

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 May 2006 (11.05.2006)

PCT

(10) International Publication Number
WO 2006/048657 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/GB2005/004257

(22) International Filing Date:
4 November 2005 (04.11.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0424494.3 5 November 2004 (05.11.2004) GB

(71) Applicant (for all designated States except US): **UNIVERSITY OF SHEFFIELD** [GB/GB]; Western Bank, Sheffield S10 2TN (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **QWARNSTROM, Eva** [SE/GB]; Hopwood House, High Moscar, Sheffield S6 6GR (GB).

(74) Agent: **HARRISON GODDARD FOOTE**; 31 St. Saviourgate, York YO1 8NQ (GB).

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

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

Published:

— without international search report and to be republished upon receipt of that report

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

(54) Title: THERAPEUTIC METHOD

(57) Abstract: The present invention relates to the use of an isolated nucleic acid sequence encoding a cell surface protein in the treatment of, for example by way of gene therapy, disorders associated with impaired cell surface interactions and wound healing.



WO 2006/048657 A2

THERAPEUTIC METHOD

The invention relates to the use of a nucleic acid sequence encoding a cell surface protein in the treatment of, *inter alia*, disorders associated with impaired cell surface interactions including wound healing.

The *Fras 1* gene encodes one member of a family of putative extracellular matrix proteins (McGregor et al., Nature Genetics, 34:203-208 (2003)). The Fras1 protein has been implicated in epithelial basement membrane and cell/matrix interactions with a role in tissue structure development and signal propagation in the extra cellular matrix. Mutations in the *Fras 1* gene has been found to be associated with Fraser syndrome (McGregor et al., Nature Genetics, 34:203-208 (2003)). Fraser syndrome, is a malformation condition which usually includes cryptophthalmus, syndactyly and renal defects and is believed to be caused by disrupted epithelial integrity *in utero* (McGregor et al., Nature Genetics, 34:203-208 (2003)).

The present inventors have isolated a human DNA sequence encoding a Toll-IL1 receptor regulator (TILRR) protein. Surprisingly, the TILRR protein shares considerable homology with the protein FRAS 1, and thus constitutes a FRAS related protein.

According to a first aspect of the present invention there is provided a purified and isolated nucleic acid molecule encoding a cell surface polypeptide selected from the group consisting of:

- i) a polypeptide, or fragment or variant thereof, encoded by a nucleic acid molecule consisting of a nucleic acid sequence as represented by Figure 1;
- ii) a polypeptide encoded by a nucleic acid molecule which hybridises to a nucleic acid molecule as defined in (i) above; and
- iii) a polypeptide comprising a nucleic acid which is degenerate as a result of the genetic code to the nucleic acid sequence defined in (i) and (ii)

for use as a medicament.

The expression “cell surface polypeptide” is intended to include a FRAS related protein such as the Toll-IL1 receptor regulator (TILRR). The TILRR has been shown to regulate Toll-IL-1 receptor function in response to IL-1 and/or pathogens. A “Toll-IL-1 receptor” is any receptor having a TIR domain and includes Toll-like receptors and IL-1 receptor.

As used herein “a nucleic acid sequence as represented by Figure 1” includes a nucleic acid sequence represented by Figure 1A or Figure 1B. Preferably, the nucleic acid sequence as represented by Figure 1 is that shown in Figure 1B.

The nucleic acid molecule of the first aspect of the invention may anneal under stringent hybridisation conditions to the nucleic acid sequence shown in Figure 1 or to its complementary strand.

Stringent hybridisation/washing conditions are well known in the art. For example, nucleic acid hybrids that are stable after washing in 0.1xSSC, 0.1% SDS at 60°C. It is well known in the art that optimal hybridisation conditions can be calculated if the sequences of the nucleic acid is known. For example, hybridisation conditions can be determined by the GC content of the nucleic acid subject to hybridisation. Please see Sambrook *et al* (1989) Molecular Cloning; A Laboratory Approach. A common formula for calculating the stringency conditions required to achieve hybridisation between nucleic acid molecules of a specified homology is:

$$T_m = 81.5^{\circ} \text{C} + 16.6 \text{ Log } [\text{Na}^+] + 0.41[\% \text{ G} + \text{C}] - 0.63 (\% \text{formamide}).$$

The nucleic acid molecule of the first aspect of the invention may comprise the sequence set out in Figure 1 or a sequence which is at least 25%, 30%, 40%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, for example 98%, or 99%, identical to the nucleic acid sequence set out in Figure 1 at the nucleic acid residue level.

“Identity”, as known in the art, is the relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the

sequences. In the art, identity also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Identity can be readily calculated (*Computational Molecular Biology*, Lesk, A.M. ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., AND Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or two polypeptide sequences, the term is well-known to skilled artisans (*Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods commonly employed to determine identity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403 (1990)).

The nucleic acid molecule of the first aspect of the invention may comprise a fragment of a sequence according to the first aspect which is at least 30 bases long, for example, 40, 50, 60, 70, 80 or 90 bases in length.

Nucleic acid molecules according to the first aspect of the invention may, for use in gene therapy, be provided alone or as part of a vector, such as an expression vector, examples of which are well known in the art.

Thus according to a second aspect of the invention there is provided a vector comprising a nucleic acid molecule according to the first aspect of the invention.

Preferably, in gene therapy, the nucleic acid molecule is administered such that it is expressed in the subject to be treated for example in the form of a recombinant DNA molecule comprising the nucleic acid molecule of the invention operatively linked to a nucleic acid sequence which controls expression, such as in an expression vector. Expression vectors include chromosomal, episomal and virus-derived vectors e.g. vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements such as cosmids and phagemids. Such a vector will thus include appropriate transcription control signals including a promoter region capable of expressing the coding sequence, said promoter being operable in the subject to be treated. These promoter sequences may be cell/tissue specific, inducible or constitutive. Thus for human gene therapy, the promoter is preferably a human promoter sequence from a human gene, or from a gene which is typically expressed in humans such as the promoter from human cytomegalovirus (CMV). Among other known promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs such as those of the Rous sarcoma virus and metallothionein promoters such as the mouse metallothionein-1 promoter. Moreover, the native TILRR promoter (Figure 3) may be used.

The vector according to the invention may also include transcriptional control signals and also polyadenylation signals recognizable in the subject to be treated such as, for example, the corresponding sequences from viruses such as, for human treatment, the SV40 virus. Expression vectors may also include selectable markers, such as antibiotic resistance, which enable the vectors to be propagated. Expression vectors capable in situ of synthesizing the cell surface polypeptide according to the invention may be introduced

directly into a wound site by physical methods. Examples of these include topical application of the 'naked' nucleic acid vector in an appropriate vehicle for example in a pharmaceutically acceptable excipient such as phosphate buffered saline (PBS), or administration of the vector by physical methods such as particle bombardment, also known as 'gene gun' technology, according to methods known in the art e.g. as described in US5371015 in which inert particles, such as gold beads coated with the vector are accelerated at speeds sufficient to enable them to penetrate the surface at the wound site e.g. skin cells by means of discharge under high pressure from a projecting device (particles coated with a nucleic acid molecule of the present invention are within the scope of the invention as are devices comprising such particles).

Physical methods of administering the nucleic acid molecule, or vector, of the invention, directly to a recipient include particle bombardment, ultrasound, electrical stimulation, electroporation and microseeding. The microseeding mode of delivery is a system for delivering genetic material into cells *in situ* in a patient and is described in US5,697,901.

The nucleic acid molecule of the invention for use in the therapy of the invention may also be administered by means of delivery vectors. A number of viruses are commonly used as vectors for the delivery of exogenous genes. Commonly employed vectors include recombinantly modified enveloped or non-enveloped DNA and RNA viruses, preferably selected from baculoviridae, parvoviridae, picornoviridae, herpesviridae, poxviridae, adenoviridae, or picornaviridae. Chimeric vectors may also be employed which exploit advantageous elements of each of the parent vector properties (See e.g., Feng, et al (1997) Nature Biotechnology 15:866-870). Such viral vectors may be wildtype or modified by recombinant DNA techniques to be replication deficient, conditionally replicating or replication competent. Preferred vectors are derived from the adenoviral, adeno-associated viral and retroviral genomes.

Other non-viral delivery vectors include lipid delivery vectors, including liposome delivery vehicles known in the art.

Typically, when using adenovirus-based vectors for gene therapy, the virus has to be modified to eliminate or minimise the disease-causing potential by rendering the virus replication-deficient. Typically, such a modification involves the deletion of the E1 region genes. Thus, in a further preferred embodiment of the invention the said adenovirus is made replication-deficient, preferably the adenovirus is E1 negative.

In addition, the adenovirus virus vector may harbour deletions within the E3 region or may be deficient in one or more E3 functions. Moreover, certain E3 genes, individual or as a whole, may be replaced by other "therapeutic" genes, including genes encoding antigenic proteins for vaccination.

A polypeptide encoded by a nucleic acid molecule according to the invention may also be administered to a wound site by means of transformed host cells. Such cells include cells harvested from the subject, into which the nucleic acid molecule is introduced by gene transfer methods known in the art, followed by growth of the transformed cells in culture and grafting to the subject.

Expression constructs such as those described herein may be used in a variety of ways in the therapy of the present invention. They may be directly administered to a wound site or they may be used to prepare recombinant protein encoded by the nucleic acid molecule of the invention which can then be administered to a wound site. The invention also relates to host cells which are genetically engineered with constructs which comprise the nucleic acid molecules of the present invention and to the uses of these vectors and cells in the therapeutic methods discussed herein. Representative examples of host cells for recombinant expression of a polypeptide according to the invention include bacterial cells such as streptococci, staphylococci, E.coli, streptomyces and Bacillus subtilis; fungal cells such as yeast cells and Aspergillus; insect cells; animal cells such as CHO, COS, HeLa and Bowes melanoma cells; and plant cells.

If a protein is being utilised for therapeutic purposes it is often desirable to be able to confirm and visualise its expression. This is typically achieved by the use of protein tags. The DNA sequence that codes for the therapeutic protein is tagged by fusing it to the

sequence of another protein that can be easily detected. When the organism expresses the therapeutic protein, the protein "tags" are also produced.

Proteinaceous fluorophores are known in the art. Green fluorescent protein, GFP, is a spontaneously fluorescent protein isolated from coelenterates, such as the Pacific jellyfish, *Aequoria victoria*. Its role is to transduce, by energy transfer, the blue chemiluminescence of another protein, aequorin, into green fluorescent light. GFP can function as a protein tag, as it tolerates N- and C-terminal fusions to a broad variety of proteins many of which have been shown to retain native function. Other proteinaceous fluorophores include yellow, red and blue fluorescent proteins. The tag may be part of the delivery vector e.g virus.

According to a third aspect of the invention there is provided an isolated and purified cell surface polypeptide, or fragment or variant thereof, selected from the group consisting of:

- i) a polypeptide, or fragment or variant thereof, encoded by a nucleic acid molecule consisting of a nucleic acid sequence as represented by Figure 1;
- ii) a polypeptide encoded by a nucleic acid molecule which hybridises to a nucleic acid molecule as defined in (i) above; and
- iii) a polypeptide comprising a nucleic acid which is degenerate as a result of the genetic code to the nucleic acid sequence defined in (i) and (ii)

for use as a medicament.

In a preferred aspect of the invention, said polypeptide comprises an amino acid sequence as represented in Figure 2. As used herein "an amino acid sequence as represented by Figure 2" includes an amino acid sequence represented by Figure 2A or Figure 2B. Preferably, the amino acid sequence as represented in Figure 2 is that shown in Figure 2B.

In a further preferred aspect of the invention, said polypeptide consists of an amino acid sequence selected from the group consisting of amino acid residues 280 to 380 of the amino acid sequence presented in Figure 2, or amino acid residues 600 to 716 of the

amino acid sequence presented in Figure 2, amino acid residues 4 - 9 of the amino acid sequence presented in Figure 2 or amino acid residues 77 and/or 102 and/or 219 and/or 245 and/or 398 and/or 477 and/or 548 of the amino acid sequence presented in Figure 2, or a variant polypeptide wherein said variant polypeptide sequence has been altered by addition, substitution or deletion of at least one amino acid residue.

As used herein, the term "polypeptide" means, in general terms, a plurality of amino acid residues joined together by peptide bonds. It is used interchangeably and means the same as peptide, protein, oligopeptide, or oligomer. The term "polypeptide" is also intended to include fragments, variants, analogues and derivatives of a polypeptide wherein the fragment, variant, analogue or derivative retains essentially the same biological activity or function as a reference protein, typically a FRAS related protein such as TILRR. As used herein, the term "cell surface polypeptide" includes naturally and recombinantly produced polypeptide including synthetic.

As used herein "fragment" may include a polypeptide fragment which may be at least 10, 15, 20, 30, 50 or 100 amino acids long.

A variant polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, truncations which may be present in any combination. Among preferred variants are those that vary from a reference polypeptide by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid by another amino acid of like characteristics. The following non-limiting list of amino acids are considered conservative replacements (similar): a) alanine, serine, and threonine; b) glutamic acid and aspartic acid; c) asparagine and glutamine d) arginine and lysine; e) isoleucine, leucine, methionine and valine and f) phenylalanine, tyrosine and tryptophan.

The nucleic acid molecules and polypeptides useful in the present invention are preferably provided in isolated form, and preferably are purified to homogeneity.

In a yet further aspect, the invention provides a pharmaceutical composition comprising a nucleic acid molecule according to the first aspect of the invention, or a polypeptide according to the third aspect of the invention, together with one or more pharmaceutically acceptable carriers.

When administered, the pharmaceutical compositions of the present invention are administered in pharmaceutically acceptable preparations. Such preparations may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, supplementary immune potentiating agents such as adjuvants and cytokines and optionally other therapeutic agents.

The compositions of the invention can be administered by any conventional route, example, be topical, oral, intravenous, intraperitoneal, intramuscular, intracavity, subcutaneous, or transdermal, intranasal, intracerebral or epidural. Compositions formulated for topical administration may take the form of ointments, creams, lotions, eye ointments, eye drops, impregnated dressings and sutures and aerosols.

The compositions of the invention are administered in effective amounts. An "effective amount" is the amount of a composition that alone, or together with further doses, produces the desired response. In the case of treating a particular disease, the desired response is inhibiting the progression of the disease. Such amounts will depend, of course, on the particular condition being treated, the severity of the condition, the individual patient parameters including age, physical condition, size and weight, the duration of the treatment, the nature of concurrent therapy (if any), the specific route of administration and like factors within the knowledge and expertise of the health practitioner. These factors are well known to those of ordinary skill in the art and can be addressed with no more than routine experimentation. It is generally preferred that a maximum dose of the individual components or combinations thereof be used, that is, the highest safe dose according to sound medical judgment.

The compositions used in the foregoing methods preferably are sterile and contain an effective amount of the active ingredient for producing the desired response in a unit of weight or volume suitable for administration to a patient. The response can, for example, be measured by measuring the physiological effects of the composition, such as decrease of disease symptoms etc. Other assays will be known to one of ordinary skill in the art and can be employed for measuring the level of the response.

The administration of the compositions to mammals, other than humans (e.g. for testing purposes or veterinary therapeutic purposes), is carried out under substantially the same conditions as described above. A subject, as used herein, is a mammal, preferably a human, and includes a non-human primate, cow, horse, pig, sheep, goat, dog, cat or rodent.

When administered, the pharmaceutical compositions of the invention are applied in pharmaceutically acceptable amounts and in pharmaceutically acceptable compositions. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. Such preparations may routinely contain salts, buffering agents, preservatives, compatible carriers, and optionally other therapeutic agents. When used in medicine, the salts should be pharmaceutically acceptable, but non-pharmaceutically acceptable salts may conveniently be used to prepare pharmaceutically acceptable salts thereof and are not excluded from the scope of the invention. Such pharmacologically and pharmaceutically-acceptable salts include, but are not limited to, those prepared from the following acids: hydrochloric, hydrobromic, sulfuric, nitric, phosphoric, maleic, acetic, salicylic, citric, formic, malonic, succinic, and the like. Also, pharmaceutically acceptable salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts.

The compositions may be combined, if desired, with a pharmaceutically acceptable carrier. The term "pharmaceutically-acceptable carrier" as used herein means one or more compatible solid or liquid fillers, diluents or encapsulating substances that are

suitable for administration into a human. The term "carrier" denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. The pharmaceutical compositions may contain suitable buffering agents, including: acetic acid in a salt; citric acid in a salt; boric acid in a salt; and phosphoric acid in a salt. The pharmaceutical compositions also may contain, optionally, suitable preservatives, such as: benzalkonium chloride; chlorobutanol; parabens and thimerosal.

Compositions suitable for parenteral administration conveniently comprise a sterile aqueous or non-aqueous preparation of protein, which is preferably isotonic with the blood of the recipient. This preparation may be formulated according to known methods using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation also may be a sterile injectable solution or suspension in a non-toxic parenterally-acceptable diluent or solvent, for example, as a solution in 1,3-butane diol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil may be employed including synthetic mono- or di-glycerides. In addition, fatty acids such as oleic acid may be used in the preparation of injectables. Carrier formulation suitable for topical, oral, subcutaneous, intravenous, intramuscular, etc. administrations can be found in Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA.

In a further aspect of the invention there is provided the use of a nucleic acid molecule according to the first aspect of the invention, or a polypeptide encoded thereby, in the manufacture of a medicament for the treatment or prevention of a disorder caused by disrupted epithelial integrity.

As used herein conditions associated with "disrupted epithelial integrity" relate to conditions in which cell-matrix or cell-basement interactions are impaired resulting, for example, in disrupted cell adhesion.

Alterations or disruptions in epithelial integrity are likely to involve changes in epithelial/basement and/or extracellular matrix interaction. These interactions are fundamental for normal tissue and organ function and development. Alterations in factors regulating epithelial development and protein associations involved underlie a wide variety of diseases such as sclerosing hemangioma (PSH), and those caused by changes in lung and kidney development, specifically in early embryonic stages.

The disruption to epithelial integrity may occur during embryo development, for example *in utero*.

The disorder may be a developmental disorder such as Fraser syndrome.

In a preferred method of the invention said subject is human. The method of treatment may be carried out on an embryo such as *in utero* or on a child or adult patient.

In a further aspect the invention provides a method of treatment or prevention of a condition associated with disrupted epithelial integrity in a patient comprising the step of administering to the patient a therapeutically effective amount of a nucleic acid molecule according to the first aspect of the invention.

A nucleic acid molecule according to the first aspect of the invention may be used therapeutically in the method of treatment by way of gene therapy in which the nucleic acid molecule is administered to tissues in a form in which it is capable of directing the production of the cell surface protein, or a biologically active fragment thereof, *in situ*.

In a yet further aspect of the invention there is provided the use of a nucleic acid molecule according to the first aspect of the invention, or a polypeptide encoded by said nucleic acid sequence, in the manufacture of a medicament for the treatment of wounds.

A potential difficulty associated with any treatment for burns is in promoting adherence of the epithelial layers of the skin (dermis and epidermis) to the underlying connective tissue. Current treatments involve the use of collagen clips to achieve this adherence. The present invention thus provides a means for enhancing cell surface interactions, for example, enhancing the interaction between the epithelial layer of the skin and the underlying connective tissue.

In a further aspect the invention provides a method of treatment of wounds in an animal, including human, comprising the step of administering to said animal a therapeutically effective amount of a pharmaceutical composition according to the invention.

As used herein "treatment of a wound" may include wound healing and associated conditions and therapy which promotes, augments, or accelerates healing of tissues and includes treatment of post-operative scarring, burns, psoriasis, ulcers, limb ulcerations in diabetes, acceleration of tissue remodelling, for example, post cosmetic surgery and organ transplantation, for example skin grafts, acceleration of bone repair and the promotion of angiogenesis and of re-endothelialisation following percutaneous transluminal coronary angioplasty or any postoperative or medical condition relying on any such optimally regulated healing process.

As indicated above, the cell surface protein may be administered in a composition as a polypeptide encoding nucleic acid which is transcribed and translated *in situ* in the form of gene therapy, or the polypeptide itself may be directly administered, for example at or close to a wound site.

The invention further provides a substrate to which a polypeptide, or composition, according to the invention is applied or attached. Preferably, the substrate is suitable for application to wounds or delivery to wound sites. Preferably, the substrate allows for the transfer of the polypeptides according to the invention from the substrate to a wound bed to achieve their healing effect. The substrate may be a dressing, for example, wound

dressing. The dressing may comprise a fabric material or it may be a collagen-like material.

A further aspect of the invention provides a method of diagnosis comprising using a nucleic acid molecule according to the first aspect of the invention as a probe to determine the presence or absence of a genetic defect in a patient. The method of diagnosis may be carried out on an embryo such as *in utero* or on a child or adult patient. The genetic defect may be a mutation in a gene encoding an extracellular matrix protein, similar to for example, a defect/mutation in the gene *Fras 1* or a gene encoding a related protein.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

Throughout the description and claims of this specification, the words “comprise” and “contain” and variations of the words, for example “comprising” and “comprises”, mean “including but not limited to”, and are not intended to (and do not) exclude other moieties, additives, components, integers or steps.

The invention will now be described, by way of example only, with reference to the following figures:

Figure 1 (A) Partial cDNA sequence of human TILRR; (B) Full-length cDNA sequence of human TILRR

Figure 2 (A) Amino acid sequence encoded by the cDNA sequence of Figure 1(A); (B) Amino acid sequence encoded by the cDNA sequence of Figure 1(B);

Figure 3 Promoter sequence of human cell surface protein according to the invention;

Figure 4 Features of the amino-acid sequence of the cell surface protein according to the invention;

EXAMPLES

Characterisation of the cell surface protein

MALDI TOF and Mass spec Analysis:

Preparation of Samples

For sequence analysis, the high molecular weight component of about 300-350kDa (Valles et al., (1999) J. Biol. Chem. 274, 20103-09; Valles et al., (2002) Lab. Invest.; 82: 855-62) from immunoprecipitated samples separated by SDS-gel electrophoresis were cut out and subjected to cysteine alkylation by Tributylphosphine (reduction.T 7567) and Iodoacetamide (alkylation A 3221) using the Sigma ProteoPrep Reduction and Alkylation Kit (Cat. No. PROTRA) using siliconised tubes (Sigma, T4691-500EA). Subsequently, samples were subjected to Trypsin digest using the Trypsin Profile IGD Kit from Sigma (Cat. No. PP0100), and according to manufacturers instructions.

MS Analysis

These samples were subsequently subjected to Surface Enhanced Laser Desorption Ionisation (SELDI) / Matrix Assisted Laser Desorption Ionisation (MALDI) as described (Ashcroft *Nat.Prod. Rep.*, **20**, 202-215).- Briefly, an aliquot (1 μ L) of the sample in solution (1:1 v/v aqueous methanol, with 0.1% trifluoroacetic acid) was placed on the sample target and treated with an aliquot (1 μ L) of the matrix, α -cyano-4-hydroxy cinnamic acid (CHCA), in solution (4.5 mg/ml in 1:1 v/v aqueous acetonitrile, with 0.05% trifluoroacetic acid). The mixture was

allowed to dry at room temperature and then analysed by MALDI-MS on the SELDI ProteinChip mass spectrometer (Ciphergen, USA; purchased with funds from the Wellcome Trust) using a laser intensity of 190 and a detector sensitivity of 9. Approximately 100 laser shots were accumulated and an external calibration applied to the resulting spectrum using ions from a standard peptide mixture (Arg-8-vasopressin, 1084.2 Da; somatostatin, 1637.9 Da; bovine insulin B chain, 3495.9 Da; human insulin, 5807.7 Da; hiruden, 7033.6 Da). A mass accuracy of ± 1 Da is expected.

Sequence analysis:

Peptide maps were analysed using peptide mass map fingerprinting programmes ProFound at PROWL site (prowl.rockefeller.edu/cgi-bin/ProFound) and the MASCOT, MOWSE and PeptIdent sites. Further, database (www.ensembl.org) and BLAST searching of the obtained sequence (shown in Figure 1) revealed a human protein which is the product of a unique gene, which GeneScan predictions suggests gives rise to several alternately spliced mRNAs, encoding polypeptides of the sizes reported by Valles et al., (1999) J. Biol. Chem. 274, 20103-09.

CLAIMS

1. A purified and isolated nucleic acid molecule encoding a cell surface polypeptide selected from the group consisting of:

- i) a polypeptide, or fragment or variant thereof, encoded by a nucleic acid molecule consisting of a nucleic acid sequence as represented by Figure 1;
- ii) a polypeptide encoded by a nucleic acid molecule which hybridises to a nucleic acid molecule as defined in (i) above; and
- iii) a polypeptide comprising a nucleic acid sequence which is degenerate as a result of the genetic code to the nucleic acid sequence defined in (i) and (ii)

for use as a medicament.

2. A nucleic acid molecule as claimed in claim 1 wherein the nucleic acid sequence is represented by Figure 1B.

3. A vector comprising a nucleic acid molecule as claimed in claim 1 or 2.

4. A vector as claimed in claim 3 wherein the vector is an expression vector

5. A vector as claimed in claim 3 wherein the vector is a delivery vector.

6. A vector as claimed in claim 5 wherein the delivery vector is a virus.

7. A vector as claimed in claim 6 wherein the virus is a DNA or RNA virus.

8. A vector as claimed in claim 7 wherein the virus is selected from the group consisting of baculoviridae, parvoviridae, picornaviridae, herpesviridae, poxviridae, adenoviridae, or picornaviridae.

9. A vector as claimed in claim 8 wherein the vector is a chimeric vector.

10. A vector as claimed in claim 9 wherein the virus is an adenovirus, adeno-associated virus and retrovirus.
11. A vector as claimed in claim 10 wherein the adenovirus is replication-deficient.
12. A vector as claimed in claim 11 wherein the adenovirus is E1 negative.
13. A vector as claimed in any preceding claim wherein the vector is provided with a detectable label or tag.
14. An isolated and purified cell surface polypeptide selected from the group consisting of:
 - i) a polypeptide, or fragment or variant thereof, encoded by a nucleic acid molecule consisting of a nucleic acid sequence as represented by Figure 1;
 - ii) a polypeptide encoded by a nucleic acid molecule which hybridises to a nucleic acid molecule as defined in (i) above; and
 - iii) a polypeptide comprising a nucleic acid which is degenerate as a result of the genetic code to the nucleic acid sequence defined in (i) and (ii) for use as a medicament.
15. A polypeptide as claimed in claim 14 wherein the polypeptide comprises an amino acid sequence as represented in Figure 2 or is a variant polypeptide thereof wherein said variant polypeptide sequence has been altered by addition, substitution or deletion of at least one amino acid residue.
16. A polypeptide as claimed in claim 15 wherein the polypeptide comprises the amino acid sequence shown in Figure 2B.
17. A polypeptide as claimed in claim 14 wherein said polypeptide consists of an amino acid sequence selected from the group consisting of amino acid residues 280 to

380 of the amino acid sequence presented in Figure 2, or amino acid residues 600 to 716 of the amino acid sequence presented in Figure 2, amino acid residues 4 - 9 of the amino acid sequence presented in Figure 2 or amino acid residues 77 and/or 102 and/or 219 and/or 245 and/or 398 and/or 477 and/or 548 of the amino acid sequence presented in Figure 2, or a variant polypeptide wherein said variant polypeptide sequence has been altered by addition, substitution or deletion of at least one amino acid residue.

18. A polypeptide as claimed in any of claims 14 to 17 wherein the polypeptide is provided with a radioactive and/or fluorescent and/or label or tag.

19. A pharmaceutical composition comprising a nucleic acid molecule as claimed in claim 1 or 2, a vector as claimed in any of claims 3 to 13 or a polypeptide as claimed in any of claims 14 to 18, together with one or more pharmaceutically acceptable carriers.

20. Use of a nucleic acid molecule as claimed in claim 1 or 2, or a polypeptide encoded thereby, in the manufacture of a medicament for the treatment or prevention of a disorder caused by disrupted epithelial integrity.

21. Use as claimed in claim 20 wherein the disorder is a developmental disorder.

22. Use as claimed in claim 21 wherein the disorder is Fraser syndrome.

23. A method of treatment or prevention of a condition associated with disrupted epithelial integrity in a patient comprising the step of administering to the patient a therapeutically effective amount of a nucleic acid molecule as claimed in claim 1 or 2 or a vector as claimed in any of claims 3 to 13.

24. A method as claimed in claim 23 wherein the method of treatment is by way of gene therapy in which the nucleic acid molecule is administered to tissues in a form in which it is capable of directing the production of the cell surface protein, or a biologically active fragment thereof, in situ.

25. Use of a nucleic acid molecule as claimed in claim 1 or 2, or a polypeptide encoded by said nucleic acid molecule, in the manufacture of a medicament for the treatment of a wound.
26. A method of treatment of a wound in an animal, including human, comprising the step of administering to said animal a therapeutically effective amount of nucleic acid molecule as claimed in claim 1 or 2 or a polypeptide encoded thereby.
27. A method as claimed in claim 26 wherein the nucleic acid molecule or polypeptide is administered at or close to a wound.
28. A method as claimed in claim 26 or 27 wherein treatment of a wound includes wound healing and associated conditions and therapy which promotes, augments, or accelerates healing of tissues and includes post-operative scarring, burns, psoriasis, acceleration of tissue remodeling.
29. A method of diagnosis comprising using a nucleic acid molecule as claimed in claim 1 or 2 as a probe to determine the presence or absence of a genetic defect in a patient.
30. A method as claimed in claim 29 wherein the method is carried out on an embryo *in utero*.
31. A method as claimed in claim 29 or 30 wherein the genetic defect is a mutation in a gene encoding an extracellular matrix protein.
32. A method as claimed in claim 31 wherein the gene encoding an extracellular matrix protein is *Fras 1* or a FRAS related protein.

1/8

Figure 1

(A)

ATGAACCTCTGAGTTGGGGGGC
TGCGAATGCCGTGCTGCTGCTCCTCCTGGCCTGGGCCAGCCCCACCTTCATCAGCATCAACCGCGG
GTGAGGGTGATGAAGGGCCACTCTGCCTTCCTGTGAGGAGATGACCTGAAGTTTGCCATCCCTAAAGAGA
AAGATGCCTGCAAAGTGAAGTTGTGATGAATGAGCCAATAACCCAGAGGGTTGGGAACTCACTCCACA
GGTCTTTGACTGCCATTTCTTCCCAACGAAGTCAAGTATGTTCAATGGTTGTCCAATTCCTTGATGAA
GACACAGTGAAGCTCAGACTTTACAGATTTACTGAAAGAGATACCTTCATAGAACTTTTATCCTGTGGG
TCTATCTCCTGGAACCAGACTGTAACATCATCCATATGAGTAACAATGTGCTGGAGGTGCCTGAATTCAA
TGGCTTGTCCCAAGCGATTGATAAAAACTGCTCAGATTCGATTATGATAGGATGGCTAGCCTGGAATGT
ACCGTCAGCCTGGACACTGCGGAGAACTCGGCTGCCAGCCCATGGCCAGATGGTTCTCGGGGAGCCTCGAC
CAGAAGAACCTCGTGGAGATCAGCCACACAGTTTTCCTCCAGAGTCCCACTGAGAGCAAACTGAAGTG
TCCAGGTGGGAGCTGTACCCAGGATTAAGAAAAATAGGAAGCCTCAAAGTGAGCTGTGAGGAGTTCCTG
CTGATGGGCCCTTCGTTATCAGCATCTGGATCCCCCTTCACCCAACATTGATTATATCTCCATCCAACCTGG
ACCTCACAGATACCAGGAGCAAAATGTGTACAAGTCAGAGAGTGCCTGGCTGCCTGTCTATATCAGAGC
TGGAAATTCGAATCAGATTCCAAAGGCTGCATTATGCGCGTGTATTATCTGGAAGTGGATCAGTTCATC
CTGACCTCCTTGACTACATCAGTTCTGGACTGTGAAGAAGATGAGACCCCTAAACCTTGCTGGTGTTC
ACATTACTAAAGCCCCGCTCCAGGGCTATGTGACTCACCTGTTGGATCACACCAGACCAATCTCCTCATT
CACCTGGAAAGATCTCAGTGACATGCAGATCGCTATCAGCCACCAACAGCAGCCATTCTGAGAGGAGA
CATGATGAGGTAGAATTGGAGGTATACGACTTCTTCTTTGAAAGGAGTGCACCTATGACAGTCCACATCT
CCATCAGAACAGCAGATACAAATGCCCCCGTGTATCCTGGAATACAGGTCTGAGTCTCCTTGAGGGGCA
GTCTCGAGCCATCACTTGGGAACAGTTTCAGGTGTGCGACATGACGACATTGGTGTGTCCGGCTAGTC
ACCGTTGGTGGCCTGCAGCATGGATGGCTGACTTTAAGAGGGGGGAAAGGGTTTCTCTTCACCGTGGCTG
ACCTCCAGGTGGAGTTGTTCGCTATCATATGATGACAGCGACTCCACCAAGAGCTTCGTGGTCTTCGG
GGTATTTGATGGCCATCACAGCATCCGTCAAAATTCCTCCATCAACGTCTTGCCCAAGATGATAGTCCC
CCGTTCTCTATAACCAATGTTGTGATTGAAGTGGAGGGGGCAGACCATCCTGATCCAGGGATCCATGC
TGCGAGCTTCAGATGTGGACGCCAGTGATGACTACATCTTCTTCAATATCACAAAGCCTCCACAGGCTGG
GGAGATCATGAAGAAGCCAGGGCCAGGACTGATAGGCTATCCTGTCCATGGCTTCCTTCAGAGGGATTG
TTAATGGAATCATTTATATCGTCATTTTGGTGGAGAAATCTTTGAAGATTCTTTCAATTTGTCTGT
GGGACAGCCATGAACCTCCAAATCTTTCAGTGCCACAGGTGGCAACAATCCATATACTCCAGTGGATGA
CCAGCTTCCAAAGAGGCTCCTGGAGTTTCTCGGCATTTGGTTGTCAAGGAACTGAGGTGGCCTATATA
ACTAAGAAACAGCTACATTTTATAGATTGAGAAATCATATGACAGGGAGCTGGTCTACACAATACTACTC
CTCCATTTTCTCCTTCAGTCACAGACACTTGGATGCTGGGAAATTTATTTATGGTGGACAGCATACCAAA
AGTAGTTAAGAAATCCTACGGCCCTGGAGCTGAGGTCACTTCACTCAGCATGCTGTGAACATATGAAAGTG
GCCTACATGCCCCCATGCAAGACATTGGTCCCCATTGTCAGAGATGTCCAGTTCACATTTTCTGTCTAGTA
ACCAACATGGCGGTACTTTGCATGGGATCTGCTTTAACATTACAATCCTCCAGTGGACAATCAAGTTCC
TGAGGCGTTTACCAACCTCTGAAAGTGACTGAGGGAGGTCAAAGCATCATCAGCACAGAGCACATTCTA
ATTTCTGATGAGATACCAAGCTGGACAATATTGACCTCTCCTGCGGGAATTGGCTCTGCACGGAAGGG
TGGAGCTGAATGGATTTCTCTAAATTCAGGGGGCACATTTTCTTGGGGCGATCTCCATACCTTAAAGT
TAGGTATCAACATGATGGAAGTGAAGTTCTTCAGGATGACCTAGTCTTGGAGGTACCGATGGCACAAAT
TCAGCAGAATTGTACTACATGTTGAGGTATTCCTGTCAACGATGAGCCACCAGTCTTAAAGGCTGACC
TCATGCCTGTGATGAATTGCTCAGAGGGAGGAGAGGTGGTCACTCACCTCTGAATACATTTTGTCTACTGA
TGTGGACAGCGATAACTTGAAGTTGATGTTTGTGATTGCTCGCGAACCTCAGCATGGGGTGGTGGAGAGA
GCTGGAGTCACAGTGGATCAGTTCTCTCAGAGAGATGTTATCTCAGAGGCCGTGACATACAAACACACAG
GTGGCGAGATTGGCCTGATGCCTTGTGTTGACACCATTACATTTGGTGGTTTCGGATGGAGAGGCTGGCCC
TTTTGTGAATGGCTGTTGCTACAATGGACCCCAATCCATCTGTTCCACTTCATGCGTCTTTCCAGTGTAT
GATCTCAACATCACGGTATACCCAGTAGACAACAGCCACCTTCCATTGCAATAGGTCCCGTGTGTTGTTG
TAGATGAGGGCCGCTCAACAGCCCTTACTGTTAACCATTTGTCCGCCACTGACCTGACACTGCAGCAGA
TGACTTGAATTTGTTTTGGTTTTCTCCTCCTCAGTTTGGCTACCTCGAAAATATACTCCCTTCTGTGGGT
TTTGAAGAAAGCAATATTGGCATAAGTATAGATTCAATTCAGTGGAAAGACATGAACGCTTTTCACATTA
ACTATGTGCAGTCCAGGATCTGAGGATAGAACCAACTGCCGACAGTTACGGGTGTACGTACAGATGG
GAAGCATCACTCCTTGGAGATACCATTTTCTATTATAATCAACCCACAAATGATGAAGCTCCTGACTTT
GTAGTGCAGAAATATTACTGTGTGTGAGGGTCAGATGAAAGAGCTGGACTTTCCATCATCAGCGCTGTGG
ACCTGGACATTTCCCAGGATGCCCTGCTGTTGAGCATCACTCAAAGCCACGCCATGGCCTCCTCATCGA
TAGGGGGTTTAGCAAAGACTTCTCTGAGAATAAGCAGCCAGCCAACCTCACCAGAAACATGCACCTGTT
CAGACTTTTCCATGGAACTCCTCAAGACTGGAATGAGGTTGACGTACATGATGACTCAGAGAGCC
TTGCTGATGATTTTACAATCCAATTGTGATGAGGAAACATAAGATACTTAAACCATTTTCAGTGGAGGT
CATCCCAGTTAATGATGAAAAACCAATGCTGAGCAAGAAGGCTGAAATTGCAATGAATATGGGTGAAACT

2/8

CGTATTATTTCCAGTGCTATTCTTTTCAGCCATAGATGAAGACTCACCCAGGGAGAAGATTTACTATGTAT
TTGAAAGGCTTCCCCAAATGGGCAACTTCAGCTTAAGGAATGTCAACTAAGATGTGGAATATAG

(B)

ATGAACCTCTCTGAGTTGGGGGGCTGCGAATGCCGTGCTGCTGCTCCTCCTGGCCTGGGCCAGCCCCA
CCTTCATCAGCATCAACCGCGGGGTGAGGGTGATGAAGGGCCACTCTGCCTTCCTGTGAGGAGATGACCT
GAAGTTTGCCATCCCTAAAGAGAAAGATGCCTGCAAAGTGGAAGTTGTGATGAATGAGCCAATAACCCAG
AGGGTTGGGAAACTCACTCCACAGGTCTTTGACTGCCATTTCTTCCCAACGAAGTCAAGTATGTTTACACA
ATGGTTGTCCAATTCTTGATGAAGACACAGTGAAGCTCAGACTTTACAGATTTACTGAAAGAGATACCTT
CATAGAACTTTTATCCTGTGGGTCTATCTCCTGGAACCAGACTGTAAACATCATCCATATGAGTAACAAT
GTGCTGGAGGTGCCTGAATTCAATGGCTTGTCCCAAGCGATTGATAAAAATCTGCTCAGATTCGATTATG
ATAGGATGGCTAGCCTGGAATGTACCGTCAGCCTGGACACTGCGAGAACTCGGCTGCCAGCCCATGGCCA
GATGGTTCTCGGGGAGCCTCGACCAGAAGAACCCTCGTGGAGATCAGCCACACAGTTTTTTCCAGAGTCC
CAACTGAGAGCAAACTGAAGTGTCCAGGTGGGAGCTGTACCCAGGATTAAAGAAAATAGGAAGCCTCA
AAGTGAGCTGTGAGGAGTTCTGCTGATGGGCCCTTCGTTATCAGCATCTGGATCCCCCTTACCCAAACAT
TGATTATATCTCCATCCAACCTGGACCTCACAGATACCAGGAGCAAAATTTGTGTACAAGTCAGAGAGTGCG
TGGCTGCCTGTCTATATCAGAGCTGGAATTCGGAATCAGATTCCAAAGGCTGCATTATGGCCGTGTTTA
TTCTGGAAGTGGATCAGTTTATCCTGACCTCCTTGACTACATCAGTTCTGGACTGTGAAGAAGATGAGAC
CCCTAAACCTTGTGCTGTTCAACATTACTAAAGCCCCGCTCCAGGGCTATGTGACTCACCTGTTGGAT
CACACCAGACCAATCTCCTCATTACCTGGAAAGATCTCAGTGACATGCAGATCGCCTATCAGCCACCAA
ACAGCAGCCATTTCTGAGAGGAGACATGATGAGGTAGAATTTGGAGGTATACGACTTCTTCTTTGAAAGGAG
TGCACCTATGACAGTCCACATCTCCATCAGAACAGCAGATACAAATGCCCCCGTGTATCCTGGAATACA
GGTCTGAGTCTCCTTGAGGGGAGTCTCGAGCCATCACTTGGGAACAGTTTCAGCTTGTGACAAATGACG
ACATTGGTGCTGTCCGGCTAGTCACCGTTGGTGGCCTGCAGCATGGATGGCTGACTTTAAGAGGGGGGAA
AGGGTTTCTCTTACCGTGGCTGACCTCCAGGCTGGAGTTGTTGCTATCATCATGATGACAGCGACTCC
ACCAAAGACTTCGTGGTCTTCCGGATATTTGATGGCCATCACAGCATCCGTACAAATTTCCCATCAACG
TCTTGCCCAAAGATGATAGTCCCCGTTCCCTATAACCAATGTTGTGATTGAAGTGGAGGAGGGGAGAG
CATCCTGATCCAGGGATCCATGCTGCGAGCTTCAGATGTGGACGCCAGTGATGACTACATCTTCTTCAAT
ATCACAAGCCTCCACAGGCTGGGGAGATCATGAAGAAGCCAGGGCCAGGACTGATAGGCTATCCTGTCC
ATGGCTTCCTTCAGAGGGATTTGTTTAATGGAATCATTTATTATCGTCATTTTGGTGGAGAAATCTTTGA
AGATTCTTTTCAATTTGTCTGTGGGACAGCCATGAACCTCCAAATCTTTCAGTGCCACAGGTGGCAACA
ATCCATATAACTCCAGTGGATGACCAGCTTCCAAAAGAGGCTCCTGGAGTTTCTCGGCATTTGGTTGTCA
AGGAAACTGAGGTGGCCTATATAACTAAGAAACAGCTACATTTTATAGATTGAGAATCATATGACAGGGA
GCTGGTCTACACAATAACTACTCCTCCATTTTCTCCTTCAGTCACAGACACTTGGATGCTGGGAAATTA
TTTATGGTGGACAGCATACCAAAAGTAGTTAAGAATCCTACGGCCCTGGAGCTGAGGTCAATCACTCAGC
ATGCTGTGAAGTATATGAAGTGGCTACATGCCCCCATGCAAGACATTGGTCCCCATTGCAGAGATGT
CCAGTTACATTTTCTGTGCTAGTAACCAACATGGCGGTACTTTGCATGGGATCTGCTTTAATACATCAATC
CTCCAGTGGACAATCAAGTTCTTGGAGGCTTCCCAACCCCTCTGAAAGTGAAGTGGAGGAGGTCAAAGCA
TCATCAGCACAGAGCACATTCTAATTTCTGATGCAGATACCAAGCTGGACAATATTGACCTCTCCCTGCG
GGAATTGCCTCTGCACGGAAGGGTGGAGCTGAATGGATTTCCTCTAAATTCAGGGGGCACATTTTCTTGG
GGGATCTCCATACCTTAAAAGTTAGGTATCAACATGATGGAAGTGAAGTTCTTCCAGGATGACCTAGTCT
TGGAGGTACCGATGGCACAAATTCAGCAGAATTTGTACTACATGTTGAGGTATTCCTGTCAACGATGA
GCCAACCAGTCTTAAAGGCTGACCTCATGCTGTGATGAATTGCTCAGAGGGAGGAGAGGTGGTCATCACC
TCTGAATACATTTTTGCTACTGATGTGGACAGCGATAACTTGAAGTTGATGTTTGTGATTGCTCGCGAAC
CTCAGCATGGGGTGGTGGAGAGCTGGAGTCAAGTGGATCAGTTCTCTCAGAGAGATGTTATCTCAGA
GGCGGTGACATACAAACACACAGGTGGCGAGATTGGCCTGATGCCTTGTTTTGACACCATTACATTTGGTG
GTTTCGGATGGAGAGGTGGCCCTTTTGTGAATGGCTGTTGCTACAATGGACCCCAATCCATCTGTTCCAC
TTCATGCGTCTTTTCCAGTGTATGATCTCAACATCAGGTATACCCAGTAGACAACCAAGCCATTTGTCGCC
TGCAATAGGTCCCGTGTGTTGTTGATGAGGGCTGCTCAACAGCCCTTACTGTTAACCATTTGTCGCC
ACTGACCTGACACTGCAGCAGATGACTTGAATTTGTTTTGGTTTCTCCTCCTCAGTTTGGCTACCTCG
AAAAATACTCCCTTCTGTGGGTTTTGAAAAAGCAATATTGGCATAAGTATAGATTCAATTCAGTGGAA
AGACATGAACGCTTTTACATTAACTATGTGCAGTCCAGGCATCTGAGGATAGAACCAACTGCCGACCAG
TTCACGGTGTACGTACAGATGGGAAGCATCACTCCTTGGAGATACCATTTTCTATTATAATCAACCCCA

3/8

CAAATGATGAAGCTCCTGACTTTGTAGTGCAGAATATTACTGTGTGTGAGGGTCAGATGAAAGAGCTGGA
CTCTTCCATCATCAGCGCTGTGGACCTGGACATTCCCCAGGATGCCCTGCTGTTTCAGCATCACTCAAAAG
CCACGCCATGGCCCTCCTCATCGATAGGGGGTTTAGCAAAGACTTCTCTGAGAATAAGCAGCCAGCCAACC
CTCACCAGAAACATGCACCTGTTTCACAGCTTTTCCATGGAACCTCCTCAAGACTGGAATGAGGTTGACGTA
CATGCATGATGACTCAGAGAGCCTTGCTGATGATTTTACAATCCAATTGTCAGATGGGAAACATAAGATA
CTTAAACCATTTCAGTAGAGGTCATCCCAGTTAATGATGAAAAACCAATGCTGAGCAAGAAGGCTGAAA
TTGCAATGAATATGGGTGAAACTCGTATTATTTCCAGTGCTATTCTTTTCAGCCATAGATGAAGACTCACC
CAGGGAGAAGATTTACTATGTATTTGAAAGGCTTCCCCAAAATGGGCAACTTCAGCTTAAGATAGGGAGG
GACTGGGTTCCTCTCTCCCCGTCATGAAATGCACCTCAGGAGGAAGTGGATCTGAACCTTGCTGAGATACA
CACACACCGGGCAATGGATTCCCAGAATCAAGATAGCTTCACCTTCTACCTTTGGGATGGCAACAACAG
GTCCCCCTGCTCTTGACTGTCAAATCACCATCAAGGACATGGAAAAAGGTGATATTGTCATCCTTACAAAA
CCACTCGTGGTGTCTAAAGGTGATAGAGGTTTCTTAACAACCACCACGCTCCTGGCTGTGGACGGAACAG
ACAAGCCTGAGGAAGTCTCTATGTTCATCACCTCCCCTCCGCGATATGGCCAGATCGAATATGTTCACTA
TCCTGGAGTTCCCATTAACAACTTCAGCCAAATGGATGTAGTGGGGCAGACAGTTTGCTATGTACACAAG
AGCAAGGTGACTGTCTCCAGTGACAGATTGAGATTCATCATCAGCAATGGACTGCGGACCGAGCACGGGG
TGTTTTGAGATCACACTGGAGACTGTGGACAGAGCCCTGCCCTGTGGTAACCAGGAACAAGGGGTTGAGACT
GGCCCAAGGGGCCGTGGGCTGCTTTCCCCTGACCTCCTTCAGCTGACCGACCTGATACACCTGCGGAG
AACCTCACCTTCTCTTGGTTCAGCTCCCCCAGCATGGCCAGCTCTACCTGTGGGGGACAGGGCTACTTC
AACACAATTTACCCAGCAGGATGTGGACAGCAAGAATGTGGCCTATCGGCACTCAGGAGGGGACTCCCA
GACTGACTGCTTTACTTTTCATGGCCACAGACGGGACAAACCAAGGCTTTATTGTGAATGGGAGAGTGTGG
GAAGAACCTGTTTTATTACCATTCAGGTAGACCAATTGGACAAAACAGCTCCTCGTATCACACTCTTGC
ATTCCTCTCTCAAGTGGGGCTCCTGAAAAATGGCTGCTACGGGATTTACATCACTTCCCGCGTGTGAA
GGCATCAGACCTTGACACTGAGGACGATCAGATCATCTTTAAATTTCTACAAGGGCCAAAACATGGACAT
TTGGAGAACACAACAACAGGTGAATTTATCCATGAGAAATTTAGCCAAAAGGACTTAAACAGTAAGACTA
TTCTTTTACATATAAACCCTCTTTGGAAGTAAATTCAGATACCGTGGAAATTTCAAATCATGGACCCAC
AGGGAACCTCGGCCACTCCTCAAATTTTGGAACTGAAGTGGTCTCATATTGAATGGTCACAGACCGAATAT
GAAGTCTGTGAGAAATGTGGGTTTGTGGCCCTTGGAAATTTATCAGAAGGGGATATTCCATGGACTCGGCCT
TTGTGGGTATAAAGGTCAACCAAGTGTGAGCTGCAGTTGGAAAAGATTTACCCGTGATTCCATCTAACT
GATTGAGTTTGACCCAGGAATGTCAACTAAGATGTGGAATATAGCAATTACCTATGACGGATTAGAGGAA
GATGATGAGGTCTTTGAAGTAATTCTGAACTCCCCTGTGAATGCAGTTCTTGGCACAAAGACAAAAGCTG
CAGTGAAAATTTTGACTCAAAAGGAGGACAATGCCATCCTTCATATTCTCCAACCAAAGCAAGCACAG
CACATGGGAGAAGGGCATTGTCATCTGCTGCCCCCAGGGTCTTCTCATCCACCACTTCTGGTTCTTTT
CATCTGGAAAGAAGACCTCTTCCATCTTCCATGCAGCTAGCAGTCATCAGGGGAGACACCCTGCGGGGCT
TTGATTCTACAGATCTTTCTCAAAGGAAGCTTAGGACCCGTGGGAATGGCAAAACAGTTTCGTCCATCCTC
TGTTTTATAGAAATGGAACAGACATCATCTATAATTATCATGGGATAGTTTCCTTGAAACTGGAGGATGAC
AGTTTCCCAACTCACAAAAGGAAGGCCAAAGTATCCATCATTAGTCAGCCACAAAAGACAATCAAAGTGG
CAGAACTGCCTCAAGCAGATAAGGTGGAATCCACAACCTGACTCACACTTCCCCAGACAGGACAGTTGCC
CTCATTTCCAAAGAACTGCACTCTGGAATTAAGGGGACTCTTCCATTTTGAAGAAGGCATCCAGAAGCTG
TATCAGTGCAATGGGATCGCCTGGAAAGCCTGGAGTCCCCAAACCAAGGATGTGGAAGACAAATCCTGTC
CAGCCGGGTGGCACCAGCACTCAGGCTACTGTACATCTTGATCACAGAGCAGAAAGGCACCTGGAATGC
GGCTGCCCCAAGCTTGACAGGGAACAATACCTGGGCAACCTTGTAAGTGTATTCTCCAGGCAGCACATGCGG
TGGCTCTGGGACATTGGTGGGAGAAAGTCTTTTGGATAGGTTTGAACGACCAAGTGCATGCTGGCCACT
GGGAGTGGATCGGTGGTGAACCTGTTGCCTTACCAATGGGAGAGAGGGCCCTCTCCACGCTCCAAGCT
TGGAAGAGCTGTGTTTGGTTCAAAGACAAGGGAATGGCAAAACAAAGACTGTAGGAGAGCCAAACCT
CATAATTATGTGTGTTCCAGAAAACCTCTAA

4/8

Figure 2

(A)

MVTQESMLKAALPLFTRFII SNGLRTEHGVFEITLETVDRALPVVTRNKGLRLAQGAVGL
 LSPDLLQLTDPDTPAENLTFLLVQLPQHGLYLWGTGLLQHNFTQQDVDSKNVAYRHSGG
 DSQTDCTTFMATDGTNQGFI VNGRVWEEPVLFTIQANQLDKTAPRITLLHSPSQVGLLKN
 GCGIYITSRVLKASDPDTEDDQIIFKILQGPKHGHLENTTTGEFIHEKFSQKDLNSKTI
 LYIINPSLEVNSDTEVFQIMDPTGNSATPQILELKWSHIEWSQTEYEVCEENVGLLPLEII
 RRGYSMDSAFVGKVNQVSAAVGKDFTVIPSKLIQFDPGMSTKMWNIAITYDGLEEDDEV
 FEVILNSPVNAVLGKTKAAVKILDSKGGQCHPSYSSNQSKHSTWEKGIWHLLPPGSSSS
 TTSGSFHLERRPLPSSMQLAVIRGDTLRGFDSTDL SQRKLRTRGNGKTVRPSSVYRNGTD
 IINYHGHIVSLKLEDDSFPTHKRKAKVSIISQPQKTIKVAELPQADKVESTTDSHFPRQD
 QLPSFPKNCTLELKGFLHFEEGIQKLYQCNGIAWKAWSPQTKDVEDKSCPAGWHQHSGYC
 HILITEQKGTWNAAAQACREQYLGNLVTVFSRQHMRWLWDIGGRKSFWIGKLNQOVHAGH
 WEWIGGEPVAFTNGRRGPSQRSKLGKSCVLVQRQGWQTKDCRRAPHPHYVCSRKL

(B)

mnsIswgaan avllllllaw asptfisnr gvrvmkghsa flsgddlkfa ipkekackv
 61 evvmnepitq rvgkltpqv dchflpnevkv yvhngcpild edtvklrlrlyr fterdtfiet
 121 filwvyllep dcniihmsnn vlevpefngl sqaidknllr fdydrmasle ctvsldtart
 181 rlpahgqmv lgeprpeerg dqphsfppes qlraklkcp gscptglkki gskvsceef
 241 llmglyrql dppspnidyi siqlldtdtr skivyksesa wlpvyiragi pnqipkaafm
 301 avfilevdqf iltslttsvl dceedetpkp llvfnitkap lqgyvthld htrpissftw
 361 kdlsdmqiay qpnnsshser rhdevelevy dfffersapm tvhisirtad tnprvswnt
 421 glsllegqsr aitweqfqlv dnddigavrl vtvggqlhgw ltlrggkgl ftvadlqagv
 481 vryhhdssds tkdfvvrif dghhsirhkf pinvlpkdds ppflitnvvi eleegqtili
 541 qgsmlrasdv dasddyiffn itkppqagei mkkpgpglig ypvhgfqlrd lfngiiyyrh
 601 fggeifedsf qfvlwdshep pnlsvpqvat ihitpvddql pkeapgvsrh lvvketeav
 661 itkklqlhfid sesydrelvy tittppffsf shrhldagkl fmvdslpkvv knptalelr
 721 ftqhavnymk vaymppmqdi gphcrdvqft fsvsnqhgg lhgicfiniti lpvndqvpea
 781 fnplkvtegg qsiistehti lisdadtld nidlslrelp lhgrelngf plnsggtfsw
 841 gdlhtlkvry qhdgtevlqd dlvlvtdgt nsaevlhve vfpvndeppv lkadlmpvmn
 901 cseggevvit seyifadtdv sdnklmfvi arepqhgvvr ragvtvdqfs qrdviseavt
 961 ykhtggeigl mpcfdtitlv vsdgeagpfv ngccyngnp svplhasfpv ydlnitvypv
 1021 dnqppsiaig pvvfvdegcs taltvnhsa tdpdtaaddl efvlvsppqf gylenilpsv
 1081 gfeksnigis idsfqwkdmn afhinyvqsr hlriepadq ftvyvtdgkh hseipfsii
 1141 inptndeapd fvvqnitvce gqmkelldssi isavldipq dallfsitqk prhgllidrg
 1201 fskdfsenkq panphqkhap vhsfsmellk tgmrltymhd dsesladdft iqlsdgkhki
 1261 lktisvevip vndekpmlsk kaeiamnmge triissails aidedsprek iyyvferlpq
 1321 ngqlqlkigr dwvplspgmkt qteevdlnl lrythtgamd sqnqdsftfy lwdgnnrspa
 1381 ldcqititkdm ekgdviltk plvsvkgdrg fltttllav dgtldkpeell yvitsppryg
 1441 qieyvhypgv pitnfsqmdv vggvtvcyvhk skvtvssdrf rfiisnglrl ehgvfeitle
 1501 tvdralpvt rnkglrlaag avgllspdll qltdpdpae ntlflvqlp qhgqlylwtg
 1561 gllqhnftqq dvdsknvayr hsggdsqtdc ftinatdgtm qgfivngrvw eepvlftiqv
 1621 dqldktapri tllhspsqvg llkngcygiy itsrvlkasd ptedddqiif kilqgpkgh
 1681 lentttgefi hekfsqkdl sktilyiinp slevnsdtve fqimdpdgns atpqilelkw
 1741 shiewsqtey evcenvglp leiirrgysm dsafvgikvn qvsaavgkdf tvipskliqf

518

1801 dpgmstkmwn iaitydglee ddevfevln spvnavlgtk tkaavkilds kggqchpsys
1861 snqskhstwe kgiwhllppg ssssttsgsf hlerrlpss mqlavirgdt lrgfdstdls
1921 qrklrtgng ktrvpssvyr ngtdiiynyh givskledd sfpthkrkak vsiisqpqkt
1981 ikvaelpqad kvesttdshf prqdqlpsfp knctilelkgf fhfeegiqkl yqcngiawka
2041 wspqtkdved kscpagwhqh sgychilite qkgtwnaaq acreqylgnl vtvfsrqhmr
2101 wlwdiggrks fwiglndqvh aghwewigge pvafngrrg psprsklgks cvlvqrqgkw
2161 qtkderrakp hnyvcsrkl

6/8

Figure 3

GCTGGTTGGGTCCG CAGGGACAAAAC T GAGCTTTTGT TAA T GAGGT CGGT TGCCTCTTCCAGTCTCTTCCAGCT
TCCTTTCTCTACTACTAGTTTGTACTACTGCTCCGCCACTCGCCGAGCCAGAGCAGCCTCTCCCGCTGCAGCAC
ACAGGCACACACACACTCAGACACAGGCACACACACATTCACAAATGGGCACACAGTCACATACAGACACACCCAC
TCACAGAGGTACACACACAGGCACACATACAAACAGATAGACACACAATCACCACACACAGACACATACACAGGC
AAATGCCTTCACATGCATGCACACATAGTGACCCACCCACACATGTGCATGCATTCTCACGCATGCTCCACACAC
ACACGGGGGCACACACATACGCACAGGCCGCACATGCCACACCCACACATGTACATACTCACACAGACTCTTCCA
ACACACTTTTTTTCTGCCCTTCTTTTCTTCCAACACATTCCCATCTTCAAAGGTTTTTATTTAAATTTAGATGAGA
AGCCCAGCATTCAGAAGGGACAGGGCCCGTTTTAACTTTTGTCTATGGCATTTAAGGAAAAAGATAGGATGGAAC TA
GATCCTAAGGTGCTGTTTTTAAGTACCACCTTAGTTCTTCCATTGTAGGCTTTTTGTGTGTCTGGCTCTTTTCG
AAAACCCATAAAAAATCCCATTAATCTGATCCATTG CATATAGT GCGTGATCCCAATAATGCAGATATATCACTGG
CCAAGAAGTGTCTCTCAACAGAGATTACATAGTTAACTTGT CAGTATT CAGTGGTCTGTGAAAGTCTTAATCCT
TAACAAGTGTTAATAAAATGCTTTCTGTATTTTATTTCCACCAGGAAAAAAAAGATAATGGCCTACCATCTGG
CCTTCTAAATTAACCATAAAGAGAGAATTTGAAGTATTAAC TCACTTTGGGGGAACGT TTTATATTAAGACATGT C
TGGCTATCCTAAACCTTTGCCAAAAATATAGAACACTGTGAATCCAATTCAACCTAGATGTGGTTTCCCGTAACAA
GGTCACAACACCATATGCCCTTCTTTGGAAGCCATAAGGAAAAAGAGAGAGGAAAAATGAAAAAGGAGAAGAAAAATTA
CATGAGGAAGGAAGAAAGAAAAAGAAAAATGTAAGAAGAAAAAGAAATGCATAAGAGGGAGAAAAAAATTTATGTA
TTTTAAAAACATGCTTTGTGGTTACTTTAGCAATGAATGGTGCATATACACTTACTCAAAGTCATCGGAGATTTCA
AATCTTCAAGAAATTTCTTGATTCATTTTATTTGCTCTGAGAATAAGTAGGTGAAAGTTAAGAATCACTAAGCTCTT
TATAAGTGAACAGATATTATTTACTAAACACAGTTATTATTAATAAATATAGTGACAATGTGTGAGTCATATATT
AATCCTGAAAAGGCCAAGAAATAAAGCCTTAATAATGACAGTTAAGAAACACAGCATAATAGACAATTTTATTTCCC
AAGTATTTTCAAATTTGCCAGGTTGTGAATATCTTACCTGCAAGAAATGAAGATGGTTTTAGAAAGCTTTAGT
TGAGACTCGGATTCCATTTGACAAAGGAATTTGTATACTATGTACAAACGGTTGCATCAACATGCTTATTTCTAACAT
CTAAATAAGGTTTATAGATTAGTCTAGAAACTAATAAGGTTTATAGATTAGTTTATAGATAGATTATAAAGTTTAT
AGATAGATTAAAGAGCTAAATAAGGTTTACAGATTAGTCTAGAAACAGAGATAAATGCTTTTATATCACTTCTATGG
GAAAAATAACTTAAGCATGCTAAGCAAAAAATAACAACTTAGATTGTAAAGTATTTCAAAAAGTCCGCAGGAGAT
ATAATCTTTCCAAAGTTAATTATATACATGAAGCTTTGGCCTAACTGTGCTGAACATATTCTGAAATAGGCGAGTTAA
GAGTTAGATCAGTCAAAATGTCTGAAATCTGAAGTACCTGTTTATACAAATGTCAATTCCTTCTATCTGCCACAAA
TGAATAAGGGCCTGGAATAATACAATTATCAACTTCAATCAATAAAATACTGTTTGAACAGATTGAAATGCCAAG
CTAGGGAATTTAAAAAGAACTAGTCAAAGAAACAACAAAGAAATCTTCCCCTAAAAGCATGTGCTGCTGAACCTTAG
TGCTGGCACAAAATTCATTTTTAAAGAAATCATTTTACTTCTAGCCCTAGTTGGAGTGATAATGAAAAATATCAAA
TGTGCTGTTGGAGGTTGGCATATCATTTTGGATTGTAAATATGAATGATTTTTGTGTAATTGCTGTAGAATGTAGCA
TTCTCTGAAAGGCATAGCCAATGCCCTTACCACAAATTCATATAAAAAATTAGGAATGTGCCATTATTAATTTTGAC
GATGGCAGCAGACAGAAGTGTTATCTATTGCAACGTAAAGATAGTTTTCCACTGAGTCTCTAACACTAGATGGCCA
TGGGAAGCAATACTATGGCAAGAGAGAGGATGAGTCCAGGAATAGAAAGGTTATACATACCACAATTGAATAAT
AAACATGAAAACAATAAATAAAGAAAAGCAATACAAACACCTCTGTGTACCAGAAAAGAGCAATAACACAGCCGA
TGAGAGGCTTTATAGGAAGAGCCTTGCTGACCTCAGACTTGGGAGCTGACAACGTGATTAGGGGTTATCAGGAAT
AAAACCACCCCTGAGCAGGGGCACAAGAGCAGCGTGCTGCTTGCCATGAGAGCGCTACCGCCAAATTTGACTGTC
AACATGAATGCAAAATCCTAAATAACTAATTTACAAATCAAATTCAGCTCTGCTTTTTAAAGACAACATGACCAA
TAGCGATGAATTTTTCTTTCAGGATTGCAAGGATGTTTCAAATCAGAACATCTAGCAGTTTAAATATATTAAAGAA
GGAAAAAGGCATGACTATCTTAATAAGTGCAGAAAAAGACATT CAGTAAATTTAATACTCATTCAATTAATAAAA
CCAAAAACTGAAAAC TATTAGCCAAC TAGTCTGGAAGGGAAGTTCC TAAATCCAATTAAGACATCTTACTCCAA
TCTCCACAATTCAACATTTGGATAGGAGATCCTAACCAACACAGTAAGTCAAGATAAAAAAATGAGGCATAAAAAAT
TAAAGGAAAAATAAACCTTTTATTATTTGCAAAAGATAAGAGTCTCTGTACAAATATGCAAAAGAATTTACTGTT
TGAAC TATTCAAAGAGTT CAGCAAGTTTGTGTTGACATAAGGTCAACATACAAAAATAGATGGCGTT CATATATACA
AGTAACAAC TAAAT TACTGAAA TACTTTTTAAAGACAGCTTTTATAATAGCAGCAAACTTTAATGAGCCTGGTA
ATTCATCTAAATTAATTAATTTGGGTAAAAATTT CATGAAGAAAAATAAAAAATTAATAAAGGGCATAAAAAATCTA
GATAAATGGGAAGATAATATCAAAC TGGTTGAGAAGACTTAATAATGTAAAAATGTCAATTTCTCCCAAAATTA AAA
TTTCAATGCAGTGTCAATTTGGAATGCCAACAAGGCATTTTATATAAATATGTGTACACATATACACATTTATGTACA
TATATGTTAT
TGTGTGTGTGTGTATGTGTGTGTGTAT
GATACTAAAAATTTATCTAAAAGATAAAAAACCAGGAATGGCCAAGACAAATTGAGAAGATACATTTAATGATTAA
TTTCCCTGCTAAATATTTGATATTTATATTATAAAGTTTAAACAATTAAGAAATTTGTGGAATTTT CACAGTGATAG
AAGGCAAGAAAGAGTCTAGAAAAGCTTGGAAAGCAGGTCACTCACATATAGCAACCTGGTGTATGACCAAGGT CAG
AATAGAAGTCAGTGAAGAAATAATAGACAATCAATAATACGTTGACAGGACAATAGGTTATCCATATGAAAAACAT
AGCATCAGAACTCTACTTCATATCATATATCAAAAAACATTAGATCTTTACCTCATACTATACACCAAAACAATTT
GCAGTTGATTTGAGTATCTTAATTTGATAACAAAGCCTTTAACTTTTGGAAACCAAAAAATCAAGGAATATCTTCT
TAATATCAAGGGAAAAATTTCTTAAGTGTGATACAATAAGTAAAGGCTATAAAAGAAAACTTGATATATTTTATAT
AATAATTAATAATTTCCCATGGAGCAAAAACTTAAAGTTTTTTTTTAAAAAAGAAACAGAAATTTGGGAGAAGGTA
CTGAAATGAAAATAAATTACAAAATCTTTATGCTCAAATTTAAAGAGCTTCTGAAGCCTATACAAATCATTAA

7/8

GAAAAAATCTTAATATTCTCCTTTTATTAAAGGAGAAGACACACAAATGGCTAATAAATACATGAAAGCATCCAA
TTACACAAACAATCAGGCCAGTACAAATGGAGACTAAAAATGAGACTCTCTCAGATAAATAAACGTTTAAAAATCT
TACAATAAATTAATCTGGTAAATTAATAAATTAATCTGGTAAAGATAAGCGGGGGAAAACTCTTAGTTGCT
ACTGGAAGTTTAAATAGACACAAACACTTAGAAAGTAGTCTGACCATTCTAGTAAAGTAAATATTCCAAAAAC
CTTAAATTAGGCATTTTAACTTCTAGAGTTACACTTAGATTCTCATATGTGTACAAAAAGTCTTATACCTGACATT
CTTTGCCATTGTAATTGAAAAATATTGAAACAACCATCATCTAAACCAAAAAGATGTTGAGTAAAGAAAGCAAATT
GCAGCCAGGTGCAGTGGCTCACACCTGTAATCCTAGCACTTTGGGAGGCCAGGCGGGTGGATCACTTGAGGTGAG
GAGTTTGAAACCAGACTGGCTAACATGGTGAAACCCCATCTCTACTAAAAATAGCAAAAAATTAGCTGGGCATGG
TGGTAGGTGCTGTAATCCCAGCTACTTGGGAAGCTGAAGCAGGAGAATCACTTGAACCTAGATGTGGTGGTTGCA
GTGAGCCAAGATCACACCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAGAAAAA
AAGGAAATTGCAAAAGTACAGGAACATGCTTTAATAAGGATTACACTGAATCTGTAGATCACTTTGGGTAGTATGG
ACATTTTAACTATATTAAGTCTTCCAATCCATGTACATGGGCTGTCTTTCCAATTATTGTTCCTTCTTTAATATT
TTTATCAATGATTTGTAGTTTTTAGTGTACAGCGTATGAGACTTTTCATCTTGATTACATTTTACCTAAGTAAAGA
GGACTTTTAAAAAAAGTGTTTTTTAAAAAACTTTACATGGATTTCAAAAATTCTTTGCACCAAAATAAATCTGTA
CTAACTTGTATAACATGTGCGAACGTCTAGCTTGAGGCACTAACAACTATAAGACACCAGTTTGAAAGAAGTGC
TATCAGAGCAACATGAATTCTGTCTAAATTTGAAATAAGAGCAAAGATCAAATTTACGGTGGAGAAGTAGTGAAATC
ATTGGTGTCTTATGAAAAGTTTGTGGGGAAAATGCCCCACAGAAATCAGCAGTTTGCAAATGGATAACTCATTTT
AAGAAGAGATGAGACAATGTTGAAGATGAAGCTCACAGGGGAGACCATCCACATCAAAATGCGAAAAAATTC
ATTTAGTTCATACCTTAATTGAAGAGAATCAATGATTAATAGCACAGACAACAGTCTACACCTAGACACCTCAGT
TGGTTCAGCTTACACAAATCTGACCAAAAAATTAAGTCAAGCAAACCTTACCATTGAGTGGTGCCAAAATGATTG
TGCTGTGATCGGCTGTAGACAAGAACAGAGTTTCAATGGAAATTTTAAACAAGGAGGATCAAGATCCTGAAGGAT
TTCTTTAAATAAATTACAACAGAAAGATAAAACACTGCTTTACCAGTATGCCCTGAAGACAAGGAACAATCAAAGCA
ATGACTACCAAAAGGTGAAAGTTGTCCAGCCAAAGCAAAGCAGACCAAGTCAAGAGCAAAGATCATAGCAAACTTG
TTTCGGATGTTCAAGGAATTTTGCTTGTTAACTTTCTGGAGGACCAAGAACGATAAGATTTGCATACATGAGAG
TGTTTTAAGAAAGTTAGCCAAAGCTTTAGCAGAAAAACAGCAGGAAAGCTTTGCCAGAGTCCACCTCCACCAGGA
CAATGCTGCTGCTCATTCCTTTAATTGAACAATGGCCATTTTGCAAGACTTTTGATGGGAAATCATTATACATCCA
TCTTACAGTCTCAAATTTGGCTCCTTCTGACTTCTTTTTTTCCCTAACCTTAAAAAGTCTTTAAAGAGAACCCA
TTTTTCTTCATCTAATAATGTTAAAAAGACTGCATTGATATGGCTAAACTGCCAGTATTTCCAGGTCTTTAGGGAC
AGACTAATTAGCTGGTATTATCATTTACAAAAGTATCTTGAACCTTCATGGAGTTTATATTGAGAAATAAAGTATGT
ATTTTTTATTTTTATTTTTTAAATTTAATTTTCCACAATTTTAAAAAGTCCCCTCATATGGCATTTATTGTGTTTA
GGTAAATTCCTTTTATACCCAGAAAAAATCATCAAATTCATATGAAACCACAAAAGATTTGAAATGGCCAAAAACA
ATCTTAAAAAAGAAAGAAACTGGAGACACAACACCTCCTGATTTCAAATATAATACAAAGCTATATTAATC
AAAACAGTGTGATACTGACATAAAACAGAGATAGAGACCAATGGAACAGAAATAAGAGCCCAGAAATAAACCAG
GCATGTACAGTCAACTGATCTTTGACAAGGGTTACACAAATACGCAATGGGGAAGAGACAGTTCTTCAACTGATGG
TGTTGGCAAAACTGAATATTCACATGCAGAAGAATGAAATTTGGGCCCTTAGCTTACACAATATACAAAAATCACCT
CAAAATAGATTAAAACTTAAGTGTACACATGAAACTATGAACTTGTAGAAGAAAACGTAGGGGAAAAGCTTTTA
TGACATTAGGCTTGACAATGATCTGTTGGATATAATACCAAAATAATATAAGTAACATTGGAACAAGAATCACTG
AGGAGGAAAATTCAAAAAAGAAAAAAGAAAAAAGCGCCAGAAGTAACAAAAGCAAATTAGACAAGTA
AAACTACATCAGACTAAAAGTCTTCTGCACAGTAAACAAAACACTCAACAGATTAAGGTAACCCACAGAATAGG
GAAAAATATTGCAAAATCATATTTCTGATAAGGGATTAATATCAAAATATACAAGGAACCTTGCACTCAATTA
AAAAATGGGCAAAATCTCAATAGACATTTTAAATGGGCAAAAAAATTCACAGACATTTCTTTAAGATATAAAAA
TGGCCAAAAGGTATATAAAAAATGCTTG

8/8

Figure 4

