RECEIVED Sec 1 55 63 46 1

DESCRIPTION

Impaired Dominant Selectable Marker Sequence And Intronic Insertion Strategies For Enhancement Of Expression Of Gene Product And Expression Vector Systems Comprising Same

37 C.F.R. §1.74(d)/(e) Copyright Notice

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner does not object to the reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent files or records, but otherwise reserves all copyright rights whatsoever.

Related Applications

This patent document is a Continuation-In-Part of U.S. Serial No. 07/977,691 filed November 13, 1992, pending. This patent document is related to "THERAPEUTIC APPLICATION OF CHIMERIC ANTIBODY TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA," having U.S. Serial No. 07/978,891, filed November 13, 1992 (pending), and "THERAPEUTIC APPLICATION OF CHIMERIC AND RADIOLABLED ANTIBODIES TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA", having U.S. Serial No. _____, filed simultaneously herewith. This patent document is related to commonly assigned United States Serial No. 07/912,292 and entitled "RECOMBINANT ANTIBODIES FOR HUMAN THERAPY," filed July 10, 1992. These documents are incorporated herein by reference.

This Sheet Has Been Replaced



FEB-19-96 22:41 FROM: ID: PAGE 5

INTENTIONALLY
LEFT BLANK
PER EXAMINER
M. HOGARTH

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:42 FROM: ID: PAGE 6

As noted, the advent of the biotechnology industry has allowed for the production of large quantities of proteins. Proteins are the essential constituents of all living cells and proteins are comprised of combinations of 20 naturally occurring amino acids; each amino acid molecule is defined ("encoded") by groupings ("codons") of three deoxyribonucleic acid ("DNA") molecules; a string of DNA molecules ("DNA macromolecule") provides, in essence, a blueprint for the production of specific sequences of amino acids specified by that blueprint. Intimately involved in this process is ribonucleic acid ("RNA"); three types of RNA (messenger RNA; transfer RNA; and ribosomal RNA) convert the information encoded by the DNA into, *eg* a protein. Thus, genetic information is generally transferred as follows: DNA →RNA → protein.

In accordance with a typical strategy involving recombinant DNA technology, a DNA sequence which encodes a desired protein material ("cDNA") is identified and either isolated from a natural source or synthetically produced. By manipulating this piece of genetic material, the ends thereof are tailored to be ligated, or "fit," into a section of a small circular molecule of double stranded DNA. This circular molecule is typically referred to as a "DNA expression vector," or simply a "vector." The combination of the vector and the genetic material can be referred to as a "plasmid" and the plasmid can be replicated in a prokaryotic host (*ie* bacterial in nature) as an autonomous circular DNA molecule as the prokaryotic host replicates. Thereafter, the circular DNA plasmid can be isolated and introduced into a eukaryotic host (*ie* mammalian in nature) and host cells which have incorporated the plasmid DNA are selected. While some plasmid vectors will replicate as an autonomous circular DNA molecule in mammalian cells, (*eg* plasmids comprising Epstein Barr virus ("EBV") and Bovine Papilloma virus ("BPV") based vectors), most plasmids including DNA vectors, and all plasmids including RNA retroviral vectors, are integrated into the cellular DNA such that

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:42 FROM: ID: PAGE 7

when the cellular DNA of the eukaryotic host cell replicates, the plasmid DNA will also replicate. Accordingly, as eukaryotic cells grow and divide, there is a corresponding increase in cells containing the integrated plasmid which leads to the production ("expression") of the protein material of interest. By subjecting the host cells containing the plasmid to favorable growth conditions, significant amounts of the host, and hence the protein of interest, are produced. Typically, the Chinese Hamster Ovary ("CHO") cell line is utilized as a eukaryotic host cell, while *E. coli* is utilized as a prokaryotic host cell.

The vector plays a crucial role in the foregoing - manipulation of the vector can allow for variability as to where the cDNA is inserted, means for determining whether the cDNA was, in fact, properly inserted within the vector, the conditions under which expression of the genetic material will or will not occur, etc. However, most of the vector manipulations are geared toward a single goal-increasing expression of a desired gene product, *ie* protein of interest. Stated again, most vector manipulation is conducted so that an "improved" vector will allow for production of a gene product at significantly higher levels when compared to a "non-improved" vector. Thus, while certain of the features/aspects/characteristics of one vector may appear to be similar to the features/aspects/characteristics of another vector, it is often necessary to examine the result of the overall goal of the manipulation - improved production of a gene product of interest.

While one "improved" vector may comprise characteristics which are desirable for one set of circumstances, these characteristics may not necessarily be desirable under other circumstances. However, one characteristic is desirable for all vectors: increased efficiency, *ie* the ability to increase the amount of protein of interest produced while at the same time decreasing the number of host cells to be screened which

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:42 FROM: ID: PAGE 8

do not generate a sufficient amount of this protein. Such increased efficiency would have several desirable advantages, including reducing manufacturing costs and decreasing the time spent by technicians in screening for viable colonies which are expressing the protein of interest. Accordingly, what would be desirable and what would significantly improve the state of the art are expression vectors with such efficiency characteristics.

Summary Of The Invention

The invention disclosed herein satisfies these and other needs. Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector; preferably, the dominant selectable marker comprises either a natural intronic insertion region or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within such insertion region.

As used herein, a "dominant selectable marker" is a gene sequence or protein encoded by that gene sequence; expression of the protein encoded by the dominant selectable marker assures that a host cell transfected with an expression vector which includes the dominant selectable marker will survive a selection process which would otherwise kill a host cell not containing this protein. As used herein, a "transcriptional cassette" is DNA encoding for a protein product (*eg* a dominant selectable marker) and the genetic elements necessary for production of the protein product in a host cell (*ie* promoter; transcription start site; polyadenylation region; etc.). These vectors are most preferably utilized in the expression of proteins in mammalian expression systems where integration of the vector into host cellular DNA occurs. Beneficially, the use of such fully impaired consensus Kozak sequences improves the efficiency of protein expression by significantly decreasing the number

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:43 FROM: ID: PAGE 9

of viable colonies while at the same time, significantly increasing the amount of protein expressed by such viable colonies. As used herein, a "natural intronic insertion region" is a region of DNA naturally present within a gene in this case, typically a dominant selectable marker, which can be utilized for insertion of DNA encoding a gene product of interest; an "artificial intronic insertion region" is a region of DNA which is selectively created in a gene (again, most typically, a dominant selectable marker) which can be utilized for insertion of DNA encoding a gene product of interest. Information regarding intronic positioning is described in Abrams, J.M. et al. "Intronic Positioning Maximizes Co-expression and Co-amplification of Nonselectable Heterologous Genes. "*J. Bio. Chem., 264124*:14016 (1989), and U.S. Patent No. 5,043,270 (both documents being incorporated herein by reference).

As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T)/ uracil (U); "Py" is a pyrimidine nucleotide, *ie* C or T/U; "ATG" is a codon encoding for the amino acid methionine, the so-called "start" codon; and -3 and +1 are directional reference points vis-a-vis ATG, *ie* -3 is meant to indicate three nucleotides upstream of ATG and +1 is meant to indicate one nucleotide downstream of ATG.

Preferably, the fully impaired consensus Kozak is part of a methionine start codon that initiates translation of a dominant selectable marker portion of a transcriptional cassette which is part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:43 FROM: ID: PAGE 10

transferase; hygromycin B phosphotransferase; and neomycin phosphotransferase. Most preferably, the dominant selectable marker is neomycin phosphotransferase.

In particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without an in-frame stop codon being located between the upstream start codon and the fully impaired consensus Kozak start codon. As used herein, the term "stop codon" is meant to indicate a codon which does not encode an amino acid such that translation of the encoded material is terminated; this definition includes, in particular, the traditional stop codons TAA, TAG and TGA. As used herein, the terms "in-frame" and "out-of-frame" are relative to the fully impaired consensus Kozak start codon. By way of example, in the following sequence:

____ -3 +1

GAC CAT GGC Cxx ATG Cxx

the underlined portion of the sequence is representative of a fully impaired consensus Kozak (where "x" represents a nucleotide) and the codons GAC, CAT and GCC are "in-frame" codons relative to the ATG start codon. The above-lined nucleotides represent an "out-of-frame" start codon which is upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is a part of a consensus Kozak. By way of example, the sequence set forth above satisfies this criteria: the -5 nucleotide is a purine (G); nucleotide -6, -7 and -8 encode an out-of-frame start codon (ATG); and nucleotide -11 is a purine (A).

Additionally, utilization of a fully impaired consensus Kozak within a secondary structure (*ie* a so-

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:43 FROM: ID: PAGE 11

called "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable number. In such an embodiment, the start codon of the fully impaired consensus Kozak is most preferably located within the stem of a stem loop.

Particularly preferred expression vectors which incorporate these aspects of the invention disclosed herein are referred to as "TCAE," and "ANEX" and "NEOSPLA" vectors; particularly preferred vectors are referred to as ANEX 1, ANEX 2 and NEOSPLA3F.

These and other aspects of the invention disclosed herein will be delineated in further detail in the sections to follow.

Brief Description Of The Drawings

Figure 1 provides the relevant portion of a consensus Kozak and several particularly preferred fully impaired consensus Kozak sequences;

Figure 2 provides a diagrammatic representation of the vectors TCAE 5.2 and ANEX 1 (TCAE 12) designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 3 is a histogram comparing protein expression levels with the vectors TCAE 5.2 and ANEX 1;

Figure 4 provides a diagrammatic representation of the vector ANEX 2 designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 5 is a histogram comparing protein expression levels with the vectors TCAE 5.2, ANEX 1 and ANEX 2;

Figure 6 provides a diagrammatic representation of a NEOSPLA vector designed for expression of mouse/human chimeric immunoglobulin; and

This Sheet Has Been Replaced



FEB-19-96 22:44 FROM: ID: PAGE 12

Figures 7A, 7B and 7C are histograms comparing protein expression levels with the vectors TCAE 5.2 vs. NEOSPLA 3F (7A). ANEX 2 vs. NEOSPLA3F (7B); and GK-NEOSPLA3F vs. NEOSPLA3F (7C).

Detailed Description Of Preferred Embodiments

Disclosed herein are nucleic acid sequence arrangements which impair translation and initiation of, most preferably, dominant selectable markers incorporated into mammalian expression vectors and which are preferably, but not necessarily, co-linked to an encoding sequence for a gene product of interest. Preferably, the dominant selectable marker comprises at least one natural or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within at least one such intron. Such arrangements have the effect of increasing expression efficiency of the gene product of interest by, *inter alia*, decreasing the number of viable colonies obtained from an equivalent amount of plasmid DNA transfected per cell, while increasing the amount of gene product expressed in each clone.

For purpose of brevity and presentational efficiency, the focus of this section of the patent disclosure will be principally directed to a specific dominant selectable marker, neomycin phosphotransferase, which is incorporated into a mouse/human chimeric immunoglobulin expression vector. It is to be understood, however, that the invention disclosed herein is not intended, nor is it to be construed, as limited to these particular systems. To the contrary, the disclosed invention is applicable to mammalian expression systems *in toto*, where vector DNA is integrated into host cellular DNA.

One of the most preferred methods utilized by those in the art for producing a mammalian cell line that produces a high level of a protein (ie "production cell line") involves random integration of DNA coding for the desired gene product (ie "exogenous DNA") by using, most typically, a drug resistant gene, referred to as a

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:44 FROM: ID: PAGE 13

"dominant selectable marker," that allows for selection of cells that have integrated the exogenous DNA. Stated again, those cells which properly incorporate the exogenous DNA including, *eg*, the drug resistant gene, will maintain resistance to the corresponding drug. This is most typically followed by co-amplification of the DNA encoding for the desired gene product in the transfected cell by amplifying an adjacent gene that also encodes for drug resistance ("amplification gene"), *eg* resistance to methotrexate (MTX) in the case of dehydrofolate reductase (DHFR) gene. The amplification gene can be the same as the dominant selectable marker gene, or it can be a separate gene. (As those in the art appreciate, "transfection" is typically utilized to describe the process or state of introduction of a plasmid into a mammalian host cell, while "transformation" is typically utilized to describe the process or state of introduction of a plasmid into a bacterial host cell).

Two amplification approaches are typically employed by those in the art. In the first, the entire population of transfected and drug resistant cells (each cell comprising at least one integration of the gene encoding for drug resistance) is amplified; in the second, individual clones derived from a single cell are amplified. Each approach has unique advantages.

With respect to the first approach, it is somewhat "easier" to amplify the entire population (typically referred to as a "shotgun" approach, an apt description) compared to individual clones. This is because amplification of individual clones initially involves, *inter alia*, screening of hundreds of isolated mammalian colonies (each derived from a single cell, most of which being single copy integrants of the expression plasmid) in an effort to isolate the one or two "grail" colonies which secrete the desired gene product at a "high" level, *ie* at a level which is (typically) three orders of magnitude higher than the lowest detectable expression level. These cells are also often found to have only a single copy

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:44 FROM: ID: PAGE 14

integration of the expression plasmid. Additionally, amplifying individual clones results in production cell lines which contain fewer copies of the amplified gene as compared to amplification of all transfected cells (typically, 10-20 versus 500-1000).

With respect to the second approach, production cell lines derived from amplifying individual clones are typically derived in lower levels of the drug(s) used to select for those colonies which comprise the gene for drug(s) resistance and the exogenous gene product (*ie* in the case of methotrexate and DHFR, 5nM versus 1μM). Furthermore, individual clones can typically be isolated in a shorter period of time (3-6 months versus 6-9 months).

Ideally then, the tangible benefits of both approaches should be merged: at a practical level, this would involve decreasing the number of colonies to be screened, and increasing the amount product secreted by these colonies. The present invention accomplishes this task.

The position where the DNA of the dominant selectable marker of the plasmid DNA is integrated within the cellular DNA of the host cell determines the level of expression of the dominant selectable marker protein, as is recognized by those in the art. It is assumed that the expression of a gene encoding a protein of interest which is either co-linked to or positioned near the dominant selectable marker DNA is proportional to the expression of the dominant selectable marker protein. While not wishing to be bound by any particular theory, the inventor has postulated that if the gene used to select for the integration of the exogenous DNA in the mammalian cell (*ie* the dominant selectable marker) was designed such that translation of that dominant selectable marker was impaired, then only those plasmids which could overcome such impairment by over-production of the gene product of the dominant selectable marker would survive, *eg*, the drug-screening process. By associating the exogenous DNA

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-98 22:44 FROM: ID: PAGE 15

with the dominant selectable marker, then, *a fiorti*, over-production of the gene product of the dominant selectable marker would also result in over-production of the gene product derived from the associated exogenous DNA. In accordance with this postulated approach, impairment of translation of the dominant selectable marker gene would be necessary, and an avenue for such impairment was the consensus Kozak portion of the gene.

By comparing several hundred vertebrate mRNAs, Marilyn Kozak in "Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes," *Nuc. Acids Res. 9*: 5233-5252 (1981) and "Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs," *Nuc. Acids Res. 12*: 857-872 (1984), proposed the following "consensus" sequence for initiation of translation in higher eukaryotes:

```
          -3      +1
       cc Acc AUG G
```

(As those in the art appreciate, uracil, U, replaces the deoxynucleotide thymine, T, in RNA.) In this sequence, referred to as a "consensus Kozak," the most highly conserved nucleotides are the purines, A and G, shown in capital letters above; mutational analysis confirmed that these two positions have the strongest influence on initiation. *See, eg,* Kozak, M. "Effects of intercistronic length on the efficiency of reinitiation of eukaryotic ribosomes." *Mol. Cell Bio. 7/10*: 3438-3445 (1987). Kozak further determined that alterations in the sequence upstream of the consensus Kozak can effect translation. For example, in "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes." *PNAS 83*: 2850-2854 (1986) Kozak describes the "artificial" introduction of a secondary hairpin structure region upstream from the consensus Kozak in several plasmids that encoded preproinsulin; it was experimentally determined that a stable stem loop structure inhibited translation of the

This Sheet Has Been Replaced



FEB-19-96 22:45 FROM: ID: PAGE 16

preproinsulin gene, reducing the yield of proinsulin by 85-95%.

Surprisingly, it was discovered by the inventor that by changing the purines A(-3 vis-a-vis ATG start codon) and G(+1) to pyrimidines, translation impairment was significant: when the consensus Kozak for the neomycin phosphotransferase gene was subjected to such alterations (as will be set forth in detail below), the number of G418 resistant colonies significantly decreased; however, there was a significant increase in the amount of gene product expressed by the individual G418 resistant clones. As those in the art will recognize, this has the effect of increasing the efficiency of the expression system—there are less colonies to screen, and most of the colonies that are viable produce significantly more product than would ordinarily be obtained. Confirmation of the inventor's postulated theory was thus experimentally determined.

As noted, for purposes of this patent document a, "consensus Kozak" comprises the following sequence:

-3 +1

Pxx ATG Pxx

a "partially impaired consensus Kozak" comprises the following sequence

-3 +1

P/Pyxx ATG P/Pyxx

and a disclosed and claimed "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T) (uracil, U, in the case of RNA); "P" is a purine, *ie* A or G, "Py" is a pyridine, *ie* C or T/U; ATG is a conventional start codon which encodes for the amino acid methionine (Met); the numerical designations are relative to the ATG codon, *ie* a negative number indicates "upstream" of ATG and a positive number indicates

This Sheet Has Been Replaced

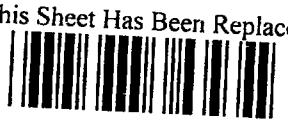
SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:45 FROM: ID: PAGE 17

"downstream" of ATG; and for the partially impaired consensus Kozak, the following proviso is applicable--only one of the -3 or +1 nucleotides is a pyridine, *eg*, if -3 is a pyridine, then +1 must be a purine or if -3 is a purine, then +1 must be a pyridine. Most preferably, the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker which is preferably (but not necessarily) co-linked to exogenous DNA which encodes for a gene product of interest. As used herein, "nucleotide" is meant to encompass natural and synthetic deoxy- and ribonucleotides as well as modified deoxy- and ribonucleotides, *ie* where the 3' OH, 5'OH, sugar and/or heterocyclic base are modified, as well as modification of the phosphate backbone, *eg* methyl phosphates, phosphorothioates and phosphoramidites.

Information regarding the gene sequence of the dominant selectable marker is preferably known; however, in lieu of the entire sequence, information regarding the nucleic acid sequence (or amino acid sequence) at the site of translation initiation of the dominant selectable marker must be known. Stated again, in order to effectuate a change in the consensus or partially impaired consensus Kozak, one must know the sequence thereof. Changing the consensus or partially impaired consensus Kozak to a fully impaired consensus Kozak sequence can be accomplished by a variety of approaches well known to those in the art including, but not limited to, site specific mutagenesis and mutation by primer-based amplification (*eg* PCR); most preferably, such change is accomplished via mutation by primer-based amplification. This preference is principally based upon the comparative "ease" in accomplishing the task, coupled with the efficacy associated therewith. For ease of presentation, a description of the most preferred means for accomplishing the change to a fully impaired consensus Kozak will be provided.

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:45 FROM: ID: PAGE 18

In essence, mutation by primer-based amplification relies upon the power of the amplification process itself--as PCR is routinely utilized, focus will be directed thereto. However, other primer-based amplification techniques (*eg* ligase chain reaction, etc.) are applicable. One of the two PCR primers ("mutational primer") incorporates a sequence which will ensure that the resulting amplified DNA product will incorporate the fully impaired consensus Kozak within the transcriptional cassette incorporating the dominant selectable marker of interest; the other PCR primer is complementary to another region of the dominant selectable marker; a transcriptional cassette incorporating the dominant selectable marker; or a vector which comprises the transcriptional cassette. By way of example, the complement to a dominant selectable marker which includes a consensus Kozak could have the following sequence (SEQ ID NO: 1):

3'-tagctaggTccTACCcc-5'

In order to create a fully impaired consensus Kozak, the mutational primer could have the following sequence (SEQ ID NO: 2) (for convenience, SEQ ID NO: 1 is placed over the mutational primer for comparative purposes):

5'-atcgatccTggATGCgg-3'
3'-tagctaggTccTACCcc-5'
* *

As is evident, complementarity is lacking in the primer (*see* the "*" symbols). By utilizing excess mutational primer in the PCR reaction, when the sequence including the consensus Kozak is amplified, the resulting amplified DNA products will incorporate the mutations such that as the amplified DNA products are in turn amplified, the mutations will predominate such that a fully impaired consensus Kozak will be incorporated into the amplification product.

Two criteria are required for the mutational primer--first, the length thereof must be sufficient such that hybridization to the target will result. As will be

SUBSTITUTE SHEET (RULE 26)

This Sheet Has Been Replaced



FEB-19-96 22:45 FROM: ID: PAGE 19

appreciated, the mutational primer will not be 100% complementary to the target. Thus, a sufficient number of complementary bases are required in order to ensure the requisite hybridization. Preferably, the length of the mutational primer is between about 15 and about 60 nucleotides, more preferably between about 18 and about 40 nucleotides, although longer and shorter lengths are viable. (To the extent that the mutational primer is also utilized to incorporate an out-of-frame start codon or secondary structure, the length of the mutational primer can correspondingly increase). Second, the ratio of mutational primer to target must be sufficiently excessive to "force" the mutation. Preferably, the ratio of mutational primer to target is between about 250:1 to about 5000:1, more preferably between about 400:1 to about 2500:1, and most preferably between about 500:1 to about 1000:1.

Because the parameters of a PCR reaction are considered to be well within the level of skill of those in the art, details regarding the particulars of that reaction are not set forth herein; the skilled artisan is readily credited with recognizing the manner in which this type of mutation can be accomplished using PCR techniques - the foregoing is provided as a means of providing elucidation as opposed to detailed edification.

As noted, it is most preferred that the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker incorporated into a transcriptional cassette which forms a part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl transferase ("XGPRT"); hygromycin B phosphotransferase; and neomycin phosphotransferase ("NEO"). Most preferably, the dominant selectable marker is NEO.

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:46 FROM: ID: PAGE 20

The dominant selectable marker herpes simplex virus thymidine kinase is reported as having the following partially impaired consensus Kozak:

-3 +1

cg Cgt ATG Gct

See, Heller, S. "Insertional Activities of a Promotorless Thymidine Kinase Gene," *Mol. & Cell. Bio. 8/8*:3218-3302, Figure 4, nucleotide 764. By changing the +1 purine (G) to a pyrimidine (C or T/U), a fully impaired Kozak as defined herein is generated (the -3 of the herpes simplex virus thymidine kinase is a pyrimidine). Changing +1 purine to a pyrimidine also has the effect of changing the encoded amino acid from alanine (GCT) to proline (CCT) or serine (TCT); it is preferred that conservative amino acid changes result from the changes to the nucleotides. Thus, it is preferred that the change to TCT be made because the change from alanine to serine is a more conservative amino acid change than changing alanine to proline.

Histidinol dehydrogenase is another dominant selectable marker. *See*, Hartmen, S.C. and Mulligan, R.C. "Two dominant acting selectable markers for gene transfer studies in mammalian cells," *PNAS 85*:8047-8051 (1988). The *his D* gene of Salmonella typhimunium has the following partially impaired consensus Kozak:

-3 +1

gc Aga ATG Tta

As -3 is a purine, changing -3 to a pyrimidine (C or T/U) results in a fully impaired consensus Kozak; as is appreciated, because these nucleotides are upstream of the start codon, no impact on amino acid translation results from this change.

Hygromycin B phosphotransferase is another dominant selectable marker; the reported sequence for the *hph* gene (*see*, Gritz, L. and Davies, J. "Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expressing in *Escherichia*

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:46 FROM: ID: PAGE 21

coli and *Succharomyces cerevisiae*" *Gene 25*:179-188, 1983) indicates that the consensus Kozak is:

-3 +1

ga Gat ATG Aaa

Both -3 and +1 are purines; thus changing -3 and +1 to pyrimidines results in a fully impaired consensus Kozak (this results in the following encoded amino acids: +1 to C - glutamine; +1 to T - stop codon. Because this codon is downstream of the start codon, the change to the stop codon TAA should not be accomplished).

XGPRT is another dominant selectable marker. The reported partially impaired consensus Kozak of XGPRT has the following sequence:

-3 +1

tt Cac ATG Agc

See, Mulligan, R.C. and Berg, P. "Factors governing the expression of a bacterial gene in mammalian cells." *Mol. & Cell Bio. 1/5*:449-459 (1981), Figure 6. By changing the +1 purine to a pyrimidine, a fully impaired consensus Kozak is created; the effect on the encoded amino acid (AGC-serine) is as follows: CGC-arginine; TGC-cysteine.

Adenosine deaminase (ADA) can also be utilized as a dominant selectable marker. The reported consensus Kozak sequence for adenosine deaminase is:

-3 +1

ga Acc ATG Gcc

See, Yeung, C.Y. *et al.*, "Identification of functional murine adenosine deaminase cDNA clones by complementation in Echerichia coli," *J. Bio. Chem. 260/18*:10299-10307 (1985), Figure 3. By changing both -3 and +1 purines to pyrimidines, fully impaired consens Kozak sequences result. The encoded amino acid corresponding to GCC (alanine) is changed to either proline (CCC) or serine (TCC), with the change to serine being preferred, due to the conservative nature of this change.

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:46 FROM: ID: PAGE 22

The reported partially impaired consensus Kozak for asparagine synthetase is as follows:

-3 +1

gc Acc ATG Tgt

See, Andrulis, I.L. *et al.*, "Isolation of human cDNAs for asparagine synthetase and expression in Jensen rat sarcoma cells," *Mol. Cell. Bio.* 7/7:2435-2443 (1987). Changing the +3 purine to a pyrimidine results in a fully impaired consensus Kozak.

The partially impaired consensus Kozak for neomycin phosphotransferase (which includes an upstream out of frame start codon) is as follows:

-3 +1

ggA TGg gga tcg ttt Cgc ATG Att

Changing the +1 purine to a pyrimidine has the effect of creating a fully impaired consensus Kozak (changes to the encoded amino acid isoleucine, ATT are as follows: CTT - leucine and TTT-phenylalanine, with the change to leucine being preferred, due to the conservative nature of this change).

The foregoing is not intended, nor is it to be construed as limiting; rather, in the context of the disclosed invention, the foregoing is presented in an effort to provide equivalent examples of changes in the reported consensus Kozak sequences or partially impaired consensus Kozak sequences of several well-known dominant selectable markers.

As noted, a most preferred dominant selectable marker is NEO. Particularly preferred fully impaired consensus sequences for NEO are as follows:

Txx ATG Ctt (SEQ ID NO: 3)
Cxx ATG Ctt (SEQ ID NO: 4)
Txx ATG Ttt (SEQ ID NO: 5)
Cxx ATG Ttt (SEQ ID NO: 6)

where x are nucleotides. SEQ. ID. NO. 3 is most preferred; and xx are preferably CC.

Other transcriptional cassettes, which may or may not include a fully impaired consensus Kozak, can be

SUBSTITUTE SHEET (RULE 26)

This Sheet Has Been Replaced



FEB-19-96 22:47 FROM: ID: PAGE 23

incorporated into a vector which includes transcriptional cassettes containing the disclosed and claimed fully impaired consensus Kozak; such "other transcriptional cassettes" typically are utilized to allow for "enhancement," "amplification" or "regulation" of gene product repression. For example, co-transfection of the exogenous DNA with the dehydrofolate reductase (DHFR) gene is exemplary. By increasing the levels the antifolate drug methotrexate (MTX), a competitive inhibitor of DHFR, presented to such cells, an increase in DHFR production can occur via amplification of the DHFR gene. Beneficially, extensive amounts of flanking exogenous DNA will also become amplified; therefore, exogenous DNA inserted co-linear with an expressible DHFR gene will also become overexpressed. Additionally, transciptional cassettes which allow for regulation of expression are available. For example, temperature sensitive COS cells, derived by placing SV40ts mutant large T antigen gene under the direction of Rous sarcoma virus LTR (insensitive to feedback repression by T antigen), has been described. *See*, 227 *Science* 23-28 (1985). These cells support replication from SV40 *ori* at 33°C but not at 40°C and allow regulation of the copy number of transfected SV40 *ori*-containing vectors. The foregoing is not intended, nor is it to be construed, as limiting; rather the foregoing is intended to be exemplary of the types of cassettes which can be incorporated into expression vectors comprising the disclosed fully impaired consensus Kozak. The skilled artisan is credited with the ability to determine the specific type of other transciptional cassettes, vis-a-vis the objective of the expression system, which are applicable and which can be advantageously exploited.

As indicated above, in particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without a stop codon being located between the out-of-frame start codon and the

This Sheet Has Been Replaced



FEB-19-96 22:47 FROM: ID: PAGE 24

fully impaired consensus Kozak start codon. The intent of the out-of-frame start codon is to, in effect, further impair translation of the dominant selectable marker.

As used herein, the term "stop codon" is meant to indicate a "nonsense codon," *ie* a codon which does not encode one of the 20 naturally occurring amino acids such that translation of the encoded material terminates at the region of the stop codon. This definition includes, in particular, the traditional stop codons TAA, TAG and TGA.

As used herein, the term "out-of-frame" is relative to the fully impaired consensus Kozak start codon. As those in the art appreciate, in any DNA macromolecule (or RNA macromolecule) for every in-frame sequence, there are two out-of-frame sequences. Thus, for example, with respect to the following sequence incorporating a fully impaired consensus Kozak:

```
                    -3      +1
          gcA TGc cAT Ggc Cxx ATG Cxx
```

the in-frame codons are separated by triplets, *eg*, gcA, TGc, cAT and Ggc; the out-of-frame codons would include, *eg* cAT, ATG, Gcc, ccA, ATG and TGg. Thus, two start codons (in capital letters and underlined) are out-of-frame relative to the start codon of the fully impaired consensus Kozak.

When such an out-of-frame start codon is utilized, it is preferred that this be within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides of the fully impaired consensus Kozak start codon.

As is appreciated, the upstream sequence can be manipulated to achieve positioning at least one out-of-frame sequence upstream of the fully impaired consensus Kozak start codon using (most preferably) a mutational primer used in the type of amplification protocol described above.

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:47 FROM: ID: PAGE 25

Utilization of a fully impaired consensus Kozak start codon located within a secondary structure (*ie* a "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable marker. In such an embodiment, it is preferred that this start codon be located within the stem of a stem loop secondary structure. These, by way of schematic example, in such an embodiment, the start codon of the fully impaired consensus Kozak is positioned as follows:

```
          T
        x   x
         x x
         x A
         x T
         x G
      xA TGxx Cxx
```

(An out-of-frame start codon which is not part of the secondary structure is also represented.) As is appreciated, within the stem loop, complementarity along the stem is, by definition, typically required. For exemplary methodologies regarding, *inter alia*, introduction of such secondary structures into the sequence, as well as information regarding secondary structure stability, *see*, Kozak, *PNAS*, 1986, supra.

As noted, it is preferred that a dominant selectable market with a naturally occurring intronic insertion region or an artificially created intronic insertion region be utilized and at least one gene product of interest inserted within this region. While not wishing to be bond by any particular theory, the inventor postulates that such an arrangement increases expression efficiency because the number of viable colonies that survive the selection process via-a-vis the dominant selectable marker will decrease; the colonies that do survive the selection process will, by definition, have expressed the protein necessary for survival, and in conjunction therewith, the gene product of interest will

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:48 FROM: ID: PAGE 28

have a greater tendency to be expressed. As further postulated, the RNA being transcribed from the gene product of interest within the intronic insertion region interferes with completion of transcription (elongation of RNA) of the dominant selectable marker; therefore, the position that the dominant selectable marker is integrated within the cellular DNA is likely to be a position where a larger amount of RNA is initially transcribed.

As is appreciated, prokaryotic proteins do not typically include splices and introns. However, the majority of dominant selectable markers which are preferred for expression vector technology are derived from prokaryotic systems. Thus, when prokaryotic-derived dominant selectable markers are utilized, as is preferred, it is often necessary to generate an artificial splice within the gene so to create a location for insertion of an intron comprising the gene product of interest. It is noted that while the following rules are provided for selection of a splice site in prokaryotic genes, they can be readily applied to eukaryotic genes.

A general mechanism for the splicing of messenger RNA precursors in eukaryotic cells is delineated and summarized in Sharp, Philip A. "Splicing of Messenger RNA Precursors" *Science*, *235*: 736-771 (1987) (*see*, in particular, Figure 1) which is incorporated herein by reference. Based upon Sharp, there are four minimum criteria in the nucleic acid sequence which are necessary for a splice: (a) 5' splice donor; (b) 3' splice acceptor; (c) branch point, and (d) polypyrimidine tract. The consensus sequences for the 5' splice donor is reported to be

$$\begin{matrix} C & & A \\ AAG/GTGAGT & & \end{matrix}$$

and for the 3' splice acceptor, NCAG/G (where a "/" symbol indicates the splice site); *see*, Mount, S.M. "A Catalogue of Splice Junction Sequences" *Nuc. Acids. Res. 10/2*:459-472 (1992) which is incorporated herein by reference. The consensus sequence for the branch point, *ie.*the location of the lariat formation with the 5' splice donor, is

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:48 FROM: ID: PAGE 27

reported as PyNPyPAPy; and the reported preferred branch site for mammalian RNA splicing is TACTAAC (Zhuang, Y. et al. "UACUAAC is the preferred branch site for mammalian mRNA splicing" *PNAS 86*: 2752-2756 (1989), incorporated herein by reference). Typically, the branch point is located at least approximately 70 to about 80 base-pairs from the 5'-splice donor (there is no defined upper limit to this distance). The poly pyrimidine tract typically is from about 15 to about 30 base pairs and is most typically bounded by the branch point and the 3' splice acceptor.

The foregoing is descriptive of the criteria imposed by nature on naturally occurring splicing mechanisms. Because there is no exact upper limit on the number of base pairs between the 5' splice donor and the branch point, it is preferred that the gene product of interest be inserted within this region in situations where a natural intron exists within the dominant selectable marker. However, as noted, such introns do not exist within most of the preferred dominant selectable markers; as such, utilization of artificial introns are preferably utilized with these markers.

In order to generate an artificial intron, a "splice donor:splice acceptor" site must be located within the encoding region of the dominant selectable marker. Based upon Sharp and Mount, it is most preferred that the following sequence function as the splice donor:splice acceptor site -- CAGG (with the artificial splice occurring at the GG region). A preferred sequence is AAGG.

Focusing in on the most preferred sequence CAGG, the following codons and amino acids can be located within the encoding region of the dominant selectable marker for generation of the artificial intronoc insertion region:

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:48 FROM: ID: PAGE 28

A		B		C	
CAG/	GNN	NCA	G/GN	NNC	AG/G
Gln	Ala	Ala	Gly	Ala	Leu Arg
	Asp	Pro		Arg	Phe
	Gly	Ser		Asn	Pro
	Glu	Thr		Asp	Ser
	Val			Cys	Thr
				Gly	Tyr
				His	Val
					Ile

(As will be appreciated, the same approach to determining viable amino acid residues can be utilized for the preferred sequence of AAGG). The most preferred codon group for derivation of the splice donor:splice acceptor site is group A. Once these amino acid sequences are located, a viable point for generation of an artificial intronic insertion region can be defined.

Focusing on the preferrred NEO dominant selectable marker, amino acid residues Gln Asp (codon group A) are located at the positions 51 and 52 of NEO and amino acid residues Ala Arg (codon group C) are located at positions 172 and 173 (as is appreciated, multiple artificial intronic insertion regions may be utilized). Focusing on residues 50 - 53 of NEO, the nucleic acid and amino acid sequences are as follows:

	50	51	52	53	
5'	CTG	CAG	GAC	GAG	3'
	Leu	Gln	Asp	Glu	

Accordingly, an artificial intronic insertion region can be generated between residues 51 and 52 of NEO. This region most preferably comprises a branch point, a polyprimidine tract and, preferably, a region for insertion of a gene product of interest, *ie* a region amendable to enzymatic digestion.

Two criteria are import for the artificial intronic insertion region: the first two nucleic acid residues of the 5' splice site (*eg* abutting CAG) are most preferably

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:48 FROM: ID: PAGE 29

GT and the first two nucleic acid residues of the 3' splice site (*eg* abutting G) are most preferably AG.

Using the criteria defined abouve, an artificial intronic insertion region was between amino acid residues 51 and 52 of NEO:

50 NEO 51 52 NEO 53

LEU GLN Branch Polypyrimide Tract ASP GLU 3'

5' CTG CAG/GTAAGT GCGGCCGC TACTAAC $(TC)_3CT(C)_3TCC(T)_5$ C CTGCAG/GAC GAG

Not I

(Details regarding the methology for creating this artificial intronic insertion region are set forth in the Example Section to follow). The Not I site was created as the region where the gene product of interest can be incorporated. Therefore, upon incorporation, the gene product of interest is located between amino acid residues 51 and 52 of NEO, such that during NEO transmission, the gene product of interest will be "spliced-out".

The host cell line is most preferably of mammalian origin; those skilled in the art are credited with ability to preferentially determine particular host cell lines which are best suited for the desired gene product to be expressed therein. Exemplary host cell lines include, but are not limited to, DG44 and DXB11 (Chinese Hamster Ovary lines, DHFR minus), HELA (human cervical carcinoma), CV1 (monkey kidney line), COS (a derivative of CV1 with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3x63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte) and 293 (human kidney). Host cell lines are typically available from commercial services, the American Tissue Culture Collection or from published literature.

Preferably the host cell line is either DG44 or SP2/O. *See*, Urland, G. *et al.*, "Effect of gamma rays and the dihydrofolate reductase locus: deletions and inversions." *Som. Cell & Mol. Gen. 12/6*:555-566 (1986) and Shulman, M. *et al.*, "A better cell line for making

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

PAGE 30 ID: FEB-19-96 22:49 FROM:

hybridomas secreting specific antibodies." *Nature 276:269* (1978), respectively. Most preferably, the host cell line is DG44. Transfection of the plasmid into the host cell can be accomplished by any technique available to those in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), cell fusion with enveloped DNA, microinjection, and infection with intact virus. *See*, Ridgway, A.A.G. "Mammalian Expression Vectors." Chapter 24.2, pp. 470-472 *Vectors*, Rodriguez and Denhardt, Eds. (Butterworths, Boston, MA 1988). Most preferably, plasmid introduction into the host is via electroporation.

Examples

The following examples are not intended, nor are they to be construed, as limiting the invention; the examples are intended to demonstrate the applicability of an embodiment of the invention disclosure herein. The disclosed fully impaired consensus Kozak sequence is intended to be broadly applied as delineated above. However, for presentational efficiency, exemplary uses of particularly preferred embodiments of fully impaired consensus Kozak sequences are utilized in conjunction with tandem chimeric antibody expression vectors (also referred to herein as antibody expression vectors) as disclosed below.

I. Tandem Chimeric Antibody Expression ("TCAE") Vector

B cell lymphocytes arise from pluripotent stem cells and proceed through ontogeny to fully matured antibody secreting plasma cells. The human B lymphocyte-restricted differentiation antigen Bp35, referred to in the art as "CD20," is a cell surface non-glycosylated phosphoprotein of 35,000 Daltons; CD20 is expressed during early pre-B cell development just prior to the expression of cytoplasmic μ heavy chains. CD20 is expressed consistently until the plasma cell differentiation stage. The CD20 molecule regulates a step in the activation

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

PAGE 31 ID: FEB-19-96 22:49 FROM:

process which is required for cell cycle initiation and differentiation. Because CD20 is expressed on neoplastic B cells, CD20 provides a promising target for therapy of B cell lymphomas and leukemias. The CD20 antigen is especially suitable as a target for anti-CD20 antibody mediated therapy because of accessibility and sensitivity of hematopoietic tumors to lysis via immune effector mechanisms. Anti-CD20 antibody mediated therapy , *inter alia*, is disclosed in co-pending Serial No. 07/978,891 and Serial No. _____ , filed simultaneously herewith. The antibodies utilized are mouse/human chimeric anti-CD20 antibodies expressed at high levels in mammalian cells (chimeric anti-CD20"). This antibody was derived using vectors disclosed herein, to wit: TCAE 5.2; ANEX 1; ANEX 2; GKNEOSPLA3F; and NEOSPLA3F (an additional vector, TCAE 8, was also utilized to derive chimeric anti-CD20 antibody -- TCAE 8 is identical to TCAE 5.2 except that the NEO translational start site is a partially impaired consensus Kozak. TCAE 8 is described in the co-pending patent document filed herewith.).

In commonly-assigned United States Serial Number 07/912,292, disclosed, *inter alia*, are human/Old World monkey chimeric antibodies; an embodiment of the invention disclosed therein are human/macaque chimeric anti-CD4 antibodies in vector TCAE 6 (*see*, Figure 6 of Serial Number 07/912,292, and corresponding discussion). TCAE 6 is substantially identical to TCAE 5.2; TCAE 6 contains human lambda constant region, while TCAE 5.2 contains human kappa constant region. TCAE 5.2 and ANEX 1 (referred to in that patent document as TCAE 12) are disclosed as vectors which can be utilized in conjunction with human/Old World monkey chimeric antibodies. The comparative data set forth in Serial No. 07/912,292 vis-a-vis TCAE 5.2 and ANEX 1 is relative to expression of chimeric anti-CD20 antibody.

TCAE 5.2 was derived from the vector CLDN, a derivative of the vector RLDN10b (*see*, 253 *Science* 77-91, 1991). RLDN10b is a derivative of the vector TND (*see*,

This Sheet Has Been Replaced



PAGE 32 ID: FEB-19-96 22:49 FROM:

7*DNA* 651-661, 1988). *Ie* the vector "family line" is as follows: TDN → RLDN10b → CLDN → TCAE 5.2 → ANEX 1 (the use of the "→" symbol is not intended, nor is it to be construed, as an indication of the effort necessary to achieve the changes from one vector to the next; *eg* to the contrary, the number and complexity of the steps necessary to generate TCAE 5.2 from CLDN were extensive).

TND was designed for high level expressions of human tissue plasminogen activator. RLDN10b differs from TND in the following ways: the dihydrofolate reductase ("DHFR") transcriptional cassette (comprising promoter, murine DHFR cDNA, and polyadenylation region) was placed in between the tissue plasminogen activator cassette ("t-PA expression cassette") and the neomycin phosphotransferase ("NEO" cassette") so that all three cassettes were in tandem and in the same transcriptional orientation. The TND vector permitted selection with G418 for cells carrying the DHFR, NEO and t-PA genes prior to selection for DHFR gene amplification in response to methotrexate, MTX. The promoter in front of the DHFR gene was changed to the mouse beta globin major promoter (*see*, 3 *Mol. Cell Bio.* 1246-1254, 1983). Finally, the t-PA cDNA was replaced by a polylinker such that different genes of interest can be inserted in the polylinker. All three eukaryotic transcriptional cassettes (t-PA, DHFR, NEO) of the TND vector can be separated from the bacterial plasmid DNA (pUC19 derivative) by digestion with the restriction endonuclease Not I.

CLDN differs from RLDN10b in the following ways: The Rous LTR, positioned in front of the polylinker, was replaced by the human cytomegalovirus immediate early gene promoter enhancer ("CMV"), (*see*, 41 *Cell* 521, 1985), from the Spe I site at -581 to the Sst I site at -16 (these numbers are from the *Cell* reference).

As the name indicates, TCAE vectors were designed for high level expressions of chimeric antibody. TCAE 5.2 differs from CLDN in the following ways:

SUBSTITUTE SHEET (RULE 26)

This Sheet Has Been Replaced



FEB-19-96 22:50 FROM: ID: PAGE 33

A. TCAE 5.2 comprises four (4) transcriptional cassettes, as opposed to three (3), and these are in tandem order, *ie* a human immunoglobulin light chain absent a variable region; a human immunoglobulin heavy chain absent a variable region; DHFR; and NEO. Each transcriptional cassette contains its own eukaryotic promoter and polyadenylatin region (reference is made to Figure 2 which is a diagrammatic representation of the TCAE 5.2 vector). The CMV promoter/enhancer in front of the immunoglobulin heavy chain is a truncated version of the promoter/enhancer in front of the light chain, from the Nhe I site at -350 to the Sst I site at -16 (the numbers are from the *Cell* reference, *supra*). Specifically,

1) A human immunoglobulin light chain constant region was derived via amplification of cDNA by a PCR reaction. In TCAE 5.2, this was the human immunoglobulin light chain kappa constant region (Kabat numbering, amino acids 108-214, allotype Km 3), and the human immunoglobulin heavy chain gamma 1 constant region (Kabat numbering amino acids 114-478, allotype Gm1a, Gm1z). The light chain was isolated from normal human blood (IDEC Pharmaceuticals Corporation, La Jolla, CA); RNA therefrom was used to synthesize cDNA which was then amplified using PCR techniques (primers were derived vis-a-vis the consensus Kabat). The heavy chain was isolated (using PCR techniques) from cDNA prepared from RNA which was in turn derived from cells transfected with a human IgG1 vector (*see*, 3 *Prot. Eng.* 531, 1990; vector $pN_{\gamma 1}62$). Two amino acids were changed in the isolated human IgG1 to match the consensus amino acid sequence in Kabat, to wit: amino acid 225 was changed from valine to alanine (GTT to GCA), and amino acid 287 was changed from methionine to lysine (ATG to AAG);

2) The human immunoglobulin light and heavy chain cassettes contain synthetic signal sequences for secretion of the immunoglobulin chains;

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:50 FROM: ID: PAGE 34

3) The human immunoglobulin light and heavy chain cassettes contain specific DNA restriction sites which allow for insertion of light and heavy immunoglobulin variable regions which maintain the transitional reading frame and do not alter the amino acids normally found in immunoglobulin chains;

4) The DHFR cassette contained its own eukaryotic promoter (mouse beta globin major promoter, "BETA") and polyadenylation region (bovine growth hormone polyadenylation, "BGH"); and

5) The NEO cassette contained its own eukaryotic promoter (BETA) and polyadenylation region (SV40 early polyadenylation, "SV").

With respect to the TCAE 5.2 and the NEO cassette, the Kozak region was a consensus Kozak (which included an upstream Cla I site SEQ ID NO: 7):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccAcc ATG Gtt

ANEX 1 (previously named TCAE 12 in the referenced case) is identical to TCAE 5.2 except that in the NEO cassette, the Kozak region was fully impaired (SEQ ID NO: 8):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccTcc ATG Ctt

As disclosed in the commonly-assigned referenced case, the impact of utilization of the fully impaired consensus Kozak was striking: relative to TCAE 5.2, there was a significant (8-fold) reduction in the number of ANEX 1 G418 resistant colonies (258 from two electroporations versus 98 from six electroporations) from the same amount of plasmid DNA transfected per cell; and, there was a significant increase in the amount of co-linked gene product expressed in each of the ANEX 1 clones. Referencing the histogram of Figure 3 (Figure 16 of the commonly assigned referenced case), 258 colonies were derived from 2 electroporations of 25 µg of DNA containing a neomycin phosphotransferase gene with a consensus Kozak at the translation start site. Two-hundred and one (201)

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:50 FROM: ID: PAGE 35

of these colonies did not express any detectable gene product (less than 25 ng/ml of chimeric immunoglobulin), and only 8 colonies expressed more than 100 ng/ml. Again, referencing Figure 3, 98 colonies were derived from 6 electroporations for ANEX 1 of 25 µg of DNA containing a neomycin phosphotransferase gene with the fully impaired consensus Kozak at the translation start site (6 electroporations were utilized in order to generate statistically comparative values; this was because on average, each electroporation for ANEX 1 yielded about 16 colonies, as opposed to about 129 colonies per electroporation for TCAE 5.2). Eight (8) of the ANEX 1 colonies did not express any detectable gene product (less than 25 ng/ml), while 62 of these colonies were expressing greater than 100 ng/ml; of these 62 colonies, nearly 23 were expressing over 250 ng/ml (23%), with 6 expressing greater than 1000 ng/ml (6%).

The foregoing evidences, *inter alia*, the following: 1) because the difference between TCAE 5.2 and ANEX 1 was limited to the Kozak translation start site of the NEO gene, and because the gene product of interest (chimeric anti-CD20 antibody) was co-linked to the NEO gene, a conclusion to be drawn is that these differences in results are attributed solely to the differences in the Kozak translation start site; 2) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable number resulted in significantly less viable colonies; 3) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable marker co-linked to a desired gene product significantly increased the amount of expressed gene product. Thus, the number of colonies to be screened decreased while the amount of expressed gene product increased.

This Sheet Has Been Replaced



FEB-19-96 22:50 FROM: ID: PAGE 36

II. Impact Of Out-Of-Frame Start Sequence

Conceptually, further impairment of translation initiation of the dominant selectable marker of ANEX 1 could be effectuated by utilization of at least one out-of-frame ATG start codon upstream of the neomycin phosphotransferase start codon. Taking this approach one step further, utilization of a secondary structure ("hairpin") which incorporated the *neo* start codon within the stem thereof, would be presumed to further inhibit translation initiation. Thus, when the out-of-frame start codon/fully impaired consensus Kozak was considered, this region was designed such that the possibility of such secondary structures was increased.

As indicated previously, the Kozak region for the *neo* start codon in the ANEX 1 vector is:

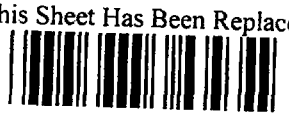
TTGGGAGCTTGG ATCGAT CC Tcc ATG Ctt

The desired sequence for a vector identical to ANEX 1 but incorporating the above-identified changes vis-a-vis the *neo* start codon, referred to as ANEX 2, is as follows (SEQ ID NO: 9):

CCA GCA TGG AGG A ATCGAT CC Tcc ATG Ctt

(The out-of-frame start codon is underlined.) The fully impaired consensus Kozak of ANEX 2 is identical to that of ANEX 1. The principal difference is the inclusion of the upstream out-of-frame start codon. A possible difference is the formation of a secondary structure involving this sequence, proposed as follows:

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:51 FROM: ID: PAGE 37

CG)
T A) CLA I site
A T)
AT
GC
GC
AT
GC
GC
TA
AT
CG
GC
AT
CT
CG

The sequence in bold, ATG, is the upstream out-of-frame start codon; the "loop" portion of the secondary structure is the CLA I site; and the sequence between the "T" and "C" (italics and bold) is the start codon (underlined) of the fully impaired consensus Kozak.

In order to effectuate this change, a PCR fragment was cloned into anti-CD20 in ANEX 1 from Xho I (5520) to Cla I (5901); see, Figure 4. Primers were as follows:

3'-Primer 489 (SEQ ID NO: 10):

5'-GGA GGA TCG ATT CCT CCA TGC TGG
CAC AAC TAT GTC AGA AGC AAA TGT
GAG C-3'

The upper-lined portion of Primer 489 is a Cla I site; the under-lined portion is the fully impaired consensus Kozak translation start site.

5'-Primer 488 (SEQ. ID. NO. 11):

5'-CTG GGG CTC GAG CTT TGC-3'

The upper-lined portion of Primer 485 is an Xho I site.

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

These primers were prepared using an ABI 391 PCR MATE™ DNA synthesizer (Applied Biosystems, Foster City, CA). Phosphoramidites were obtained from Cruachem (Glasgow, Scotland): dA(bz) - Prod. No. 20-8120-21; dG(ibu) - Prod. No. 20-8110-21; dC(bz) - Prod. No. 20-8130-21; T - Prod. No. 20-8100-21.

Conditions for the PCR reaction using these primers were as follows: 2λ ("microliters") of anti-CD20 in TCAE 5.2 in plasmid grown in *E. coli* strain GM48 (obtained from the ATCC) was admixed with 77λ of deionized water; 2λ of Primer 488 (64 pmoles); and 4λ of primer 489 (56 pmoles). This was followed by a denaturation step (94°C, 5 min.) and a renaturation step (54°C, 5 min.). Thereafter, 4λ of 5 mn dNTPS (Promega, Madison, WI: dATP, Prod. No. U1201; dCTP, Prod. No. U1221; dGTP, Prod. No. U1211; dTTp, Prod. No. U1231), 1λ of Pfu DNA polymerase (Stratagene, La Jolla, CA Prod. No. 600135, 2.5 U/ml), and 50λ of mineral oil overlay was added thereto, followed by 30 cycles, with each cycle comprising the following: 72°C, 2 min.; 94°C, 1 min.; 54°C, 1 min. Ten microliters (10λ) of this admixture was analyzed by agarose gel electrophoresis (results not shown); a single band was found at about 400 base pairs.

The PCR product and the vector were prepared for ligation as follows: Anti-CD20 in ANEX 1 plasmid grown in *E. coli* bacterial strain GM48 was digested with Cla 1 and Xho 1 as follows: 20λ of anti-CD20 in ANEX 1 was admixed with 10λ of 10XNEB4 buffer (New England Biolabs, Beverly, MA; hereinafter, NEB); 5λ Cla 1 (NEB, Prod. No. 197 S, 60u); and 64λ deionized water. This admixture incubated overnight at 37°C, followed by the addition of 5λ Xho 1 (NEB, Prod. No. 146 S, 100 u) and incubation at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut ANEX 1". The approximate 400 base pair PCR fragment was prepared and digested with Cla 1 and Xho 1 as follows: 90λ of the PCR fragment was admixed with 10λ of 3M NaOAc; 1λ 10% sodium dodecyl sulfate (SDS); and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30

This Sheet Has Been Replaced SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:51 FROM: ID: PAGE 39

sec. followed by a 1 min. spin (1700 RPM). The aqueous phase was subjected to a spin column which resulted in 85λ total admixture. To this admixture was added 10λ 10XNEB4, 1λ bovine serum albumin (BSA,100X; NEB), 2λ Cla 1 (24u), and 2λXho 1 (40 u). This admixture was incubated at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut PCR 488/489". Both Cla 1/Xho 1 cut ANEX 1 and Cla 1/Xho 1 cut PCR 488/489 were analyzed by agarose gel electrophoresis and the resulting bands were observed at the same relative location on the gel (results not shown).

Ligation of Cla 1/Xho 1 cut PCR 488/489 and Cla 1/Xho 1 cut ANEX 1 was accomplished as follows: 1λ of tRNA (Sigma, St. Louis, MO, Prod. No. R-8508) was admixed with 1λ 10% SDS; 10λ 3M NaOAc; 45λ of Cla 1/Xho 1 cut PCR 488/489 (about 22.5 ng); a 1:4 dilution (0.25λ) of Cla 1/Xho 1 cut ANEX 1 (about 32 ng) in 0.75λ tris-hydroxymethyl aminomethane ethylenediamine tetracetic acid (TE); and 42λ TE. To this admixture was added 90λ phenyl/$CHCl_3$/isoamyl, followed by a 30 sec. vortex and a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube, followed by addition of 270λ of 100% EtOH (-20° C), 10 min. spin (13,000 RPM) followed by addition of another 270λ of 100% EtOH (-20° C) and another 1 min. spin (13,000 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 17λ TE, 2λ ligase buffer (Promega, T4 DNA Ligase kit, Prod. No. M180) and 1λ ligase (Promega Ligase kit). This ligation mix incubated at 14°C overnight. Twenty microliters (20λ) of the ligation mix was admixed with 10λ 3M NaOAc, 1λ 10% SDS, 69λ TE, and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30 sec., followed by a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube and 270λ of 100% EtOH (-20°C) was added thereto, followed by a 10 min. spin (1700 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 20λ TE. Ten microliters of the resuspended admixture was transformed in *E. coli* X-L1 blue™ (Stratagene, La Jolla, CA), following manufacturer

This Sheet Has Been Replaced



PAGE 40 ID: FEB-19-98 22:52 FROM:

instructions. Ten (10) bacterial colonies were inoculated in LB Broth (Gibco BRL, Grand Island, NY, Prod. No. M27950B) including ampicillin (50µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification System (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the ANEX 2 vector, depending on the sufficiency of the foregoing.

ANEX 2 includes a Hinf I site ("GAATC") upstream of the *neo* start site (-9 to -13 relative to the *neo* start codon); ANEX 1 does not include this Hinf I site. The purified plasmids comprising putative ANEX 2, and previously purified ANEX 1 standard , were subjected to Hinf 1 digestion as follows: 2λ of each isolate was admixed with 8λ of Hinf I digestion buffer (15λ 10 x NEB2 buffer; 15λ Hinf I (NEB, Prod. No. 155S, 10u/λ); and 90λ H_2O). This admixture incubated for 3 hrs. at 37°C and each isolate was analyzed via agarose gel electrophoresis (results not shown); nine (9) of the bands were substantially identical to the ANEX 1 standard, one(1) showed a slight difference in band pattern. For this single isolate, the first two bands were at 1691 and 670 kB; for the ANEX 1 Hinf I digested product, the first three bands were at 1691, 766, and 670 kB. The missing band at 766 kB for the single isolate was attributed to the presence of the Hinf I site therein, indicating that the desired change to ANEX 1 was incorporated into this vector. This vector was designated "Anti-CD20 in ANEX 2 (G1,K)," and is generally referred to by the inventor as ANEX 2.

Electroporation of anti-CD20 in ANEX 2 was accomplished as follows: two-hundred and forty microliters (240λ) of the anti-CD20 in ANEX 2 DNA (400µg) was admixed with 100λ of 10 X NEB2 buffer ; 100λ of Stu I (NEB, Prod. No. 187S, 1000 u); and 560λ TE, and incubated at 37°C for 2hrs. This admixture was then placed over 8 spin columns (125λ each), followed by addition of 110λ 10X Not I buffer (NEB); 10λ 100X BSA; and 20λ of Not I (NEB,

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:52 FROM: ID: PAGE 41

Prod. No. 189S, 800u). This admixture was incubated at 37°C for 3 hrs., followed by the addition of 120λ of 3M NaOAC and 12λ of 10% SDS. The admixture was transferred to 2 vortex tubes and 500λ of phenyl/$CHCl_3$/isoamyl was added to each, followed by a 30 sec. vortex and 1 min. spin (1700 RPM). The aqueous phase was removed from the tubes and segregated into 3 tubes, followed by the addition to each tube of -20°C 100% ETOH, followed by 10 min. spin (13,000 RPM). Thereafter, -20°C 70% ETOH was added to each tube, followed by 1 min. spin (13,000 RPM). The tubes were then placed in a Speed VAC™ for drying, followed by resuspension of the contents in 100λ TE in a sterile hood. Five microliters (5λ) of the resuspended DNA was admixed with 995λ of deionized water (1:200 dilution). An optical density reading was taken (OD=260) and the amount of DNA present was calculated to be 0.75μg/λ. In order to utilize 25 μg of DNA for electroporation, 32λ of the 1:200 dilution of the DNA was utilized (25 μg was utilized as this was the amount of DNA utilized for TCAE 5.2 and ANEX 1 in the foregoing Example 1). The 1:200 dilution of DNA was formally referred to as "Stu 1, Not I cut anti-CD20 in ANEX 2 (25μg) in TE" and generally referred to as "anti-CD20 in ANEX 2."

Host cells utilized was DG44 CHO ("CHO") (*see*, Urlaub, G. *Somatic Cell*, 1986 *supra*). One hundred milliliters of 6.6×10^5 cells/ml (84%) were subjected to a 2 min. spin at 1000 RPM. These were washed with 50 ml sucrose buffered solution, followed by 5 min. spin at 1000 RPM; the material was then resuspended in 4.5 ml of the sucrose buffered solution. Thereafter, cells were counted and 0.4 ml of CHO cells (4.0×10^6 cells) were admixed with 32λ of the anti-CD20 in ANEX 2 in BTX sterile, disposable electroporation cuvettes. Electroporation settings were as follows: 210 volts; 400 microfaraday capacitance; 13 ohms resistance, using a BTX 600™ electro cell manipulator (BTX, San Diego, CA). Nine (9) electroporations were conducted; actual voltage delivered over actual times were as follows: 1-199V, 4.23 msec; 2-

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:52 FROM: ID: PAGE 42

188V, 4.57 msec; 3-189V, 4.24 msec, 4-200V; 4.26 msec, 5-200V, 4.26 msec; 6-199V, 4.26 msec; 7-189V, 4.59 msec; 8-189V, 4.57 msec; 9-201 V, 4.24 msec. (As noted in Example I, the difference in number of performed electroporations was attributed to the need to achieve a statistically significant number of viable colonies for each of the three conditions, TCAE 5.2, ANEX 1 and ANEX 2; the amount of DNA used for each electroporation (25 µg) was the same for each, and the same number of cells were electroporated.

Thereafter, the electroporation material was admixed with 20 ml of G418 Growth Media (CHO-S-SFM II minus hypoxanthine and thymidine (Gibco, Grand Island, NT, Form No. 91-0456PK) including 50 µM hypoxanthine and 8 µM thymidine). The admixture was gently agitated, followed by plating 200 µl of the admixture per well into 96-well plates, one plate for each electroporation (nine). Beginning on day 2 after electroporation, through day 17, 150µl of each well was removed, and 150 µl of fresh G418 Growth Media containing 400 µg/ml G418 was added thereto. Colonies were analyzed on day 25.

One hundred and twenty one (121) colonies expressed anti-CD20 antibody (*ie* , 13 colonies per electroporation). Of these, 63 (52%) expressed over 250µg /ml of protein; of the 63, 20 of the colonies (16.5%) expressed over 1000µg/ml of protein. Only 5 of the 121 colonies (4.1%) expressed less than 25µg/ml of protein. Figure 5 provides a histogram comparing expression of protein per colonies derived from the vectors TCAE 5.2, ANEX 1 and ANEX 2.

The foregoing data indicates that, *inter alia*, as between ANEX 1 and ANEX 2, the use of at least one out-of-frame start codon upstream of a fully impaired consensus Kozak associated with the translation initiation of a dominant selectable marker decreases the number of viable colonies expressing co-linked gene product and significantly increases the amount of expressed co-linked gene product.

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:53 FROM: ID: PAGE 43

III. Impact Of Insertion Of Gene Product Of Interest Within At Least One Artificial Intronic Insertion Region Of A Dominant Selectable Marker

Building further upon the ANEX 2 vector, an artificial splice was generated between amino acid residues 51 and 52 of the NEO coding region of ANEX 2, followed by insertion therein of the anti-CD20 encoding region. Two such vectors were generated: the first, comprising a consensus Kozak sequence for the NEO translation initiation codon and not comprising an out-of-frame start codon, is referred to as "GKNEOSPLA3F;" the second, comprising a fully impaired consensus Kozak and an out-of frame start condon, is referred to as "NEOSPLA3F."

Both GKNEOSPLA3F and NEOSPLA3F contain the following artificial intron sequence between amino acid residues 51 and 52 of NEO:

51 52

5'CTG CAG/GTAAGT GCGGCCGC TACTAAC $(TC)_3$ CT $(C)_3$ TCC $(T)_5$ C CTGCAG/GAC GAG 3'

The underlined portion represents a sequence amenable to digestion with Not I enzyme; the encoding region for anti-CD20, *inter alia*, was inserted within this region.

Although not wishing to be bound by any particular theory, the inventor postulates that during expression, the inclusion of a gene product of interest within an artificial intronic insertion region of a dominant selectable marker (*e.g.* the NEO gene) should significantly decrease the number of viable colonies producing, in the case of the disclosed GKNEOSPA3F and NEOSPLA3F vectors, anti-CD20 antibody. This is predicated upon two points: first, only those vectors which are able to transcribe and correctly splice-out the antibody encoding region and correctly translate NEO will be G418 resistant; second, because each antibody cassette has its own promoter and polyadenylation region, transcription and translation of the antibody is independent of translation of NEO.

The GKNEOSPLA3F and NEOSPLA3F vectors were constructed in the following manner:

This Sheet Has Been Replaced



FEB-19-96 22:53 FROM: ID: PAGE 44

Anti-CD20 in ANEX 2 was digested with Not I and Xho I in order to isolate the 1503 bp NEO cassette DNA fragment (see Figure 4 between "Not I 7023" and "Xho I 5520") as follows: 10 µl of anti-CD20 in ANEX 2 was admixed with 6 µl deionized H_2O ("dH_2O"); 1 µl Not I enzyme (NEB, Prod. No. 189S); 2 µl of 10X Not I digestion buffer (NEB; provided with enzyme); and 1 µ Xho I enzyme (Promega, Madison, WS, Prod. No. R4164). This digestion mixture was incubated overnight at 37°C. The resulting digested DNA was size fractionated by 0.8% agarose gel electrophoresis and the desired fragment migrating at 1503 was isolated via the GlassMAX™ method (Gibco BRL, Grand Island, NY, Prod. No. 15590-011) for insertion into pBluescript SK(-) plasmid DNA (Stratagene, La Jolla, CA).

pBluescript SK (-) was previously prepared for acceptance of the NEO cassette by double digestion with Not I and Xho I using the same conditions as above for anti-CD20 in ANEX 2. Digested pBluescript SK (-) was then collected by ethanol precipitation by the addition of 70 µl dH_2O; 2 µl tRNA (Sigma, St. Louis, MO, Prod. No. R-8508); 10 µl of 3M NaOAc; and 300 µl 100% ETOH (-20°C). This was followed by a 10 min spin (13,000 RPM), decanting the supernatant, rinsing with 70% ETOH, decanting the liquid, drying in a SpeedVAC™ and resuspending in 20 µl 1 x TE.

Ligation of the NEO cassette DNA fragment into prepared pBluescript SK (-) vector was accomplished as follows: 10 µl of NEO fragment DNA was admixed with 6 µl dH_2O; 1 µl cut pBluescript SK (-) vector DNA; 2 µl 10 x ligation buffer (Promega, supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega, Prod. No. M1801) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:53 FROM: ID: PAGE 45

M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO+ depending on the sufficiency of the foregoing. (BlueNEO+ was confirmed due to the sufficiency of the following procedure.)

BlueNEO+ contains a Not I restriction recognition sequence reformed upon ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector. This site was destroyed by the following: 1 µl of BlueNEO+ DNA was admixed with 16 µl dH_2O; 2 µ 10 x Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB). This was followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume. This 15 µl Not I digested DNA was "blunt-ended" by admixing with 4 µl 5X Klenow buffer (20 mM Tris-HCL, pH 8.0, 100mM $MgCl_2$) and 1 µl DNA Polymerase I Large (Klenow) Fragment (Promega, Prod. No. M2201). This admixture was incubated at room temperature for 30 minutes. Blunt-ended DNA was then purified by spin column fractionation, giving a final volume of 15 µl.

Ligation of the blunt-ended DNA was performed in an analogous way as to the ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector except that the final DNA was resuspended in 17 µl of 1 X TE.

Following ligation, the DNA was subjected to a second restriction digestion with Not I by mixing the 17 µl of DNA with 2 µl 10 X Not I digestion buffer and 1 µl Not I enzyme (NEB). Digestion was allowed to proceed at 37°C for 60 minutes. Following digestion, the admixture was purified by spin column fractionation resulting in 15 µl final volume.

Ten (10) µl of the purified DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL including ampicillin (50 µg/ml;

SUBSTITUTE SHEET (RULE 26)

This Sheet Has Been Replaced



FEB-19-96 22:54 FROM: ID: PAGE 46

Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO- depending on the sufficiency of the foregoing. (BlueNEO- was confirmed due to the sufficency of the following procedure.)

BlueNEO- contains a unique Pst I restriction site spanning the codons for amino acid residues 51 and 52. BlueNEO- was digested with Pst I as follows: an admixture was formed containing 15 µl dH_2O; 1 µl BlueNEO- DNA, 2 µl digestion buffer 3 (NEB) and 2 µl Pst I enzyme (NEB, Prod. No. 140S). This admixture was incubate at 37°C for 3 hrs. Digested DNA was then purified by spin column fractionation. The following synthetic oligonucleotide was then ligated to the Pst I cohesive ends of BlueNEO-: 5' GGTAAGTGCGGCCGCTACTAACTCTCTCCTCCCTCCTTTTTCCTGCA 3' (SEQ ID NO: 12) and its complementary sequence: 5' GGAAAAAGGAGGGAGGAGAGAGTTAGTAGCGGCCGCACTTACCTGCA 3' (SEQ ID NO: 13). Insertion of this linker creates a consensus 5' splice donor site (by ligation) followed by a Not I site, followed by a consensus splice branch point, followed by a synthetic polypyrimidine tract, followed by a consensus 3' splice acceptor site, as indicated above.

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Pst I linearized BlueNEO- DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

The foregoing (and following) synthetic oligo nucleotides were chemically synthesized using an Applied Biosystems 391 PCR MATE™ DNA Synthesizer (Applied Biosystems, Foster City, CA). All reagents for the synthesis were purchased from Applied Biosystems.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene),

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:54 FROM: ID: PAGE 47

following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA and/or NEOSPLA- depending on the sufficiency of the foregoing and the orientation of the insertion of the oligonucleotides.

Determination of orientation of the splice junction linker was preformed by nucleic acid sequencing using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical, Cleveland, OH, Prod. No. 70770) following manufacturer instructions. Upon determination of linker orientation within six independent plasmid isolates, identification of NEOSPLA was made such that the inserted splice junction sequences are in the correct forward orientation with respect to the direction of NEO transcription.

NEOSPLA was digested with Xho I by forming an admixture of 15 µl dH_2O; 1 µl NEOSPLA DNA; 2 µl 10 X digestion buffer D (Promega, supplied with enzyme); and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide having the following sequence: 5' TCGATTAATTAA 3' (SEQ ID NO: 14). Insertion of this sequence effectively changes the Xho I site to a Pac I restriction site (as underlined in SEQ ID NO: 14).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized NEOSPLA DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:54 FROM: ID: PAGE 48

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA3 depending on the sufficiency of the foregoing. (NEOSPLA3 was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K) contains the anti-CD20 light chain and heavy chain immunoglobulin cassettes and a DHFR cassette bounded by a Not I site at the 5' end and an Xho I site at the 3' end. Anti-CD20 in ANEX 2(G1,K) was digested with Xho I by forming an admixture of 15µldH_2O, 1 µl anti-CD20 in ANEX 2(G1,K) DNA, 2 µl 10 X digestion buffer D (Promega, supplied with enzyme) and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3 hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide of the following sequence: 5' TCGAAGCGGCCGCT 3' (SEQ ID NO: 15). Insertion of this sequence effectively changes the Xho I site to a Not I restriction site (as underlined in SEQ ID NO: 15).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized anti-CD20 in ANEX 2 DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

SUBSTITUTE SHEET (RULE 26) This Sheet Has Been Replaced



FEB-19-96 22:54 FROM: ID: PAGE 49

M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as Anti-CD20 in ANEX 2(G1,K)A depending on the sufficiency of the foregoing. (This was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K)A was digested with Not I and Xho I by forming an admixture of 6 µl dH_2O; 10 µl Anti-CD20 in ANEX 2(G1,K); 2 µl 10 x Not I digestion buffer (NEB, supplied with Not I enzyme); and 1 µl Not I enzyme (NEB). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 5515 base pairs by was isolated via the GlassMAX method for insertion into NEOSPLA3.

NEOSPLA3 was previously prepared for acceptance of the anti-CD20 cassette by digestion of 1 µl of DNA with Not I using an admixture comprising 16 µl dH_2O;

2 µl 10X Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB); followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume.

Ligation of the anti-CD20 DNA fragment into prepared NEOSPLA3 vector was accomplished as follows: 10 µl of anti-CD20 fragment DNA was admixed with 6 µl dH_2O; 1 µl cut NEOSPLA3 vector DNA; 2 µl 10 x ligation buffer (Promega supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega); followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmids referred to as anti-CD20 in NEOSPLA3F and anti-CD20 in NEOSPLA3R depending on the sufficiency of the foregoing and relative orientation of the inserted fragment with respect to NEO transcription.

Determination of orientation of the anti-CD20 cassette insertion was preformed by double digestion with KpnI and SpeI (NEB, Prod. No. 1335) in NEB buffer 1 plus acetated BSA as follows: an admixture comprising 4 µl DNA; 2 µl NEB buffer 1; 1 µl Kpn I; 1 µl Spe I; 2 µl BSA; and 10 µl dH_20 was formed. The admixture was digested at 37°C for 2 hrs, followed by size fractionation on an 0.8% agarose gel electrophoresis. Upon determination of anti-CD20 insert orientation within six independent plasmid isolates, identification of anti-CD20 in NEOSPLA3F was made such that the inserted sequences are in the forward orientation with respect to the direction of NEO transcription.

The 5515 bp anti-CD20 fragment contains the SV40 origin, a chimeric mouse human immunoglobulin light chain transcriptional cassette, a chimeric mouse human immunoglubulin heavy chain transcriptional cassette, and a murine dihydrofolate reductase transcriptional cassette (*see*, Figure 4).

Anti-CD20 in NEOSPLA3F was doubly digested with Kpn I and Stu I by creating the admixture consisting of 14 µl dH_2O, 1 µl anti-CD20 in NEOSPLA3F, 2 µl 10 x digestion buffer 1 (NEB, supplied with enzyme), 2 µl 10 x acetylated BSA (NEB supplied with Kpn I enzyme), 1 µl Kpn I enzyme, 1 µl Stu I enzyme (NEB, Prod. Nos. 142S and 187S respectively). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 9368 base pairs by was isolated via the GlassMAX method.

A PCR fragment of DNA was generated from TCAE 5.2. The two following synthetic oligonucleotide primers were utilized in the PCR reaction:

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)



FEB-19-96 22:55 FROM: ID: PAGE 51

5' primer: 5' GCA TGC GGT ACC GGA TCC ATC GAG CTA CTA GCT TTG C 3'

(SEQ ID NO: 16);

3' primer: 5' CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC TGC AGT TCA TCC AGG GC 3' (SEQ ID NO: 17)

The underlined portion of SEQ ID NO: 16 represents a Kpn I site, and the underlined portion of SEQ ID NO: 17 represents a Stu I site.

The PCR product was digested with Kpn I and Stu I and then ligated into prepared anti-CD20 in NEOSPLA3F.

Ligation of the 627 bp fragment into prepared anti-CD20 in NEOSPLA3F was accomplished as follows:

2 µl anti-CD20 in NEOSPLA3F; 1 µl SDS; 1 µl tRNA (Sigma); 11 µl 3M sodium acetate (pH 4.5) were admixed. Following phenol/chloroform isoamyl extraction of the admixture, the DNA was precipitated from the aqueous phase by addition of 270 µl ethanol (ice-cold) and this was spun at 13,000 rpm for 10 min. Following a 70% ETOH wash, the DNA was resuspended in 16 µlTE, 10 µl of PCR fragment DNA was admixed with 6 µl dH_2O, 1 µl cut anti-CD20 in NEOSPLA3F vector DNA, 2 µl 10 x ligation buffer (Promega, supplied with enzyme) and 1 µl T-4 DNA Ligase (Promega) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No. M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as anti-CD20 in GKNEOSPLA3F depending on the sufficiency of the foregoing. (Confirmation was based upon sequence determination of the different regions of GKNEOSPLA3F vs. NEOSPLA3F.)

This Sheet Has Been Replaced SUBSTITUTE SHEET (RULE 26)



FEB-19-96 22:55 FROM: ID: PAGE 52

The new plasmid differs from anti-CD20 in NEOSPLA3F in its Kozak sequence for the NEO gene which is:

```
                                        -3          +1
TGT GTT GGG AGC TTG GAT CGAT cc Acc ATG Gtt
                          Cla I   Start NEO
```

for Anti-CD20 in GKNEOSPLA3F, and

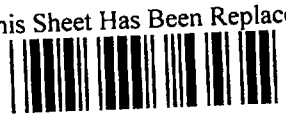
```
                                      -3    +1
TGT G CCA GCA TGG AGG AAT CGA Tcc Tcc ATG Ctt
          upstream Start            Start NEO
```

for Anti-CD20 in NEOSPLA3F.

Comparative analysis of expression of anti-CD20 in TCAE 5 vector (comprising NEO with consensus Kozak); ANEX 2 vector (comprising NEO with fully impaired Kozak, and upstream out-of-frame start sequence); NEOPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has fully impaired Kozak and an upstream out-of-frame start sequence); and GKNEOSPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has consensus Kozak).

Twenty-five (25) µg of each plasmid (digested as follows: anti-CD20 in TCAE5 and ANEX2 - Not I; anti-CD20 in NEOSPLA3F - Pac I; anti-CD20 in GKNEOSPLA3F - Pac I and Kpn I) was electroporated into 4 x 10^6 CHO cells; these digestions were utilized to separate the genes expressed in mammalian cells from the DNA used to grow the plasmid in bacteria. Following digestion, EtOH precipitation of the DNA, and drying thereof, the DNA was resuspended in sterile TE at a concentration of 1 µg/µl. Electroporation conditions were as described in Example II, except that 230 volts was utilized and, following electroporation, the mixture of cells and DNA was maintained for 10 min. at room temperature in the sterile, disposable electroporation cuvette.

This Sheet Has Been Replaced



FEB-19-96 22:56 FROM: ID: PAGE 53

Following electroporation, cells were plated into 96 well dishes as shown below in Table I, based upon the expected frequency of G418 resistant colonies (as derived from preliminary experiments; data not shown):

Table I
Comparative Expression

Plasmid	No. Transfections	No. Cells Plated	No. 96 Well Plates
TCAE 5	1	4×10^5	5
ANEX 2	1	2×10^6	5
GKNEOSPLA3F	1	2×10^6	5
NEOSPLA3F	5	2×10^7	5

Table 1 (continued)

Plasmid	No. G418 Resistant Colonies	Frequency of G418 Resistant Colony per Transfected Cell
TCAE 5	16	1 in 20,000
ANEX 2	16	1 in 100,000
GKNEOSPLA3F	16	1 in 100,000
NEOSPLA3F	16	1 in 1,000,000

(Cells were fed with G418 containing media on days 2, 5, 7, 9, 12, 14, 18, 22, 26, 30 and 34; supernatant from colonies was assayed for immunoglobulin production and the colonies became confluent in the wells on days 18, 22, 26, 30 and 34).

Figures 7A to 7C provide histogram results and evidence the percentage of colonies at a particular level of expression.

The Examples provided herein are not to be construed as limited to the specific vectors, fully impaired consensus Kozak sequences, dominant selectable markers, transcriptional cassettes, and/or expressed proteins. The fully impaired consensus Kozak, and the utilization

SUBSTITUTE SHEET (RULE 26)

This Sheet Has Been Replaced



FEB-19-96 22:56 FROM: ID: PAGE 54

thereof, are not to be construed as limited to ANEX 1 and ANEX 2 vectors. Similarly, the preferred fully impaired consensus Kozak sequences and vectors in no way constitute an admission, either actual or implied, that these are the only sequences or vectors to which the inventor is entitled. The inventor is entitled to the full breadth of protection under applicable patent laws. Preferred vectors incorporating fully impaired consensus Kozak sequences have been identified by the inventor on ANEX 1 and ANEX 2 for purposes of claiming these vectors by designating plasmids comprising these vectors and anti-CD20 were deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland, 20852, under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The plasmids were tested by the ATCC on November 9, 1992, and determined to be viable on that date. The ATCC has assigned these plasmids the following ATCC deposit numbers 69120 (anti-CD20 in TCAE 12(ANEX 1)) and 69118 (anti-CD20 in ANEX 2 (G1,K)); for purposes of this deposit, these plasmids were transformed into *E. coli*.

Although the invention has been described in considerable detail with regard to certain preferred embodiments thereof, other embodiments within the scope of the teachings of the present invention are possible. Accordingly, neither the disclosure nor the claims to follow, are intended, nor should be construed to be, limited by the descriptions of the preferred embodiments contained here.

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:56 FROM: ID: PAGE 55

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Reff, Mitchell E.

(ii) TITLE OF INVENTION: Impaired Dominant Selectable Marker Sequence and Intronic Insertion Strategies for Enhancement of Expression of Gene Product and Expression Vector Systems Comprising Same

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDING ADDRESS:

(A) ADDRESSEE: IDEC Pharmaceuticals Corporation

(B) STREET: 11011 Torreyana Road

(C) CITY: San Diego

(D) STATE: California

(E) COUNTRY: USA

(F) ZIP: 92121

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb

(B) COMPUTER: Macintosh

(C) OPERATING SYSTEM: MS.DOS

(D) SOFTWARE: Microsoft Word 5.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Burgoon, Richard P. Jr.

(B) REGISTRATION NUMBER: 34,787

(C) REFERENCE/DOCKET NUMBER:

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (619) 550-8500

(B) TELEFAX: (619) 550 8750

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)



FEB-19-96 22:56 FROM: ID: PAGE 56

(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
TAG CTA GGT CCT ACC CC 17

(3) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
ATC GAT CCT GGA TGC GG

(4) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
TNN ATG CTT 9

(5) INFORMATION FOR SEQ ID NO: 4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single, including nick
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: (DNA (genomic)

This Sheet Has Been Replaced

SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:57 FROM: ID: PAGE 57

(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
CNN ATG CTT 9

(6) INFORMATION FOR SEQ ID NO: 5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
TNN ATG TTT 9

(7) INFORMATION FOR SEQ ID NO: 6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
CNN ATG TTT 9

(8) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
TTG GGA GCT TGG ATC GAT 18
CCA CCA TGG TT 11

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:57 FROM: ID: PAGE 58

(9) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
TTG GGA GCT TGG ATC GAT 18
CCT CCA TGC TT 11

(10) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
CCA GCA TGG AGG AAT CGA TCC 21
TCC ATG CTT 9

(11) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
GGA GGA TCG ATT CCT CCA TGC TGG 24
CAC AAC TAT GTC AGA AGC AAA TGT GAG C 28

(12) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:

This Sheet Has Been Replaced



FEB-19-96 22:57 FROM: ID: PAGE 59

(A) LENGTH: 18 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
CTG GGG CTC GAG CTT TGC 18

(13) INFORMATION FOR SEQ ED NO: 12
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no62
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
GGT AAG TGC GGC CGC TAC TAA CTC TCT CCT 30
CCC TCC TTT TTC CTG GA 17

(14) INFORMATION FOR SEQ ID NO: 13
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:
GGA AAA AGG AGG GAG GAG AGA GTT AGT AGC GGC 33
CGC ACT TAC CTG CA 14

(15) INFORMATION FOR SEQ ID NO: 14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single

This Sheet Has Been Replaced



FEB-19-96 22:57 FROM: ID: PAGE 60

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCG ATT AAT TAA 12

(16) INFORMATION FOR SEQ ID NO: 15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCG AAG CGG CCG CT 14

(17) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16

GCA TGC GGT ACC GGA TCC ATC GAG CTA 27

CTA GCT TTG C 10

(18) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

This Sheet Has Been Replaced



FEB-19-96 22:58 FROM: ID: PAGE 81

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17

CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC 30

TGC AGT TCA TCC AGG GC 17

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:58 FROM: ID: PAGE 62

Claims

1. An expression vector for expressing a protein of interest by recombinant deoxyribonucleic acid techniques, said vector comprising at least one dominant selectable marker, wherein the translation initiation start site of said marker comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where "Py" is a pyrimidine nucleotide; "x" is a nucleotide; and the numerical designations are relative to the codon "ATG".

2. The expression vector of claim 1 wherein the nucleic acid sequence encoding for the protein of interest is co-linked to said dominant selectable marker.

3. The expression vector of claim 1 wherein said dominant selectable marker is selected from the group consisting of: herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

4. The expression vector of claim 1 wherein said translation initiation start site sequence is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that the codon "Txx" downstream of the ATG codon does not encode a stop codon.

5. The expression vector of claim 1 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

6. The expression vector of claim 1 wherein said translation initiation start site sequence is TCCATGCTT.

7. The expression vector of claim 1 wherein said translation initiation start site sequence is located within a secondary structure.

8. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

9. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

10. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

11. The expression vector of claims 8, 9 and 10 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

12. The expression vector of claim 10 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

13. The expression vector of claims 8, 9, 10 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

14. A dominant selectable marker encoded by a nucleic acid sequence, wherein the translation initiation start site of said dominant selectable marker is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that "Txx" downstream of the ATG codon does not encode a stop codon.

15. The material of claim 14 wherein said dominant selectable marker is selected from the group consisting of herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

16. The material of claim 14 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

17. The material of claim 14 wherein said translation initiation start site sequence is TCCATGCTT.

18. The material of claim 14 wherein said translation initiation start site sequence is located within a secondary structure.

19. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

20. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

21. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

22. The material of claims 19, 20 and 21 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

23. The material of claim 21 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

24. The material of claims 19, 20, 21 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

25. An expression vector selected from the group consisting of ANEX 1 (included within American Type Culture Collection deposit number 69120) and ANEX 2 (included within ATCC deposit number 69118).

26. A plasmid comprising the expression vector of claim 1 wherein the nucleic acid sequence encoding for said protein of interest is co-linked to said dominant selectable marker.

27. The plasmid of claim 26 integrated within the cellular deoxyribonucleic acid of a mammalian host cell.

28. The plasmid of claim 27 wherein said host cell is selected from the group consisting of DG44, DXB11, CV1, COS, R1610, SP2/0, P3x633-Ag8.653, BFA-1c1BPT, RAJI, and 293.

29. The expression vector of claim 1 further comprising an artificial intronic insertion region within said dominant selectable marker, wherein an encoding sequence for a protein of interest is located within said insertion region.

30. The dominant selectable marker of claim 14 further comprising an artificial instronic insertion region.

This Sheet Has Been Replaced



SUBSTITUTE SHEET (RULE 26)

FEB-19-96 22:59 FROM: ID: PAGE 67

CORRECTED VERSION*

PCT

204/180

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
C12N 15/85, 15/13, 15/65, 15/67	A3	WO 94/11523
		(43) International Publication Date: 26 May 1994 (26.05.94)

(21) International Application Number: PCT/US93/11221

(22) International Filing Date: 12 November 1993 (12.11.93)

(30) Priority data:
07/977,691 13 November 1992 (13.11.92) US

(71) Applicant: IDEC PHARMACEUTICALS CORPORATION [US/US]; 11011 Torreyana Road, San Diego, CA 92121-1104 (US).

(72) Inventor: REFF, Mitchell, E. ; 4166 Combe Way, San Diego, CA 92122 (US).

(74) Agents: WEISS, Steven, M. et al.; Lyon & Lyon, 611 West Sixth Street, 34th floor, Los Angeles, CA 90017 (US).

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, UZ, VN. European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, 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 the receipt of amendments.

(88) Date of publication of the international search report:
7 July 1994 (07.07.94)

RECEIVED

AUG - 8 1994

LYON & LYON
INTERNATIONAL DEPT

(54) Title: FULLY IMPAIRED CONSENSUS KOZAC SEQUENCES FOR MAMMALIAN EXPRESSION

Neomycin phosphotransferase gene

		Cla I	−3		+1
TCAE 5.2	TTGGGAGCTTGG	ATCGAT	CC	A cc ATG	Gtt
				Met	Val
ANEX 1	TTGGGAGCTTGG	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu
ANEX 2	CCaGCATGgAGGA	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu

This Sheet Has Been Replaced



−3	+1	(I)
Pyxx	ATG	Pyxx

(57) Abstract

Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector. These vectors are most typically utilized in the expression of proteins in mammalian expression systems. As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises sequence (I), where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T)/uracil (U); "Py" is a pyrimidine nucleotide, *ie* C or TU; "ATG" is a codon encoding for the amino acid methionine, the so-called "start" codon; and -3 and +1 are directional reference points vis-á-vis ATG, *ie* -3 is meant to indicate three nucleotides upstream of ATG and +1 is meant to indicate one nucleotide downstream of ATG. Dominant selectable markers further comprising artificial intronic insertion regions are further disclosed.

* (Referred to in PCT Gazette No. 17/1994, Section II)

FEB-19-96 22:41 FROM: ID: PAGE 4

This Sheet Has Been Replaced



Neomycin phosphotransferase gene

		Cla I	-3		+1
TCAE 5.2	TTGGGAGCTTGG	ATCGAT	CC	A cc ATG	Gtt
				Met	Val
ANEX 1	TTGGGAGCTTGG	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu
ANEX 2	CCaGCATGgAGGA	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu

FIG. 1

FEB-19-96 22:59 FROM: ID: PAGE 68

This Sheet Has Been Replaced

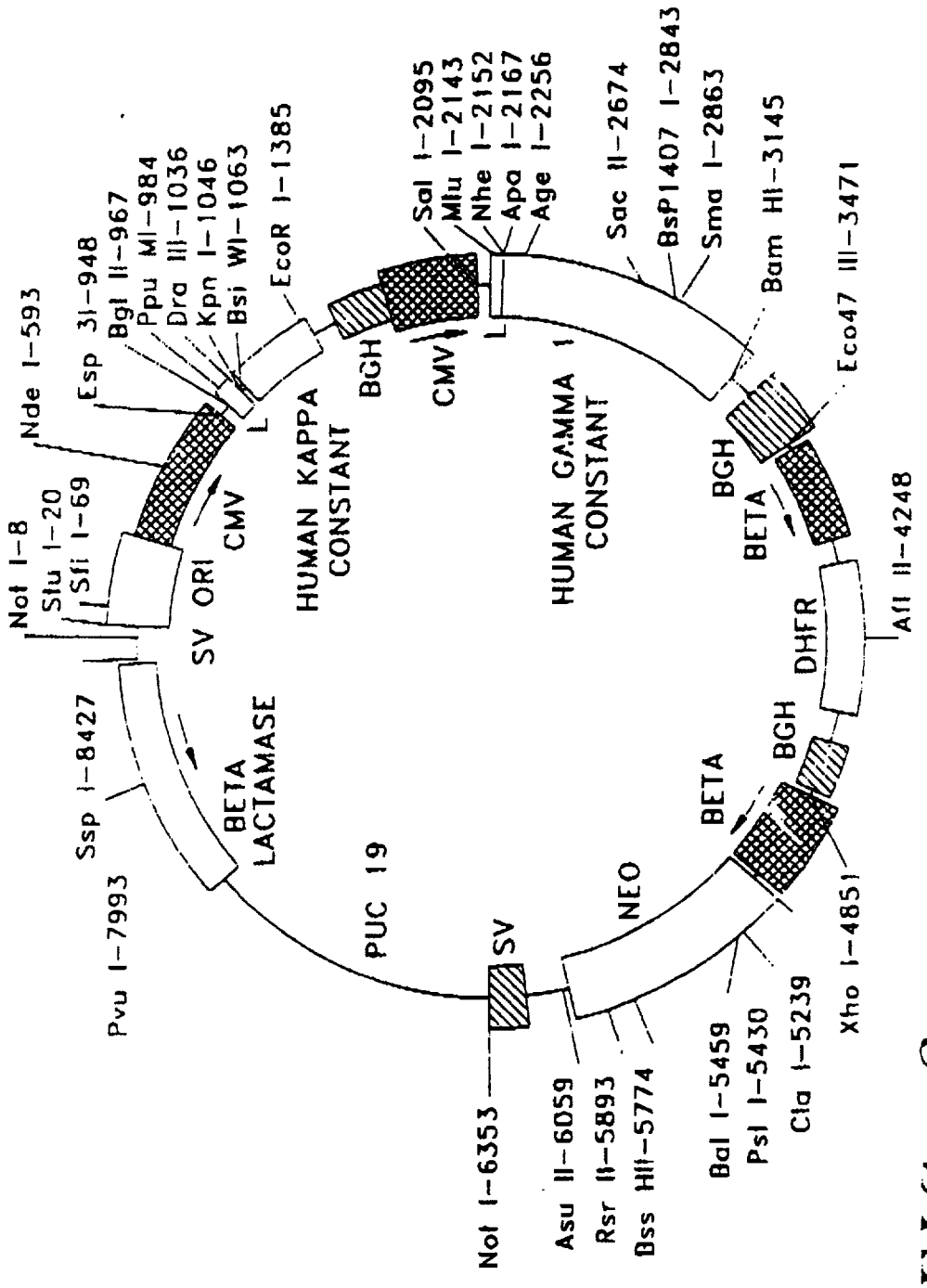


FIG. 2

FEB-19-96 22:59 FROM: ID: PAGE 69

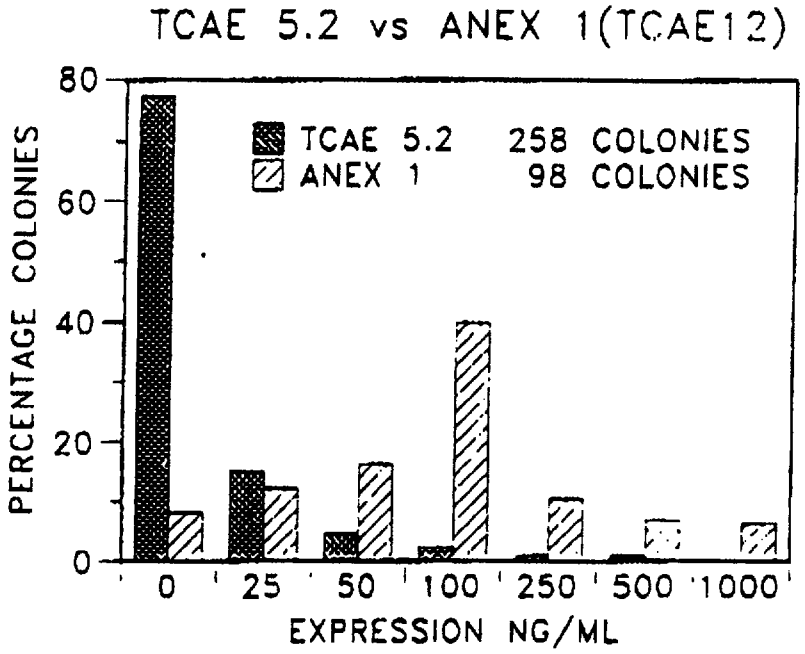


FIG. 3

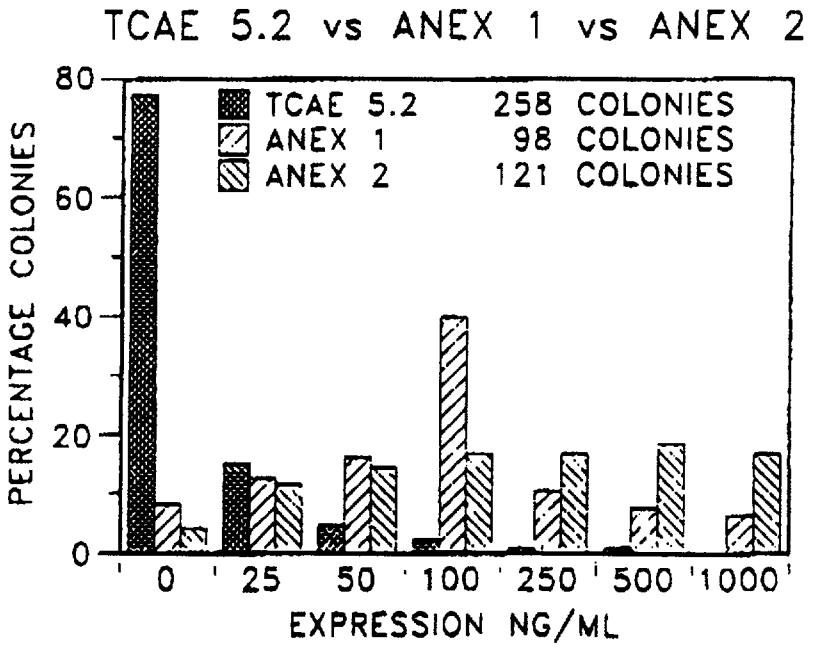


FIG. 5

FEB-19-96 23:00 FROM: ID: PAGE 70

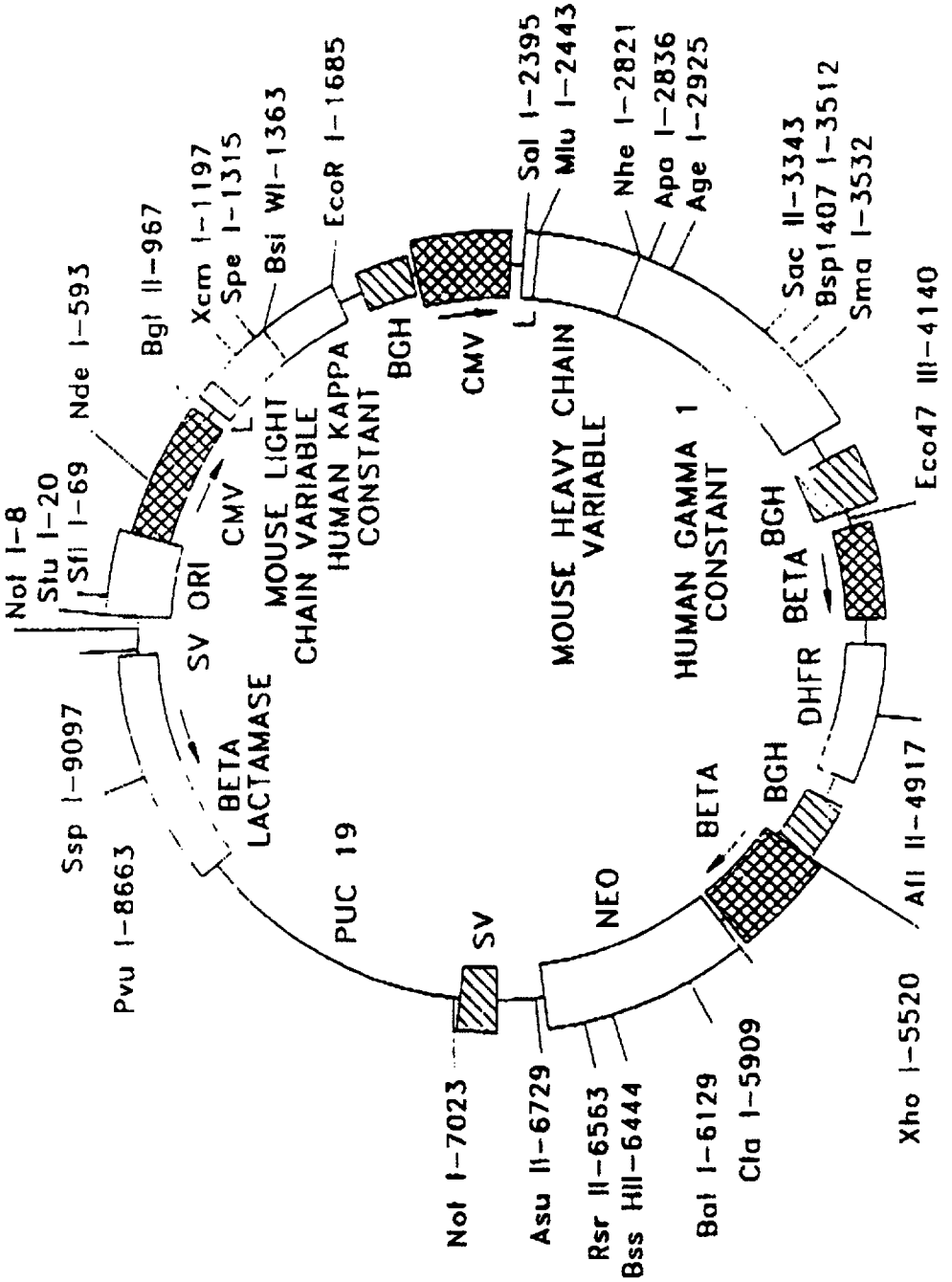


FIG. 4

FEB-19-96 23:00 FROM: ID: PAGE 71

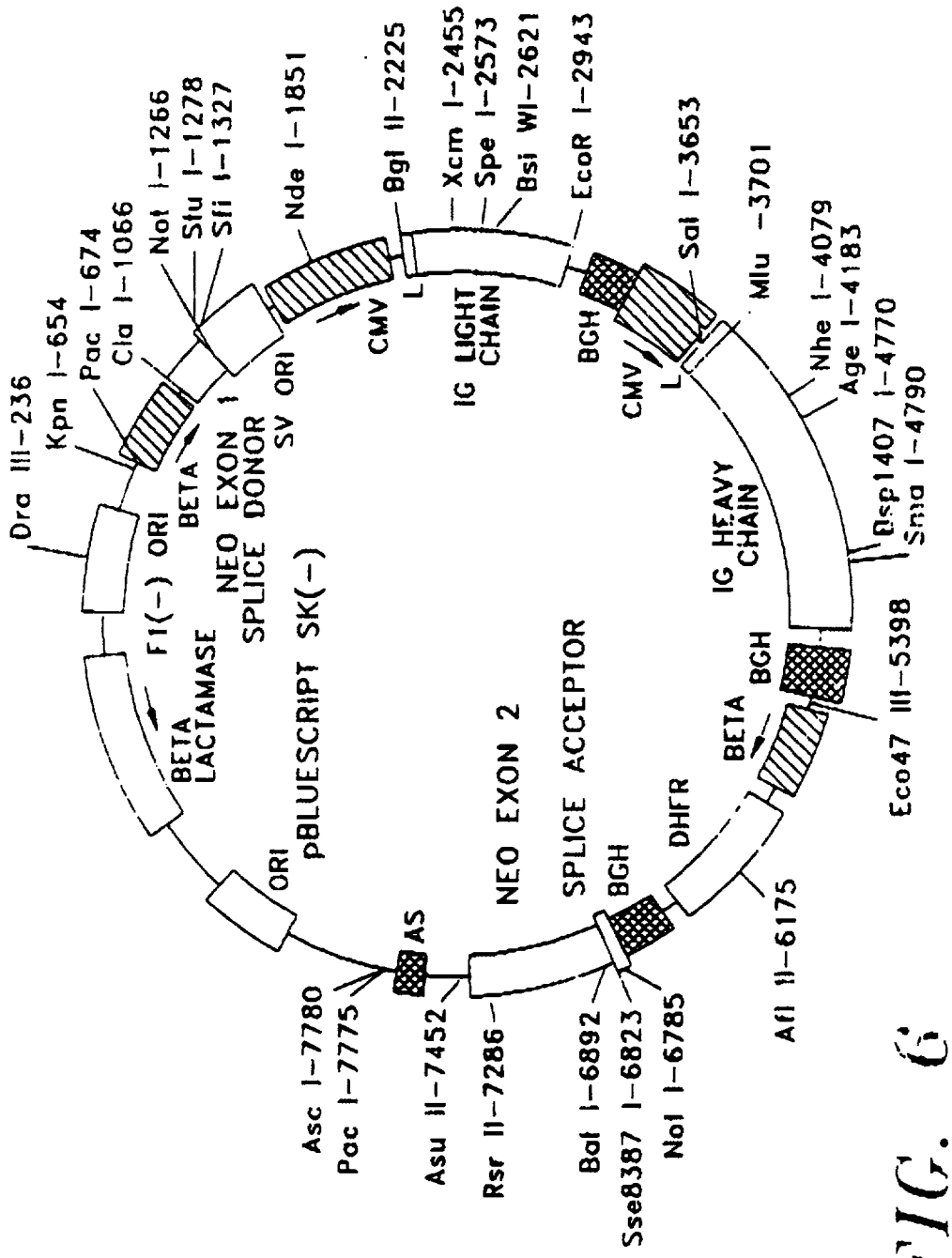


FIG. 6

FEB-19-96 23:00 FROM: ID: PAGE 72

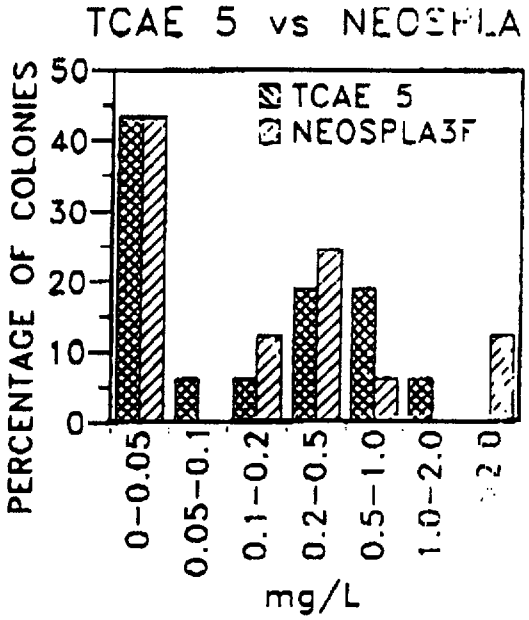


FIG. 7A

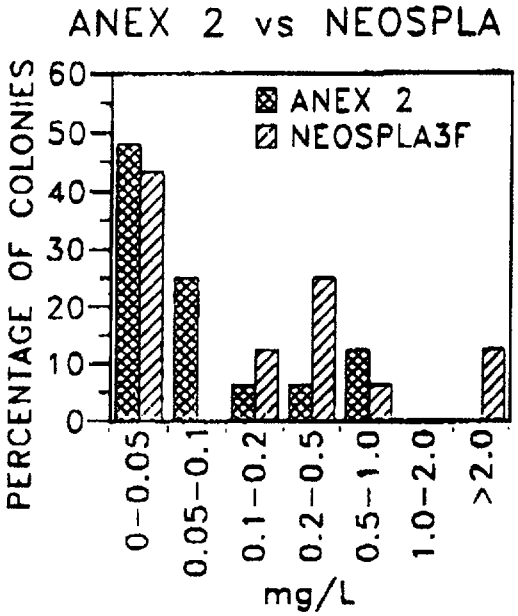


FIG. 7B

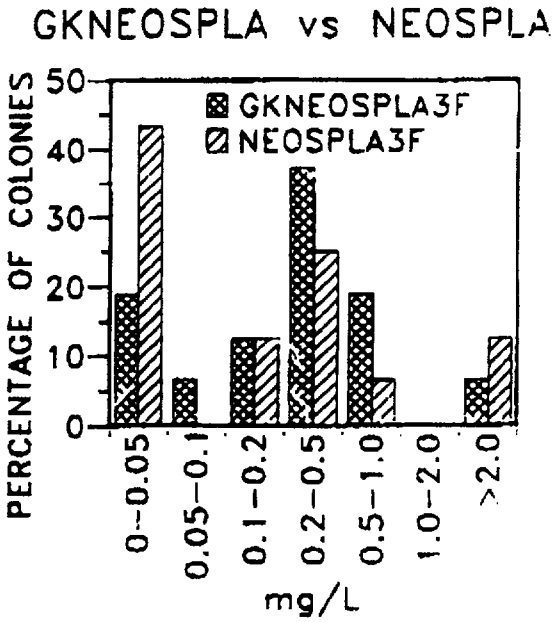


FIG. 7C

FEB-19-96 23:00 FROM: ID: PAGE 73

This Sheet Has Been Replaced

DESCRIPTION

Impaired Dominant Selectable Marker Sequence And Intronic Insertion Strategies For Enhancement Of Expression Of Gene Product And Expression Vector Systems Comprising Same

37 C.F.R. §1.74(d)/(e) Copyright Notice

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner does not object to the reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent files or records, but otherwise reserves all copyright rights whatsoever.

Related Applications

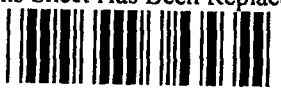
This patent document is a Continuation-In-Part of U.S. Serial No. 07/977,691 filed November 13, 1992, pending. This patent document is related to "THERAPEUTIC APPLICATION OF CHIMERIC ANTIBODY TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA," having U.S. Serial No. 07/978,891, filed November 13, 1992 (pending), and "THERAPEUTIC APPLICATION OF CHIMERIC AND RADIOLABLED ANTIBODIES TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA", having U.S. Serial No. _____, filed simultaneously herewith. This patent document is related to commonly assigned United States Serial No. 07/912,292 and entitled "RECOMBINANT ANTIBODIES FOR HUMAN THERAPY," filed July 10, 1992. These documents are incorporated herein by reference.

This is the replacement sheet

Date: 11 FEB 1997

INTENTIONALLY
LEFT BLANK
PER EXAMINER
M. HOGARTH

This Sheet Has Been Replaced



This is the replacement sheet

11 FEB 1997

As noted, the advent of the biotechnology industry has allowed for the production of large quantities of proteins. Proteins are the essential constituents of all living cells and proteins are comprised of combinations of 20 naturally occurring amino acids; each amino acid molecule is defined ("encoded") by groupings ("codons") of three deoxyribonucleic acid ("DNA") molecules; a string of DNA molecules ("DNA macromolecule") provides, in essence, a blueprint for the production of specific sequences of amino acids specified by that blueprint. Intimately involved in this process is ribonucleic acid ("RNA"); three types of RNA (messenger RNA; transfer RNA; and ribosomal RNA) convert the information encoded by the DNA into, *eg* a protein. Thus, genetic information is generally transferred as follows: DNA →RNA → protein.

In accordance with a typical strategy involving recombinant DNA technology, a DNA sequence which encodes a desired protein material ("cDNA") is identified and either isolated from a natural source or synthetically produced. By manipulating this piece of genetic material, the ends thereof are tailored to be ligated, or "fit," into a section of a small circular molecule of double stranded DNA. This circular molecule is typically referred to as a "DNA expression vector," or simply a "vector." The combination of the vector and the genetic material can be referred to as a "plasmid" and the plasmid can be replicated in a prokaryotic host (*ie* bacterial in nature) as an autonomous circular DNA molecule as the prokaryotic host replicates. Thereafter, the circular DNA plasmid can be isolated and introduced into a eukaryotic host (*ie* mammalian in nature) and host cells which have incorporated the plasmid DNA are selected. While some plasmid vectors will replicate as an autonomous circular DNA molecule in mammalian cells, (*eg* plasmids comprising Epstein Barr virus ("EBV") and Bovine Papilloma virus ("BPV") based vectors), most plasmids including DNA vectors, and all plasmids including RNA retroviral vectors, are integrated into the cellular DNA such that

This Sheet Has Been Replaced



when the cellular DNA of the eukaryotic host cell replicates, the plasmid DNA will also replicate. Accordingly, as eukaryotic cells grow and divide, there is a corresponding increase in cells containing the integrated plasmid which leads to the production ("expression") of the protein material of interest. By subjecting the host cells containing the plasmid to favorable growth conditions, significant amounts of the host, and hence the protein of interest, are produced. Typically, the Chinese Hamster Ovary ("CHO") cell line is utilized as a eukaryotic host cell, while *E. coli* is utilized as a prokaryotic host cell.

The vector plays a crucial role in the foregoing - manipulation of the vector can allow for variability as to where the cDNA is inserted, means for determining whether the cDNA was, in fact, properly inserted within the vector, the conditions under which expression of the genetic material will or will not occur, etc. However, most of the vector manipulations are geared toward a single goal-increasing expression of a desired gene product, *ie* protein of interest. Stated again, most vector manipulation is conducted so that an "improved" vector will allow for production of a gene product at significantly higher levels when compared to a "non-improved" vector. Thus, while certain of the features/aspects/characteristics of one vector may appear to be similar to the features/aspects/characteristics of another vector, it is often necessary to examine the result of the overall goal of the manipulation - improved production of a gene product of interest.

While one "improved" vector may comprise characteristics which are desirable for one set of circumstances, these characteristics may not necessarily be desirable under other circumstances. However, one characteristic is desirable for all vectors: increased efficiency, *ie* the ability to increase the amount of protein of interest produced while at the same time decreasing the number of host cells to be screened which

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

do not generate a sufficient amount of this protein. Such increased efficiency would have several desirable advantages, including reducing manufacturing costs and decreasing the time spent by technicians in screening for viable colonies which are expressing the protein of interest. Accordingly, what would be desirable and what would significantly improve the state of the art are expression vectors with such efficiency characteristics.

Summary Of The Invention

The invention disclosed herein satisfies these and other needs. Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector; preferably, the dominant selectable marker comprises either a natural intronic insertion region or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within such insertion region.

As used herein, a "dominant selectable marker" is a gene sequence or protein encoded by that gene sequence; expression of the protein encoded by the dominant selectable marker assures that a host cell transfected with an expression vector which includes the dominant selectable marker will survive a selection process which would otherwise kill a host cell not containing this protein. As used herein, a "transcriptional cassette" is DNA encoding for a protein product (*eg* a dominant selectable marker) and the genetic elements necessary for production of the protein product in a host cell (*ie* promoter; transcription start site; polyadenylation region; etc.). These vectors are most preferably utilized in the expression of proteins in mammalian expression systems where integration of the vector into host cellular DNA occurs. Beneficially, the use of such fully impaired consensus Kozak sequences improves the efficiency of protein expression by significantly decreasing the number

This Sheet Has Been Replaced



This is the replacement sheet

11 FEB 1997

of viable colonies while at the same time, significantly increasing the amount of protein expressed by such viable colonies. As used herein, a "natural intronic insertion region" is a region of DNA naturally present within a gene in this case, typically a dominant selectable marker, which can be utilized for insertion of DNA encoding a gene product of interest; an "artificial intronic insertion region" is a region of DNA which is selectively created in a gene (again, most typically, a dominant selectable marker) which can be utilized for insertion of DNA encoding a gene product of interest. Information regarding intronic positioning is described in Abrams, J.M. et al. "Intronic Positioning Maximizes Co-expression and Co-amplification of Nonselectable Heterologous Genes. "*J. Bio. Chem., 264124*:14016 (1989), and U.S. Patent No. 5,043,270 (both documents being incorporated herein by reference).

As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T)/ uracil (U); "Py" is a pyrimidine nucleotide, *ie* C or T/U; "ATG" is a codon encoding for the amino acid methionine, the so-called "start" codon; and -3 and +1 are directional reference points vis-a-vis ATG, *ie* -3 is meant to indicate three nucleotides upstream of ATG and +1 is meant to indicate one nucleotide downstream of ATG.

Preferably, the fully impaired consensus Kozak is part of a methionine start codon that initiates translation of a dominant selectable marker portion of a transcriptional cassette which is part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl

This Sheet Has Been Replaced



transferase; hygromycin B phosphotransferase; and neomycin phosphotransferase. Most preferably, the dominant selectable marker is neomycin phosphotransferase.

In particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without an in-frame stop codon being located between the upstream start codon and the fully impaired consensus Kozak start codon. As used herein, the term "stop codon" is meant to indicate a codon which does not encode an amino acid such that translation of the encoded material is terminated; this definition includes, in particular, the traditional stop codons TAA, TAG and TGA. As used herein, the terms "in-frame" and "out-of-frame" are relative to the fully impaired consensus Kozak start codon. By way of example, in the following sequence:

____ -3 +1

GAC CAT GGC Cxx ATG Cxx

the underlined portion of the sequence is representative of a fully impaired consensus Kozak (where "x" represents a nucleotide) and the codons GAC, CAT and GCC are "in-frame" codons relative to the ATG start codon. The above-lined nucleotides represent an "out-of-frame" start codon which is upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is a part of a consensus Kozak. By way of example, the sequence set forth above satisfies this criteria: the -5 nucleotide is a purine (G); nucleotide -6, -7 and -8 encode an out-of-frame start codon (ATG); and nucleotide -11 is a purine (A).

Additionally, utilization of a fully impaired consensus Kozak within a secondary structure (*ie* a so-

This Sheet Has Been Replaced  This is the replacement sheet

Date: 11 FEB 1997

called "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable number. In such an embodiment, the start codon of the fully impaired consensus Kozak is most preferably located within the stem of a stem loop.

Particularly preferred expression vectors which incorporate these aspects of the invention disclosed herein are referred to as "TCAE," and "ANEX" and "NEOSPLA" vectors; particularly preferred vectors are referred to as ANEX 1, ANEX 2 and NEOSPLA3F.

These and other aspects of the invention disclosed herein will be delineated in further detail in the sections to follow.

Brief Description Of The Drawings

Figure 1 provides the relevant portion of a consensus Kozak and several particularly preferred fully impaired consensus Kozak sequences;

Figure 2 provides a diagrammatic representation of the vectors TCAE 5.2 and ANEX 1 (TCAE 12) designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 3 is a histogram comparing protein expression levels with the vectors TCAE 5.2 and ANEX 1;

Figure 4 provides a diagrammatic representation of the vector ANEX 2 designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 5 is a histogram comparing protein expression levels with the vectors TCAE 5.2, ANEX 1 and ANEX 2;

Figure 6 provides a diagrammatic representation of a NEOSPLA vector designed for expression of mouse/human chimeric immunoglobulin; and

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

Figures 7A, 7B and 7C are histograms comparing protein expression levels with the vectors TCAE 5.2 vs. NEOSPLA 3F (7A). ANEX 2 vs. NEOSPLA3F (7B); and GK-NEOSPLA3F vs. NEOSPLA3F (7C).

Detailed Description Of Preferred Embodiments

Disclosed herein are nucleic acid sequence arrangements which impair translation and initiation of, most preferably, dominant selectable markers incorporated into mammalian expression vectors and which are preferably, but not necessarily, co-linked to an encoding sequence for a gene product of interest. Preferably, the dominant selectable marker comprises at least one natural or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within at least one such intron. Such arrangements have the effect of increasing expression efficiency of the gene product of interest by, *inter alia*, decreasing the number of viable colonies obtained from an equivalent amount of plasmid DNA transfected per cell, while increasing the amount of gene product expressed in each clone.

For purpose of brevity and presentational efficiency, the focus of this section of the patent disclosure will be principally directed to a specific dominant selectable marker, neomycin phosphotransferase, which is incorporated into a mouse/human chimeric immunoglobulin expression vector. It is to be understood, however, that the invention disclosed herein is not intended, nor is it to be construed, as limited to these particular systems. To the contrary, the disclosed invention is applicable to mammalian expression systems *in toto*, where vector DNA is integrated into host cellular DNA.

One of the most preferred methods utilized by those in the art for producing a mammalian cell line that produces a high level of a protein (*ie* "production cell line") involves random integration of DNA coding for the desired gene product (*ie* "exogenous DNA") by using, most typically, a drug resistant gene, referred to as a

This Sheet Has Been Replaced



"dominant selectable marker," that allows for selection of cells that have integrated the exogenous DNA. Stated again, those cells which properly incorporate the exogenous DNA including, *eg*, the drug resistant gene, will maintain resistance to the corresponding drug. This is most typically followed by co-amplification of the DNA encoding for the desired gene product in the transfected cell by amplifying an adjacent gene that also encodes for drug resistance ("amplification gene"), *eg* resistance to methotrexate (MTX) in the case of dehydrofolate reductase (DHFR) gene. The amplification gene can be the same as the dominant selectable marker gene, or it can be a separate gene. (As those in the art appreciate, "transfection" is typically utilized to describe the process or state of introduction of a plasmid into a mammalian host cell, while "transformation" is typically utilized to describe the process or state of introduction of a plasmid into a bacterial host cell).

Two amplification approaches are typically employed by those in the art. In the first, the entire population of transfected and drug resistant cells (each cell comprising at least one integration of the gene encoding for drug resistance) is amplified; in the second, individual clones derived from a single cell are amplified. Each approach has unique advantages.

With respect to the first approach, it is somewhat "easier" to amplify the entire population (typically referred to as a "shotgun" approach, an apt description) compared to individual clones. This is because amplification of individual clones initially involves, *inter alia*, screening of hundreds of isolated mammalian colonies (each derived from a single cell, most of which being single copy integrants of the expression plasmid) in an effort to isolate the one or two "grail" colonies which secrete the desired gene product at a "high" level, *ie* at a level which is (typically) three orders of magnitude higher than the lowest detectable expression level. These cells are also often found to have only a single copy

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

integration of the expression plasmid. Additionally, amplifying individual clones results in production cell lines which contain fewer copies of the amplified gene as compared to amplification of all transfected cells (typically, 10-20 versus 500-1000).

With respect to the second approach, production cell lines derived from amplifying individual clones are typically derived in lower levels of the drug(s) used to select for those colonies which comprise the gene for drug(s) resistance and the exogenous gene product (*ie* in the case of methotrexate and DHFR, 5nM versus 1μM). Furthermore, individual clones can typically be isolated in a shorter period of time (3-6 months versus 6-9 months).

Ideally then, the tangible benefits of both approaches should be merged: at a practical level, this would involve decreasing the number of colonies to be screened, and increasing the amount product secreted by these colonies. The present invention accomplishes this task.

The position where the DNA of the dominant selectable marker of the plasmid DNA is integrated within the cellular DNA of the host cell determines the level of expression of the dominant selectable marker protein, as is recognized by those in the art. It is assumed that the expression of a gene encoding a protein of interest which is either co-linked to or positioned near the dominant selectable marker DNA is proportional to the expression of the dominant selectable marker protein. While not wishing to be bound by any particular theory, the inventor has postulated that if the gene used to select for the integration of the exogenous DNA in the mammalian cell (*ie* the dominant selectable marker) was designed such that translation of that dominant selectable marker was impaired, then only those plasmids which could overcome such impairment by over-production of the gene product of the dominant selectable marker would survive, *eg*, the drug-screening process. By associating the exogenous DNA

This Sheet Has Been Replaced

11 FEB 1997

with the dominant selectable marker, then, *a fiorti*, over-production of the gene product of the dominant selectable marker would also result in over-production of the gene product derived from the associated exogenous DNA. In accordance with this postulated approach, impairment of translation of the dominant selectable marker gene would be necessary, and an avenue for such impairment was the consensus Kozak portion of the gene.

By comparing several hundred vertebrate mRNAs, Marilyn Kozak in "Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes," *Nuc. Acids Res. 9*: 5233-5252 (1981) and "Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs," *Nuc. Acids Res. 12*: 857-872 (1984), proposed the following "consensus" sequence for initiation of translation in higher eukaryotes:

```
      -3    +1
   cc Acc AUG G
```

(As those in the art appreciate, uracil, U, replaces the deoxynucleotide thymine, T, in RNA.) In this sequence, referred to as a "consensus Kozak," the most highly conserved nucleotides are the purines, A and G, shown in capital letters above; mutational analysis confirmed that these two positions have the strongest influence on initiation. *See, eg,* Kozak, M. "Effects of intercistronic length on the efficiency of reinitiation of eukaryotic ribosomes." *Mol. Cell Bio. 7/10*: 3438-3445 (1987). Kozak further determined that alterations in the sequence upstream of the consensus Kozak can effect translation. For example, in "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes." *PNAS 83*: 2850-2854 (1986) Kozak describes the "artificial" introduction of a secondary hairpin structure region upstream from the consensus Kozak in several plasmids that encoded preproinsulin; it was experimentally determined that a stable stem loop structure inhibited translation of the

This Sheet Has Been Replaced



preproinsulin gene, reducing the yield of proinsulin by 85-95%.

Surprisingly, it was discovered by the inventor that by changing the purines A(-3 vis-a-vis ATG start codon) and G(+1) to pyrimidines, translation impairment was significant: when the consensus Kozak for the neomycin phosphotransferase gene was subjected to such alterations (as will be set forth in detail below), the number of G418 resistant colonies significantly decreased; however, there was a significant increase in the amount of gene product expressed by the individual G418 resistant clones. As those in the art will recognize, this has the effect of increasing the efficiency of the expression system—there are less colonies to screen, and most of the colonies that are viable produce significantly more product than would ordinarily be obtained. Confirmation of the inventor's postulated theory was thus experimentally determined.

As noted, for purposes of this patent document a, "consensus Kozak" comprises the following sequence:

-3 +1

Pxx ATG Pxx

a "partially impaired consensus Kozak" comprises the following sequence

-3 +1

P/Pyxx ATG P/Pyxx

and a disclosed and claimed "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T) (uracil, U, in the case of RNA); "P" is a purine, *ie* A or G, "Py" is a pyridine, *ie* C or T/U; ATG is a conventional start codon which encodes for the amino acid methionine (Met); the numerical designations are relative to the ATG codon, *ie* a negative number indicates "upstream" of ATG and a positive number indicates

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

"downstream" of ATG; and for the partially impaired consensus Kozak, the following proviso is applicable--only one of the -3 or +1 nucleotides is a pyridine, *eg*, if -3 is a pyridine, then +1 must be a purine or if -3 is a purine, then +1 must be a pyridine. Most preferably, the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker which is preferably (but not necessarily) co-linked to exogenous DNA which encodes for a gene product of interest. As used herein, "nucleotide" is meant to encompass natural and synthetic deoxy- and ribonucleotides as well as modified deoxy- and ribonucleotides, *ie* where the 3' OH, 5'OH, sugar and/or heterocyclic base are modified, as well as modification of the phosphate backbone, *eg* methyl phosphates, phosphorothioates and phosphoramidites.

Information regarding the gene sequence of the dominant selectable marker is preferably known; however, in lieu of the entire sequence, information regarding the nucleic acid sequence (or amino acid sequence) at the site of translation initiation of the dominant selectable marker must be known. Stated again, in order to effectuate a change in the consensus or partially impaired consensus Kozak, one must know the sequence thereof. Changing the consensus or partially impaired consensus Kozak to a fully impaired consensus Kozak sequence can be accomplished by a variety of approaches well known to those in the art including, but not limited to, site specific mutagenesis and mutation by primer-based amplification (*eg* PCR); most preferably, such change is accomplished via mutation by primer-based amplification. This preference is principally based upon the comparative "ease" in accomplishing the task, coupled with the efficacy associated therewith. For ease of presentation, a description of the most preferred means for accomplishing the change to a fully impaired consensus Kozak will be provided.

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

In essence, mutation by primer-based amplification relies upon the power of the amplification process itself--as PCR is routinely utilized, focus will be directed thereto. However, other primer-based amplification techniques (*eg* ligase chain reaction, etc.) are applicable. One of the two PCR primers ("mutational primer") incorporates a sequence which will ensure that the resulting amplified DNA product will incorporate the fully impaired consensus Kozak within the transcriptional cassette incorporating the dominant selectable marker of interest; the other PCR primer is complementary to another region of the dominant selectable marker; a transcriptional cassette incorporating the dominant selectable marker; or a vector which comprises the transcriptional cassette. By way of example, the complement to a dominant selectable marker which includes a consensus Kozak could have the following sequence (SEQ ID NO: 1):

3'-tagctaggTccTACCcc-5'

In order to create a fully impaired consensus Kozak, the mutational primer could have the following sequence (SEQ ID NO: 2) (for convenience, SEQ ID NO: 1 is placed over the mutational primer for comparative purposes):

5'-atcgatccTggATGCgg-3'
3'-tagctaggTccTACCcc-5'
* *

As is evident, complementarity is lacking in the primer (*see* the "*" symbols). By utilizing excess mutational primer in the PCR reaction, when the sequence including the consensus Kozak is amplified, the resulting amplified DNA products will incorporate the mutations such that as the amplified DNA products are in turn amplified, the mutations will predominate such that a fully impaired consensus Kozak will be incorporated into the amplification product.

Two criteria are required for the mutational primer--first, the length thereof must be sufficient such that hybridization to the target will result. As will be

This Sheet Has Been Replaced



appreciated, the mutational primer will not be 100% complementary to the target. Thus, a sufficient number of complementary bases are required in order to ensure the requisite hybridization. Preferably, the length of the mutational primer is between about 15 and about 60 nucleotides, more preferably between about 18 and about 40 nucleotides, although longer and shorter lengths are viable. (To the extent that the mutational primer is also utilized to incorporate an out-of-frame start codon or secondary structure, the length of the mutational primer can correspondingly increase). Second, the ratio of mutational primer to target must be sufficiently excessive to "force" the mutation. Preferably, the ratio of mutational primer to target is between about 250:1 to about 5000:1, more preferably between about 400:1 to about 2500:1, and most preferably between about 500:1 to about 1000:1.

Because the parameters of a PCR reaction are considered to be well within the level of skill of those in the art, details regarding the particulars of that reaction are not set forth herein; the skilled artisan is readily credited with recognizing the manner in which this type of mutation can be accomplished using PCR techniques - the foregoing is provided as a means of providing elucidation as opposed to detailed edification.

As noted, it is most preferred that the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker incorporated into a transcriptional cassette which forms a part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl transferase ("XGPRT"); hygromycin B phosphotransferase; and neomycin phosphotransferase ("NEO"). Most preferably, the dominant selectable marker is NEO.

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

The dominant selectable marker herpes simplex virus thymidine kinase is reported as having the following partially impaired consensus Kozak:

```
      -3    +1
cg Cgt ATG Gct
```

See, Heller, S. "Insertional Activities of a Promotorless Thymidine Kinase Gene," *Mol. & Cell. Bio. 8/8*:3218-3302, Figure 4, nucleotide 764. By changing the +1 purine (G) to a pyrimidine (C or T/U), a fully impaired Kozak as defined herein is generated (the -3 of the herpes simplex virus thymidine kinase is a pyrimidine). Changing +1 purine to a pyrimidine also has the effect of changing the encoded amino acid from alanine (GCT) to proline (CCT) or serine (TCT); it is preferred that conservative amino acid changes result from the changes to the nucleotides. Thus, it is preferred that the change to TCT be made because the change from alanine to serine is a more conservative amino acid change than changing alanine to proline.

Histidinol dehydrogenase is another dominant selectable marker. *See*, Hartmen, S.C. and Mulligan, R.C. "Two dominant acting selectable markers for gene transfer studies in mammalian cells," *PNAS 85*:8047-8051 (1988). The *his D* gene of Salmonella typhimunium has the following partially impaired consensus Kozak:

```
      -3    +1
gc Aga ATG Tta
```

As -3 is a purine, changing -3 to a pyrimidine (C or T/U) results in a fully impaired consensus Kozak; as is appreciated, because these nucleotides are upstream of the start codon, no impact on amino acid translation results from this change.

Hygromycin B phosphotransferase is another dominant selectable marker; the reported sequence for the *hph* gene (*see*, Gritz, L. and Davies, J. "Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expressing in *Escherichia*

This Sheet Has Been Replaced



This is the replacement sheet

11 FEB 1997

coli and *Succharomyces cerevisiae*" *Gene 25*:179-188, 1983) indicates that the consensus Kozak is:

-3 +1

ga Gat ATG Aaa

Both -3 and +1 are purines; thus changing -3 and +1 to pyrimidines results in a fully impaired consensus Kozak (this results in the following encoded amino acids: +1 to C - glutamine; +1 to T - stop codon. Because this codon is downstream of the start codon, the change to the stop codon TAA should not be accomplished).

XGPRT is another dominant selectable marker. The reported partially impaired consensus Kozak of XGPRT has the following sequence:

-3 +1

tt Cac ATG Agc

See, Mulligan, R.C. and Berg, P. "Factors governing the expression of a bacterial gene in mammalian cells." *Mol. & Cell Bio. 1/5*:449-459 (1981), Figure 6. By changing the +1 purine to a pyrimidine, a fully impaired consensus Kozak is created; the effect on the encoded amino acid (AGC-serine) is as follows: CGC-arginine; TGC-cysteine.

Adenosine deaminase (ADA) can also be utilized as a dominant selectable marker. The reported consensus Kozak sequence for adenosihe deaminase is:

-3 +1

ga Acc ATG Gcc

See, Yeung, C.Y. *et al.*, "Identification of functional murine adenosine deaminase cDNA clones by complementation in Echerichia coli," *J. Bio. Chem. 260/18*:10299-10307 (1985), Figure 3. By changing both -3 and +1 purines to pyrimidines, fully impaired consens Kozak sequences result. The encoded amino acid corresponding to GCC (alanine) is changed to either proline (CCC) or serine (TCC), with the change to serine being preferred, due to the conservative nature of this change.

This Sheet Has Been Replaced



The reported partially impaired consensus Kozak for asparagine synthetase is as follows:

-3 +1

gc Acc ATG Tgt

See, Andrulis, I.L. *et al.*, "Isolation of human cDNAs for asparagine synthetase and expression in Jensen rat sarcoma cells," *Mol. Cell. Bio.* 7/7:2435-2443 (1987). Changing the +3 purine to a pyrimidine results in a fully impaired consensus Kozak.

The partially impaired consensus Kozak for neomycin phosphotransferase (which includes an upstream out of frame start codon) is as follows:

-3 +1

ggA TGg gga tcg ttt Cgc ATG Att

Changing the +1 purine to a pyrimidine has the effect of creating a fully impaired consensus Kozak (changes to the encoded amino acid isoleucine, ATT are as follows: CTT - leucine and TTT-phenylalanine, with the change to leucine being preferred, due to the conservative nature of this change).

The foregoing is not intended, nor is it to be construed as limiting; rather, in the context of the disclosed invention, the foregoing is presented in an effort to provide equivalent examples of changes in the reported consensus Kozak sequences or partially impaired consensus Kozak sequences of several well-known dominant selectable markers.

As noted, a most preferred dominant selectable marker is NEO. Particularly preferred fully impaired consensus sequences for NEO are as follows:

Txx ATG Ctt (SEQ ID NO: 3)
Cxx ATG Ctt (SEQ ID NO: 4)
Txx ATG Ttt (SEQ ID NO: 5)
Cxx ATG Ttt (SEQ ID NO: 6)

where x are nucleotides. SEQ. ID. NO. 3 is most preferred; and xx are preferably CC.

Other transcriptional cassettes, which may or may not include a fully impaired consensus Kozak, can be

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

incorporated into a vector which includes transcriptional cassettes containing the disclosed and claimed fully impaired consensus Kozak; such "other transcriptional cassettes" typically are utilized to allow for "enhancement," "amplification" or "regulation" of gene product repression. For example, co-transfection of the exogenous DNA with the dehydrofolate reductase (DHFR) gene is exemplary. By increasing the levels the antifolate drug methotrexate (MTX), a competitive inhibitor of DHFR, presented to such cells, an increase in DHFR production can occur via amplification of the DHFR gene. Beneficially, extensive amounts of flanking exogenous DNA will also become amplified; therefore, exogenous DNA inserted co-linear with an expressible DHFR gene will also become overexpressed. Additionally, transciptional cassettes which allow for regulation of expression are available. For example, temperature sensitive COS cells, derived by placing SV40ts mutant large T antigen gene under the direction of Rous sarcoma virus LTR (insensitive to feedback repression by T antigen), has been described. *See*, 227 *Science* 23-28 (1985). These cells support replication from SV40 *ori* at 33°C but not at 40°C and allow regulation of the copy number of transfected SV40 *ori*-containing vectors. The foregoing is not intended, nor is it to be construed, as limiting; rather the foregoing is intended to be exemplary of the types of cassettes which can be incorporated into expression vectors comprising the disclosed fully impaired consensus Kozak. The skilled artisan is credited with the ability to determine the specific type of other transciptional cassettes, vis-a-vis the objective of the expression system, which are applicable and which can be advantageously exploited.

As indicated above, in particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without a stop codon being located between the out-of-frame start codon and the

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

fully impaired consensus Kozak start codon. The intent of the out-of-frame start codon is to, in effect, further impair translation of the dominant selectable marker.

As used herein, the term "stop codon" is meant to indicate a "nonsense codon," *ie* a codon which does not encode one of the 20 naturally occurring amino acids such that translation of the encoded material terminates at the region of the stop codon. This definition includes, in particular, the traditional stop codons TAA, TAG and TGA.

As used herein, the term "out-of-frame" is relative to the fully impaired consensus Kozak start codon. As those in the art appreciate, in any DNA macromolecule (or RNA macromolecule) for every in-frame sequence, there are two out-of-frame sequences. Thus, for example, with respect to the following sequence incorporating a fully impaired consensus Kozak:

```
                    -3      +1
gcA TGc cAT Ggc Cxx ATG Cxx
```

the in-frame codons are separated by triplets, *eg*, gcA, TGc, cAT and Ggc; the out-of-frame codons would include, *eg* cAT, ATG, Gcc, ccA, ATG and TGg. Thus, two start codons (in capital letters and underlined) are out-of-frame relative to the start codon of the fully impaired consensus Kozak.

When such an out-of-frame start codon is utilized, it is preferred that this be within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides of the fully impaired consensus Kozak start codon.

As is appreciated, the upstream sequence can be manipulated to achieve positioning at least one out-of-frame sequence upstream of the fully impaired consensus Kozak start codon using (most preferably) a mutational primer used in the type of amplification protocol described above.

This Sheet Has Been Replaced



Utilization of a fully impaired consensus Kozak start codon located within a secondary structure (*ie* a "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable marker. In such an embodiment, it is preferred that this start codon be located within the stem of a stem loop secondary structure. These, by way of schematic example, in such an embodiment, the start codon of the fully impaired consensus Kozak is positioned as follows:

```
          T
        x   x
         x x
         x A
         x T
         x G
      xA TGxx Cxx
```

(An out-of-frame start codon which is not part of the secondary structure is also represented.) As is appreciated, within the stem loop, complementarity along the stem is, by definition, typically required. For exemplary methodologies regarding, *inter alia*, introduction of such secondary structures into the sequence, as well as information regarding secondary structure stability, *see*, Kozak, *PNAS*, 1986, supra.

As noted, it is preferred that a dominant selectable market with a naturally occurring intronic insertion region or an artificially created intronic insertion region be utilized and at least one gene product of interest inserted within this region. While not wishing to be bond by any particular theory, the inventor postulates that such an arrangement increases expression efficiency because the number of viable colonies that survive the selection process via-a-vis the dominant selectable marker will decrease; the colonies that do survive the selection process will, by definition, have expressed the protein necessary for survival, and in conjunction therewith, the gene product of interest will

This Sheet Has Been Replaced



This is the replacement sheet

Date:

have a greater tendency to be expressed. As further postulated, the RNA being transcribed from the gene product of interest within the intronic insertion region interferes with completion of transcription (elongation of RNA) of the dominant selectable marker; therefore, the position that the dominant selectable marker is integrated within the cellular DNA is likely to be a position where a larger amount of RNA is initially transcribed.

As is appreciated, prokaryotic proteins do not typically include splices and introns. However, the majority of dominant selectable markers which are preferred for expression vector technology are derived from prokaryotic systems. Thus, when prokaryotic-derived dominant selectable markers are utilized, as is preferred, it is often necessary to generate an artificial splice within the gene so to create a location for insertion of an intron comprising the gene product of interest. It is noted that while the following rules are provided for selection of a splice site in prokaryotic genes, they can be readily applied to eukaryotic genes.

A general mechanism for the splicing of messenger RNA precursors in eukaryotic cells is delineated and summarized in Sharp, Philip A. "Splicing of Messenger RNA Precursors" *Science, 235*: 736-771 (1987) (*see*, in particular, Figure 1) which is incorporated herein by reference. Based upon Sharp, there are four minimum criteria in the nucleic acid sequence which are necessary for a splice: (a) 5' splice donor; (b) 3' splice acceptor; (c) branch point, and (d) polypyrimidine tract. The consensus sequences for the 5' splice donor is reported to be

$$\begin{matrix} C & & A \\ AAG/GTGAGT & & \end{matrix}$$

and for the 3' splice acceptor, NCAG/G (where a "/" symbol indicates the splice site); *see*, Mount, S.M. "A Catalogue of Splice Junction Sequences" *Nuc. Acids. Res. 10/2*:459-472 (1992) which is incorporated herein by reference. The consensus sequence for the branch point, *ie.* the location of the lariat formation with the 5' splice donor, is

This Sheet Has Been Replaced

This is the replacement sheet

11 FEB 1997

Date

reported as PyNPyPAPy; and the reported preferred branch site for mammalian RNA splicing is TACTAAC (Zhuang, Y. et al. "UACUAAC is the preferred branch site for mammalian mRNA splicing" *PNAS 86*: 2752-2756 (1989), incorporated herein by reference). Typically, the branch point is located at least approximately 70 to about 80 base-pairs from the 5'-splice donor (there is no defined upper limit to this distance). The poly pyrimidine tract typically is from about 15 to about 30 base pairs and is most typically bounded by the branch point and the 3' splice acceptor.

The foregoing is descriptive of the criteria imposed by nature on naturally occurring splicing mechanisms. Because there is no exact upper limit on the number of base pairs between the 5' splice donor and the branch point, it is preferred that the gene product of interest be inserted within this region in situations where a natural intron exists within the dominant selectable marker. However, as noted, such introns do not exist within most of the preferred dominant selectable markers; as such, utilization of artificial introns are preferably utilized with these markers.

In order to generate an artificial intron, a "splice donor:splice acceptor" site must be located within the encoding region of the dominant selectable marker. Based upon Sharp and Mount, it is most preferred that the following sequence function as the splice donor:splice acceptor site -- CAGG (with the artificial splice occurring at the GG region). A preferred sequence is AAGG.

Focusing in on the most preferred sequence CAGG, the following codons and amino acids can be located within the encoding region of the dominant selectable marker for generation of the artificial intronoc insertion region:

This Sheet Has Been Replaced



11 FEB 1997

A		B		C		
CAG/	GNN	NCA	G/GN	NNC		AG/G
Gln	Ala	Ala	Gly	Ala	Leu	Arg
	Asp	Pro		Arg	Phe	
	Gly	Ser		Asn	Pro	
	Glu	Thr		Asp	Ser	
	Val			Cys	Thr	
				Gly	Tyr	
				His	Val	
					Ile	

(As will be appreciated, the same approach to determining viable amino acid residues can be utilized for the preferred sequence of AAGG). The most preferred codon group for derivation of the splice donor:splice acceptor site is group A. Once these amino acid sequences are located, a viable point for generation of an artificial intronic insertion region can be defined.

Focusing on the preferrred NEO dominant selectable marker, amino acid residues Gln Asp (codon group A) are located at the positions 51 and 52 of NEO and amino acid residues Ala Arg (codon group C) are located at positions 172 and 173 (as is appreciated, multiple artificial intronic insertion regions may be utilized). Focusing on residues 50 - 53 of NEO, the nucleic acid and amino acid sequences are as follows:

	50	51	52	53	
5'	CTG	CAG	GAC	GAG	3'
	Leu	Gln	Asp	Glu	

Accordingly, an artificial intronic insertion region can be generated between residues 51 and 52 of NEO. This region most preferably comprises a branch point, a polyprimidine tract and, preferably, a region for insertion of a gene product of interest, *ie* a region amendable to enzymatic digestion.

Two criteria are import for the artificial intronic insertion region: the first two nucleic acid residues of the 5' splice site (*eg* abutting CAG) are most preferably

This Sheet Has Been Replaced

This is the replacement sheet

Date:

GT and the first two nucleic acid residues of the 3' splice site (*eg* abutting G) are most preferably AG.

Using the criteria defined abouve, an artificial intronic insertion region was between amino acid residues 51 and 52 of NEO:

```
   50 NEO 51                                                    52  NEO 53
   LEU GLN                        Branch          Polypyrimide Tract ASP GLU 3'
5' CTG CAG/GTAAGT GCGGCCGC TACTAAC (TC)3CT(C)3TCC(T)5 C CTGCAG/GAC  GAG
                  Not I
```

(Details regarding the methology for creating this artificial intronic insertion region are set forth in the Example Section to follow). The Not I site was created as the region where the gene product of interest can be incorporated. Therefore, upon incorporation, the gene product of interest is located between amino acid residues 51 and 52 of NEO, such that during NEO transmission, the gene product of interest will be "spliced-out".

The host cell line is most preferably of mammalian origin; those skilled in the art are credited with ability to preferentially determine particular host cell lines which are best suited for the desired gene product to be expressed therein. Exemplary host cell lines include, but are not limited to, DG44 and DXB11 (Chinese Hamster Ovary lines, DHFR minus), HELA (human cervical carcinoma), CV1 (monkey kidney line), COS (a derivative of CV1 with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3x63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte) and 293 (human kidney). Host cell lines are typically available from commercial services, the American Tissue Culture Collection or from published literature.

Preferably the host cell line is either DG44 or SP2/O. *See*, Urland, G. *et al.*, "Effect of gamma rays and the dihydrofolate reductase locus: deletions and inversions." *Som. Cell & Mol. Gen. 12/6*:555-566 (1986) and Shulman, M. *et al.*, "A better cell line for making

This Sheet Has Been Replaced This is the replacement sheet

Date: 11 FEB 1997

hybridomas secreting specific antibodies." *Nature 276*:269 (1978), respectively. Most preferably, the host cell line is DG44. Transfection of the plasmid into the host cell can be accomplished by any technique available to those in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), cell fusion with enveloped DNA, microinjection, and infection with intact virus. *See*, Ridgway, A.A.G. "Mammalian Expression Vectors." Chapter 24.2, pp. 470-472 *Vectors*, Rodriguez and Denhardt, Eds. (Butterworths, Boston, MA 1988). Most preferably, plasmid introduction into the host is via electroporation.

Examples

The following examples are not intended, nor are they to be construed, as limiting the invention; the examples are intended to demonstrate the applicability of an embodiment of the invention disclosure herein. The disclosed fully impaired consensus Kozak sequence is intended to be broadly applied as delineated above. However, for presentational efficiency, exemplary uses of particularly preferred embodiments of fully impaired consensus Kozak sequences are utilized in conjunction with tandem chimeric antibody expression vectors (also referred to herein as antibody expression vectors) as disclosed below.

I. Tandem Chimeric Antibody Expression ("TCAE") Vector

B cell lymphocytes arise from pluripotent stem cells and proceed through ontogeny to fully matured antibody secreting plasma cells. The human B lymphocyte-restricted differentiation antigen Bp35, referred to in the art as "CD20," is a cell surface non-glycosylated phosphoprotein of 35,000 Daltons; CD20 is expressed during early pre-B cell development just prior to the expression of cytoplasmic μ heavy chains. CD20 is expressed consistently until the plasma cell differentiation stage. The CD20 molecule regulates a step in the activation

This Sheet Has Been Replaced

11 FEB 1997

process which is required for cell cycle initiation and differentiation. Because CD20 is expressed on neoplastic B cells, CD20 provides a promising target for therapy of B cell lymphomas and leukemias. The CD20 antigen is especially suitable as a target for anti-CD20 antibody mediated therapy because of accessibility and sensitivity of hematopoietic tumors to lysis via immune effector mechanisms. Anti-CD20 antibody mediated therapy , *inter alia*, is disclosed in co-pending Serial No. 07/978,891 and Serial No. _____ , filed simultaneously herewith. The antibodies utilized are mouse/human chimeric anti-CD20 antibodies expressed at high levels in mammalian cells (chimeric anti-CD20"). This antibody was derived using vectors disclosed herein, to wit: TCAE 5.2; ANEX 1; ANEX 2; GKNEOSPLA3F; and NEOSPLA3F (an additional vector, TCAE 8, was also utilized to derive chimeric anti-CD20 antibody -- TCAE 8 is identical to TCAE 5.2 except that the NEO translational start site is a partially impaired consensus Kozak. TCAE 8 is described in the co-pending patent document filed herewith.).

In commonly-assigned United States Serial Number 07/912,292, disclosed, *inter alia*, are human/Old World monkey chimeric antibodies; an embodiment of the invention disclosed therein are human/macaque chimeric anti-CD4 antibodies in vector TCAE 6 (*see*, Figure 6 of Serial Number 07/912,292, and corresponding discussion). TCAE 6 is substantially identical to TCAE 5.2; TCAE 6 contains human lambda constant region, while TCAE 5.2 contains human kappa constant region. TCAE 5.2 and ANEX 1 (referred to in that patent document as TCAE 12) are disclosed as vectors which can be utilized in conjunction with human/Old World monkey chimeric antibodies. The comparative data set forth in Serial No. 07/912,292 vis-a-vis TCAE 5.2 and ANEX 1 is relative to expression of chimeric anti-CD20 antibody.

TCAE 5.2 was derived from the vector CLDN, a derivative of the vector RLDN10b (*see*, 253 *Science* 77-91, 1991). RLDN10b is a derivative of the vector TND (*see*,

This Sheet Has Been Replaced

This is the replacement sheet

Date:

7*DNA* 651-661, 1988). *Ie* the vector "family line" is as follows: TDN → RLDN10b → CLDN → TCAE 5.2 → ANEX 1 (the use of the "→" symbol is not intended, nor is it to be construed, as an indication of the effort necessary to achieve the changes from one vector to the next; *eg* to the contrary, the number and complexity of the steps necessary to generate TCAE 5.2 from CLDN were extensive).

TND was designed for high level expressions of human tissue plasminogen activator. RLDN10b differs from TND in the following ways: the dihydrofolate reductase ("DHFR") transcriptional cassette (comprising promoter, murine DHFR cDNA, and polyadenylation region) was placed in between the tissue plasminogen activator cassette ("t-PA expression cassette") and the neomycin phosphotransferase ("NEO" cassette") so that all three cassettes were in tandem and in the same transcriptional orientation. The TND vector permitted selection with G418 for cells carrying the DHFR, NEO and t-PA genes prior to selection for DHFR gene amplification in response to methotrexate, MTX. The promoter in front of the DHFR gene was changed to the mouse beta globin major promoter (*see*, 3 *Mol. Cell. Bio.* 1246-1254, 1983). Finally, the t-PA cDNA was replaced by a polylinker such that different genes of interest can be inserted in the polylinker. All three eukaryotic transcriptional cassettes (t-PA, DHFR, NEO) of the TND vector can be separated from the bacterial plasmid DNA (pUC19 derivative) by digestion with the restriction endonuclease Not I.

CLDN differs from RLDN10b in the following ways: The Rous LTR, positioned in front of the polylinker, was replaced by the human cytomegalovirus immediate early gene promoter enhancer ("CMV"), (*see*, 41 *Cell* 521, 1985), from the Spe I site at -581 to the Sst I site at -16 (these numbers are from the *Cell* reference).

As the name indicates, TCAE vectors were designed for high level expressions of chimeric antibody. TCAE 5.2 differs from CLDN in the following ways:

This Sheet Has Been Replaced

This is the replacement sheet

Date:

A. TCAE 5.2 comprises four (4) transcriptional cassettes, as opposed to three (3), and these are in tandem order, *ie* a human immunoglobulin light chain absent a variable region; a human immunoglobulin heavy chain absent a variable region; DHFR; and NEO. Each transcriptional cassette contains its own eukaryotic promoter and polyadenylatin region (reference is made to Figure 2 which is a diagrammatic representation of the TCAE 5.2 vector). The CMV promoter/enhancer in front of the immunoglobulin heavy chain is a truncated version of the promoter/enhancer in front of the light chain, from the Nhe I site at -350 to the Sst I site at -16 (the numbers are from the *Cell* reference, *supra*). Specifically,

1) A human immunoglobulin light chain constant region was derived via amplification of cDNA by a PCR reaction. In TCAE 5.2, this was the human immunoglobulin light chain kappa constant region (Kabat numbering, amino acids 108-214, allotype Km 3), and the human immunoglobulin heavy chain gamma 1 constant region (Kabat numbering amino acids 114-478, allotype Gm1a, Gm1z). The light chain was isolated from normal human blood (IDEC Pharmaceuticals Corporation, La Jolla, CA); RNA therefrom was used to synthesize cDNA which was then amplified using PCR techniques (primers were derived vis-a-vis the consensus Kabat). The heavy chain was isolated (using PCR techniques) from cDNA prepared from RNA which was in turn derived from cells transfected with a human IgG1 vector (*see*, 3 *Prot. Eng.* 531, 1990; vector $pN_{\gamma 1}62$). Two amino acids were changed in the isolated human IgG1 to match the consensus amino acid sequence in Kabat, to wit: amino acid 225 was changed from valine to alanine (GTT to GCA), and amino acid 287 was changed from methionine to lysine (ATG to AAG);

2) The human immunoglobulin light and heavy chain cassettes contain synthetic signal sequences for secretion of the immunoglobulin chains;

This Sheet Has Been Replaced

11 FEB 1997

3) The human immunoglobulin light and heavy chain cassettes contain specific DNA restriction sites which allow for insertion of light and heavy immunoglobulin variable regions which maintain the transitional reading frame and do not alter the amino acids normally found in immunoglobulin chains;

4) The DHFR cassette contained its own eukaryotic promoter (mouse beta globin major promoter, "BETA") and polyadenylation region (bovine growth hormone polyadenylation, "BGH"); and

5) The NEO cassette contained its own eukaryotic promoter (BETA) and polyadenylation region (SV40 early polyadenylation, "SV").

With respect to the TCAE 5.2 and the NEO cassette, the Kozak region was a consensus Kozak (which included an upstream Cla I site SEQ ID NO: 7):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccAcc ATG Gtt

ANEX 1 (previously named TCAE 12 in the referenced case) is identical to TCAE 5.2 except that in the NEO cassette, the Kozak region was fully impaired (SEQ ID NO: 8):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccTcc ATG Ctt

As disclosed in the commonly-assigned referenced case, the impact of utilization of the fully impaired consensus Kozak was striking: relative to TCAE 5.2, there was a significant (8-fold) reduction in the number of ANEX 1 G418 resistant colonies (258 from two electroporations versus 98 from six electroporations) from the same amount of plasmid DNA transfected per cell; and, there was a significant increase in the amount of co-linked gene product expressed in each of the ANEX 1 clones. Referencing the histogram of Figure 3 (Figure 16 of the commonly assigned referenced case), 258 colonies were derived from 2 electroporations of 25 µg of DNA containing a neomycin phosphotransferase gene with a consensus Kozak at the translation start site. Two-hundred and one (201)

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

of these colonies did not express any detectable gene product (less than 25 ng/ml of chimeric immunoglobulin), and only 8 colonies expressed more than 100 ng/ml. Again, referencing Figure 3, 98 colonies were derived from 6 electroporations for ANEX 1 of 25 µg of DNA containing a neomycin phosphotransferase gene with the fully impaired consensus Kozak at the translation start site (6 electroporations were utilized in order to generate statistically comparative values; this was because on average, each electroporation for ANEX 1 yielded about 16 colonies, as opposed to about 129 colonies per electroporation for TCAE 5.2). Eight (8) of the ANEX 1 colonies did not express any detectable gene product (less than 25 ng/ml), while 62 of these colonies were expressing greater than 100 ng/ml; of these 62 colonies, nearly 23 were expressing over 250 ng/ml (23%), with 6 expressing greater than 1000 ng/ml (6%).

The foregoing evidences, *inter alia*, the following: 1) because the difference between TCAE 5.2 and ANEX 1 was limited to the Kozak translation start site of the NEO gene, and because the gene product of interest (chimeric anti-CD20 antibody) was co-linked to the NEO gene, a conclusion to be drawn is that these differences in results are attributed solely to the differences in the Kozak translation start site; 2) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable number resulted in significantly less viable colonies; 3) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable marker co-linked to a desired gene product significantly increased the amount of expressed gene product. Thus, the number of colonies to be screened decreased while the amount of expressed gene product increased.

This Sheet Has Been Replaced



This is an [illegible] document sheet

Date: [illegible]

II. Impact Of Out-Of-Frame Start Sequence

Conceptually, further impairment of translation initiation of the dominant selectable marker of ANEX 1 could be effectuated by utilization of at least one out-of-frame ATG start codon upstream of the neomycin phosphotransferase start codon. Taking this approach one step further, utilization of a secondary structure ("hairpin") which incorporated the *neo* start codon within the stem thereof, would be presumed to further inhibit translation initiation. Thus, when the out-of-frame start codon/fully impaired consensus Kozak was considered, this region was designed such that the possibility of such secondary structures was increased.

As indicated previously, the Kozak region for the *neo* start codon in the ANEX 1 vector is:

TTGGGAGCTTGG ATCGAT CC Tcc ATG Ctt

The desired sequence for a vector identical to ANEX 1 but incorporating the above-identified changes vis-a-vis the *neo* start codon, referred to as ANEX 2, is as follows (SEQ ID NO: 9):

CCA GCA TGG AGG A ATCGAT CC Tcc ATG Ctt

(The out-of-frame start codon is underlined.) The fully impaired consensus Kozak of ANEX 2 is identical to that of ANEX 1. The principal difference is the inclusion of the upstream out-of-frame start codon. A possible difference is the formation of a secondary structure involving this sequence, proposed as follows:

This Sheet Has Been Replaced



11 FEB 1997

```
      CG   )
     T  A  )  CLA I site
     A  T  )
      AT
      GC
      GC
      AT
      GC
      GC
      TA
      AT
      CG
      GC
      AT
      CT
      CG
```

The sequence in bold, ATG, is the upstream out-of-frame start codon; the "loop" portion of the secondary structure is the CLA I site; and the sequence between the "T" and "C" (italics and bold) is the start codon (underlined) of the fully impaired consensus Kozak.

In order to effectuate this change, a PCR fragment was cloned into anti-CD20 in ANEX 1 from Xho I (5520) to Cla I (5901); see, Figure 4. Primers were as follows:

3'-Primer 489 (SEQ ID NO: 10):

```
5'-GGA GGA TCG ATT CCT CCA TGC TGG
   CAC AAC TAT GTC AGA AGC AAA TGT
   GAG C-3'
```

The upper-lined portion of Primer 489 is a Cla I site; the under-lined portion is the fully impaired consensus Kozak translation start site.

5'-Primer 488 (SEQ. ID. NO. 11):

```
5'-CTG GGG CTC GAG CTT TGC-3'
```

The upper-lined portion of Primer 485 is an Xho I site.

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

These primers were prepared using an ABI 391 PCR MATE™ DNA synthesizer (Applied Biosystems, Foster City, CA). Phosphoramidites were obtained from Cruachem (Glasgow, Scotland): dA(bz) - Prod. No. 20-8120-21; dG(ibu) - Prod. No. 20-8110-21; dC(bz) - Prod. No. 20-8130-21; T - Prod. No. 20-8100-21.

Conditions for the PCR reaction using these primers were as follows: 2λ ("microliters") of anti-CD20 in TCAE 5.2 in plasmid grown in *E. coli* strain GM48 (obtained from the ATCC) was admixed with 77λ of deionized water; 2λ of Primer 488 (64 pmoles); and 4λ of primer 489 (56 pmoles). This was followed by a denaturation step (94°C, 5 min.) and a renaturation step (54°C, 5 min.). Thereafter, 4λ of 5 mn dNTPS (Promega, Madison, WI: dATP, Prod. No. U1201; dCTP, Prod. No. U1221; dGTP, Prod. No. U1211; dTTp, Prod. No. U1231), 1λ of Pfu DNA polymerase (Stratagene, La Jolla, CA Prod. No. 600135, 2.5 U/ml), and 50λ of mineral oil overlay was added thereto, followed by 30 cycles, with each cycle comprising the following: 72°C, 2 min.; 94°C, 1 min.; 54°C, 1 min. Ten microliters (10λ) of this admixture was analyzed by agarose gel electrophoresis (results not shown); a single band was found at about 400 base pairs.

The PCR product and the vector were prepared for ligation as follows: Anti-CD20 in ANEX 1 plasmid grown in *E. coli* bacterial strain GM48 was digested with Cla 1 and Xho 1 as follows: 20λ of anti-CD20 in ANEX 1 was admixed with 10λ of 10XNEB4 buffer (New England Biolabs, Beverly, MA; hereinafter, NEB); 5λ Cla 1 (NEB, Prod. No. 197 S, 60u); and 64λ deionized water. This admixture incubated overnight at 37°C, followed by the addition of 5λ Xho 1 (NEB, Prod. No. 146 S, 100 u) and incubation at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut ANEX 1". The approximate 400 base pair PCR fragment was prepared and digested with Cla 1 and Xho 1 as follows: 90λ of the PCR fragment was admixed with 10λ of 3M NaOAc; 1λ 10% sodium dodecyl sulfate (SDS); and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30

This Sheet Has Been Replaced

sec. followed by a 1 min. spin (1700 RPM). The aqueous phase was subjected to a spin column which resulted in 85λ total admixture. To this admixture was added 10λ 10XNEB4, 1λ bovine serum albumin (BSA,100X; NEB), 2λ Cla 1 (24u), and 2λXho 1 (40 u). This admixture was incubated at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut PCR 488/489". Both Cla 1/Xho 1 cut ANEX 1 and Cla 1/Xho 1 cut PCR 488/489 were analyzed by agarose gel electrophoresis and the resulting bands were observed at the same relative location on the gel (results not shown).

Ligation of Cla 1/Xho 1 cut PCR 488/489 and Cla 1/Xho 1 cut ANEX 1 was accomplished as follows: 1λ of tRNA (Sigma, St. Louis, MO, Prod. No. R-8508) was admixed with 1λ 10% SDS; 10λ 3M NaOAc; 45λ of Cla 1/Xho 1 cut PCR 488/489 (about 22.5 ng); a 1:4 dilution (0.25λ) of Cla 1/Xho 1 cut ANEX 1 (about 32 ng) in 0.75λ tris-hydroxymethyl aminomethane ethylenediamine tetracetic acid (TE); and 42λ TE. To this admixture was added 90λ phenyl/$CHCl_3$/isoamyl, followed by a 30 sec. vortex and a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube, followed by addition of 270λ of 100% EtOH (-20° C), 10 min. spin (13,000 RPM) followed by addition of another 270λ of 100% EtOH (-20° C) and another 1 min. spin (13,000 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 17λ TE, 2λ ligase buffer (Promega, T4 DNA Ligase kit, Prod. No. M180) and 1λ ligase (Promega Ligase kit). This ligation mix incubated at 14°C overnight. Twenty microliters (20λ) of the ligation mix was admixed with 10λ 3M NaOAc, 1λ 10% SDS, 69λ TE, and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30 sec., followed by a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube and 270λ of 100% EtOH (-20°C) was added thereto, followed by a 10 min. spin (1700 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 20λ TE. Ten microliters of the resuspended admixture was transformed in *E. coli* X-L1 blue™ (Stratagene, La Jolla, CA), following manufacturer

This Sheet Has Been Replaced

11 FEB 1997

instructions. Ten (10) bacterial colonies were inoculated in LB Broth (Gibco BRL, Grand Island, NY, Prod. No. M27950B) including ampicillin (50µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification System (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the ANEX 2 vector, depending on the sufficiency of the foregoing.

ANEX 2 includes a Hinf I site ("GAATC") upstream of the *neo* start site (-9 to -13 relative to the *neo* start codon); ANEX 1 does not include this Hinf I site. The purified plasmids comprising putative ANEX 2, and previously purified ANEX 1 standard , were subjected to Hinf 1 digestion as follows: 2λ of each isolate was admixed with 8λ of Hinf I digestion buffer (15λ 10 x NEB2 buffer; 15λ Hinf I (NEB, Prod. No. 155S, 10u/λ); and 90λ H_2O). This admixture incubated for 3 hrs. at 37°C and each isolate was analyzed via agarose gel electrophoresis (results not shown); nine (9) of the bands were substantially identical to the ANEX 1 standard, one(1) showed a slight difference in band pattern. For this single isolate, the first two bands were at 1691 and 670 kB; for the ANEX 1 Hinf I digested product, the first three bands were at 1691, 766, and 670 kB. The missing band at 766 kB for the single isolate was attributed to the presence of the Hinf I site therein, indicating that the desired change to ANEX 1 was incorporated into this vector. This vector was designated "Anti-CD20 in ANEX 2 (G1,K)," and is generally referred to by the inventor as ANEX 2.

Electroporation of anti-CD20 in ANEX 2 was accomplished as follows: two-hundred and forty microliters (240λ) of the anti-CD20 in ANEX 2 DNA (400µg) was admixed with 100λ of 10 X NEB2 buffer ; 100λ of Stu I (NEB, Prod. No. 187S, 1000 u); and 560λ TE, and incubated at 37°C for 2hrs. This admixture was then placed over 8 spin columns (125λ each), followed by addition of 110λ 10X Not I buffer (NEB); 10λ 100X BSA; and 20λ of Not I (NEB,

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

Prod. No. 189S, 800u). This admixture was incubated at 37°C for 3 hrs., followed by the addition of 120λ of 3M NaOAC and 12λ of 10% SDS. The admixture was transferred to 2 vortex tubes and 500λ of phenyl/$CHCl_3$/isoamyl was added to each, followed by a 30 sec. vortex and 1 min. spin (1700 RPM). The aqueous phase was removed from the tubes and segregated into 3 tubes, followed by the addition to each tube of -20°C 100% ETOH, followed by 10 min. spin (13,000 RPM). Thereafter, -20°C 70% ETOH was added to each tube, followed by 1 min. spin (13,000 RPM). The tubes were then placed in a Speed VAC™ for drying, followed by resuspension of the contents in 100λ TE in a sterile hood. Five microliters (5λ) of the resuspended DNA was admixed with 995λ of deionized water (1:200 dilution). An optical density reading was taken (OD=260) and the amount of DNA present was calculated to be 0.75μg/λ. In order to utilize 25 μg of DNA for electroporation, 32λ of the 1:200 dilution of the DNA was utilized (25 μg was utilized as this was the amount of DNA utilized for TCAE 5.2 and ANEX 1 in the foregoing Example 1). The 1:200 dilution of DNA was formally referred to as "Stu 1, Not I cut anti-CD20 in ANEX 2 (25μg) in TE" and generally referred to as "anti-CD20 in ANEX 2."

Host cells utilized was DG44 CHO ("CHO") (*see*, Urlaub, G. *Somatic Cell*, 1986 *supra*). One hundred milliliters of 6.6×10^5 cells/ml (84%) were subjected to a 2 min. spin at 1000 RPM. These were washed with 50 ml sucrose buffered solution, followed by 5 min. spin at 1000 RPM; the material was then resuspended in 4.5 ml of the sucrose buffered solution. Thereafter, cells were counted and 0.4 ml of CHO cells (4.0×10^6 cells) were admixed with 32λ of the anti-CD20 in ANEX 2 in BTX sterile, disposable electroporation cuvettes. Electroporation settings were as follows: 210 volts; 400 microfaraday capacitance; 13 ohms resistance, using a BTX 600™ electro cell manipulator (BTX, San Diego, CA). Nine (9) electroporations were conducted; actual voltage delivered over actual times were as follows: 1-199V, 4.23 msec; 2-

This Sheet Has Been Replaced



This is the replacement sheet

Date:

188V, 4.57 msec; 3-189V, 4.24 msec, 4-200V; 4.26 msec, 5-200V, 4.26 msec; 6-199V, 4.26 msec; 7-189V, 4.59 msec; 8-189V, 4.57 msec; 9-201 V, 4.24 msec. (As noted in Example I, the difference in number of performed electroporations was attributed to the need to achieve a statistically significant number of viable colonies for each of the three conditions. TCAE 5.2, ANEX 1 and ANEX 2; the amount of DNA used for each electroporation (25 µg) was the same for each, and the same number of cells were electroporated.

Thereafter, the electroporation material was admixed with 20 ml of G418 Growth Media (CHO-S-SFM II minus hypoxanthine and thymidine (Gibco, Grand Island, NT, Form No. 91-0456PK) including 50 µM hypoxanthine and 8 µM thymidine). The admixture was gently agitated, followed by plating 200 µl of the admixture per well into 96-well plates, one plate for each electroporation (nine). Beginning on day 2 after electroporation, through day 17, 150µl of each well was removed, and 150 µl of fresh G418 Growth Media containing 400 µg/ml G418 was added thereto. Colonies were analyzed on day 25.

One hundred and twenty one (121) colonies expressed anti-CD20 antibody (*ie* , 13 colonies per electroporation). Of these, 63 (52%) expressed over 250µg /ml of protein; of the 63, 20 of the colonies (16.5%) expressed over 1000µg/ml of protein. Only 5 of the 121 colonies (4.1%) expressed less than 25µg/ml of protein. Figure 5 provides a histogram comparing expression of protein per colonies derived from the vectors TCAE 5.2, ANEX 1 and ANEX 2.

The foregoing data indicates that, *inter alia*, as between ANEX 1 and ANEX 2, the use of at least one out-of-frame start codon upstream of a fully impaired consensus Kozak associated with the translation initiation of a dominant selectable marker decreases the number of viable colonies expressing co-linked gene product and significantly increases the amount of expressed co-linked gene product.

This Sheet Has Been Replaced



III. Impact Of Insertion Of Gene Product Of Interest Within At Least One Artificial Intronic Insertion Region Of A Dominant Selectable Marker

Building further upon the ANEX 2 vector, an artificial splice was generated between amino acid residues 51 and 52 of the NEO coding region of ANEX 2, followed by insertion therein of the anti-CD20 encoding region. Two such vectors were generated: the first, comprising a consensus Kozak sequence for the NEO translation initiation codon and not comprising an out-of-frame start codon, is referred to as "GKNEOSPLA3F;" the second, comprising a fully impaired consensus Kozak and an out-of frame start condon, is referred to as "NEOSPLA3F."

Both GKNEOSPLA3F and NEOSPLA3F contain the following artificial intron sequence between amino acid residues 51 and 52 of NEO:

51 52

5'CTG CAG/GTAAGT GCGGCCGC TACTAAC $(TC)_3$ CT $(C)_3$ TCC $(T)_5$ C CTGCAG/GAC GAG 3'

The underlined portion represents a sequence amenable to digestion with Not I enzyme; the encoding region for anti-CD20, *inter alia*, was inserted within this region.

Although not wishing to be bound by any particular theory, the inventor postulates that during expression, the inclusion of a gene product of interest within an artificial intronic insertion region of a dominant selectable marker (*e.g.* the NEO gene) should significantly decrease the number of viable colonies producing, in the case of the disclosed GKNEOSPA3F and NEOSPLA3F vectors, anti-CD20 antibody. This is predicated upon two points: first, only those vectors which are able to transcribe and correctly splice-out the antibody encoding region and correctly translate NEO will be G418 resistant; second, because each antibody cassette has its own promoter and polyadenylation region, transcription and translation of the antibody is independent of translation of NEO.

The GKNEOSPLA3F and NEOSPLA3F vectors were constructed in the following manner:

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

Anti-CD20 in ANEX 2 was digested with Not I and Xho I in order to isolate the 1503 bp NEO cassette DNA fragment (see Figure 4 between "Not I 7023" and "Xho I 5520") as follows: 10 µl of anti-CD20 in ANEX 2 was admixed with 6 µl deionized H_2O ("dH_2O"); 1 µl Not I enzyme (NEB, Prod. No. 189S); 2 µl of 10X Not I digestion buffer (NEB; provided with enzyme); and 1 µ Xho I enzyme (Promega, Madison, WS, Prod. No. R4164). This digestion mixture was incubated overnight at 37°C. The resulting digested DNA was size fractionated by 0.8% agarose gel electrophoresis and the desired fragment migrating at 1503 was isolated via the GlassMAX™ method (Gibco BRL, Grand Island, NY, Prod. No. 15590-011) for insertion into pBluescript SK(-) plasmid DNA (Stratagene, La Jolla, CA).

pBluescript SK (-) was previously prepared for acceptance of the NEO cassette by double digestion with Not I and Xho I using the same conditions as above for anti-CD20 in ANEX 2. Digested pBluescript SK (-) was then collected by ethanol precipitation by the addition of 70 µl dH_2O; 2 µl tRNA (Sigma, St. Louis, MO, Prod. No. R-8508); 10 µl of 3M NaOAc; and 300 µl 100% ETOH (-20°C). This was followed by a 10 min spin (13,000 RPM), decanting the supernatant, rinsing with 70% ETOH, decanting the liquid, drying in a SpeedVAC™ and resuspending in 20 µl 1 x TE.

Ligation of the NEO cassette DNA fragment into prepared pBluescript SK (-) vector was accomplished as follows: 10 µl of NEO fragment DNA was admixed with 6 µl dH_2O; 1 µl cut pBluescript SK (-) vector DNA; 2 µl 10 x ligation buffer (Promega, supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega, Prod. No. M1801) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

This Sheet Has Been Replaced

This is the replacement sheet

FEB 1997

Date

M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO+ depending on the sufficiency of the foregoing. (BlueNEO+ was confirmed due to the sufficiency of the following procedure.)

BlueNEO+ contains a Not I restriction recognition sequence reformed upon ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector. This site was destroyed by the following: 1 µl of BlueNEO+ DNA was admixed with 16 µl dH_2O; 2 µ 10 x Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB). This was followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume. This 15 µl Not I digested DNA was "blunt-ended" by admixing with 4 µl 5X Klenow buffer (20 mM Tris-HCL, pH 8.0, 100mM $MgCl_2$) and 1 µl DNA Polymerase I Large (Klenow) Fragment (Promega, Prod. No. M2201). This admixture was incubated at room temperature for 30 minutes. Blunt-ended DNA was then purified by spin column fractionation, giving a final volume of 15 µl.

Ligation of the blunt-ended DNA was performed in an analogous way as to the ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector except that the final DNA was resuspended in 17 µl of 1 X TE.

Following ligation, the DNA was subjected to a second restriction digestion with Not I by mixing the 17 µl of DNA with 2 µl 10 X Not I digestion buffer and 1 µl Not I enzyme (NEB). Digestion was allowed to proceed at 37°C for 60 minutes. Following digestion, the admixture was purified by spin column fractionation resulting in 15 µl final volume.

Ten (10) µl of the purified DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL including ampicillin (50 µg/ml;

This Sheet Has Been Replaced



Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO- depending on the sufficiency of the foregoing. (BlueNEO- was confirmed due to the sufficency of the following procedure.)

BlueNEO- contains a unique Pst I restriction site spanning the codons for amino acid residues 51 and 52. BlueNEO- was digested with Pst I as follows: an admixture was formed containing 15 µl dH_2O; 1 µl BlueNEO- DNA, 2 µl digestion buffer 3 (NEB) and 2 µl Pst I enzyme (NEB, Prod. No. 140S). This admixture was incubate at 37°C for 3 hrs. Digested DNA was then purified by spin column fractionation. The following synthetic oligonucleotide was then ligated to the Pst I cohesive ends of BlueNEO-: 5' GGTAAGTGCGGCCGCTACTAACTCTCTCCTCCCTCGTTTTTCCTGCA 3' (SEQ ID NO: 12) and its complementary sequence: 5'GGAAAAAGGAGGGAGGAGAGAGTTAGTAGCGGCCGCACTTACCTGCA 3' (SEQ ID NO: 13). Insertion of this linker creates a consensus 5' splice donor site (by ligation) followed by a Not I site, followed by a consensus splice branch point, followed by a synthetic polypyrimidine tract, followed by a consensus 3' splice acceptor site, as indicated above.

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Pst I linearized BlueNEO- DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

The foregoing (and following) synthetic oligo nucleotides were chemically synthesized using an Applied Biosystems 391 PCR MATE™ DNA Synthesizer (Applied Biosystems, Foster City, CA). All reagents for the synthesis were purchased from Applied Biosystems.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene),

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA and/or NEOSPLA- depending on the sufficiency of the foregoing and the orientation of the insertion of the oligonucleotides.

Determination of orientation of the splice junction linker was preformed by nucleic acid sequencing using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical, Cleveland, OH, Prod. No. 70770) following manufacturer instructions. Upon determination of linker orientation within six independent plasmid isolates, identification of NEOSPLA was made such that the inserted splice junction sequences are in the correct forward orientation with respect to the direction of NEO transcription.

NEOSPLA was digested with Xho I by forming an admixture of 15 µl dH_2O; 1 µl NEOSPLA DNA; 2 µl 10 X digestion buffer D (Promega, supplied with enzyme); and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide having the following sequence: 5' TCGATTAATTAA 3' (SEQ ID NO: 14). Insertion of this sequence effectively changes the Xho I site to a Pac I restriction site (as underlined in SEQ ID NO: 14).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized NEOSPLA DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

This Sheet Has Been Replaced



This is the replacement sheet

11 FEB 1997

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA3 depending on the sufficiency of the foregoing. (NEOSPLA3 was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K) contains the anti-CD20 light chain and heavy chain immunoglobulin cassettes and a DHFR cassette bounded by a Not I site at the 5' end and an Xho I site at the 3' end. Anti-CD20 in ANEX 2(G1,K) was digested with Xho I by forming an admixture of 15µldH_2O, 1 µl anti-CD20 in ANEX 2(G1,K) DNA, 2 µl 10 X digestion buffer D (Promega, supplied with enzyme) and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3 hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide of the following sequence: 5' TCGAAGCGGCCGCT 3' (SEQ ID NO: 15). Insertion of this sequence effectively changes the Xho I site to a Not I restriction site (as underlined in SEQ ID NO: 15).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized anti-CD20 in ANEX 2 DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

This Sheet Has Been Replaced



M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as Anti-CD20 in ANEX 2(G1,K)A depending on the sufficiency of the foregoing. (This was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K)A was digested with Not I and Xho I by forming an admixture of 6 µl dH_2O; 10 µl Anti-CD20 in ANEX 2(G1,K); 2 µl 10 x Not I digestion buffer (NEB, supplied with Not I enzyme); and 1 µl Not I enzyme (NEB). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 5515 base pairs by was isolated via the GlassMAX method for insertion into NEOSPLA3.

NEOSPLA3 was previously prepared for acceptance of the anti-CD20 cassette by digestion of 1 µl of DNA with Not I using an admixture comprising 16 µl dH_2O;

2 µl 10X Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB); followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume.

Ligation of the anti-CD20 DNA fragment into prepared NEOSPLA3 vector was accomplished as follows: 10 µl of anti-CD20 fragment DNA was admixed with 6 µl dH_2O; 1 µl cut NEOSPLA3 vector DNA; 2 µl 10 x ligation buffer (Promega supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega); followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmids referred to as anti-CD20 in NEOSPLA3F and anti-CD20 in NEOSPLA3R depending on the sufficiency of the foregoing and relative orientation of the inserted fragment with respect to NEO transcription.

Determination of orientation of the anti-CD20 cassette insertion was preformed by double digestion with KpnI and SpeI (NEB, Prod. No. 1335) in NEB buffer 1 plus acetated BSA as follows: an admixture comprising 4 µl DNA; 2 µl NEB buffer 1; 1 µl Kpn I; 1 µl Spe I; 2 µl BSA; and 10 µl dH_20 was formed. The admixture was digested at 37°C for 2 hrs, followed by size fractionation on an 0.8% agarose gel electrophoresis. Upon determination of anti-CD20 insert orientation within six independent plasmid isolates, identification of anti-CD20 in NEOSPLA3F was made such that the inserted sequences are in the forward orientation with respect to the direction of NEO transcription.

The 5515 bp anti-CD20 fragment contains the SV40 origin, a chimeric mouse human immunoglobulin light chain transcriptional cassette, a chimeric mouse human immunoglulbulin heavy chain transcriptional cassette, and a murine dihydrofolate reductase transcriptional cassette (*see*, Figure 4).

Anti-CD20 in NEOSPLA3F was doubly digested with Kpn I and Stu I by creating the admixture consisting of 14 µl dH_2O, 1 µl anti-CD20 in NEOSPLA3F, 2 µl 10 x digestion buffer 1 (NEB, supplied with enzyme), 2 µl 10 x acetylated BSA (NEB supplied with Kpn I enzyme), 1 µl Kpn I enzyme, 1 µl Stu I enzyme (NEB, Prod. Nos. 142S and 187S respectively). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 9368 base pairs by was isolated via the GlassMAX method.

A PCR fragment of DNA was generated from TCAE 5.2. The two following synthetic oligonucleotide primers were utilized in the PCR reaction:

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1994

5' primer: 5' GCA TGC GGT ACC GGA TCC ATC GAG CTA CTA GCT TTG C 3'

(SEQ ID NO: 16);

3' primer: 5' CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC TGC AGT TCA TCC AGG GC 3' (SEQ ID NO: 17)

The underlined portion of SEQ ID NO: 16 represents a Kpn I site, and the underlined portion of SEQ ID NO: 17 represents a Stu I site.

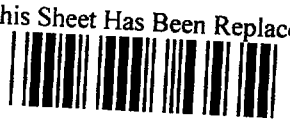
The PCR product was digested with Kpn I and Stu I and then ligated into prepared anti-CD20 in NEOSPLA3F.

Ligation of the 627 bp fragment into prepared anti-CD20 in NEOSPLA3F was accomplished as follows:

2 µl anti-CD20 in NEOSPLA3F; 1 µl SDS; 1 µl tRNA (Sigma); 11 µl 3M sodium acetate (pH 4.5) were admixed. Following phenol/chloroform isoamyl extraction of the admixture, the DNA was precipitated from the aqueous phase by addition of 270 µl ethanol (ice-cold) and this was spun at 13,000 rpm for 10 min. Following a 70% ETOH wash, the DNA was resuspended in 16 µlTE, 10 µl of PCR fragment DNA was admixed with 6 µl dH_2O, 1 µl cut anti-CD20 in NEOSPLA3F vector DNA, 2 µl 10 x ligation buffer (Promega, supplied with enzyme) and 1 µl T-4 DNA Ligase (Promega) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No. M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as anti-CD20 in GKNEOSPLA3F depending on the sufficiency of the foregoing. (Confirmation was based upon sequence determination of the different regions of GKNEOSPLA3F vs. NEOSPLA3F.)

This Sheet Has Been Replaced



The new plasmid differs from anti-CD20 in NEOSPLA3F in its Kozak sequence for the NEO gene which is:

```
                                        -3          +1
TGT GTT GGG AGC TTG GAT CGAT cc Acc ATG Gtt
                          Cla I    Start NEO
```

for Anti-CD20 in GKNEOSPLA3F, and

```
                                      -3     +1
TGT G CCA GCA TGG AGG AAT CGA Tcc Tcc ATG Ctt
          upstream Start            Start NEO
```

for Anti-CD20 in NEOSPLA3F.

Comparative analysis of expression of anti-CD20 in TCAE 5 vector (comprising NEO with consensus Kozak); ANEX 2 vector (comprising NEO with fully impaired Kozak, and upstream out-of-frame start sequence); NEOPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has fully impaired Kozak and an upstream out-of-frame start sequence); and GKNEOSPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has consensus Kozak).

Twenty-five (25) µg of each plasmid (digested as follows: anti-CD20 in TCAE5 and ANEX2 - Not I; anti-CD20 in NEOSPLA3F - Pac I; anti-CD20 in GKNEOSPLA3F - Pac I and Kpn I) was electroporated into 4 x 10^6 CHO cells; these digestions were utilized to separate the genes expressed in mammalian cells from the DNA used to grow the plasmid in bacteria. Following digestion, EtOH precipitation of the DNA, and drying thereof, the DNA was resuspended in sterile TE at a concentration of 1 µg/µl. Electroporation conditions were as described in Example II, except that 230 volts was utilized and, following electroporation, the mixture of cells and DNA was maintained for 10 min. at room temperature in the sterile, disposable electroporation cuvette.

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

Following electroporation, cells were plated into 96 well dishes as shown below in Table I, based upon the expected frequency of G418 resistant colonies (as derived from preliminary experiments; data not shown):

Table I
Comparative Expression

Plasmid	No. Transfections	No. Cells Plated	No. 96 Well Plates
TCAE 5	1	4×10^5	5
ANEX 2	1	2×10^6	5
GKNEOSPLA3F	1	2×10^6	5
NEOSPLA3F	5	2×10^7	5

Table 1 (continued)

Plasmid	No. G418 Resistant Colonies	Frequency of G418 Resistant Colony per Transfected Cell
TCAE 5	16	1 in 20,000
ANEX 2	16	1 in 100,000
GKNEOSPLA3F	16	1 in 100,000
NEOSPLA3F	16	1 in 1,000,000

(Cells were fed with G418 containing media on days 2, 5, 7, 9, 12, 14, 18, 22, 26, 30 and 34; supernatant from colonies was assayed for immunoglobulin production and the colonies became confluent in the wells on days 18, 22, 26, 30 and 34).

Figures 7A to 7C provide histogram results and evidence the percentage of colonies at a particular level of expression.

The Examples provided herein are not to be construed as limited to the specific vectors, fully impaired consensus Kozak sequences, dominant selectable markers, transcriptional cassettes, and/or expressed proteins. The fully impaired consensus Kozak, and the utilization

This Sheet Has Been Replaced

This is a replacement sheet

Date: 11 FEB 1997

thereof, are not to be construed as limited to ANEX 1 and ANEX 2 vectors. Similarly, the preferred fully impaired consensus Kozak sequences and vectors in no way constitute an admission, either actual or implied, that these are the only sequences or vectors to which the inventor is entitled. The inventor is entitled to the full breadth of protection under applicable patent laws. Preferred vectors incorporating fully impaired consensus Kozak sequences have been identified by the inventor on ANEX 1 and ANEX 2 for purposes of claiming these vectors by designating plasmids comprising these vectors and anti-CD20 were deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland, 20852, under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The plasmids were tested by the ATCC on November 9, 1992, and determined to be viable on that date. The ATCC has assigned these plasmids the following ATCC deposit numbers 69120 (anti-CD20 in TCAE 12(ANEX 1)) and 69118 (anti-CD20 in ANEX 2 (G1,K)); for purposes of this deposit, these plasmids were transformed into *E. coli*.

Although the invention has been described in considerable detail with regard to certain preferred embodiments thereof, other embodiments within the scope of the teachings of the present invention are possible. Accordingly, neither the disclosure nor the claims to follow, are intended, nor should be construed to be, limited by the descriptions of the preferred embodiments contained here.

This Sheet Has Been Replaced



SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Reff, Mitchell E.

(ii) TITLE OF INVENTION: Impaired Dominant Selectable Marker Sequence and Intronic Insertion Strategies for Enhancement of Expression of Gene Product and Expression Vector Systems Comprising Same

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDING ADDRESS:

(A) ADDRESSEE: IDEC Pharmaceuticals Corporation

(B) STREET: 11011 Torreyana Road

(C) CITY: San Diego

(D) STATE: California

(E) COUNTRY: USA

(F) ZIP: 92121

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb

(B) COMPUTER: Macintosh

(C) OPERATING SYSTEM: MS.DOS

(D) SOFTWARE: Microsoft Word 5.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Burgoon, Richard P. Jr.

(B) REGISTRATION NUMBER: 34,787

(C) REFERENCE/DOCKET NUMBER:

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (619) 550-8500

(B) TELEFAX: (619) 550 8750

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
TAG CTA GGT CCT ACC CC 17

(3) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
ATC GAT CCT GGA TGC GG

(4) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
TNN ATG CTT 9

(5) INFORMATION FOR SEQ ID NO: 4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single, including nick
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: (DNA (genomic)

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 FEB 1997

(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
CNN ATG CTT 9

(6) INFORMATION FOR SEQ ID NO: 5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
TNN ATG TTT 9

(7) INFORMATION FOR SEQ ID NO: 6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
CNN ATG TTT 9

(8) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
TTG GGA GCT TGG ATC GAT 18
CCA CCA TGG TT 11

This Sheet Has Been Replaced



(9) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
TTG GGA GCT TGG ATC GAT 18
CCT CCA TGC TT 11

(10) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
CCA GCA TGG AGG AAT CGA TCC 21
TCC ATG CTT 9

(11) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
GGA GGA TCG ATT CCT CCA TGC TGG 24
CAC AAC TAT GTC AGA AGC AAA TGT GAG C 28

(12) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:

This Sheet Has Been Replaced

This is the replacement sheet

Date:

(A) LENGTH: 18 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
CTG GGG CTC GAG CTT TGC 18

(13) INFORMATION FOR SEQ ED NO: 12
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no62
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
GGT AAG TGC GGC CGC TAC TAA CTC TCT CCT 30
CCC TCC TTT TTC CTG GA 17

(14) INFORMATION FOR SEQ ID NO: 13
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:
GGA AAA AGG AGG GAG GAG AGA GTT AGT AGC GGC 33
CGC ACT TAC CTG CA 14

(15) INFORMATION FOR SEQ ID NO: 14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single

This Sheet Has Been Replaced

This is the replacement sheet

Date: 11 Feb 1997

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCG ATT AAT TAA 12

(16) INFORMATION FOR SEQ ID NO: 15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCG AAG CGG CCG CT 14

(17) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16

GCA TGC GGT ACC GGA TCC ATC GAG CTA 27

CTA GCT TTG C 10

(18) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

This Sheet Has Been Replaced



11 Feb 1997

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17

CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC 30

TGC AGT TCA TCC AGG GC 17

This Sheet Has Been Replaced



This is the replacement sheet

Date:

This Sheet Has Been Replaced

Claims

1. An expression vector for expressing a protein of interest by recombinant deoxyribonucleic acid techniques, said vector comprising at least one dominant selectable marker, wherein the translation initiation start site of said marker comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where "Py" is a pyrimidine nucleotide; "x" is a nucleotide; and the numerical designations are relative to the codon "ATG".

2. The expression vector of claim 1 wherein the nucleic acid sequence encoding for the protein of interest is co-linked to said dominant selectable marker.

3. The expression vector of claim 1 wherein said dominant selectable marker is selected from the group consisting of: herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

4. The expression vector of claim 1 wherein said translation initiation start site sequence is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that the codon "Txx" downstream of the ATG codon does not encode a stop codon.

5. The expression vector of claim 1 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

* C L M S *

This is the replacement sheet

11 FEB 1997

6. The expression vector of claim 1 wherein said translation initiation start site sequence is TCCATGCTT.

7. The expression vector of claim 1 wherein said translation initiation start site sequence is located within a secondary structure.

8. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

9. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

10. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

11. The expression vector of claims 8, 9 and 10 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

12. The expression vector of claim 10 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

13. The expression vector of claims 8, 9, 10 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

14. A dominant selectable marker encoded by a nucleic acid sequence, wherein the translation initiation start site of said dominant selectable marker is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that "Txx" downstream of the ATG codon does not encode a stop codon.

15. The material of claim 14 wherein said dominant selectable marker is selected from the group consisting of herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

16. The material of claim 14 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

17. The material of claim 14 wherein said translation initiation start site sequence is TCCATGCTT.

18. The material of claim 14 wherein said translation initiation start site sequence is located within a secondary structure.

19. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

20. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

21. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

22. The material of claims 19, 20 and 21 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

23. The material of claim 21 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

24. The material of claims 19, 20, 21 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

25. An expression vector selected from the group consisting of ANEX 1 (included within American Type Culture Collection deposit number 69120) and ANEX 2 (included within ATCC deposit number 69118).

26. A plasmid comprising the expression vector of claim 1 wherein the nucleic acid sequence encoding for said protein of interest is co-linked to said dominant selectable marker.

27. The plasmid of claim 26 integrated within the cellular deoxyribonucleic acid of a mammalian host cell.

28. The plasmid of claim 27 wherein said host cell is selected from the group consisting of DG44, DXB11, CV1, COS, R1610, SP2/0, P3x633-Ag8.653, BFA-1c1BPT, RAJI, and 293.

29. The expression vector of claim 1 further comprising an artificial intronic insertion region within said dominant selectable marker, wherein an encoding sequence for a protein of interest is located within said insertion region.

30. The dominant selectable marker of claim 14 further comprising an artificial instronic insertion region.

ABSTRACT

IMPAIRED DOMINANT SELECTABLE MARKER SEQUENCE AND INTRONIC INSERTION STRATEGIES FOR ENHANCEMENT OF EXPRESSION OF GENE PRODUCT AND EXPRESSION VECTOR SYSTEMS COMPRISING SAME

Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector. These vectors are most typically utilized in the expression of proteins in mammalian expression systems. As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises sequence (1), where; "x" is a nucleotide selected from the group consisting of adenine (A), quanine (G), cytosine (C) or thymine (T)/uracil (U); "Py" is a pyrimidine nucleotide, ie C or TU; "ATG" is a codon encoding for the amino acid methionine, the so-called "start codon; and - 3 and + 1 are directional reference points vis-a-vis ATG. ie - 3 is meant to indicate three nucleotides upstream of ATG and + 1 is meant to indicate one nucleotide downstream of ATG. Dominant selectable markers further comprising artificial intronic insertion regions are further disclosed.

(Figure 1)

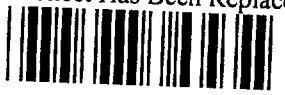
This Sheet Has Been Replaced



This is the replacement sheet

Date:

This Sheet Has Been Replaced



Neomycin phosphotransferase gene

		Cla I	-3		+1
TCAE 5.2	TTGGGAGCTTGG	ATCGAT	CC	A cc ATG	Gtt
				Met	Val
ANEX 1	TTGGGAGCTTGG	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu
ANEX 2	CCaGCATGgAGGA	ATCGAT	CC	I cc ATG	Ctt
				Met	Leu

FIG. 1

This is the replacement sheet

11 FEB 1997

This Sheet Has Been Replaced

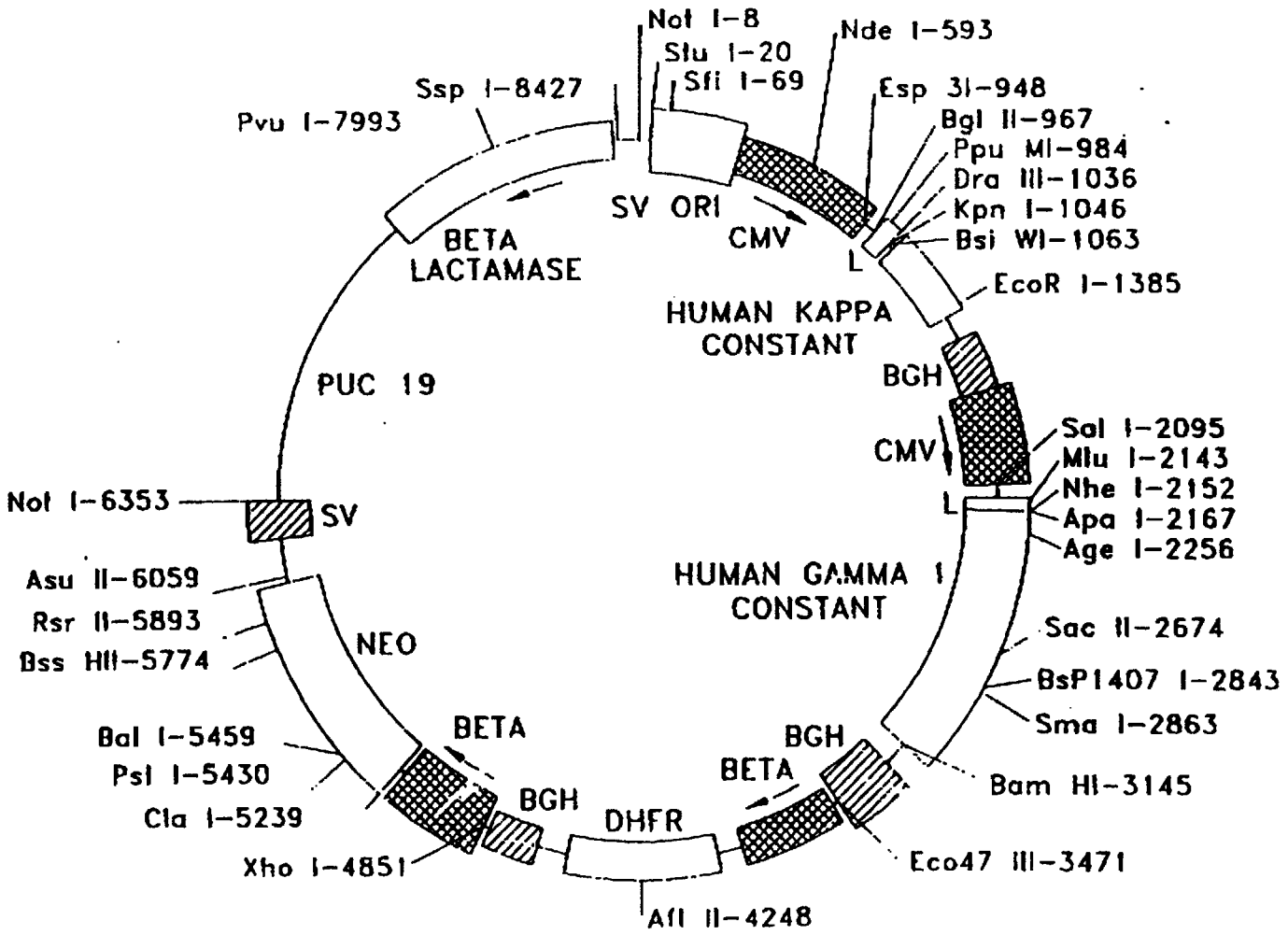


FIG. 2

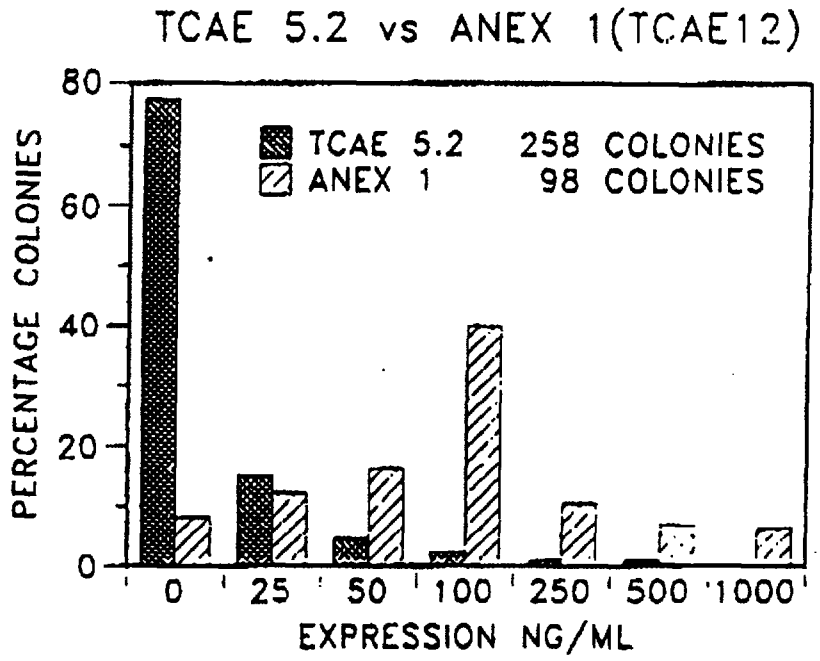


FIG. 3

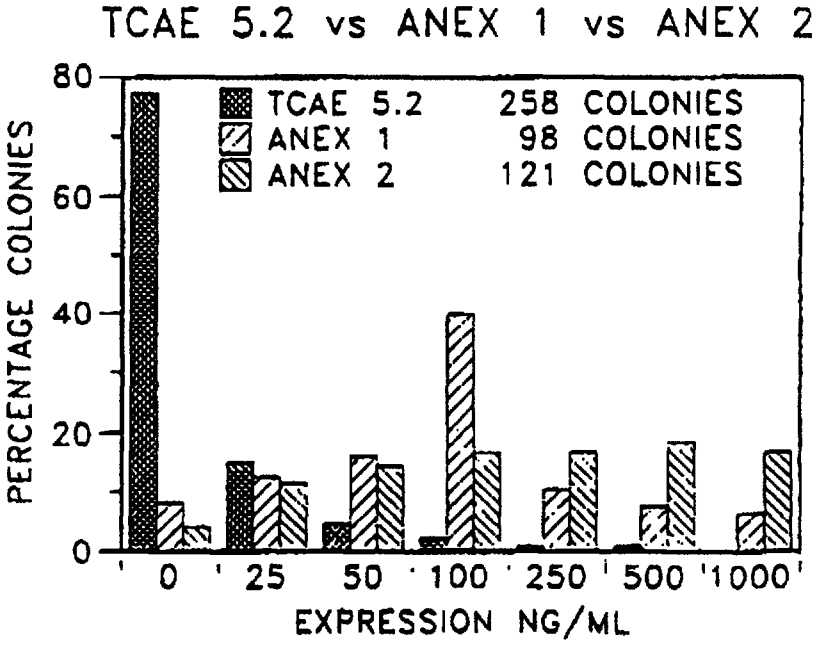


FIG. 5

This Sheet Has Been Replaced

This is the replacement sheet

Date:

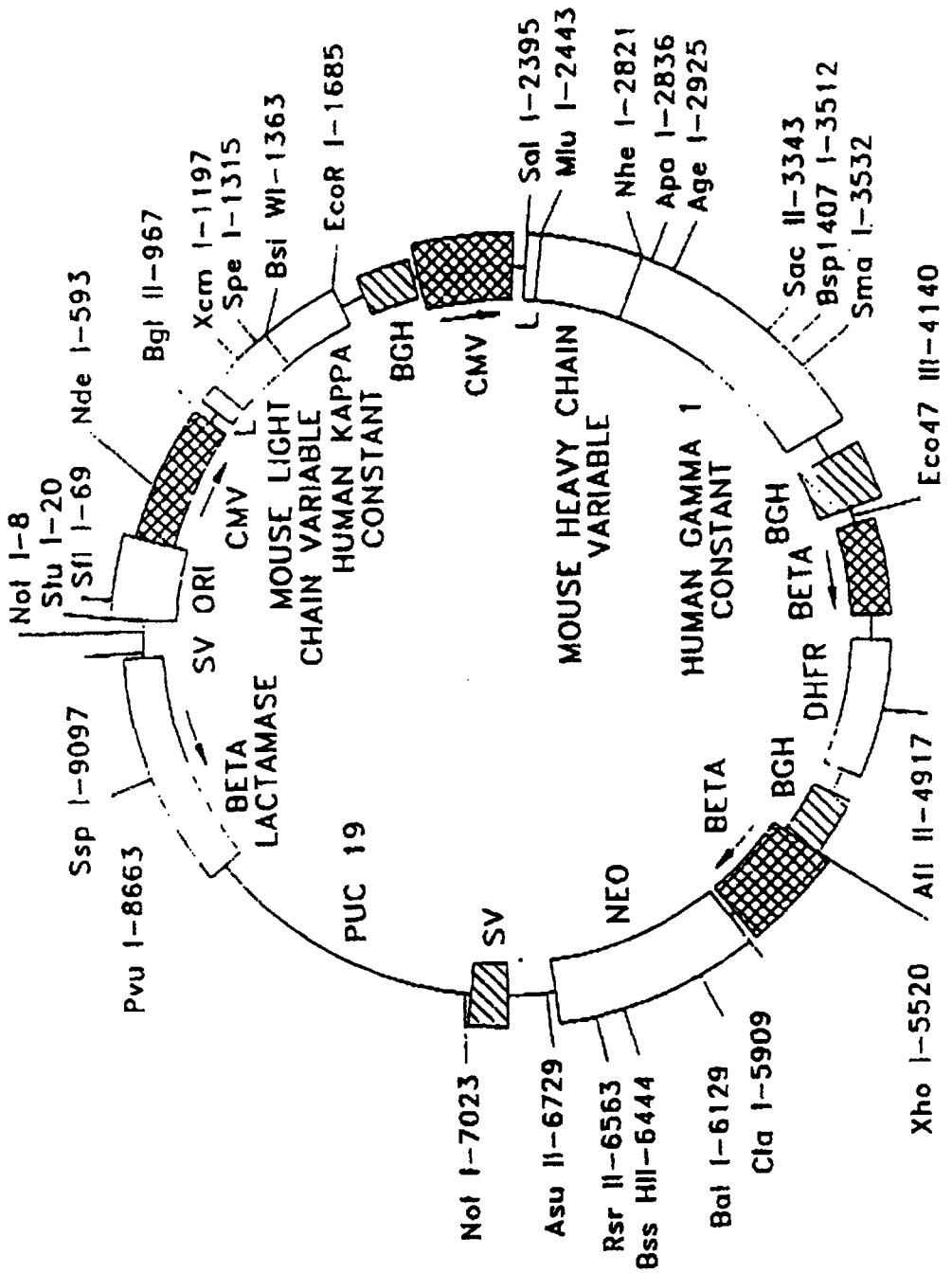


FIG. 4

This Sheet Has Been Replaced



This is the replacement sheet

Date: 11 FEB 1997

This Sheet Has Been Replaced

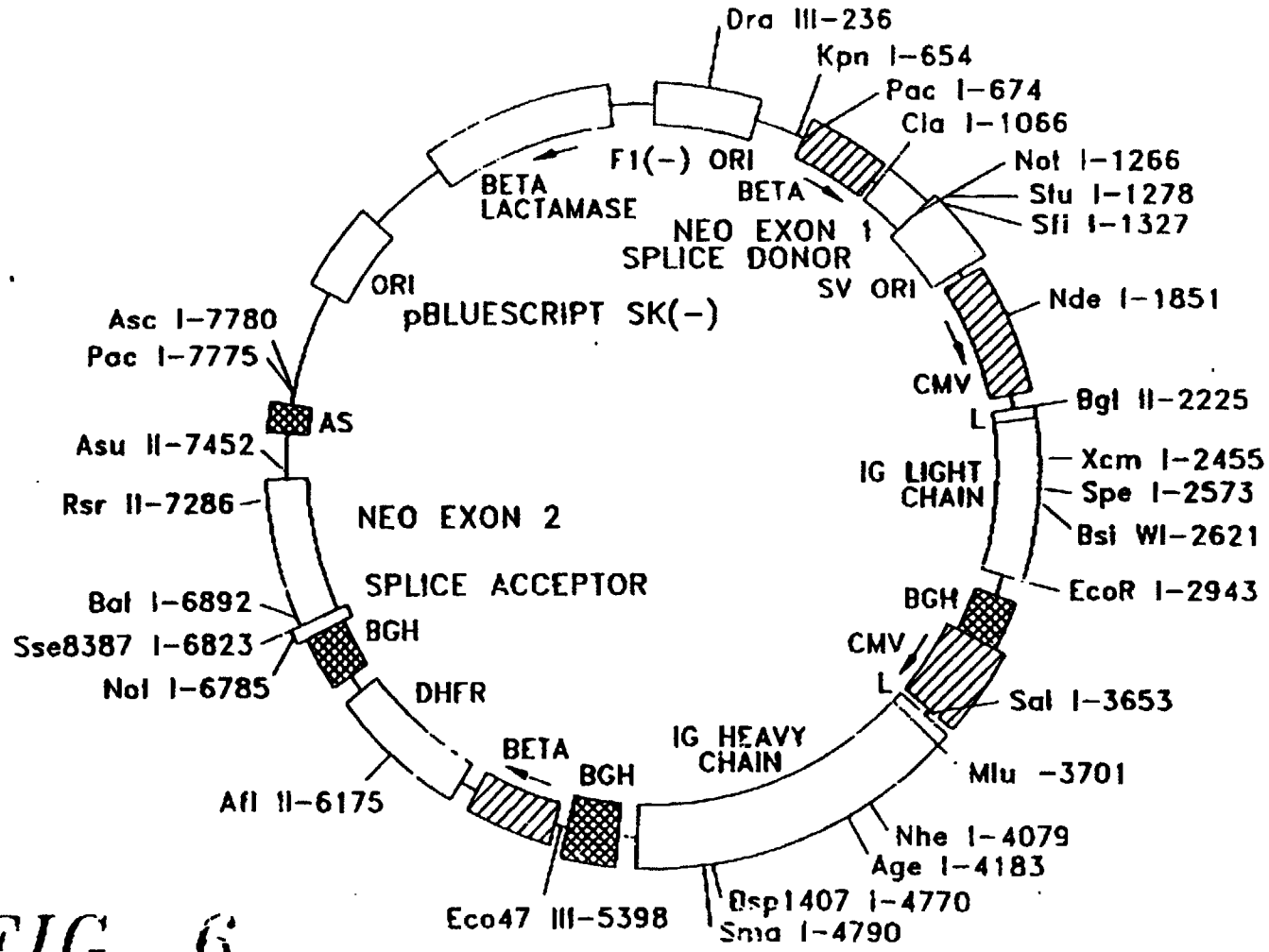


FIG. 6

FEB 1997

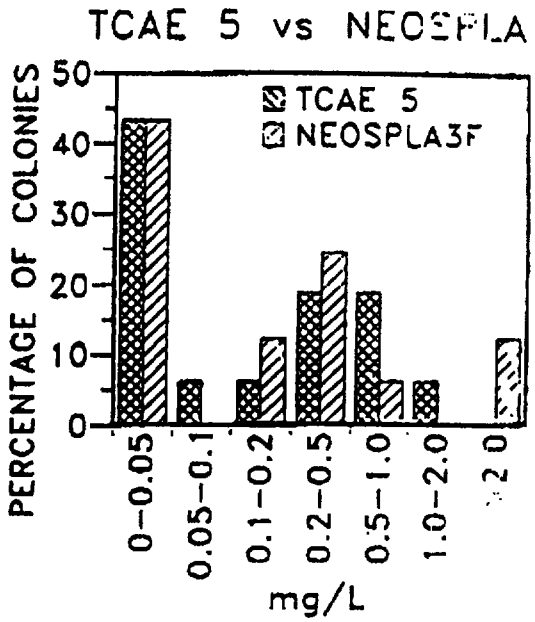


FIG. 7A

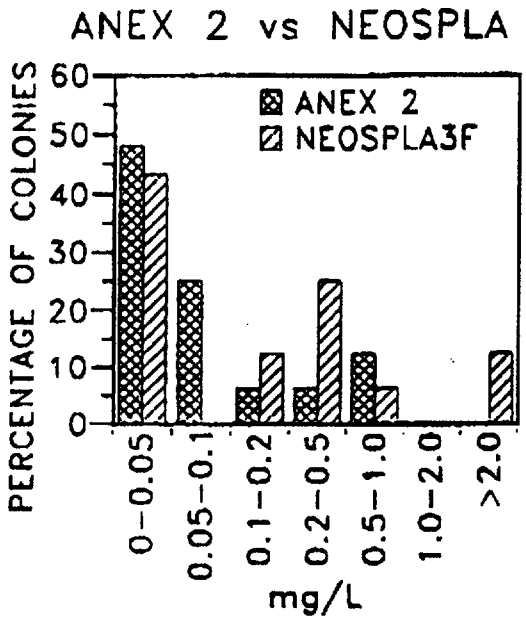


FIG. 7B

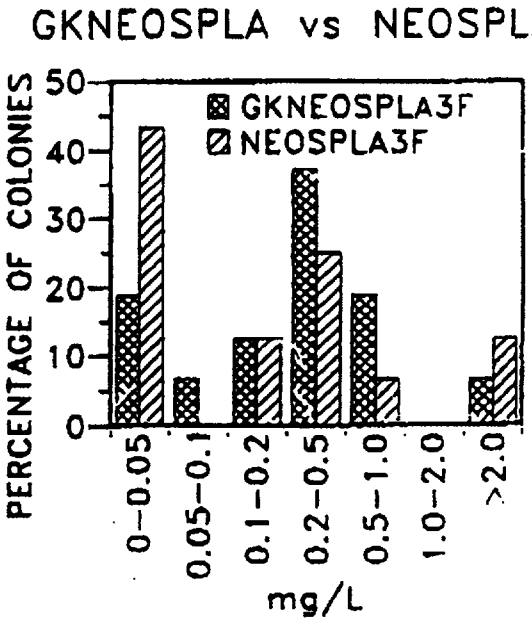


FIG. 7C

This Sheet Has Been Replaced



This is the replacement sheet

Date:

960/98C-5 IR-5

DESCRIPTION

Impaired Dominant Selectable Marker Sequence And Intronic Insertion Strategies For Enhancement Of Expression Of Gene Product And Expression Vector Systems Comprising Same

37 C.F.R. §1.74(d)/(e) Copyright Notice

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner does not object to the reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent files or records, but otherwise reserves all copyright rights whatsoever.

Related Applications

This patent document is a Continuation-In-Part of U.S. Serial No. 07/977,691 filed November 13, 1992, pending. This patent document is related to "THERAPEUTIC APPLICATION OF CHIMERIC ANTIBODY TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA," having U.S. Serial No. 07/978,891, filed November 13, 1992 (pending), and "THERAPEUTIC APPLICATION OF CHIMERIC AND RADIOLABLED ANTIBODIES TO HUMAN B LYMPHOCYTE RESTRICTED DIFFERENTIATION ANTIGEN FOR TREATMENT OF B CELL LYMPHOMA", having U.S. Serial No. _____, filed simultaneously herewith. This patent document is related to commonly assigned United States Serial No. 07/912,292 and entitled "RECOMBINANT ANTIBODIES FOR HUMAN THERAPY," filed July 10, 1992. These documents are incorporated herein by reference.

This is the replacement sheet

14 MAR 1997

INTENTIONALLY
LEFT BLANK
PER EXAMINER
M. HOGARTH

This is the replacement sheet

Date: 14 MAR 1997

As noted, the advent of the biotechnology industry has allowed for the production of large quantities of proteins. Proteins are the essential constituents of all living cells and proteins are comprised of combinations of 20 naturally occurring amino acids; each amino acid molecule is defined ("encoded") by groupings ("codons") of three deoxyribonucleic acid ("DNA") molecules; a string of DNA molecules ("DNA macromolecule") provides, in essence, a blueprint for the production of specific sequences of amino acids specified by that blueprint. Intimately involved in this process is ribonucleic acid ("RNA"); three types of RNA (messenger RNA; transfer RNA; and ribosomal RNA) convert the information encoded by the DNA into, *eg* a protein. Thus, genetic information is generally transferred as follows: DNA →RNA → protein.

In accordance with a typical strategy involving recombinant DNA technology, a DNA sequence which encodes a desired protein material ("cDNA") is identified and either isolated from a natural source or synthetically produced. By manipulating this piece of genetic material, the ends thereof are tailored to be ligated, or "fit," into a section of a small circular molecule of double stranded DNA. This circular molecule is typically referred to as a "DNA expression vector," or simply a "vector." The combination of the vector and the genetic material can be referred to as a "plasmid" and the plasmid can be replicated in a prokaryotic host (*ie* bacterial in nature) as an autonomous circular DNA molecule as the prokaryotic host replicates. Thereafter, the circular DNA plasmid can be isolated and introduced into a eukaryotic host (*ie* mammalian in nature) and host cells which have incorporated the plasmid DNA are selected. While some plasmid vectors will replicate as an autonomous circular DNA molecule in mammalian cells, (*eg* plasmids comprising Epstein Barr virus ("EBV") and Bovine Papilloma virus ("BPV") based vectors), most plasmids including DNA vectors, and all plasmids including RNA retroviral vectors, are integrated into the cellular DNA such that

when the cellular DNA of the eukaryotic host cell replicates, the plasmid DNA will also replicate. Accordingly, as eukaryotic cells grow and divide, there is a corresponding increase in cells containing the integrated plasmid which leads to the production ("expression") of the protein material of interest. By subjecting the host cells containing the plasmid to favorable growth conditions, significant amounts of the host, and hence the protein of interest, are produced. Typically, the Chinese Hamster Ovary ("CHO") cell line is utilized as a eukaryotic host cell, while *E. coli* is utilized as a prokaryotic host cell.

The vector plays a crucial role in the foregoing - manipulation of the vector can allow for variability as to where the cDNA is inserted, means for determining whether the cDNA was, in fact, properly inserted within the vector, the conditions under which expression of the genetic material will or will not occur, etc. However, most of the vector manipulations are geared toward a single goal—increasing expression of a desired gene product, *ie* protein of interest. Stated again, most vector manipulation is conducted so that an "improved" vector will allow for production of a gene product at significantly higher levels when compared to a "non-improved" vector. Thus, while certain of the features/aspects/characteristics of one vector may appear to be similar to the features/aspects/characteristics of another vector, it is often necessary to examine the result of the overall goal of the manipulation - improved production of a gene product of interest.

While one "improved" vector may comprise characteristics which are desirable for one set of circumstances, these characteristics may not necessarily be desirable under other circumstances. However, one characteristic is desirable for all vectors: increased efficiency, *ie* the ability to increase the amount of protein of interest produced while at the same time decreasing the number of host cells to be screened which

This is the replacement sheet

Date: 14 MAR 1997

do not generate a sufficient amount of this protein. Such increased efficiency would have several desirable advantages, including reducing manufacturing costs and decreasing the time spent by technicians in screening for viable colonies which are expressing the protein of interest. Accordingly, what would be desirable and what would significantly improve the state of the art are expression vectors with such efficiency characteristics.

Summary Of The Invention

The invention disclosed herein satisfies these and other needs. Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector; preferably, the dominant selectable marker comprises either a natural intronic insertion region or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within such insertion region.

As used herein, a "dominant selectable marker" is a gene sequence or protein encoded by that gene sequence; expression of the protein encoded by the dominant selectable marker assures that a host cell transfected with an expression vector which includes the dominant selectable marker will survive a selection process which would otherwise kill a host cell not containing this protein. As used herein, a "transcriptional cassette" is DNA encoding for a protein product (*eg* a dominant selectable marker) and the genetic elements necessary for production of the protein product in a host cell (*ie* promoter; transcription start site; polyadenylation region; etc.). These vectors are most preferably utilized in the expression of proteins in mammalian expression systems where integration of the vector into host cellular DNA occurs. Beneficially, the use of such fully impaired consensus Kozak sequences improves the efficiency of protein expression by significantly decreasing the number

This is the enhancement sheet

Date: 14 MAR 1997

of viable colonies while at the same time, significantly increasing the amount of protein expressed by such viable colonies. As used herein, a "natural intronic insertion region" is a region of DNA naturally present within a gene in this case, typically a dominant selectable marker, which can be utilized for insertion of DNA encoding a gene product of interest; an "artificial intronic insertion region" is a region of DNA which is selectively created in a gene (again, most typically, a dominant selectable marker) which can be utilized for insertion of DNA encoding a gene product of interest. Information regarding intronic positioning is described in Abrams, J.M. et al. "Intronic Positioning Maximizes Co-expression and Co-amplification of Nonselectable Heterologous Genes. "*J. Bio. Chem.*, *264124*:14016 (1989), and U.S. Patent No. 5,043,270 (both documents being incorporated herein by reference).

As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T)/ uracil (U); "Py" is a pyrimidine nucleotide, *ie* C or T/U; "ATG" is a codon encoding for the amino acid methionine, the so-called "start" codon; and -3 and +1 are directional reference points vis-a-vis ATG, *ie* -3 is meant to indicate three nucleotides upstream of ATG and +1 is meant to indicate one nucleotide downstream of ATG.

Preferably, the fully impaired consensus Kozak is part of a methionine start codon that initiates translation of a dominant selectable marker portion of a transcriptional cassette which is part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl

transferase; hygromycin B phosphotransferase; and neomycin phosphotransferase. Most preferably, the dominant selectable marker is neomycin phosphotransferase.

In particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without an in-frame stop codon being located between the upstream start codon and the fully impaired consensus Kozak start codon. As used herein, the term "stop codon" is meant to indicate a codon which does not encode an amino acid such that translation of the encoded material is terminated; this definition includes, in particular, the traditional stop codons TAA, TAG and TGA. As used herein, the terms "in-frame" and "out-of-frame" are relative to the fully impaired consensus Kozak start codon. By way of example, in the following sequence:

-3 +1

GAC CAT GGC Cxx ATG Cxx

the underlined portion of the sequence is representative of a fully impaired consensus Kozak (where "x" represents a nucleotide) and the codons GAC, CAT and GCC are "in-frame" codons relative to the ATG start codon. The above-lined nucleotides represent an "out-of-frame" start codon which is upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides upstream of the fully impaired consensus Kozak start codon. Preferably, the out-of-frame start codon is a part of a consensus Kozak. By way of example, the sequence set forth above satisfies this criteria: the -5 nucleotide is a purine (G); nucleotide -6, -7 and -8 encode an out-of-frame start codon (ATG); and nucleotide -11 is a purine (A).

Additionally, utilization of a fully impaired consensus Kozak within a secondary structure (*ie* a so-

This is the replacement sheet

Date: 14 MAR 1997

called "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable number. In such an embodiment, the start codon of the fully impaired consensus Kozak is most preferably located within the stem of a stem loop.

Particularly preferred expression vectors which incorporate these aspects of the invention disclosed herein are referred to as "TCAE," and "ANEX" and "NEOSPLA" vectors; particularly preferred vectors are referred to as ANEX 1, ANEX 2 and NEOSPLA3F.

These and other aspects of the invention disclosed herein will be delineated in further detail in the sections to follow.

Brief Description Of The Drawings

Figure 1 provides the relevant portion of a consensus Kozak and several particularly preferred fully impaired consensus Kozak sequences;

Figure 2 provides a diagrammatic representation of the vectors TCAE 5.2 and ANEX 1 (TCAE 12) designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 3 is a histogram comparing protein expression levels with the vectors TCAE 5.2 and ANEX 1;

Figure 4 provides a diagrammatic representation of the vector ANEX 2 designed for expression of mouse/human chimeric immunoglobulin, where the immunoglobulin genes are arranges in a tandem configuration using neomycin phosphotransferase as the dominant selectable marker;

Figure 5 is a histogram comparing protein expression levels with the vectors TCAE 5.2, ANEX 1 and ANEX 2;

Figure 6 provides a diagrammatic representation of a NEOSPLA vector designed for expression of mouse/human chimeric immunoglobulin; and

This is the replacement sheet

Date: 14 MAR 1997

Figures 7A, 7B and 7C are histograms comparing protein expression levels with the vectors TCAE 5.2 vs. NEOSPLA 3F (7A). ANEX 2 vs. NEOSPLA3F (7B); and GK-NEOSPLA3F vs. NEOSPLA3F (7C).

Detailed Description Of Preferred Embodiments

Disclosed herein are nucleic acid sequence arrangements which impair translation and initiation of, most preferably, dominant selectable markers incorporated into mammalian expression vectors and which are preferably, but not necessarily, co-linked to an encoding sequence for a gene product of interest. Preferably, the dominant selectable marker comprises at least one natural or artificial intronic insertion region, and at least one gene product of interest is encoded by DNA located within at least one such intron. Such arrangements have the effect of increasing expression efficiency of the gene product of interest by, *inter alia*, decreasing the number of viable colonies obtained from an equivalent amount of plasmid DNA transfected per cell, while increasing the amount of gene product expressed in each clone.

For purpose of brevity and presentational efficiency, the focus of this section of the patent disclosure will be principally directed to a specific dominant selectable marker, neomycin phosphotransferase, which is incorporated into a mouse/human chimeric immunoglobulin expression vector. It is to be understood, however, that the invention disclosed herein is not intended, nor is it to be construed, as limited to these particular systems. To the contrary, the disclosed invention is applicable to mammalian expression systems *in toto*, where vector DNA is integrated into host cellular DNA.

One of the most preferred methods utilized by those in the art for producing a mammalian cell line that produces a high level of a protein (*ie* "production cell line") involves random integration of DNA coding for the desired gene product (*ie* "exogenous DNA") by using, most typically, a drug resistant gene, referred to as a

"dominant selectable marker," that allows for selection of cells that have integrated the exogenous DNA. Stated again, those cells which properly incorporate the exogenous DNA including, *eg*, the drug resistant gene, will maintain resistance to the corresponding drug. This is most typically followed by co-amplification of the DNA encoding for the desired gene product in the transfected cell by amplifying an adjacent gene that also encodes for drug resistance ("amplification gene"), *eg* resistance to methotrexate (MTX) in the case of dehydrofolate reductase (DHFR) gene. The amplification gene can be the same as the dominant selectable marker gene, or it can be a separate gene. (As those in the art appreciate, "transfection" is typically utilized to describe the process or state of introduction of a plasmid into a mammalian host cell, while "transformation" is typically utilized to describe the process or state of introduction of a plasmid into a bacterial host cell).

Two amplification approaches are typically employed by those in the art. In the first, the entire population of transfected and drug resistant cells (each cell comprising at least one integration of the gene encoding for drug resistance) is amplified; in the second, individual clones derived from a single cell are amplified. Each approach has unique advantages.

With respect to the first approach, it is somewhat "easier" to amplify the entire population (typically referred to as a "shotgun" approach, an apt description) compared to individual clones. This is because amplification of individual clones initially involves, *inter alia*, screening of hundreds of isolated mammalian colonies (each derived from a single cell, most of which being single copy integrants of the expression plasmid) in an effort to isolate the one or two "grail" colonies which secrete the desired gene product at a "high" level, *ie* at a level which is (typically) three orders of magnitude higher than the lowest detectable expression level. These cells are also often found to have only a single copy

This is the replacement sheet

Date: 14 MAR 1997

integration of the expression plasmid. Additionally, amplifying individual clones results in production cell lines which contain fewer copies of the amplified gene as compared to amplification of all transfected cells (typically, 10-20 versus 500-1000).

With respect to the second approach, production cell lines derived from amplifying individual clones are typically derived in lower levels of the drug(s) used to select for those colonies which comprise the gene for drug(s) resistance and the exogenous gene product (*ie* in the case of methotrexate and DHFR, 5nM versus 1μM). Furthermore, individual clones can typically be isolated in a shorter period of time (3-6 months versus 6-9 months).

Ideally then, the tangible benefits of both approaches should be merged: at a practical level, this would involve decreasing the number of colonies to be screened, and increasing the amount product secreted by these colonies. The present invention accomplishes this task.

The position where the DNA of the dominant selectable marker of the plasmid DNA is integrated within the cellular DNA of the host cell determines the level of expression of the dominant selectable marker protein, as is recognized by those in the art. It is assumed that the expression of a gene encoding a protein of interest which is either co-linked to or positioned near the dominant selectable marker DNA is proportional to the expression of the dominant selectable marker protein. While not wishing to be bound by any particular theory, the inventor has postulated that if the gene used to select for the integration of the exogenous DNA in the mammalian cell (*ie* the dominant selectable marker) was designed such that translation of that dominant selectable marker was impaired, then only those plasmids which could overcome such impairment by over-production of the gene product of the dominant selectable marker would survive, *eg*, the drug-screening process. By associating the exogenous DNA

This is the replacement sheet
Date: 14 MAR 1997

with the dominant selectable marker, then, *a fiorti*, over-production of the gene product of the dominant selectable marker would also result in over-production of the gene product derived from the associated exogenous DNA. In accordance with this postulated approach, impairment of translation of the dominant selectable marker gene would be necessary, and an avenue for such impairment was the consensus Kozak portion of the gene.

By comparing several hundred vertebrate mRNAs, Marilyn Kozak in "Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes," *Nuc. Acids Res. 9*: 5233-5252 (1981) and "Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs," *Nuc. Acids Res. 12*: 857-872 (1984), proposed the following "consensus" sequence for initiation of translation in higher eukaryotes:

```
      -3      +1
cc Acc AUG G
```

(As those in the art appreciate, uracil, U, replaces the deoxynucleotide thymine, T, in RNA.) In this sequence, referred to as a "consensus Kozak," the most highly conserved nucleotides are the purines, A and G, shown in capital letters above; mutational analysis confirmed that these two positions have the strongest influence on initiation. *See, eg,* Kozak, M. "Effects of intercistronic length on the efficiency of reinitiation of eukaryotic ribosomes." *Mol. Cell Bio. 7/10*: 3438-3445 (1987). Kozak further determined that alterations in the sequence upstream of the consensus Kozak can effect translation. For example, in "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes." *PNAS 83*: 2850-2854 (1986) Kozak describes the "artificial" introduction of a secondary hairpin structure region upstream from the consensus Kozak in several plasmids that encoded preproinsulin; it was experimentally determined that a stable stem loop structure inhibited translation of the

preproinsulin gene, reducing the yield of proinsulin by 85-95%.

Surprisingly, it was discovered by the inventor that by changing the purines A(-3 vis-a-vis ATG start codon) and G(+1) to pyrimidines, translation impairment was significant: when the consensus Kozak for the neomycin phosphotransferase gene was subjected to such alterations (as will be set forth in detail below), the number of G418 resistant colonies significantly decreased; however, there was a significant increase in the amount of gene product expressed by the individual G418 resistant clones. As those in the art will recognize, this has the effect of increasing the efficiency of the expression system--there are less colonies to screen, and most of the colonies that are viable produce significantly more product than would ordinarily be obtained. Confirmation of the inventor's postulated theory was thus experimentally determined.

As noted, for purposes of this patent document a, "consensus Kozak" comprises the following sequence:

-3 +1

Pxx ATG Pxx

a "partially impaired consensus Kozak" comprises the following sequence

-3 +1

P/Pyxx ATG P/Pyxx

and a disclosed and claimed "fully impaired consensus Kozak" comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where: "x" is a nucleotide selected from the group consisting of adenine (A), guanine (G), cytosine (C) or thymine (T) (uracil, U, in the case of RNA); "P" is a purine, *ie* A or G, "Py" is a pyridine, *ie* C or T/U; ATG is a conventional start codon which encodes for the amino acid methionine (Met); the numerical designations are relative to the ATG codon, *ie* a negative number indicates "upstream" of ATG and a positive number indicates

This is the replacement sheet

Date: 14 MAR 1997

"downstream" of ATG; and for the partially impaired consensus Kozak, the following proviso is applicable--only one of the -3 or +1 nucleotides is a pyridine, *eg*, if -3 is a pyridine, then +1 must be a purine or if -3 is a purine, then +1 must be a pyridine. Most preferably, the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker which is preferably (but not necessarily) co-linked to exogenous DNA which encodes for a gene product of interest. As used herein, "nucleotide" is meant to encompass natural and synthetic deoxy- and ribonucleotides as well as modified deoxy- and ribonucleotides, *ie* where the 3' OH, 5'OH, sugar and/or heterocyclic base are modified, as well as modification of the phosphate backbone, *eg* methyl phosphates, phosphorothioates and phosphoramidites.

Information regarding the gene sequence of the dominant selectable marker is preferably known; however, in lieu of the entire sequence, information regarding the nucleic acid sequence (or amino acid sequence) at the site of translation initiation of the dominant selectable marker must be known. Stated again, in order to effectuate a change in the consensus or partially impaired consensus Kozak, one must know the sequence thereof. Changing the consensus or partially impaired consensus Kozak to a fully impaired consensus Kozak sequence can be accomplished by a variety of approaches well known to those in the art including, but not limited to, site specific mutagenesis and mutation by primer-based amplification (*eg* PCR); most preferably, such change is accomplished via mutation by primer-based amplification. This preference is principally based upon the comparative "ease" in accomplishing the task, coupled with the efficacy associated therewith. For ease of presentation, a description of the most preferred means for accomplishing the change to a fully impaired consensus Kozak will be provided.

This is the replacement sheet

Date: 14 MAR 1997

In essence, mutation by primer-based amplification relies upon the power of the amplification process itself--as PCR is routinely utilized, focus will be directed thereto. However, other primer-based amplification techniques (*eg* ligase chain reaction, etc.) are applicable. One of the two PCR primers ("mutational primer") incorporates a sequence which will ensure that the resulting amplified DNA product will incorporate the fully impaired consensus Kozak within the transcriptional cassette incorporating the dominant selectable marker of interest; the other PCR primer is complementary to another region of the dominant selectable marker; a transcriptional cassette incorporating the dominant selectable marker; or a vector which comprises the transcriptional cassette. By way of example, the complement to a dominant selectable marker which includes a consensus Kozak could have the following sequence (SEQ ID NO: 1):

3'-tagctaggTccTACCcc-5'

In order to create a fully impaired consensus Kozak, the mutational primer could have the following sequence (SEQ ID NO: 2) (for convenience, SEQ ID NO: 1 is placed over the mutational primer for comparative purposes):

5'-atcgatccTggATGCgg-3'

3'-tagctaggTccTACCcc-5'

* *

As is evident, complementarity is lacking in the primer (see the "*" symbols). By utilizing excess mutational primer in the PCR reaction, when the sequence including the consensus Kozak is amplified, the resulting amplified DNA products will incorporate the mutations such that as the amplified DNA products are in turn amplified, the mutations will predominate such that a fully impaired consensus Kozak will be incorporated into the amplification product.

Two criteria are required for the mutational primer--first, the length thereof must be sufficient such that hybridization to the target will result. As will be

appreciated, the mutational primer will not be 100% complementary to the target. Thus, a sufficient number of complementary bases are required in order to ensure the requisite hybridization. Preferably, the length of the mutational primer is between about 15 and about 60 nucleotides, more preferably between about 18 and about 40 nucleotides, although longer and shorter lengths are viable. (To the extent that the mutational primer is also utilized to incorporate an out-of-frame start codon or secondary structure, the length of the mutational primer can correspondingly increase). Second, the ratio of mutational primer to target must be sufficiently excessive to "force" the mutation. Preferably, the ratio of mutational primer to target is between about 250:1 to about 5000:1, more preferably between about 400:1 to about 2500:1, and most preferably between about 500:1 to about 1000:1.

Because the parameters of a PCR reaction are considered to be well within the level of skill of those in the art, details regarding the particulars of that reaction are not set forth herein; the skilled artisan is readily credited with recognizing the manner in which this type of mutation can be accomplished using PCR techniques - the foregoing is provided as a means of providing elucidation as opposed to detailed edification.

As noted, it is most preferred that the fully impaired consensus Kozak is associated with the site of translation initiation of a dominant selectable marker incorporated into a transcriptional cassette which forms a part of an expression vector. Preferred dominant selectable markers include, but are not limited to: herpes simplex virus thymidine kinase; adenosine deaminase; asparagine synthetase; *Salmonella* his D gene; xanthine guanine phosphoribosyl transferase ("XGPRT"); hygromycin B phosphotransferase; and neomycin phosphotransferase ("NEO"). Most preferably, the dominant selectable marker is NEO.

This is the replacement sheet

Date: 14 MAR 1997

The dominant selectable marker herpes simplex virus thymidine kinase is reported as having the following partially impaired consensus Kozak:

-3 +1

cg Cgt ATG Gct

See, Heller, S. "Insertional Activities of a Promotorless Thymidine Kinase Gene," *Mol. & Cell. Bio. 8/8*:3218-3302, Figure 4, nucleotide 764. By changing the +1 purine (G) to a pyrimidine (C or T/U), a fully impaired Kozak as defined herein is generated (the -3 of the herpes simplex virus thymidine kinase is a pyrimidine). Changing +1 purine to a pyrimidine also has the effect of changing the encoded amino acid from alanine (GCT) to proline (CCT) or serine (TCT); it is preferred that conservative amino acid changes result from the changes to the nucleotides. Thus, it is preferred that the change to TCT be made because the change from alanine to serine is a more conservative amino acid change than changing alanine to proline.

Histidinol dehydrogenase is another dominant selectable marker. *See*, Hartmen, S.C. and Mulligan, R.C. "Two dominant acting selectable markers for gene transfer studies in mammalian cells," *PNAS 85*:8047-8051 (1988). The *his D* gene of Salmonella typhimunium has the following partially impaired consensus Kozak:

-3 +1

gc Aga ATG Tta

As -3 is a purine, changing -3 to a pyrimidine (C or T/U) results in a fully impaired consensus Kozak; as is appreciated, because these nucleotides are upstream of the start codon, no impact on amino acid translation results from this change.

Hygromycin B phosphotransferase is another dominant selectable marker; the reported sequence for the *hph* gene (*see*, Gritz, L. and Davies, J. "Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expressing in *Escherichia*

This is the replacement sheet

Date: 14 MAR 1997

coli and *Succharomyces cerevisiae*" *Gene 25*:179-188, 1983) indicates that the consensus Kozak is:

-3 +1

ga Gat ATG Aaa

Both -3 and +1 are purines; thus changing -3 and +1 to pyrimidines results in a fully impaired consensus Kozak (this results in the following encoded amino acids: +1 to C - glutamine; +1 to T - stop codon. Because this codon is downstream of the start codon, the change to the stop codon TAA should not be accomplished).

XGPRT is another dominant selectable marker. The reported partially impaired consensus Kozak of XGPRT has the following sequence:

-3 +1

tt Cac ATG Agc

See, Mulligan, R.C. and Berg, P. "Factors governing the expression of a bacterial gene in mammalian cells." *Mol. & Cell Bio. 1/5*:449-459 (1981), Figure 6. By changing the +1 purine to a pyrimidine, a fully impaired consensus Kozak is created; the effect on the encoded amino acid (AGC-serine) is as follows: CGC-arginine; TGC-cysteine.

Adenosine deaminase (ADA) can also be utilized as a dominant selectable marker. The reported consensus Kozak sequence for adenosihe deaminase is:

-3 +1

ga Acc ATG Gcc

See, Yeung, C.Y. et *al.*, "Identification of functional murine adenosine deaminase cDNA clones by complementation in Echerichia coli," *J. Bio. Chem. 260/18*:10299-10307 (1985), Figure 3. By changing both -3 and +1 purines to pyrimidines, fully impaired consens Kozak sequences result. The encoded amino acid corresponding to GCC (alanine) is changed to either proline (CCC) or serine (TCC), with the change to serine being preferred, due to the conservative nature of this change.

The reported partially impaired consensus Kozak for asparagine synthetase is as follows:

-3 +1

gc Acc ATG Tgt

See, Andrulis, I.L. *et al.*, "Isolation of human cDNAs for asparagine synthetase and expression in Jensen rat sarcoma cells," *Mol. Cell. Bio.* 7/7:2435-2443 (1987). Changing the +3, purine to a pyrimidine results in a fully impaired consensus Kozak.

The partially impaired consensus Kozak for neomycin phosphotransferase (which includes an upstream out of frame start codon) is as follows:

-3 +1

ggA TGg gga tcg ttt Cgc ATG Att

Changing the +1 purine to a pyrimidine has the effect of creating a fully impaired consensus Kozak (changes to the encoded amino acid isoleucine, ATT are as follows: CTT - leucine and TTT-phenylalanine, with the change to leucine being preferred, due to the conservative nature of this change).

The foregoing is not intended, nor is it to be construed as limiting; rather, in the context of the disclosed invention, the foregoing is presented in an effort to provide equivalent examples of changes in the reported consensus Kozak sequences or partially impaired consensus Kozak sequences of several well-known dominant selectable markers.

As noted, a most preferred dominant selectable marker is NEO. Particularly preferred fully impaired consensus sequences for NEO are as follows:

Txx ATG Ctt (SEQ ID NO: 3)
Cxx ATG Ctt (SEQ ID NO: 4)
Txx ATG Ttt (SEQ ID NO: 5)
Cxx ATG Ttt (SEQ ID NO: 6)

where x are nucleotides. SEQ. ID. NO. 3 is most preferred; and xx are preferably CC.

Other transcriptional cassettes, which may or may not include a fully impaired consensus Kozak, can be

This is the replacement sheet

Date: 14 MAR 1997

incorporated into a vector which includes transcriptional cassettes containing the disclosed and claimed fully impaired consensus Kozak; such "other transcriptional cassettes" typically are utilized to allow for "enhancement," "amplification" or "regulation" of gene product repression. For example, co-transfection of the exogenous DNA with the dehydrofolate reductase (DHFR) gene is exemplary. By increasing the levels the antifolate drug methotrexate (MTX), a competitive inhibitor of DHFR, presented to such cells, an increase in DHFR production can occur via amplification of the DHFR gene. Beneficially, extensive amounts of flanking exogenous DNA will also become amplified; therefore, exogenous DNA inserted co-linear with an expressible DHFR gene will also become overexpressed. Additionally, transciptional cassettes which allow for regulation of expression are available. For example, temperature sensitive COS cells, derived by placing SV40ts mutant large T antigen gene under the direction of Rous sarcoma virus LTR (insensitive to feedback repression by T antigen), has been described. *See*, 227 *Science* 23-28 (1985). These cells support replication from SV40 *ori* at 33°C but not at 40°C and allow regulation of the copy number of transfected SV40 *ori*-containing vectors. The foregoing is not intended, nor is it to be construed, as limiting; rather the foregoing is intended to be exemplary of the types of cassettes which can be incorporated into expression vectors comprising the disclosed fully impaired consensus Kozak. The skilled artisan is credited with the ability to determine the specific type of other transciptional cassettes, vis-a-vis the objective of the expression system, which are applicable and which can be advantageously exploited.

As indicated above, in particularly preferred embodiments of the invention, at least one out-of-frame start codon (*ie* ATG) is located upstream of the fully impaired consensus Kozak start codon, without a stop codon being located between the out-of-frame start codon and the

This is the replacement sheet
14 MAR 1997
Date

fully impaired consensus Kozak start codon. The intent of the out-of-frame start codon is to, in effect, further impair translation of the dominant selectable marker.

As used herein, the term "stop codon" is meant to indicate a "nonsense codon," *ie* a codon which does not encode one of the 20 naturally occurring amino acids such that translation of the encoded material terminates at the region of the stop codon. This definition includes, in particular, the traditional stop codons TAA, TAG and TGA.

As used herein, the term "out-of-frame" is relative to the fully impaired consensus Kozak start codon. As those in the art appreciate, in any DNA macromolecule (or RNA macromolecule) for every in-frame sequence, there are two out-of-frame sequences. Thus, for example, with respect to the following sequence incorporating a fully impaired consensus Kozak:

```
                    -3      +1
gcA TGc cAT Ggc Cxx ATG Cxx
```

the in-frame codons are separated by triplets, *eg*, gcA, TGc, cAT and Ggc; the out-of-frame codons would include, *eg* cAT, ATG, Gcc, ccA, ATG and TGg. Thus, two start codons (in capital letters and underlined) are out-of-frame relative to the start codon of the fully impaired consensus Kozak.

When such an out-of-frame start codon is utilized, it is preferred that this be within about 1000 nucleotides upstream of the fully impaired consensus Kozak start codon, more preferably within about 350 nucleotides upstream of the fully impaired consensus Kozak start codon, and most preferably within about 50 nucleotides of the fully impaired consensus Kozak start codon.

As is appreciated, the upstream sequence can be manipulated to achieve positioning at least one out-of-frame sequence upstream of the fully impaired consensus Kozak start codon using (most preferably) a mutational primer used in the type of amplification protocol described above.

Utilization of a fully impaired consensus Kozak start codon located within a secondary structure (*ie* a "stem-loop" or "hairpin") is beneficially viable to impairment of translation of the protein encoded by the dominant selectable marker. In such an embodiment, it is preferred that this start codon be located within the stem of a stem loop secondary structure. These, by way of schematic example, in such an embodiment, the start codon of the fully impaired consensus Kozak is positioned as follows:

```
          T
        x   x
         x x
         x A
         x T
         x G
      xA TGxx Cxx
```

(An out-of-frame start codon which is not part of the secondary structure is also represented.) As is appreciated, within the stem loop, complementarity along the stem is, by definition, typically required. For exemplary methodologies regarding, *inter alia*, introduction of such secondary structures into the sequence, as well as information regarding secondary structure stability, *see*, Kozak, *PNAS*, 1986, supra.

As noted, it is preferred that a dominant selectable market with a naturally occurring intronic insertion region or an artificially created intronic insertion region be utilized and at least one gene product of interest inserted within this region. While not wishing to be bond by any particular theory, the inventor postulates that such an arrangement increases expression efficiency because the number of viable colonies that survive the selection process via-a-vis the dominant selectable marker will decrease; the colonies that do survive the selection process will, by definition, have expressed the protein necessary for survival, and in conjunction therewith, the gene product of interest will

This is the replacement sheet

Date: 14 MAR 1997

have a greater tendency to be expressed. As further postulated, the RNA being transcribed from the gene product of interest within the intronic insertion region interferes with completion of transcription (elongation of RNA) of the dominant selectable marker; therefore, the position that the dominant selectable marker is integrated within the cellular DNA is likely to be a position where a larger amount of RNA is initially transcribed.

As is appreciated, prokaryotic proteins do not typically include splices and introns. However, the majority of dominant selectable markers which are preferred for expression vector technology are derived from prokaryotic systems. Thus, when prokaryotic-derived dominant selectable markers are utilized, as is preferred, it is often necessary to generate an artificial splice within the gene so to create a location for insertion of an intron comprising the gene product of interest. It is noted that while the following rules are provided for selection of a splice site in prokaryotic genes, they can be readily applied to eukaryotic genes.

A general mechanism for the splicing of messenger RNA precursors in eukaryotic cells is delineated and summarized in Sharp, Philip A. "Splicing of Messenger RNA Precursors" *Science, 235*: 736-771 (1987) (*see*, in particular, Figure 1) which is incorporated herein by reference. Based upon Sharp, there are four minimum criteria in the nucleic acid sequence which are necessary for a splice: (a) 5' splice donor; (b) 3' splice acceptor; (c) branch point, and (d) polypyrimidine tract. The consensus sequences for the 5' splice donor is reported to be

$$\overset{C\quad A}{AAG/GTGAGT}$$

and for the 3' splice acceptor, NCAG/G (where a "/" symbol indicates the splice site); *see*, Mount, S.M. "A Catalogue of Splice Junction Sequences" *Nuc. Acids. Res. 10/2*:459-472 (1992) which is incorporated herein by reference. The consensus sequence for the branch point, *ie.*the location of the lariat formation with the 5' splice donor, is

This is the replacement sheet

14 MAR 1997

reported as PyNPyPAPy; and the reported preferred branch site for mammalian RNA splicing is TACTAAC (Zhuang, Y. et al. "UACUAAC is the preferred branch site for mammalian mRNA splicing" *PNAS 86*: 2752-2756 (1989), incorporated herein by reference). Typically, the branch point is located at least approximately 70 to about 80 base-pairs from the 5'-splice donor (there is no defined upper limit to this distance). The poly pyrimidine tract typically is from about 15 to about 30 base pairs and is most typically bounded by the branch point and the 3' splice acceptor.

The foregoing is descriptive of the criteria imposed by nature on naturally occurring splicing mechanisms. Because there is no exact upper limit on the number of base pairs between the 5' splice donor and the branch point, it is preferred that the gene product of interest be inserted within this region in situations where a natural intron exists within the dominant selectable marker. However, as noted, such introns do not exist within most of the preferred dominant selectable markers; as such, utilization of artificial introns are preferably utilized with these markers.

In order to generate an artificial intron, a "splice donor:splice acceptor" site must be located within the encoding region of the dominant selectable marker. Based upon Sharp and Mount, it is most preferred that the following sequence function as the splice donor:splice acceptor site -- CAGG (with the artificial splice occurring at the GG region). A preferred sequence is AAGG.

Focusing in on the most preferred sequence CAGG, the following codons and amino acids can be located within the encoding region of the dominant selectable marker for generation of the artificial intronoc insertion region:

A		B		C		
CAG/	GNN	NCA	G/GN	NNC		AG/G
Gln	Ala	Ala	Gly	Ala	Leu	Arg
	Asp	Pro		Arg	Phe	
	Gly	Ser		Asn	Pro	
	Glu	Thr		Asp	Ser	
	Val			Cys	Thr	
				Gly	Tyr	
				His	Val	
					Ile	

(As will be appreciated, the same approach to determining viable amino acid residues can be utilized for the preferred sequence of AAGG). The most preferred codon group for derivation of the splice donor:splice acceptor site is group A. Once these amino acid sequences are located, a viable point for generation of an artificial intronic insertion region can be defined.

Focusing on the preferrred NEO dominant selectable marker, amino acid residues Gln Asp (codon group A) are located at the positions 51 and 52 of NEO and amino acid residues Ala Arg (codon group C) are located at positions 172 and 173 (as is appreciated, multiple artificial intronic insertion regions may be utilized). Focusing on residues 50 - 53 of NEO, the nucleic acid and amino acid sequences are as follows:

	50	51	52	53	
5'	CTG	CAG	GAC	GAG	3'
	Leu	Gln	Asp	Glu	

Accordingly, an artificial intronic insertion region can be generated between residues 51 and 52 of NEO. This region most preferably comprises a branch point, a polyprimidine tract and, preferably, a region for insertion of a gene product of interest, *ie* a region amendable to enzymatic digestion.

Two criteria are import for the artificial intronic insertion region: the first two nucleic acid residues of the 5' splice site (*eg* abutting CAG) are most preferably

This is the replacement sheet

14 MAR 1997

Date:

GT and the first two nucleic acid residues of the 3' splice site (*eg* abutting G) are most preferably AG.

Using the criteria defined abouve, an artificial intronic insertion region was between amino acid residues 51 and 52 of NEO:

```
   50 NEO 51                                                  52 NEO 53
   LEU GLN                    Branch        Polypyrimide Tract ASP GLU 3'
5' CTG CAG/GTAAGT GCGGCCGC TACTAAC (TC)3CT(C)3TCC(T)5 C CTGCAG/GAC GAG
                   Not I
```

(Details regarding the methology for creating this artificial intronic insertion region are set forth in the Example Section to follow). The Not I site was created as the region where the gene product of interest can be incorporated. Therefore, upon incorporation, the gene product of interest is located between amino acid residues 51 and 52 of NEO, such that during NEO transmission, the gene product of interest will be "spliced-out".

The host cell line is most preferably of mammalian origin;those skilled in the art are credited with ability to preferentially determine particular host cell lines which are best suited for the desired gene product to be expressed therein. Exemplary host cell lines include, but are not limited to, DG44 and DXB11 (Chinese Hamster Ovary lines, DHFR minus), HELA (human cervical carcinoma), CV1 (monkey kidney line), COS (a derivative of CV1 with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3x63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte) and 293 (human kidney). Host cell lines are typically available from commercial services, the American Tissue Culture Collection or from published literature.

Preferably the host cell line is either DG44 or SP2/O. *See*, Urland, G. *et al.*, "Effect of gamma rays and the dihydrofolate reductase locus: deletions and inversions." *Som. Cell & Mol. Gen. 12/6*:555-566 (1986) and Shulman, M. *et al.*, "A better cell line for making

This is the replacement sheet
14 MAR 1997
Date

hybridomas secreting specific antibodies." *Nature 276*:269 (1978), respectively. Most preferably, the host cell line is DG44. Transfection of the plasmid into the host cell can be accomplished by any technique available to those in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), cell fusion with enveloped DNA, microinjection, and infection with intact virus. *See*, Ridgway, A.A.G. "Mammalian Expression Vectors." Chapter 24.2, pp. 470-472 *Vectors*, Rodriguez and Denhardt, Eds. (Butterworths, Boston, MA 1988). Most preferably, plasmid introduction into the host is via electroporation.

Examples

The following examples are not intended, nor are they to be construed, as limiting the invention; the examples are intended to demonstrate the applicability of an embodiment of the invention disclosure herein. The disclosed fully impaired consensus Kozak sequence is intended to be broadly applied as delineated above. However, for presentational efficiency, exemplary uses of particularly preferred embodiments of fully impaired consensus Kozak sequences are utilized in conjunction with tandem chimeric antibody expression vectors (also referred to herein as antibody expression vectors) as disclosed below.

I. Tandem Chimeric Antibody Expression ("TCAE") Vector

B cell lymphocytes arise from pluripotent stem cells and proceed through ontogeny to fully matured antibody secreting plasma cells. The human B lymphocyte-restricted differentiation antigen Bp35, referred to in the art as "CD20," is a cell surface non-glycosylated phosphoprotein of 35,000 Daltons; CD20 is expressed during early pre-B cell development just prior to the expression of cytoplasmic μ heavy chains. CD20 is expressed consistently until the plasma cell differentiation stage. The CD20 molecule regulates a step in the activation

process which is required for cell cycle initiation and differentiation. Because CD20 is expressed on neoplastic B cells, CD20 provides a promising target for therapy of B cell lymphomas and leukemias. The CD20 antigen is especially suitable as a target for anti-CD20 antibody mediated therapy because of accessibility and sensitivity of hematopoietic tumors to lysis via immune effector mechanisms. Anti-CD20 antibody mediated therapy , *inter alia*, is disclosed in co-pending Serial No. 07/978,891 and Serial No. _____ , filed simultaneously herewith. The antibodies utilized are mouse/human chimeric anti-CD20 antibodies expressed at high levels in mammalian cells (chimeric anti-CD20"). This antibody was derived using vectors disclosed herein, to wit: TCAE 5.2; ANEX 1; ANEX 2; GKNEOSPLA3F; and NEOSPLA3F (an additional vector, TCAE 8, was also utilized to derive chimeric anti-CD20 antibody -- TCAE 8 is identical to TCAE 5.2 except that the NEO translational start site is a partially impaired consensus Kozak. TCAE 8 is described in the co-pending patent document filed herewith.).

In commonly-assigned United States Serial Number 07/912,292, disclosed, *inter alia*, are human/Old World monkey chimeric antibodies; an embodiment of the invention disclosed therein are human/macaque chimeric anti-CD4 antibodies in vector TCAE 6 (*see*, Figure 6 of Serial Number 07/912,292, and corresponding discussion). TCAE 6 is substantially identical to TCAE 5.2; TCAE 6 contains human lambda constant region, while TCAE 5.2 contains human kappa constant region. TCAE 5.2 and ANEX 1 (referred to in that patent document as TCAE 12) are disclosed as vectors which can be utilized in conjunction with human/Old World monkey chimeric antibodies. The comparative data set forth in Serial No. 07/912,292 vis-a-vis TCAE 5.2 and ANEX 1 is relative to expression of chimeric anti-CD20 antibody.

TCAE 5.2 was derived from the vector CLDN, a derivative of the vector RLDN10b (*see*, 253 *Science* 77-91, 1991). RLDN10b is a derivative of the vector TND (*see*,

This is the replacement sheet

Date: 14 MAR 1997

7*DNA* 651-661, 1988). *Ie* the vector "family line" is as follows: TDN → RLDN10b → CLDN → TCAE 5.2 → ANEX 1 (the use of the "→" symbol is not intended, nor is it to be construed, as an indication of the effort necessary to achieve the changes from one vector to the next; *eg* to the contrary, the number and complexity of the steps necessary to generate TCAE 5.2 from CLDN were extensive).

TND was designed for high level expressions of human tissue plasminogen activator. RLDN10b differs from TND in the following ways: the dihydrofolate reductase ("DHFR") transcriptional cassette (comprising promoter, murine DHFR cDNA, and polyadenylation region) was placed in between the tissue plasminogen activator cassette ("t-PA expression cassette") and the neomycin phosphotransferase ("NEO" cassette") so that all three cassettes were in tandem and in the same transcriptional orientation. The TND vector permitted selection with G418 for cells carrying the DHFR, NEO and t-PA genes prior to selection for DHFR gene amplification in response to methotrexate, MTX. The promoter in front of the DHFR gene was changed to the mouse beta globin major promoter (*see*, 3 *Mol. Cell. Bio.* 1246-1254, 1983). Finally, the t-PA cDNA was replaced by a polylinker such that different genes of interest can be inserted in the polylinker. All three eukaryotic transcriptional cassettes (t-PA, DHFR, NEO) of the TND vector can be separated from the bacterial plasmid DNA (pUC19 derivative) by digestion with the restriction endonuclease Not I.

CLDN differs from RLDN10b in the following ways: The Rous LTR, positioned in front of the polylinker, was replaced by the human cytomegalovirus immediate early gene promoter enhancer ("CMV"), (*see*, 41 *Cell* 521, 1985), from the Spe I site at -581 to the Sst I site at -16 (these numbers are from the *Cell* reference).

As the name indicates, TCAE vectors were designed for high level expressions of chimeric antibody. TCAE 5.2 differs from CLDN in the following ways:

This is the replacement sheet

Date: 14 MAR 1991

A. TCAE 5.2 comprises four (4) transcriptional cassettes, as opposed to three (3), and these are in tandem order, *ie* a human immunoglobulin light chain absent a variable region; a human immunoglobulin heavy chain absent a variable region; DHFR; and NEO. Each transcriptional cassette contains its own eukaryotic promoter and polyadenylatin region (reference is made to Figure 2 which is a diagrammatic representation of the TCAE 5.2 vector). The CMV promoter/enhancer in front of the immunoglobulin heavy chain is a truncated version of the promoter/enhancer in front of the light chain, from the Nhe I site at -350 to the Sst I site at -16 (the numbers are from the *Cell* reference, *supra*). Specifically,

1) A human immunoglobulin light chain constant region was derived via amplification of cDNA by a PCR reaction. In TCAE 5.2, this was the human immunoglobulin light chain kappa constant region (Kabat numbering, amino acids 108-214, allotype Km 3), and the human immunoglobulin heavy chain gamma 1 constant region (Kabat numbering amino acids 114-478, allotype Gm1a, Gm1z). The light chain was isolated from normal human blood (IDEC Pharmaceuticals Corporation, La Jolla, CA); RNA therefrom was used to synthesize cDNA which was then amplified using PCR techniques (primers were derived vis-a-vis the consensus Kabat). The heavy chain was isolated (using PCR techniques) from cDNA prepared from RNA which was in turn derived from cells transfected with a human IgG1 vector (*see*, 3 *Prot. Eng.* 531, 1990; vector $pN_{\gamma 1}62$). Two amino acids were changed in the isolated human IgG1 to match the consensus amino acid sequence in Kabat, to wit: amino acid 225 was changed from valine to alanine (GTT to GCA), and amino acid 287 was changed from methionine to lysine (ATG to AAG);

2) The human immunoglobulin light and heavy chain cassettes contain synthetic signal sequences for secretion of the immunoglobulin chains;

3) The human immunoglobulin light and heavy chain cassettes contain specific DNA restriction sites which allow for insertion of light and heavy immunoglobulin variable regions which maintain the transitional reading frame and do not alter the amino acids normally found in immunoglobulin chains;

4) The DHFR cassette contained its own eukaryotic promoter (mouse beta globin major promoter, "BETA") and polyadenylation region (bovine growth hormone polyadenylation, "BGH"); and

5) The NEO cassette contained its own eukaryotic promoter (BETA) and polyadenylation region (SV40 early polyadenylation, "SV").

With respect to the TCAE 5.2 and the NEO cassette, the Kozak region was a consensus Kozak (which included an upstream Cla I site SEQ ID NO: 7):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccAcc ATG Gtt

ANEX 1 (previously named TCAE 12 in the referenced case) is identical to TCAE 5.2 except that in the NEO cassette, the Kozak region was fully impaired (SEQ ID NO: 8):

ClaI -3 +1

TTGGGAGCTTGG ATCGAT ccTcc ATG Ctt

As disclosed in the commonly-assigned referenced case, the impact of utilization of the fully impaired consensus Kozak was striking: relative to TCAE 5.2, there was a significant (8-fold) reduction in the number of ANEX 1 G418 resistant colonies (258 from two electroporations versus 98 from six electroporations) from the same amount of plasmid DNA transfected per cell; and, there was a significant increase in the amount of co-linked gene product expressed in each of the ANEX 1 clones. Referencing the histogram of Figure 3 (Figure 16 of the commonly assigned referenced case), 258 colonies were derived from 2 electroporations of 25 µg of DNA containing a neomycin phosphotransferase gene with a consensus Kozak at the translation start site. Two-hundred and one (201)

This is the replacement sheet

14 MAR 1997

Date:

of these colonies did not express any detectable gene product (less than 25 ng/ml of chimeric immunoglobulin), and only 8 colonies expressed more than 100 ng/ml. Again, referencing Figure 3, 98 colonies were derived from 6 electroporations for ANEX 1 of 25 µg of DNA containing a neomycin phosphotransferase gene with the fully impaired consensus Kozak at the translation start site (6 electroporations were utilized in order to generate statistically comparative values; this was because on average, each electroporation for ANEX 1 yielded about 16 colonies, as opposed to about 129 colonies per electroporation for TCAE 5.2). Eight (8) of the ANEX 1 colonies did not express any detectable gene product (less than 25 ng/ml), while 62 of these colonies were expressing greater than 100 ng/ml; of these 62 colonies, nearly 23 were expressing over 250 ng/ml (23%), with 6 expressing greater than 1000 ng/ml (6%).

The foregoing evidences, *inter alia*, the following: 1) because the difference between TCAE 5.2 and ANEX 1 was limited to the Kozak translation start site of the NEO gene, and because the gene product of interest (chimeric anti-CD20 antibody) was co-linked to the NEO gene, a conclusion to be drawn is that these differences in results are attributed solely to the differences in the Kozak translation start site; 2) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable number resulted in significantly less viable colonies; 3) it was experimentally confirmed that utilization of a fully impaired consensus Kozak in conjunction with a dominant selectable marker co-linked to a desired gene product significantly increased the amount of expressed gene product. Thus, the number of colonies to be screened decreased while the amount of expressed gene product increased.

This is not an official document sheet

Date: 14 MAR 1997

II. Impact Of Out-Of-Frame Start Sequence

Conceptually, further impairment of translation initiation of the dominant selectable marker of ANEX 1 could be effectuated by utilization of at least one out-of-frame ATG start codon upstream of the neomycin phosphotransferase start codon. Taking this approach one step further, utilization of a secondary structure ("hairpin") which incorporated the *neo* start codon within the stem thereof, would be presumed to further inhibit translation initiation. Thus, when the out-of-frame start codon/fully impaired consensus Kozak was considered, this region was designed such that the possibility of such secondary structures was increased.

As indicated previously, the Kozak region for the *neo* start codon in the ANEX 1 vector is:

TTGGGAGCTTGG ATCGAT CC Tcc ATG Ctt

The desired sequence for a vector identical to ANEX 1 but incorporating the above-identified changes vis-a-vis the *neo* start codon, referred to as ANEX 2, is as follows (SEQ ID NO: 9):

CCA GCA TGG AGG A ATCGAT CC Tcc ATG Ctt

(The out-of-frame start codon is underlined.) The fully impaired consensus Kozak of ANEX 2 is identical to that of ANEX 1. The principal difference is the inclusion of the upstream out-of-frame start codon. A possible difference is the formation of a secondary structure involving this sequence, proposed as follows:

```
CG   )
T  A  )  CLA I site
A  T  )
 AT
 GC
 GC
 AT
 GC
 GC
 TA
 AT
 CG
 GC
 AT
 CT
 CG
```

The sequence in bold, ATG, is the upstream out-of-frame start codon; the "loop" portion of the secondary structure is the CLA I site; and the sequence between the "T" and "C" (italics and bold) is the start codon (underlined) of the fully impaired consensus Kozak.

In order to effectuate this change, a PCR fragment was cloned into anti-CD20 in ANEX 1 from Xho I (5520) to Cla I (5901); *see*, Figure 4. Primers were as follows:

3'-Primer 489 (SEQ ID NO: 10):

```
5'-GGA GGA TCG ATT CCT CCA TGC TGG
   CAC AAC TAT GTC AGA AGC AAA TGT
   GAG C-3'
```

The upper-lined portion of Primer 489 is a Cla I site; the under-lined portion is the fully impaired consensus Kozak translation start site.

5'-Primer 488 (SEQ. ID. NO. 11):

```
5'-CTG GGG CTC GAG CTT TGC-3'
```

The upper-lined portion of Primer 485 is an Xho I site.

This is the replacement sheet

Date: 14 MAR 1997

These primers were prepared using an ABI 391 PCR MATE™ DNA synthesizer (Applied Biosystems, Foster City, CA). Phosphoramidites were obtained from Cruachem (Glasgow, Scotland): dA(bz) - Prod. No. 20-8120-21; dG(ibu) - Prod. No. 20-8110-21; dC(bz) - Prod. No. 20-8130-21; T - Prod. No. 20-8100-21.

Conditions for the PCR reaction using these primers were as follows: 2λ ("microliters") of anti-CD20 in TCAE 5.2 in plasmid grown in *E. coli* strain GM48 (obtained from the ATCC) was admixed with 77λ of deionized water; 2λ of Primer 488 (64 pmoles); and 4λ of primer 489 (56 pmoles). This was followed by a denaturation step (94°C, 5 min.) and a renaturation step (54°C, 5 min.). Thereafter, 4λ of 5 mn dNTPS (Promega, Madison, WI: dATP, Prod. No. U1201; dCTP, Prod. No. U1221; dGTP, Prod. No. U1211; dTTp, Prod. No. U1231), 1λ of Pfu DNA polymerase (Stratagene, La Jolla, CA Prod. No. 600135, 2.5 U/ml), and 50λ of mineral oil overlay was added thereto, followed by 30 cycles, with each cycle comprising the following: 72°C, 2 min.; 94°C, 1 min.; 54°C, 1 min. Ten microliters (10λ) of this admixture was analyzed by agarose gel electrophoresis (results not shown); a single band was found at about 400 base pairs.

The PCR product and the vector were prepared for ligation as follows: Anti-CD20 in ANEX 1 plasmid grown in *E. coli* bacterial strain GM48 was digested with Cla 1 and Xho 1 as follows: 20λ of anti-CD20 in ANEX 1 was admixed with 10λ of 10XNEB4 buffer (New England Biolabs, Beverly, MA; hereinafter, NEB); 5λ Cla 1 (NEB, Prod. No. 197 S, 60u); and 64λ deionized water. This admixture incubated overnight at 37°C, followed by the addition of 5λ Xho 1 (NEB, Prod. No. 146 S, 100 u) and incubation at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut ANEX 1". The approximate 400 base pair PCR fragment was prepared and digested with Cla 1 and Xho 1 as follows: 90λ of the PCR fragment was admixed with 10λ of 3M NaOAc; 1λ 10% sodium dodecyl sulfate (SDS); and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30

This is the replacement sheet

Date: 14 MAY 1997

sec. followed by a 1 min. spin (1700 RPM). The aqueous phase was subjected to a spin column which resulted in 85λ total admixture. To this admixture was added 10λ 10XNEB4, 1λ bovine serum albumin (BSA,100X; NEB), 2λ Cla 1 (24u), and 2λXho 1 (40 u). This admixture was incubated at 37°C for 2 hrs. The resulting material is designated herein as "Cla 1/Xho 1 cut PCR 488/489". Both Cla 1/Xho 1 cut ANEX 1 and Cla 1/Xho 1 cut PCR 488/489 were analyzed by agarose gel electrophoresis and the resulting bands were observed at the same relative location on the gel (results not shown).

Ligation of Cla 1/Xho 1 cut PCR 488/489 and Cla 1/Xho 1 cut ANEX 1 was accomplished as follows: 1λ of tRNA (Sigma, St. Louis, MO, Prod. No. R-8508) was admixed with 1λ 10% SDS; 10λ 3M NaOAc; 45λ of Cla 1/Xho 1 cut PCR 488/489 (about 22.5 ng); a 1:4 dilution (0.25λ) of Cla 1/Xho 1 cut ANEX 1 (about 32 ng) in 0.75λ tris-hydroxymethyl aminomethane ethylenediamine tetracetic acid (TE); and 42λ TE. To this admixture was added 90λ phenyl/$CHCl_3$/isoamyl, followed by a 30 sec. vortex and a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube, followed by addition of 270λ of 100% EtOH (-20° C), 10 min. spin (13,000 RPM) followed by addition of another 270λ of 100% EtOH (-20° C) and another 1 min. spin (13,000 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 17λ TE, 2λ ligase buffer (Promega, T4 DNA Ligase kit, Prod. No. M180) and 1λ ligase (Promega Ligase kit). This ligation mix incubated at 14°C overnight. Twenty microliters (20λ) of the ligation mix was admixed with 10λ 3M NaOAc, 1λ 10% SDS, 69λ TE, and 90λ phenyl/$CHCl_3$/isoamyl. This admixture was vortexed for 30 sec., followed by a 1 min. spin (1700 RPM). The aqueous phase was transferred to a new tube and 270λ of 100% EtOH (-20°C) was added thereto, followed by a 10 min. spin (1700 RPM). This admixture was dried in a SpeedVAC™ and resuspended in 20λ TE. Ten microliters of the resuspended admixture was transformed in *E. coli* X-L1 blue™ (Stratagene, La Jolla, CA), following manufacturer

instructions. Ten (10) bacterial colonies were inoculated in LB Broth (Gibco BRL, Grand Island, NY, Prod. No. M27950B) including ampicillin (50µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification System (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the ANEX 2 vector, depending on the sufficiency of the foregoing.

ANEX 2 includes a Hinf I site ("GAATC") upstream of the *neo* start site (-9 to -13 relative to the *neo* start codon); ANEX 1 does not include this Hinf I site. The purified plasmids comprising putative ANEX 2, and previously purified ANEX 1 standard , were subjected to Hinf 1 digestion as follows: 2λ of each isolate was admixed with 8λ of Hinf I digestion buffer (15λ 10 x NEB2 buffer; 15λ Hinf I (NEB, Prod. No. 155S, 10u/λ); and 90λ H_2O). This admixture incubated for 3 hrs. at 37°C and each isolate was analyzed via agarose gel electrophoresis (results not shown); nine (9) of the bands were substantially identical to the ANEX 1 standard, one(1) showed a slight difference in band pattern. For this single isolate, the first two bands were at 1691 and 670 kB; for the ANEX 1 Hinf I digested product, the first three bands were at 1691, 766, and 670 kB. The missing band at 766 kB for the single isolate was attributed to the presence of the Hinf I site therein, indicating that the desired change to ANEX 1 was incorporated into this vector. This vector was designated "Anti-CD20 in ANEX 2 (G1,K)," and is generally referred to by the inventor as ANEX 2.

Electroporation of anti-CD20 in ANEX 2 was accomplished as follows: two-hundred and forty microliters (240λ) of the anti-CD20 in ANEX 2 DNA (400µg) was admixed with 100λ of 10 X NEB2 buffer ; 100λ of Stu I (NEB, Prod. No. 187S, 1000 u); and 560λ TE, and incubated at 37°C for 2hrs. This admixture was then placed over 8 spin columns (125λ each), followed by addition of 110λ 10X Not I buffer (NEB); 10λ 100X BSA; and 20λ of Not I (NEB,

This is the replacement sheet

Date: 14 MAR 1997

Prod. No. 189S, 800u). This admixture was incubated at 37°C for 3 hrs., followed by the addition of 120λ of 3M NaOAC and 12λ of 10% SDS. The admixture was transferred to 2 vortex tubes and 500λ of phenyl/$CHCl_3$/isoamyl was added to each, followed by a 30 sec. vortex and 1 min. spin (1700 RPM). The aqueous phase was removed from the tubes and segregated into 3 tubes, followed by the addition to each tube of -20°C 100% ETOH, followed by 10 min. spin (13,000 RPM). Thereafter, -20°C 70% ETOH was added to each tube, followed by 1 min. spin (13,000 RPM). The tubes were then placed in a Speed VAC™ for drying, followed by resuspension of the contents in 100λ TE in a sterile hood. Five microliters (5λ) of the resuspended DNA was admixed with 995λ of deionized water (1:200 dilution). An optical density reading was taken (OD=260) and the amount of DNA present was calculated to be 0.75μg/λ. In order to utilize 25 μg of DNA for electroporation, 32λ of the 1:200 dilution of the DNA was utilized (25 μg was utilized as this was the amount of DNA utilized for TCAE 5.2 and ANEX 1 in the foregoing Example 1). The 1:200 dilution of DNA was formally referred to as "Stu 1, Not I cut anti-CD20 in ANEX 2 (25μg) in TE" and generally referred to as "anti-CD20 in ANEX 2."

Host cells utilized was DG44 CHO ("CHO") (*see*, Urlaub, G. *Somatic Cell*, 1986 *supra*). One hundred milliliters of 6.6×10^5 cells/ml (84%) were subjected to a 2 min. spin at 1000 RPM. These were washed with 50 ml sucrose buffered solution, followed by 5 min. spin at 1000 RPM; the material was then resuspended in 4.5 ml of the sucrose buffered solution. Thereafter, cells were counted and 0.4 ml of CHO cells (4.0×10^6 cells) were admixed with 32λ of the anti-CD20 in ANEX 2 in BTX sterile, disposable electroporation cuvettes. Electroporation settings were as follows: 210 volts; 400 microfaraday capacitance; 13 ohms resistance, using a BTX 600™ electro cell manipulator (BTX, San Diego, CA). Nine (9) electroporations were conducted; actual voltage delivered over actual times were as follows: 1-199V, 4.23 msec; 2-

This is the replacement sheet
14 MAR 1997
Date:

188V, 4.57 msec; 3-189V, 4.24 msec, 4-200V; 4.26 msec, 5-200V, 4.26 msec; 6-199V, 4.26 msec; 7-189V, 4.59 msec; 8-189V, 4.57 msec; 9-201 V, 4.24 msec. (As noted in Example I, the difference in number of performed electroporations was attributed to the need to achieve a statistically significant number of viable colonies for each of the three conditions, TCAE 5.2, ANEX 1 and ANEX 2; the amount of DNA used for each electroporation (25 µg) was the same for each, and the same number of cells were electroporated.

Thereafter, the electroporation material was admixed with 20 ml of G418 Growth Media (CHO-S-SFM II minus hypoxanthine and thymidine (Gibco, Grand Island, NT, Form No. 91-0456PK)including 50 µM hypoxanthine and 8 µM thymidine). The admixture was gently agitated, followed by plating 200 µl of the admixture per well into 96-well plates, one plate for each electroporation (nine). Beginning on day 2 after electroporation, through day 17, 150µl of each well was removed, and 150 µl of fresh G418 Growth Media containing 400 µg/ml G418 was added thereto. Colonies were analyzed on day 25.

One hundred and twenty one (121) colonies expressed anti-CD20 antibody (*ie* , 13 colonies per electroporation). Of these, 63 (52%) expressed over 250µg /ml of protein; of the 63, 20 of the colonies (16.5%) expressed over 1000µg/ml of protein. Only 5 of the 121 colonies (4.1%) expressed less than 25µg/ml of protein. Figure 5 provides a histogram comparing expression of protein per colonies derived from the vectors TCAE 5.2, ANEX 1 and ANEX 2.

The foregoing data indicates that, *inter alia*, as between ANEX 1 and ANEX 2, the use of at least one out-of-frame start codon upstream of a fully impaired consensus Kozak associated with the translation initiation of a dominant selectable marker decreases the number of viable colonies expressing co-linked gene product and significantly increases the amount of expressed co-linked gene product.

III. Impact Of Insertion Of Gene Product Of Interest Within At Least One Artificial Intronic Insertion Region Of A Dominant Selectable Marker

Building further upon the ANEX 2 vector, an artificial splice was generated between amino acid residues 51 and 52 of the NEO coding region of ANEX 2, followed by insertion therein of the anti-CD20 encoding region. Two such vectors were generated: the first, comprising a consensus Kozak sequence for the NEO translation initiation codon and not comprising an out-of-frame start codon, is referred to as "GKNEOSPLA3F;" the second, comprising a fully impaired consensus Kozak and an out-of frame start condon, is referred to as "NEOSPLA3F."

Both GKNEOSPLA3F and NEOSPLA3F contain the following artificial intron sequence between amino acid residues 51 and 52 of NEO:

51 52

5'CTG CAG/GTAAGT GCGGCCGC TACTAAC $(TC)_3$ CT $(C)_3$ TCC $(T)_5$ C CTGCAG/GAC GAG 3'

The underlined portion represents a sequence amenable to digestion with Not I enzyme; the encoding region for anti-CD20, *inter alia*, was inserted within this region.

Although not wishing to be bound by any particular theory, the inventor postulates that during expression, the inclusion of a gene product of interest within an artificial intronic insertion region of a dominant selectable marker (*e.g.* the NEO gene) should significantly decrease the number of viable colonies producing, in the case of the disclosed GKNEOSPA3F and NEOSPLA3F vectors, anti-CD20 antibody. This is predicated upon two points: first, only those vectors which are able to transcribe and correctly splice-out the antibody encoding region and correctly translate NEO will be G418 resistant; second, because each antibody cassette has its own promoter and polyadenylation region, transcription and translation of the antibody is independent of translation of NEO.

The GKNEOSPLA3F and NEOSPLA3F vectors were constructed in the following manner:

This is the replacement sheet

Date: 14 MAR 1997

Anti-CD20 in ANEX 2 was digested with Not I and Xho I in order to isolate the 1503 bp NEO cassette DNA fragment (see Figure 4 between "Not I 7023" and "Xho I 5520") as follows: 10 µl of anti-CD20 in ANEX 2 was admixed with 6 µl deionized H_2O ("dH_2O"); 1 µl Not I enzyme (NEB, Prod. No. 189S); 2 µl of 10X Not I digestion buffer (NEB; provided with enzyme); and 1 µ Xho I enzyme (Promega, Madison, WS, Prod. No. R4164). This digestion mixture was incubated overnight at 37°C. The resulting digested DNA was size fractionated by 0.8% agarose gel electrophoresis and the desired fragment migrating at 1503 was isolated via the GlassMAX™ method (Gibco BRL, Grand Island, NY, Prod. No. 15590-011) for insertion into pBluescript SK(-) plasmid DNA (Stratagene, La Jolla, CA).

pBluescript SK (-) was previously prepared for acceptance of the NEO cassette by double digestion with Not I and Xho I using the same conditions as above for anti-CD20 in ANEX 2. Digested pBluescript SK (-) was then collected by ethanol precipitation by the addition of 70 µl dH_2O; 2 µl tRNA (Sigma, St. Louis, MO, Prod. No. R-8508); 10 µl of 3M NaOAc; and 300 µl 100% ETOH (-20°C). This was followed by a 10 min spin (13,000 RPM), decanting the supernatant, rinsing with 70% ETOH, decanting the liquid, drying in a SpeedVAC™ and resuspending in 20 µl 1 x TE.

Ligation of the NEO cassette DNA fragment into prepared pBluescript SK (-) vector was accomplished as follows: 10 µl of NEO fragment DNA was admixed with 6 µl dH_2O; 1 µl cut pBluescript SK (-) vector DNA; 2 µl 10 x ligation buffer (Promega, supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega, Prod. No. M1801) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

This is a replacement sheet
14 MAR 1997

M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO+ depending on the sufficiency of the foregoing. (BlueNEO+ was confirmed due to the sufficiency of the following procedure.)

BlueNEO+ contains a Not I restriction recognition sequence reformed upon ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector. This site was destroyed by the following: 1 µl of BlueNEO+ DNA was admixed with 16 µl dH_2O; 2 µ 10 x Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB). This was followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume. This 15 µl Not I digested DNA was "blunt-ended" by admixing with 4 µl 5X Klenow buffer (20 mM Tris-HCL, pH 8.0, 100mM $MgCl_2$) and 1 µl DNA Polymerase I Large (Klenow) Fragment (Promega, Prod. No. M2201). This admixture was incubated at room temperature for 30 minutes. Blunt-ended DNA was then purified by spin column fractionation, giving a final volume of 15 µl.

Ligation of the blunt-ended DNA was performed in an analogous way as to the ligation of the NEO cassette fragment DNA into the pBluescript SK (-) vector except that the final DNA was resuspended in 17 µl of 1 X TE.

Following ligation, the DNA was subjected to a second restriction digestion with Not I by mixing the 17 µl of DNA with 2 µl 10 X Not I digestion buffer and 1 µl Not I enzyme (NEB). Digestion was allowed to proceed at 37°C for 60 minutes. Following digestion, the admixture was purified by spin column fractionation resulting in 15 µl final volume.

Ten (10) µl of the purified DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL including ampicillin (50 µg/ml;

Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmid referred to as BlueNEO- depending on the sufficiency of the foregoing. (BlueNEO- was confirmed due to the sufficency of the following procedure.)

BlueNEO- contains a unique Pst I restriction site spanning the codons for amino acid residues 51 and 52. BlueNEO- was digested with Pst I as follows: an admixture was formed containing 15 µl dH_2O; 1 µl BlueNEO- DNA, 2 µl digestion buffer 3 (NEB) and 2 µl Pst I enzyme (NEB, Prod. No. 140S). This admixture was incubate at 37°C for 3 hrs. Digested DNA was then purified by spin column fractionation. The following synthetic oligonucleotide was then ligated to the Pst I cohesive ends of BlueNEO-: 5' GGTAAGTGCGGCCGCTACTAACTCTCTCCTCCCTCCTTTTTCCTGCA 3' (SEQ ID NO: 12) and its complementary sequence: 5'GCAAAAAGGAGGGAGGAGAGAGTTAGTAGCGGCCGCACTTACCTGCA 3' (SEQ ID NO: 13). Insertion of this linker creates a consensus 5' splice donor site (by ligation) followed by a Not I site, followed by a consensus splice branch point, followed by a synthetic polypyrimidine tract, followed by a consensus 3' splice acceptor site, as indicated above.

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Pst I linearized BlueNEO- DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

The foregoing (and following) synthetic oligo nucleotides were chemically synthesized using an Applied Biosystems 391 PCR MATE™ DNA Synthesizer (Applied Biosystems, Foster City, CA). All reagents for the synthesis were purchased from Applied Biosystems.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene),

This is the replacement sheet

Date: 14 MAR 1997

following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA and/or NEOSPLA- depending on the sufficiency of the foregoing and the orientation of the insertion of the oligonucleotides.

Determination of orientation of the splice junction linker was preformed by nucleic acid sequencing using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical, Cleveland, OH, Prod. No. 70770) following manufacturer instructions. Upon determination of linker orientation within six independent plasmid isolates, identification of NEOSPLA was made such that the inserted splice junction sequences are in the correct forward orientation with respect to the direction of NEO transcription.

NEOSPLA was digested with Xho I by forming an admixture of 15 µl dH_2O; 1 µl NEOSPLA DNA; 2 µl 10 X digestion buffer D (Promega, supplied with enzyme); and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide having the following sequence: 5' TCGATTAATTAA 3' (SEQ ID NO: 14). Insertion of this sequence effectively changes the Xho I site to a Pac I restriction site (as underlined in SEQ ID NO: 14).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized NEOSPLA DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

This is the replacement sheet

Date 14 MAR 1997

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from the 10 cultures with a Promega DNA purification system, following manufacturer instructions; these plasmids may have comprised the plasmid referred to as NEOSPLA3 depending on the sufficiency of the foregoing. (NEOSPLA3 was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K) contains the anti-CD20 light chain and heavy chain immunoglobulin cassettes and a DHFR cassette bounded by a Not I site at the 5' end and an Xho I site at the 3' end. Anti-CD20 in ANEX 2(G1,K) was digested with Xho I by forming an admixture of 15µldH$_2$O, 1 µl anti-CD20 in ANEX 2(G1,K) DNA, 2 µl 10 X digestion buffer D (Promega, supplied with enzyme) and 2 µl Xho I enzyme (Promega, Prod. No. R6161). This admixture was digested at 37°C for 3 hrs followed by DNA purifcation by spin column fractionation. Into this site was ligated a self complementary synthetic oligonucleotide of the following sequence: 5' TCGAAGCGGCCGCT 3' (SEQ ID NO: 15). Insertion of this sequence effectively changes the Xho I site to a Not I restriction site (as underlined in SEQ ID NO: 15).

Ligation was performed as described above for the ligation of the NEO cassette into pBluescript SK (-) except using 2 µl of Xho I linearized anti-CD20 in ANEX 2 DNA and 14 µl (175 pmoles) of annealed complementary oligonucleotides.

Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No.

M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as Anti-CD20 in ANEX 2(G1,K)A depending on the sufficiency of the foregoing. (This was confirmed due to the sufficency of the following procedure.)

Anti-CD20 in ANEX 2(G1,K)A was digested with Not I and Xho I by forming an admixture of 6 µl dH_2O; 10 µl Anti-CD20 in ANEX 2(G1,K); 2 µl 10 x Not I digestion buffer (NEB, supplied with Not I enzyme); and 1 µl Not I enzyme (NEB). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 5515 base pairs by was isolated via the GlassMAX method for insertion into NEOSPLA3.

NEOSPLA3 was previously prepared for acceptance of the anti-CD20 cassette by digestion of 1 µl of DNA with Not I using an admixture comprising 16 µl dH_2O;

2 µl 10X Not I digestion buffer (NEB); 1 µl Not I enzyme (NEB); followed by incubation at 37°C for 2 hrs. This digested DNA was then purified by spin column fractionation resulting in 15 µl final volume.

Ligation of the anti-CD20 DNA fragment into prepared NEOSPLA3 vector was accomplished as follows: 10 µl of anti-CD20 fragment DNA was admixed with 6 µl dH_2O; 1 µl cut NEOSPLA3 vector DNA; 2 µl 10 x ligation buffer (Promega supplied with enzyme); and 1 µl T-4 DNA Ligase (Promega); followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL) including ampicillin (50 µg/ml; Sigma). Plasmids were isolated from

This is the replacement sheet

Date: 14 MAR 1997

the 10 cultures with a Promega DNA purification system following manufacturer instructions; these plasmids may have comprised the plasmids referred to as anti-CD20 in NEOSPLA3F and anti-CD20 in NEOSPLA3R depending on the sufficiency of the foregoing and relative orientation of the inserted fragment with respect to NEO transcription.

Determination of orientation of the anti-CD20 cassette insertion was preformed by double digestion with KpnI and SpeI (NEB, Prod. No. 1335) in NEB buffer 1 plus acetated BSA as follows: an admixture comprising 4 µl DNA; 2 µl NEB buffer 1; 1 µl Kpn I; 1 µl Spe I; 2 µl BSA; and 10 µl dH_20 was formed. The admixture was digested at 37°C for 2 hrs, followed by size fractionation on an 0.8% agarose gel electrophoresis. Upon determination of anti-CD20 insert orientation within six independent plasmid isolates, identification of anti-CD20 in NEOSPLA3F was made such that the inserted sequences are in the forward orientation with respect to the direction of NEO transcription.

The 5515 bp anti-CD20 fragment contains the SV40 origin, a chimeric mouse human immunoglobulin light chain transcriptional cassette, a chimeric mouse human immunoglubulin heavy chain transcriptional cassette, and a murine dihydrofolate reductase transcriptional cassette (see, Figure 4).

Anti-CD20 in NEOSPLA3F was doubly digested with Kpn I and Stu I by creating the admixture consisting of 14 µl dH_2O, 1 µl anti-CD20 in NEOSPLA3F, 2 µl 10 x digestion buffer 1 (NEB, supplied with enzyme), 2 µl 10 x acetylated BSA (NEB supplied with Kpn I enzyme), 1 µl Kpn I enzyme, 1 µl Stu I enzyme (NEB, Prod. Nos. 142S and 187S respectively). This admixture was digested at 37°C for 3 hrs followed by size fractionation by 0.8% agarose gel electrophoresis and the desired fragment migrating at 9368 base pairs by was isolated via the GlassMAX method.

A PCR fragment of DNA was generated from TCAE 5.2. The two following synthetic oligonucleotide primers were utilized in the PCR reaction:

This is the replacement sheet
Date: 14 MAR 1997

5' primer: 5' GCA TGC GGT ACC GGA TCC ATC GAG CTA CTA GCT TTG C 3'

(SEQ ID NO: 16);

3' primer: 5' CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC TGC AGT TCA TCC AGG GC 3' (SEQ ID NO: 17)

The underlined portion of SEQ ID NO: 16 represents a Kpn I site, and the underlined portion of SEQ ID NO: 17 represents a Stu I site.

The PCR product was digested with Kpn I and Stu I and then ligated into prepared anti-CD20 in NEOSPLA3F.

Ligation of the 627 bp fragment into prepared anti-CD20 in NEOSPLA3F was accomplished as follows:

2 µl anti-CD20 in NEOSPLA3F; 1 µl SDS; 1 µl tRNA (Sigma); 11 µl 3M sodium acetate (pH 4.5) were admixed. Following phenol/chloroform isoamyl extraction of the admixture, the DNA was precipitated from the aqueous phase by addition of 270 µl ethanol (ice-cold) and this was spun at 13,000 rpm for 10 min. Following a 70% ETOH wash, the DNA was resuspended in 16 µlTE, 10 µl of PCR fragment DNA was admixed with 6 µl dH_2O, 1 µl cut anti-CD20 in NEOSPLA3F vector DNA, 2 µl 10 x ligation buffer (Promega, supplied with enzyme) and 1 µl T-4 DNA Ligase (Promega) followed by incubation at 14°C overnight. Ligated DNA was collected by ethanol precipitation as described above for the preparation of pBluescript SK (-) vector DNA.

Ten (10) µl of the resuspended ligated DNA was transformed into *E. coli* XL-1 Blue™ (Stratagene), following manufacturer instructions. Ten (10) bacterial colonies were inoculated in LB broth (Gibco BRL, Prod. No. M27950B) including ampicillin (50 µg/ml; Sigma, Prod. No. A-9393). Plasmids were isolated from the 10 cultures with a Promega DNA purification system (Prod. No. PR-A7100), following manufacturer instructions; these plasmids may have comprised the plasmid referred to as anti-CD20 in GKNEOSPLA3F depending on the sufficiency of the foregoing. (Confirmation was based upon sequence determination of the different regions of GKNEOSPLA3F vs. NEOSPLA3F.)

The new plasmid differs from anti-CD20 in NEOSPLA3F in its Kozak sequence for the NEO gene which is:

```
                                        -3          +1
TGT GTT GGG AGC TTG GAT CGAT cc Acc ATG Gtt
                          Cla I    Start NEO
```

for Anti-CD20 in GKNEOSPLA3F, and

```
                                      -3      +1
TGT G CCA GCA TGG AGG AAT CGA Tcc Tcc ATG Ctt
          upstream Start            Start NEO
```

for Anti-CD20 in NEOSPLA3F.

Comparative analysis of expression of anti-CD20 in TCAE 5 vector (comprising NEO with consensus Kozak); ANEX 2 vector (comprising NEO with fully impaired Kozak, and upstream out-of-frame start sequence); NEOPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has fully impaired Kozak and an upstream out-of-frame start sequence); and GKNEOSPLA3F (anti-CD20 inserted via artificial intronic insertion region between amino acids 51 and 52 of NEO; NEO has consensus Kozak).

Twenty-five (25) µg of each plasmid (digested as follows: anti-CD20 in TCAE5 and ANEX2 - Not I; anti-CD20 in NEOSPLA3F - Pac I; anti-CD20 in GKNEOSPLA3F - Pac I and Kpn I) was electroporated into 4 x 10^6 CHO cells; these digestions were utilized to separate the genes expressed in mammalian cells from the DNA used to grow the plasmid in bacteria. Following digestion, EtOH precipitation of the DNA, and drying thereof, the DNA was resuspended in sterile TE at a concentration of 1 µg/µl. Electroporation conditions were as described in Example II, except that 230 volts was utilized and, following electroporation, the mixture of cells and DNA was maintained for 10 min. at room temperature in the sterile, disposable electroporation cuvette.

This is the replacement sheet

Date: 14 MAR 1997

Following electroporation, cells were plated into 96 well dishes as shown below in Table I, based upon the expected frequency of G418 resistant colonies (as derived from preliminary experiments; data not shown):

Table I
Comparative Expression

Plasmid	No. Transfections	No. Cells Plated	No. 96 Well Plates
TCAE 5	1	4×10^5	5
ANEX 2	1	2×10^6	5
GKNEOSPLA3F	1	2×10^6	5
NEOSPLA3F	5	2×10^7	5

Table 1 (continued)

Plasmid	No. G418 Resistant Colonies	Frequency of G418 Resistant Colony per Transfected Cell
TCAE 5	16	1 in 20,000
ANEX 2	16	1 in 100,000
GKNEOSPLA3F	16	1 in 100,000
NEOSPLA3F	16	1 in 1,000,000

(Cells were fed with G418 containing media on days 2, 5, 7, 9, 12, 14, 18, 22, 26, 30 and 34; supernatant from colonies was assayed for immunoglobulin production and the colonies became confluent in the wells on days 18, 22, 26, 30 and 34).

Figures 7A to 7C provide histogram results and evidence the percentage of colonies at a particular level of expression.

The Examples provided herein are not to be construed as limited to the specific vectors, fully impaired consensus Kozak sequences, dominant selectable markers, transcriptional cassettes, and/or expressed proteins. The fully impaired consensus Kozak, and the utilization

This is the replacement sheet
Date: 14 MAR 1997

thereof, are not to be construed as limited to ANEX 1 and ANEX 2 vectors. Similarly, the preferred fully impaired consensus Kozak sequences and vectors in no way constitute an admission, either actual or implied, that these are the only sequences or vectors to which the inventor is entitled. The inventor is entitled to the full breadth of protection under applicable patent laws. Preferred vectors incorporating fully impaired consensus Kozak sequences have been identified by the inventor on ANEX 1 and ANEX 2 for purposes of claiming these vectors by designating plasmids comprising these vectors and anti-CD20 were deposited with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland, 20852, under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The plasmids were tested by the ATCC on November 9, 1992, and determined to be viable on that date. The ATCC has assigned these plasmids the following ATCC deposit numbers 69120 (anti-CD20 in TCAE 12(ANEX 1)) and 69118 (anti-CD20 in ANEX 2 (G1,K)); for purposes of this deposit, these plasmids were transformed into *E. coli*.

Although the invention has been described in considerable detail with regard to certain preferred embodiments thereof, other embodiments within the scope of the teachings of the present invention are possible. Accordingly, neither the disclosure nor the claims to follow, are intended, nor should be construed to be, limited by the descriptions of the preferred embodiments contained here.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Reff, Mitchell E.

(ii) TITLE OF INVENTION: Impaired Dominant Selectable Marker Sequence and Intronic Insertion Strategies for Enhancement of Expression of Gene Product and Expression Vector Systems Comprising Same

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDING ADDRESS:

(A) ADDRESSEE: IDEC Pharmaceuticals Corporation

(B) STREET: 11011 Torreyana Road

(C) CITY: San Diego

(D) STATE: California

(E) COUNTRY: USA

(F) ZIP: 92121

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb

(B) COMPUTER: Macintosh

(C) OPERATING SYSTEM: MS.DOS

(D) SOFTWARE: Microsoft Word 5.0

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Burgoon, Richard P. Jr.

(B) REGISTRATION NUMBER: 34,787

(C) REFERENCE/DOCKET NUMBER:

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (619) 550-8500

(B) TELEFAX: (619) 550 8750

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

This is the replacement sheet

Date: 14 MAR 1997

(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
TAG CTA GGT CCT ACC CC 17

(3) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
ATC GAT CCT GGA TGC GG

(4) INFORMATION FOR SEQ ID NO: 3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
TNN ATG CTT 9

(5) INFORMATION FOR SEQ ID NO: 4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single, including nick
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: (DNA (genomic)

This is the replacement sheet
Date: 14 MAR 1997

(iii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
CNN ATG CTT 9

(6) INFORMATION FOR SEQ ID NO: 5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
TNN ATG TTT 9

(7) INFORMATION FOR SEQ ID NO: 6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
CNN ATG TTT 9

(8) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:
TTG GGA GCT TGG ATC GAT 18
CCA CCA TGG TT 11

(9) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
TTG GGA GCT TGG ATC GAT 18
CCT CCA TGC TT 11

(10) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
CCA GCA TGG AGG AAT CGA TCC 21
TCC ATG CTT 9

(11) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 bases
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: yes
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
GGA GGA TCG ATT CCT CCA TGC TGG 24
CAC AAC TAT GTC AGA AGC AAA TGT GAG C 28

(12) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:

14 MAR 1997

(A) LENGTH: 18 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
CTG GGG CTC GAG CTT TGC 18

(13) INFORMATION FOR SEQ ED NO: 12
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: no62
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
GGT AAG TGC GGC CGC TAC TAA CTC TCT CCT 30
CCC TCC TTT TTC CTG GA 17

(14) INFORMATION FOR SEQ ID NO: 13
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 47 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(ii) HYPOTHETICAL: no
(iv) ANTI-SENSE: yes
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:
GGA AAA AGG AGG GAG GAG AGA GTT AGT AGC GGC 33
CGC ACT TAC CTG CA 14

(15) INFORMATION FOR SEQ ID NO: 14
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 bases
(B) TYPE: Nucleic acid
(C) STRANDEDNESS: single

This is the replacement sheet

Date: 14 MAR 1997

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCG ATT AAT TAA 12

(16) INFORMATION FOR SEQ ID NO: 15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCG AAG CGG CCG CT 14

(17) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16

GCA TGC GGT ACC GGA TCC ATC GAG CTA 27

CTA GCT TTG C 10

(18) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 bases

(B) TYPE: Nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ii) HYPOTHETICAL: no

(iv) ANTI-SENSE: no

This is the replacement sheet

Date: / /

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17

CTG ACT AGG CCT AGA GCG GCC GCA CTT ACC 30

TGC AGT TCA TCC AGG GC 17

14 MAR 1997

This Sheet Has Been Replaced

Claims

1. An expression vector for expressing a protein of interest by recombinant deoxyribonucleic acid techniques, said vector comprising at least one dominant selectable marker, wherein the translation initiation start site of said marker comprises the following sequence:

-3 +1

Pyxx ATG Pyxx

where "Py" is a pyrimidine nucleotide; "x" is a nucleotide; and the numerical designations are relative to the codon "ATG".

2. The expression vector of claim 1 wherein the nucleic acid sequence encoding for the protein of interest is co-linked to said dominant selectable marker.

3. The expression vector of claim 1 wherein said dominant selectable marker is selected from the group consisting of: herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

4. The expression vector of claim 1 wherein said translation initiation start site sequence is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that the codon "Txx" downstream of the ATG codon does not encode a stop codon.

5. The expression vector of claim 1 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

This is the replacement sheet

* C L M S *

Date: 14 MAR 1997

This Sheet Has Been Replaced

6. The expression vector of claim 1 wherein said translation initiation start site sequence is TCCATGCTT.

7. The expression vector of claim 1 wherein said translation initiation start site sequence is located within a secondary structure.

8. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

9. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

10. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

11. The expression vector of claims 8, 9 and 10 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

12. The expression vector of claim 10 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

* C L M S *

This Sheet Has Been Replaced

13. The expression vector of claims 8, 9, 10 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

14. A dominant selectable marker encoded by a nucleic acid sequence, wherein the translation initiation start site of said dominant selectable marker is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that "Txx" downstream of the ATG codon does not encode a stop codon.

15. The material of claim 14 wherein said dominant selectable marker is selected from the group consisting of herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, *Salmonella* his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

16. The material of claim 14 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

17. The material of claim 14 wherein said translation initiation start site sequence is TCCATGCTT.

18. The material of claim 14 wherein said translation initiation start site sequence is located within a secondary structure.

19. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

* C L M S *

14 MAR 1997

This Sheet Has Been Replaced



20. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

21. The material of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

22. The material of claims 19, 20 and 21 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

23. The material of claim 21 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

24. The material of claims 19, 20, 21 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

25. An expression vector selected from the group consisting of ANEX 1 (included within American Type Culture Collection deposit number 69120) and ANEX 2 (included within ATCC deposit number 69118).

26. A plasmid comprising the expression vector of claim 1 wherein the nucleic acid sequence encoding for said protein of interest is co-linked to said dominant selectable marker.

This is the replacement sheet



Date: 14 MAR 1997

27. The plasmid of claim 26 integrated within the cellular deoxyribonucleic acid of a mammalian host cell.

28. The plasmid of claim 27 wherein said host cell is selected from the group consisting of DG44, DXB11, CV1, COS, R1610, SP2/0, P3x633-Ag8.653, BFA-1c1BPT, RAJI, and 293.

29. The expression vector of claim 1 further comprising an artificial intronic insertion region within said dominant selectable marker, wherein an encoding sequence for a protein of interest is located within said insertion region.

30. The dominant selectable marker of claim 14 further comprising an artificial instronic insertion region.

This Sheet Has Been Replaced





* C L M S *

This is the replacement sheet

Date

ABSTRACT

IMPAIRED DOMINANT SELECTABLE MARKER SEQUENCE AND INTRONIC INSERTION STRATEGIES FOR ENHANCEMENT OF EXPRESSION OF GENE PRODUCT AND EXPRESSION VECTOR SYSTEMS COMPRISING SAME

Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector. These vectors are most typically utilized in the expression of proteins in mammalian expression systems. As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises sequence (1), where; "x" is a nucleotide selected from the group consisting of adenine (A), quanine (G), cytosine (C) or thymine (T)/uracil (U); "Py" is a pyrimidine nucleotide, ie C or TU; "ATG" is a codon encoding for the amino acid methionine, the so-called "start codon; and - 3 and + 1 are directional reference points vis-a-vis ATG. ie - 3 is meant to indicate three nucleotides upstream of ATG and + 1 is meant to indicate one nucleotide downstream of ATG. Dominant selectable markers further comprising artificial intronic insertion regions are further disclosed.

(Figure 1)

This Sheet Has Been Replaced



Neomycin phosphotransferase gene

		Cla I	-3		+1
TCAE 5.2	TTGGGAGCTTGG	ATCGAT	CC	A cc ATG	Gtt
				Met	Val
ANEX 1	TTGGGAGCTTGG	ATCGAT	CC	I cc ATG	Gtt
				Met	Leu
ANEX 2	CCaGCATGgAGGA	ATCGAT	CC	I cc ATG	Gtt
				Met	Leu

FIG. 1

This is the replacement sheet

Date: 14 MAR 1997

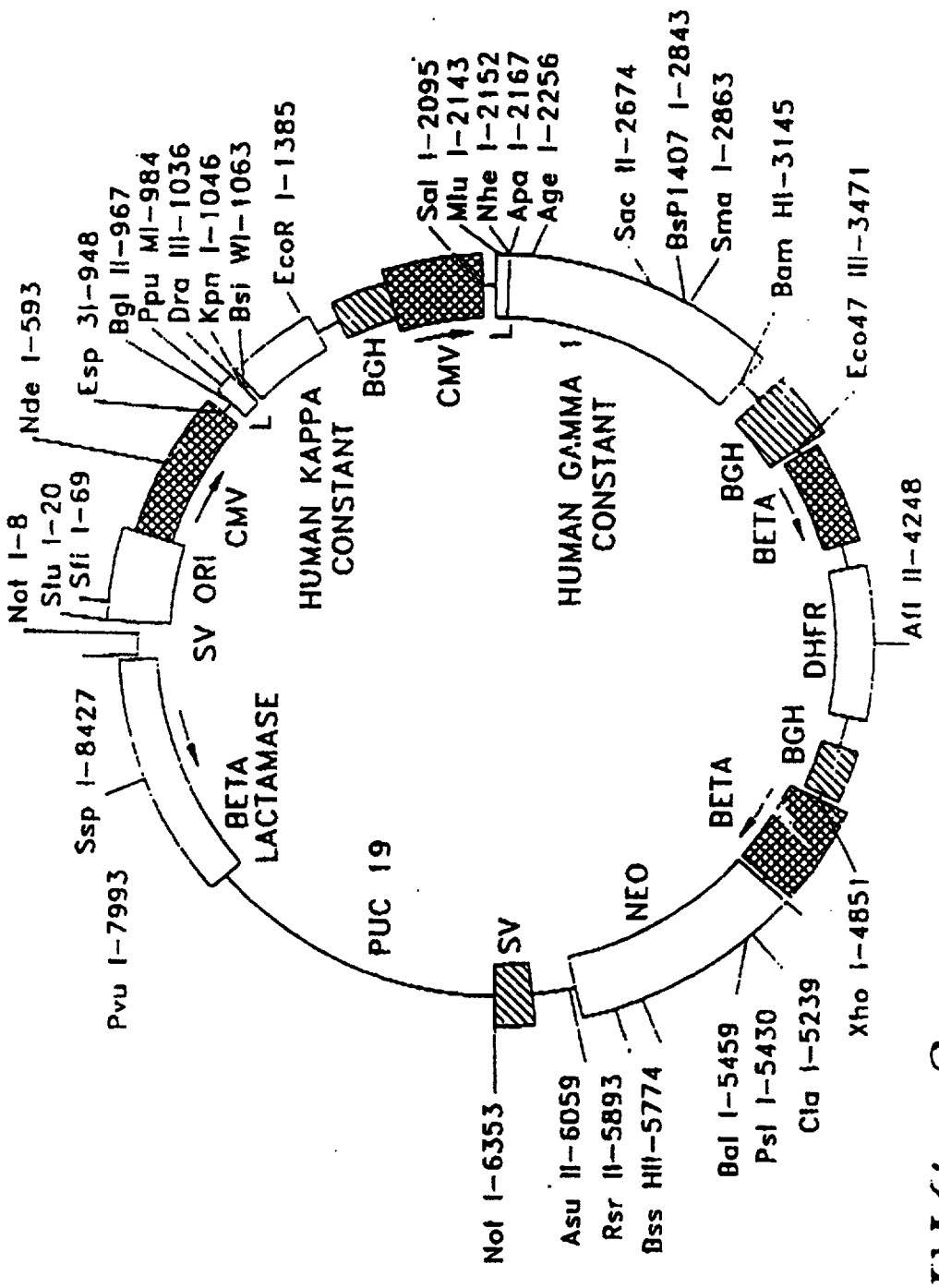


FIG. 2

This is the replacement sheet

Date: 14 MAR 1997

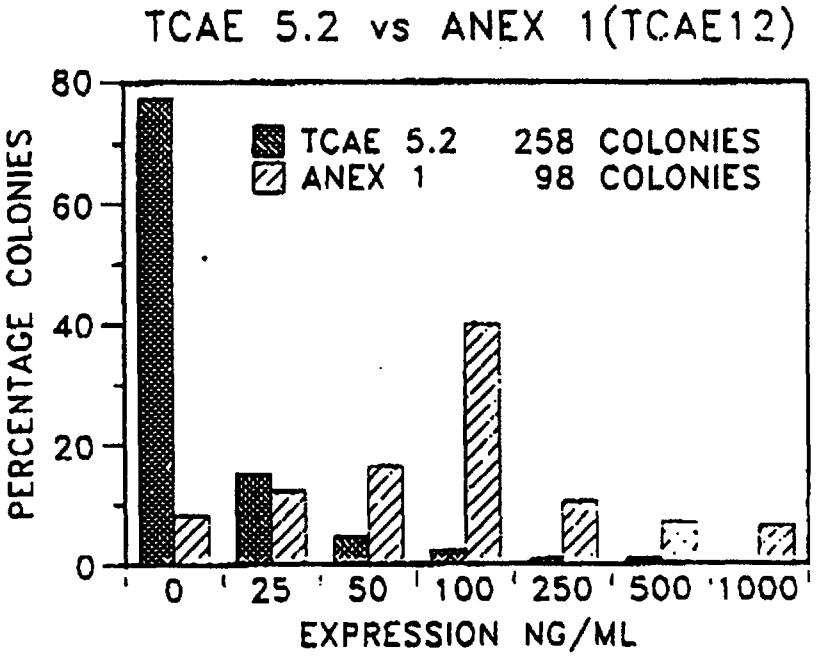


FIG. 3

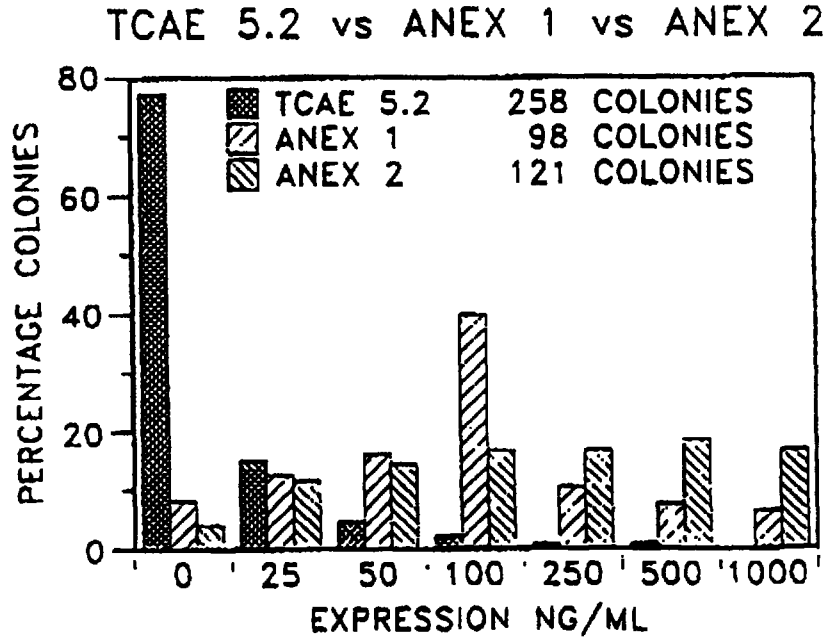


FIG. 5

This is the replacement sheet

Date:

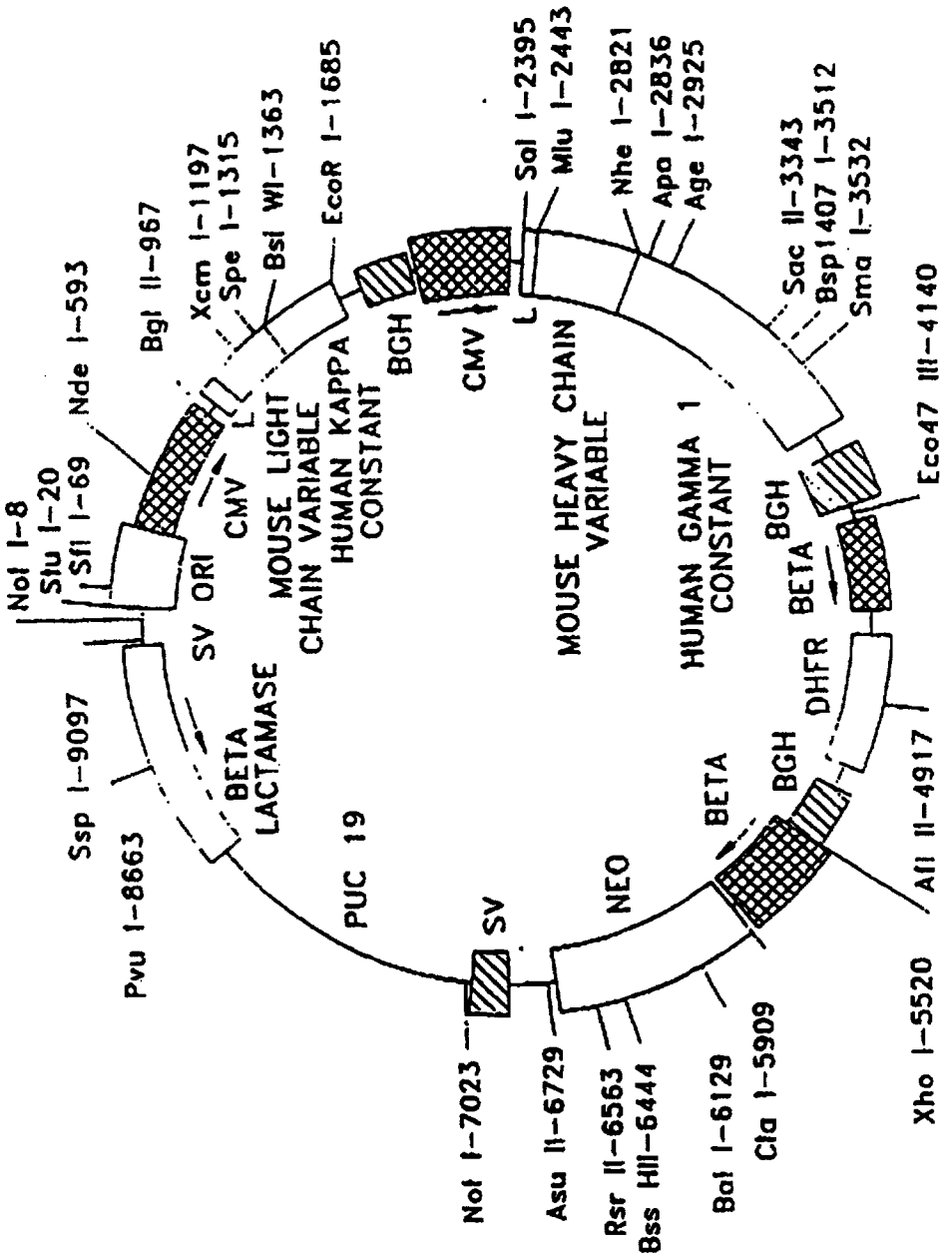


FIG. 4

This is an replacement sheet

Date: 14 MAR 1997

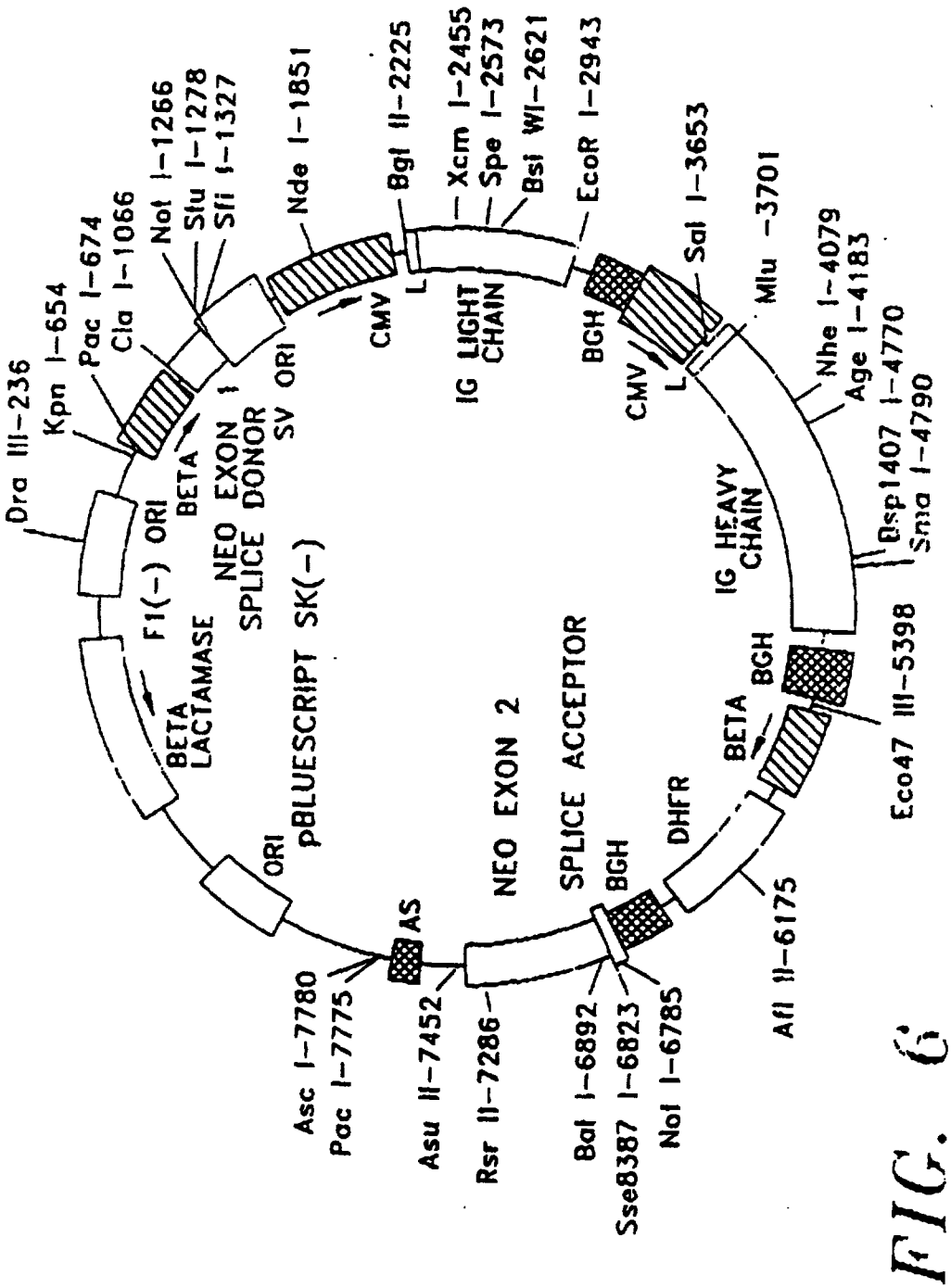


FIG. 6

14 MAR 1997

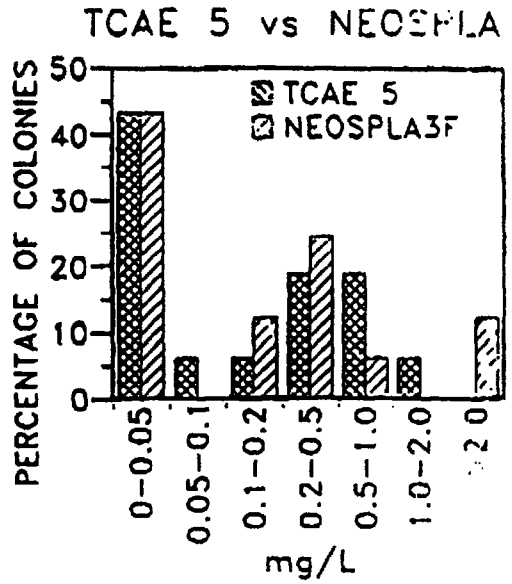


FIG. 7A

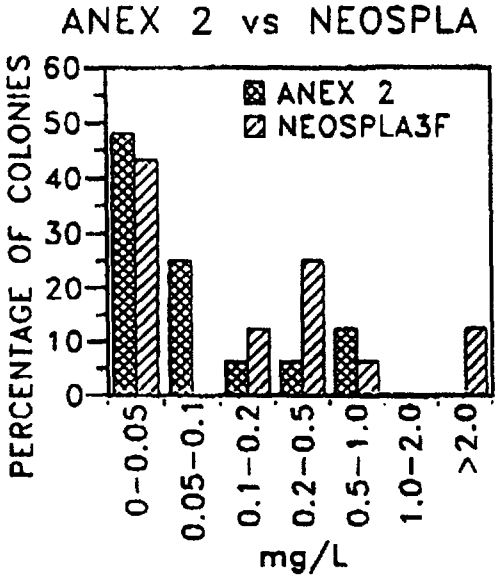


FIG. 7B

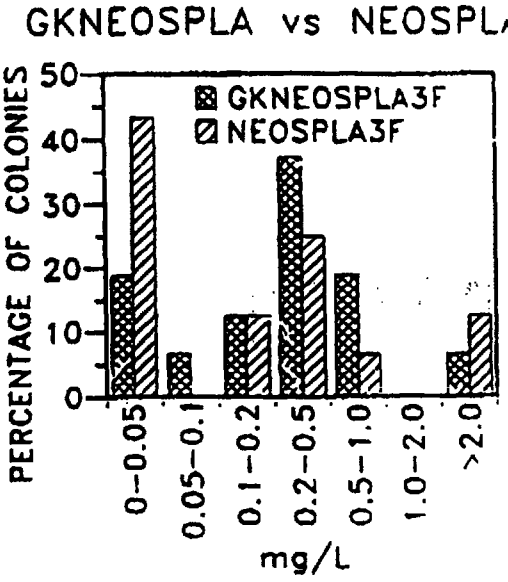


FIG. 7C

This is the replacement sheet

Date: 14 MAR 1997

What is claimed is: 14/5/99

1. An expression vector for expressing a protein of interest by recombinant deoxyribonucleic acid techniques, said vector comprising at least one dominant selectable marker, wherein the translation initiation start site of said marker comprises the following sequence:

-3 +1
Pyxx ATG Pyxx

where "Py" is a pyrimidine nucleotide; "x" is a nucleotide; and the numerical designations are relative to the codon "ATG".

2. The expression vector of claim 1 wherein the vector comprises a nucleic add sequence encoding the protein of interest is co-linked to said dominant selectable marker.

3. The expression vector of claim 1 wherein said dominant selectable marker is selected from the group consisting of: herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, Salmonella his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

4. The expression vector of claim 1 wherein said translation initiation start site sequence is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that the codon "Txx" downstream of the ATG codon does not encode a stop codon.

5. The expression vector of claim 1 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

* C L M S *

This is the replacement sheet

Date: 14 MAY 1999

6. The expression vector of claim 1 wherein said translation initiation start site sequence is TCCATGCTT.

7. The expression vector of claim 1 wherein said translation initiation start site sequence is located within a secondary structure.

8. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

9. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

10. The expression vector of claim 1 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

11. The expression vector of claims 8, 9 or 10 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

12. The expression vector of claim 10 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

This is the replacement sheet

* C L M S *

Date: 14 MAY

13. The expression vector of claims 8, 9 or 10 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

14. A dominant selectable marker encoded by a nucleic acid sequence, wherein the translation initiation start site of said dominant selectable marker is selected from the group consisting of TxxATGCxx; CxxATGCxx; CxxATGTxx; and TxxATGTxx, where "x" is a nucleotide, with the proviso that "Txx" downstream of the ATG codon does not encode a stop codon.

15. The marker of claim 14 wherein said dominant selectable marker is selected from the group consisting of herpes simplex virus thymidine kinase, adenosine deaminase, asparagine synthetase, Salmonella his D gene, xanthine guanine phosphoribosyl transferase, hygromycin B phosphotransferase, and neomycin phosphotransferase.

16. The marker of claim 14 wherein said translation initiation start site sequence is TxxATGCxx, where "x" is a nucleotide.

17. The marker of claim 14 wherein said translation initiation start site sequence is TCCATGCTT.

18. The marker of claim 14 wherein said translation initiation start site sequence is located within a secondary structure.

19. The marker of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 1000 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 1000 nucleotides.

* C L M S *

14 MAY 1999

20. The marker of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 350 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 350 nucleotides.

21. The marker of claim 14 wherein said translation initiation start site sequence further comprises at least one out-of-frame start codon within about 50 nucleotides of the ATG start codon of said start site, with the proviso that no in-frame stop codon is located within said 50 nucleotides.

22. The marker of claims 19, 20 and 21 wherein said out-of-frame start codon is part of a consensus Kozak sequence.

23. The marker of claim 21 wherein said out-of-frame start codon and said translation initiation start site sequence are both included as part of a secondary structure.

24. The marker of claims 19, 20, 21 wherein said translation initiation start site sequence is part of a secondary structure and said out-of-frame start codon is not part of said secondary structure.

25. An expression vector selected from the group consisting of ANEX 1 and ANEX 2.

26. A plasmid comprising the expression vector of claim 1 wherein the nucleic acid sequence encoding for said protein of interest is co-linked to said dominant selectable marker.

27. A mammalian host cell containing the plasmid of claim 26 wherein said plasmid is integrated

This is the replacement sheet

* C L M S *

Date: 14 MAY 1990

within the cellular deoxyribonucleic acid of said mammalian host cell.

28. The mammalian host cell of claim 27 wherein said mammalian host cell is selected from the group consisting of DG44, DXB11, CV1, COS, R1610, SP2/O, P3x633-Ag8.653, BPA-1c1BPT, RAJI, and 293.

29. The expression vector of claim 1 further comprising an artificial intronic insertion region within said dominant selectable marker, wherein an encoding sequence for a protein of interest is located within said insertion region.

30. The dominant selectable marker of claim 14 further comprising an artificial instronic insertion region.

31. An expression vector which expresses at least one protein of interest in a recombinant host cell wherein said expression vector comprises:

(i) a translationally impaired neomycin phosphotransferase (NEO) dominant selectable marker gene which has been translationally impaired by modification of the region of the NEO gene which includes the NEO translation initiation start codon such that said modified region of the NEO gene which includes the NEO translation initiation start codon has the following nucleotide sequence:

CCA GCA TGG AGG A ATCGAT CC TCC ATG CTT (SEQ ID NO: 17)

which translationally impaired NEO gene is operably linked to a promoter and polyadenylation sequence; and

(ii) at least one heterologous DNA which encodes for at least one protein of interest, wherein said heterologous DNA is operably linked to a promoter and polyadenylation sequence different from the promoter and

This is the replacement sheet

* C L M S *

Date:

polydenylation sequence operably linked to the NEO gene, and wherein said heterologous DNA and said promoter and polyadenylation sequence operably linked to said heterologous are inserted into an intronic insertion region contained in the NEO gene.

32. The vector of claim 31 which further comprises a second dominant selectable marker gene.

33. The vector of claim 31wherein the intronic insertion region is a synthetic intron sequence which is inserted between the CAG which encodes glycine at position 61 and the GAC which encodes aspartic acid at position 62 of the NEO gene.

34. The vector of claim 31 wherein the protein of interest is an antibody.

35. A recombinant host cell which contains the expression vector according to claim 31.

36. A recombinant host cell which contains the expression vector according to claim 32.

37. A recombinant host cell which contains the expression vector according to claim 33.

38. A recombinant host cell which contains the expression vector according to claim 34.

39. The recombinant host cell of claim 35 wherein said host cell is a mammalian cell.

40. The recombinant host cell of claim 39 wherein said mammalian cell is a Chinese hamster ovary cell.

This is the replacement sheet

* C L M S *

Date 14 MAY 1999

ABSTRACT

IMPAIRED DOMINANT SELECTABLE MARKER SEQUENCE AND INTRONIC INSERTION STRATEGIES FOR ENHANCEMENT OF EXPRESSION OF GENE PRODUCT AND EXPRESSION VECTOR SYSTEMS COMPRISING SAME

Disclosed herein are fully impaired consensus Kozak sequences which are most typically used with dominant selectable markers of transcriptional cassettes which are a part of an expression vector. These vectors are most typically utilized in the expression of proteins in mammalian expression systems. As defined, disclosed and claimed herein, a "fully impaired consensus Kozak" comprises sequence (1), where; "x" is a nucleotide selected from the group consisting of adenine (A), quanine (G), cytosine (C) or thymine (T)/uracil (U); "Py" is a pyrimidine nucleotide, ie C or TU; "ATG" is a codon encoding for the amino acid methionine, the so-called "start codon; and - 3 and + 1 are directional reference points vis-a-vis ATG. ie - 3 is meant to indicate three nucleotides upstream of ATG and + 1 is meant to indicate one nucleotide downstream of ATG. Dominant selectable markers further comprising artificial intronic insertion regions are further disclosed.

(Figure 1)

This is the replacement sheet

Date: 14 MAY 1999