in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes {e.g., Zn-protein complexes); and/or non-ionic surfactants such as TWEEN™, PLURONICS™ or polyethylene glycol (PEG).

The molecules also may be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization (for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacylate) microcapsules, respectively), in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules), or in macroemulsions. Such techniques are disclosed in Remington's Pharmaceutical Sciences, *supra*.

The molecules containing surrogate light chains disclosed herein may also be formulated as immunoliposomes. Liposomes containing the molecules are prepared by methods known in the art, such as described in Epstein et al, *Proc. Natl. Acad. Sci.* USA 82:3688 (1985); Hwang et al, *Proc. Natl Acad. Sci. USA* 77:4030 (1980); U.S. Patent Nos. 4,485,045 and 4,544,545; and WO97/38731 published October 23, 1997. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol and PEG- derivatized phosphatidyl ethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fragments of the molecules of the present invention can be conjugated to the liposomes via a disulfide interchange reaction (Martin et al. J. Biol. Chem. 257:286-288 (1982). A chemotherapeutic agent is optionally contained within the liposome. See Gabizon et al. J. National Cancer Inst. 81(19)1484 (1989).

For the prevention or treatment of disease, the appropriate dosage of molecule will depend on the type of infection to be treated the severity and course of the disease, and whether the antibody is administered for preventive or therapeutic purposes. The molecule is suitably administered to the patient at one time or over a series of treatments. Depending on the type and severity of the disease, about 1 μg/kg to about 15 mg/kg of antibody is a typical initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion.

Molecules containing a surrogate light chain of the present invention are suitable for use in the treatment or prevention of diseases. In one embodiment, the present invention provides a surrogate light chain-containing molecule for use as a medicament, or for the treatment of a disease. In another embodiment, the present invention provides the use of a surrogate light chain-containing molecule for the manufacture of a medicament for treating disease. The molecule may be a nucleic acid encoding an SLC polypeptide or SLC fusion.

In one aspect, the invention provides methods useful for treating a disease in a mammal, the methods including the step of administering a therapeutically effective amount of a surrogate light chain-containing molecule to the mammal. The therapeutic compositions can be administered short term (acute) or chronic, or intermittent as directed by physician.

The invention also provides kits and articles of manufacture containing materials useful for the treatment, prevention and/or diagnosis of disease. The kit includes a container and a label, which can be located on the container or associated with the container. The container may be a bottle, vial, syringe, or any other suitable container, and may be formed from various materials, such as glass or plastic. The container holds a composition having a surrogate light chain-containing molecule as described herein, and may have a sterile access port. Examples of containers include an intravenous solution bag or a vial with a stopper that can be pierced by a hypodermic injection needle. The kits may have additional containers that hold various reagents, e.g., diluents and buffers. The label may provide a description of the composition as well as instructions for the intended use. Kits containing the molecules find use, e.g., for cellular assays, for purification or immunoprecipitation of a polypeptide from cells. For example, for isolation and purification of a protein, the kit can contain a surrogate light chain-containing molecule that binds the protein coupled to beads (e.g., sepharose beads). Kits can be provided which contain the molecules for detection and quantitation of the protein in vitro, e.g., in an ELISA or a Western blot. Such molecules useful for detection may be provided with a label such as a fluorescent or radiolabel.

The kit has at least one container that includes a molecule comprising a surrogate light chain described herein as the active agent. A label may be provided indicating that the composition may be used to treat a disease. The label may also provide instructions for administration to a subject in need of treatment. The kit may further contain an additional container having a pharmaceutically-acceptable buffer, such as bacteriostatic water for injection (BWFI), phosphate-buffered saline, Ringer's solution and dextrose solution. Finally, the kit may also contain any other suitable materials, including other buffers, diluents, filters, needles, and syringes.

Further details of the invention are provided in the following non-limiting examples.

Example 1 - Transient Expression in HEK293 cells

Surrobody supernatants were transiently produced in human embryonic kidney 293 (HEK 293) cells. A pTT5 plasmid (National Research Council, Canada) was used to provide recombinant Surrobody light chain constructs. The surrogate light chain nucleic acid sequences of the plasmids were provided with or without the substitution of the native leader sequence with a heterologous leader sequence METDTLLLWVLLLWVPGSTG (SEQ ID NO:36 - murine Ig κ leader sequence). pTT5 plasmids containing the following nucleic acid sequences were used: (a) human VpreB1 with a native leader sequence (SEQ ID NO:1 - Figure 1) or the heterologous murine Ig κ leader sequence of SEQ ID NO:36, (b) human λ5 with a native leader sequence (SEQ ID NO:8 - Figure 2) or the heterologous murine Ig κ leader sequence of SEQ ID NO:36, or (c) a fusion of VpreB1 and λ5 with a native VpreB1 leader sequence or the heterologous murine Ig κ leader sequence of SEQ ID NO:36 (SEQ ID NO:35 - Figure 3). The plasmids of (a) and (b) correspond to 3-piece Surrobody formats while the plasmids of (c) correspond to 3-piece Surrobody formats. These plasmids were co-transfected with pTT5 plasmids containing an antibody heavy chain.

Surrobodies were transiently produced in HEK293 Freestyle-based systems (Invitrogen) essentially as previously described in Xu et al., (2008). *Proc. Natl Acad. Sci. USA*, 105, 10756–10761; Kashyap et al., (2008). *Proc. Natl Acad. Sci. USA*, 105, 5986–5991. The HEK 293 cells were propagated in growth medium at densities between 0.25 – 2.0 x 10^6 cells/ml and then inoculated one day prior to transfection into a fresh shake flask containing 90 ml growth medium at a density of 0.75 x10^6 cells/ml. After overnight growth, the cell density was verified at between 1-1.5 x10^6 cells/ml. For expression, a pTT5 expression vector was used.. Next a DNA-Transfection agent mixture was prepared as follows. A DNA solution corresponding to a total

of 0.1 mg plasmid DNA (pTT5-SLC molecule) was mixed in a 10 ml centrifuge tube, to a final volume of 5 mL with growth media. For the 2-piece Surrobody format, 0.05 mg of a plasmid containing an antibody heavy chain was mixed with 0.05 mg of a VpreB1-λ5 chimeric plasmid (Surrobody fusion). For the 3-piece Surrobody format, 0.033 mg of a plasmid containing an antibody heavy chain was mixed with 0.033 mg of a VpreB1 plasmid, and 0.033 mg of a λ5 plasmid . Next, a polyethylenimine (PEI) transfection solution was prepared by combining 4.8 ml growth media with 0.2 ml PEI, which is added to the plasmid DNA solutions. The mixture was vigorously vortexed for 1-2 seconds. After incubating at room temperature for 15 minutes, the plasmid DNA-PEI mixture was transferred with 10 ml pipette to a flask containing HEK 293 cells at a density of between 1-1.5 x10^6 cells/ml. The flask was immediately swirled and transferred to a shaking incubator. The cells are grown in a humidified incubator at 37°C and 5% CO^2 with a shaker platform at 125 rpm for six days. Protein production levels in the resulting culture supernatants were determined by quantitative kinetic analysis (ForteBio – Octet: Anti-Fc sensors). As shown in the table below, substituting a murine Ig κ light chain leader sequence improves transient recombinant Surrobody expression levels. Protein levels were improved by at least 20-fold as shown in Table 1 (mg per L).

Table 1

	mg/L	
	endogenous leader	murine Ig kappa leader
3 - piece	0.2	7.4
	0.1	26.4
		4.9
2 - piece	6.7	104.6

Further analysis of purified proteins from multiple transfections of 3-piece and 2-piece Surrobodies support high level yields. Protein yields from independent transfections, using the heterologous leader, were monitored over a 4 month period. The proteins were purified with a fast protein liquid chromatography (FPLC) system using either Protein A or Protein G chromatographic supports and low pH elution. In either format the average yields were substantially higher than that seen using the endogenous surrogate light chain leader sequence, as shown in Table 2 (mg per L).

Table 2

2 piece average mg/L (n=47)	3 piece average mg/L (n=15)
49.2	30.1

Table 3 below provides the individual concentrations measured for various 2-piece and 3-piece SLC formats that are averaged in Table 2. In general, as described above, the 2-piece format includes a Surrobody light chain fusion and an antibody heavy chain, while the 3-piece format includes two SLC polypeptides and an antibody heavy chain. Column 1, rows 1-47 correspond to the 47 different constructs having a 2-piece format and column 1, and rows 49-63 correspond to the 15 different constructs having a 3-piece format. The fourth column provides some of the features of the surrobodies tested. A "Surrobody" is a Surrobody construct made up of two SLC polypeptides and a heavy chain. A "fusion" is a Surrobody construct made up of a fusion of two SLC polypeptides and a heavy chain. A "fusion with peptide tag" is a Surrobody construct in which an epitope tag is incorporated. A "functional peptide fusion" is a Surrobody construct in which at least one non-SLC polypeptide sequence with a certain function has been incorporated.

Table 3

Protein prep #	Surrobody (SgG) format (2- or 3-piece)	Concentration (mg/L)	Features
1	2	91.6	Fusion with peptide tag
2	2	42.8	Functional peptide fusion
3	2	40.8	Functional peptide fusion
4	2	27.2	Fusion with peptide tag
5	2	97.9	Fusion with peptide tag
6	2	54.0	Fusion
7	2	32.8	Functional peptide fusion
8	2	32.8	Functional peptide fusion
9	2	18.2	Functional peptide fusion
10	2	36.0	Functional peptide fusion
11	2	23.3	Fusion
12	2	20.0	Fusion
13	2	65.6	Fusion with peptide tag

14	2	16.0	Fusion with peptide tag
28	2	56.2	Fusion with peptide tag
29	2	27.2	Fusion with peptide tag
30	2	31.6	Fusion with peptide tag
31	2	99.2	Functional peptide fusion
32	2	139.2	Functional peptide fusion
33	2	70.4	Functional peptide fusion
34	2	121.6	Functional peptide fusion
35	2	14.4	Fusion
36	2	91.1	Fusion with peptide tag
37	2	14.4	Functional peptide fusion
38	2	5.1	Functional peptide fusion
39	2	2.1	Functional peptide fusion
40	2	0.7	Functional peptide fusion
41	2	16.7	Fusion with peptide tag
42	2	13.3	Fusion with peptide tag
43	2	67.5	Fusion
44	2	7.4	Fusion`
45	2	81.6	Fusion
46	2	124.8	Fusion
47	2	17.5	Fusion
1	3	44.6	Surrobody
2	3	43.4	Functional peptide fusion
3	3	18.2	Functional peptide fusion
4	3	97.9	Surrobody
5	3	16.0	Functional peptide fusion
6	3	39.4	Functional peptide fusion
7	3	18.0	Surrobody
8	3	5.3	Surrobody
9	3	6.2	Surrobody
10	3	36.7	Surrobody
11	3	3.8	Surrobody
12	3	12.8	Surrobody
13	3	22.4	Surrobody
14	3	54.7	Surrobody
15	3	32.0	Surrobody

Table 3 provides evidence that improved yields can be obtained for multiple Surrobody formats, including Surrobodies that comprise SLC polypeptides, SLC fusion polypeptides, and SLC fusion polypeptides that contain non-SLC molecules.

Surrobody molecules may also be transiently expressed in Chinese hamster ovary K1 (CHO-K1) cells. Surrobody supernatants may be transiently produced in (CHO-K1) cells. Plasmids of recombinant SLC polypeptides or non-SLC polypeptides are cotransfected with plasmids containing antibody heavy chain using Lipofectamine-2000 in Dulbecco's modified Eagle's medium/F12mediumsupplemented with 10%fetal bovine serum following manufacturer's instruction (Invitrogen, catalog no. 11668-027). After overnight incubation at 37 °C, with 5% CO2, the medium is replaced with fresh Opti-MEMI Reduced-SerumMediumwith Glutamax-1 (Invitrogen, catalog no.12362). The transfected supernatants are harvested 72 hours later and filtered through a 0.22-μm filter unit.

Example 2 - Stable Expression in CHO cells

Mammalian expressed surrogate light chain constructs or generated by *de novo* synthesis as eukaryotic codon optimized soluble secreted genes (DNA 2.0) are subcloned into a pCI plasmid (Promega) for mammalian protein expression. The sequence is verified before transfection into Chinese hamster ovary (CHO-K1) cells (Invitrogen) according to manufacturers guidelines. A transfection of 80% confluent cells in T-75 flasks are performed using equal amounts of desired surrogate light chains totaling 32 μg of DNA and Lipofectamine 2000 (Invitrogen) according to manufacturers guidelines. Cells are allowed to produce proteins into 20 ml of Opti-MEM I per transfection. After 4 days the secreted Surrobodies are purified from the culture supernatents using nickel chelate chromatography (Ni-NTA agarose, Qiagen). The resulting purified Surrobodies are buffer exchanged into sterile PBS using centrifugal size filtration (Centricon Plus-20) and their protein concentrations determined by A280 readings, SDS gel, or Western blot analysis compared to known standards.

Although in the foregoing description the invention is illustrated with reference to certain embodiments, it is not so limited. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

All publications, patents and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual publication,

patent or patent application were specifically and individually indicated to be so incorporated by reference.

What is claimed is:

1. An isolated nucleic acid molecule encoding a surrogate light chain (SLC) polypeptide, wherein the native secretory leader sequence of the polypeptide is replaced by a heterologous secretory leader sequence.

2. The isolated nucleic acid molecule of claim 1, wherein the SLC polypeptide comprises a VpreB polypeptide.

3. The isolated nucleic acid molecule of claim 1, wherein the SLC polypeptide is a λ5 polypeptide.

4. The isolated nucleic acid molecule of claim 2, wherein the VpreB polypeptide is selected from the group consisting of a native VpreB1 sequence, a native VpreB2 sequence, a native VpreB3 sequence, and fragments and variants thereof.

5. The isolated nucleic acid molecule of claim 4,wherein the native VpreB sequence is selected from the group consisting of human VpreB1 of SEQ ID NO: 1, mouse VpreB2 of SEQ ID NOS: 2 and 3, human VpreB3 of SEQ ID NO: 4, human VpreB-like polypeptide of SEQ ID NO:5, human VpreB dTail polypeptide of SEQ ID NO:6 and fragments and variants thereof.

6. The isolated nucleic acid molecule of claim 3, wherein the λ5 polypeptide is selected from the group consisting of a human λ5-like of SEQ ID NO: 7; a human λ5 polypeptide of SEQ ID NO: 8, a human λ5 dTail polypeptide of SEQ ID NO:9, and fragments and variants thereof.

7. The isolated nucleic acid molecule of claim 1, wherein the SLC polypeptide comprises a VpreB-λ5 polypeptide fusion.

8. The isolated nucleic acid molecule of claim 7, wherein fusion of the VpreB polypeptide sequence and λ5 polypeptide sequence takes place at or around the CDR3 analogous regions of said VpreB sequence and said λ5 sequence respectively.

9. The isolated nucleic acid molecule of claim 7, wherein the VpreB polypeptide sequence is linked at its carboxy terminus to the amino terminus of the λ5 polypeptide sequence.

10. The isolated nucleic acid molecule of claim 1, wherein the SLC polypeptide comprises a Vκ -like polypeptide.

11. The isolated nucleic acid molecule of claim 1 or claim 10, wherein the SLC polypeptide comprises a JCκ polypeptide.

12. The isolated nucleic acid molecule of claim 10, wherein the Vκ -like polypeptide sequence is selected from the group consisting of SEQ ID NOS: 12-24, and fragments and variants thereof.

13. The isolated nucleic acid molecule of claim 11, wherein the JCκ polypeptide sequence is selected from the group consisting of SEQ ID NOS:26-39, and fragments and variants thereof.

14. The isolated nucleic acid molecule of claim 1, wherein the SLC polypeptide is a Vκ-like-JCκ polypeptide fusion.

15. The isolated nucleic acid molecule of claim 14, wherein the fusion of the Vκ-like polypeptide sequence and JCκ polypeptide sequence takes place at or around the CDR3 analogous regions of said Vκ-like sequence and said JCκ sequence respectively.

16. The isolated nucleic acid molecule of claim 14, wherein the Vκ-like polypeptide sequence is fused at its carboxy terminus to the amino terminus of the JCκ polypeptide sequence.

17. The isolated nucleic acid molecule of any one of claims 1 to 16 wherein said heterologous secretory leader sequence is a leader sequence of a secreted polypeptide selected from the group consisting of antibodies, cytokines, lymphokines, monokines, chemokines, polypeptide hormones, digestive enzymes, and components of the extracellular matrix.

18. The isolated nucleic acid molecule of claim 17 wherein said cytokine is selected from the group consisting of growth hormone, such as human growth hormone, N-methionyl human growth hormone, and bovine growth hormone; parathyroid hormone; thyroxine; insulin; proinsulin; relaxin; prorelaxin; glycoprotein hormones such as follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH), and luteinizing hormone (LH); hepatic growth factor; fibroblast growth factor; prolactin; placental lactogen; tumor necrosis factor-α and -β (TNF-α and -β); mullerian-inhibiting substance; mouse gonadotropin-associated peptide;

inhibin; activin; vascular endothelial growth factor; integrin; thrombopoietin (TPO); nerve growth factors such as NGF-β; platelet-growth factor; transforming growth factors (TGFs) such as TGF-α and TGF-β; insulin-like growth factor-I and -II; erythropoietin (EPO); osteoinductive factors; interferons such as interferon-α, -β and -γ; colony stimulating factors (CSFs) such as macrophage-CSF (M-CSF); granulocyte-macrophage-CSF (GM-CSF); and granulocyte-CSF (G-CSF); interleukins (ILs) such as IL-1, IL-1a, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12; a tumor necrosis factor such as TNF-α or TNF-β; MIP-1α; MIP-1β; and other polypeptide factors including LIF and kit ligand (KL).

19. The isolated nucleic acid molecule of any one of claims 1 to 16, wherein said secretory leader sequence is selected from the group consisting of leader sequences of human and non-human mammalian albumin, transferrin, CD36, growth hormone, tissue plasminogen activator (t-PA), erythropoietin (EPO), and neublastin.

20. The isolated nucleic acid molecule of any one of claims 1 to 16, wherein said secretory leader sequence is a synthetic sequence.

21. The isolated nucleic acid molecule of any one of claims 1 to 16, wherein said secretory leader sequence is a consensus sequence of native secretory leader sequences.

22. An isolated nucleic acid molecule according to any one of claims 1 to 21 encoding a surrogate light chain (SLC) construct.

23. A vector comprising the nucleic acid molecule of any one of claims 1 to 22.

24. A recombinant host cell transformed with the nucleic acid molecule of any one of claims 1 to 22 or the vector of claim 23.

25. A method for the expression of a surrogate light chain (SLC) polypeptide or SLC construct in a recombinant host cell comprising transforming said recombinant host cell with a nucleic acid molecule encoding an SLC polypeptide or SLC construct, wherein the native secretory leader sequence of the polypeptide is replaced by a heterologous secretory leader sequence.

26. The method of claim 25 wherein said recombinant host cell is an eukaryotic cell.

27. The method of claim 25 wherein said recombinant host cell is a Chinese Hamster Ovary (CHO) cell.

28. The method of claim 25 wherein said recombinant host cell is a human embryonic kidney (HEK) 293 cell.

29. The method of claim 25, wherein the SLC polypeptide or SLC construct is selected from the group consisting of an SLC polypeptide comprising one or more of a VpreB polypeptide, a λ5 polypeptide, a VpreB-λ5 polypeptide fusion, a Vκ -like polypeptide, a JCκ polypeptide, and a Vκ-like-JCκ polypeptide fusion.

MSWAPVLLMLFVYCTGCGPQPVLHQPPAMSSALGTTIRLTCTLRNDHDIGVYSVYWYQ
QRPGHPPRFLLRYFSQSDKSQGPQVPPRFSGSKDVARNRGYLSISELQPEDEAMYYCAM
GARSSEKEERERERWEEEMEPTAARTRVP

(SEQ ID NO:1)

MAWTSVLLMLLAHLTGCGPQPMVHQPPSASSSLGATIRLSCTLSNDHNIGIYSIYWYQQ
RPGHPPRFLLRYFSHSDKHQGPDIPPRFSGSKDTARNLGYLSISELQPEDEAVYYCAVGL
RSHEKKRMEREWEGEKSYTDLGS

(SEQ ID NO:2)

MAWTSVLLMLLAHLTGKGTLGVQGFLAPPVALLCPSDGHASIFSGCGPQPMVHQPPSA
SSSLGATIRLSCTLSNDHNIGIYSIYWYQQRPGHPPRFLLRYFSHSDKHQGPDIPPRFSGSK
DTARNLGYLSISELQPEDEAVYYCAVGLRSHEKKRMEREWEGEKSYTDLGS

(SEQ ID NO:3)

MACRCLSFLLMGTFLSVSQTVLAQLDALLVFPGQVAQLSCTLSPQHVTIRDYGVSWYQ
QRAGSAPRYLLYYRSEEDHHRPADIPDRFSAAKDEAHNACVLTISPVQPEDDADYYCSV
GYGFSP

(SEQ ID NO:4)

MSWAPVLLMLFVYCTGCGPQPVLHQPPAMSSALGTTIRLTCTLRNDHDIGVYSVYWYQ
QRPGHPPRFLLRYFSQSDKSQGPQVPPRFSGSKDVARNRGYLSISELQPEDEAMYYCAM
GA

(SEQ ID NO:5)

METDTLLLWVLLLWVPGSTGQPVLHQPPAMSSALGTTIRLTCTLRNDHDIGVYSVYWY
QQRPGHPPRFLLRYFSQSDKSQGPQVPPRFSGSKDVARNRGYLSISELQPEDEAMYYCA
MGARSSEKEERERERWEEEMEPTAARTRVP

(SEQ ID NO:6)

Figure 1

MKLRVGQTLGTIPRQCEVLLLLLLLGLVDGVHHILSPSSAERSRAVGPGASVGSNRPSL
WALPGRLLFQIIPRGAGPRCSPHRLPSKPQFWYVFGGGTQLTILGQPKSDPLVTLFLPSLK
NLQPTRPHVVCLVSEFYPGTLVVDWKVDGVPVTQGVETTQPSKQTNNKYMVSSYLTLI
SDQWMPHSRYSCRVTHEGNTVEKSVSPAECS

(SEQ ID NO:7)

MRPGTGQGGLEAPGEPGPNLRQRWPLLLLGLAVVTHGLLRPTAASQSRALGPGAPGGS
SRSSLRSRWGRFLLQRGSWTGPRCWPRGFQSKHNSVTHVFGSGTQLTVLSQPKATPSVT
LFPPSSEELQANKATLVCLMNDFYPGILTVTWKADGTPITQGVEMTTPSKQSNNKYAAS
SYLSLTPEQWRSRRSYSCQVMHEGSTVEKTVAPAECS

(SEQ ID NO:8)

MRPGTGQGGLEAPGEPGPNLRQRWPLLLLGLAVVTHGSVTHVFGSGTQLTVLSQPKAT
PSVTLFPPSSEELQANKATLVCLMNDFYPGILTVTWKADGTPITQGVEMTTPSKQSNNK
YAASSYLSLTPEQWRSRRSYSCQVMHEGSTVEKTVAPAECS

(SEQ ID NO:9)

METDTLLLWVLLLWVPGSTGSVTHVFGSGTQLTVLSQPKATPSVTLFPPSSEELQANKA
TLVCLMNDFYPGILTVTWKADGTPITQGVEMTTPSKQSNNKYAASSYLSLTPEQWRSRR
SYSCQVMHEGSTVEKTVAPAECS

(SEQ ID NO:10)

Figure 2

METDTLLLWVLLLWVPGSTGQPVLHQPPAMSSALGTTIRLTCTLRNDHDIGVYSVYWY
QQRPGHPPRFLLRYFSQSDKSQGPQVPPRFSGSKDVARNRGYLSISELQPEDEAMYYCA
MGARSSVTHVFGSGTQLTVLSQPKATPSVTLFPPSSEELQANKATLVCLMNDFYPGILTV
TWKADGTPITQGVEMTTPSKQSNNKYAASSYLSLTPEQWRSRRSYSCQVMHEGSTVEK
TVAPAECS

(SEQ ID NO:35)

Figure 3

```
          M  V  L     Q  T  Q  V  F     I  S  L  L  L     W  I  S  G  A     Y  G  D  I  V  .
1   CAGCAAGATGGTGTT GCAGACCCAGGTCTT CATTTCTCTGTTGCT CTGGATCTCTGGTGC CTACGGGGACATCGT
    GTCGTTCTACCACAA CGTCTGGGTCCAGAA GTAAAGAGACAACGA GACCTAGAGACCACG GATGCCCCTGTAGCA

          .  M  T  Q  S  P     D  S  L  A  V     S  L  G  E  R     A  T  I  N  C     K  S  S  Q  S  .
76  GATGACCCAGTCTCC AGACTCCCTGGCTGT GTCTCTGGGCGAGAG GGCCACCATCAACTG CAAGTCCAGCCAGAG
    CTACTGGGTCAGAGG TCTGAGGGACCGACA CAGAGACCCGCTCTC CCGGTGGTAGTTGAC GTTCAGGTCGGTCTC

          .  V  L  Y  S  S     N  N  K  N  Y     L  A  W  Y  Q     Q  K  P  G  Q     P  P  K  L  L  .
151 TGTTTTATACAGCTC CAACAATAAGAACTA CTTAGCTTGGTACCA GCAGAAACCAGGACA GCCTCCTAAGCTGCT
    ACAAAATATGTCGAG GTTGTTATTCTTGAT GAATCGAACCATGGT CGTCTTTGGTCCTGT CGGAGGATTCGACGA

          .  I  Y  W  A  S     T  R  E  S  G     V  P  D  R  F     S  G  S  G  S     G  T  D  F  T  .
226 CATTTACTGGGCATC TACCCGGGAATCCGG GGTCCCTGACCGATT CAGTGGCAGCGGGTC TGGGACAGATTTCAC
    GTAAATGACCCGTAG ATGGGCCCTTAGGCC CCAGGGACTGGCTAA GTCACCGTCGCCCAG ACCCTGTCTAAAGTG

          .  L  T  I  S  S     L  Q  A  E  D     V  A  V  Y  Y     C  Q  Q  Y  Y     S  T  P  P  T  .
301 TCTCACCATCAGCAG CCTGCAGGCTGAAGA TGTGGCAGTTTATTA CTGTCAGCAATATTA TAGTACTCCTCCCAC
    AGAGTGGTAGTCGTC GGACGTCCGACTTCT ACACCGTCAAATAAT GACAGTCGTTATAAT ATCATGAGGAGGGTG

          .  V  L  Q  P  R     T  Q  T  S  S     P  Y  A  G  P     V  G  L  C  C     S  S  C  F  L  .
376 AGTGCTTCAGCCTCG AACACAAACCTCCTC CCCATACGCTGGGCC AGTAGGTCTTTGCTG CAGCAGCTGCTTCCT
    TCACGAAGTCGGAGC TTGTGTTTGGAGGAG GGGTATGCGACCCGG TCATCCAGAAACGAC GTCGTCGACGAAGGA

          .  C  T  Q  P  P     T  C  M  L  P     L  C  V  G  E     V  T  L  L  I     Y  S  L  E  G  .
451 CTGCACACAGCCCCC AACATGCATGCTTCC TCTGTGTGTTGGGGA GGTCACTCTCTTGAT TTATTCGTTGGAGGG
    GACGTGTGTCGGGGG TTGTACGTACGAAGG AGACACACAACCCCT CCAGTGAGAGAACTA AATAAGCAACCTCCC

          .  L  Q  G  P  G     L  N  *
526 TTTGCAGGGCCCAGG ATTAAATTAAGAGAC TTGACTTTTGCTGGA TCTCTTTTTGTAGAA GATTATTAAAGCAAA
    AAACGTCCCGGGTCC TAATTTAATTCTCTG AACTGAAAACGACCT AGAGAAAAACATCTT CTAATAATTTCGTTT

601 ATGTTGTAAAGATCC CTTAGAGACATTGTC AGGAGTTTTTGTGTT ACAGGAACCTGCATG TTTCACATGGACACA
    TACAACATTTCTAGG GAATCTCTGTAACAG TCCTCAAAAACACAA TGTCCTTGGACGTAC AAAGTGTACCTGTGT

676 TCACATGACCGAGCC AAATAGATTTATCTT TACTCT
    AGTGTACTGGCTCGG TTTATCTAAATAGAA ATGAGA
```

(SEQ ID NOS:11-12)

Figure 4A

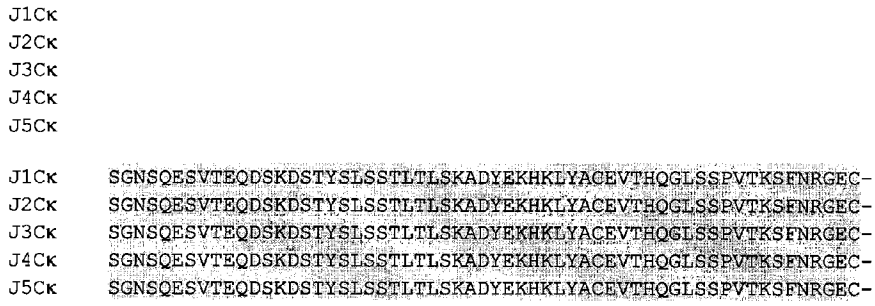
(SEQ ID NOS:13-24)

Figure 4B

```
      V  R  R  V  F  V  Q  Q  D  N  G  E  L  T  L  W  W  T  E  G
    1 GTGAGAAGGG TTTTTGTTCA GCAAGACAAT GGAGAGCTCA CACTGTGGTG GACGTTCGGC
      CACTCTTCCC AAAAACAAGT CGTTCTGTTA CCTCTCGAGT GTGACACCAC CTGCAAGCCG

      Q  G  T  K  V  E  I  K  R  T  V  A  A  P  S  V  F  I  F  P
   61 CAAGGGACCA AGGTGGAAAT CAAACGAACT GTGGCTGCAC CATCTGTCTT CATCTTCCCG
      GTTCCCTGGT TCCACCTTTA GTTTGCTTGA CACCGACGTG GTAGACAGAA GTAGAAGGGC

      P  S  D  E  Q  L  K  S  G  T  A  S  V  V  C  L  L  N  N  F
  121 CCATCTGATG AGCAGTTGAA ATCTGGAACT GCCTCTGTTG TGTGCCTGCT GAATAACTTC
      GGTAGACTAC TCGTCAACTT TAGACCTTGA CGGAGACAAC ACACGGACGA CTTATTGAAG

      Y  P  R  E  A  K  V  Q  W  K  V  D  N  A  L  Q  S  G  N  S
  181 TATCCCAGAG AGGCCAAAGT ACAGTGGAAG GTGGATAACG CCCTCCAATC GGGTAACTCC
      ATAGGGTCTC TCCGGTTTCA TGTCACCTTC CACCTATTGC GGGAGGTTAG CCCATTGAGG

      Q  E  S  V  T  E  Q  D  S  K  D  S  T  Y  S  L  S  S  T  L
  241 CAGGAGAGTG TCACAGAGCA GGACAGCAAG GACAGCACCT ACAGCCTCAG CAGCACCCTG
      GTCCTCTCAC AGTGTCTCGT CCTGTCGTTC CTGTCGTGGA TGTCGGAGTC GTCGTGGGAC

      T  L  S  K  A  D  Y  E  K  H  K  L  Y  A  C  E  V  T  H  Q
  301 ACGCTGAGCA AAGCAGACTA CGAGAAACAC AAACTCTACG CCTGCGAAGT CACCCATCAG
      TGCGACTCGT TTCGTCTGAT GCTCTTTGTG TTTGAGATGC GGACGCTTCA GTGGGTAGTC

      G  L  S  S  P  V  T  K  S  F  N  R  G  E  C  *
  361 GGCCTGAGCT CGCCCGTCAC AAAGAGCTTC AACAGGGGAG AGTGTTAG
      CCGGACTCGA GCGGGCAGTG TTTCTCGAAG TTGTCCCCTC TCACAATC
```

(SEQ ID NOS:25-26)

Figure 5A



(SEQ ID NOS:27-31)

Figure 5B

METDTLLLWVLLLWVPGSTGVRRVFVQQDNGELTLWWTFGQGTKVEIKRTVAAPSVFI
FPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQGNSQESVTEQDSKDSTYSLSS
TLTLSKADYEKHKLYACEVTHQGLSSPVTKSFNRGEC

METDTLLLWVLLLWVPGSTGWTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVC
LLNNFYPREAKVQWKVDNALQGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKLYA
CEVTHQGLSSPVTKSFNRGEC

METDTLLLWVLLLWVPGSTGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYP
REAKVQWKVDNALQGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKLYACEVTHQG
LSSPVTKSFNRGEC

(SEQ ID NOS:32-34)

Figure 5C

Figure 6

Surrogate light chain and fusion constructs
VpreB
Lambda 5
Surrogate light chain
• Two polypeptides
21
121
87
209
30
146
Surrogate light chain fusions
• Fusion at CDR3 analagous site
• Optionally fuse VpreB to CDR3 of lambda light chain
21
121
87
209
21
121
Classic recombined light chain
• Single protein derived from V-J joining
CDRs
Variable light
Constant light

Figure 7

Surrogate light chain deletion and single chain constructs

Can be used individually or with another protein such as a heavy chain

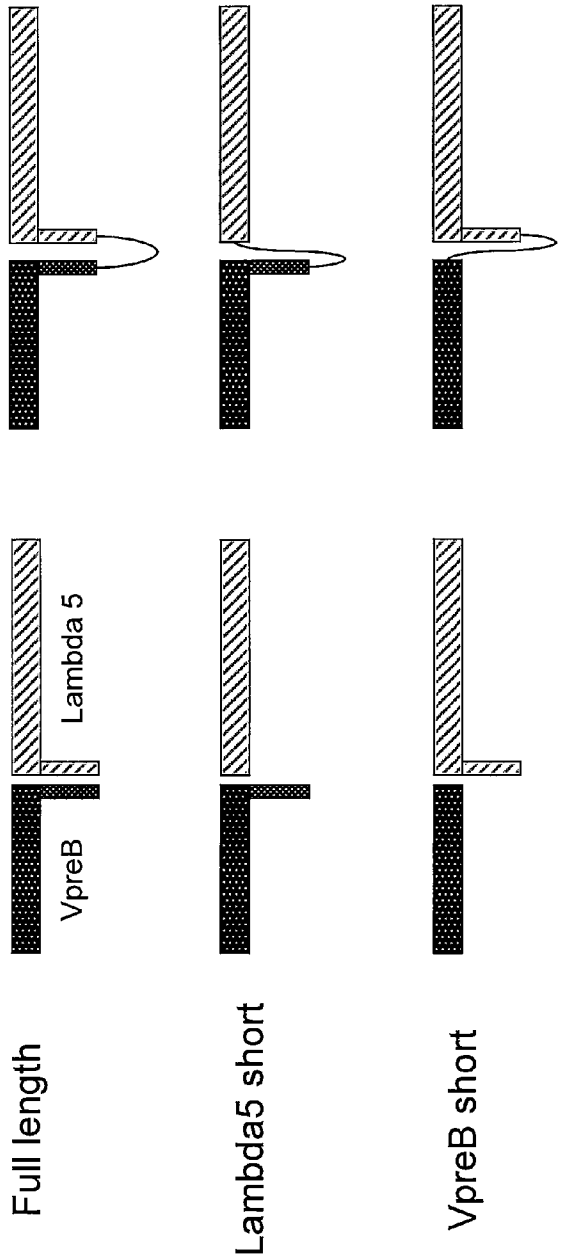


Figure 8

Incorporating combinatorial functional diversity into Surrogate Light Chain constructs

Note: Red lines indicate appended diversity such as a peptide library

Figure 9

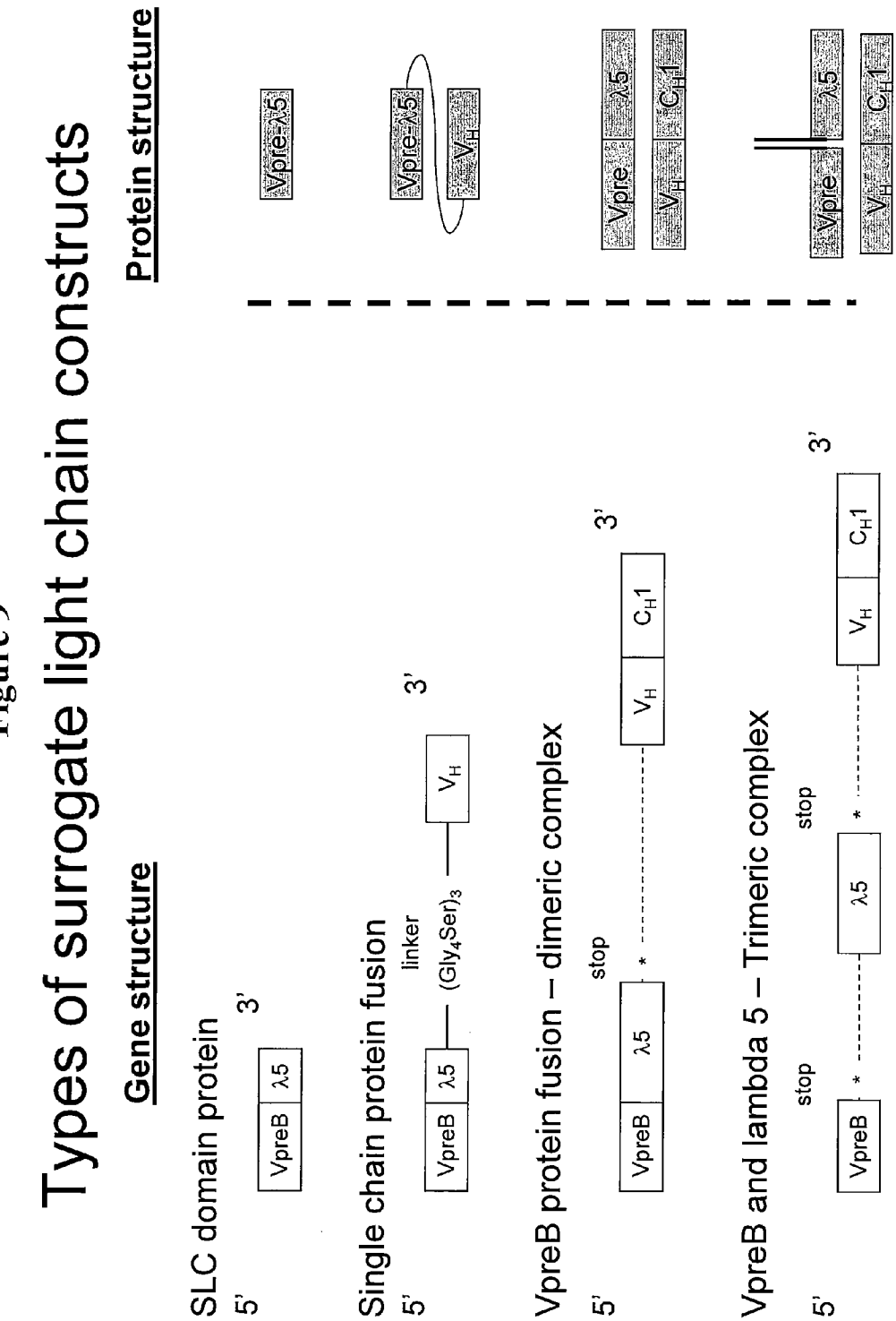

Types of surrogate light chain constructs
Gene structure
Protein structure
SLC domain protein
5'
3'
VpreB
λ5
Vpre-λ5
Single chain protein fusion
linker
(Gly4Ser)3
VH
VpreB protein fusion – dimeric complex
stop
*
CH1
Vpre
VpreB and lambda 5 – Trimeric complex

Figure 10

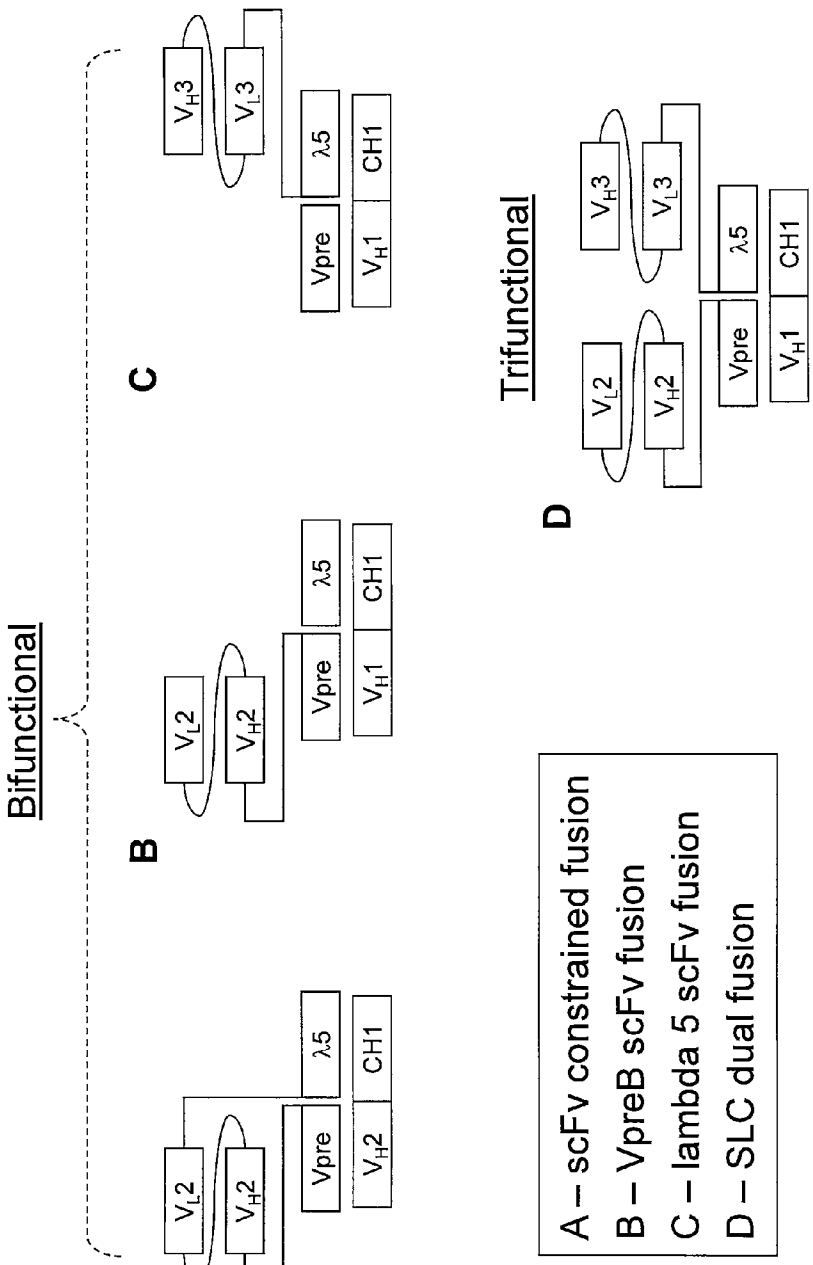
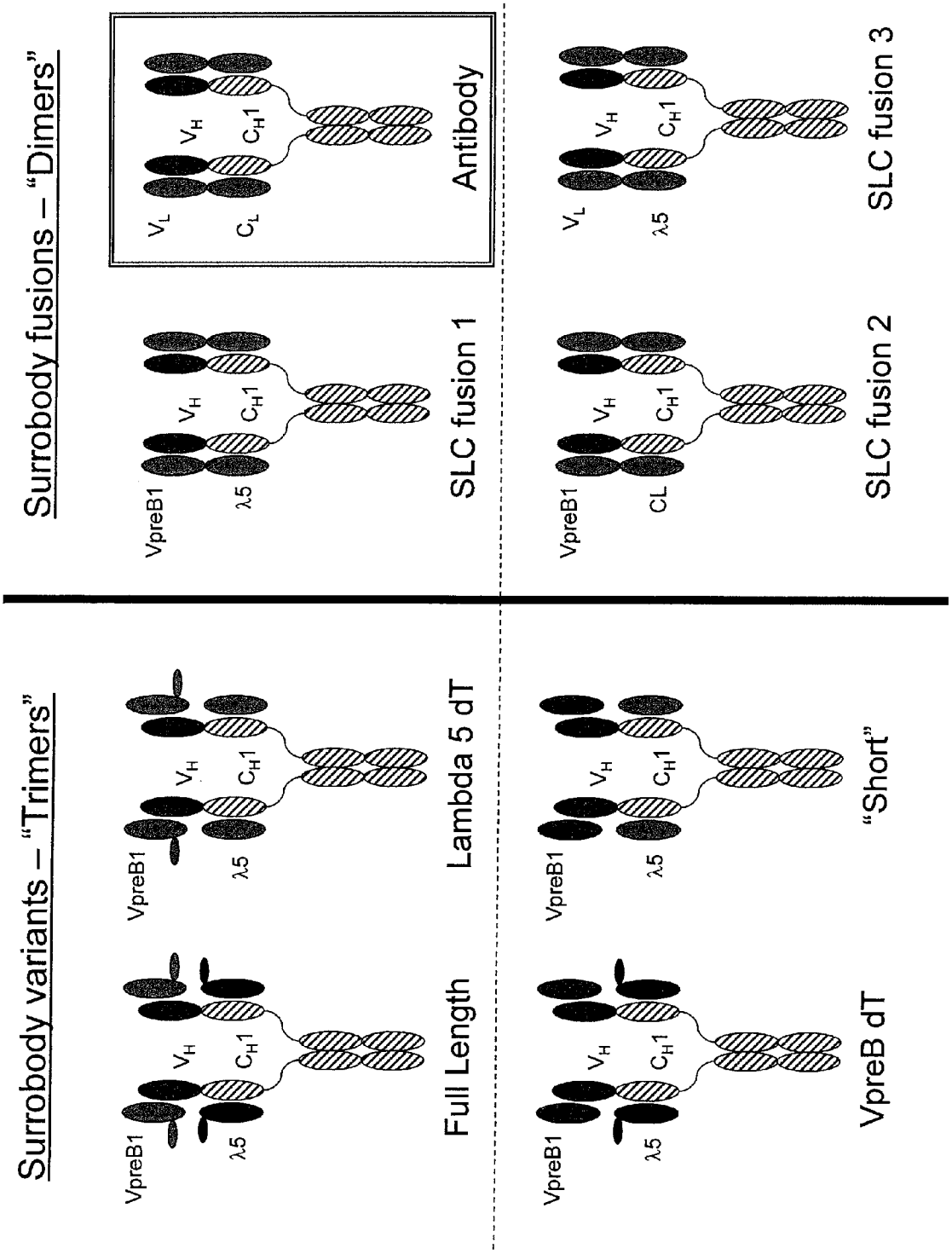

Adding functionality to SLC components
Bifunctional
A
B
C
D
Trifunctional
A – scFv constrained fusion
B – VpreB scFv fusion
C – lambda 5 scFv fusion
D – SLC dual fusion
VL2
VH2
Vpre
λ5
CH1
VH1
VH3
VL3

Figure 11


Surrobody variants – "Trimers"
Surrobody fusions – "Dimers"
VpreB1
λ5
VH
CH1
VL
CL
Full Length
Lambda 5 dT
VpreB dT
"Short"
SLC fusion 1
SLC fusion 2
SLC fusion 3
Antibody

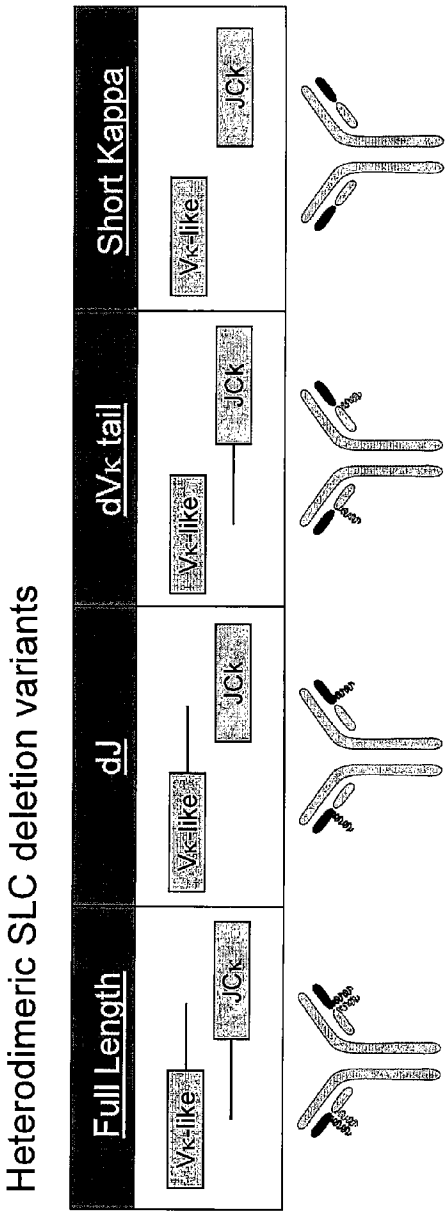

Kappa Surrobody variants
Heterodimeric SLC deletion variants
Full Length
dJ
dVκ tail
Short Kappa
Vκ-like
JCκ
Vκ-like
JCκ
Vκ-like
JCκ
Vκ-like
JCκ

FIG. 12

Kappa-like light chain deletion and single chain constructs

Can be used individually or with another protein such as a heavy chain

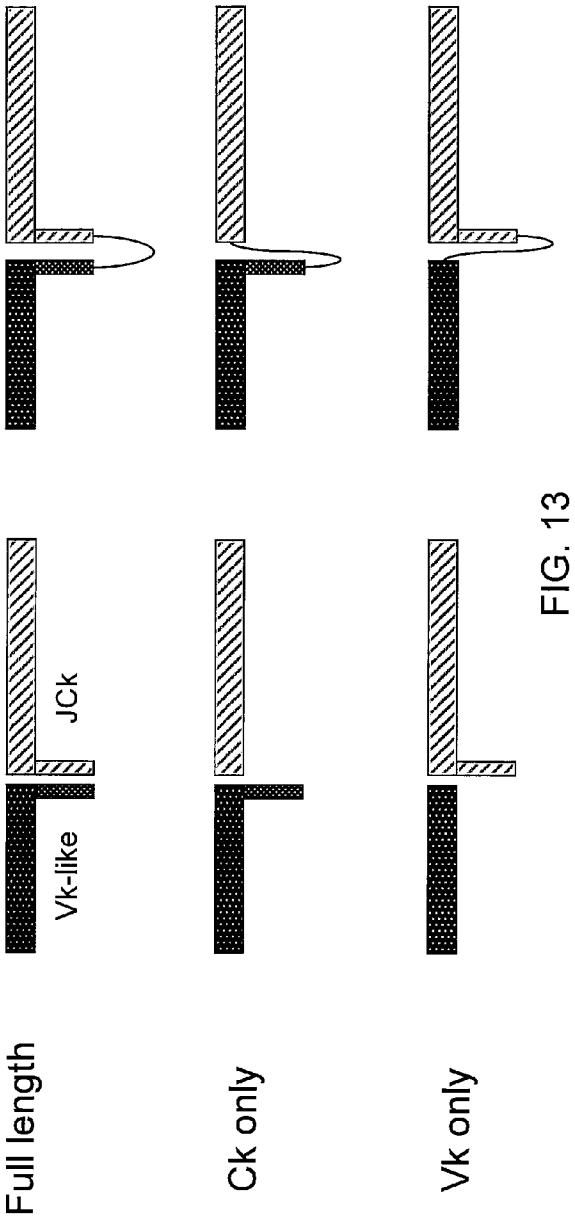
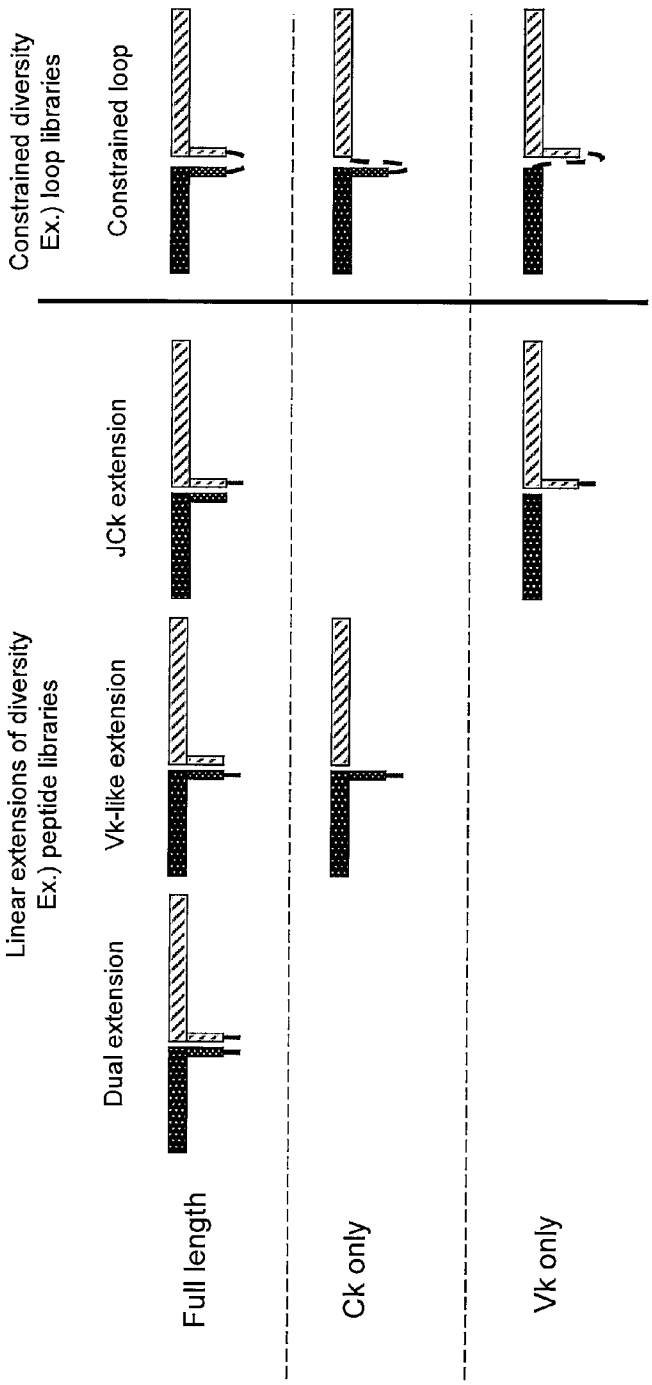


FIG. 13

Incorporating combinatorial functional diversity into Kappa-like SLC constructs



Note: Red lines indicate appended diversity such as a peptide library

FIG. 14

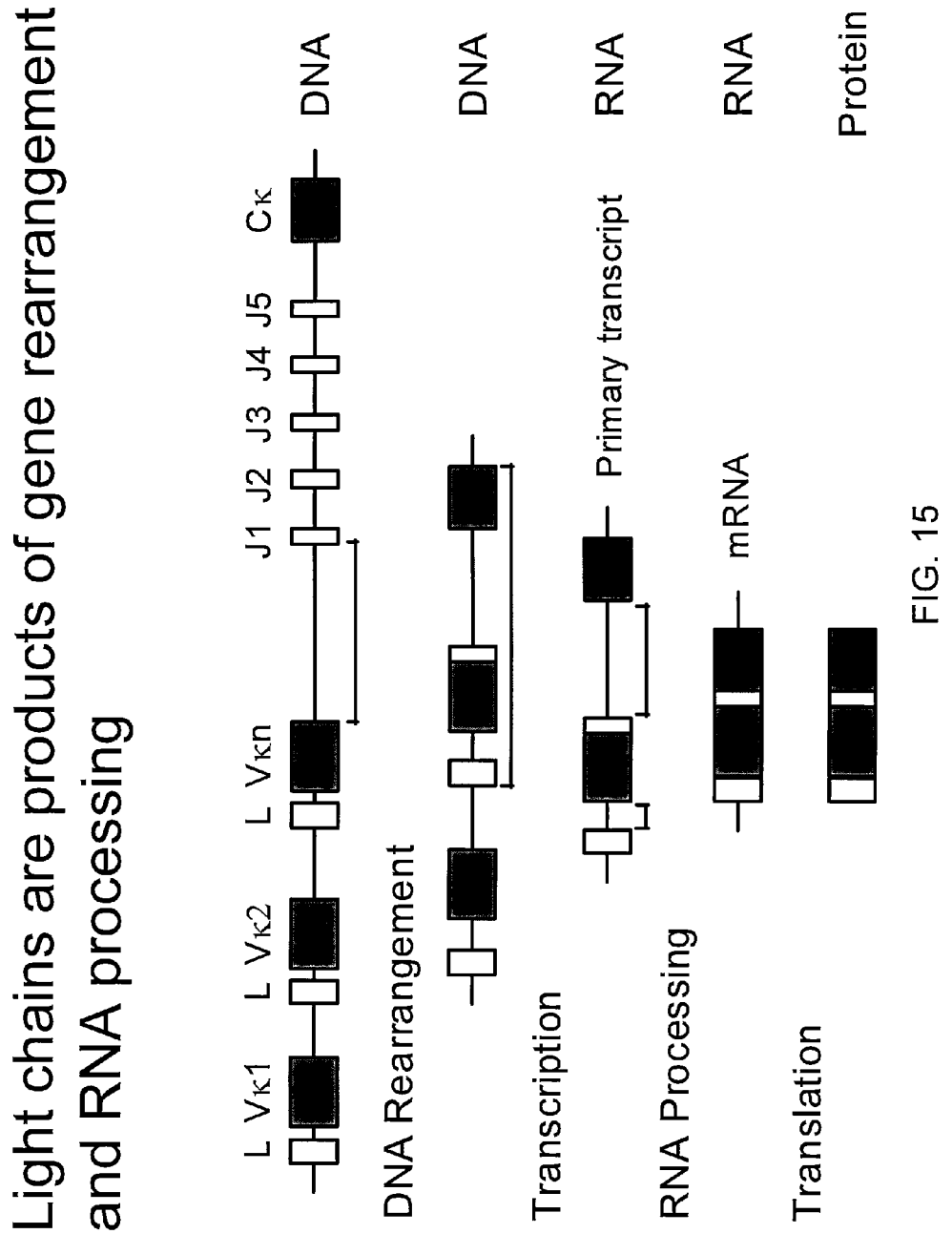

Light chains are products of gene rearrangement and RNA processing
L Vκ1
L Vκ2
L Vκn
J1 J2 J3 J4 J5
Cκ
DNA
DNA Rearrangement
DNA
Transcription
Primary transcript
RNA
RNA Processing
mRNA
RNA
Translation
Protein

FIG. 15

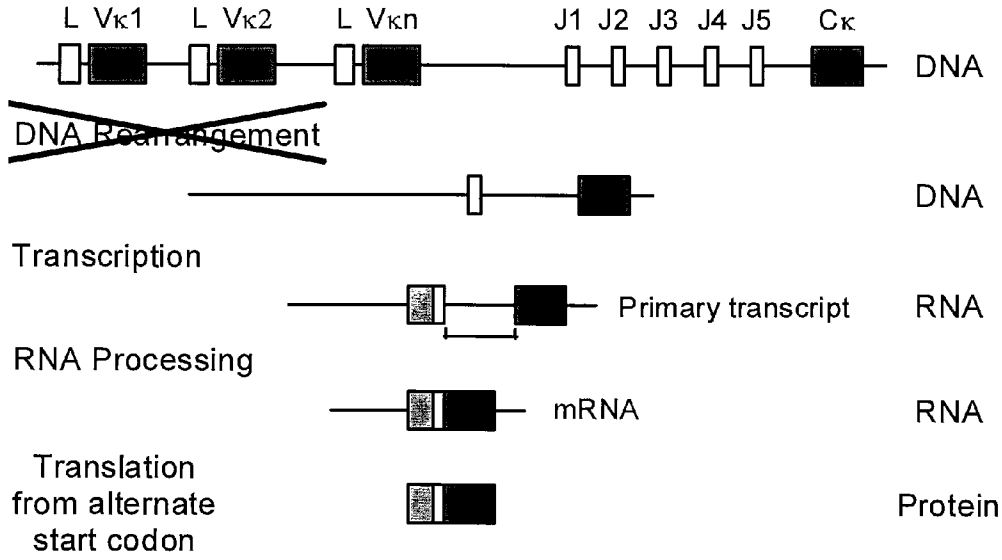
Vκ-like protein is derived from unrearranged VκIV-gene transcription and translation

VκIV is one of seventy-one VL germline genes

- There are an additional 70 VL germline genes capable of creating Vκ-like proteins
  - 39 more kappa V genes
  - 31 more lambda V genes

FIG. 16A

JCk is a product of processed RNA from unrearranged J and C germlines



JCκ is one of forty-five JC germline combinations

- There are an additional 44 VL germline genes capable of creating JCκ-like proteins
  - 4 more Jκ genes to combine with Cκ
  - 4 Jλ genes to combine with 10 Cλ genes (40 total)

FIG. 16B

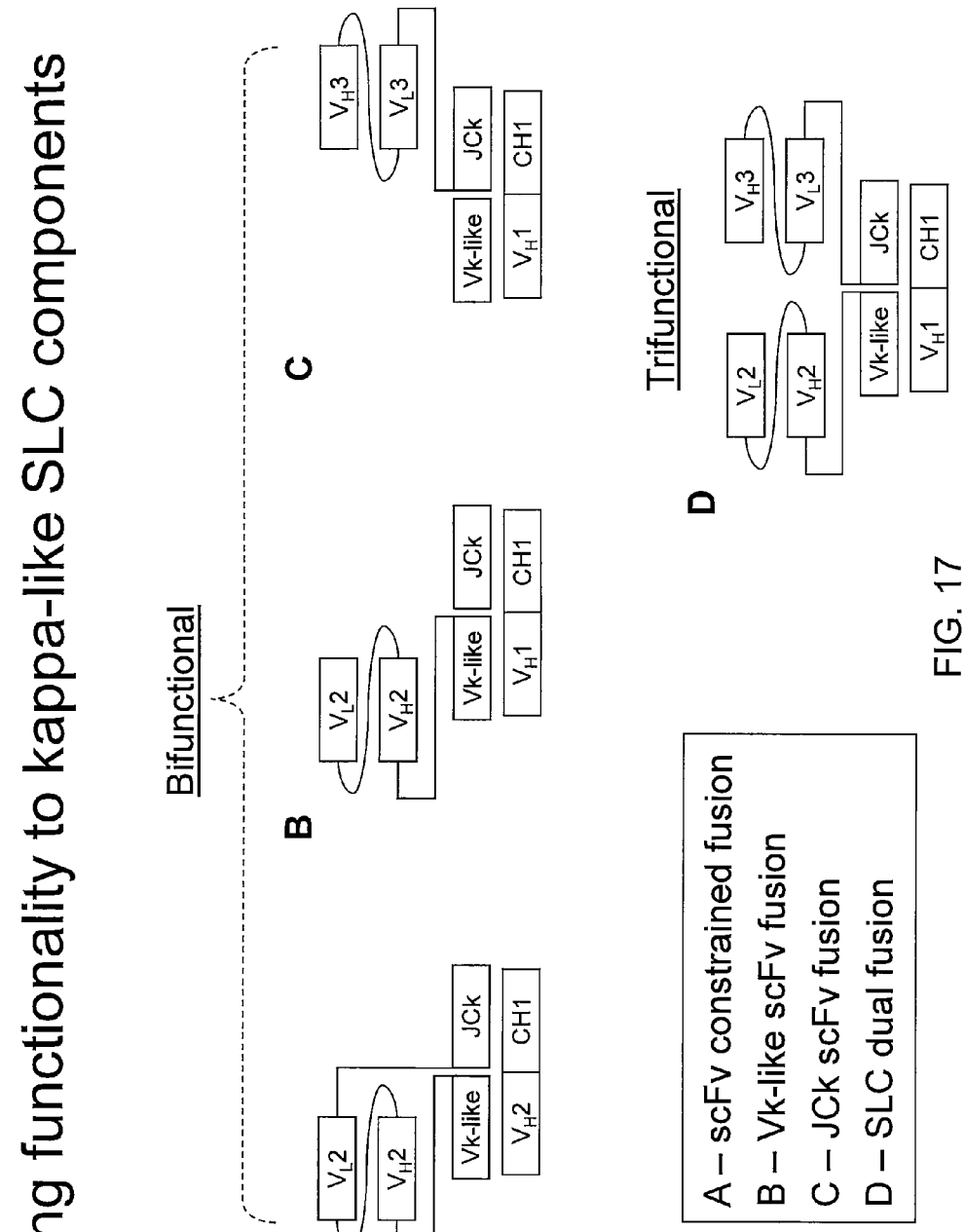

Adding functionality to kappa-like SLC components
Bifunctional
A
B
C
D
Trifunctional
VL2
VH2
Vk-like
JCk
VH1
CH1
VH3
VL3
A – scFv constrained fusion
B – Vk-like scFv fusion
C – JCk scFv fusion
D – SLC dual fusion

FIG. 17

Types SLC functional tail extensions

SLC Tail fusions	Additional or Enhanced Function
Metal coordinating peptide	Enhanced binding
Growth Factor or Cytokine	Secondary activity or singly to create proteins with Enhanced PK
Target-based (dominant negative whole protiens or fragments)	Guide binding to disrupt protein-protein interactions
Proteases	Targeted processive proteolysis
Glycopeptides	Glycospecific interactions and/or PK enhancement
Membrane interactive peptides	Cell surface anchoring or cellular translocation

FIG. 18

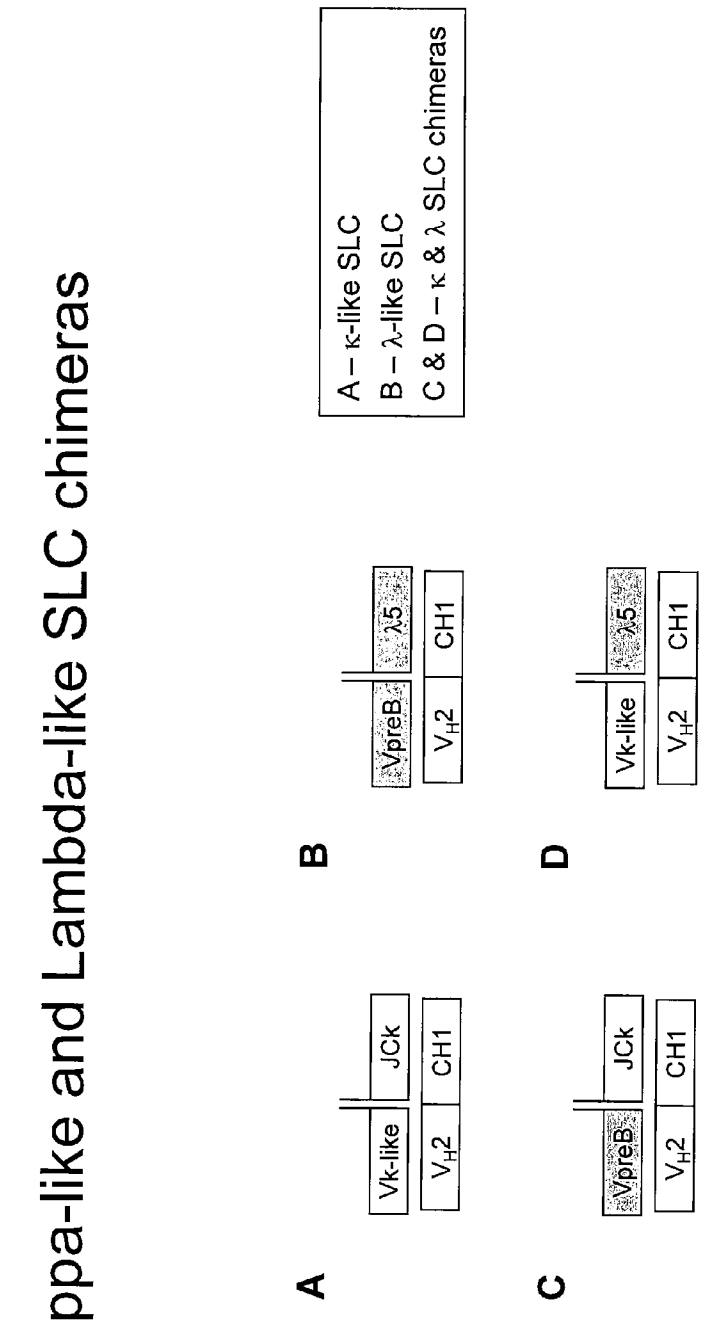

Kappa-like and Lambda-like SLC chimeras
A – κ-like SLC
B – λ-like SLC
C & D – κ & λ SLC chimeras
A
Vk-like
JCk
VH2
CH1
B
VpreB
λ5
VH2
CH1
C
VpreB
JCk
VH2
CH1
D
Vk-like
λ5
VH2
CH1

FIG. 19

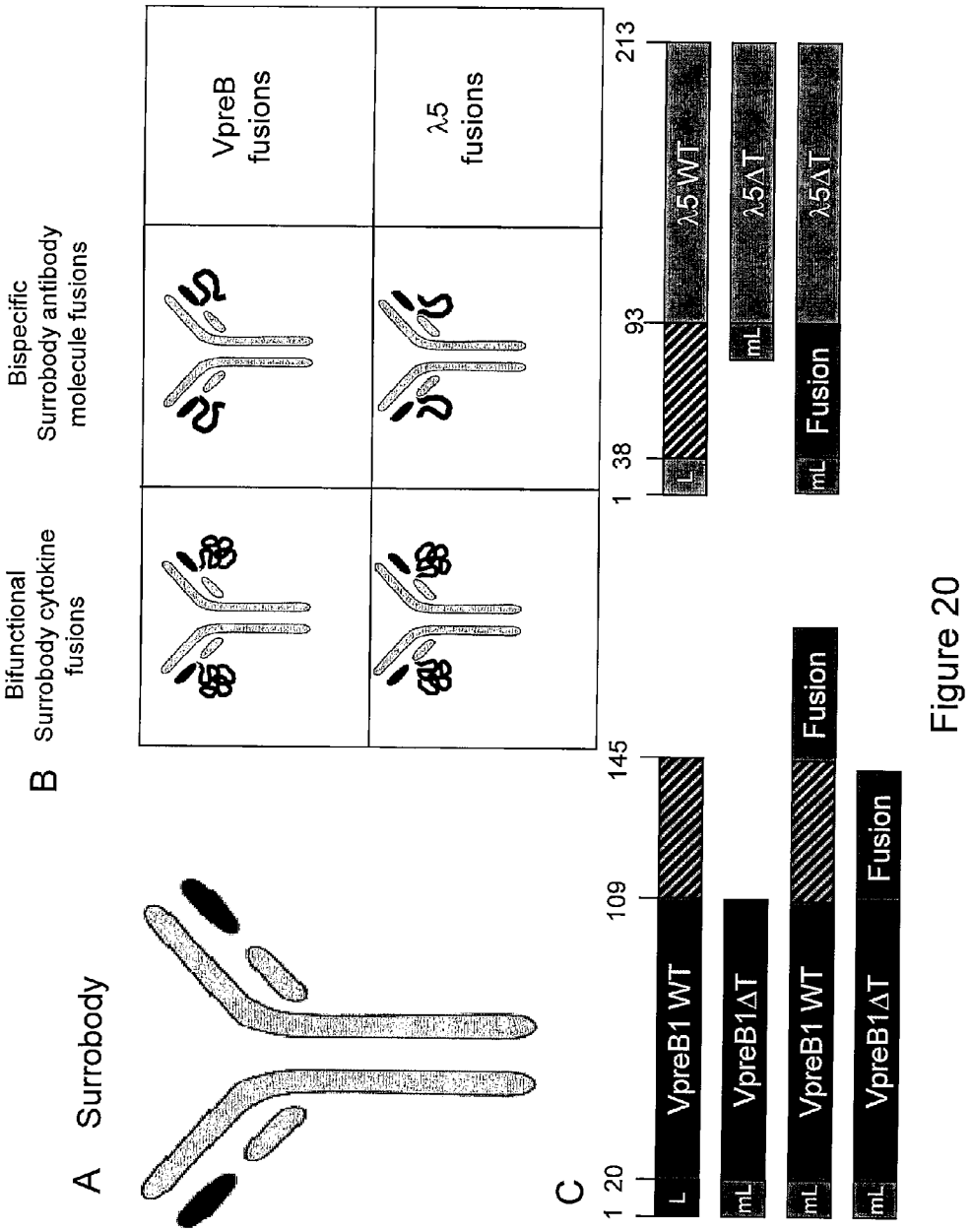

A Surrobody
B
Bifunctional Surrobody cytokine fusions
Bispecific Surrobody antibody molecule fusions
VpreB fusions
λ5 fusions
C
1 20
109
145
L
VpreB1 WT
mL
VpreB1ΔT
mL
VpreB1 WT
Fusion
mL
VpreB1ΔT
Fusion
1 38
93
213
L
λ5 WT
mL
λ5ΔT
mL
Fusion
λ5ΔT

Figure 20

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/040052

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/118970 A2 (SEA LANE BIOTECHNOLOGIES LLC [US]; BHATT RAMESH [US]; HOROWITZ LAWRENC) 2 October 2008 (2008-10-02)	1-16, 19-29
Y	page 31, lines 11-16; example 7	17,18
Y	XU LI ET AL: "Combinatorial surrobody libraries" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS), NATIONAL ACADEMY OF SCIENCE, US LNKD-DOI:10.1073/PNAS.0805293105, vol. 105, no. 31, 5 August 2008 (2008-08-05), pages 10756-10761, XP002498064 ISSN: 0027-8424 the whole document	1-29

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 August 2010	11/10/2010

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

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/040052

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KAGANMAN I: "A surrogate scaffold tested" NATURE METHODS 2008 GB LNKD- DOI:10.1038/NMETH1008-861, vol. 5, no. 10, 2008, page 861, XP002597770 the whole document	1-29
A	MOLHOJ MICHAEL ET AL: "Leader sequences are not signal peptides." NATURE BIOTECHNOLOGY DEC 2004 LNKD- PUBMED:15583649, vol. 22, no. 12, December 2004 (2004-12), page 1502, XP002597940 ISSN: 1087-0156 the whole document	1-29
X,P	XU L ET AL: "Surrobodies with Functional Tails" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD- DOI:10.1016/J.JMB.2010.01.036, vol. 397, no. 1, 19 March 2010 (2010-03-19), pages 352-360, XP026929320 ISSN: 0022-2836 [retrieved on 2010-01-25] the whole document	1-29

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/040052

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purpose of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2010/040052

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008118970 A2	02-10-2008	AU 2008230795 A1	02-10-2008
		CA 2680237 A1	02-10-2008
		CN 101679974 A	24-03-2010
		EP 2132312 A2	16-12-2009
		JP 2010522566 T	08-07-2010
		KR 20100015902 A	12-02-2010
		US 2010004139 A1	07-01-2010