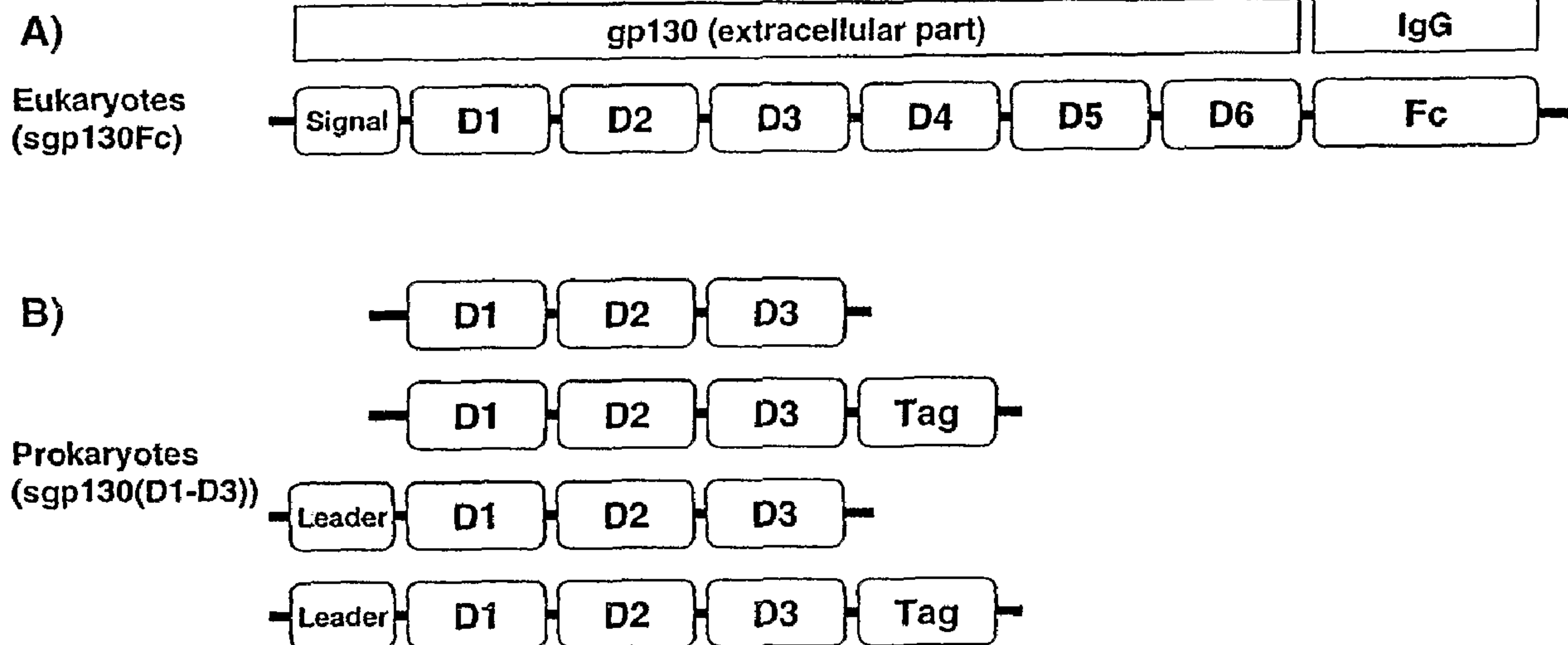




(86) Date de dépôt PCT/PCT Filing Date: 2005/08/26  
(87) Date publication PCT/PCT Publication Date: 2006/03/02  
(85) Entrée phase nationale/National Entry: 2007/02/01  
(86) N° demande PCT/PCT Application No.: EP 2005/009247  
(87) N° publication PCT/PCT Publication No.: 2006/021453  
(30) Priorité/Priority: 2004/08/27 (EP04020455.4)

(51) Cl.Int./Int.Cl. *C12N 15/67* (2006.01),  
*C07K 14/715* (2006.01)  
(71) Demandeur/Applicant:  
CONARIS RESEARCH INSTITUTE AG, DE  
(72) Inventeurs/Inventors:  
SEEGERT, DIRK, DE;  
WAETZIG, GEORG, DE;  
RAHAUS, NIKOLAUS, DE;  
DAECKE, JESSICA, DE;  
ROSE-JOHN, STEFAN, DE  
(74) Agent: SMART & BIGGAR

(54) Titre : SEQUENCES NUCLEOTIDIQUES OPTIMISEES CODANT SGP 130  
(54) Title: OPTIMIZED NUCLEOTIDE SEQUENCES ENCODING SGP 130



(57) **Abrégé/Abstract:**

Described are codon optimized sgpl30 encoding nucleic acid molecules as well as a method for the highly efficient recombinant production of sgpl30 in mammalian cells or bacteria using a nucleic acid molecule of the invention.



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

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 March 2006 (02.03.2006)

PCT

(10) International Publication Number  
**WO 2006/021453 A3**

## (51) International Patent Classification:

C12N 15/67 (2006.01) C07K 14/715 (2006.01)

## (21) International Application Number:

PCT/EP2005/009247

## (22) International Filing Date: 26 August 2005 (26.08.2005)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

04020455.4 27 August 2004 (27.08.2004) EP

## (71) Applicant (for all designated States except US):

**CONARIS RESEARCH INSTITUTE AG** [DE/DE];  
Schauenburgerstrasse 116, 24118 Kiel (DE).

## (72) Inventors; and

(75) **Inventors/Applicants (for US only):** **SEEGERT, Dirk** [DE/DE]; Freesenberg 148, 24161 Altenholz (DE). **WÄTZIG, Georg** [DE/DE]; Heiderstrasse 19, 24106 Kiel (DE). **RAHAUS, Nikolaus** [DE/DE]; Brammerkamp 12, 24159 Kiel (DE). **DAECKE, Jessica** [DE/DE]; Jungmannstrasse, 24118 Kiel (DE). **ROSE-JOHN, Stefan** [DE/DE]; Theodor-Heuss-Weg 6, 24211 Schellhorn (DE).

(74) **Agent:** **SCHÜSSLER, Andrea**; Huber & Schüssler, Truderinger Strasse 246, 81825 Munich (DE).

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

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

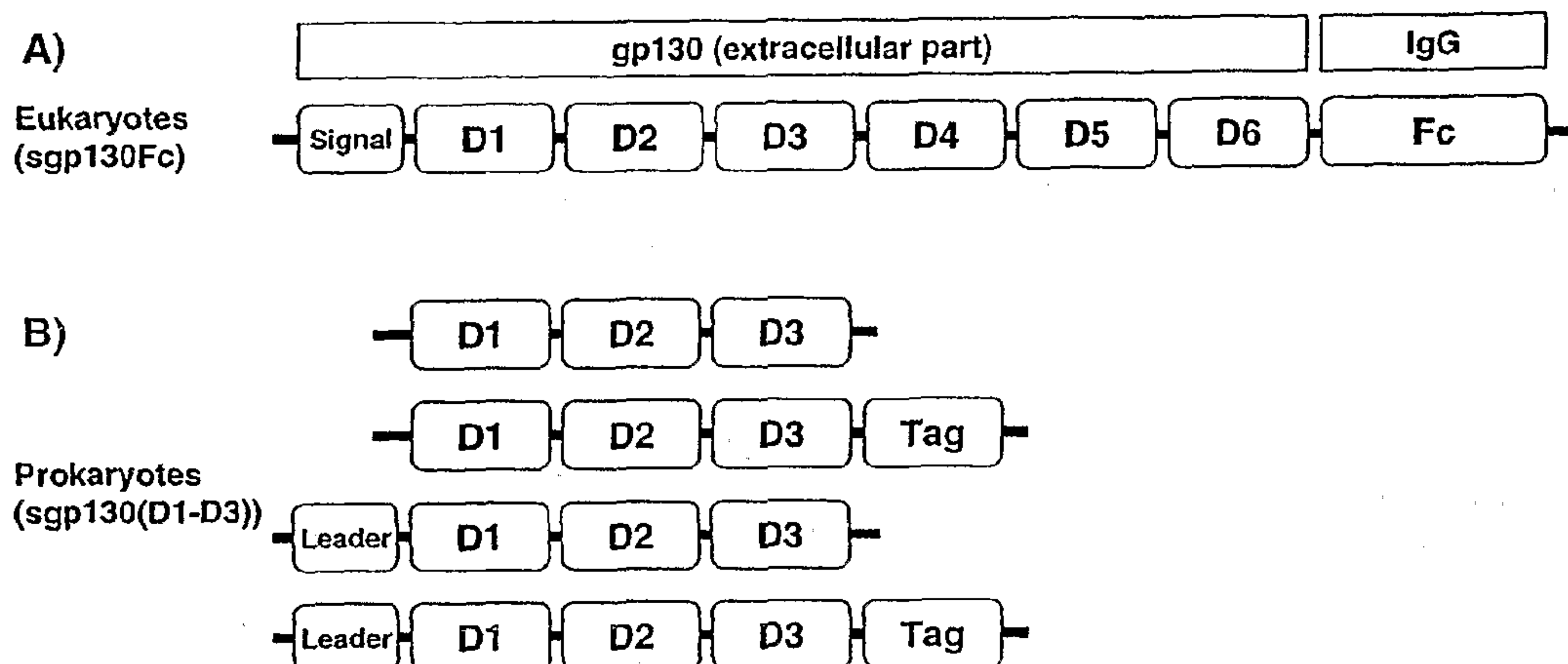
**Published:**

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

(88) **Date of publication of the international search report:**  
22 June 2006

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

(54) **Title:** OPTIMIZED NUCLEOTIDE SEQUENCES ENCODING SGP 130



(57) **Abstract:** Described are codon optimized sgp130 encoding nucleic acid molecules as well as a method for the highly efficient recombinant production of sgp130 in mammalian cells or bacteria using a nucleic acid molecule of the invention.

WO 2006/021453 A3

### Optimized nucleotide sequences encoding sgp130

The present invention relates to codon optimized sgp130 encoding nucleic acid molecules as well as a method for the highly efficient recombinant production of sgp130 in mammalian cells or bacteria using a nucleic acid molecule of the invention.

For the treatment of various diseases such as Crohn's disease etc. the specific blocking of IL-6 responses dependent on soluble IL-6R might be desirable for treatment. It was found that a soluble gp130-dimer, in particular an IgG-Fc fusion protein or a PEGylated version of sgp130, efficiently inhibits the anti-apoptotic effect of sIL-6R from LPMC from Crohn's disease (CD) patients and that, thus, said compound is useful for the treatment of said disease and related diseases like, e.g., colitis or rheumatoid arthritis. Unfortunately, so far the recombinant production of sgp130 is difficult in particular due to the fact that only low amounts of protein can be obtained

Thus, the technical problem underlying the present invention was to provide means allowing to improve the efficiency of recombinant production of sgp130Fc or sgp130(D1-D3).

The solution of the said technical problem is achieved by providing the embodiments characterized in the claims. During the experiments leading to the present invention it was found that by use of particular codon optimized versions of the DNA encoding sgp130Fc the yields of recombinant protein can be increased at least 10- to 20-fold compared to the unmodified version of the DNA. In case of the prokaryotic sgp130(D1-D39) version, the optimization of the DNA led to the reduction of undesired shorter side products.

### Brief description of the drawings

#### Figure 1: Schematic presentation of the constructs

Grey shading marks the parts of the protein which have been optimized. (A) Eukaryotic construct comprising a signal peptide, six extracellular gp130 domains and the IgG-Fc part. (B) Variations of the sgp130 protein expressed in prokaryotic cells. sgp130 can be expressed with or without N-terminal leader sequence and/or C-terminal Tag for purification purposes.

#### Figure 2: sgp130(D1-D3) (nucleotide sequence and amino acid sequence) for expression in bacterial cells

An alignment of the nucleotide sequence with optimized codons (sgp130(D1-D3)\_opt) vs. the original sequence (sgp130(D1-D3)\_wt) is shown.

#### Figure 3: sgp130(Fc) (nucleotide sequence and amino acid sequence) for expression in mammalian cells

An alignment of the nucleotide sequence with optimized codons (sgp130Fc\_opt) vs. the original sequence (sgp130Fc\_wt) is shown.

#### Figure 4: Detection of sgp130Fc after transient transfection of HEK293 cells (A) or CHO cells (B) with wildtype or optimized (opt) sgp130Fc expression plasmids

The position of sgp130Fc is indicated by arrows (◄). Wildtype and optimized sgp130Fc expression was detected in two independent transfection experiments each. The different sizes of the protein (left panel) result from the leader sequence which has been partially cleaved off after secretion into the medium. The right panel represents the results derived from

whole cell extracts from CHO cells (sgp130Fc with leader sequence).

Figure 5: Detection of RNA transcribed from transfected plasmid DNA (sgp130Fc; neomycin resistance gene (NeoR)) by gene-specific RT-PCR

HEK293 cells were transfected with an expression plasmid encoding either wildtype or optimized sgp130Fc. Transfection of the empty vector (mock) or non-transfected cells (control) served as negative controls.  $\beta$ -actin was amplified from total RNA to demonstrate the use of equal amounts of RNA in each single experiment.

Figure 6: Expression of sgp130(D1-D3) in BL21(DE3)pLys bacteria

The cDNA encoding sgp130(D1-D3) was cloned into the expression plasmid pET22b (Invitrogen, Carlsbad, CA, USA) which in addition encodes a leading pelB sequence and a C-terminal His-tag. sgp130(D1-D3) was detected by western blot with a His-specific antibody and marked with an arrow ( $\blacktriangleleft$ ).

Figure 7: Comparison and selection of different optimized sgp130Fc sequences

The wt cDNA encoding sgp130Fc was optimized using different optimization algorithms ("Java Codon Adaption Tool" (JCat, available at <http://www.prodoric.de/JCat/> and "UpGene" Gao et al. (2004), Biotechnol. Proc. 20(2): 443-8) and the resulting expression constructs were transfected into HEK293 cells. The expression of sgp130Fc was determined on RNA level by RT-PCR and on protein level by western blot as described earlier. The figure represents an example of the comparison of two different optimized sgp130Fc sequences (Opt1 and Opt2) with a sgp130Fc sequence optimized according to the present invention (Opt CONARIS).

Figure 8: Alignment of different optimized sgp130Fc cDNA sequences and comparison with the unmodified wild type (wt) sequence

The shaded regions are showing a few examples of codons which either have been left unmodified (grey) as compared to the wild type (wt) or have been changed (partly in different ways) by the algorithms.

Figure 9: The table indicates the numbers of base pairs which are different between two compared sequences

sgp130Fc sequence optimized according to the present invention (opt\_CON), JCat (opt\_JCat) or UpGene (opt\_Upgene).

Thus, the present invention relates to a nucleic acid molecule encoding sgp130 comprising the nucleic acid sequence (a) as depicted in Figure 2 (sgp130(D1-3)\_opt) or Figure 3 (sgp130Fc\_opt) or (b) a fragment or analogue thereof which maintains the codon usage pattern thereof.

The letter "s" of sgp130 means "soluble". The term "soluble" as used herein refers to a gp130 molecule lacking the intracellular domain and the transmembrane domain. The domains utilized in sgp130(D1-D3)\_opt consist of the first three extracellular domains D1-D3 of gp130.

The term "fragment" as used herein refers to sgp130 fragments which comprise the entire or smaller parts of the optimized cDNA encoding the extracellular domain of gp130. Preferably, such fragments show a biological activity of the full length molecule, e. g. maintain the ability to inhibit the activity of the agonistic complex IL-6/sIL-6R. For the expression in bacteria such fragment also comprises sgp130 without the eukaryotic secretory leader sequence (MLTLQTWVVQALFIFLTTESTG,

Pos. 1 to 22). Moreover, a prokaryotic secretory leader sequence, e. g. *pelB* or *OmpA* could be cloned in front of the *sgp130* sequence or parts of it and could be derived from the respective suitable expression plasmid, e. g. pET22b (Merck Biosciences GmbH, Bad Soden, Germany), pASK-IBA2, pASK-IBA12 (IBA, Goettingen, Germany). In addition the *sgp130* protein can be expressed with or without Tag for purification purposes, e. g. His<sub>6</sub>, Strep, Myc or others.

The term "analogue" as used herein refers to a nucleic acid molecule which encodes the same amino acid sequence but which, through the redundancy of the genetic code, has a different nucleotide sequence. The term "codon usage pattern" as used herein refers to the average frequencies in the nucleotide sequence, e.g., highly expressed mammalian genes. Codon usage patterns for mammals, including humans can be found in the literature; see, e.g., Nakamura et al., Nucleic Acids Research 1996, 24:214-5. In the nucleic acid molecules of the present invention, the codon usage pattern is altered to more closely represent the codon bias of the host organism, e.g. a mammalian cell.

Alternatively, the present invention relates to a nucleic acid molecule, wherein at least 80%, preferably at least 90%, more preferably at least 95% and, most preferably at least 98% of the codons altered in the nucleic acid sequence of Figure 2 or 3 vs. the wild type sequence are present.

In a preferred embodiment, the nucleic acid molecule of the present invention is a DNA molecule.

The present invention includes expression vectors that comprise the nucleic acid molecules of the invention. The expression vectors can be constructed according to methods well known to the person skilled in the art; see, e.g.,

Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. The "control elements" or "regulatory sequences" used for recombinant expression are those non-translated regions of the vector-enhancers, promoters, 5' and 3' untranslated regions which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which expression is designed. For example, mammalian promoters include the metallothionein promoter, which can be induced in response to heavy metals such as cadmium, and the  $\beta$ -actin 30 promoter. Viral promoters such as the SV40 large T antigen promoter, human cytomegalovirus (CMV) immediate early (IE) promoter, rous sarcoma virus LTR promoter, adenovirus promoter, or a HPV promoter, particularly the HPV upstream regulatory region (URR) may also be used. All these promoters are well described and readily available in the art.

In mammalian host cells, a number of viral-based expression systems may be utilised. In cases where an adenovirus is used as an expression vector, sequences encoding the polypeptide(s) of the present invention may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the antibody in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV)

enhancer, may be used to increase expression in mammalian host cells.

Further examples of suitable viral vectors, include herpes simplex viral vectors, vaccinia or alpha-virus vectors and retroviruses, including lentiviruses and adeno-associated viruses. Gene transfer techniques using these viruses are known to those skilled in the art. Retrovirus vectors for example may be used to stably integrate the nucleic acid molecules of the invention into the host genome, although such recombination is not preferred. Replication-defective adenovirus vectors by contrast remain episomal and therefore allow transient expression. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding the spg130 polypeptides, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10 M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the spg130, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in case where only a coding sequence for a fragment is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert.

Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed polypeptide chains in the desired fashion. Post-translational processing which cleaves a "prepro" form of the polypeptide may also be used to facilitate correct insertion, folding and/or function. Different mammalian host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, 293, COS-7 and W138), are available from the American Type Culture Collection (ATCC; Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign polypeptide chains.

For long-term, high-yield production of sgp130, stable expression in mammalian cells is preferred. For example, cell lines which stably express sgp130Fc may be transfected using expression vectors which may contain viral origins of replication and/or endogenous expression elements and one or more selectable marker genes on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

After the introduction of the recombinant vector(s), the host cells are grown in a selective medium, which selects for the growth of vector-containing cells. Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilise indole in place of tryptophan, or hisD, which allows cells to utilise histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) Methods Mol. Biol. 55:121-131).

The person skilled also knows vectors and host cells for bacterial expression, e.g. bacteriophage, plasmid, or cosmid DNA expression vectors. Vectors suitable for use in the present invention include, but are not limited to the pSKK expression vector for expression in bacteria. Depending on the

vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript.RTM. phagemid (Stratagene, LaJolla, Calif.) or pSport1.TM. plasmid (Gibco BRL) and the like may be used.

Purification of the recombinant sgp130 is carried out by any one of the methods known for this purpose, i.e., any conventional procedure involving extraction, precipitation, chromatography, electrophoresis, or the like. A further purification procedure that may be used is affinity chromatography using monoclonal antibodies which bind the target polypeptide or a Tag fused to it, e.g., His, Strep or Myc, and which are produced and immobilized on a gel matrix contained within a column. Impure preparations containing the recombinant sgp130 are passed through the column. The sgp130 will be bound to the column by the specific antibody while the impurities will pass through. After washing the polypeptide is eluted from the gel by a change in pH or ionic strength and can then, if desired, dimerized and/or PEGylated.

Accordingly, the present invention also relates to a method of producing the sgp130 of the present invention, comprising culturing a host cell transformed with a nucleic acid molecule of the invention and recovering the sgp130 polypeptide from said host cell or the culture.

The sgp130 polypeptide produced from a nucleic acid molecule of the present invention is useful in the treatment and/or prevention of all the pathologies, in which the activity of the agonistic complex IL-6/sIL6R must be inhibited, e.g., for the treatment/prevention of bone resorption, hypercalcemia,

cachexia, tumours, autoimmune diseases such as Crohn's disease and bacterial or viral infections.

The below examples explain the invention in more detail.

### **Example 1: Material and Methods**

#### (A) Construct and transfection

The cDNA encoding either wildtype or optimized sgp130Fc was cloned into the expression plasmid pDEST40 (Invitrogen, Carlsbad, CA, USA) according to standard procedures. The wildtype sequence was derived from an expression plasmid which has been described in Jostock et al., Eur. J. Biochem. 268 (2001), 160-7 (Figure 1; upper panel). The construct was sequence verified.  $3 \times 10^5$  HEK293 cells were transiently transfected with 1  $\mu$ g of plasmid and 3  $\mu$ l of Fugene (Roche Diagnostics, Mannheim, Germany) in 3 ml of medium according to the manufacturer's manual. The cells were subsequently incubated for 24 h at 37°C and supernatants and cells were harvested for further preparations of either total proteins or RNA, respectively.

One set of cells was transfected with the empty vector (mock), another set of cells was left untransfected (control). Both sets served as negative controls.

#### (B) Protein extraction and western blot

The sgp130Fc protein was precipitated from the cell supernatants by adding 20  $\mu$ l of Protein-A/G-Plus Agarose (Santa Cruz, CA, USA). The slurry was incubated overnight at 4°C and finally centrifuged. Bound proteins were extracted by boiling the agarose pellet in SDS sample buffer for 5 minutes at 100°C. In parallel the cells were scraped from the plates

using a rubber policeman, harvested in 100  $\mu$ l of PBS and centrifuged. All protein samples were separated on a standard acrylamide gel, transferred to a PDVF-membrane by semi-dry blotting and stained with a gp130-specific antibody (Hözel Diagnostika, Köln, Germany). 50 ng of recombinant sgp130 served as positive control (sgp130).

#### (C) RNA extraction

Total RNA was extracted from the cell pellets using a RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. The following primers were used to determine RNA transcribed from the transfected plasmid DNA by RT-PCR: sgp130Fc\_f: 5'-ATGAGGTGTGAGTGGGATGG-3'; sgp130Fc\_r: 5'-ACCTTGCACTTGTACTCCTTGC-3'; neomycin resistance gene NeoR\_s: 5'-GATGCCTGCTTGCCGAATATC-3'; NeoR\_r: 5'-CGCCAAGCTCTTCAGCAATATC-3'. Total RNA was initially reverse transcribed and the cDNA was amplified by 30 cycles of 30 seconds at 95°C followed by 2 minutes at 57°C and a final elongation step of 5 minutes at 72°C. Expected amplicon sizes: gp130: 1.712 bp, NeoR: 133 bp. The amplification of NeoR was performed to document an equal transfection efficiency of the plasmid. In addition,  $\beta$ -actin was amplified to demonstrated the use of equal amounts of total RNA in each experiment.

#### **Example 2: Highly efficient recombinant production of sgp130Fc in HEK293 cells**

Figure 4 demonstrates that in comparison to the wildtype expression plasmid the production of sgp130Fc was increased at least 10 to 20-fold in HEK293 cells transfected with the optimized sgp130Fc expression plasmid. On the RNA level (Figure 5) a similar increase of sgp130Fc expression was detected with the optimized construct. This elevation of sgp130Fc RNA amounts was not due to a different transfection

efficiency as shown by equal amounts of RNA encoded by the neomycin resistance gene which was also located on the expression plasmid.

The results indicate that the significant increase of sgp130Fc production after optimization of the cDNA sequence is partially based on an improved codon usage during translation but is mainly derived from the elevation of the corresponding RNA levels. This might be due to a more efficient transcription or a higher stability of the RNA.

**Example 3: Highly efficient recombinant production of  
sgp130(D1-D3) in bacteria**

(A) Constructs and transformation

The cDNA encoding either wildtype or optimized sgp130(D1-D3) was cloned into the prokaryotic expression plasmid pET22b (Merck Biosciences GmbH, Bad Soden, Germany) according to standard procedures. The D1-D3 fragment was amplified afore from the pSVL-sgp130Fc plasmid described in Jostock et al. 2001. The construct was sequence verified and transformed into BL21(DE3)pLys bacteria (Invitrogen, Carlsbad, CA, USA).

(B) Protein expression and western blot

10 ml of bacterial suspension were diluted at 1:100 with LB-medium and grown at 30°C overnight until the OD<sub>600 nm</sub> reached a value of 0.3 (250 rpm). Protein expression was induced by the addition of 0.3 mM of IPTG (Isopropyl-beta-D-thiogalactopyranoside) (Qiagen, Hilden, Germany) and further incubation of the cells overnight at 25°C. The cells were pelleted by centrifugation at 4°C and 4600 rpm for 30 minutes and the pellet was resuspended in 1 ml PBS (PAA Laboratories GmbH, Cölbe, Germany). Disruption of the cells was performed

by sonication (3 x 30 sec, 10% cycle, 20% power) with a Bandelin Sonoplus HD 2070 sonicator. Insoluble material was pelleted at 13.000 rpm and 4°C for 30 min and the pellet was resuspended in 1 ml of urea buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 8 M urea, pH8). An aliquot was diluted at 1:100 and subsequently analyzed by SDS PAGE according to standard protocols. His-tagged target proteins were detected with an anti-PentaHis antibody (Qiagen, Hilden, Germany).

### (C) Results

Whereas the wildtype sequence generated a second shorter form of sgp130(D1-D3) (Figure 6, left lane) this by-product was not observed with the optimized cDNA (Figure 6, right lane). This unwanted variation of sgp130(D1-D3) is generated by further alternative transcriptional and translational start sites which have been eliminated by codon modifications in the optimized cDNA sequence. Subsequently the efficiency to produce the desired protein with the right size was increased at least at a factor of 3-fold.

### **Example 4: Identification of the best optimized sgp130Fc cDNA sequence**

The wt sgp130Fc sequence was optimized using different optimization algorithms. The resulting sequences were synthesized, cloned into respective expression vectors and transfected into certain expression cells. The best sgp130Fc sequence was identified by detection of sgp130Fc expression on the RNA and the protein level. Figure 7 shows an example of these sets of experiments and demonstrates that although three different optimized sequences were used only one sequence (namely "Opt CONARIS"), i.e., a sequence optimized according

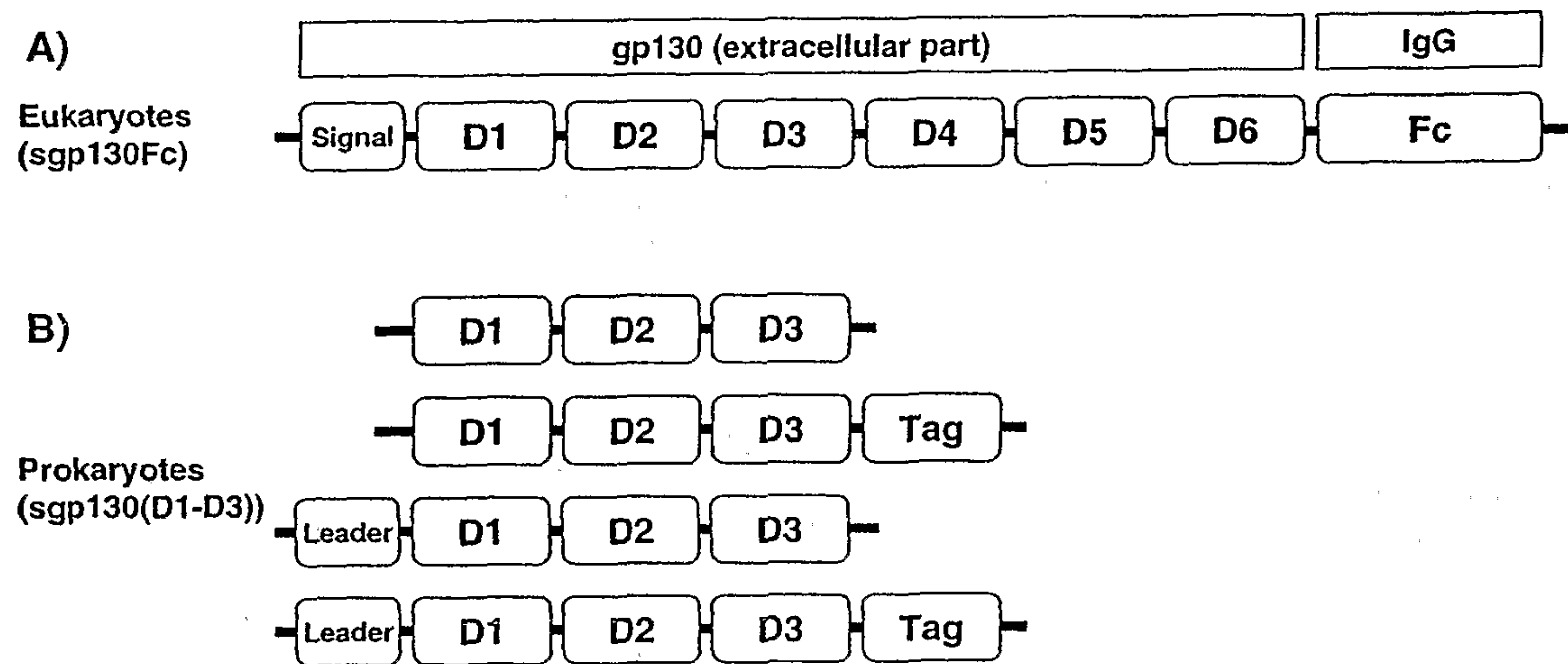
to the present invention, significantly increased the expression of sgp130Fc protein by HEK293 cells. Figure 8 represents an alignment of the same sequences. Although in all cases the optimization approach was based on an optimal codon usage in eukaryotes, the figure clearly demonstrates that the prediction for the optimal codon at a certain position is often completely different. Figure 9 summarizes the findings of these alignments by showing the number of different base pairs between each of the sequences.

These results clearly demonstrate that the computer aided prediction of cDNA sequences for optimal protein expression in a certain organism has only an extremely limited value. The development of best optimized sequences makes it necessary to choose individual research approaches and must be accompanied by high innovative technologies and inventive power.

### Claims

1. A nucleic acid molecule encoding sgp130 comprising (a) the nucleic acid sequence as (a<sub>1</sub>) depicted in Figure 2 (sgp130(D1-3)\_opt) or (a<sub>2</sub>) Figure 3 (sgp130Fc\_opt) or (b) a fragment or analogue thereof which maintains the codon usage pattern thereof.
2. The nucleic acid molecule of claim 1, wherein at least 80% of the codons altered in the nucleic acid sequence of Figure 2 or 3 vs. the wild type sequence are present.
3. The nucleic acid molecule of claim 1 or 2 which is a DNA molecule.
4. An expression vector containing a nucleic acid molecule of any one of claims 1 to 3.
5. A host cell containing an expression vector of claim 4.
6. The host cell of claim 5 which is a mammalian host cell.
7. The host cell of claim 6 which is a CHO or HEK293 cell.
8. The host cell of claim 5 which is a prokaryotic cell.
9. The host cell of claim 8 which is a bacterial cell.
10. A method of producing a sgp130 polypeptide comprising culturing a host cell of any one of claims 5 to 9 and recovering the polypeptide from said host cell or the culture.

Figure 1



|                    |      |                                                                                                                                                                                                       |
|--------------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sgp130 (D1-D3)_wt  | #1   | ATGGAACCTC TAGATCCATG TGGTTATATC AGTCTCTGAAT CTCACGTTGT ACAACTTCAT TCTAATTCA CTGCAGTTTG TGTGCTAAAG GAAAAATGTA TGGATTATTT<br>M E L L D P C G Y I S P E S P V V Q L H S N F T A V C V L K E K C M D Y F |
| sgp130 (D1-D3)_opt | #1   | ATGGAACCTC TGGACCCGTG CGGTTATATC AGCCCGGAAA GCCCAGTTGT TCAACTGCAT AGCAATTCA CCGCGTTTG TGTGCTGAAA GAGAAGTGA TGGATTATTT<br>M E L L D P C G Y I S P E S P V V Q L H S N F T A V C V L K E K C M D Y F    |
| sgp130 (D1-D3)_wt  | #111 | TCAATGTAAT GCTAATTACA TTGCTCTGGA AACAACCAT TTACTATTTC CTAAGGAGCA ATATACTATC ATAAACAGAA CAGCATCCAG TGTACCTTT ACAGATATAG<br>H V N A N Y I V W K T N H F T I P K E Q Y T I I N R T A S S V T F T D I A   |
| sgp130 (D1-D3)_opt | #111 | CCACGTCAAC GCGAAGTACA TTGCTCTGGA AACCAACCAT TTCACCATCC CGAAGAGACA GTATACCATC ATTAACCGTA CCGGAGCAG CGTTACCTTT ACCGATATCG<br>H V N A N Y I V W K T N H F T I P K E Q Y T I I N R T A S S V T F T D I A  |
| sgp130 (D1-D3)_wt  | #221 | CTTCATTAAA TATTCAGCTC ACTTGCAACA TTCTTACATT CGGACAGCTT GAACAGAATG TTTATGGAAT CACAATAATT TCGGGCTTGC CTCACAGAAA ACCTAAAAAT<br>S L N I Q L T C N I L T F G Q L E Q N V Y G I T I I S G L P P E K P K N   |
| sgp130 (D1-D3)_opt | #221 | CGAGCCTGAA CATTCAACTG ACCTGCAACA TTCTGACCTT CGGTCAACTG GAACAGAATG TTTATGGCAT CACCATCATT AGCGGCCTGC CGCCGGGAAA ACCGAAAAAT<br>S L N I Q L T C N I L T F G Q L E Q N V Y G I T I I S G L P P E K P K N   |
| sgp130 (D1-D3)_wt  | #331 | TTGAGTTGCA TTGTGAACGA GGGGAAGPAA ATGAGGTG AGTGGGATGG TGGAGGGAA ACACACTGG AGACAACTT CACTTAAAA TCTGAATGG CAACACACAA<br>L S C I V N E G K K M R C E W D G G R E T H L E T N F T L K S E W A T H K        |
| sgp130 (D1-D3)_opt | #331 | CTGAGCTGCA TTGTGAACGA AGGCAAAAAA ATGCGCTGCG AATGGGATGG TGGTCGGAA ACCCACCTGG AAACCACTT CACCTGAAA AGCGAATGG CGACCCACAA<br>L S C I V N E G K K M R C E W D G G R E T H L E T N F T L K S E W A T H K     |

Fig. 2

|                    |      |                                                                                                                                                                                                        |
|--------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| sgp130 (D1-D3)_wt  | #441 | GTTTGCTGAT TGCAAAGCAA AAGTGACAC CCCACCTCA TGCACGTGTG ATTATCTAC TGTGTATTTT GTCAACATTG AAGTCTGGGT AGAAGCAGAG AATGCCCTTG<br>F A D C K A K R D T P T S C T V D Y S T V Y F V N I E V W V E A E N A L G     |
| sgp130 (D1-D3)_opt | #441 | ATTTGCGGAT TGCAAAGCGA AAGCGGATAC CCGACCTCT TGCACCGTTG ATTACAGCAC CGTGTACTTC GTGAACATTG AAGTGTGGGT GGAAGCGGAA AAGGCCCTGG<br>F A D C K A K R D T P T S C T V D Y S T V Y F V N I E V W V E A E N A L G   |
| sgp130 (D1-D3)_wt  | #551 | GGAAGGTAC ATCAGATCAT ATCAATTG ATCCTGTATA TAAAGTGAAG CCCAATCCGC CACATAATTT ATCAGTGATC AACTCAGAGG AACGTGCTAG TATCTTAAAA<br>K V T S D H I N F D P V Y K V K P N P P H N L S V I N S E E L S S I L K       |
| sgp130 (D1-D3)_opt | #551 | GTAAGTCAC CAGCGATCAT ATCAACTTG ATCCGGTGTA CAAAGTGAAA CCGAATCCGC CGCATAATCT GAGCGTGATC AACAGCGAAG AACTGAGCAG CATCCTGAAA<br>K V T S D H I N F D P V Y K V K P N P P H N L S V I N S E E L S S I L K      |
| sgp130 (D1-D3)_wt  | #661 | TTGACATGGA CCAACCCAAG TATTAGAGT GTTATAATAC TAAATATATA CATTCAATAT AGGACCAAAG ATGCCCTCAAC TTGGAGCCAG ATTCTCTCTG AAGACACAGC<br>L T W T N P S I K S V I I L K Y N I Q Y R T K D A S T W S Q I P P E D T A  |
| sgp130 (D1-D3)_opt | #661 | CTGACCTGGA CCAACCGGAG CATTAGAGC GTTATCATCC TGAATATCAA CATTCAATAT CGCACCAAAG ATGCCAGCAC CTGGAGCCAG ATTCCGCGGG AAGATACCGC<br>L T W T N P S I K S V I I L K Y N I Q Y R T K D A S T W S Q I P P E D T A   |
| sgp130 (D1-D3)_wt  | #771 | ATCCACCGGA TCCTCATTTCA CTGTCCAAGA CCTTAAACCT TTTACAGAAAT ATGTGTTTAG GATTGCTGT ATGAAGGAAG ATGGTAAGGG ATACTGGAGT GACTGGAGTG<br>S T R S S F T V Q D L K P F T E Y V F R I R C M K E D G K G Y W S D W S E |
| sgp130 (D1-D3)_opt | #771 | GAGCACCGT AGCAGCTTCA CCGTTCAGGA TCIGAAACCG TTTACCGAAT ATGTGTTTCG CATTCGCTGT ATGAAGAAG ATGGTAAAGG CTACTGGAGC GATTGGAGCG<br>S T R S S F T V Q D L K P F T E Y V F R I R C M K E D G K G Y W S D W S E    |
|                    |      | .....                                                                                                                                                                                                  |

sgp130 (D1-D3) wt #881 AAGAAGCAAG TGGGATCACC TATGAAGATA GACCATGA

F A S G I T Y E D R P .

sgp130 (D1-D3) opt #881 AAGAAGCGAG CGGTATCACC TATGAAGATC GTCCGTGA

E A S G I T Y E D R P .

bioRxiv preprint doi: <https://doi.org/10.1101/2020.05.14.242800>; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

○ ○ ○ ○ ○

|              |      |                                                                                                                                             |
|--------------|------|---------------------------------------------------------------------------------------------------------------------------------------------|
| sgp130Fc_wt  | #1   | ATGTTGACCT TGCAGACTTG GCTAGTGCAA GCCTTGTTTA TTTTCCTCAC CACTGAATCT ACAGTGAAC TTCTAGATCC ATGTGGTTAT ATCAGTCCTG AATCTCCAGT TGTACAACCT          |
|              |      | M L T L Q T W L V Q A L F I F L T T E S T G E L L D P C G Y I S P E S P V V Q L                                                             |
| sgp130Fc_opt | #1   | ATGCTGACAC TGCAGACATG GCTGGTGAC GCCCTGTTA TCTTCTGAC CACCGAGTCT ACAGGAGAGC TGCTGGATCC TTGGGGCTAT ATCTCCCTCTG AGTCTCCTGT GGTGCAGCTG           |
|              |      | M L T L Q T W L V Q A L F I F L T T E S T G E L L D P C G Y I S P E S P V V Q L                                                             |
| sgp130Fc_wt  | #121 | CATTCTAATT TCACTCGAGT TTGTGTGCTA AAGGAAAAT GTATGGAATA TTTTCATGTA AATGCTAATT ACATTGTCTG GAAAACAAAC CATTCTACTA TTCTTAAGGA GCAATATACT          |
|              |      | H S N F T A V C V L K E K C M D Y F H V N A N Y I V W K T N H F T I P K E Q Y T                                                             |
| sgp130Fc_opt | #121 | CATTCTAACT TCACCGCCGT GTGTGTGCTG AAGGAAAAGT GCATGGACTA CTTCACGTG AACGCCAAT ACATCGTGTG GAAAACCAAC CACTTCACCA TCCCCAAGGA GCAGTACACC           |
|              |      | H S N F T A V C V L K E K C M D Y F H V N A N Y I V W K T N H F T I P K E Q Y T                                                             |
| sgp130Fc_wt  | #241 | ATCATAAACA GAACAGCATC CAGTGTACAC TTTACAGATA TAGCTTCATT AAATATTTCAG CTCACITGCA ACATCTTAC ATTGGGACAG CTGGAACAGA ATGTTTATGG AATCACAATA         |
|              |      | I I N R T A S S V T F T D I A S L N I Q L T C N I L T F G Q L E Q N V Y G I T I                                                             |
| sgp130Fc_opt | #241 | ATCATCAACC GGACCGCTTC TTCTGTGACC TTCAACGATA TCGCTCCCT GAATATCCAG CTGACCTGCA ACATCTCTGAC CTGAGGACAG ATGTTGACGG CATCACCATC                    |
|              |      | I I N R T A S S V T F T D I A S L N I Q L T C N I L T F G Q L E Q N V Y G I T I                                                             |
| sgp130Fc_wt  | #361 | ATTTGGGCT TGCCTCCAGA AAAACCTAAA AATTIGAGIT GCATTGTGAA CGAGGGGAAG AAAATGAGGT GTGAGTGGGA TGGTGGAGG GAAACACACT TGGAGACAAA CTTCACCTTA           |
|              |      | I S G L P P E K P K N L S C I V N E G K K M R C E W D G G R E T H L E T N F T L                                                             |
| sgp130Fc_opt | #361 | ATCTCTGGCC TGCCTCCAGA GAAGCCTAAG AACCTGTCT GCATCTGTA AACCTGTCTG TGGGGCAAG AAGATGAGGT GTGACTGGGA TGGCGGCAGA GAGACACATC TGGAGACCAA CTTCACCTTG |
|              |      | I S G L P P E K P K N L S C I V N E G K K M R C E W D G G R E T H L E T N F T L                                                             |

Fig. 3

sgp130Fc\_wt #481 AAATCTGAAT GGGCAACACA CAAATTGCT GATTGCAAG CAAACGGTGA CACCCCCACC TCATGCACTG TTGATTAATC TACGTGTAT TTTGTCAACA TTGAAGTCTG GGTAGAAGCA  
K S E W A T H K P A D C K A K R D T P T S C T V D Y S T V Y F V N I E V W V E A

sgp130Fc\_opt #481 AAGTCTGAGT GGGCCACCCA CAAATTGCT GACTGCAAG CCAAGAGAGA TACCCCTAAC TCCTGACCG TCGACTACTC CACCGTGTAC TTGTTGAACA TCGAGGTGTG GGTGAGGCT  
K S E W A T H K F A D C K A K R D T P T S C T V D Y S T V Y F V N I E V W V E A

sgp130Fc\_wt #601 GAGAATGCCC TTGGGAAGGT TACATCGAT CATATCAATT TTGATCCGT ATATAAAGTG AAGCCCAATC CGCCACATAA TTATCAGTG ATCAACTCAG AGGAAGTGT TAGTATCTA  
E N A L G K V T S D H I N F D P V Y K V K P N P P H N L S V I N S E E L S S I L

sgp130Fc\_opt #601 GAGAATGCTC TTGGCAAGGT GACTCTGAC CACATCAACT TCGACCCCGT GTACAAGTG AAGCCCAATC CTCTCACA CCGTTCGCG ATCAACTCTG AGGAGTGTG CTCTATCCG  
E N A L G K V T S D H I N F D P V Y K V K P N P P H N L S V I N S E E L S S I L

sgp130Fc\_wt #721 AAATGACAT GGACCAACCC AAGTATTAG AGTGTATAA TACTAAATA TAACATTCAA TATAGGACCA AAGATGCCCT AACTGGAGC CAGATTCCTC CTGAAGACAC AGCATCCACC  
K L T W T N P S I K S V I I L K Y N I Q Y R T K D A S T W S Q I P P E D T A S T

sgp130Fc\_opt #721 AAGCTGACCT GGACCAACCC TTCCATCAAG TCCGTGATCA TCGTGAAGTA CAACATCCAG TACAGGACCA AGGATGCTTC TACTGTGTCT CAGATCCCTC CTGAGGATAC CCGTTCACCC  
K L T W T N P S I K S V I I L K Y N I Q Y R T K D A S T W S Q I P P E D T A S T

sgp130Fc\_wt #841 CGATCTTCAT TCACGTGTCU AGACCTTAAA CCTTTTACAG AATATGTAG TAGGATTCGC TGTATGAGG AAGATGTAA GGGTACTGG AGTACTGGA GTGAACAAGC AAGTGGATC  
R S S F T V Q D L K P F T E Y V F R I R C M K E D G K G Y W S D W S E E A S G I

sgp130Fc\_opt #841 AGATCCAGCT TCACAGTGA GGACCTGAAG CCTTTTACCG AGTACTGTG CAGATCCCG TGCATGAGG AGGATGCAA GGGCTATTGG TCTGACTGGT CTGAGGAGGC TTCTGGCATC  
R S S F T V Q D L K P F T E Y V F R I R C M K E D G K G Y W S D W S E E A S G I

sgp130Fc\_wt #961 ACCTATGAAG ATAGACCATC TAAAGCACCA AGTTTCCTGGT ATAAATAGA TCCATCCCAT ACTCAAGCT ACAGAAGCT TGAAGACAT TGCCTCCTTT TGAAGCCAAT  
T Y E D R P S K A P S F W Y K I D P S H T Q G Y R T V Q L V W K T L P P F E A N

sgp130Fc\_opt #961 ACCTAGGAG ACAGACCCTC TAAGGCCCT AGCTTCCTGGT ACAAGATGA CCTTCTCAG ACCCAGGCT ATAGAAGCT GCAGCTGGT TGGAAAACCC TGCCTCCATT CGAGGCTAAT  
T Y E D R P S K A P S F W Y K I D P S H T Q G Y R T V Q L V W K T L P P F E A N

sgp130Fc\_wt #1081 GGAAAAATCT TGGATTATGA AGTGACTCTC ACAAGATGA AATCACATTT ACAAAATTAC ACAGTTAATG CCACAAAATG GACAGTAAAT CTCACAAATG ATCGCTATCT AGCAACCCTA  
G K I L D Y E V T L T R W K S H L Q N Y T V N A T K L T V N L T N D R Y L A T L

sgp130Fc\_opt #1081 GGCAAGATCC TGGACTATGA GGTGACCCCTG ACCAGATGA AGTCTCACCT GCAGAACTAC ACCGTGAAC CTACCAAGCT GACCTGAAC CTGACCAACG ATAGATACCT GGCTACCCCTG  
G K I L D Y E V T L T R W K S H L Q N Y T V N A T K L T V N L T N D R Y L A T L

sgp130Fc\_wt #1201 ACAGTAAGAA ATCTTGTTGG CAAATCAGAT GCAGCTGTTT TAACTATCCC TGCCGTGAC TTTCAGCTA CTCACCCCTGT AATGATCTT AAAGCATTCC CCAAGATAA CATGCTTTGG  
T V R N L V G K S D A A V L T I P A C D F Q A T H P V M D L K A F P K D N M L W

sgp130Fc\_opt #1201 ACCGTGAGAA ATCTGGTGG CAAGTCTGAT GCTGCTGTGC TGACCATCCC TGCCCTGAT TTTCAGCTA CCAACCCCTGT GATGATCTG AAGCCTTCC CCAAGATAA CATGCTGTGG  
T V R N L V G K S D A A V L T I P A C D F Q A T H P V M D L K A F P K D N M L W

sgp130Fc\_wt #1321 GTGGAATGGA CTACTCCAAG GGAATCTGTA AAGAAATATA TACTTGAGTG GTGTGTGTA TCAGATAAG CACCTGTAT CACAGACTGG CAACAAGAAG ATGGTACCGT GCATCGCACC  
V E W T T P R E S V K K Y I L E W C V L S D K A P C I T D W Q Q E D G T V H R T

sgp130Fc\_opt #1321 GTGGAGTGA CAACACTTAG AGAGTCCCTG AAGAAATACA TCCCTGAGTG GTGGGTGCTG TCTGATAAG CCACTTGAT CACAGATTGG CAGCAGGAGG ATGGCACCGT GCATAGAAC  
V E W T T P R E S V K K Y I L E W C V L S D K A P C I T D W Q Q E D G T V H R T

sgp130Fc\_wt #1441 TATTTAAGAG GGAACCTTAGC AGAGAGCAAA TGCTATTGA TAACAGTTAC TCCAGTATAT GCTGATGGAC CAGGAAGCCC TGAATCCATA AAGGCATACC TTAAACAAGC TCCACCTTCC  
Y L R G N L A E S K C Y L I T V T P V Y A D G P G S P E S I K A Y L K Q A P P S

sgp130Fc\_opt #1441 TACCTGAGAG GCAATCTGGC CGAGTCTAAG TGCTATCTGA TCACCGTGAC CCCTGTGTAT GCTGATGGAC CTGGCTCTCC TGAGTCTATC AAGGCCTACC TGAAGCAGGC TCCTCCATCT  
Y L R G N L A E S K C Y L I T V T P V Y A D G P G S P E S I K A Y L K Q A P P S

sgp130Fc\_wt #1561 AAAGGACCTA CTGTTGGGAC AAAAAAGTA GGGAAAAACG AAGCTGTCTT AGAGTGGGAC CAACCTTCCIG TTGATGTTCA GAATGGATTT ATCAGAAATTT ATACTATATT TTATAGAACC  
K G P T V R T K K V G K N E A V L E W D Q L P V D V Q N G F I R N Y T I F Y R T

sgp130Fc\_opt #1561 AAGGACCTA CCGTGAGGAC AAAGAAGGTG GGCAAGAACG AGGCTGTGCT GGAGTGGGAT CAGCTGCCIG TGGATGTSCA GAACGGCTTC ATCCGGAAT ACACCATCTT CTACCGGACC  
K G P T V R T K K V G K N E A V L E W D Q L P V D V Q N G F I R N Y T I F Y R T

sgp130Fc\_wt #1681 ATCATGGGAA ATGAACCTGC TGTGAATGIG GATTCITCCC ACACAGAATA TACATTGTCC TCTTTGACTA GTGACACATT GTACATGGTA CGAATGGCAG CATACACAGA TGAAGGTGGG  
I I G N E T A V N V D S S H T E Y T L S S L T S D T L Y M V R M A A Y T D E G G

sgp130Fc\_opt #1681 ATCATCGSCA ATGAGACCGC CGTGAACGIG GATTCITCCC ACACCGAGTA CACACTGTCC TCTCTGACCT CTGACACCCCT GTACATGGIG AGAATGGCCG CTTATATACCA TGAGGGCGGC  
I I G N E T A V N V D S S H T E Y T L S S L T S D T L Y M V R M A A Y T D E G G

sgp130Fc\_wt #1801 AAGGATGGTC CAGAATTGAG ATCTTGTGAC AAAACTCACA CATGCCACCC GTGCCAGCA CCTGAAGCCG AGGGCGGCGC GTCAGTCTTC CTCTTCCCCC CAAAACCCCA GCACACCTTC  
K D G P E F R S C D K T H T C P P C P A P E A E G A P S V F L F P P K P K D T L

sgp130Fc\_opt #1801 AAGGATGGAC CTGAGTTGAG ATCTTGTGAC AAGACCCACA CCTGTCTTCC TTGTCTGCT CCTGAGGCTG AGGGCGCTCC TTCTGTGTTT CTGTCTCCCC CAAAGCCTAA GGATACCTTG  
K D G P E F R S C D K T H T C P P C P A P E A E G A P S V F L F P P K P K D T L

|              |       |                                                                                                                                     |
|--------------|-------|-------------------------------------------------------------------------------------------------------------------------------------|
| sgp130Fc_wt  | #1921 | ATGATCTCCC GGACCCCTGA GGTACATGC GTGGTGGTG ACGTGAGCCA CGAAGACCTT GAGTCAAGT TCAACTGGTA CGTGGACGGC GTGGAGGTGC ATAATGCCAA GACAAAGCCG    |
|              |       | M I S R T P E V T C V V V D V S H E D P E V K F N W Y V D G V E V H N A K T K P                                                     |
| sgp130Fc_opt | #1921 | ATGATCTCCA GAACCCCTGA GGTACATGT GTGGTGGTG ATGTGTCTCA TGAGACCCC GAGTGAAGT TCAACTGGTA CGTGGATGGC GTGGAGGTGC ACAATGCTAA GACCAAGCCT     |
|              |       | M I S R T P E V T C V V V D V S H E D P E V K F N W Y V D G V E V H N A K T K P                                                     |
| sgp130Fc_wt  | #2041 | CGGGAGGAGC AGTACAACAG CACGTACCGT GTGGTCAGCG TCCTCACCGT CCTGCACCCAG GACTGGCTGA ATGGCAAGGA GTACAACTGC AAGTCTCCA ACAAGCCCTT CCCAGCCCCC |
|              |       | R E E Q Y N S T Y R V V S V L T V L H Q D W L N G K E Y K C K V S N K A L P A P                                                     |
| sgp130Fc_opt | #2041 | AGGGAGGAGC AGTACAACAC CACCTACAGA GTGGTGTCTG TGCTGACAGT GCTGCTCAG GATTGGCTGA ACGGCAAGGA GTACAACTGC AAGTGTCCA ACAAGGCTCT GCCTGCTCCT   |
|              |       | R E E Q Y N S T Y R V V S V L T V L H Q D W L N G K E Y K C K V S N K A L P A P                                                     |
| sgp130Fc_wt  | #2161 | ATCGAGAAA CCATCTCCA AGCCAAAGG CAGCCCCGAG AACCAAGGT GTACACCTCG CCCCCATCC GGGAGGAGAT GACCAGAAC CAGGTCAGCC TGACCTGCCT GGTCAAAGGC       |
|              |       | I E K T I S K A K G Q P R E P Q V Y T L P P S R E E M T K N Q V S L T C L V K G                                                     |
| sgp130Fc_opt | #2161 | ATCGAAAAGA CCATCTCCA GGCTAAGGA CAGCCTAGAG AGCCTCAGGT GTACACACTG CCTCCATCTA GGGAGGAGAT GACCAGAAT CAGGTGTCCC TGACCTGTCT GGTGAAGGGC    |
|              |       | I E K T I S K A K G Q P R E P Q V Y T L P P S R E E M T K N Q V S L T C L V K G                                                     |
| sgp130Fc_wt  | #2281 | TTCTATCCCA GCGACATCGC CGTGGAGTGG GAGAGCAATG GGCAGCCGGA GAACAACCTAC AAGACCAGC CTCCCCTGCT GGACTCCGAC GGCTCTTCT TCCTCTATAG CAAGTCACC   |
|              |       | F Y P S D I A V E W E S N G Q P E N N Y K T T P P V L D S D G S F F L Y S K L T                                                     |
| sgp130Fc_opt | #2281 | TTCTACCCCT CTGATATCGC TGTGGAGTGG GAGTCTAATG GCCAGCCCGA GAACAATTAC AAGACCACCC CTCCTGTGCT GGATCTGAC GGCTCTTCT TCCTGTACTC CAAACTGACC   |
|              |       | F Y P S D I A V E W E S N G Q P E N N Y K T T P P V L D S D G S F F L Y S K L T                                                     |



Figure 4

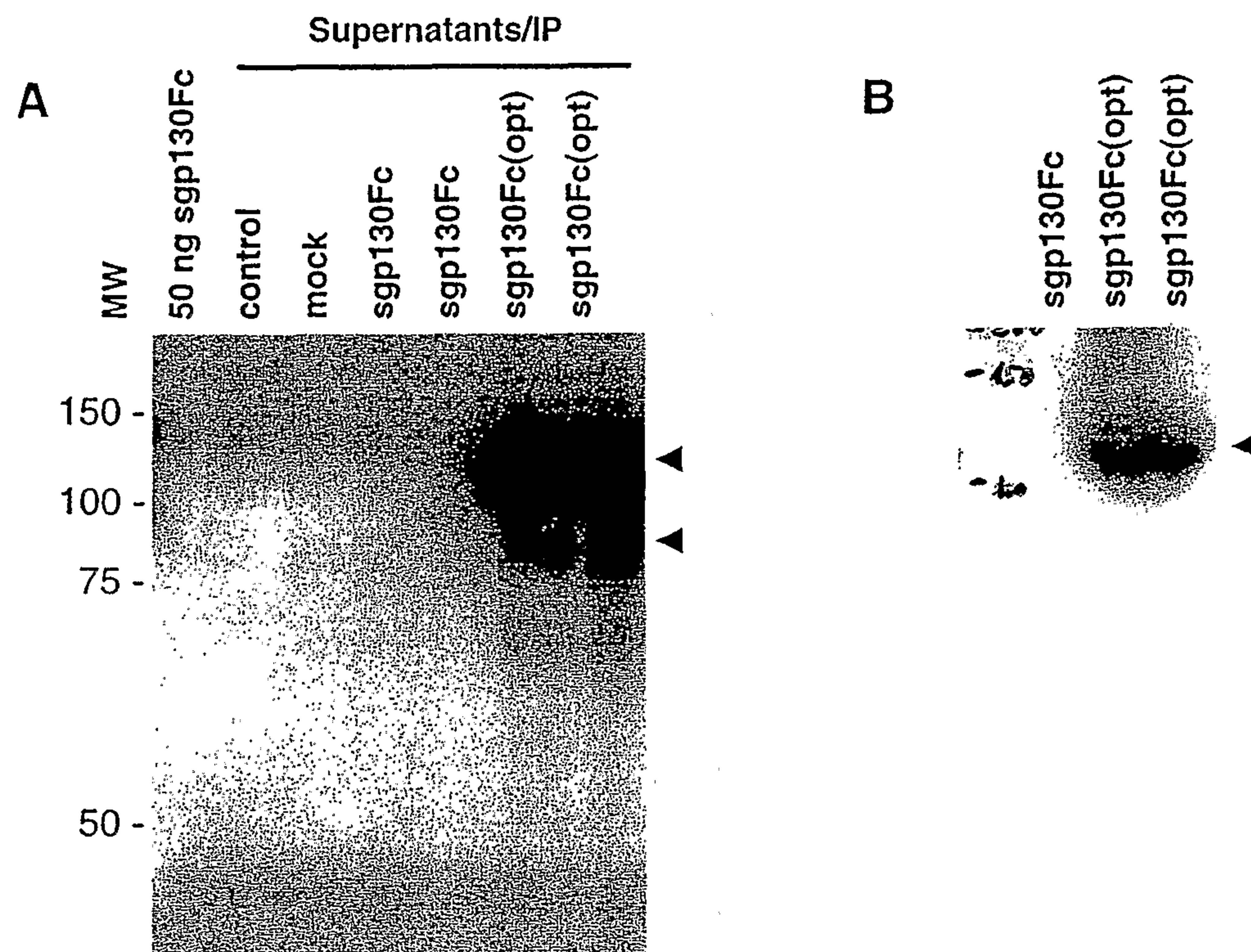
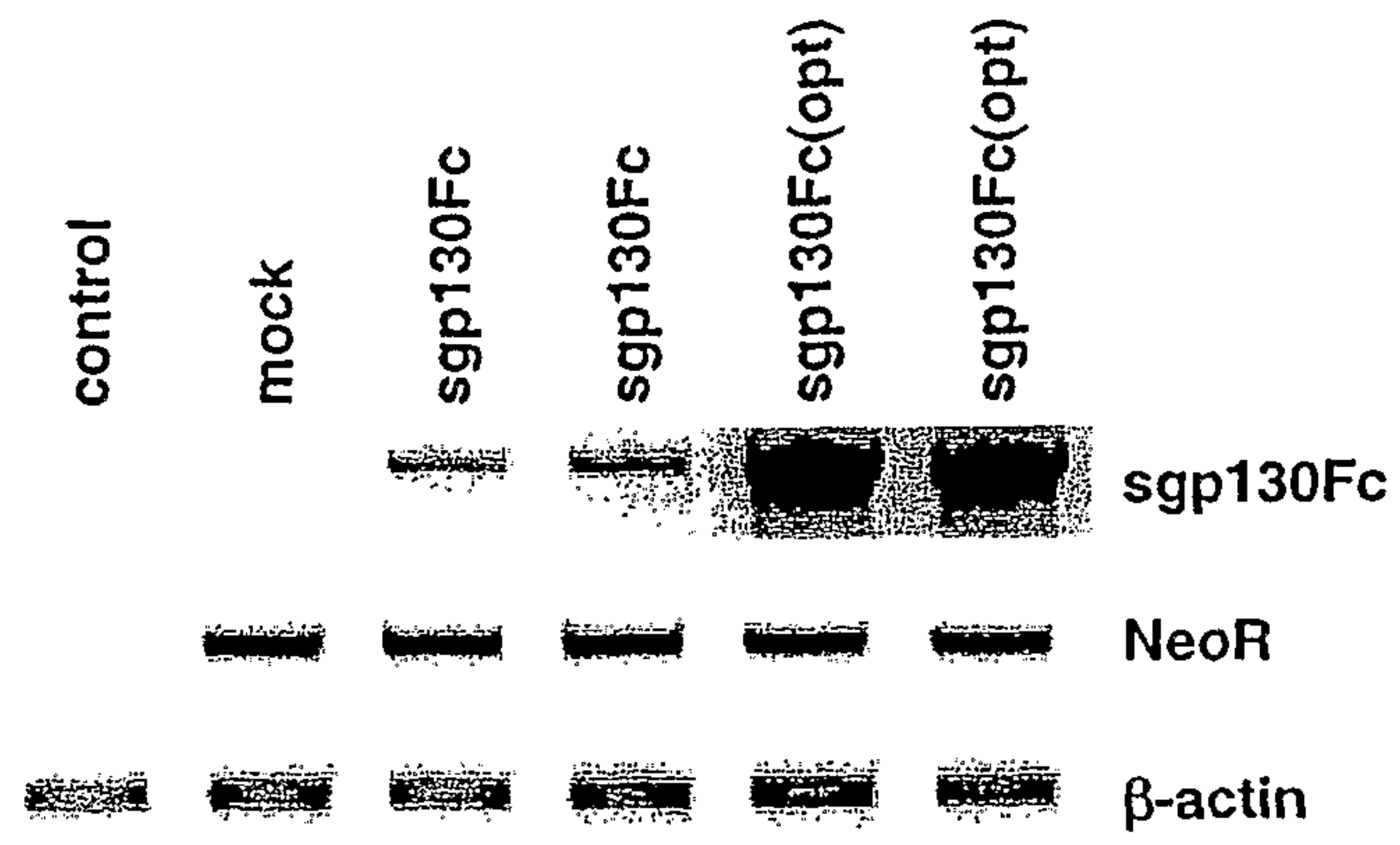


Figure 5



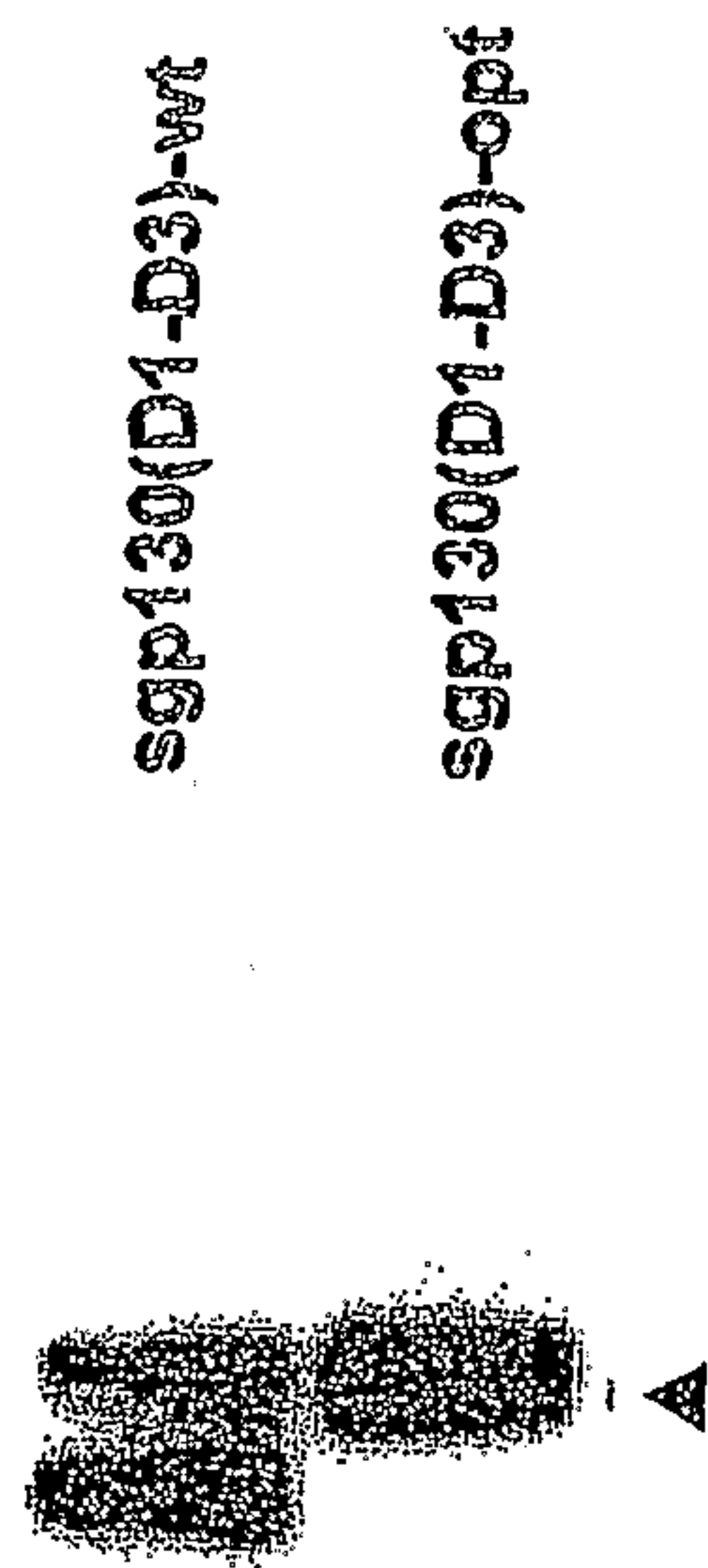


Fig. 6

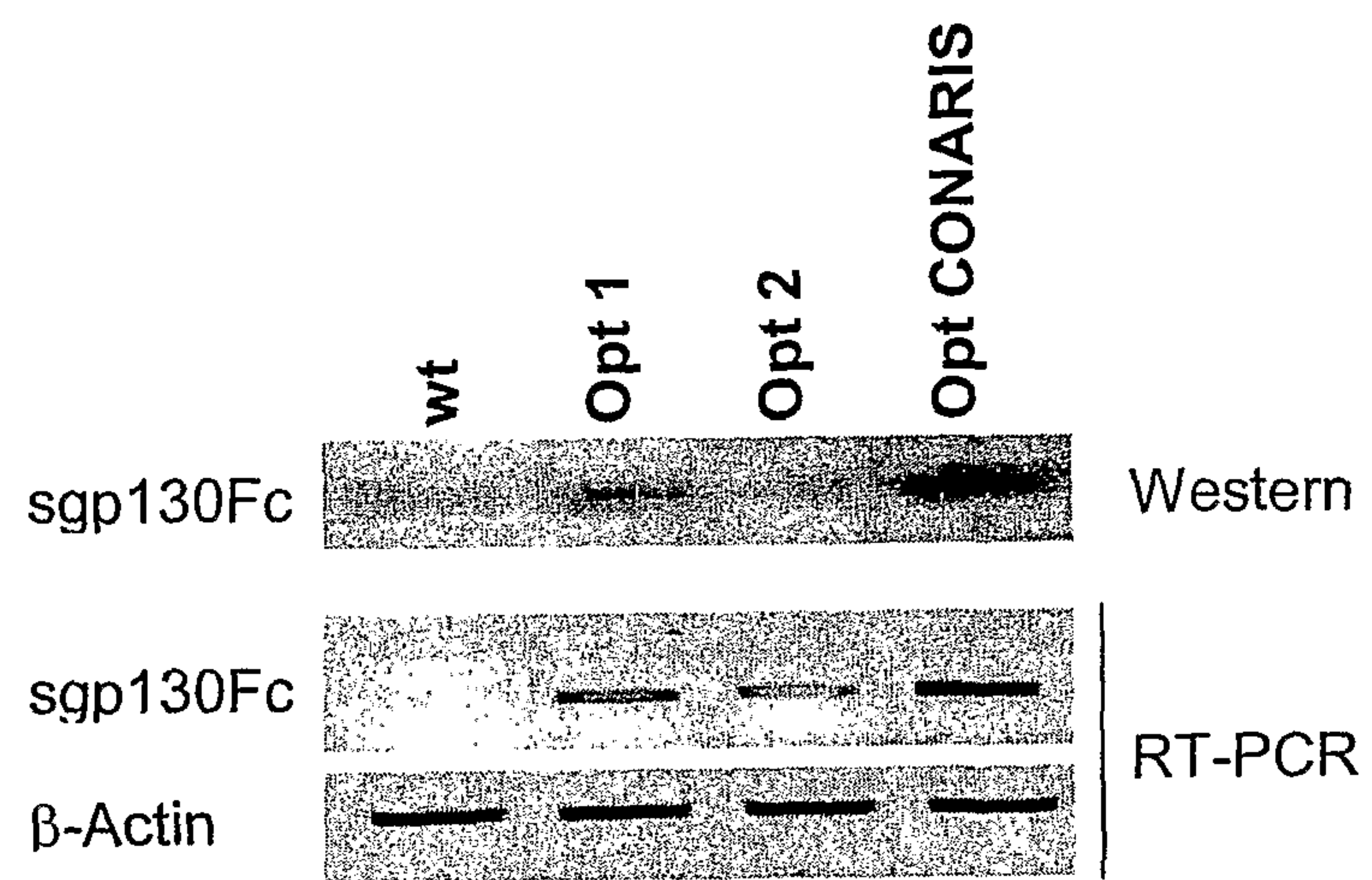
**Figure 7**

Figure 8

|                      |      |                                                       |                                                        |
|----------------------|------|-------------------------------------------------------|--------------------------------------------------------|
| >sgp130Fc_wt         | #1   | ATGTTGACGT TGCAGAGTTG GCTAGTGCAA TTTTCTCAG CACTGAATG  | ACAGGTGAAC TTCTAGATCC ATGTGTTAT ATCAGTCCTG AATCTCCAGT  |
|                      | #1   | ATGTTGACAC TGCAGACATG GCTGGTGCAG TTTTCTCAG CACTGAATG  | ACAGGAGAGC TGCTGGATCC TTGGGGCTAT ATCTCCCTG AGTCTCCTGT  |
|                      | #1   | ATGTTGACCC TGCAGACATG GCTGGTGCAG TTTTCTCAG CACTGAATG  | ACCGGGAGC TGCTGGATCC CTGGGGCTAC ATCAGCCCC AGAGCCCCGT   |
|                      | #1   | ATGTTGACAC TCCAGACATG GCTGGTGCAG TTTTCTCAG CACTGAATG  | ACAGGGAGC TGCTGGATCC CTGGGGCTAT ATTTCCCCCG AGTCTCCCCGT |
| >sgp130Fc_opt_Upgene | #1   | ATGCTGACAC TGCAGACNTG GCTGGTGCAG GCTGTTTAA TTTTCTCAG  | ACAGGGAGC TGCTGGATCC CTGGGGCTAT ATCWSCCCCG AGTCTCCCCGT |
|                      | #111 | TGTACAACTT CATTCTAAT TCACTGCAGT TTGTGTGCTA AAGGAAAAAT | TTTTCATGTA AATGCTAAT ACATTGTCTG GAAAACAAAC CATTTTACTA  |
|                      | #111 | GGTGACCTG CATTCTAAT TCACCGCCGT GTGTGTGCTG AAGGAAAAAT  | CTTCCACGTG AACGCCAAT ACATCGTGTG GAAAACAAAC CACTTCACCA  |
|                      | #111 | GGTGACCTG CACAGCAAT TCACCGCCGT GTGTGTGCTG AAGGAAAAAT  | CTTCCACGTG AACGCCAAT ACATCGTGTG GAAAACAAAC CACTTCACCA  |
| >sgp130Fc_opt_Upgene | #111 | GGTGACCTG CACTCGAAT TCACGGCCGT ATCGTCTCTC AAGGAGAAAT  | CTTCCACGTG AACGCCAAT ATATCGTTT GAAAACAAAC CACTTTACCA   |
|                      | #111 | GGTGACCTG CAYTCTAAT TCACCGCCGT GTGTGTGCTG AAGGAAAAAT  | CTTCCACGTG AACGCCAAT ACATCGTGTG GAAAACAAAC CACTTYACCA  |
|                      | #111 | GGTGACCTG CAYTCTAAT TCACCGCCGT GTGTGTGCTG AAGGAAAAAT  | CTTCCACGTG AACGCCAAT ACATCGTGTG GAAAACAAAC CACTTYACCA  |
|                      | #111 | GGTGACCTG CAYTCTAAT TCACCGCCGT GTGTGTGCTG AAGGAAAAAT  | CTTCCACGTG AACGCCAAT ACATCGTGTG GAAAACAAAC CACTTYACCA  |
| >sgp130Fc_wt         | #221 | TTCTTAAGGA GCAATATACT ATCATAAACA GAACAGCATC CAGTGTACC | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
| >sgp130Fc_opt_Upgene | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
|                      | #221 | TCCCAAGGA GCAGTACACC ATCATCAACC GGACGGTTC TTCTGTACC   | TTTACAGATA TAGCTTCATT AAATATTGAG CTCACTTGA ACATTCTTAC  |
| >sgp130Fc_wt         | #331 | CTTGAACAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
| >sgp130Fc_opt_Upgene | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
|                      | #331 | CTGAGGAGA ATGTTTATGG AATCACAATA ATTTGGGGT TGCCTCCAGA  | AAACCTAAA AATTGAGTT GCATTGTGAA CGAGGGGAAG AAAATGAGGT   |
| >sgp130Fc_wt         | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
| >sgp130Fc_opt_Upgene | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |
|                      | #441 | TGGTGAAGG GAAACACACT TGGAGACAAA CTTACCTTTA AAATCTGAAT | GGGCAACACA CAAGTTTGGT GATTGCAAA GAAAACGTGA CACCCCCACC  |





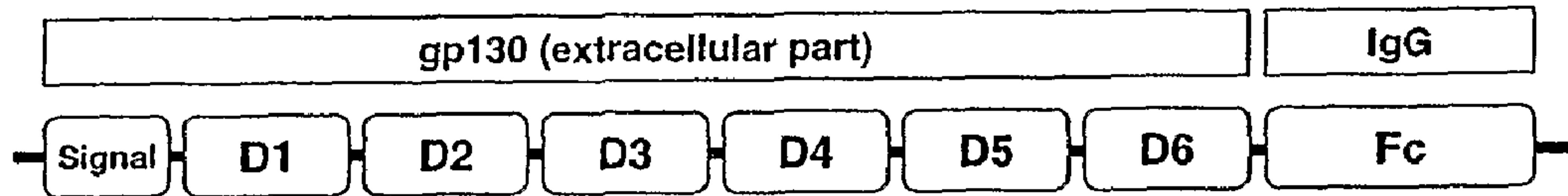
|                      |       |             |             |             |            |            |            |            |            |            |            |            |
|----------------------|-------|-------------|-------------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|
| >sgp130Fc_wt         | #1871 | AGGGCGCGCC  | GTGAGTCTTC  | CTCTTCCCCC  | CAAAACCCAA | GGACACCCTC | ATGATCTCCC | GGACCCCTGA | GGTCACATGC | GTGGTGTGG  | ACGTGAGCCA | CGAAGACCTT |
| >sgp130Fc_opt_CON    | #1871 | AGGGCGCTCC  | TTCTGTGTTT  | CTGTTCCCCC  | CAAAGCCTAA | GGATACCCTG | ATGATCTCCA | GAACCCCTGA | GGTGACATGT | GTGGTGTGG  | ATGTGTCTCA | TGAGGACCCC |
| >sgp130Fc_opt_JCat   | #1871 | AGGGCGGCCC  | CAGCGTGTTC  | CTGTTCCCCC  | CCAAGCCCAA | GGACACCCTG | ATGATCAGCC | GGACCCCGA  | GGTGACCTGC | GTGGTGTGG  | ACGTGAGCCA | CGAGGACCCC |
| >sgp130Fc_opt_Upgene | #1871 | AGGGGGGTCC  | CTCCGTGTTT  | CTCTTCCCCC  | CCAAGCCGAA | AGACACCCTC | ATGATCTCAC | GGACCCCGCA | GGTGACATGC | GTGGTGTGG  | ATGTGTCA   | CGAGGACCCC |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #1871 | AGGGCGSTCC  | CTCCGTGTTT  | CTSTTCCCCC  | CMAAGCCCAA | GGACACCCTA | ATGATCTCCC | GGACCCCYGA | GGTGACATGC | GTGGTGTGG  | AYGTGWSCCA | CGAGGACCCC |
|                      |       | * **        | ****        | * * *       | * * *      | * * *      | ****       | * * *      | * * *      | * * *      | ***        | * * *      |
| >sgp130Fc_wt         | #1981 | GAGGTCAAAGT | TCAACTGGTA  | CGTGGACGGC  | GTGGAGGTGC | ATAATGCCAA | GACAAAGCCG | CGGGAGGAGC | AGTACAACAG | CAGTACCCT  | GTGGTCAGCG | TCCTACCCT  |
| >sgp130Fc_opt_CON    | #1981 | GAGGTCAAAGT | TCAACTGGTA  | CGTGGATGGC  | GTGGAGGTGC | ACAATGCTAA | GACCAAGCCT | AGGGAGGAGC | AGTACAATC  | CACCTACAGA | GTGGTGTCTG | TGCTGACAGT |
| >sgp130Fc_opt_JCat   | #1981 | GAGGTCAAAGT | TCAACTGGTA  | CGTGGACGGC  | GTGGAGGTGC | ACAACGCCAA | GACCAAGCCC | CGCGAGGAGC | AGTACAACAG | CACCTACCGC | GTGGTGTGG  | TGCTGACCTG |
| >sgp130Fc_opt_Upgene | #1981 | GAGGTCAAAGT | TCAACTGGTA  | CGTGGACGGC  | GTGGAGGTGC | ACAATGCCAA | GACCAAGCCA | CGGGAGGAGC | AGTACAATC  | CACCTACAGA | GTAGTCAGTG | TGCTGACCTG |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #1981 | GAGGTSAAAGT | TCAACTGGTA  | CGTGGACGGC  | GTGGAGGTGC | ACAATGCCAA | GACCAAGCCN | CGGGAGGAGC | AGTACAACWS | CACCTACMGA | GTGGTSAGYG | TGCTGACCTG |
|                      |       | * *         |             | *           | *          | * * *      | *          | * * *      | *          | *          | ****       | * * *      |
| >sgp130Fc_wt         | #2091 | CCTGCACCAG  | GACTGGCTGA  | ATGGCAAGGA  | GTACAAGTGC | AAGGTCTCCA | ACAAAGCCCT | CCCAGCCCC  | ATCGAGAAAA | CCATCTCCAA | AGCCAAAGGG | CAGCCCCGAG |
| >sgp130Fc_opt_CON    | #2091 | GCTGCATCAG  | GATTGGCTGA  | ACGGCAAGGA  | GTACAAGTGC | AAGGTGTCCA | ACAAAGCTCT | GCCTGCTCT  | ATCGAAAA   | CCATCTCCAA | GGCTAAGGA  | CAGCCTAGAG |
| >sgp130Fc_opt_JCat   | #2091 | GCTGCACCAG  | GACTGGCTGA  | ACGGCAAGGA  | GTACAAGTGC | AAGGTGAGCA | ACAAAGCCCT | GCCCGCCCC  | ATCGAGAA   | CCATCAGCAA | GGCCAAAGGC | CAGCCCCGCG |
| >sgp130Fc_opt_Upgene | #2091 | CCTCCACCAG  | GACTGGCTCA  | ACGGCAAGGA  | GTACAAGTGC | AAGGTGTCCA | ACAAAGCCCT | GCCAGCCCC  | ATCGAGAAAA | CCATTTCCAA | GGCCAAAGGT | CAGCCCCAGG |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #2091 | SCTGCACCAG  | GACTGGCTGA  | ACGGCAAGGA  | GTACAAGTGC | AAGGTGTCCA | ACAAAGCCCT | GCCAGCCCC  | ATCGAGAA   | CCATCTCCAA | GGCCAAAGGN | CAGCCCMGAG |
|                      |       | * * *       | *           | *           | *          | ***        | *          | * * *      | *          | ***        | * * *      | ** *       |
| >sgp130Fc_wt         | #2201 | AACCACAGGT  | GTACACCCCTG | CCCCCATCCC  | GGGAGGAGAT | GACCAAGAAC | CAGGTGAGAT | TGACCTGCT  | GGTCAAAGGC | TTCTATCCCA | GGCACATCGC | CGTGGAGTGG |
| >sgp130Fc_opt_CON    | #2201 | AGCCTCAGGT  | GTACACACTG  | CCTCCATCTA  | GGGAGGAGAT | GACCAAGAAT | CAGGTGTGCT | TGACCTGCT  | GGTGAAGGC  | TTCTACCTT  | CTGATATCGC | TGTGGAGTGG |
| >sgp130Fc_opt_JCat   | #2201 | AGCCTCAGGT  | GTACACCCCTG | CCCCCAGCC   | GGGAGGAGAT | GACCAAGAAC | CAGGTGAGAT | TGACCTGCT  | GGTGAAGGC  | TTCTACCTT  | CTGATATCGC | TGTGGAGTGG |
| >sgp130Fc_opt_Upgene | #2201 | AGCCTCAGGT  | GTACACCCCTG | CCTCCGCTCA  | GAGAGGAAAT | GACCAAGAAT | CAGGTGTGCT | TGACCTGCT  | GGTGAAGGC  | TTCTACCTT  | CTGATATCGC | TGTGGAGTGG |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #2201 | AGCCTCAGGT  | GTACACCCCTG | CCYCCATCCM  | GGGAGGAGAT | GACCAAGAAY | CAGGTGAGAT | TGACCTGCT  | GGTGAAGGC  | TTCTACCTT  | CTGATATCGC | YGTGGAGTGG |
|                      |       | * *         | *           | * * * * *   | *          | *          | *          | *          | *          | *          | ** *       | *          |
| >sgp130Fc_wt         | #2311 | GAGAGCAATG  | GGCAGCCCGA  | GAACAACCTAC | AAGACCACGC | CTCCCGTGCT | GGACTCCGAC | GGCTCCTTCT | TCCTCTATAG | CAAGCTCACC | GTGGACAAGA | GCAGGTGGCA |
| >sgp130Fc_opt_CON    | #2311 | GAGTCTAATG  | GGCAGCCCGA  | GAACAATTAC  | AAGACCACCC | CTCCTGTGCT | GGATTCTGAC | GGCTCCTTCT | TCCTGTACTC | CAAACTGACC | GTGGACAAGT | CTAGATGGCA |
| >sgp130Fc_opt_JCat   | #2311 | GAGAGCAACG  | GGCAGCCCGA  | GAACAACCTAC | AAGACCACCC | CCCCCGTGCT | GGACAGCGAC | GGCAGCTTCT | TCCTGTACAG | CAAGCTGACC | GTGGACAAGA | GCCGCTGGCA |
| >sgp130Fc_opt_Upgene | #2311 | GAGTCCGAACG | GGCAGCCCTGA | GAATAATTAT  | AAGACAACCC | CCCCCGTGCT | GGACTCCGAC | GGCTCCTTCT | TCCTCTACTC | TAAGCTGACC | GTGGACAAGA | GTAGGTGGCA |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #2311 | GAGWSCAAYG  | GSCAGCCCGA  | GAACAAYTAC  | AAGACCACCC | CYCCCGTGCT | GGACTCCGAC | GGCTCCTTCT | TCCTSTACWS | CAAGCTGACC | GTGGACAAGA | GYAGGTGGCA |
|                      |       | ***         | *           | *           | *          | *          | *          | **         | *          | *          | *          | ***        |
| >sgp130Fc_wt         | 2421  | GCAGGGGAAC  | GTCTTCTCAT  | GCTCCGTGAT  | GCATGAGGCT | CTGCACAACC | ACTACACGCA | GAAGAGCCTC | TCCTGTCTCT | CGGGTAAATG | A          | A          |
| >sgp130Fc_opt_CON    | 2421  | GCAGGGCAAC  | GTGTTCTCTT  | GTTCGGTGAT  | GCACGAGGCT | CTGCACAATC | ACTATACCCA | GAAGTCCCTG | TCTGTCTCTC | CTGGCAAGTG | A          | A          |
| >sgp130Fc_opt_JCat   | 2421  | GCAGGGCAAC  | GTGTTCTCAT  | GCAGCGTGAT  | GCACGAGGCT | CTGCACAACC | ACTACACCCA | GAAGAGCCTG | AGCCTGAGCC | CCGGCAAGTA | A          | A          |
| >sgp130Fc_opt_Upgene | 2421  | GCAGGGGAAC  | GTCTTCTCTT  | GTTCGGTGAT  | GCACGAGGCT | CTGCACAACC | ATTATACTCA | GAAGTCCCTG | TGCTCAGTC  | CCGGTAAATG | A          | A          |
|                      |       | .....       | .....       | .....       | .....      | .....      | .....      | .....      | .....      | .....      | .....      | .....      |
|                      | #2421 | GCAGGSAAC   | GTSTTCTCTT  | GYTCCGTGAT  | GCACGAGGCT | CTGCACAACC | ACTAYACCCA | GAAGWSCTG  | TCCTGWSCTC | CCGGYAARTG | A          | A          |
|                      |       | *           | *           | ***         | ***        | *          | *          | *          | *          | *          | *          | *          |

|            | wt  | opt_Con | opt_JCat | opt_UpGene |
|------------|-----|---------|----------|------------|
| wt         | -   | 561     | 626      | 588        |
| opt_Con    | 561 | -       | 383      | 405        |
| opt_JCat   | 626 | 383     | -        | 404        |
| opt_UpGene | 588 | 405     | 404      | -          |

**Figure 9**

**A)**

**Eukaryotes  
(sgp130Fc)**



**B)**

**Prokaryotes  
(sgp130(D1-D3))**

