Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2889980 A1 2014/05/08

(21) 2 889 980

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2013/10/22
(87) **Date publication PCT/PCT Publication Date:** 2014/05/08
(85) **Entrée phase nationale/National Entry:** 2015/04/29
(86) **N° demande PCT/PCT Application No.:** IB 2013/059525
(87) **N° publication PCT/PCT Publication No.:** 2014/068438
(30) **Priorité/Priority:** 2012/10/30 (US61/720,136)

(51) **Cl.Int./Int.Cl.** ***C12N 15/29*** (2006.01),
***C12P 7/64*** (2006.01)

(71) **Demandeur/Applicant:**
AGRESEARCH LIMITED, NZ

(72) **Inventeurs/Inventors:**
ROBERTS, NICHOLAS JOHN, NZ;
CURRAN, AMY CHRISTINA, US;
WINICHAYAKUL, SOMRUTAI, NZ;
ROLDAN, MARISSA, NZ;
SCOTT, RICHARD WILLIAM, NZ

(74) **Agent:** SIM & MCBURNEY

(54) **Titre : NOUVEAUX POLYNUCLEOTIDES A ACTIVITE ACYLTRANSFERASE, POLYPEPTIDES ET PROCEDES D'UTILISATION**
(54) **Title: NOVEL ACYLTRANSFERASE POLYNUCLEOTIDES, POYPEPTIDES, AND METHODS OF USE**

(57) **Abrégé/Abstract:**
The invention provides a novel DGAT1 protein with improved properties over known DGAT proteins, particularly known DGAT1 proteins from plants. The novel DGAT1 protein of the invention can be expressed in cells to increase cellular lipid accumulation. Expression of the DGAT1 protein of the invention in cells results in a higher level of lipid than any of several other plant DGAT1 proteins tested by the applicants. The invention provides polynucleotides encoding the novel DGAT1 protein of SEQ ID NO:39, constructs, cells, plant, plant parts and progeny comprising the polynucleotides, and methods of use of the polynucleotides and polypeptides of the invention.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 May 2014 (08.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/068438 A3

(51) **International Patent Classification**:
*C12N 15/29* (2006.01) *C12P 7/64* (2006.01)

(21) **International Application Number**:
PCT/IB2013/059525

(22) **International Filing Date**:
22 October 2013 (22.10.2013)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/720,136 30 October 2012 (30.10.2012) US

(71) **Applicant**: **AGRESEARCH LIMITED** [NZ/NZ]; 5th Floor, Tower Block, Ruakura Research Centre, East Street, Hamilton (NZ).

(72) **Inventors**: **ROBERTS, Nicholas John**; 97 McKAY LINE, RD 7, Feilding, 4777 (NZ). **CURRAN, Amy Christina**; 3121 Haidas Avenue, San Diego, California 92117 (US). **WINICHAYAKUL, Somrutai**; 26 Geneva Terrace, Kelvin Grove, Palmerston North, 4414 (NZ). **ROLDAN, Marissa**; 12 Brooklyn Heights Drive, Kelvin Grove, Palmerston North, 4414 (NZ). **SCOTT, Richard William**; 25 Colombo Street, Palmerston North, 4410 (NZ).

(74) **Agents**: **FOOTE, Humphrey, C** et al.; A J Park, Level 22, State Insurance Tower, 1 Willis Street, P O Box 949, Wellington, 6015 (NZ).

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

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

**Published**:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

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

(88) **Date of publication of the international search report**:
24 July 2014

WO 2014/068438 A3

(54) **Title**: NOVEL ACYLTRANSFERASE POLYNUCLEOTIDES, POLYPEPTIDES, AND METHODS OF USE

(57) **Abstract**: The invention provides a novel DGAT1 protein with improved properties over known DGAT proteins, particularly known DGAT1 proteins from plants. The novel DGAT1 protein of the invention can be expressed in cells to increase cellular lipid accumulation. Expression of the DGAT1 protein of the invention in cells results in a higher level of lipid than any of several other plant DGAT1 proteins tested by the applicants. The invention provides polynucleotides encoding the novel DGAT1 protein of SEQ ID NO:39, constructs, cells, plant, plant parts and progeny comprising the polynucleotides, and methods of use of the polynucleotides and polypeptides of the invention.

# Novel acyltransferase polynucleotides, poypeptides, and methods of use

## TECHNICAL FIELD

The invention relates to compositions and methods for the manipulation of cellular lipid production and/or cellular lipid profile.

## BACKGROUND

Plant oil is an economically important product not only due to its broad utilization in the food industry and as a component of feed ingredients but it also has a wide range of applications as biofuels or in the manufacture of various nutraceutical and industrial products. Within the plant itself, oil is essential to carry out a number of metabolic processes which are vital to growth and development particularly during seed germination and early plant growth stages. Considering its value, there is a growing research interest within the biotechnology field to improve plant oil production and make the supply more sustainable.

The major component of plant oil is triacylglyceride (TAG). It is the main form of storage lipid in oil seeds and the primary source of energy for seed germination and seedling development. TAG biosynthesis via the Kennedy pathway involves sequential acylation steps starting from the precursor *sn*-glycerol-3-phosphate (G3P). Firstly, G3P is esterified by an acyl-CoA to form *lyso*phosphatidic acid (LPA) in a reaction catalyzed by glycerol-3-phosphate acyltransferase (GPAT, EC 2.3.1.15). This is followed by a second acylation step catalyzed by *lyso*phosphatidic acid acyltransferase (LPAT; EC 2.3.1.51) forming phosphatidic acid (PA), a key intermediate in the biosynthesis of glycerolipids. The PA is then dephosphorylated by the enzyme phosphatidic acid phosphatase (PAP; EC3.1.3.4) to release the immediate precursor for TAG, the *sn*- 1,2-diacylglycerol (DAG). Finally, DAG is acylated in the *sn-3* position by the enzyme diacylglycerol acyltransferase (DGAT; EC 2.3.1.20) to form TAG.

Since this last catalytic action is the only unique step in TAG biosynthesis, DGAT is termed as the committed triacylglycerol-forming enzyme. As DAG is located at the branch point between TAG and membrane phospholipid biosyntheses, DGAT potentially plays a decisive role in regulating the formation of TAG in the glycerolipid synthesis pathway (Lung and Weselake, 2006, Lipids. Dec 2006;41(12):1073-88). There are two different families of DGAT proteins. The first family of DGAT proteins ("DGAT1") is related to the acyl-coenzyme A:cholesterol acyltransferase ("ACAT") and has been desbried in the U.A. at. 6,100,077 and 6,344,548. A

second family of DGAT proteins ("DGAT2") is unrelated to the DGAT1 family and is described in PCT Patention Publication WO 2004/011671 published Feb. 5, 2004. Other references to DGAT genes and their use in plants include PCT Publication Nos. WO2004/011,671, WO1998/055,631, and WO2000/001,713, and US Patent Publication No. 20030115632.

DGAT1 is typically the major TAG synthesising enzyme in both the seed and senescing leaf (Kaup *et al.*, 2002, Plant Physiol. 129(4):1616-26; for reviews see Lung and Weselake 2006, Lipids. 41(12):1073-88; Cahoon *et al.*, 2007, Current Opinion in Plant Biology. 10:236-244; and Li *et al.*, 2010, Lipids. 45:145-157).

Raising the yield of oilseed crops (canola, sunflower, safflower, soybean, corn, cotton, linseed, flax etc) has been a major target for the agricultural industry for decades. Many approaches (including traditional and mutational breeding as well as genetic engineering) have been tried, typically with modest success (Xu *et al.*, 2008, Plant Biotechnol J., 6:799-818 and references therein).

Although liquid biofuels offer considerable promise the reality of utilising biological material is tempered by competing uses and the quantities available. Consequently, engineering plants and microorganisms to address this is the focus of multiple research groups; in particular the accumulation of triacylglcerol (TAG) in vegetative tissues and oleaginous yeasts and bacteria (Fortman *et al.*, 2008, Trends Biotechnol 26, 375-381; Ohlrogge *et al.*, 2009, Science 324, 1019-1020). TAG is a neutral lipid with twice the energy density of cellulose and can be used to generate biodiesel a high energy density desirable biofuel with one of the simplest and most efficient manufacturing processes. Engineering TAG accumulation in leaves has so far resulted in a 5-20 fold increase over WT utilising a variety of strategies which includes: the over-expression of seed development transcription factors (LEC1, LEC2 and WRI1); silencing of APS (a key gene involved in starch biosynthesis); mutation of CGI-58 (a regulator of neutral lipid accumulation); and upregulation of the TAG synthesising enzyme DGAT (diacylglycerol O acyltransferase, EC 2.3.1.20) in plants and also in yeast (Andrianov *et al.*, 2009, Plant Biotech J 8, 1-11; Mu *et al.*, 2008, Plant Physiol 148, 1042-1054; Sanjaya *et al* 2011, Plant Biotech J 9, 874-883; Santos-Mendoza *et al.*, 2008, Plant J 54, 608-620; James *et al.*, 2010, Proc Natl Acad Sci U S A 107, 17833–17838; Beopoulos *et al.*, 2011, Appl Microbiol Biotechnol 90, 1193-1206; Bouvier-Navé *et al.*, 2000, Eur J Biochem 267, 85-96; Durrett *et al.*, 2008, Plant J 54, 593-607. However, it has been acknowledged that to achieve further increases in TAG, preventing its catabolism may

be crucial within non oleaginous tissues and over a range of developmental stages (Yang and Ohlrogge, 2009, Plant Physiol 150, 1981–1989.

Positively manipulating the yield and quality of triacylglycderides (TAG) in eukaryotes is difficult to achieve. The enzyme diacylglycerol-*O*-acyltransferase (DGAT) has the lowest specific activity of the Kennedy pathway enzymes and is regarded as a 'bottleneck' in TAG synthesis.

Attempts have been made previously to improve DGAT1 by biotechnological methods, with limited success. For example Nykiforuk *et al.*, (2002, Biochimica et Biophysica Acta 1580:95-109) reported N-terminal truncation of the *Brassica napus* DGAT1 but reported approximately 50% lower activity. McFie *et al.*, (2010, JBC., 285:37377-37387) reported that N-terminal truncation of the mouse DGAT1 resulted in increased specific activity of the enzyme, but also reported a large decline in the level of protein that accumulated.

Xu *et al.*, (2008, Plant Biotechnology Journal, 6:799-818) recently identified a consensus sequence (X-Leu-X-Lys-X-X-Ser-X-X-X-Val) within *Tropaeolum majus* (garden nasturtium) DGAT1 (TmDGAT1) sequences as a targeting motif typical of members of the SNF1-related protein kinase-1 (SnRK1) with Ser being the residue for phosphorylation. The SnRK1 proteins are a class of Ser/Thr protein kinases that have been increasingly implicated in the global regulation of carbon metabolism in plants, e.g. the inactivation of sucrose phosphate synthase by phosphorylation (Halford & Hardie 1998, Plant Mol Biol. 37:735-48. Review). Xu *et al.*, (2008, Plant Biotechnology Journal, 6:799-818) performed site-directed mutagenesis on six putative functional regions/motifs of the TmDGAT1 enzyme. Mutagenesis of a serine residue (S197) in a putative SnRK1 target site resulted in a 38%–80% increase in DGAT1 activity, and over-expression of the mutated TmDGAT1 in Arabidopsis resulted in a 20%–50% increase in oil content on a per seed basis.

It would be beneficial to provide improved forms of DGAT1, which overcome one or more of the deficiencies in the prior art, and which can be used to increase cellular oil production.

It is an object of the invention to provide an improved DGAT1 protein and methods for its use to increase cellular lipid production and/or at least to provide the public with a useful choice.

## SUMMARY OF THE INVENTION

The inventors provide a novel DGAT1 protein with improved properties over known DGAT1 proteins, particularly known DGAT1 proteins from plants. The novel DGAT1 protein of the invention can be expressed in cells to increase cellular lipid accumulation. Expression of the DGAT1 protein of the invention in cells results in a higher level of lipid than any of several other plant DGAT1 proteins tested by the applicants.

*Polynucleotide encoding a polypeptide*

In the first aspect the invention provides an isolated polynucleotide encoding a DGAT1 polypeptide comprising the sequence of SEQ ID NO:39 (ZmDGAT1-long) or a variant or fragment thereof.

In one embodiment the variant has at least 70% identity to SEQ ID NO:39. In a further embodiment the variant has DGAT1 activity.

In a further embodiment the DGAT1 polypeptide has higher DGAT1 activity than at least one other DGAT1 protein.

In one embodiment the DGAT1 polypeptide has at least 5%, more preferably at least 10%, more preferably at least 15%, more preferably at least 20%, more preferably at least 25%, more preferably at least 30%, more preferably at least 35%, more preferably at least 40%, more preferably at least 45%, more preferably at least 50% higher DGAT1 activity relative to the at least one other DGAT1 protein.

Preferably the at least one other DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In one embodiment the DGAT1 polypeptide has the higher DGAT1 activity when expressed in a cell.

In a further embodiment the DGAT1 polypeptide has higher DGAT1 activity than any previously known DGAT1 protein.

In one embodiment the DGAT1 polypeptide has the higher DGAT1 activity when expressed in a cell.

In one embodiment the DGAT1 polypeptide has at least 5%, more preferably at least 10%, more preferably at least 15%, more preferably at least 20%, more preferably at least 25%, more preferably at least 30%, more preferably at least 35%, more preferably at least 40%, more preferably at least 45%, more preferably at least 50% higher DGAT1 activity than any previously known DGAT1 protein.

In one embodiment the DGAT1 polypeptide has the higher DGAT1 activity when expressed in a cell.

In a further embodiment the polypeptide of the invention has altered substrate specificity relative to at least one other DGAT1 protein.

In one embodiment the DGAT1 polypeptide has the altered substrate specificity when expressed in a cell.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the polypeptide of the invention, when expressed in the cell, has altered substrate specificity relative to any previously known plant DGAT1 protein.

In one embodiment the DGAT1 polypeptide has the altered substrate specificity when expressed in a cell.

In a further embodiment the DGAT1 protein of the invention is not expressed in naturally occurring plants.

*Polypeptide fragment*

Preferably the fragment comprises at least 50 contiguous amino acids, more preferably at least 100 contiguous amino acids, more preferably at least 150 contiguous amino acids, more preferably at least 200 contiguous amino acids, more preferably at least 250 contiguous amino acids, more preferably at least 300 contiguous amino acids, more preferably at least 350 contiguous amino acids, more preferably at least 400 contiguous amino acids, more preferably at least 450 contiguous amino acids of the polypeptide of the invention.

In one embodiment the fragment of the DGAT1 polypeptide of the invention can confer increased DGAT1 activity when added to at least part of another DGAT1 polypeptide.

*Polynucleotide*

In a further aspect the invention provides an isolated polynucleotide comprising the sequence of SEQ ID NO:10 (ZmDGAT1-long) or a variant or fragment thereof.

In one embodiment the variant has at least 70% identity to SEQ ID NO:10. In a further embodiment the variant encodes a polypeptide with DGAT1 activity.

*Polynucleotide fragment*

In a preferred embodiment, the fragment of the polynucleotide of the invention, encodes a fragment of the polypeptide of the invention.

*Polypeptide*

In a further aspect the invention provides a polypeptide with the sequence of SEQ ID NO:39 (ZmDGAT1-long) or a variant or fragment thereof.

In one embodiment the variant has at least 70% identity to SEQ ID NO:39. In a further embodiment the variant has DGAT1 activity.

In a further embodiment the DGAT1 polypeptide has higher DGAT1 activity than at least one other DGAT1 protein.

In one embodiment the DGAT1 polypeptide has at least 5%, more preferably at least 10%, more preferably at least 15%, more preferably at least 20%, more preferably at least 25%, more preferably at least 30%, more preferably at least 35%, more preferably at least 40%, more preferably at least 45%, more preferably at least 50% higher DGAT1 activity relative to the at least one other DGAT1 protein.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the DGAT1 polypeptide has higher DGAT1 activity than any previously known DGAT1 protein.

In a further embodiment the polypeptide of the invention, when expressed in the cell, has altered substrate specificity relative to at least one other DGAT1 protein.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the polypeptide of the invention, when expressed in the cell, has altered substrate specificity relative to any previously known plant DGAT1 protein.

Preferably the fragment comprises at least 50 contiguous amino acids, more preferably at least 100 contiguous amino acids, more preferably at least 150 contiguous amino acids, more preferably at least 200 contiguous amino acids, more preferably at least 250 contiguous amino acids, more preferably at least 300 contiguous amino acids, more preferably at least 350 contiguous amino acids, more preferably at least 400 contiguous amino acids, more preferably at least 450 contiguous amino acids of the polypeptide of the invention.

In one embodiment the fragment of the DGAT1 polypeptide of the invention can confer increased DGAT1 activity when added to at least part of an other DGAT1 polypeptide.

*Construct*

In a further embodiment the invention provides a genetic construct comprising a polynucleotide of the invention.

*Cells*

In a further embodiment the invention provides a cell comprising a polynucleotide of the invention. Preferably the cell, or its predecessor, is transformed to comprise the polynucleotide of the invention.

In a further embodiment the invention provides a cell comprising a genetic construct of the invention.

In a preferred embodiment the cell expresses the polynucleotide of the invention.

In a preferred embodiment the cell expresses the polypeptide of the invention.

In a preferred embodiment the cell, or its predecessor, is transformed or genetically modified to expresses the polynucleotide or polypeptide of the invention.

In one embodiment the polypeptide of the invention, when expressed in the cell, has increased DGAT1 activity relative to at least one other DGAT1 protein.

In one embodiment the DGAT1 polypeptide has at least 5%, more preferably at least 10%, more preferably at least 15%, more preferably at least 20%, more preferably at least 25%, more preferably at least 30%, more preferably at least 35%, more preferably at least 40%, more preferably at least 45%, more preferably at least 50% higher DGAT1 activity relative to the at least one other DGAT1 protein.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the polypeptide of the invention, when expressed in the cell, has increased DGAT1 activity relative to any previously known plant DGAT1 protein.

In a further embodiment the cell produces more lipid than does a control cell.

In one embodiment the cell produces at least 5% more, preferably at least 10% more, preferably at least 15% more, preferably at least 20% more, preferably at least 25% more, preferably at least 30% more, preferably at least 35% more, preferably at least 40% more, preferably at least 45% more, preferably at least 50% more, preferably at least 55% more, preferably at least 60% more, preferably at least 65% more, preferably at least 70% more, preferably at least 75% more, preferably at least 80% more, preferably at least 85% more, preferably at least 90% more, preferably at least 95% more preferably at least 100% more, preferably at least 105% more, preferably at least 110% more, preferably at least 115% more, preferably at least 120% more, preferably at least 125% more, preferably at least 130% more, preferably at least 135% more, preferably at least 140% more, preferably at least 145% more, preferably at least 150% more lipid than does a control cell.

In a further embodiment the polypeptide of the invention, when expressed in the cell, has altered substrate specificity relative to at least one other DGAT1 protein.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the polypeptide of the invention, when expressed in the cell, has altered substrate specificity relative to any previously known plant DGAT1 protein.

In a further embodiment the cell has an altered lipid profile relative to a control cell.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered relative to that in a control cell.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least

8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control cell.

In a further embodiment the altered lipid profile has a proportion of 16:0 in the triacylglycerol in the range 6% to 16%. In this embodiment the proportion of 16:0 in the triacylglycerol is altered within the range 6% to 16%.

In a further embodiment the proportion of 18:0 in the triacylglycerol is altered relative to that in a control cell.

In one embodiment the proportion of 18:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control cell.

In a further embodiment the altered lipid profile has a proportion of 18:0 in the triacylglycerol in the range 7% to 15%. In this embodiment the proportion of 18:0 in the triacylglycerol is altered within the range 7% to 15%.

In a further embodiment the proportion of 18:1 in the triacylglycerol is altered relative to that in a control cell.

In one embodiment the proportion of 18:1 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more

preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control cell.

In a further embodiment the altered lipid profile has a proportion of 18:1 in the triacylglycerol in the range 39% to 55%. In this embodiment the proportion of 18:1 in the triacylglycerol is altered within the range 39% to 55%.

The control cell may be any cell of the same type that is not transformed with the polynucleotide, or construct, of the invention to express the polypeptide of the invention. The control cell may also be transformed with an "empty" vector, wherein the empty vector does not include an insert sequence corresponding to a polynucleotide of the invnetion or expressing a polypeptide of the invention.

In one embodiment the control cell is an untransformed cell. In a further embodiment the control cell is transformed cell to express the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short). In a further embodiment the control cell is transformed cell to express any previously known plant DGAT1 protein.

*Cells also transformed to express an oleosin*

In one embodiment the cell is also transformed to express at least one of: an oleosin, steroleosin, caloleosin, polyoleosin, and an oleosin including at least one artificially introduced cysteine (WO2011/053169).

In one embodiment the cell is a plant cell.

*Plant*

In a further embodiment the invention provides a plant comprising a polynucleotide of the invention. Preferably the plant, or its predecessor, is transformed to comprise the polynucleotide of the invention.

In a further embodiment the invention provides a plant comprising a genetic construct of the invention.

Ina further embodiment the the invention provides a plant comprising a plant cell of the invention.

In a preferred embodiment the plant expresses the polynucleotide of the invention.

In a preferred embodiment the plant expresses the polypeptide of the invention.

In a preferred embodiment the plant, or its predecessor, is transformed or genetically modified to expresses the polynucleotide or polypeptide of the invention.

In one embodiment the polypeptide of the invention, when expressed in the plant, has increased DGAT1 activity relative to at least one other DGAT1 protein.

In one embodiment the DGAT1 protein of the invention has at least 5%, more preferably at least 10%, more preferably at least 15%, more preferably at least 20%, more preferably at least 25%, more preferably at least 30%, more preferably at least 35%, more preferably at least 40%, more preferably at least 45%, more preferably at least 50% higher DGAT1 activity relative to the at least one other DGAT1 protein.

Preferably the at least one DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

In a further embodiment the polypeptide of the invention, when expressed in the plant, has increased DGAT1 activity relative to any previously know plant DGAT1 protein.

In a further embodiment the plant produces more lipid, in at least one of its tissues or parts, than does the equivalent tissue or part in a control plant.

In one embodiment the plant produces at least 5% more, preferably at least 10% more, preferably at least 15% more, preferably at least 20% more, preferably at least 25% more, preferably at least 30% more, preferably at least 35% more, preferably at least 40% more,

preferably at least 45% more, preferably at least 50% more, preferably at least 55% more, preferably at least 60% more, preferably at least 65% more, preferably at least 70% more, preferably at least 75% more, preferably at least 80% more, preferably at least 85% more, preferably at least 90% more, preferably at least 95% more preferably at least 100% more, preferably at least 105% more, preferably at least 110% more, preferably at least 115% more, preferably at least 120% more, preferably at least 125% more, preferably at least 130% more, preferably at least 135% more, preferably at least 140% more, preferably at least 145% more, preferably at least 150% more lipid than does a control plant.

In one embodiment the tissue is a vegetative tissue. In one embodiment the part is a leaf. In a further embodiment the part is a root. In a further embodiment the part is a tuber. In a further embodiment the part is a corm. In a further embodiment the part is a stalk. In a further embodiment the part is a stalk of a monoct plant. In a further embodiment the part is from stover (the dried stalks and leaves of a field crop). Stover is often used as animal fodder, for example, after the grain of the crop has been harvested

In a preferred embodiment the tissue is seed tissue. In a preferred embodiment the part is a seed. In a preferred embodiment the tissue is endosperm tissue.

In a further embodiment the plant as a whole produces more lipid than does the control plant as a whole.

In a further embodiment the plant has an altered lipid profile, in at least one of its tissues or parts, relative to a control plant.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered relative to that in a control plant.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more

preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control plant.

In a further embodiment the altered lipid profile has a proportion of 16:0 in the triacylglycerol in the range 6% to 16%. In this embodiment the proportion of 16:0 in the triacylglycerol is altered within the range 6% to 16%.

In a further embodiment the proportion of 18:0 in the triacylglycerol is altered relative to that in a control plant.

In one embodiment the proportion of 18:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control plant.

In a further embodiment the altered lipid profile has a proportion of 18:0 in the triacylglycerol in the range 7% to 15%. In this embodiment the proportion of 18:0 in the triacylglycerol is altered within the range 7% to 15%.

In a further embodiment the proportion of 18:1 in the triacylglycerol is altered relative to that in a control plant.

In one embodiment the proportion of 18:1 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control plant.

In a further embodiment the altered lipid profile has a proportion of 18:1 in the triacylglycerol in the range 39% to 55%. In this embodiment the proportion of 18:1 in the triacylglycerol is altered within the range 39% to 55%.

In one embodiment the tissue is a vegetative tissue. In one embodiment the part is a leaf. In a further embodiment the part is a root. In a further embodiment the part is a tuber. In a further embodiment the part is a corm. In a further embodiment the part is a stalk. In a further embodiment the part is a stalk of a monoct plant. In a further embodiment the part is a stovum (stalk and leaf blade).

In a preferred embodiment the tissue is seed tissue. In a preferred embodiment the part is a seed. In a preferred embodiment the tissue is endosperm tissue.

In a further embodiment the plant as a whole has an altered lipid profile relative to the control plant as a whole.

The control plant may be any plant of the same type that is not transformed with the polynucleotide, or construct, of the invention to express the polypeptide of the invention. The control plant may also be transformed with an "empty" vector, wherein the empty vector does not include an insert sequence corresponding to a polynucleotide of the invnetion or expressing a polypeptide of the invention.

In one embodiment the control plant is an untransformed plant. In a further embodiment the control cell is transformed cell to express the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

*Plant also transformed to express an oleosin*

In one embodiment the plant is also transformed to express at least one of: an oleosin, a steroleosin, a caloleosin, a polyoleosin, and an oleosin including at least one artificially introduced cysteine (WO 2011/053169).

*Plant parts*

In a further embodiment the invention provides a part, propagule or progeny of a plant of the invention.

In a preferred embodiment the part, propagule or progeny comprises at least one of a polynucleotide, construct, or polypeptide of the invention. Preferably the part, propagule or progeny, or its predecessor plant, is transformed to comprise the polynucleotide of the invention.

In a preferred embodiment the part, propagule or progeny expresses at least one of: a polynucleotide and a polypeptide of the invention.

In a preferred embodiment the part, propagule or progeny expresses a polypeptide of the invention.

In a preferred embodiment the part, propagule or progeny, or its predecessor plant, is transformed or genetically modified to expresses the polynucleotide or polypeptide of the invention.

In a further embodiment the part, propagule or progeny produces more lipid than does a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In one embodiment the part, propagule or progeny produces at least 5% more, preferably at least 10% more, preferably at least 15% more, preferably at least 20% more, preferably at least 25% more, preferably at least 30% more, preferably at least 35% more, preferably at least 40% more, preferably at least 45% more, preferably at least 50% more, preferably at least 55% more, preferably at least 60% more, preferably at least 65% more, preferably at least 70% more, preferably at least 75% more, preferably at least 80% more, preferably at least 85% more, preferably at least 90% more, preferably at least 95% more preferably at least 100% more, preferably at least 105% more, preferably at least 110% more, preferably at least 115% more, preferably at least 120% more, preferably at least 125% more, preferably at least 130% more, preferably at least 135% more, preferably at least 140% more, preferably at least 145% more,

preferably at least 150% more lipid than does a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In a further embodiment the part, propagule or progeny has an altered lipid profile relative to a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered relative to that in a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In one embodiment the proportion of 16:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In a further embodiment the altered lipid profile has a proportion of 16:0 in the triacylglycerol in the range 6% to 16%. In this embodiment the proportion of 16:0 in the triacylglycerol is altered within the range 6% to 16%.

In a further embodiment the proportion of 18:0 in the triacylglycerol is altered relative to that in control part, propagule or progeny, or part, propagule or progeny of a control plant.

In one embodiment the proportion of 18:0 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in control part, propagule or progeny, or part, propagule or progeny of a control plant.

In a further embodiment the altered lipid profile has a proportion of 18:0 in the triacylglycerol in the range 7% to 15%. In this embodiment the proportion of 18:0 in the triacylglycerol is altered within the range 7% to 15%.

In a further embodiment the proportion of 18:1 in the triacylglycerol is altered relative to that in a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In one embodiment the proportion of 18:1 in the triacylglycerol is altered by at least 1%, preferably at least 2%, more preferably at least 3%, more preferably at least 4%, more preferably at least 5%, more preferably at least 6%, more preferably at least 7%, more preferably at least 8%, more preferably at least 9%, more preferably at least 10%, more preferably at least 11%, more preferably at least 12%, more preferably at least 13%, more preferably at least 14%, more preferably at least 15%, more preferably at least 16%, more preferably at least 17%, more preferably at least 18%, more preferably at least 19%, more preferably at least 20%, relative to that in a control part, propagule or progeny, or part, propagule or progeny of a control plant.

In a further embodiment the altered lipid profile has a proportion of 18:1 in the triacylglycerol in the range 39% to 55%. In this embodiment the proportion of 18:1 in the triacylglycerol is altered within the range 39% to 55%.

The control plant may be any plant of the same type that is not transformed with the polynucleotide, or construct, of the invention to express the polypeptide of the invention.

In one embodiment the control plant is an untransformed plant. In a further embodiment the control plant is transformed plant to express the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

Preferably the control the part, propagule or progeny is from a control plant as described above.

In one embodiment the part is from a vegetative tissue. In one embodiment the part is a leaf. In a further embodiment the part is a root. In a further embodiment the part is a tuber. In a further embodiment the part is a corm. In a further embodiment the part is a stalk. In a further embodiment the part is a stalk of a monocot plant. In a further embodiment the part is a stovum (stalk and leaf blade).

In a further embodiment the part is from a reproductive tissue. In a further embodiment the part is a seed. In a preferred embodiment the part is from or includes endosperm tissue.

*Animal feed*

In a further aspect the invention provides an animal feedstock comprising at least one of a polynucleotide, polypeptide,construct, cell, plant cell, plant part, plant, propagule and progeny of the invention.

*Biofuel feedstock*

In a further aspect the invention provides a biofuel feedstock comprising at least one of a polynucleotide, polypeptide, construct, cell, plant cell, plant part, plant, propagule and progeny of the invention.

*Lipid*

In one embodiment the lipid is an oil. In a further embodiment the lipid is triacylglycerol (TAG)

*Method for producing lipid/ oil*

In a further aspect the invention provides a method for producing a lipid, the method comprising growing a cell, plant cell or plant that is transformed, or genetically modified, to express and polynucleotide or polypeptide of the invention wherein the plant produces oil through the activity of the expressed polypeptide.

In one embodiment the cell, plant cell or plant produces the lipid as a result of the DGAT1 activity of the polypeptide.

In a further aspect the invention provides a method for producing lipid, the method comprising extracting lipid from at least one of a cell, plant cell, plant, plant part, propagule and progeny of the invention.

In one embodiment the lipid is triacylglycerol (TAG).

In a further embodiment the lipid is processed into at least one of:

a) a fuel,

b) an oleochemical,

c) a nutritional oil,

d) a cosmetic oil,

e) a polyunsaturated fatty acid (PUFA), and

f) a combination of any of a) to e).

In a further aspect the invention provides a method for producing lipid, the method comprising expressing a DGAT1 protein of the invention in a cell, plant cell or plant.

In a preferred embodiment expressing the DGAT1 protein of the invention in the plant leads production of the lipid in the cell, plant cell or plant.

In one embodiment the method includes the step of transforming a cell, plant cell or plant with a polynucleotide of the invention encoding the DGAT1 protein.

In a further embodiment the method includes the step of extracting the lipid from the cell, plant cell, or plant, or from a part, propagule or progeny of the plant.

In one embodiment the lipid is an oil. In a further embodiment the lipid is triacylglycerol (TAG)

In a further embodiment the lipid is processed into at least one of:

a) a fuel,

b) an oleochemical,

c) a nutritional oil,

d) a cosmetic oil,

e) a polyunsaturated fatty acid (PUFA), and

f) a combination of any of a) to e).

## DETAILED DESCRIPTION OF THE INVENTION

In this specification where reference has been made to patent specifications, other external documents, or other sources of information, this is generally for the purpose of providing a context for discussing the features of the invention. Unless specifically stated otherwise, reference to such external documents is not to be construed as an admission that such documents, or such sources of information, in any jurisdiction, are prior art, or form part of the common general knowledge in the art.

The term "comprising" as used in this specification means "consisting at least in part of". When interpreting each statement in this specification that includes the term "comprising", features other than that or those prefaced by the term may also be present. Related terms such as "comprise" and "comprises" are to be interpreted in the same manner. In some embodiments, the term "comprising" (and related terms such as "comprise and "comprises") can be replaced by "consisting of" (and related terms "consist" and "consists").

The term "DGAT1" as used herein means acyl CoA: diacylglycerol acyltransferase (EC 2.3.1.20) DGAT1 is typically the major TAG synthesising enzyme in both the seed and senescing leaf (Kaup *et al.*, 2002, Plant Physiol. 129(4):1616-26; for reviews see Lung and Weselake 2006, Lipids. Dec 2006;41(12):1073-88; Cahoon *et al.*, 2007, Current Opinion in Plant Biology. 10:236-244; and Li *et al.*, 2010, Lipids. 45:145-157).

DGAT1 contains approximately 500 amino acids and has 10 predicted transmembrane domains whereas DGAT2 has only 320 amino acids and is predicted to contain only two transmembrane domains; both proteins were also predicted to have their N- and C-termini located in the cytoplasm (Shockey *et al.*, 2006, Plant Cell 18:2294-2313). Both *DGAT1* and *DGAT2* have orthologues in animals and fungi and are transmembrane proteins located in the ER.

In most dicotyledonous plants *DGAT1* & *DGAT2* appear to be single copy genes whereas there are typically two versions of each in the grasses which presumably arose during the duplication of the grass genome (Salse *et al.*, 2008, Plant Cell, 20:11-24).

The phrase "increased DGAT1 activity" means increased specific activity relative to that of the first and/or DGAT1 protein.

An art skilled worker would know how to test the "specific activity" of a DGAT1 protein or variant thereof of the invnetion. This may typically be done by isolating, enriching and quantifying the recombinant DGAT1 then using this material to determine either the rate of triaclyglyceride formation and/or the disappearance of precursor substrates (including various forms of acyl-CoA and DAG) as per Xu *et al.*, (2008), Plant Biotechnology Journal. 6:799-818.

*Lipid*

In one embodiment the lipid is an oil. In a further embodiment the oil is triacylglycerol (TAG)

*Lipid production*

In certain embodiments the cell, cells, tissues, plants and plant parts of the invention produces more lipid than control cells, tissues, plants and plant parts.

Those skilled in the art are well aware of methods for measuring lipid production. This may typically be done by quantitative fatty acid methyl ester gas chromatography mass spectral analysis (FAMES GC-MS). Suitable methods are also described in the examples section of this specification.

*Substrate specificity*

In certain embodiments, the polypeptides of the invention have altered substrate specificity relative to other DGAT1 proteins. Plant DGAT1 proteins are relatively promiscuous in terms of the fatty acid substrates and DAG species they are capable of utilisting to generate TAG. As such they can be considered to have relatively low substrate specificity. However, this can be modified such that certain fatty acids become a preferred substrate over others. This leads to an increase in the proportions of the preferred fatty acids in the TAG and decreases in the proportions of the non preferred fatty acid species. Substrate specificity can be determined by *in vitro* quantitiative analysis of TAG production following the addition of specific and known

quantities of purified substrates to known quantities of recombinant DGAT, as per Xu et al., (2008), Plant Biotechnology Journal. 6:799-818.

*Lipid profile*

In a further embodiment the cell, cells, tissues, plants and plant parts of the invention have an altered lipid profile relative to the control cells, tissues, plants and plant parts.

Those skilled in the art are well aware of methods for assessing lipid profile. This may involve assessing the proportion or percentage of at least one of the 16:0, 16:1, 18:0, 18:1c9 fatty acid species present in the lipid. This may typically be done by fatty acid methyl ester (FAME) analysis (Browse *et al.*, 1986, Anal. Biochem. 152, 141-145). Suitable methods are also described in the examples section of this specification.

*Cells*

The DGAT1 polypeptide of the invention, or as used in the methods of the invention, may be expressed in any cell type.

In one embodiment the cell is a prokaryotic cell. In a further embodiment the cell is a eukaryotic cell. In one embodiment the cell is selected from a bacterial cell, a yeast cell, a fungal cell, an insect cell, algal cell, and a plant cell. In one embodiment the cell is a bacterial cell. In a further embodiment the cell is a yeast cell. In one embodiment the yeast cell is a *S. ceriviseae* cell. In further embodiment the cell is a fungal cell. In further embodiment the cell is an insect cell. In further embodiment the cell is an algal cell. In a further embodiment the cell is a plant cell.

In one embodiment the cell is a non-plant cell. In one embodiment the non-plant is selected from *E. coli, P. pastoris, S. ceriviseae, D. salina* and *C. reinhardtii.* In a further embodiment the non-plant is selected from *P. pastoris, S. ceriviseae, D. Salina* and *C. reinhardtii.*

In one embodiment the cell is a microbial cell. In another embodiment, the microbial cell is an algal cell of the division of Chlorophyta (green algae), Rhodophyta (red algae), Phaeophyceae (brown algae), Bacillariophycaeae (diatoms), or Dinoflagellata (dinoflagellates). In another embodiment, the microbial cell is an algal cell of the species *Chlamydomonas, Dunaliella, Botrycoccus, Chlorella, Crypthecodinium, Gracilaria, Sargassum, Pleurochrysis, Porphyridium*, *Phaeodactylum*,

*Haematococcus*, *Isochrysis*, *Scenedesmus*, *Monodus*, *Cyclotella*, *Nitzschia*, or *Parietochloris*. In another embodiment, the algal cell is *Chlamydomonas reinhardtii*. In yet another embodiment, the cell is from the genus *Yarrowia*, *Candida*, *Rhodotorula*, *Rhodosporidium*, *Cryptococcus*, *Trichosporon*, *Lipomyces*, *Pythium*, *Schizochytrium*, *Thraustochytrium*, or *Ulkenia*. In yet another embodiment, the cell is a bacterium of the genus *Rhodococcus*, *Escherichia*, or a cyanobacterium. In yet another embodiment, the cell is a yeast cell. In yet another embodiment, the cell is a synthetic cell.

*Plants*

The variant DGAT1 sequences of the invention may be naturally-occurring DGAT1sequences. Preferably the variant DGAT1 sequences are from plants. In certain embodiments the cells into which the DGAT1 proteins of the invention are expressed are from plants. In other embodiments the DGAT1 proteins of the invention are expressed in plants.

The plant cells, from which the DGAT1 proteins of the invention are derived, the plants from which the plant cells are derived, and the plants in which the DGAT1 proteins of the invention are expressed may be from any plant species.

In one embodiment the plant cell or plant, is derived from a gymnosperm plant species.

In a further embodiment the plant cell or plant, is derived from an angiosperm plant species.

In a further embodiment the plant cell or plant, is derived from a from dicotyledonous plant species.

In a further embodiment the plant cell or plant, is derived from a monocotyledonous plant species.

Other preferred plants are forage plant species from a group comprising but not limited to the following genera: *Zea, Lolium, Hordium, Miscanthus, Saccharum, Festuca*, *Dactylis*, *Bromus*, *Thinopyrum*, *Trifolium*, *Medicago*, *Pheleum*, *Phalaris*, *Holcus*, *Glycine*, *Lotus*, *Plantago* and *Cichorium*.

Other preferred plants are leguminous plants. The leguminous plant or part thereof may encompass any plant in the plant family Leguminosae or Fabaceae. For example, the plants may be selected from forage legumes including, alfalfa, clover; leucaena; grain legumes including, beans, lentils, lupins, peas, peanuts, soy bean; bloom legumes including lupin, pharmaceutical or industrial legumes; and fallow or green manure legume species.

A particularly preferred genus is *Trifolium*. Preferred *Trifolium* species include *Trifolium repens*; *Trifolium arvense*; *Trifolium affine*; and *Trifolium occidentale*. A particularly preferred *Trifolium* species is *Trifolium repens*.

Another preferred genus is *Medicago*. Preferred *Medicago* species include *Medicago sativa* and *Medicago truncatula*. A particularly preferred *Medicago* species is *Medicago sativa*, commonly known as alfalfa.

Another preferred genus is *Glycine*. Preferred *Glycine* species include *Glycine max* and *Glycine wightii* (also known as *Neonotonia wightii*). A particularly preferred *Glycine* species is *Glycine max*, commonly known as soy bean. A particularly preferred *Glycine* species is *Glycine wightii*, commonly known as perennial soybean.

Another preferred genus is *Vigna*. A particularly preferred *Vigna* species is *Vigna unguiculata* commonly known as cowpea.

Another preferred genus is *Mucana*. Preferred *Mucana* species include *Mucana pruniens*. A particularly preferred *Mucana* species is *Mucana pruniens* commonly known as velvetbean.

Another preferred genus is *Arachis*. A particularly preferred *Arachis* species is *Arachis glabrata* commonly known as perennial peanut.

Another preferred genus is *Pisum*. A preferred *Pisum* species is *Pisum sativum* commonly known as pea.

Another preferred genus is *Lotus*. Preferred Lotus species include *Lotus corniculatus*, *Lotus pedunculatus*, *Lotus glabar*, *Lotus tenuis* and *Lotus uliginosus*. A preferred *Lotus* species is *Lotus corniculatus* commonly known as Birdsfoot Trefoil. Another preferred *Lotus* species is *Lotus glabar* commonly known as Narrow-leaf Birdsfoot Trefoil. Another preferred preferred *Lotus* species is *Lotus pedunculatus* commonly known as Big trefoil. Another preferred *Lotus* species is *Lotus tenuis* commonly known as Slender trefoil.

Another preferred genus is *Brassica*. A preferred *Brassica* species is *Brassica oleracea*, commonly known as forage kale and cabbage. A preferred *Brassica* genus is *Camelina*. A preferred *Camelina* species is *Camelina sativa*.

Other preferred species are oil seed crops including but not limited to the following genera: *Brassica*, *Carthumus*, *Helianthus*, *Zea* and *Sesamum*.

A preferred oil seed genera is *Brassica*. A preferred oil seed species is *Brassica napus*.

A preferred oil seed genera is *Brassica*. A preferred oil seed species is *Brassica oleraceae*.

A preferred oil seed genera is *Carthamus*. A preferred oil seed species is *Carthamus tinctorius*.

A preferred oil seed genera is *Helianthus*. A preferred oil seed species is *Helianthus annuus*.

A preferred oil seed genera is *Zea*. A preferred oil seed species is *Zea mays*.

A preferred oil seed genera is *Sesamum*. A preferred oil seed species is *Sesamum indicum*.

A preferred silage genera is *Zea*. A preferred silage species is *Zea mays*.

A preferred grain producing genera is *Hordeum*. A preferred grain producing species is *Hordeum vulgare*.

A preferred grazing genera is *Lolium*. A preferred grazing species is *Lolium perenne*.

A preferred grazing genera is *Lolium*. A preferred grazing species is *Lolium arundinaceum*.

A preferred grazing genera is *Trifolium*. A preferred grazing species is *Trifolium repens*.

A preferred grazing genera is *Hordeum*. A preferred grazing species is *Hordeum vulgare*.

Preferred plants also include forage, or animal feedstock plants. Such plants include but are not limited to the following genera: *Miscanthus*, *Saccharum*, *Panicum*.

A preferred biofuel genera is *Miscanthus*. A preferred biofuel species is *Miscanthus giganteus*.

A preferred biofuel genera is *Saccharum*. A preferred biofuel species is *Saccharum officinarum*.

A preferred biofuel genera is *Panicum*. A preferred biofuel speices is *Panicum virgatum*.

*Plant parts, propagues and progeny*

The term "plant" is intended to include a whole plant, any part of a plant, a seed, a fruit, propagules and progeny of a plant.

The term 'propagule' means any part of a plant that may be used in reproduction or propagation, either sexual or asexual, including seeds and cuttings.

The plants of the invention may be grown and either self-ed or crossed with a different plant strain and the resulting progeny, comprising the polynucleotides or constructs of the invention, and/or expressing the DGAT1 sequences of the invention, also form an part of the present invention.

Preferably the plants, plant parts, propagules and progeny comprise a polynucleotide or construct of the invention, and/or express a DGAT1 sequence of the invention.

*Polynucleotides and fragments*

The term "polynucleotide(s)," as used herein, means a single or double-stranded deoxyribonucleotide or ribonucleotide polymer of any length but preferably at least 15 nucleotides, and include as non-limiting examples, coding and non-coding sequences of a gene, sense and antisense sequences complements, exons, introns, genomic DNA, cDNA, pre-mRNA, mRNA, rRNA, siRNA, miRNA, tRNA, ribozymes, recombinant polypeptides, isolated and purified naturally occurring DNA or RNA sequences, synthetic RNA and DNA sequences, nucleic acid probes, primers and fragments.

A "fragment" of a polynucleotide sequence provided herein is a subsequence of contiguous nucleotides.

The term "primer" refers to a short polynucleotide, usually having a free 3'OH group, that is hybridized to a template and used for priming polymerization of a polynucleotide complementary to the target.

The term "probe" refers to a short polynucleotide that is used to detect a polynucleotide sequence that is complementary to the probe, in a hybridization-based assay. The probe may consist of a "fragment" of a polynucleotide as defined herein.

*Polypeptides and fragments*

The term "polypeptide", as used herein, encompasses amino acid chains of any length but preferably at least 5 amino acids, including full-length proteins, in which amino acid residues are linked by covalent peptide bonds. Polypeptides of the present invention, or used in the methods of the invention, may be purified natural products, or may be produced partially or wholly using recombinant or synthetic techniques.

A "fragment" of a polypeptide is a subsequence of the polypeptide that preferably performs a function/activity of and/or influences three dimensional structure of the polypeptide.

The term "isolated" as applied to the polynucleotide or polypeptide sequences disclosed herein is used to refer to sequences that are removed from their natural cellular environment. The isolated sequence is preferably separated from the sequences that may be found flanking the sequence in its naturally occurring environment. An isolated molecule may be obtained by any method or combination of methods including biochemical, recombinant, and synthetic techniques.

The term "recombinant" refers to a polynucleotide sequence that is removed from sequences that surround it in its natural context and/or is recombined with sequences that are not present in its natural context.

A "recombinant" polypeptide sequence is produced by translation from a "recombinant" polynucleotide sequence.

The term "derived from" with respect to polynucleotides or polypeptides of the invention being derived from a particular genera or species, means that the polynucleotide or polypeptide has the same sequence as a polynucleotide or polypeptide found naturally in that genera or species. The polynucleotide or polypeptide, derived from a particular genera or species, may therefore be produced synthetically or recombinantly.

*Variants*

As used herein, the term "variant" refers to polynucleotide or polypeptide sequences different from the specifically identified sequences, wherein one or more nucleotides or amino acid residues is deleted, substituted, or added. Variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variants may be from the same or from other species and may encompass homologues, paralogues and orthologues. In certain embodiments, variants of the inventive polypeptides and polypeptides possess biological activities that are the same or similar to those of the inventive polypeptides or polypeptides. The term "variant" with reference to polypeptides and polypeptides encompasses all forms of polypeptides and polypeptides as defined herein.

*Polynucleotide variants*

Variant polynucleotide sequences preferably exhibit at least 50%, more preferably at least 51%, more preferably at least 52%, more preferably at least 53%, more preferably at least 54%, more

preferably at least 55%, more preferably at least 56%, more preferably at least 57%, more preferably at least 58%, more preferably at least 59%, more preferably at least 60%, more preferably at least 61%, more preferably at least 62%, more preferably at least 63%, more preferably at least 64%, more preferably at least 65%, more preferably at least 66%, more preferably at least 67%, more preferably at least 68%, more preferably at least 69%, more preferably at least 70%, more preferably at least 71%, more preferably at least 72%, more preferably at least 73%, more preferably at least 74%, more preferably at least 75%, more preferably at least 76%, more preferably at least 77%, more preferably at least 78%, more preferably at least 79%, more preferably at least 80%, more preferably at least 81%, more preferably at least 82%, more preferably at least 83%, more preferably at least 84%, more preferably at least 85%, more preferably at least 86%, more preferably at least 87%, more preferably at least 88%, more preferably at least 89%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, and most preferably at least 99% identity to a sequence of the present invention. Identity is found over a comparison window of at least 20 nucleotide positions, preferably at least 50 nucleotide positions, more preferably at least 100 nucleotide positions, and most preferably over the entire length of a polynucleotide of the invention.

Polynucleotide sequence identity can be determined in the following manner. The subject polynucleotide sequence is compared to a candidate polynucleotide sequence using BLASTN (from the BLAST suite of programs, version 2.2.5 [Nov 2002]) in bl2seq (Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250), which is publicly available from the NCBI website on the World Wide Web at ftp://ftp.ncbi.nih.gov/blast/. The default parameters of bl2seq are utilized except that filtering of low complexity parts should be turned off.

The identity of polynucleotide sequences may be examined using the following unix command line parameters:

bl2seq –i nucleotideseq1 –j nucleotideseq2 –F F –p blastn

The parameter –F F turns off filtering of low complexity sections. The parameter –p selects the appropriate algorithm for the pair of sequences. The bl2seq program reports sequence identity as both the number and percentage of identical nucleotides in a line “Identities = “.

Polynucleotide sequence identity may also be calculated over the entire length of the overlap between a candidate and subject polynucleotide sequences using global sequence alignment programs (e.g. Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). A full implementation of the Needleman-Wunsch global alignment algorithm is found in the needle program in the EMBOSS package (Rice,P. Longden,I. and Bleasby,A. EMBOSS: The European Molecular Biology Open Software Suite, Trends in Genetics June 2000, vol 16, No 6. pp.276-277) which can be obtained from the World Wide Web at http://www.hgmp.mrc.ac.uk/Software/EMBOSS/. The European Bioinformatics Institute server also provides the facility to perform EMBOSS-needle global alignments between two sequences on line at http:/www.ebi.ac.uk/emboss/align/.

Alternatively the GAP program may be used which computes an optimal global alignment of two sequences without penalizing terminal gaps. GAP is described in the following paper: Huang, X. (1994) On Global Sequence Alignment. Computer Applications in the Biosciences 10, 227-235.

A preferred method for calculating polynucleotide % sequence identity is based on aligning sequences to be compared using Clustal X (Jeanmougin *et al.*, 1998, Trends Biochem. Sci. 23, 403-5.)

Polynucleotide variants of the present invention also encompass those which exhibit a similarity to one or more of the specifically identified sequences that is likely to preserve the functional equivalence of those sequences and which could not reasonably be expected to have occurred by random chance. Such sequence similarity with respect to polypeptides may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from the NCBI website on the World Wide Web at ftp://ftp.ncbi.nih.gov/blast/.

The similarity of polynucleotide sequences may be examined using the following unix command line parameters:

bl2seq –i nucleotideseq1 –j nucleotideseq2 –F F –p tblastx

The parameter –F F turns off filtering of low complexity sections. The parameter –p selects the appropriate algorithm for the pair of sequences. This program finds regions of similarity between the sequences and for each such region reports an "E value" which is the expected number of times one could expect to see such a match by chance in a database of a fixed reference size containing random sequences. The size of this database is set by default in the bl2seq program.

For small E values, much less than one, the E value is approximately the probability of such a random match.

Variant polynucleotide sequences preferably exhibit an E value of less than 1 x 10 -6 more preferably less than 1 x 10 -9, more preferably less than 1 x 10 -12, more preferably less than 1 x 10 -15, more preferably less than 1 x 10 -18, more preferably less than 1 x 10 -21, more preferably less than 1 x 10 -30, more preferably less than 1 x 10 -40, more preferably less than 1 x 10 -50, more preferably less than 1 x 10 -60, more preferably less than 1 x 10 -70, more preferably less than 1 x 10 -80, more preferably less than 1 x 10 -90 and most preferably less than 1 x 10-100 when compared with any one of the specifically identified sequences.

Alternatively, variant polynucleotides of the present invention, or used in the methods of the invention, hybridize to the specified polynucleotide sequences, or complements thereof under stringent conditions.

The term "hybridize under stringent conditions", and grammatical equivalents thereof, refers to the ability of a polynucleotide molecule to hybridize to a target polynucleotide molecule (such as a target polynucleotide molecule immobilized on a DNA or RNA blot, such as a Southern blot or Northern blot) under defined conditions of temperature and salt concentration. The ability to hybridize under stringent hybridization conditions can be determined by initially hybridizing under less stringent conditions then increasing the stringency to the desired stringency.

With respect to polynucleotide molecules greater than about 100 bases in length, typical stringent hybridization conditions are no more than 25 to 30° C (for example, 10° C) below the melting temperature (Tm) of the native duplex (see generally, Sambrook *et al.,* Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Ausubel *et al.*, 1987, Current Protocols in Molecular Biology, Greene Publishing,). Tm for polynucleotide molecules greater than about 100 bases can be calculated by the formula Tm = 81. 5 + 0. 41% (G + C-log (Na+). (Sambrook *et al.,* Eds, 1987, Molecular Cloning, A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press; Bolton and McCarthy, 1962, PNAS 84:1390). Typical stringent conditions for polynucleotide of greater than 100 bases in length would be hybridization conditions such as prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65° C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C.

With respect to polynucleotide molecules having a length less than 100 bases, exemplary stringent hybridization conditions are 5 to 10° C below Tm. On average, the Tm of a polynucleotide molecule of length less than 100 bp is reduced by approximately (500/oligonucleotide length)° C.

With respect to the DNA mimics known as peptide nucleic acids (PNAs) (Nielsen *et al.*, Science. 1991 Dec 6;254(5037):1497-500) Tm values are higher than those for DNA-DNA or DNA-RNA hybrids, and can be calculated using the formula described in Giesen *et al.*, Nucleic Acids Res. 1998 Nov 1;26(21):5004-6. Exemplary stringent hybridization conditions for a DNA-PNA hybrid having a length less than 100 bases are 5 to 10° C below the Tm.

Variant polynucleotides of the present invention, or used in the methods of the invention, also encompasses polynucleotides that differ from the sequences of the invention but that, as a consequence of the degeneracy of the genetic code, encode a polypeptide having similar activity to a polypeptide encoded by a polynucleotide of the present invention. A sequence alteration that does not change the amino acid sequence of the polypeptide is a "silent variation". Except for ATG (methionine) and TGG (tryptophan), other codons for the same amino acid may be changed by art recognized techniques, e.g., to optimize codon expression in a particular host organism.

Polynucleotide sequence alterations resulting in conservative substitutions of one or several amino acids in the encoded polypeptide sequence without significantly altering its biological activity are also included in the invention. A skilled artisan will be aware of methods for making phenotypically silent amino acid substitutions (see, e.g., Bowie *et al.*, 1990, Science 247, 1306).

Variant polynucleotides due to silent variations and conservative substitutions in the encoded polypeptide sequence may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from the NCBI website on the World Wide Web at ftp://ftp.ncbi.nih.gov/blast/ via the tblastx algorithm as previously described.

*Polypeptide variants*

The term "variant" with reference to polypeptides encompasses naturally occurring, recombinantly and synthetically produced polypeptides. Variant polypeptide sequences preferably exhibit at least 50%, more preferably at least 51%, more preferably at least 52%, more preferably at least 53%, more preferably at least 54%, more preferably at least 55%, more preferably at least 56%, more preferably at least 57%, more preferably at least 58%, more

preferably at least 59%, more preferably at least 60%, more preferably at least 61%, more preferably at least 62%, more preferably at least 63%, more preferably at least 64%, more preferably at least 65%, more preferably at least 66%, more preferably at least 67%, more preferably at least 68%, more preferably at least 69%, more preferably at least 70%, more preferably at least 71%, more preferably at least 72%, more preferably at least 73%, more preferably at least 74%, more preferably at least 75%, more preferably at least 76%, more preferably at least 77%, more preferably at least 78%, more preferably at least 79%, more preferably at least 80%, more preferably at least 81%, more preferably at least 82%, more preferably at least 83%, more preferably at least 84%, more preferably at least 85%, more preferably at least 86%, more preferably at least 87%, more preferably at least 88%, more preferably at least 89%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, and most preferably at least 99% identity to a sequences of the present invention. Identity is found over a comparison window of at least 20 amino acid positions, preferably at least 50 amino acid positions, more preferably at least 100 amino acid positions, and most preferably over the entire length of a polypeptide of the invention.

Polypeptide sequence identity can be determined in the following manner. The subject polypeptide sequence is compared to a candidate polypeptide sequence using BLASTP (from the BLAST suite of programs, version 2.2.5 [Nov 2002]) in bl2seq, which is publicly available from the NCBI website on the World Wide Web at ftp://ftp.ncbi.nih.gov/blast/. The default parameters of bl2seq are utilized except that filtering of low complexity regions should be turned off.

Polypeptide sequence identity may also be calculated over the entire length of the overlap between a candidate and subject polynucleotide sequences using global sequence alignment programs. EMBOSS-needle (available at http:/www.ebi.ac.uk/emboss/align/) and GAP (Huang, X. (1994) On Global Sequence Alignment. Computer Applications in the Biosciences 10, 227-235.) as discussed above are also suitable global sequence alignment programs for calculating polypeptide sequence identity.

A preferred method for calculating polypeptide % sequence identity is based on aligning sequences to be compared using Clustal X (Jeanmougin *et al.*, 1998, Trends Biochem. Sci. 23, 403-5.)

Polypeptide variants of the present invention, or used in the methods of the invention, also encompass those which exhibit a similarity to one or more of the specifically identified sequences that is likely to preserve the functional equivalence of those sequences and which could not reasonably be expected to have occurred by random chance. Such sequence similarity with respect to polypeptides may be determined using the publicly available bl2seq program from the BLAST suite of programs (version 2.2.5 [Nov 2002]) from the NCBI website on the World Wide Web at ftp://ftp.ncbi.nih.gov/blast/. The similarity of polypeptide sequences may be examined using the following unix command line parameters:

bl2seq –i peptideseq1 –j peptideseq2 -F F –p blastp

Variant polypeptide sequences preferably exhibit an E value of less than 1 x 10 -6 more preferably less than 1 x 10 -9, more preferably less than 1 x 10 -12, more preferably less than 1 x 10 -15, more preferably less than 1 x 10 -18, more preferably less than 1 x 10 -21, more preferably less than 1 x 10 -30, more preferably less than 1 x 10 -40, more preferably less than 1 x 10 -50, more preferably less than 1 x 10 -60, more preferably less than 1 x 10 -70, more preferably less than 1 x 10 -80, more preferably less than 1 x 10 -90 and most preferably 1x10-100 when compared with any one of the specifically identified sequences.

The parameter –F F turns off filtering of low complexity sections. The parameter –p selects the appropriate algorithm for the pair of sequences. This program finds regions of similarity between the sequences and for each such region reports an "E value" which is the expected number of times one could expect to see such a match by chance in a database of a fixed reference size containing random sequences. For small E values, much less than one, this is approximately the probability of such a random match.

Conservative substitutions of one or several amino acids of a described polypeptide sequence without significantly altering its biological activity are also included in the invention. A skilled artisan will be aware of methods for making phenotypically silent amino acid substitutions (see, e.g., Bowie *et al.*, 1990, Science 247, 1306).

*Constructs, vectors and components thereof*

The term "genetic construct" refers to a polynucleotide molecule, usually double-stranded DNA, which may have inserted into it another polynucleotide molecule (the insert polynucleotide molecule) such as, but not limited to, a cDNA molecule. A genetic construct may contain the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally,

translating the transcript into a polypeptide. The insert polynucleotide molecule may be derived from the host cell, or may be derived from a different cell or organism and/or may be a recombinant polynucleotide. Once inside the host cell the genetic construct may become integrated in the host chromosomal DNA. The genetic construct may be linked to a vector.

The term "vector" refers to a polynucleotide molecule, usually double stranded DNA, which is used to transport the genetic construct into a host cell. The vector may be capable of replication in at least one additional host system, such as *E. coli*.

The term "expression construct" refers to a genetic construct that includes the necessary elements that permit transcribing the insert polynucleotide molecule, and, optionally, translating the transcript into a polypeptide. An expression construct typically comprises in a 5' to 3' direction:

a) a promoter functional in the host cell into which the construct will be transformed,
b) the polynucleotide to be expressed, and
c) a terminator functional in the host cell into which the construct will be transformed.

The term "coding region" or "open reading frame" (ORF) refers to the sense strand of a genomic DNA sequence or a cDNA sequence that is capable of producing a transcription product and/or a polypeptide under the control of appropriate regulatory sequences. The coding sequence may, in some cases, identified by the presence of a 5' translation start codon and a 3' translation stop codon. When inserted into a genetic construct, a "coding sequence" is capable of being expressed when it is operably linked to promoter and terminator sequences.

"Operably-linked" means that the sequenced to be expressed is placed under the control of regulatory elements that include promoters, tissue-specific regulatory elements, temporal regulatory elements, enhancers, repressors and terminators.

The term "noncoding region" refers to untranslated sequences that are upstream of the translational start site and downstream of the translational stop site. These sequences are also referred to respectively as the 5' UTR and the 3' UTR. These regions include elements required for transcription initiation and termination, mRNA stability, and for regulation of translation efficiency.

Terminators are sequences, which terminate transcription, and are found in the 3' untranslated ends of genes downstream of the translated sequence. Terminators are important determinants of mRNA stability and in some cases have been found to have spatial regulatory functions.

The term "promoter" refers to nontranscribed cis-regulatory elements upstream of the coding region that regulate gene transcription. Promoters comprise cis-initiator elements which specify the transcription initiation site and conserved boxes such as the TATA box, and motifs that are bound by transcription factors. Introns within coding sequences can also regulate transcription and influence post-transcriptional processing (including splicing, capping and polyadenylation).

A promoter may be homologous with respect to the polynucleotide to be expressed. This means that the promoter and polynucleotide are found operably linked in nature.

Alternatively the promoter may be heterologous with respect to the polynucleotide to be expressed. This means that the promoter and the polynucleotide are not found operably linked in nature.

In certain embodiments the DGAT1 polynucleotides/polypeptides of the invention may be andvantageously expessed under the contol of selected promoter sequences as described below.

*Vegetative tissue specific promoters*

An example of a vegetative specific promoter is found in US 6,229,067; and US 7,629,454; and US 7,153,953; and US 6,228,643.

*Pollen specific promoters*

An example of a pollen specific promoter is found in US 7,141,424; and US 5,545,546; and US 5,412,085; and US 5,086,169; and US 7,667,097.

*Seed specific promoters*

An example of a seed specific promoter is found in US 6,342,657; and US 7,081,565; and US 7,405,345; and US 7,642,346; and US 7,371,928. A preferred seed specific promoter is the napin promoter of *Brassica napus* (Josefsson *et al.*, 1987, J Biol Chem. 262(25):12196-201; Ellerström *et al.*, 1996, Plant Molecular Biology, Volume 32, Issue 6, pp 1019-1027).

*Fruit specific promoters*

An example of a fruit specific promoter is found in US 5,536,653; and US 6,127,179; and US 5,608,150; and US 4,943,674.

*Non-photosynthetic tissue preferred promoters*

Non-photosynthetic tissue preferred promoters include those preferentially expressed in non-photosynthetic tissues/organs of the plant.

Non-photosynthetic tissue preferred promoters may also include light repressed promoters.

*Light repressed promoters*

An example of a light repressed promoter is found in US 5,639,952 and in US 5,656,496.

*Root specific promoters*

An example of a root specific promoter is found in US 5,837,848; and US 2004/0067506 and US 2001/0047525.

*Tuber specific promoters*

An example of a tuber specific promoter is found in US 6,184,443.

*Bulb specific promoters*

An example of a bulb specific promoter is found in Smeets *et al.*, (1997) Plant Physiol. 113:765-771.

*Rhizome preferred promoters*

An example of a rhizome preferred promoter is found Seong Jang *et al.*, (2006) Plant Physiol. 142:1148-1159.

*Endosperm specific promoters*

An example of an endosperm specific promoter is found in US 7,745,697.

*Corm promoters*

An example of a promoter capable of driving expression in a corm is found in Schenk *et al.*, (2001) Plant Molecular Biology, 47:399-412.

*Photosythetic tissue preferred promoters*

Photosythetic tissue preferred promoters include those that are preferrentially expressed in photosynthetic tissues of the plants. Photosynthetic tissues of the plant include leaves, stems, shoots and above ground parts of the plant. Photosythetic tissue preferred promoters include light regulated promoters.

*Light regulated promoters*

Numerous light regulated promoters are known to those skilled in the art and include for example chlorophyll a/b (Cab) binding protein promoters and Rubisco Small Subunit (SSU) promoters. An example of a light regulated promoter is found in US 5,750,385. Light regulated in this context means light inducible or light induced.

A "transgene" is a polynucleotide that is taken from one organism and introduced into a different organism by transformation. The transgene may be derived from the same species or from a different species as the species of the organism into which the transgene is introduced.

*Host cells*

Host cells may be derived from, for example, bacterial, fungal, yeast, insect, mammalian, algal or plant organisms. Host cells may also be synthetic cells. Preferred host cells are eukaryotic cells. A particularly preferred host cell is a plant cell, particularly a plant cell in a vegetative tissue of a plant.

A "transgenic plant" refers to a plant which contains new genetic material as a result of genetic manipulation or transformation. The new genetic material may be derived from a plant of the same species as the resulting transgenic plant or from a different species. Subsequent offspring or generations of the plant that still contain the new genetic material are also transgenic plants according to the invention.

*Methods for isolating or producing polynucleotides*

The polynucleotide molecules of the invention can be isolated by using a variety of techniques known to those of ordinary skill in the art. By way of example, such polypeptides can be isolated through use of the polymerase chain reaction (PCR) described in Mullis *et al.*, Eds. 1994 The Polymerase Chain Reaction, Birkhauser, incorporated herein by reference. The polypeptides of the invention can be amplified using primers, as defined herein, derived from the polynucleotide sequences of the invention.

Further methods for isolating polynucleotides of the invention include use of all, or portions of, the polypeptides having the sequence set forth herein as hybridization probes. The technique of hybridizing labelled polynucleotide probes to polynucleotides immobilized on solid supports such as nitrocellulose filters or nylon membranes, can be used to screen the genomic or cDNA libraries. Exemplary hybridization and wash conditions are: hybridization for 20 hours at 65°C in 5. 0 X SSC, 0. 5% sodium dodecyl sulfate, 1 X Denhardt's solution; washing (three washes of twenty minutes each at 55°C) in 1. 0 X SSC, 1% (w/v) sodium dodecyl sulfate, and optionally one wash (for twenty minutes) in 0. 5 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C. An optional further wash (for twenty minutes) can be conducted under conditions of 0.1 X SSC, 1% (w/v) sodium dodecyl sulfate, at 60°C.

The polynucleotide fragments of the invention may be produced by techniques well-known in the art such as restriction endonuclease digestion, oligonucleotide synthesis and PCR amplification.

A partial polynucleotide sequence may be used, in methods well-known in the art to identify the corresponding full length polynucleotide sequence. Such methods include PCR-based methods, 5'RACE (Frohman MA, 1993, Methods Enzymol. 218: 340-56) and hybridization- based method, computer/database –based methods. Further, by way of example, inverse PCR permits acquisition of unknown sequences, flanking the polynucleotide sequences disclosed herein, starting with primers based on a known region (Triglia *et al.*, 1998, Nucleic Acids Res 16, 8186, incorporated herein by reference). The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template. Divergent primers are designed from the known region. In order to physically assemble full-length clones, standard molecular biology approaches can be utilized (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987).

It may be beneficial, when producing a transgenic plant from a particular species, to transform such a plant with a sequence or sequences derived from that species. The benefit may be to alleviate public concerns regarding cross-species transformation in generating transgenic organisms. For these reasons among others, it is desirable to be able to identify and isolate orthologues of a particular gene in several different plant species.

Variants (including orthologues) may be identified by the methods described.

*Methods for identifying variants*

*Physical methods*

Variant polypeptides may be identified using PCR-based methods (Mullis *et al.*, Eds. 1994 The Polymerase Chain Reaction, Birkhauser). Typically, the polynucleotide sequence of a primer, useful to amplify variants of polynucleotide molecules of the invention by PCR, may be based on a sequence encoding a conserved region of the corresponding amino acid sequence.

Alternatively library screening methods, well known to those skilled in the art, may be employed (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987). When identifying variants of the probe sequence, hybridization and/or wash stringency will typically be reduced relatively to when exact sequence matches are sought.

Polypeptide variants may also be identified by physical methods, for example by screening expression libraries using antibodies raised against polypeptides of the invention (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987) or by identifying polypeptides from natural sources with the aid of such antibodies.

*Computer based methods*

The variant sequences of the invention, including both polynucleotide and polypeptide variants, may also be identified by computer-based methods well-known to those skilled in the art, using public domain sequence alignment algorithms and sequence similarity search tools to search sequence databases (public domain databases include Genbank, EMBL, Swiss-Prot, PIR and others). See, e.g., Nucleic Acids Res. 29: 1-10 and 11-16, 2001 for examples of online resources. Similarity searches retrieve and align target sequences for comparison with a sequence to be analyzed (i.e., a query sequence). Sequence comparison algorithms use scoring matrices to assign an overall score to each of the alignments.

An exemplary family of programs useful for identifying variants in sequence databases is the BLAST suite of programs (version 2.2.5 [Nov 2002]) including BLASTN, BLASTP, BLASTX, tBLASTN and tBLASTX, which are publicly available from (ftp://ftp.ncbi.nih.gov/blast/) or from the National Center for Biotechnology Information (NCBI), National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894 USA. The NCBI server also provides the facility to use the programs to screen a number of publicly available sequence databases.

BLASTN compares a nucleotide query sequence against a nucleotide sequence database. BLASTP compares an amino acid query sequence against a protein sequence database. BLASTX compares a nucleotide query sequence translated in all reading frames against a protein sequence database. tBLASTN compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames. tBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. The BLAST programs may be used with default parameters or the parameters may be altered as required to refine the screen.

The use of the BLAST family of algorithms, including BLASTN, BLASTP, and BLASTX, is described in the publication of Altschul *et al.*, Nucleic Acids Res. 25: 3389-3402, 1997.

The "hits" to one or more database sequences by a queried sequence produced by BLASTN, BLASTP, BLASTX, tBLASTN, tBLASTX, or a similar algorithm, align and identify similar portions of sequences. The hits are arranged in order of the degree of similarity and the length of sequence overlap. Hits to a database sequence generally represent an overlap over only a fraction of the sequence length of the queried sequence.

The BLASTN, BLASTP, BLASTX, tBLASTN and tBLASTX algorithms also produce "Expect" values for alignments. The Expect value (E) indicates the number of hits one can "expect" to see by chance when searching a database of the same size containing random contiguous sequences. The Expect value is used as a significance threshold for determining whether the hit to a database indicates true similarity. For example, an E value of 0.1 assigned to a polynucleotide hit is interpreted as meaning that in a database of the size of the database screened, one might expect to see 0.1 matches over the aligned portion of the sequence with a similar score simply by chance. For sequences having an E value of 0.01 or less over aligned and matched portions, the probability of finding a match by chance in that database is 1% or less using the BLASTN, BLASTP, BLASTX, tBLASTN or tBLASTX algorithm.

Multiple sequence alignments of a group of related sequences can be carried out with CLUSTALW (Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673-4680, http://www-igbmc.u-strasbg.fr/BioInfo/ClustalW/Top.html) or T-COFFEE (Cedric Notredame, Desmond G. Higgins, Jaap Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, J. Mol. Biol. (2000) 302: 205-217)) or PILEUP, which uses progressive, pairwise alignments. (Feng and Doolittle, 1987, J. Mol. Evol. 25, 351).

Pattern recognition software applications are available for finding motifs or signature sequences. For example, MEME (Multiple Em for Motif Elicitation) finds motifs and signature sequences in a set of sequences, and MAST (Motif Alignment and Search Tool) uses these motifs to identify similar or the same motifs in query sequences. The MAST results are provided as a series of alignments with appropriate statistical data and a visual overview of the motifs found. MEME and MAST were developed at the University of California, San Diego.

PROSITE (Bairoch and Bucher, 1994, Nucleic Acids Res. 22, 3583; Hofmann *et al.*, 1999, Nucleic Acids Res. 27, 215) is a method of identifying the functions of uncharacterized proteins translated from genomic or cDNA sequences. The PROSITE database (www.expasy.org/prosite) contains biologically significant patterns and profiles and is designed so that it can be used with appropriate computational tools to assign a new sequence to a known family of proteins or to determine which known domain(s) are present in the sequence (Falquet *et al.*, 2002, Nucleic Acids Res. 30, 235). Prosearch is a tool that can search SWISS-PROT and EMBL databases with a given sequence pattern or signature.

*Methods for isolating polypeptides*

The polypeptides of the invention, or used in the methods of the invention, including variant polypeptides, may be prepared using peptide synthesis methods well known in the art such as direct peptide synthesis using solid phase techniques (e.g. Stewart *et al.*, 1969, in Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco California, or automated synthesis, for example using an Applied Biosystems 431A Peptide Synthesizer (Foster City, California). Mutated forms of the polypeptides may also be produced during such syntheses.

The polypeptides and variant polypeptides of the invention, or used in the methods of the invention, may also be purified from natural sources using a variety of techniques that are well

known in the art (e.g. Deutscher, 1990, Ed, Methods in Enzymology, Vol. 182, Guide to Protein Purification,).

Alternatively the polypeptides and variant polypeptides of the invention, or used in the methods of the invention, may be expressed recombinantly in suitable host cells and separated from the cells as discussed below.

*Methods for producing constructs and vectors*

The genetic constructs of the present invention comprise one or more polynucleotide sequences of the invention and/or polynucleotides encoding polypeptides of the invention, and may be useful for transforming, for example, bacterial, fungal, insect, mammalian or plant organisms. The genetic constructs of the invention are intended to include expression constructs as herein defined.

Methods for producing and using genetic constructs and vectors are well known in the art and are described generally in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987 ; Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing, 1987).

*Methods for producing host cells comprising polynucleotides, constructs or vectors*

The invention provides a host cell which comprises a genetic construct or vector of the invention.

Host cells comprising genetic constructs, such as expression constructs, of the invention are useful in methods well known in the art (e.g. Sambrook *et al.*, Molecular Cloning : A Laboratory Manual, 2nd Ed. Cold Spring Harbor Press, 1987 ; Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing, 1987) for recombinant production of polypeptides of the invention. Such methods may involve the culture of host cells in an appropriate medium in conditions suitable for or conducive to expression of a polypeptide of the invention. The expressed recombinant polypeptide, which may optionally be secreted into the culture, may then be separated from the medium, host cells or culture medium by methods well known in the art (e.g. Deutscher, Ed, 1990, Methods in Enzymology, Vol 182, Guide to Protein Purification).

*Methods for producing plant cells and plants comprising constructs and vectors*

The invention further provides plant cells which comprise a genetic construct of the invention, and plant cells modified to alter expression of a polynucleotide or polypeptide of the invention, or used in the methods of the invention. Plants comprising such cells also form an aspect of the invention.

Methods for transforming plant cells, plants and portions thereof with polypeptides are described in Draper *et al.*, 1988, Plant Genetic Transformation and Gene Expression. A Laboratory Manual. Blackwell Sci. Pub. Oxford, p. 365; Potrykus and Spangenburg, 1995, Gene Transfer to Plants. Springer-Verlag, Berlin.; and Gelvin *et al.*, 1993, Plant Molecular Biol. Manual. Kluwer Acad. Pub. Dordrecht. A review of transgenic plants, including transformation techniques, is provided in Galun and Breiman, 1997, Transgenic Plants. Imperial College Press, London.

*Methods for genetic manipulation of plants*

A number of plant transformation strategies are available (e.g. Birch, 1997, Ann Rev Plant Phys Plant Mol Biol, 48, 297, Hellens *et al.*, (2000) Plant Mol Biol 42: 819-32, Hellens *et al.*, Plant Meth 1: 13). For example, strategies may be designed to increase expression of a polynucleotide/polypeptide in a plant cell, organ and/or at a particular developmental stage where/when it is normally expressed or to ectopically express a polynucleotide/polypeptide in a cell, tissue, organ and/or at a particular developmental stage which/when it is not normally expressed. The expressed polynucleotide/polypeptide may be derived from the plant species to be transformed or may be derived from a different plant species.

Transformation strategies may be designed to reduce expression of a polynucleotide/polypeptide in a plant cell, tissue, organ or at a particular developmental stage which/when it is normally expressed. Such strategies are known as gene silencing strategies.

Genetic constructs for expression of genes in transgenic plants typically include promoters for driving the expression of one or more cloned polynucleotide, terminators and selectable marker sequences to detect presence of the genetic construct in the transformed plant.

The promoters suitable for use in the constructs of this invention are functional in a cell, tissue or organ of a monocot or dicot plant and include cell-, tissue- and organ-specific promoters, cell cycle specific promoters, temporal promoters, inducible promoters, constitutive promoters that are active in most plant tissues, and recombinant promoters. Choice of promoter will depend upon the temporal and spatial expression of the cloned polynucleotide, so desired. The

promoters may be those normally associated with a transgene of interest, or promoters which are derived from genes of other plants, viruses, and plant pathogenic bacteria and fungi. Those skilled in the art will, without undue experimentation, be able to select promoters that are suitable for use in modifying and modulating plant traits using genetic constructs comprising the polynucleotide sequences of the invention. Examples of constitutive plant promoters include the CaMV 35S promoter, the nopaline synthase promoter and the octopine synthase promoter, and the Ubi 1 promoter from maize. Plant promoters which are active in specific tissues respond to internal developmental signals or external abiotic or biotic stresses are described in the scientific literature. Exemplary promoters are described, e.g., in WO 02/00894 and WO2011/053169, which is herein incorporated by reference.

Exemplary terminators that are commonly used in plant transformation genetic construct include, e.g., the cauliflower mosaic virus (CaMV) 35S terminator, the *Agrobacterium tumefaciens* nopaline synthase or octopine synthase terminators, the *Zea mays* zein gene terminator, the *Oryza sativa* ADP-glucose pyrophosphorylase terminator and the *Solanum tuberosum* PI-II terminator.

Selectable markers commonly used in plant transformation include the neomycin phophotransferase II gene (NPT II) which confers kanamycin resistance, the aadA gene, which confers spectinomycin and streptomycin resistance, the phosphinothricin acetyl transferase (*bar* gene) for Ignite (AgrEvo) and Basta (Hoechst) resistance, and the hygromycin phosphotransferase gene ( hpt) for hygromycin resistance.

Use of genetic constructs comprising reporter genes (coding sequences which express an activity that is foreign to the host, usually an enzymatic activity and/or a visible signal (e.g., luciferase, GUS, GFP) which may be used for promoter expression analysis in plants and plant tissues are also contemplated. The reporter gene literature is reviewed in Herrera-Estrella *et al.*, 1993, Nature 303, 209, and Schrott, 1995, In: Gene Transfer to Plants (Potrykus, T., Spangenberg. Eds) Springer Verlag. Berline, pp. 325-336.

The following are representative publications disclosing genetic transformation protocols that can be used to genetically transform the following plant species: Rice (Alam *et al.*, 1999, Plant Cell Rep. 18, 572); apple (Yao *et al.*, 1995, Plant Cell Reports 14, 407-412); maize (US Patent Serial Nos. 5, 177, 010 and 5, 981, 840); wheat (Ortiz *et al.*, 1996, Plant Cell Rep. 15, 1996, 877); tomato (US Patent Serial No. 5, 159, 135); potato (Kumar *et al.*, 1996 Plant J. 9, : 821); cassava (Li *et al.*, 1996 Nat. Biotechnology 14, 736); lettuce (Michelmore *et al.*, 1987, Plant Cell Rep. 6, 439); tobacco (Horsch *et al.*, 1985, Science 227, 1229); cotton (US Patent Serial Nos. 5, 846, 797

and 5, 004, 863); grasses (US Patent Nos. 5, 187, 073 and 6. 020, 539); peppermint (Niu *et al.*, 1998, Plant Cell Rep. 17, 165); citrus plants (Pena *et al.*, 1995, Plant Sci.104, 183); caraway (Krens *et al.*, 1997, Plant Cell Rep, 17, 39); banana (US Patent Serial No. 5, 792, 935); soybean (US Patent Nos. 5, 416, 011 ; 5, 569, 834 ; 5, 824, 877 ; 5, 563, 04455 and 5, 968, 830); pineapple (US Patent Serial No. 5, 952, 543); poplar (US Patent No. 4, 795, 855); monocots in general (US Patent Nos. 5, 591, 616 and 6, 037, 522); brassica (US Patent Nos. 5, 188, 958 ; 5, 463, 174 and 5, 750, 871); cereals (US Patent No. 6, 074, 877); pear (Matsuda *et al.*, 2005, Plant Cell Rep. 24(1):45-51); Prunus (Ramesh *et al.*, 2006 Plant Cell Rep. 25(8):821-8; Song and Sink 2005 Plant Cell Rep. 2006 ;25(2):117-23; Gonzalez Padilla *et al.*, 2003 Plant Cell Rep.22(1):38-45); strawberry (Oosumi *et al.*, 2006 Planta. 223(6):1219-30; Folta *et al.*, 2006 Planta Apr 14; PMID: 16614818), rose (Li *et al.*, 2003), Rubus (Graham *et al.*, 1995 Methods Mol Biol. 1995;44:129-33), tomato (Dan *et al.*, 2006, Plant Cell Reports V25:432-441), apple (Yao *et al.*, 1995, *Plant Cell Rep.* **14**, 407–412), Canola (Brassica napus L.).(Cardoza and Stewart, 2006 Methods Mol Biol. 343:257-66), safflower (Orlikowska *et al.*, 1995, Plant Cell Tissue and Organ Culture 40:85-91), ryegrass (Altpeter *et al.*, 2004 Developments in Plant Breeding 11(7):255-250), rice (Christou *et al.*, 1991 Nature Biotech. 9:957-962), maize (Wang *et al.*, 2009 In: Handbook of Maize pp. 609-639) and *Actinidia eriantha* (Wang *et al.*, 2006, Plant Cell Rep. 25,5: 425-31). Transformation of other species is also contemplated by the invention. Suitable methods and protocols are available in the scientific literature.

## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the nucleic acid sequence and three frame translation of the *Arabidopsis thaliana* DGAT1 transcribed region (SEQ ID NO:116). Exon coding sequences are shown in bold face, underlined, grey blocks.

Figure 2 shows the nucleic acid sequence and three frame translation of the *Zea mays* short DGAT1 transcribed region (SEQ ID NO:117). This genomic sequence has F469 deleted and Q67 added compared to the cDNA (EU039830) and peptide (ABV91586) sequences actually used in this patent. Exon coding sequences are shown in bold face, underlined, grey blocks.

Figure 3 shows the nucleic acid sequence and three frame translation of the *Zea mays* long DGAT1 transcribed region (SEQ ID NO:118) derived from CHORI-201 Maize B73 BAC Library (available from the World Wide Web at http://www.ncbi.nlm.nih.gov/nuccore/AC204647; http://bacpac.chori.org/maize201.htm). Exon coding sequences are shown in bold face, underlined, grey blocks.

Figure 4 shows the peptide sequence of the N-terminal cytoplasmic region of a number of plant DGAT1s including both long and short versions from the grasses as well as examples from dicotyledonous species. Left hand box represents acyl-CoA binding site (Nykiforuk *et al.*, 2002, Biochimica et Biophysica Acta 1580:95-109). Right hand box represents first transmembrane region (McFie *et al.*, 2010, JBC., 285:37377-37387). Left hand arrow represents boundary between exon 1 and exon 2. Right hand arrow represents boundary between exon 2 and exon 3. The sequences are AtDGAT1 (SEQ ID NO:119), BjDGAT1 (SEQ ID NO:120), BnDGAT1-AF (SEQ ID NO:121), BjDGAT1 (SEQ ID NO:122), TmajusDGAT1 (SEQ ID NO:123), EpDGAT1 (SEQ ID NO:124), VgDGAT1 (SEQ ID NO:125), NtDGAT1 (SEQ ID NO:126), PfDGAT1 (SEQ ID NO:127), ZmL (SEQ ID NO:128), SbDGAT1 (SEQ ID NO:129), OsL (SEQ ID NO:130), OsS (SEQ ID NO:131), SbDGAT1 (SEQ ID NO:132), ZmS (SEQ ID NO:133), PpDGAT1 (SEQ ID NO:132), SmDGAT1 (SEQ ID NO:135), EaDGAT1 (SEQ ID NO:136), VvDGAT1 (SEQ ID NO:137), GmDGAT1 (SEQ ID NO:138), GmDGAT1 (SEQ ID NO:139), LjDGAT1 (SEQ ID NO:140), MtDGAT1 (SEQ ID NO:141), JcDGAT1 (SEQ ID NO:142), VfDGAT1 (SEQ ID NO:143), RcDGAT1 (SEQ ID NO:144), PtDGAT1 (SEQ ID NO:145), Pt DGAT1 (SEQ ID NO:146).

Figure 5 shows the line-bond structures of the amino acid residues lysine (K) and arginine (R).

## EXAMPLES

### Example 1: Identification of the DGAT1 sequence of the invention

Several nucleic acid sequences and polypeptide sequences for the plant type 1 DGATs can be found by accession number in public domain libraries (Table 1). For creating initial alignments we used ClustalW (Thompson *et al.*, 1994, Nucleic Acids Res., 22, 4673-4680); these were manually edited and used to create the models to search the DGAT sequences, using the HMMER2 package (HMMER 2.3.2 (Oct 2003) Copyright © 1992-2003 HHMI/Washington University School of Medicine, available from the World Wide Web at http://hmmer.org). Initial matching of protein sequences against genomic DNA with splice prediction was performed with the GeneWise package (Birney *et al.*, 2004, Genome Res. 14: 988-995). Some of the sequences retrieved appeared to have errors; in particular incorrectly predicted splice sites which would result in internal deletions that would likely result in non-functional proteins. While both dicotyledonous and monocotyledonous type 1 DGATs have 16 exons there are some differences in the position of the splicing. Exon 8 in the dicotylendonous DGAT1 gene corresponds to exons 8 and 9 in the monocotyledonous DGAT1 gene, while exon 14 in the monocotyledonous gene corresponds to exons 13 and 14 in the dicotyledonous gene. We have found that the most accurate method for determing the likely genuine coding sequence from genomic data has been to use Vector NTI Advance (TM) 11.0 (© 2008 Invitrogen Corporation) to translate the genome in the three forward reading frames and align these with demonstrated functional DGAT1s from dicotyledonous or monocotyledous species as appropriate (for example *A. thaliana* cDNA NM_127503, protein NP_179535 and *Z. mays* cDNA EU039830, protein ABV91586). The genomic sequence and corresponding exon/intron boundary positions for *Arabidopsis thaliana* encoding NP_179535 and *Zea mays* encoding ABV91586 that can be used as a template for determining other plant DGAT coding regions are shown in Figures 1 and Figures 2, respectivlely.

Using this method, the applicants have assembled/identified a novel DGAT1 sequence from *Z. mays* DGAT1 (SEQ ID NO: 10 and SEQ ID NO: 39 [Figure 3]). To the best of the applicant's knowledge, a functional portion of sequence is not present in any public cDNA database, which may indicate that the functional protein is not expressed in naturally occurring plants.

The applicants designated this sequence *Zea mays* DGAT-Long (ZmDGAT1-L or Zm-L DGAT1) because the encoded polypeptides is longer than the known *Zea mays* of SEQ ID NO:

44 (referred to as *Zea mays* DGAT1-short or ZmDGAT1-S or Zm-S DGAT1) as indicated in Figure 4.

A similar relationship exists between *Oryza sativa* DGAT1-short, or OsDGAT1-S, or Os-S DGAT1 (SEQ ID NO:41, and *Oryza sativa* DGAT1-long, or OsDGAT1-L, or Os-L DGAT1 (SEQ ID NO:42).

Table 1

| **DGAT1 Species Source** | **DNA accession #s & BAC #** | **SEQ ID NO:** | **PROTEIN accession #s & BAC #** | **SEQ ID NO:** |
|---|---|---|---|---|
| *A. thaliana* | NM_127503 | 1 | NP_179535 | 30 |
| *B. juncea* | AF164434 | 2 | AAY40784 | 31 |
| *B. napus* | AF164434_1 | 3 | AAD45536.1 | 32 |
| *B. juncea* | DQ016107 | 4 | AAY40785 | 33 |
| *T. majus* | AY084052 | 5 | AAM03340 | 34 |
| *E. pitardii* | FJ226588 | 6 | ACO55635 | 35 |
| *V. galamensis* | EF653276 | 7 | ABV21945 | 36 |
| *N. tabacum* | AF129003_1 | 8 | AAF19345.1 | 37 |
| *P. frutescens* | AF298815_1 | 9 | AAG23696.1 | 38 |
| *Z. mays* | From: CHORI-201 Maize B73 BAC | 10 | From: CHORI-201 Maize B73 BAC | 39 |
| *S. bicolor* | XM_002439374 | 11 | XP_002439419 | 40 |
| *O. sativa* | Os05g0196800 | 12 | NP_001054869 | 41 |
| *O. sativa* | From: AP003714.1 | 13 | From: AP003714.1 | 42 |
| *S. bicolor* | XM_002437120.1 | 14 | XP_002437165 | 43 |
| *Z. mays* | EU039830 | 15 | ABV91586 | 44 |
| *P. patens* | XM_001770877.1 | 16 | XP_001770929 | 45 |
| *S. moellendorffii* | XM_002964119 | 17 | XP_002964165 | 46 |
| *E. alatus* | AY751297 | 18 | AAV31083 | 47 |
| *V. vinifera* | XM_002279309 | 19 | XP_002279345 | 48 |
| *G. max* | AY496439 | 20 | AAS78662 | 49 |
| *G. max* | AB257590 | 21 | BAE93461 | 50 |
| *L. japonicus* | AY859489 | 22 | AAW51456 | 51 |
| *M. truncatula* | AC174465.2 | 23 | ABN09107 | 52 |
| *J. curcas* | DQ278448.1 | 24 | ABB84383 | 53 |
| *V. fordii* | DQ356680.1 | 25 | ABC94472 | 54 |
| *V. galamensis* | EF653276.1 | 26 | ABV21945 | 55 |
| *R. communis* | XM_002514086.1 | 27 | XP_002514132 | 56 |
| *P. trichocarpa* | XM_002308242.1 | 28 | XP_002308278 | 57 |
| *P. trichocarpa* | XM_002330474.1 | 29 | XP_002330510 | 58 |

**Example 2: The DGAT1 sequence of the invention has surprisingly high activity in increasing cellular lipid content, and fragments of the DGAT1 sequence of the invention are useful in conferring increased activity to other DGAT1 proteins.**

*Summary*

The applicants compared the activity of the DGAT1 sequence of the invention to other known DGAT1 sequences. Surprisingly the DGAT1 sequence of the invention showed higher activity, in increasing cellular lipid content, than any of the other DGAT1 sequences tested.

Furthermore the applicants have shown that fragments of the DGAT1 protein of the invention are useful in conferring increase activity on at least parts of other DGAT proteins.

*Materials and methods*

Nucleic acid constructs encoding the amino acid sequences, SEQ ID NO: 30 (*A.* thaliana DGAT1), 34 (*T. magus* DGAT1), 39 (*Zea mays* DGAT1-L), 41 (*O. sativa* DGAT1-S), 42 (*O. sativa DGAT1-L)* and 44 (Table 1) were optimised for expression in *Saccharomyces cerevisiae* by GeneArt AG (Germany). These were engineered to have an internal *Xho*I site within exon 1 encoding the conserved *N*-terminal acyl-Co binding region (identified by Lung and Weselake, 2006, Lipids. Dec 2006;41(12):1073-88) without altering the amino acid sequence leucine-serine-serine (LSS). Figure 4 shows alignment of a number of DGAT1 sequences from plants. The left box shows the position of the Acyl-CoA binding site.

An *Eco*RI site was engineered upstream of the 5′ coding sequence while an *Xba*I site was placed downstream of the 3′ stop codon. The internal *Xho*I and flanking *Eco*RI and *Xba*I sites were used to generate chimeras between the DGAT1 sequence of the invention and each of the other DGAT1 clones; essentially this fused the *N*-terminal reputed cytoplasmic region (based on Lung and Weselake, 2006, Lipids. Dec 2006;41(12):1073-88 and McFie *et al.*, 2010, JBC., 285:37377-37387) from one DGAT1 with the *C*-terminal ER luminal region of a different DGAT1. In some combinations this resulted in one amino acid change in the remaining cytoplasmic region downstream of the engineered XhoI site. The putative acyl-Co binding region the *A. thaliana* DGAT1, *T.majus* DGAT1, *Z.mays*-L DGAT1and *O. sativa*-L DGAT1 have an identical amino acid sequence down stream of the *Xho*I site (LSSDAIFKQSHA). While in the *Z.mays*-S DGAT1

and *O. sativa*-S DGAT1 the lysine (K) residue is replaced by an arginine (R) residue (LSSDAIFRQSHA). Since the position of this residue is located 3′ to the *Xho* I site encoded by LLS then chimeras deriving from one parent containing the lysine and one parent containing the arginine residue will effectively result in a substitution of this residue. This was considered to be a minimal disruption since both lysine and arginine are large, positively charged, hydrophilic, basic amino acids containing a free amine or guanidinium group, respectively at the end of an aliphatic side chain (Figure 5). The *N*-terminal region / *C*-terminal region domain swapping constructs, and the parent constructs (highlighted in bold) are shown in in Table 2 , with their corresponding SEQ ID NOs.

Table 2

| **DGAT1 N-terminal parent** | **DGAT1 C-terminal parent** | **C-terminal Tail Fusion** | **SEQ ID NO:** |
|---|---|---|---|
| ***A. thaliana*** | ***A. thaliana*** | **V5-6xHis** | **59** |
| *A. thaliana* | *Z. mays*-L | V5-6xHis | 63 |
| ***O. sativa*-S** | ***O. sativa*-S** | **V5-6xHis** | **65** |
| *O. sativa*-S | *Z. mays*-L | V5-6xHis | 69 |
| ***O. sativa*-L** | ***O. sativa*-L** | **V5-6xHis** | **71** |
| *O. sativa*-L | *Z. mays*-L | V5-6xHis | 75 |
| ***Z. mays*-S** | ***Z. mays*-S** | **V5-6xHis** | **77** |
| *Z. mays*-S | *Z. mays*-L | V5-6xHis | 81 |
| **Z. mays-L** | ***Z. mays*-L** | **V5-6xHis** | **83** |
| *Z. mays*-L | *A. thaliana* | V5-6xHis | 84 |
| *Z. mays*-L | *O. sativa*-S | V5-6xHis | 85 |
| *Z. mays*-L | *O. sativa*-L | V5-6xHis | 86 |
| *Z. mays*-L | *Z. mays*-S | V5-6xHis | 87 |
| *Z. mays*-L | *T. majus* | V5-6xHis | 88 |
| ***T. majus*** | ***T. majus*** | **V5-6xHis** | **89** |
| *T. majus* | *Z. mays*-L | V5-6xHis | 94 |

Sequences were synthesised either by GENEART AG (Germany) or GeneScript (USA). Sequences were optimised for expression in *Saccharomyces cerevisiae* and flanked with appropriate incorporated appropriate restriction sites to facilitate the cloning into the pYES2.1 vector (Invitrogen).

*Expression of constructs in S. cerevisiae*

The parent DGAT1 constructs and chimeric DGAT1 contstructs were placed into the galactose-inducible yeast expression vector pYES2.1/V5-His TOPO® (Invitrogen). This resulted in the addition of an inframe *C*-terminal V5 epitope and 6x histidine tag. The chimeric constructs and the number of their corresponding polypeptide sequences are shown in Table 2 above.

The *Saccharomyces cerevisiae* quadruple mutant (H1246) in which all four neutral lipid biosynthesis genes have been disrupted (Sandager *et al.*, 2002, The Journal of Biological Chemistry, 277:6478-6482) was transformed as per Elble (1992, BioTechniques 13, 18-20) and selected by the ability to grow in the absence of uracil. Routinely, yeast cells were grown aerobically overnight in a synthetic medium with 0.67% YNB, without uracil (SC-U) and containing 2% glucose. Cells from overnight culture were used to inoculate 200 mL of induction medium (SC-U containing 2% galactose and 1% raffinose) to an initial $OD_{600}$ of 0.4. Cells were allowed to further grow at 30°C, with shaking at 200 rpm until late stationary phase, normally 48 h. Cells were harvested by centrifugation at 1500 x *g* for 5 min, then cell pellets were washed with distilled water and either used immediately for subsequent analysis or kept in -80°C until required. Cell pellets for neutral lipid extraction were freeze-dried for 48 h and stored in -20°C freezer until required.

*Lipid analysis of S. cerevisiae*

Approximately 10 mg of freeze-dried yeast cell material was accurately weighed then disrupted using glass beads by vortexing for 1 minute. This lysate was extracted in hot methanolic HCl for fatty acid methyl ester (FAME) analysis (Browse *et al.*, 1986, Anal. Biochem. 152, 141-145).

For FA profile analysis approximately 50 mg freeze dried yeast was placed in a 13-mm screw cap tube, and an equal volume of glass beads added before vortexing at high speed in 3x 1 min bursts. Following addition of 50 µg of 19:0 TAG internal standard, 2.4 mL of 0.17 M NaCl in MeOH was added and the mixture vortexed for 15 sec followed by the addition of then 4.8 mL of heptane and the entire contents mixed.

The solution was then incubated in 80°C water bath for 2 h without shaking. After incubation, the solution was cooled to room temperature. After cooling, the upper phase (lipidic phase) was transferred to fresh screw-cap tube and evaporated to dryness under stream of nitrogen gas. The

dried residue was then dissolved in 1 mL heptane and mixed thoroughly for TAG SPE separation using Strata Si-1 Silica column (Phenomenwx, 8B-S012-EAK).

After preconditioning with methanol and equilibrating the Silica column with heptane the 1 mL TAG extract (including 50 μg 17:0 TAG Internal Standard was passed through the pre-equilibrated column, followed by 1.2 mL of heptane and then 2 mL of chloroform:heptane (1:9 v/v) and the eluate collected.The total eluate collected was evaporated to dryness under the stream of N gas and the residue used for FAMEs extraction.

*FAMEs of extracted TAG*

To the TAG residue above 10 μL of internal standard 15:0 FA (4 mg/mL dissolved in heptane) and 1 mL of methanolic HCl (1N) reagent containing 5% of 2,2-dimeethoxypropane (as water scavenger) were added.

The tube was then flushed with N gas, then sealed immediately with Teflon-lined cap, and heated at 80°C in a water bath for 1 h. After cooling down, 0.6 mL heptane and 1.0 mL of 0.9% (w/v) NaCl was added, the mixture vortexed then spun at 500 rpm for 1 min.

From the top heptane layer, 100 μL was collected and transfered to a flat-bottom glass insert fitted into a vial for FAMES GC/MS analysis.

*Protein extraction and Trypsin digestion*

Yeast cell pellets were washed with lysis buffer (50 mM sodium phosphate, pH 7.4, 1 mM EDTA, 5% glycerol, 1 mM PMSF) then resuspended in 500 μL lysis buffer, glass beads were added and cells disrupted by vortexing 2x at medium speed for 30 seconds. Cell debris was pelleted by centrifugation at 1000 x *g* for 5 min, the supernatant transferred to fresh tubes and total cellular membranes pelleted by ultracentrifugation at 100,000 x *g* for 1 h. Membrane proteins were resuspended in lysis buffer with or without detergent (1% Dodecyl maltoside) and quantified in a Qubit Fluorometer using the Qubit IT Quantitation Kit.

Trypsin was added to give a final concentration of 25 μg/mL to 50 μL of protein extract and the mixture incubated at 30°C for 30 min. The reaction was terminated by addition of Trypsin inhibitor from *Glycine max* (Sigma-Aldrich catalogue # T6414) to a final concentration of 0.4 μg/μL. After addition of trypsin inhibitor, 4x SDS loading dye and 10x reducing agent (Invitrogen) were added, and the protein incubated at 70°C for 10 min prior to SDS-PAGE followed by immunoblotting. The blot was probed with either Anti V5-HRP antibody (Cat #R96125, Invitrogen) at 1:2500 dilution, or anti Kar2 (y-115) antibody produced in rabbit (SC-33630, Santa Cruz Biotechnology) at 1:200 dilution. Anti Kar2 was used to detect the yeast protein Kar2, an ER luminaly-located protein (Rose *et al.*, 1989, Cell 57,1211-1221) which serves as a control to demonstrate the presence of intact microsomes.

*Expression of chimeric DGAT1 in Brassica napus*

The same strategy, as described above, was used to generate a variety of chimeric DGAT1 constructs for expression in the seeds of *Brassica napus*. This included the parent DGAT1s of *T. majus* DGAT1, *Z. mays*-L DGAT1 and *Z. mays*-S DGAT1 (amino acid SEQ ID NO: 34, 39 and 44 respecitively, Table 1) optimised for expression in *Brassica napus* by GeneArt AG. The *T. majus* construct was engineered to contain a single point mutation $S_{197}A$ (Xu *et al.*, 2008, Plant Biotechnology Journal, 6:799-818). All constructs were engineered to have an optimised Kozak, *Arabidopsis thaliana* UBQ10 intron, and tetranucleotide stop codon as per Scott *et al.*, (2010, Plant Biotechnology Journal, 8:912-917) as indicated in Table 3 below.

Table 3

| **DGAT1 Parent Species** | **Kozak, intron, stop codon** | **Residue modification** | **SEQ ID NO:** |
|---|---|---|---|
| *T. majus* | yes | S197A | 95 |
| *Z. mays*-S | yes | none | 96 |
| *Z. mays*-L | yes | none | 97 |

The same digestion pattern used to generate the chimeras for expression in *S. cerevisiae* was used on the *B. napus*-optimised constructs to generate the chimeras Tm-ZmL and ZmL-Tm(S189A); resulting in the peptide sequences listed in Table 4.

Table 4

| DGAT1 N-terminal parent | DGAT1 C-terminal parent | Residue modification | SEQ ID NO: |
|---|---|---|---|
| *T. majus* | *T. majus* | S197A | 98 |
| *Z. mays*-L | *Z. mays*-L | none | 99 |
| *T. majus* | *Z. mays*-L | none | 100 |
| *Z. mays*-L | *T. majus* | S189A | 101 |

The parent DGATs and their chimeras were transferred into the Gateway®-compatible binary vector pMD107 (courtesy of Dr Mark Smith, NRC Saskatoon, SK, Canada, S7N 0W9) which placed them under the control of the seed- specific napin promoter (Ellerström *et al.*, 1996, Plant Molecular Biology, Volume 32, Issue 6, pp 1019-1027).

*Plant transformation*

*B. napus* (cv. DH12075) was transformed via *Agrobacterium tumefaciens* (GV3101) using the cotyledon co-cultivation method (adapted from that of Maloney *et al.*, 1989, Plant Cell Rep. 8, 238-242). Control lines contained an empty-vector, and when identified, null sibling lines were subsequently used as true controls.

Approximately 200 $T_0$ transformed lines were produced and their corresponding $T_1$ selfed seeds were analysed for oil content by GC. Approximately 50 individual transgenic lines (including control lines) were selected for the next generation (10 plants/line) based on their oil content, or seed weight (8 lines).

A total of approximately $T_1$ plants were grown and screened by PCR for copy number and identification of null sibing lines. $T_2$ seeds were analysed in triplicate for oil content by NMR.

**Expression of Z.mays-L and T. majus DGAT1 in *Camelina sativa***

The strategy above can also be used to generate a variety of chimeric DGAT1 constructs for expression in the seeds of *Camelina sativa* and other plants.

Sequences with modifications were synthesised either by GENEART AG (Germany) or GeneScript (USA). Sequences with modifications were synthesised either by GENEART AG (Germany) or GeneScript (USA). Sequences were optimised for expression in *Brassica species* and included an intron (SEQ ID NO:102) from *Arabidopsis thaliana* DGAT1 – intron 3. Each sequence was flanked with appropriate attL recombination sites sites to enable the cloning Gateway® adapted vectors.

Table 5

| DGAT1 N-terminal parent | DGAT1 C-terminal parent | Residue modification | C-terminal mod | Additional information | Type of sequence | SEQ ID NO: |
|---|---|---|---|---|---|---|
| T.majus | T. majus | S197A | V5-His tag | + intron | NUCLEIC | 103 |
| T.majus | T. majus | S197A | V5-His tag | ORF only | NUCLEIC | 104 |
| T.majus | T. majus | S197A | V5-His tag | | PEPTIDE | 105 |
| Z. mays-L | Z. mays-L | None | V5-His tag | + intron | NUCLEIC | 106 |
| Z. mays-L | Z. mays-L | None | V5-His tag | ORF only | NUCLEIC | 107 |
| Z. mays-L | Z. mays-L | None | V5-His tag | | PEPTIDE | 108 |
| T.majus | Z. mays-L | None | V5-His tag | + intron | NUCLEIC | 109 |
| T.majus | Z. mays-L | None | V5-His tag | ORF only | NUCLEIC | 110 |
| T.majus | Z. mays-L | None | V5-His tag | | PEPTIDE | 111 |
| Z. mays-L | T. majus | S189A | V5-His tag | + intron | NUCLEIC | 112 |
| Z. mays-L | T. majus | S189A | V5-His tag | ORF only | NUCLEIC | 113 |
| Z. mays-L | T. majus | S189A | V5-His tag | | PEPTIDE | 114 |

The parent DGATs and their modified forms were transferred into the Gateway®-compatible binary pRSh1 Gateway adapted binary vector (Winichayakul *et al.*, 2009, Biotechnol. Appl. Biochem. 53, 111–122) modified by replacement of the CaMV35S promoter replaced with the *Brassica napus* Napin promoter (SEQ ID NO:115).

**_Camelina sativa_ transformation**

*C. sativa* (cf. Calena) were transformed via Agrobacterium tumefaciens (GV3101) using the floral dip method (adapted from that of Clough and Bent, 1998, Plant J. 16(6):735-745). Essentially

seeds were sown in potting mix in 10 cm pots in a controlled environment, approximately 6 weeks after planting the flowers were dipped for 5-14 minutes under vacuum (70-80 inch Hg) in an overnight culture of appropriated Agrobacterium GV3101 cells re-suspended in a floral dip buffer. After vacuum-transformation, plants were kept for 24 h under low light conditions by partly covering with a black plastic sheet. Vacuum transformations can be repeated three times at approximately 10-12 days intervals, corresponding to the flowering duration. Plants were grown in potting mix in a controlled environment (16-h day length, 21-24 °C, 65-70 % relative humidity).

The $T_1$ seeds produced can be collected and screened for transformants by germinating and growing seedlings at 22 °C with continuous light on a half-strength MS medium (pH 5.6) selection plate containing 1 %(w/v) sucrose, 300 mg/L Timentin, and 25 mg/L DL-phosphinothricin to select for herbicide resistance. $T_2$ selfed seed populations can also be screened by immuno blot for the presence of the V5 eptiope.

$T_2$ selfed seeds may be analysed for oil content by GC. Approximately 50 individual transgenic lines (including control lines) may be selected for the next generation (10 plants/line) based on their oil content, or seed weight. $T_2$ plants may be grown and screened by PCR for copy number and identification of null sibing lines. $T_2$ seeds may be analysed in triplicate for oil content by NMR or GC/MS.

**Results**

*Addition of fragments of the DGAT1 polypeptide of the invention to other DGAT1 sequences enhance lipid production in Saccharomyces cerevisiae relative to that with the other DGAT1 sequences alone.*

Tables 8-14 show the lipid yields of a variety of chimeric DGAT1s in which the *N*-terminal or C- terminal region has been derived from the DGAT1 sequence of the invention while the remainder of the protein has been derived from another plant DGAT1. The lipid yields are presented either as grams of lipid produced per litre (which therefore compensates for any differences in growth rate) or have been normalised as a percentage of the lipid yield of the corresponding unmodified parent DGAT1.

A comparison of parent DGAT1s with each other, and with each of the chimeric DGAT1s made using one donor parent for the *N*-terminal region, and a different donor parent for the *C*-terminal region are shown in Table 5. The parent DGAT1 sequences are highlighted in bold.

Surprisingly the DGAT1 sequence of the invention shows higher activity, in lipid yield production, than any of the other sequences tested.

The lipid yields at 32 hr have been normalised against the highest lipid-producing parent (*Z. mays*-L) and are presented in ascending order.

Table 6

| *N*-terminal region DGAT1 Parent | *C*-terminal region DGAT1parent | SEQ ID NO: | Lipid yield as % *Z.mays*-L |
|---|---|---|---|
| Vector only | Vector only | N/A | 31.96 |
| *A. thaliana* | *Z. mays*-L | 63 | 38.28 |
| ***A. thaliana*** | ***A. thaliana*** | **59** | **64.69** |
| ***T. majus*** | ***T. majus*** | **89** | **77.62** |
| ***Z. mays*-S** | ***Z. mays*-S** | **77** | **81.79** |
| *Z. mays*-L | *T. majus* | 88 | 83.39 |
| ***O. sativa*-S** | ***O. sativa*-S** | **65** | **84.76** |
| ***O. sativa*-L** | ***O. sativa*-L** | **71** | **88.33** |
| *O. sativa*-S | *Z. mays*-L | 69 | 95.81 |
| *Z. mays*-L | *O. sativa*-L | 86 | 96.17 |
| *Z. mays*-L | *A. thaliana* | 84 | 97.53 |
| ***Z. mays*-L** | ***Z. mays*-L** | **83** | **100.00** |
| *T. majus* | *Z. mays*-L | 94 | 100.71 |
| *O. sativa*-L | *Z. mays*-L | 75 | 104.29 |
| *Z. mays*-L | *O. sativa*-S | 85 | 105.02 |

The results also show that addition of the *Z. mays*-L N-terminal region to the C-terminal region of the *A. thaliana* DGAT1 parent results in increased lipid yield over the full-length *A. thaliana* DGAT1 sequence (see SEQ ID NO: 84 versus SEQ ID NO: 59).

The results also show that addition of the *Z. mays*-L N-terminal or C-terminal region to the C-terminal or N-terminal region respectively, of the *T. majus* DGAT1 sequence results in increased lipid yield over the full-length *T. majus* DGAT1 sequence (see SEQ ID NO: 88 and 94 versus SEQ ID NO: 89).

The results also show that addition of the *Z. mays*-L N-terminal or C-terminal region to the C-terminal or N-terminal region respectively, of the *O. sativa*–S DGAT1 sequence results in increased lipid yield over the full-length *O. sativa*–S DGAT1 sequence (see SEQ ID NO: 85 and 69 versus SEQ ID NO: 65).

The results also show that addition of the *Z. mays*-L N-terminal or C-terminal region to the C-terminal or N-terminal region respectively, of the *O. sativa*–L DGAT1 sequence results in increased lipid yield over the full-length O. *sativa*-L (see SEQ ID NO: 86 and 75 versus SEQ ID NO: 71).

The results also show that addition of the Z. mays-L N-terminal or C-terminal region to C-terminal or N-terminal region respectively, of the *Z. mays*-S sequence results in increased lipid yield over the full-length *Z. mays*-S sequence (see SEQ ID NO: 87 and 81 versus SEQ ID NO: 77).

In summary addition of fragments (either the N-terminal region or C-terminal region) of the *Z. mays*-L polypeptide of the invention to another DGAT1 sequence can increase the cellular lipid yield attainable by the combined sequence over that of the other DGAT1 sequence.

*Addition of fragments of the DGAT1 polypeptide of the invention to other DGAT1 sequences enhance lipid production in Brassica napus relative to that with the other DGAT1 sequences alone.*

Fragments (*N*-terminal region or C-terminal region) of the *Z. mays*-L polypeptide of the invention can also be combined with fragments of other plant DGAT1s to raise the oil content in *Brassica napus* seeds. Tables 21-22 show the seed oil contents from a variety of transgenic plants expressing such chimeric DGAT1s. In Table 6 the seed oil contents are presented both as a % of Dry Matter (DM) and as a normalised percentage of the seed oil content of the corresponding unmodified DGAT1 parents.

Table 7

| Construct description | Transgenic plant ID # | Seed Oil as % DM | Oil Increase as % of Vector Control | Oil Increase as % of *N*-terminal DGAT1 Parent | Oil Increase as % of *C*-terminal DGAT1 Parent |
|---|---|---|---|---|---|
| Vector control | CV | 37.99 | 0.00 | N/A | N/A |
| *T. majus* | N2 | 39.07 | 2.84 | N/A | N/A |
| *Z. mays*-L | N6 | 38.96 | 2.55 | N/A | N/A |
| Tm-ZmS | 182-38-4 | 44.66 | 17.56 | 14.31 | 10.96 |

| | | | | | |
|---|---|---|---|---|---|
| Tm-ZmL | 183-60-6 | 44.47 | 17.06 | 13.82 | 14.14 |
| ZmL-Tm | 185-24-5 | 45.27 | 19.16 | 16.20 | 15.87 |
| ZmL-Tm | 185-24-9 | 45.14 | 18.82 | 15.86 | 15.54 |
| ZmL-Tm | 185-22-1 | 44.23 | 16.43 | 13.53 | 13.21 |
| ZmL-Tm | 185-22-4 | 43.20 | 13.71 | 10.88 | 10.57 |
| ZmL-Tm | 185-22-9 | 43.49 | 14.48 | 11.63 | 11.31 |
| ZmL-Tm | 185-14-10 | 44.77 | 17.85 | 14.91 | 14.59 |
| ZmL-Tm | 185-9-9 | 43.73 | 15.11 | 12.24 | 11.93 |
| ZmL-Tm | 185-8-4 | 44.02 | 15.87 | 12.99 | 12.67 |
| ZmL-Tm | 185-8-7 | 45.11 | 18.74 | 15.79 | 15.46 |
| ZmL-Tm | 185-8-8 | 44.62 | 17.45 | 14.53 | 14.21 |
| ZmL-Tm | 185-8-9 | 43.48 | 14.45 | 11.60 | 11.29 |

In Table 7 the oil contents are presented both on a % of DM basis and as a normalised percentage of the seed oil content of the corresponding segregating null sibling.

Table 8

| Construct description | Transgenic ID # | Seed Oil as % DM | Oil Increase as % of Null Sibling |
|---|---|---|---|
| Tm-ZmL | 183-17-10 | 43.8 | 29.43 |
| Tm-ZmL Null Sib | 183-17-4 | 33.84 | N/A |
| ZmL-Tm | 185-24-5 | 45.27 | 19.41 |
| ZmL-Tm | 185-24-9 | 45.14 | 19.07 |
| ZmL-Tm Null Sib | 185-24-10 | 37.91 | N/A |
| ZmL-Tm | 185-22-1 | 44.23 | 30.09 |
| ZmL-Tm | 185-22-4 | 43.2 | 27.06 |
| ZmL-Tm | 185-22-9 | 43.49 | 27.91 |
| ZmL-Tm Null Sib | 185-22-2 | 34 | N/A |
| ZmL-Tm | 185-9-9 | 43.73 | 15.60 |
| ZmL-Tm Null Sib | 185-9-8 | 37.83 | N/A |

Together these results show that addition of fragments (N-terminal or C-terminal) of the *Z. mays* DGAT1-L polypeptide of the invention can be added to parts of the *T.*majus DGATS1 sequence to increase oil yield relative to that produced by the full length *T. majus* DGAT1.

*Discussion*

The applicants have thus shown that the novel *Z. mays* DGAT1-L protein of the invention can be used to manipulate cellular lipid accumulation. The DGAT1 of the invention also has higher activity in increasing cellular lipid content than any other DGAT1 proteins tested by the applicants. The applicants have also shown that subsequences, or fragments, of the DGAT1 of the invention can be combined with parts of other DGAT1 to increase activity over that shown over the other DGAT1 sequences.

PCT/IB2013/059525
Received 28/08/2014

**CLAIMS:**

1. An isolated polynucleotide encoding a DGAT1 polypeptide comprising at least one of:
   a) a sequence with at least 95% identity to the sequence of SEQ ID NO:39
   b) the sequence of SEQ ID NO:39
   c) a sequence comprising at least 450 contiguous amino acids of the sequence of a) or b).

2. The isolated polynucleotide of claim 1 or 2 wherein the polypeptide has DGAT1 activity.

3. The isolated polynucleotide of any one of claims 1 to 2, wherein the DGAT1 polypeptide, when expressed in the cell has at least one of
   a) higher DGAT1 activity than at least one other DGAT1 protein, and
   b) altered substrate specificity relative to at least one other DGAT1 protein.

4. The isolated polynucleotide of claim 3 wherein the at least one other DGAT1 protein has the amino acid sequence of the polypeptide of SEQ ID NO:44 ( ZmDGAT1-short).

5. The isolated polynucleotide of any one of claims 1 to 4, wherein the DGAT1 polypeptide, has at least one of
   a) higher DGAT1 activity than any previously known DGAT1 protein, and
   b) altered substrate specificity relative to any previously known DGAT1 protein.

6. A genetic construct comprising a polynucleotide of any one of claims 1 to 5.

7. The genetic construct of claim 6 in which the polynucleotide is operably linked to a promorer.

8. The genetic construct of claim 7 in which the polynucleotide and the promoter are not found operably linked in nature.

9. A cell comprising a polynucleotide of the invention of any one of claims 1 to 5.

PCT/IB2013/059525
Received 28/08/2014

10. The cell of claim 9 wherein the cell, or its predecessor, is transformed to comprise the polynucleotide.

11. A cell comprising a genetic construct of any one of claims 6 to 8.

12. The cell of any one of claims 9 to 11 that expresses the polynucleotide of any one of claims 1 to 5.

13. The cell of claim 12 that expresses the DGAT1 polypeptide.

14. The cell of any one of claims 9 to 13 wherein the expressed DGAT1 polypeptide has increased DGAT1 activity than one or both of:

a) at least one other DGAT1 protein, and

b) any other previously known DGAT1 protein

15. The cell of any one of claims 9 to 14 wherein the expressed DGAT1 polypeptide has altered substrate specificity relative to one or both of:

a) at least one other DGAT1 protein, and

b) any other previously known DGAT1 protein.

16. The cell of any one of claims 9 to 15 that produces more lipid than does a control cell.

17. The cell of any one of claims 9 to 16 that has an altered lipid profile relative to a control cell.

18. The cell of any one of claims 9 to 17 that is a plant cell.

19. The cell of any one of claims 9 to 18 that is also transformed to express at least one of: an oleosin, steroleosin, caloleosin, polyoleosin, and an oleosin including at least one artificially introduced cysteine.

20. A plant comprising a polynucleotide of any one of claims 1 to 5.

21. The plant of claim 20 wherein the plant, or its predecessor, has been transformed to comprise the polynucleotide.

PCT/IB2013/059525
Received 28/08/2014

22. A plant comprising a genetic construct of any one of claims 6 to 8.

23. A plant comprising a plant cell of claim 18.

24. The plant of any one of claims 20 to 23 wherein plant, or its predecessor, has been transformed or genetically modified to expresses the polynucleotide or endcoded polypeptide.

25. The plant of any one of claims 20 to 24 wherein the expressed polypeptide, has increased DGAT1 activity relative to at least one other DGAT1 protein.

26. The plant of any one of claims 20 to 25 wherein the plant produces more lipid, in at least one of its tissues or parts, or as a whole, than does a control plant.

27. The plant of any one of claims 20 to 26 wherein the plant has an altered lipid profile, in at least one of its tissues or parts, or as a whole, relative to a control plant.

28. The plant of any one of claims 20 to 27 wherein the plant is also transformed to express at least one of: an oleosin, a steroleosin, a caloleosin, a polyoleosin, and an oleosin including at least one artificially introduced cysteine.

29. An isolated polypeptide comprising at least one of:

a) a sequence with at least 95% identity to the sequence of SEQ ID NO:39

b) the sequence of SEQ ID NO:39

c) a sequence comprising at least 450 contiguous amino acids of the sequence of a) or b).

30. The polypeptide of claim 29 wherein the polypeptide has DGAT1 activity.

31. A part, propagule or progeny of the plant of any one of claims 20 to 28.

32. A part, propagule or progeny of claim 31 wherein the part, propagule or progeny comprises at least one of a polynucleotide, construct, cell, or polypeptide as defined in any one of claims 1 to 19 and 29 to 30.



AMENDED SHEET
IPEA/AU

PCT/IB2013/059525
Received 28/08/2014

33. The part, propagule or progeny of claim 31 or 32 wherein the part, propagule or progeny, or its predecessor plant, has been transformed to comprise the polynucleotide or construct as defined in any one of claims 1 to 8.

34. The part, propagule or progeny of any one of claims 31 or 33 wherein the part, propagule or progeny expresses the polynucleotide or polypeptide as defined in any one of claims 1 to 5 and 29 to 30.

35. The part, propagule or progeny of any one of claims 31 or 34 wherein the part, propagule or progeny produces more lipid than does a control part, propagule or progeny, or part, propagule or progeny of a control plant.

36. The part, propagule or progeny of any one of claims 31 or 35 wherein the part, propagule or progeny has an altered lipid profile relative to a control part, propagule or progeny, or part, propagule or progeny of a control plant.

37. An animal feedstock comprising at least one of a polynucleotide, polypeptide, construct, cell, plant cell, plant, plant part, propagule and progeny as defined in any one of claims 1 to 36.

38. A biofuel feedstock comprising at least one of a polynucleotide, polypeptide, construct, cell, plant cell, plant, plant part, propagule and progeny as defined in any one of claims 1 to 36.

39. A method for producing a lipid, the method comprising growing a cell, plant cell or plant that is transformed, or genetically modified, to express and polynucleotide or polypeptide as defined in any one of claims 1 to 5 and 29 to 30, wherein the plant produces oil through the activity of the expressed polypeptide.

40. The method of claim 39 wherein the cell, plant cell or plant produces the lipid as a result of the DGAT1 activity of the polypeptide.

41. A method for producing lipid, the method comprising extracting lipid from at least one of a cell, plant cell, plant, plant part, propagule and progeny as defined in any one of claims 9 to 36.

PCT/IB2013/059525
Received 28/08/2014

42. The method any one of of claims 39 to 41 wherein the lipid is triacylglycerol (TAG).

43. The method of any one of claims 39 70 42 wherein the lipid is processed into at least one of:

a) a fuel,

b) an oleochemical,

c) a nutritional oil,

d) a cosmetic oil,

e) a polyunsaturated fatty acid (PUFA), and

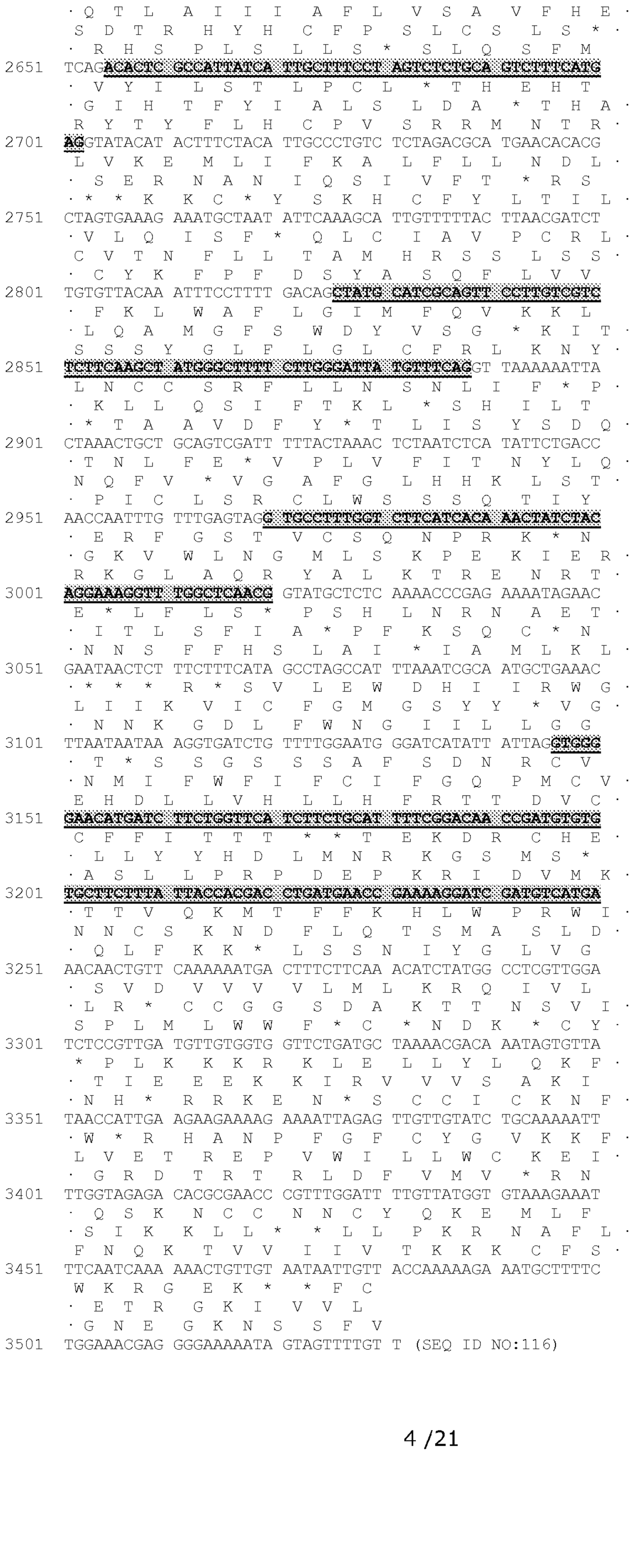
f) a combination of any of a) to e).

# Figure 1

```
      N  P  F   S  F  L   L  L  L  L   F  R  E   N  F  A
     E  S  F   F  L  S  S   S  S  S   L  Q  R   K  L  C  F ·
    *  I  L  F   P  F  F   F  F  F   S  S  E   K  T  L  L ·
 1  TGAATCCTTT TTCCTTTCTT CTTCTTCTTC TCTTCAGAGA AAACTTTGCT
     S  L  S  I   R  N  Q   T  R  I   P  F  P   P  I  S  * ·
    ·  S  F  Y   K  E  P   D  T  N   P  I  P   T  D  F  L
    ·L  F  L  *   G  T  R   H  E  S   H  S  H   R  F  L  S ·
51  TCTCTTTCTA TAAGGAACCA GACACGAATC CCATTCCCAC CGATTTCTTA
    ·L  L  P   S  I  R  S   F  P  L   H  *  I   L  F  P  L ·
     A  S  S   F  N  P  L   F  P  S   P  L  D   S  V  S  S ·
    ·  F  F  L   Q  S  A   L  S  L   S  I  R   F  C  F  L
101 GCTTCTTCCT TCAATCCGCT CTTTCCCTCT CCATTAGATT CTGTTTCCTC
    ·  S  I  S   S  A  C   F  S  I   L  S  D   A  S  F  L
    ·F  N  F  F   C  M  L   L  D  S   L  *  R   L  F  S  P ·
     F  Q  F  L   L  H  A   S  R  F   S  L  T   P  L  F  S ·
151 TTTCAATTTC TTCTGCATGC TTCTCGATTC TCTCTGACGC CTCTTTCTC
     P  T  L  F   R  Q  T   L  F  E   M  A  I   L  D  S  A ·
    ·  D  A  V   S  S  N   A  F  R   N  G  D   F  G  F  C
    ·R  R  C  F   V  K  R   F  S  K   W  R  F   W  I  L  L ·
201 CCGACGCTGT TTCGTCAAAC GCTTTTCGAA ATGGCGATTT TGGATTCTGC
    ·G  V  T  T   V  T  E   N  G  G   G  E  F   V  D  L  D ·
     W  R  Y  Y   G  D  G   E  R  W   R  R  V   R  R  S  * ·
    ·  A  L  L   R  *  R   R  T  V   A  E  S   S  S  I  L
251 TGGCGTTACT ACGGTGACGG AGAACGGTGG CGGAGAGTTC GTCGATCTTG
    ·  R  L  R   R  R  K   S  R  S   D  S  S   N  G  L  L
    ·*  A  S  S   T  E  I   E  I  G   F  F  *   R  T  S  S ·
     I  G  F  V   D  G  N   R  D  R   I  L  L   T  D  F  F ·
301 ATAGGCTTCG TCGACGGAAA TCGAGATCGG ATTCTTCTAA CGGACTTCTT
     L  S  G  S   D  N  N   S  P  S   D  D  V   G  A  P  A ·
    ·  L  W  F   R  *  *   F  S  F   G  *  C   W  S  S  R
    ·S  L  V  P   I  I  I   L  L  R   M  M  L   E  L  P  P ·
351 CTCTCTGGTT CCGATAATAA TTCTCCTTCG GATGATGTTG GAGCTCCCGC
    ·D  V  R  D   R  I  D   S  V  V   N  D  D   A  Q  G  T ·
     R  R  *  G   S  D  *   F  R  C   *  R  *   R  S  G  N ·
    ·  T  L  G   I  G  L   I  P  L   L  T  M   T  L  R  E
401 CGACGTTAGG GATCGGATTG ATTCCGTTGT TAACGATGAC GCTCAGGGAA
    ·  A  N  L   A  G  D   N  N  G   G  G  D   N  N  G  G
    ·S  Q  F  G   R  R  *   *  R  W   W  R  *   *  R  W  W ·
     Q  P  I  W   P  E  I   I  T  V   V  A  I   I  T  V  V ·
451 CAGCCAATTT GGCCGGACAT AATAACGGTG GTGGCGATAA TAACGGTGGT
     G  R  G  G   G  E  G   R  G  N   A  D  A   T  F  T  Y ·
    ·  K  R  R   R  R  R   K  R  K   R  C  Y   V  Y  V
    ·E  E  A  A   E  K  E   E  T  P   M  L  R   L  R  I ·
501 CGAAGAGGCG GCGGAGAAGC AAGAGGAAAC GCCGATGCTA CGTTTACGTA
    ·R  P  S  V   P  A  H   R  R  A   R  E  S   P  L  S  S ·
     S  T  V  G   S  S  S   S  E  G   E  R  E   S  T  *  L ·
    ·  D  R  R   F  Q  L   I  G  G   R  E  R   V  H  L  A
551 TCGACCGTCG GTTCCAGCTC ATCGGAGGGC GAGAGAGAGT CCACTTAGCT
    ·  D  A  I   F  K  Q   V  *  N   L  R  N   L  R  I  W
    ·R  R  N  L   Q  T  G   L  K  S   Q  K  S   S  N  L  V ·
     P  T  Q  S   S  N  R   F  K  I   S  E  I   F  E  F  G ·
601 CCGACGCAAT CTTCAAACAG GTTTAAAATC TCAGAAATCT TCGAATTTGG
     C  L  L  V   V  L  Y   G  I  E   F  G  D   C  F  A  L ·
    ·  F  A  C   C  F  I   W  N  *   V  W  *   L  F  C  I
    ·V  C  L  L   F  Y  M   E  L  S   L  V  I   V  L  H  C ·
651 TGTTTGCTTG TTGTTTTATA TGGAATTGAG TTTGGTGATT GTTTTGCATT
    ·Q  S  H  A   G  L  F   N  L  C   V  V  V   L  I  A  V ·
     A  E  P  C   R  I  I   Q  P  L   C  S  S   S  Y  C  C ·
    ·  R  A  M   P  D  Y   S  T  S   V  *  *   F  L  L  L
701 GCAGAGCCAT GCCGGATTAT TCAACCTCTG TGTAGTAGTT CTTATTGCTG
    ·  N  S  R   L  I  I   E  N  L   M  K  V   C  C  Y  L
    ·K  Q  *  T   H  H  R   K  S  Y   E  G  L   L  L  L  V ·
     *  T  V  D   S  S  S   K  I  L   *  R  F   A  V  T  C ·
751 TAAACAGTAG ACTCATCATC GAAAATCTTA TGAAGGTTTG CTGTTACTTG
     F  L  L  L   G  I  E   L  L  E   N  L  S   E  T  N  N ·
    ·  S  P  F   R  N  *   I  A  *   K  F  I   R  D  E  *
    ·F  S  F  *   E  L  N   C  L  K   I  Y  Q   R  R  I  T ·
801 TTTCTCCTTT TAGGAATTGA ATTGCTTGAA AATTTATCAG AGACGAATAA
```

```
     · F  V  V   A  I  I  H   V  V  W   L  V  D   Q  N  G  F ·
      L  C  C  C   Y  H  S   C  S  M   V  G  *   S  E  R  I ·
     · L  L  L   L  S  F  M   *  Y  G   W  L  I   R  T  D
851  CTTTGTTGTT GCTATCATTC ATGTAGTATG GTTGCTTGAT CAGAACGCAT
     · L  V  *   F  K  I  A   A  R  L   A  A  F   H  V  L
     · S  G  L   V  Q  D  R   C  E  I   G  R  F   S  C  V  G ·
      F  W  F   S   S  R  S   L  R  D   W  P  L   F  M  C  W ·
901  TTCTGGTTTA GTTCAAGATC GCTGCGAGAT TGGCCGCTTT TCATGTGTTG
      V  K  E   D  V  F  Y   F  Q  Q   C  Y  I   V  I  R  I ·
     · K  R  R   C  F  L  F   P  A  M   L  H  C   Y  T  Y
     · *  K  K   M  F  F  I   S  S  N   V  T  L   L  Y  V  * ·
951  GTAAAAGAAG ATGTTTTTTA TTTCCAGCAA TGTTACATTG TTATACGTAT
     · M  M  S   L  V  I  K   F  L  F   D  S  S   F  L  L  Q ·
      N  D  E   F   S  D  Q   V  P  L   *  F  F   F  L  V  A ·
     · *  *  V   *  *  S  S   S  S  S   L  I  L   L  S  C  C
1001 AATGATGAGT TTAGTGATCA AGTTCCTCTT TGATTCTTCT TTCTTGTTGC
     · Y  I  P   F  D  L  S   F  G  C   L  Y  G   *  E  I
     · V  Y  P   F  R  S  F   L  W  L   P  L  R   L  R  N  W ·
      S  I  S   L   S  I  F   P  L  A   A  F  T   V  E  K  L ·
1051 AGTATATCCC TTTCGATCTT TCCTTTGGCT GCCTTTACGG TTGAGAAATT
      G  T  S   E  I  H  I   R  T  C   E  *  L   L  F  S  S ·
     · Y  F  R   N  T  Y  Q   N  L  *   V  I  T   I  L  Q
     · V  L  Q   K  Y  I  S   E  P  V   S  N  Y   Y  S  P  A ·
1101 GGTACTTCAG AAATACATAT CAGAACCTGT GAGTAATTAC TATTCTCCAG
     · H  Y  C   N  F  Y  *   R  Q  V   C  I  M   K  N  L  Q ·
      P  L  L   *   F  L  L   K  T  S   L  Y  H   E  E  L  T ·
     · I  T  V   I  F  I  E   D  K  F   V  S  *   R  T  Y
1151 CCATTACTGT AATTTTTATT GAAGACAAGT TTGTATCATG AAGAACTTAC
     · V  L  F   *  K  C  S   R  L  S   S  F  F   I  L  L
     · S  S  V   L  K  M  L   K  V  V   I  F  L   H  I  I  I ·
      K  F  C   F   E  N  A   Q  G  C   H  L  S   S  Y  Y  Y ·
1201 AAGTTCTGTT TTGAAAATGC TCAAGGTTGT CATCTTTCTT CATATTATTA
      S  P  *   Q   R  F  C   I  Q  F   T  S  P   *  G  D  T ·
     · T  M  T   E  V  L  Y   P  V  Y   V  T  L   R  *  Y
     · H  H  D   R  G  F  V   S  S  L   R  H  P   K  V  I  L ·
1251 TCACCATGAC AGAGGTTTTG TATCCAGTTT ACGTCACCCT AAGGTGATAC
     · V  F  L   V  S  V  C   D  T  V   F  K  F   S  C  L  T ·
      C  F  S   G   L  S  L   *  Y  C   F  *  V   *  L  S  D ·
     · F  F  W   S  Q  F  V   I  L  F   L  S  L   V  V  *
1301 TGTTTTTCTG GTCTCAGTTT GTGATACTGT TTTTAAGTTT AGTTGTCTGA
     · R  *  S   *  K  W  T   G  V  I   L  L  F   Y  Q  V
     · P  V  I   L  K  M  D   R  C  D   S  A  F   L  S  G  V ·
      P  G  D   L   E  N  G   Q  V  *   F  C  F   F  I  R  C ·
1351 CCCGGTGATC TTGAAAATGG ACAGGTGTGA TTCTGCTTTT TTATCAGGTG
      S  L  *   C   S  S  L   A  L  C   G  *  S   W  F  L  M ·
     · T  L  M   L  L  T  C   I  V  W   L  K  L   V  S  Y
     · H  F  D   A  P  H  L   H  C  V   A  K  V   G  F  L  C ·
1401 TCACTTTGAT GCTCCTCACT TGCATTGTGT GGCTAAAGTT GGTTTCTTAT
     · L  I  L   A  M  T  *   D  P  *   P  M  Q   L  I  R  * ·
      A  H  T   S   Y  D  I   R  S  L   A  N  A   A  D  K  V ·
     · S  Y  *   L  *  H  K   I  P  S   Q  C  S   *  *  G
1451 GCTCATACTA GCTATGACAT AAGATCCCTA GCCAATGCAG CTGATAAGGT
     · N  T  K   K  K  R  M   Y  *  S   L  A  L   C  Y  C
     · K  Y  E   K  E  A  Y   V  L  V   T  C  T   V  L  L  F ·
      K  I  R   K   R  S  V   C  I  S   H  L  H   C  V  T  V ·
1501 AAAATACGAA AAAGAAGCGT ATGTATTAGT CACTTGCACT GTGTTACTGT
      F  N  Q   T   L  L  *   T  L  G   Q  S  *   S  L  L  L ·
     · *  P  N   T  V  M  N   F  R  P   I  L  K   S  P  T
     · L  T  K   H  C  Y  E   L  *  A   N  P  E   V  S  Y  Y ·
1551 TTTAACCAAA CACTGTTATG AACTTTAGGC CAATCCTGAA GTCTCCTACT
     · R  *  L   E  E  L  G   I  F  H   G  R  S   H  I  V  L ·
      T  L  A   *   R  A  W   H  I  S   W  S  L   P  H  C  V ·
     · V  S  L   K  S  L  A   Y  F  M   V  A  P   T  L  C
1601 ACGTTAGCTT GAAGAGCTTG GCATATTTCA TGGTCGCTCC CACATTGTGT
     · S  G  N   C  K  V  H   Q  P  F   L  Y  L   Q  E  F
     · I  R  *   L  Q  S  A   S  T  I   L  I  L   A  R  V  S ·
      Y  Q  V   T   A  K  C   I  N  H   S  Y  T   C  K  S  F ·
1651 TATCAGGTAA CTGCAAAGTG CATCAACCAT TCTTATACTT GCAAGAGTTT
      L  V  *   T   S  D  L   C  F  S   P  A  K   L  S  T  F ·
     · C  L  N   L  G  S  L   L  F  P   S  Q  V   I  H  V
     · L  S  K   P  R  I  F   A  F  P   Q  P  S   Y  P  R  S ·
1701 CTTGTCTAAA CCTCGGATCT TTGCTTTTCC CCAGCCAAGT TATCCACGTT
```

·C M Y T E G L G G S S I C K T G H·
L H V Y G R V G W L V N L Q N W S·
·A C I R K G W V A R Q F A K L V
1751 CTGCATGTAT ACGGAAGGGT TGGGTGGCTC GTCAATTTGC AAAACTGGTC
·I H R I H G I Y N R T S T F S H
·Y S P D S W D L * * N K Y V F T S·
I F T G F M G F I I E Q V R F H I·
1801 ATATTCACCG GATTCATGGC ATTTATAATA GAACAAGTAC GTTTTCACAT
L A L L V F L G E N H H P C V V T·
·C F I S F P W * K S S S L R C H
·L L Y * F S L V K I I P A L S P·
1851 CTTGCTTTAT TAGTTTTCCT TGGTGAAAAT CATCATCCCT GCGTTGTCAC
·T * L H V L L L H F G S I * I L L·
H L T S C S F V T F W Q Y I N P I·
·L D F M F F C Y I L A V Y K S Y
1901 CACTTGACTT CATGTTCTTT TGTTACATTT TGGCAGTATA TAAATCCTAT
·S G T Q S I L * K A I F Y M L L
·V R N S K H P L K G D L L Y A I E·
C Q E L K A S F E R R S S I C Y *·
1951 TGTCAGGAAC TCAAAGCATC CTTTGAAAGG CGATCTTCTA TATGCTATTG
K E C * S F Q F Q I Y M C G S A C·
·R V L K L S V P N L Y V W L C M
·K S V E A F S S K F I C V A L H V·
2001 AAAGAGTGTT GAAGCTTTCA GTTCCAAATT TATATGTGTG GCTCTGCATG
·S T A S S T F G M L * S H L F Q N·
F Y C F F H L W Y A V I P S L S K·
·L L L L P P L V C C D P I S F K
2051 TTCTACTGCT TCTTCCACCT TTGGTATGCT GTGATCCCAT CTCTTTCAAA
·N L Q I R K T E K G * I S Y E F
·* F A N S K N R K R L N L I R I *·
I I C K F E K P K K A K S H T N L·
2101 ATAATTTGCA AATTCGAAAA ACCGAAAAAG GCTAAATCTC ATACGAATTT
D I F S F L E S V M * F Q L L N A·
·Y F * F L R V G D V I S V T E R
·I F L V S * S R * C N F S Y * T Q·
2151 GATATTTTTA GTTTCTTAGA GTCGGTGATG TAATTTCAGT TACTGAACGC
·N L L S K G * T Y W Q S F S A S G·
K S L V Q R L N I L A E L L C F G·
·I S C P K V K H I G R A S L L R
2201 AAATCTCTTG TCCAAAGGTT AAACATATTG GCAGAGCTTC TCTGCTTCGG
·I V N S T K I G G M Q K V W E M
·D R E F Y K D W W N A K S V G D V·
G S * I L Q R L V E C K K C G R C·
2251 GGATCGTGAA TTCTACAAAG ATTGGTGGAA TGCAAAAAGT GTGGGAGATG
* A I L L K R K L M I F N V V V V·
·S Y F T Q K K T Y D F * C C R C
·E L F Y S K E N L * F L M L S L F·
2301 TGAGCTATTT TACTCAAAAG AAAACTTATG ATTTTTAATG TTGTCGTTGT
·F G S S N * P N S C I H C L P L S·
F W V I * L T K F M Y S L S S F I·
·L G H L T N Q I H V F T V F L Y
2351 TTTTGGGTCA TCTAACTAAC CAAATTCATG TATTCACTGT CTTCCTTTAT
·V L E N V E Y G M V L F L N I T
·S T G E C G I W Y G S L P K H H L·
Q Y W R M W N M V W F S S * T S P·
2401 CAGTACTGGA GAATGTGGAA TATGGTATGG TTCTCTTCCT AAACATCACC
F F C T Q N R R R E L I K I L F S·
·L L Y T K * K K R A N * D L V F
·S F V H K I E E E S * L R S C F P·
2451 TTCTTTTGTA CACAAAATAG AAGAAGAGAG CTAATTAAGA TCTTGTTTTC
·L T A C S * M D G S T Y I L P V L·
L D S L F I N G W F D I Y T S R A·
·* Q P V H K W M V R H I Y F P C
2501 CTTGACAGCC TGTTCATAAA TGGATGGTTC GACATATATA CTTCCGTGC
·A Q Q D T K G E * D I Y R Y A I
·C A A R Y Q R * V R Y I P I C N C·
L R S K I P K V S E I Y T D M Q L·
2551 TTGCGCAGCA AGATACCAAA GGTGAGTGAG ATATATACCG ATATGCAATT
V E I C F C D I N L T L H T L V F·
·R D L F L * Y K F N P P H T C F
·S R F V S V I * I * P S T H L F F·
2601 GTCGAGATTT GTTTCTGTGA TATAAATTTA ACCCTCCACA CACTTGTTTT


· Q T L A I I I A F L V S A V F H E ·
S D T R H Y H C F P S L C S L S * ·
· R H S P L S L L S * S L Q S F M
2651 TCAGACACTC GCCATTATCA TTGCTTTCGT AGTCTCTGCA GTCTTTCATG
· V Y I L S T L P C L * T H E H T
· G I H T F Y I A L S L D A * T H A ·
R Y T Y F L H C P V S R R M N T R ·
2701 AGGTATACAT ACTTTCTACA TTGCCCTGTC TCTAGACGCA TGAACACACG
L V K E M L I F K A L F L L N D L ·
· S E R N A N I Q S I V F T * R S
· * * K K C * Y S K H C F Y L T I L ·
2751 CTAGTGAAAG AAATGCTAAT ATTCAAAGCA TTGTTTTTAC TTAACGATCT
· V L Q I S F * Q L C I A V P C R L ·
C V T N F L L T A M H R S S L S S ·
· C Y K F P F D S Y A S Q F L V V
2801 TGTGTTACAA ATTTCCTTTT GACAGCTATG CATCGCAGTT CCTTGTCGTC
· F K L W A F L G I M F Q V K K L
· L Q A M G F S W D Y V S G * K I T ·
S S S Y G L F L G L C F R L K N Y ·
2851 TCTTCAAGCT ATGGCCTTTT CTTCGCATTA TGTTTCAGGT TAAAAAATTA
L N C C S R F L L N S N L I F * P ·
· K L L Q S I F T K L * S H I L T
· * T A A V D F Y * T L I S Y S D Q ·
2901 CTAAACTGCT GCAGTCGATT TTTACTAAAC TCTAATCTCA TATTCTGACC
· T N L F E * V P L V F I T N Y L Q ·
N Q F V * V G A F G L H H K L S T ·
· P I C L S R C L W S S S Q T I Y
2951 AACCAATTTG TTTGAGTAGC TGCCTTTGGT CTTCATCACA AACTATCTAC
· E R F G S T V C S Q N P R K * N
· G K V W L N G M L S K P E K I E R ·
R K G L A Q R Y A L K T R E N R T ·
3001 AGGAAAGCTT TGGCTCAACG GTATGCTCTC AAAACCCGAG AAAATAGAAC
E * L F L S * P S H L N R N A E T ·
· I T L S F I A * P F K S Q C * N
· N N S F F H S L A I * I A M L K L ·
3051 GAATAACTCT TTCTTTCATA GCCTAGCCAT TTAAATCGCA ATGCTGAAAC
· * * * R * S V L E W D H I I R W G ·
L I I K V I C F G M G S Y Y * V G ·
· N N K G D L F W N G I I L L G G
3101 TTAATAATAA AGGTGATCTG TTTTGGAATG GGATCATATT ATTAGGTGGG
· T * S S G S S S A F S D N R C V
· N M I F W F I F C I F G Q P M C V ·
E H D L L V H L L H F R T T D V C ·
3151 GAACATGATC TTCTGGTTCA TCTTCTGCAT TTTCGGACAA CCGATGTGTG
C F F I T T T * * T E K D R C H E ·
· L L Y Y H D L M N R K G S M S *
· A S L L P R P D E P K R I D V M K ·
3201 TGCTTCTTTA TTACCACGAC CTGATGAACC GAAAAGGATC GATGTCATGA
· T T V Q K M T F F K H L W P R W I ·
N N C S K N D F L Q T S M A S L D ·
· Q L F K K * L S S N I Y G L V G
3251 AACAACTGTT CAAAAAATGA CTTTCTTCAA ACATCTATGG CCTCGTTGGA
· S V D V V V V L M L K R Q I V L
· L R * C C G G S D A K T T N S V I ·
S P L M L W W F * C * N D K * C Y ·
3301 TCTCCGTTGA TGTTGTGGTG GTTCTGATGC TAAAACGACA AATAGTGTTA
* P L K K K R K L E L L Y L Q K F ·
· T I E E E K K I R V V V S A K I
· N H * R R K E N * S C C I C K N F ·
3351 TAACCATTGA AGAAGAAAAG AAAATTAGAG TTGTTGTATC TGCAAAAATT
· W * R H A N P F G F C Y G V K K F ·
L V E T R E P V W I L L W C K E I ·
· G R D T R T R L D F V M V * R N
3401 TTGGTAGAGA CACGCGAACC CGTTTGGATT TTGTTATGGT GTAAAGAAAT
· Q S K N C C N N C Y Q K E M L F
· S I K K L L * * L L P K R N A F L ·
F N Q K T V V I I V T K K K C F S ·
3451 TTCAATCAAA AAACTGTTGT AATAATTGTT ACCAAAAAGA AATGCTTTTC
W K R G E K * * F C
· E T R G K I V V L
· G N E G K N S S F V
3501 TGGAAACGAG GGGAAAAATA GTAGTTTTGT T (SEQ ID NO:116)

# Figure 2

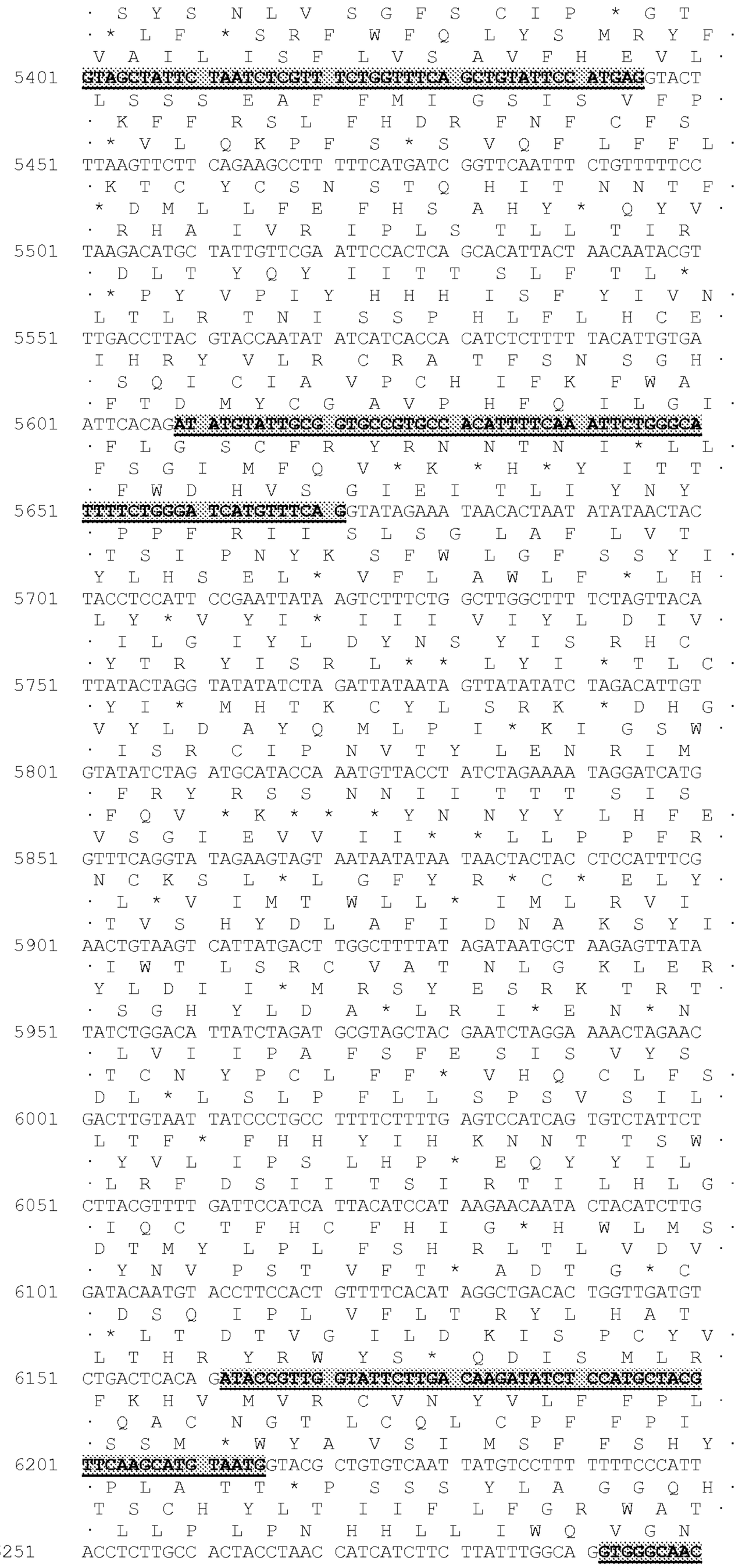
```
      G  P  A  P  L  H  A  C  R  L  R  S  R  R  P  W
     W  P  R  P  P  P  C  L  P  P  P  I  A  P  A  L  A ·
    M  A  P  P  P  S  M  P  A  A  S  D  R  A  G  P  G ·
1   ATGGCCCCGC CCCCTCCAT GCCTGCCGCC TCCGATCGCG CCGGCCCTGG
     P  R  R  G  R  L  V  L  P  S  P  P  P  R  P  L  S ·
    · A  T  R  A  T  R  P  P  F  A  S  A  A  P  P  Q
    ·R  D  A  G  D  S  S  S  L  R  L  R  R  A  P  S  A ·
51  CCGCGACGCG GGCGACTCGT CCTCCCTTCG CCTCCGCCGC GCCCCCTCAG
    ·R  R  R  R  P  C  R  R  F  L  G  R  L  A  G  E  R ·
     P  T  P  A  T  L  P  A  I  P  R  *  A  C  G  R  T ·
    · D  A  G  D  L  A  G  D  S  S  V  G  L  R  E  N
101 CCGACGCCGG CGACCTTGCC GGCGATTCCT CGGTAGGCTT GCGGGAGAAC
    · R  A  A  T  A  D  E  S  A  A  A  G  A  A  A  A
    ·A  S  R  N  R  R  R  I  R  R  R  S  S  S  S  S ·
     G  E  P  Q  P  P  T  N  P  P  P  Q  E  Q  Q  Q  Q ·
151 GGCGAGCCGC AACCGCCGAC GAATCCGCCG CCGCAGGAGC AGCAGCAGCA
     A  R  D  A  I  L  P  R  V  G  A  R  P  P  P  R  Q ·
    · T  R  C  Y  T  T  A  R  R  R  P  P  T  A  A  S
    ·H  E  M  L  Y  Y  R  A  S  A  P  A  H  R  R  V  K ·
201 GCACGAGATG CTATACTACC GCGCGTCGGC GCCCGCCCAC CGCCGCGTCA
    ·G  E  P  P  Q  L  *  R  H  L  P  A  G  E  E  T  R ·
     R  R  A  P  S  A  L  T  P  S  S  G  R  *  G  D  A ·
    · E  S  P  L  S  S  D  A  I  F  R  Q  V  R  R  R
251 AGGAGAGCCC CCTCAGCTCT GACGCCATCT TCCGGCAGGT GAGGAGACGC
    · I  L  G  S  L  F  V  S  D  C  L  I  P  A  L  V
    ·N  F  R  L  A  V  C  K  R  L  F  D  P  R  A  C  A ·
     E  F  *  A  R  C  L  *  A  I  V  *  S  P  R  L  C ·
301 GAATTTTAGG CTCGCTGTTT GTAAGCGATT GTTTGATCCC CGCGCTTGTG
     L  R  S  T  P  V  A  K  S  C  K  L  F  V  A  S  S ·
    · S  I  H  A  S  C  K  I  L  Q  I  V  C  C  F  Q
    ·F  D  P  R  Q  L  Q  N  P  A  N  C  L  L  L  P  V ·
351 CTTCGATCCA CGCCAGTTGC AAAATCCTGC AAATTGTTTG TTGCTTCCAG
    ·Q  L  C  L  C  F  F  L  V  G  V  C  V  C  V  C  V ·
     S  T  L  P  L  F  F  F  G  W  C  V  C  V  C  V  C ·
    · N  S  A  S  V  F  F  W  L  V  C  V  C  V  C  V
401 TCAACTCTGC CTCTGTTTTT TTTTGGTTGG TGTGTGTGTG TGTGTGTGTG
    · Q  I  T  L  C  A  I  G  S  L  T  L  P  V  A  I
    ·S  N  H  T  L  C  Y  R  *  L  N  T  A  G  C  H  L ·
     F  K  S  H  F  V  L  S  V  A  *  H  C  R  L  P  S ·
451 TTCAAATCAC ACTTTGTGCT ATCGGTAGCT TAACACTGCC GGTTGCCATC
     S  R  A  R  M  F  Y  C  G  P  W  A  S  E  L  W  I ·
    · A  R  T  D  V  L  L  W  A  L  G  F  G  I  V  D
    ·R  A  H  G  C  F  I  V  G  L  G  L  R  N  C  G  *  ·
501 TCGCGCGCAC GGATGTTTTA TTGTGGGCCT TGGGCTTCGG AATTGTGGAT
    ·D  C  A  R  V  L  E  W  A  Q  F  V  S  W  G  A  Y ·
     R  L  C  A  C  T  R  M  G  T  I  R  F  V  G  G  I ·
    · I  V  R  V  Y  S  N  G  H  N  S  F  R  G  G  H
551 AGATTGTGCG CGTGTACTCG AATGGGCACA ATTCGTTTCG TGGGGGGCAT
    · A  A  A  I  E  V  G  V  Y  L  F  W  D  Q  G  D
    ·C  C  C  D  *  G  R  C  L  L  V  L  G  S  G  G  P ·
     M  L  L  R  L  R  S  V  F  T  C  F  G  I  R  G  T ·
601 ATGCTGCTGC GATTGAGGTC GGTGTTTACT TGTTTTGGGA TCAGGGGGAC
     Q  C  R  C  A  G  A  R  C  M  P  R  R  I  W  H  R ·
    · V  P  V  R  G  C  Q  M  H  A  T  Q  N  L  A  S
    ·S  A  G  A  R  V  P  D  A  C  H  A  E  F  G  I  G ·
651 CAGTGCCGGT GCGCGGGTGC CAGATGCATG CCACGCAGAA TTTGGCATCG
    ·P  A  E  A  A  N  N  E  R  N  R  Y  H  W  R  S  F ·
     A  G  *  S  S  K  Q  R  A  *  P  L  P  L  E  E  L ·
    · R  L  K  Q  Q  T  T  S  V  T  V  T  T  G  G  A
701 GCCGGCTGAA GCAGCAAACA ACGAGCGTAA CCGTTACCAC TGGAGGAGCT
    · G  L  S  K  R  M  T  G  *  A  N  E  S  L  N  S
    ·W  L  V  E  T  D  D  W  M  S  E  *  I  I  E  F  I ·
     L  A  C  R  N  G  *  L  D  E  R  M  N  H  *  I  H ·
751 TTGGCTTGTC GAAACGGATG ACTGGATGAG CGAATGAATC ATTGAATTCA
     L  L  A  V  L  T  I  V  M  W  T  V  V  G  T  A  P ·
    · V  G  G  T  H  Y  S  D  V  D  S  C  W  D  S  T
    ·C  W  R  Y  S  L  *  *  C  G  Q  L  L  G  Q  H  L ·
801 TTGTTGGCGG TACTCACTAT AGTGATGTGG ACAGTTGTTG GGACAGCACC
    ·A  V  P  P  V  L  L  M  L  T  F  L  T  T  M  R  V ·
     C  S  A  P  S  I  I  N  A  D  F  S  N  Y  N  A  C ·
    · Q  C  P  Q  Y  Y  *  C  *  L  F  *  L  Q  C  V
851 TGCAGTGCCC CCAGTATTAT TAATGCTGAC TTTTCTAACT ACAATGCGTG
```

```
     · T  L  F   V  H  L   G  F  P  A   W  G  I   A  S  C
     ·Y  I  V   C  T  P  W   L  S  C   L  G  H   C  F  L  L ·
      L  H  C   L  Y  T  L   A  F  L   L  G  A   L  L  V ·
901  TTACATTGTT TGTACACCTT GGCTTTCCTG CTTGGGGCAT TGCTTCTTGT
      *  G  P   Y  N  C  A   P  T  *   N  C  I   G  P  L  V ·
     · R  T  I   *  L  C   T  Y  I  E   L  Y  W   T  T  C
     ·E  D  H   I  T  V  H   L  H  R   T  V  L   D  H  L  *  ·
951  TGAGGACCAT ATAACTGTGC ACCTACATAG AACTGTATTG GACCACTTGT
     ·S  F  N   W  L  A  L   H  F  L   I  G  I   L  L  D  N ·
      K  F  *   L  V  S  P   P  F  F   N  R  Y   I  I  R  Q ·
     · V  L  T   G  *  P   S  I  F  *   *  V  Y   Y  *  T
1001 AAGTTTTAAC TGGTTAGCCC TCCATTTTTT AATAGGTATA TTATTAGACA
     · F  Y  C   H  *  H   Y  F  C  L   L  L  S   E  P  F
     ·F  L  L   S  L  T  L   F  L  F   A  T  L   G  A  L  F ·
      I  F  I   V  I  D  I   I  F  V   C  Y  S   R  S  P  F ·
1051 ATTTTTATTG TCATTGACAT TATTTTTGTT TGCTACTCTC GGAGCCCTTT
      S  Q  C   N  L  N  R   A  Q  I   T  A  E   T  R  E  T ·
     · P  V  *   S  *  *   G  S  N  H   S  R  N   T  *  D
     ·P  S  V   I  L  I  G   L  K  S   Q  Q  K   H  V  R  R ·
1101 TCCCAGTGTA ATCTTAATAG GGCTCAAATC ACAGCAGAAA CACGTGAGAC
     ·*  F  S   S  D  T  F   I  R  L   C  C  F   C  T  Y  S ·
      V  I  F   *  *  Y  F   Y  *  T   L  L  F   L  H  I  L ·
     · N  F  L   V  I  L   L  L  D  F   V  V  S   A  H  T
1151 GTAATTTTCT AGTGATACTT TTATTAGACT TTGTTGTTTC TGCACATACT
     · K  S  V   L  K  V   G  V  L  I   W  M  I   N  N  P
     ·*  I  C   F  E  G  R   S  A  Y   L  D  D   K  *  S  S ·
      L  N  L   F  *  R  *   E  C  L   F  G  *   *  I  I  L ·
1201 CTAAATCTGT TTTGAAGGTA GGAGTGCTTA TTTGGATGAT AAATAATCCT
      L  L  V   A  *  I  F   I  H  H   M  P  P   T  W  F  L ·
     · V  S  C   M  N  I   Y  T  S  H   A  S  Y   M  V  P
     ·C  *  L   H  E  Y  L   Y  I  T   C  L  L   H  G  S  W ·
1251 CTGTTAGTTG CATGAATATT TATACATCAC ATGCCTCCTA CATGGTTCCT
     ·G  L  H   S  G  Q  R   F  D  N   *  V  H   A  N  L  I ·
      G  I  T   Q  W  T  T   L  *  *   L  S  P   C  *  L  D ·
     · D  Y  T   V  D  N   A  L  I  I   E  S  M   L  T  *
1301 GGGATTACAC AGTGGACAAC GCTTTGATAA TTGAGTCCAT GCTAACTTGA
     · I  I  Y   Q  Y  S   I  Y  H  F   I  L  Y   F  N  *
     ·Y  N  I   S  V  F  H   I  S  F   Y  L  V   L  Q  L  R ·
      L  *  Y   I  S  I  P   Y  I  I   L  S  C   T  S  T  E ·
1351 TTATAATATA TCAGTATTCC ATATATCATT TTATCTTGTA CTTCAACTGA
      D  H  P   Y  F  L  Q   T  V  F   I  G  C   S  G  E  L ·
     · S  S  L   F  F  A   N  R  I  Y   W  L  L   W  R  I
     ·I  I  L   I  F  C  K   P  Y  L   L  V  A   L  E  N  *  ·
1401 GATCATCCTT ATTTTTTGCA AACCGTATTT ATTGGTTGCT CTGGAGAATT
     ·K  S  *   N  *  A  L   L  L  I   A  E  P   C  W  S  S ·
      E  V  L   K  L  S  T   S  P  D   C  R  A   M  L  V  F ·
     · S  L  E   T  K  H  F   S  *  L   Q  S  H   A  G  L
1451 GAAGTCTTGA AACTAAGCAC TTCTCCTGAT TGCAGAGCCA TGCTGCTCTT
     · E  S  M   H  C  C  S   D  R  S   E  Q  Q   T  H  Y
     ·*  I  Y   A  L  L  F   *  S  Q   *  T  A   D  S  L  L ·
      L  N  L   C  I  V  V   L  I  A   V  N  S   R  L  I  I ·
1501 CTGAATCTAT GCATTGTTGT TCTGATCGCA GTGAACAGCA GACTCATTAT
      *  E  F   N  E  G  L   L  L  S   F  F  F   H  F  P  H ·
     · R  I  *   *  R  F  I   T  F  F   L  F  S   F  S  S
     ·E  N  L   M  K  V  Y   Y  F  L   S  F  F   I  F  L  T ·
1551 TGAGAATTTA ATGAAGGTTT ATTACTTTCT TTCTTTTTTC ATTTTCCTCA
     ·L  H  L   Q  I  P  Q   S  I  S   F  *  N   T  S  G  L ·
      P  S  F   T  D  P  S   I  H  L   L  L  K   Y  I  W  S ·
     · F  I  Y   R  S  L  N   P  S  P   S  E  I   H  L  V
1601 CCTTCATTTA CAGATCCCTC AATCCATCTC CTTCTGAAAT ACATCTGGTC
     · L  P  A   H  L  S   S  V  N  L   T  H  S   V  F  Y
     ·S  S  C   A  F  V  *   C  K  S   D  T  F   C  V  L  F ·
      F  F  L   R  I  C  L   V  *  I   *  H  I   L  C  F  I ·
1651 TTCTTCCTGC GCATTTGTCT AGTGTAAATC TGACACATTC TGTGTTTTAT
      L  N  W   L  V  Q  Y   G  L  L   I  R  A   G  F  W  F ·
     · K  L  A   G  A  V   W  P  V  D   K  S  W   I  L  V
     ·*  I  G   W  C  S  M   A  C  *   *  E  L   D  F  G  L ·
1701 TTAAATTGGC TGGTGCAGTA TGGCCTGTTG ATAAGAGCTG CATTTTGGTT
     ·S  A  R   S  L  G  D   W  P  L   L  M  C   W  *  K  L ·
      *  C  K   I  A  G  *   L  A  P   S  N  V   L  V  E  I ·
     · V  Q  D   R  W  V   T  G  P  F   *  C  A   G  R  N
1751 TAGTGCAAGA TCGCTGGGTG ACTGGCCCCT TCTAATGTGC TGGTAGAAAT
```

```
      · L  S  F  L  I  Q  M  G  F  K  *  E  L  W  S  N
      ·V  V  I  F  N  S  D  G  F  Q  I  R  T  V  E  *  S ·
       C  C  H  F  *  F  R  W  V  S  N  K  N  C  G  V  I ·
1801  TGTTGTCATT TTTAATTCAG ATGGGTTTCA AATAAGAACT GTGGAGTAAT
       Q  S  V  N  F  S  L  T  L  P  V  F  P  L  V  A  L ·
      · I  C  Q  F  Q  P  H  S  T  S  F  P  T  S  C  T
      ·N  L  S  I  S  A  S  L  Y  Q  F  S  H  *  L  H  S ·
1851  CAATCTGTCA ATTTCAGCCT CACTCTACCA GTTTTCCCAC TAGTTGCACT
      ·M  A  E  K  L  I  T  R  K  L  I  G  E  H  V  S  L ·
       H  G  *  E  A  D  H  K  K  A  H  W  *  T  C  K  F ·
      · W  L  R  S  *  S  Q  E  S  S  L  V  N  M  *  V
1901  CATGGCTGAG AAGCTGATCA CAAGAAAGCT CATTGGTGAA CATGTAAGTT
      · T  H  K  I  A  *  Y  F  V  E  K  F  S  F  V  I
      ·D  S  Q  D  C  V  V  F  C  R  E  V  L  F  C  Y  F ·
       *  L  T  R  L  R  S  I  L  *  R  S  S  L  L  L  F ·
1951  TGACTCACAA GATTGCGTAG TATTTTGTAG AGAAGTTCTC TTTTGTTATT
       S  *  V  *  V  L  R  I  E  L  D  V  K  L  D  S  P ·
      · L  G  I  S  V  E  D  *  I  R  C  K  T  R  Q  S
      ·L  R  Y  K  C  *  G  L  N  *  M  *  N  *  T  V  L ·
2001  TCTTAGGTAT AAGTGTTGAG GATTGAATTA GATGTAAAAC TAGACAGTCC
      ·L  F  C  I  F  Q  V  P  F  I  V  Y  D  F  Y  T  P ·
       S  I  L  H  L  P  G  A  I  Y  R  L  *  L  L  Y  T ·
      · Y  S  A  S  S  R  C  H  L  S  F  M  T  S  I  H
2051  TCTATTCTGC ATCTTCCAGG TGCCATTTAT CGTTTATGAC TTCTATACAC
      · L  A  G  G  Y  S  T  P  Y  H  Y  Y  N  I  C  H
      ·S  C  R  W  L  F  Y  S  I  S  L  L  Q  H  L  P  L ·
       L  L  Q  V  V  I  L  L  H  I  I  I  T  T  S  A  I ·
2101  CTCTTGCAGG TGGTTATTCT ACTCCATATC ATTATTACAA CATCTGCCAT
       C  L  S  S  C  C  D  S  *  V  S  I  S  F  C  F  A ·
      · S  I  Q  L  L  *  L  L  S  K  H  F  F  L  L  C
      ·V  Y  P  V  V  V  T  L  K  *  A  F  L  S  A  L  Q ·
2151  TGTCTATCCA GTTGTTGTGA CTCTTAAGTA AGCATTTCTT TCTGCTTTGC
      ·V  C  L  D  A  S  Y  F  D  I  R  *  A  L  V  F  H ·
       S  L  F  G  C  I  L  F  *  H  S  L  S  S  S  I  S ·
      · F  V  W  M  H  L  I  L  T  F  V  E  L  *  Y  F
2201  AGTTTGTTTG GATGCATCTT ATTTTGACAT TCGTTGAGCT CTAGTATTTC
      · G  M  E  Y  I  Q  L  I  L  F  V  I  C  C  T  S
      ·W  Y  G  I  H  S  I  N  L  V  R  N  L  L  Y  F  M ·
       M  V  W  N  T  F  N  *  S  C  S  *  F  A  V  L  H ·
2251  ATGGTATGGA ATACATTCAA TTAATCTTGT TCGTAATTTG CTGTACTTCA
       W  Y  G  G  Q  L  H  Y  C  A  P  N  I  *  S  F  P ·
      · V  W  W  P  T  T  L  L  C  P  K  H  L  V  F  P
      ·G  M  V  A  N  Y  I  I  V  P  Q  T  F  S  L  S  L ·
2301  TGGTATGGTG GCCAACTACA TTATTGTGCC CCAAACATTT AGTCTTTCCC
      ·S  R  Y  V  L  Y  Y  A  N  W  V  D  K  K  V  A  T ·
       F  K  I  R  T  I  L  C  K  L  G  G  *  K  G  S  Y ·
      · Q  D  T  Y  Y  T  M  Q  I  G  W  I  K  R  *  L
2351  TTCAAGATAC GTACTATACT ATGCAAATTG GGTGGATAAA AAGGTAGCTA
      · *  H  F  Y  L  I  V  S  G  D  S  T  L  *  Y  K
      ·I  T  L  L  F  N  C  I  W  *  L  H  T  I  I  Q  R ·
       H  N  T  F  I  *  L  Y  L  V  T  P  H  Y  N  T  K ·
2401  CATAACACTT TTATTTAATT GTATCTGGTG ACTCCACACT ATAATACAAA
       E  T  Q  L  S  S  I  F  K  K  K  M  Y  L  V  I  K ·
      · N  A  T  L  Q  H  I  Q  E  K  N  V  S  G  D  K
      ·K  R  N  S  P  A  Y  S  R  K  K  C  I  W  *  *  K ·
2451  GAAACGCAAC TCTCCAGCAT ATTCAAGAAA AAAATGTATC TGGTGATAAA
      ·I  Y  C  K  C  S  F  I  S  S  R  R  N  P  Y  Y  L ·
       N  L  L  Q  M  F  I  Y  L  *  *  K  K  S  L  L  S ·
      · S  I  A  N  V  H  L  S  L  V  E  E  I  L  T  I
2501  AATCTATTGC AAATGTTCAT TTATCTCTAG TAGAAGAAAT CCTTACTATC
      · T  L  S  *  S  V  H  *  L  H  L  I  G  K  I  C
      ·Y  S  V  L  I  C  S  L  T  A  S  N  R  E  D  L  L ·
       L  L  C  L  D  L  F  T  D  C  I  *  *  G  R  F  V ·
2551  TTACTCTGTC TTGATCTGTT CACTGACTGC ATCTAATAGG GAAGATTTGT
       *  S  I  N  I  D  T  H  F  I  M  Q  I  F  C  F  F ·
      · V  H  Q  Y  *  Y  T  F  Y  Y  A  D  I  L  F  L
      ·S  P  S  I  L  I  H  I  L  L  C  R  Y  F  V  S  F ·
2601  TAGTCCATCA ATATTGATAC ACATTTTATT ATGCAGATAT TTTGTTTCTT
      ·H  V  A  S  S  L  *  P  L  S  *  H  E  A  D  L  S ·
       S  C  S  F  *  L  V  T  P  F  L  T  *  S  *  S  F ·
      · M  *  L  L  A  C  N  P  F  P  N  M  K  L  I  F
2651  TCATGTAGCT TCTAGCTTGT AACCCCTTTC CTAACATGAA GCTGATCTTT
```

```
      · I  V  Q  E  K  L  D  I  F  V  H  M  L  G  N  *
      ·H  C  T  R  K  I  G  Y  I  C  S  H  A  W  K  L  N ·
       P  L  Y  K  K  N  W  I  Y  L  F  T  C  L  E  I  E ·
2701  CCATTGTACA AGAAAAATTG GATATATTTG TTCACATGCT TGGAAATTGA
       I  N  K  L  *  Y  F  *  C  *  C  A  S  S  R  L  W ·
      · K  Q  T  V  V  F  L  M  L  M  C  K  *  *  T  L
      · *  T  N  C  S  I  S  D  V  D  V  Q  V  V  D  F  G ·
2751  ATAAACAAAC TGTAGTATTT CTGATGTTGA TGTGCAAGTA GTAGACTTTG
      · L  S  Q  L  L  S  L  K  K  S  H  *  E  Q  V  T  F ·
       V  E  S  I  V  I  S  Q  K  E  P  L  G  A  S  Y  L ·
      · *  V  N  C  Y  L  S  K  R  A  I  R  S  K  L  P
2801  GTTGAGTCAA TTGTTATCTC TCAAAAAGAG CCATTAGGAG CAAGTTACCT
      · S  L  I  I  F  S  V  R  L  Q  E  L  R  M  L  Y
      ·F  I  D  Y  I  F  C  E  T  A  R  V  K  N  V  V  W ·
       F  H  *  L  Y  F  L  *  D  C  K  S  *  E  C  C  M ·
2851  TTTCATTGAT TATATTTTCT GTGAGACTGC AAGAGTTAAG AATGTTGTAT
       G  *  C  L  M  L  F  S  L  S  L  L  *  L  P  R  N ·
      · L  M  P  Y  A  V  *  F  K  F  V  I  I  A  K  K
      ·V  D  A  L  C  C  L  V  *  V  C  Y  N  C  Q  E  M ·
2901  GGTTGATGCC TTATGCTGTT TAGTTTAAGT TTGTTATAAT TGCCAAGAAA
      ·V  T  *  K  D  I  V  P  C  I  N  Y  G  L  S  V  Q ·
       C  Y  L  K  R  Y  C  P  M  H  Q  L  W  I  I  S  S ·
      · L  L  E  K  I  L  S  H  A  S  I  M  D  Y  Q  F
2951  TGTTACTTGA AAAGATATTG TCCCATGCAT CAATTATGGA TTATCAGTTC
      · S  Y  S  E  K  F  Q  V  *  L  S  S  T  I  W  I
      ·V  I  F  R  K  I  S  G  V  T  Q  Q  Y  Y  L  D  L ·
       S  H  I  P  K  N  F  R  C  D  S  A  V  L  S  G  F ·
3001  AGTCATATTC CGAAAAATTT CAGGTGTGAC TCAGCAGTAC TATCTGGATT
       C  A  N  V  S  C  E  H  H  V  D  E  A  C  L  L  C ·
      · C  *  C  F  L  R  A  S  C  G  *  S  L  S  L  M
      ·V  L  M  F  L  A  S  I  M  W  M  K  L  V  S  Y  A ·
3051  TGTGCTAATG TTTCTTGCGA GCATCATGTG GATGAAGCTT GTCTCTTATG
      · T  Y  K  L  *  Y  K  G  I  V  Q  K  Y  *  E  G  N ·
       H  I  Q  I  M  I  *  G  Y  C  P  K  V  L  R  R  * ·
      · H  T  N  Y  D  I  R  V  L  S  K  S  T  E  K  V
3101  CACATACAAA TTATGATATA AGGGTATTCT CCAAAAGTAC TGAGAAAGGTA
      · A  L  T  C  *  S  E  S  V  Q  I  F  C  *  H  V
      ·C  I  D  M  L  I  *  I  S  S  N  I  L  L  T  C  C ·
       M  H  *  H  V  N  L  N  Q  F  K  Y  F  V  N  M  L ·
3151  ATGCATTGAC ATGTTAATCT GAATCAGTTC AAATATTTTG TTAACATGTT
       A  H  F  S  K  L  I  C  *  R  S  N  F  S  *  N  S ·
      · P  F  L  K  I  D  L  L  T  F  K  L  F  L  K  L
      ·P  I  S  Q  N  *  F  V  D  V  Q  T  F  L  K  T  P ·
3201  GCCCATTTCT CAAAATTGAT TTGTTGACGT TCAAACTTTT CTTAAAACTC
      ·F  W  W  P  N  F  S  E  A  R  I  S  P  T  C  L  N ·
       L  L  V  A  K  F  F  *  S  *  N  I  S  H  L  F  K ·
      · F  G  G  Q  I  F  L  K  L  E  Y  L  P  L  V  *
3251  CTTTTGGTGG CCAAATTTTT CTGAAGCTAG AATATCTCCC ACTTGTTTAA
      · F  F  S  S  F  I  S  *  M  S  Y  I  *  F  Q  F
      ·L  L  F  Q  F  H  F  M  N  V  L  Y  L  V  S  I  F ·
       T  S  F  P  V  S  F  H  E  C  L  I  S  S  F  N  F ·
3301  ACTTCTTTTC CAGTTTCATT TCATGAATGT CTTATATCTA GTTTCAATTT
       L  H  R  M  K  C  G  A  N  Q  Y  T  L  P  S  R  E ·
      · A  *  D  E  M  W  C  Q  S  I  Y  V  T  I  K  R
      ·C  I  G  *  N  V  V  P  I  N  I  R  Y  H  Q  E  S ·
3351  TTGCATAGGA TGAAATGTGG TGCCAATCAA TATACGTTAC CATCAAGAGA
      · *  K  N  C  S  *  L  L  I  Q  C  F  C  Y  M  G  * ·
       V  K  K  L  F  L  T  S  H  T  V  F  L  L  H  G  L ·
      · K  K  I  V  L  N  F  S  Y  S  V  F  V  T  W  A
3401  GTAAAAAAAT TGTTCTTAAC TTCTCATACA GTGTTTTTGT TACATGGGCT
      · S  Y  I  L  S  C  V  S  L  T  V  S  V  Y  L  Y
      ·I  I  Y  T  L  M  C  *  L  N  C  *  C  I  P  L  L ·
       D  H  I  Y  S  H  V  L  A  *  L  L  V  Y  T  S  I ·
3451  GATCATATAT ACTCTCATGT GTTAGCTTAA CTGTTAGTGT ATACCTCTAT
       C  N  G  P  W  S  T  *  P  C  Y  I  N  A  F  P  T ·
      · *  W  A  L  V  H  L  T  L  L  Y  Q  C  I  P  N
      ·V  M  G  L  G  P  P  N  P  V  I  S  M  H  S  Q  P ·
3501  TGTAATGGGC CTTGGTCCAC CTAACCCTGT TATATCAATG CATTCCCAAC
      ·L  I  R  V  R  V  S  L  I  L  T  S  G  N  G  S  I ·
       P  N  *  G  *  G  F  P  H  S  N  F  R  Q  R  *  H ·
      · *  L  G  L  G  F  P  S  F  *  L  Q  A  T  V  A
3551  CCTAATTAGG GTTAGGGTTT CCCTCATTCT AACTTCAGGC AACGGTAGCA
```

```
        · H  Q  L  Y  S  H  V  L  K  C  D  G  I  L  G  L
        ·T  S  I  I  F  S  C  T  *  M  *  W  Y  F  G  A  L ·
         D  I  N  Y  I  L  M  Y  L  N  V  M  V  F  W  G  F ·
4501    GACATCAATT ATATTCTCAT GTACTTAAAT GTGATGGTAT TTTGGGGCTT
         Y  L  S  T  G  G  C  G  T  W  *  S  F  C  Y  F  Y ·
        · P  Q  Y  W  R  M  W  N  M  V  I  F  L  L  L  L
        ·T  S  V  L  E  D  V  E  H  G  N  L  F  V  T  S  I ·
4551    TACCTCAGTA CTGGAGGATC TGGAACATGG TAATCTTTTT GTTACTTCTA
        ·I  Q  I  L  Y  P  F  I  *  L  R  L  C  Y  L  T  K ·
         Y  S  D  S  I  P  F  Y  L  V  E  T  L  L  L  N  * ·
        · F  R  F  Y  T  L  L  F  S  *  D  F  V  T  *  L
4601    TATTCAGATT CTATACCCTT TTATTTAGTT GAGACTTTGT TACTTAACTA
        · D  S  C  D  G  S  G  T  L  L  F  S  *  D  F  L
        ·G  Q  L  *  W  *  W  Y  S  S  I  *  L  R  L  P  * ·
         R  T  V  V  M  V  V  V  L  F  Y  L  V  K  T  S  L ·
4651    AGGACAGTTG TGATGGTAGT GGTACTCTTC TATTTAGTTA AGACTTCCTT
         N  F  C  H  *  A  *  D  I  C  L  I  I  S  F  K  * ·
        · L  L  S  L  S  L  R  Y  L  S  N  N  I  F  Q  I
        ·T  S  V  T  E  L  E  I  F  V  *  *  Y  L  S  N  N ·
4701    AACTTCTGTC ACTGAGCTTG AGATATTTGT CTAATAATAT CTTTCAAATA
        ·L  T  I  S  L  F  F  V  S  L  F  I  S  G  S  S  D ·
         T  D  N  *  S  I  F  C  Q  P  V  H  K  W  I  I  R ·
        · *  Q  L  V  Y  F  L  S  A  C  S  *  V  D  H  Q
4751    ACTGACAATT AGTCTATTTT TTGTCAGCCT GTTCATAAGT GGATCATCAG
        · T  Y  I  F  H  V  *  G  K  A  F  P  G  *  L  L
        ·H  I  Y  F  P  C  I  R  K  G  F  S  R  V  I  A  S ·
         T  H  I  F  S  M  Y  K  E  R  L  F  Q  G  N  C  F ·
4801    ACACATATAT TTTCCATGTA TAAGCAAAGG CTTTTCCAGG GTAATTGCTT
         L  Y  V  Y  K  T  L  H  L  F  F  A  F  E  F  S  K ·
        · I  C  V  Q  N  S  T  F  V  L  C  F  *  I  L  Q
        ·Y  M  C  T  K  L  Y  I  C  S  L  L  L  N  S  P  N ·
4851    CTATATGTGT ACAAAACTCT ACATTTGTTC TTTGCTTTTG AATTCTCCAA
        ·C  S  L  V  W  N  I  D  A  I  *  N  S  Q  Y  T  N ·
         M  Q  F  S  L  E  H  R  C  N  I  E  F  T  I  Y  K ·
        · A  V  *  F  G  T  S  M  Q  Y  R  I  H  N  I  Q
4901    ATGCAGTTTA GTTTGGAACA TCGATGCAAT ATAGAATTCA CAATATACAA
        · D  V  L  *  K  M  G  K  Q  S  W  T  E  C  *  H
        ·*  C  S  L  E  N  G  E  A  E  L  D  R  V  L  A  L ·
         M  M  F  F  R  K  W  G  S  R  A  G  Q  S  V  S  T ·
4951    ATGATGTTCT TTAGAAAATG GGGAAGCAGA GCTGGACAGA GTGTTAGCAC
         S  I  V  N  L  S  *  *  *  *  I  Q  L  N  K  W  L ·
        · N  C  Q  F  V  I  I  I  M  N  T  T  E  Q  V  A
        ·Q  L  S  I  C  H  N  N  N  E  Y  N  *  T  S  G  * ·
5001    TCAATTGTCA ATTTGTCATA ATAATAATGA ATACAACTGA ACAAGTGGCT
        ·K  L  L  *  E  N  Q  N  T  S  G  Q  Y  Y  L  H  S ·
         E  T  V  V  R  K  S  E  H  *  W  S  I  L  F  A  * ·
        · N  C  C  E  K  I  R  T  L  V  V  N  I  I  C  I
5051    GAAACTGTTG TGAGAAAATC AGAACACTAG TGGTCAATAT TATTTGCATA
        · K  S  I  W  *  C  K  L  R  Y  E  V  L  T  S  Y
        ·*  I  N  L  V  M  *  I  K  I  *  S  S  Y  F  L  Y ·
         V  N  Q  F  G  N  V  N  *  D  M  K  F  L  L  L  I ·
5101    GTAAATCAAT TTGGTAATGT AAATTAAGAT ATGAAGTTCT TACTTCTTAT
         I  K  I  Y  Y  A  *  I  L  *  W  L  K  L  Y  C  S ·
        · K  D  L  L  C  L  N  F  I  V  A  E  T  L  L  F
        ·*  R  F  T  M  L  E  F  Y  S  G  *  N  F  T  V  L ·
5151    ATAAAGATTT ACTATGCTTG AATTTTATAG TGGCTGAAAC TTTACTGTTC
        ·W  I  K  I  L  N  K  N  K  G  Y  L  D  L  A  T  K ·
         L  D  K  D  F  K  *  K  Q  R  I  S  R  L  G  N  K ·
        · G  *  R  F  *  I  K  T  K  D  I  *  T  W  Q  Q
5201    TTGGATAAAG ATTTTAAATA AAAACAAAGG ATATCTAGAC TTGGCAACAA
        · C  C  L  L  L  T  G  K  S  K  L  D  N  V  N  T
        ·M  L  P  S  A  D  W  Q  K  *  I  R  Q  C  E  Y  M ·
         N  A  A  F  C  *  L  A  K  V  N  *  T  M  *  I  H ·
5251    AATGCTGCCT TCTGCTGACT GGCAAAAGTA AATTAGACAA TGTGAATACA
         W  T  Y  I  K  F  C  W  S  F  H  F  C  R  T  D  M ·
        · D  I  H  K  I  L  L  V  L  S  F  L  Q  N  *  H
        ·G  H  T  *  N  F  V  G  P  F  I  F  A  E  L  T  * ·
5301    TGGACATACA TAAAATTTTG TTGGTCCTTT CATTTTTGCA GAACTGACAT
        ·I  F  T  A  Y  F  S  N  S  Y  C  I  Y  T  A  G  C ·
         D  F  H  C  L  L  L  K  F  V  L  Y  L  H  C  R  V ·
        · F  S  L  P  T  S  Q  I  R  I  V  S  T  L  Q  G
5351    GATTTTCACT GCCTACTTCT CAAATTCGTA TTGTATCTAC ACTGCAGGGT
```


· S Y S N L V S G F S C I P * G T
· * L F * S R F W F Q L Y S M R Y F ·
V A I L I S F L V S A V F H E V L ·
5401 GTAGCTATTC TAATCTCGTT TCTGGTTTCA GCTGTATTCC ATGAGGTACT
L S S S E A F F M I G S I S V F P ·
· K F F R S L F H D R F N F C F S
· * V L Q K P F S * S V Q F L F F L ·
5451 TTAAGTTCTT CAGAAGCCTT TTTCATGATC GGTTCAATTT CTGTTTTTCC
· K T C Y C S N S T Q H I T N N T F ·
* D M L L F E F H S A H Y * Q Y V ·
· R H A I V R I P L S T L L T I R
5501 TAAGACATGC TATTGTTCGA ATTCCACTCA GCACATTACT AACAATACGT
· D L T Y Q Y I I T T S L F T L *
· * P Y V P I Y H H H I S F Y I V N ·
L T L R T N I S S P H L F L H C E ·
5551 TTGACCTTAC GTACCAATAT ATCATCACCA CATCTCTTTT TACATTGTGA
I H R Y V L R C R A T F S N S G H ·
· S Q I C I A V P C H I F K F W A
· F T D M Y C G A V P H F Q I L G I ·
5601 ATTCACAGAT ATCTATTGCG GTGCCGTGCC ACATTTTCAA ATTCTGGCCA
· F L G S C F R Y R N N T N I * L L ·
F S G I M F Q V * K * H * Y I T T ·
· F W D H V S G I E I T L I Y N Y
5651 TTTTCTGGGA TCATGTTTCA GGTATAGAAA TAACACTAAT ATATAACTAC
· P P F R I I S L S G L A F L V T
· T S I P N Y K S F W L G F S S Y I ·
Y L H S E L * V F L A W L F * L H ·
5701 TACCTCCATT CCGAATTATA AGTCTTTCTG GCTTGGCTTT TCTAGTTACA
L Y * V Y I * I I I V I Y L D I V ·
· I L G I Y L D Y N S Y I S R H C
· Y T R Y I S R L * * L Y I * T L C ·
5751 TTATACTAGG TATATATCTA GATTATAATA GTTATATATC TAGACATTGT
· Y I * M H T K C Y L S R K * D H G ·
V Y L D A Y Q M L P I * K I G S W ·
· I S R C I P N V T Y L E N R I M
5801 GTATATCTAG ATGCATACCA AATGTTACCT ATCTAGAAAA TAGGATCATG
· F R Y R S S N N I I T T T S I S
· F Q V * K * * * Y N N Y Y L H F E ·
V S G I E V V I I * * L L P P F R ·
5851 GTTTCAGGTA TAGAAGTAGT AATAATATAA TAACTACTAC CTCCATTTCG
N C K S L * L G F Y R * C * E L Y ·
· L * V I M T W L L * I M L R V I
· T V S H Y D L A F I D N A K S Y I ·
5901 AACTGTAAGT CATTATGACT TGGCTTTTAT AGATAATGCT AAGAGTTATA
· I W T L S R C V A T N L G K L E R ·
Y L D I I * M R S Y E S R K T R T ·
· S G H Y L D A * L R I * E N * N
5951 TATCTGGACA TTATCTAGAT GCGTAGCTAC GAATCTAGGA AAACTAGAAC
· L V I I P A F S F E S I S V Y S
· T C N Y P C L F F * V H Q C L F S ·
D L * L S L P F L L S P S V S I L ·
6001 GACTTGTAAT TATCCCTGCC TTTTCTTTTG AGTCCATCAG TGTCTATTCT
L T F * F H H Y I H K N N T T S W ·
· Y V L I P S L H P * E Q Y Y I L
· L R F D S I I T S I R T I L H L G ·
6051 CTTACGTTTT GATTCCATCA TTACATCCAT AAGAACAATA CTACATCTTG
· I Q C T F H C F H I G * H W L M S ·
D T M Y L P L F S H R L T L V D V ·
· Y N V P S T V F T * A D T G * C
6101 GATACAATGT ACCTTCCACT GTTTTCACAT AGGCTGACAC TGGTTGATGT
· D S Q I P L V F L T R Y L H A T
· * L T D T V G I L D K I S P C Y V ·
L T H R Y R W Y S * Q D I S M L R ·
6151 CTGACTCACA GATACCGTTG GTATTCTTGA CAAGATATCT CCATGCTACG
F K H V M V R C V N Y V L F F P L ·
· Q A C N G T L C Q L C P F F P I
· S S M * W Y A V S I M S F F S H Y ·
6201 TTCAAGCATG TAATGGTACG CTGTGTCAAT TATGTCCTTT TTTTCCCATT
· P L A T T * P S S S Y L A G G Q H ·
T S C H Y L T I I F L F G R W A T ·
· L L P L P N H H L L I W Q V G N
6251 ACCTCTTGCC ACTACCTAAC CATCATCTTC TTATTTGGCA GGTGGCAAC

```
        · D  I  L  V  L  Q  Y  S  R  T  A  D  V  C  P  S
        · *  Y  F  G  S  S  V  *  S  D  S  R  C  V  S  F  Y ·
        M  I  F  W  F  F  S  I  V  G  Q  P  M  C  V  L  L ·
6301    ATGATATTTT GGTTCTTCAG TATAGTCGGA CAGCCGATGT GTGTCCTTCT
        I  L  P  *  R  H  E  Q  A  G  P  G  K  *  I
        · T  T  M  T  S  *  T  G  R  P  R  Q  V  D
        · Y  Y  H  D  V  M  N  R  Q  A  Q  A  S  R  *
6351    ATACTACCAT GACGTCATGA ACAGGCAGGC CCAGGCAAGT AGATAG (SEQ ID NO:117)
```

# Figure 3

```
          S  S  K  A  R  S  P  R  S  S  N  G  G  L  R  R
          V  V  E  G  S  *  P  *  K  L  Q  W  R  T  P  K  T ·
           R  R  R  L  V  A  L  E  A  P  M  A  D  S  E  D ·
1        TCGTCGAA GGCTCGTAGC CCTAGAAGCT CCAATGGCGG ACTCCGAAGA
       R  A  A  S  R  A  P  P  P  T  A  P  R  S  R  C  C ·
      ·  R  R  Q  P  C  T  A  A  H  R  A  P  L  A  V  L
      ·A  P  P  A  V  H  R  R  P  P  R  P  A  R  G  A  A ·
51    CGCGCCGCCA GCCCTGCACC GCCGCCCACC GCGCCCCGCT CGCGGTGCTG
      ·C  S  P  G  L  R  G  R  V  A  P  P  A  E  I  R  R ·
       L  Q  P  R  A  S  R  P  R  C  A  A  G  *  D  P  A ·
      ·  A  A  Q  G  F  A  A  A  L  R  R  R  L  R  S  G
101   CTGCAGCCCA GGGCTTCGCG GCCGCGTTGC GCCGCCGGCT GAGATCCGGC
      ·  C  G  G  G  T  R  Q  L  C  R  R  L  R  G  R  V
      ·L  R  W  R  H  A  P  A  L  P  Q  T  P  G  T  S  L ·
       A  A  V  A  A  R  A  S  F  A  A  D  S  G  D  E  S ·
151   GCTGCCGGTGG CGGCACGCGC CAGCTTTGCC GCAGACTCCG GGGACGAGTC
       W  P  R  R  R  A  L  F  V  S  P  P  R  Q  Q  R  G  R ·
      ·  A  P  A  S  P  L  R  L  A  A  A  T  T  A  G  A
      ·G  P  G  E  P  S  S  S  R  R  R  D  N  S  G  G  A ·
201   TGGCCCCCGC CAGCCCTCTT CGTCTCGCCC CCGCCACAAC AGCGGCGCCC
      ·L  V  R  R  R  R  P  G  R  G  R  G  L  L  R  V  H ·
       P  R  P  P  P  A  A  G  P  G  Q  G  T  S  P  R  S ·
      ·  S  S  A  A  G  G  R  A  G  A  G  D  F  S  A  F
251   CCTCGTCCGC CGCCGGCGGC CGGGCCGGGG CAGGGGACTT CTCCGCGTTC
      ·  L  P  R  R  G  A  C  P  P  E  S  Q  G  E  P  S
      ·P  S  A  P  R  R  L  S  T  G  K  P  R  R  A  L  * ·
       T  F  R  A  A  A  P  V  H  R  K  A  K  E  S  P  L ·
301   ACCTTCCGCG CCGCCGGCGC TGTCCACCGG AAAGCCAAGG AGAGCCCTCT
       E  L  R  R  H  L  Q  A  G  Q  S  R  D  L  F  D  C ·
      ·  A  P  T  P  S  S  S  R  S  E  Q  G  P  I  *  L
      ·S  S  D  A  I  F  K  Q  V  R  A  G  T  Y  L  I  A ·
351   CAGCTCCGAC GCCATCTTCA AGCAGGTCAG AGCAGGGACC TATTTGATTG
      ·R  A  R  A  T  C  S  S  R  C  Q  L  F  C  G  R  F ·
       P  R  P  R  D  L  F  V  E  M  S  A  I  L  R  P  I ·
      ·  A  P  A  R  P  V  R  R  D  V  S  Y  S  A  A  D
401   CCGCGCCCGC GCGACCTGTT CGTCGAGATG TCAGCTATTC TGCGGCCGAT
      ·  *  F  L  R  H  A  C  L  V  S  S  V  T  N  L  K
      ·L  I  S  P  P  C  L  F  G  F  I  S  N  E  F  E  T ·
       F  N  F  S  A  M  L  V  W  F  H  Q  *  R  I  *  N ·
451   TTTAATTTCT CCGCCATGCT TGTTTGGTTT CATCAGTAAC GAATTTGAAA
       L  M  C  N  R  L  I  V  A  G  E  F  S  R  C  I  Y ·
      ·  D  V  *  P  I  N  R  S  R  G  I  L  E  V  H  L
      ·*  C  V  T  D  *  S  *  Q  G  N  S  R  G  A  S  T ·
501   CTGATGTGTA ACCGATTAAT CGTAGCAGGG GAATTCTCGA GGTGCATCTA
      ·A  C  G  I  K  A  S  P  F  I  Q  R  I  W  E  G  N ·
       R  M  W  D  K  G  V  A  F  Y  S  E  N  L  G  G  E ·
      ·  H  V  G  *  R  R  R  L  L  F  R  E  S  G  R  G
551   CGCATGTGGG ATAAAGGCGT CGCCTTTTAT TCAGAGAATC TGGGAGGGGA
      ·  L  T  F  R  A  A  Y  A  C  E  K  L  E  R  S  K
      ·F  N  F  S  C  C  I  C  V  R  E  A  R  T  *  Q  T ·
       I  *  L  F  V  L  H  M  R  A  R  S  S  N  V  A  N ·
601   ATTTAACTTT TCGTGCTGCA TATGCGTGCG AGAAGCTCGA ACGTAGCAAA
       Q  R  A  A  H  S  S  F  C  F  I  S  T  P  T  V  W ·
      ·  A  C  C  P  Q  F  I  L  F  Y  *  H  T  N  R  L
      ·S  V  L  P  T  V  H  S  V  L  L  A  H  Q  P  F  G ·
651   CAGCGTGCTG CCCACAGTTC ATTCTGTTTT ATTAGCACAC CAACCGTTTG
      ·*  Q  L  L  S  G  M  I  C  L  D  *  C  G  *  K  E ·
       V  T  V  T  *  W  H  D  M  S  *  L  V  W  I  K  G ·
      ·  D  S  Y  L  V  A  *  Y  V  L  I  S  V  D  K  R
701   GTGACAGTTA CTTAGTGGCA TGATATGTCT TGATTAGTGT GGATAAAAGG
      ·  L  F  Y  *  L  G  A  V  S  S  N  H  V  I  S  F
      ·T  F  L  L  V  G  S  C  E  F  K  S  R  D  L  L  Y ·
       N  F  F  T  S  W  E  L  *  V  Q  I  T  *  S  P  L ·
751   AACTTTTTTA CTAGTTGGGA GCTGTGAGTT CAAATCACGT GATCTCCTTT
       I  G  I  V  R  D  L  P  F  Y  K  W  I  Q  Y  M  I ·
      ·  W  H  R  T  *  P  A  I  L  Q  M  D  P  V  H  D
      ·L  A  S  Y  V  T  C  H  F  T  N  G  S  S  T  *  * ·
801   ATTGGCATCG TACGTGACCT GCCATTTTAC AAATGGATCC AGTACATGAT
```

```
      · I  *  Y  V  C  F  S  S  H  G  K  V  I  V  R  S  L ·
       N  M  I  C  V  L  Q  L  T  W  *  G  H  S  E  V  T ·
      · Y  D  M  C  A  S  A  H  M  V  R  S  *  *  G  H
851   AATATGATAT GTGTGCTTCA GCTCACATGG TAAGGTCATA GTGAGGTCAC
      · R  N  *  E  N  F  P  V  S  C  V  F  Y  V  K  I
      · *  E  L  G  K  F  S  C  F  L  C  F  L  C  E  N  K ·
       L  G  T  R  K  I  F  L  F  P  V  F  F  M  *  K  * ·
901   TTAGGAACTA GGAAAATTTT CCTGTTTCCT GTGTTTTTTA TGTGAAAATA
       N  *  F  F  L  Y  D  S  *  A  I  H  V  A  M  L  F ·
      · L  I  L  P  V  R  F  L  S  N  S  C  G  Y  A  V
      · T  D  S  S  C  T  I  P  K  Q  F  M  W  L  C  C  F ·
951   AACTGATTCT TCCTGTACGA TTCCTAAGCA ATTCATGTGG CTATGCTGTT
      · S  L  K  I  L  L  H  Q  R  I  L  F  Y  H  E  I  E ·
       F  P  Q  N  F  A  S  S  E  N  I  V  L  P  *  D  * ·
      · P  S  K  F  C  F  I  R  E  Y  C  F  T  M  R  L
1001  TTCCCTCAAA ATTTTGCTTC ATCAGAGAAT ATTGTTTTAC CATGAGATTG
      · S  C  *  L  I  L  L  F  W  F  L  T  E  S  C  R
      · E  L  L  A  N  T  S  I  L  V  S  D  R  V  M  Q  A ·
       R  A  V  S  *  Y  F  Y  F  G  F  *  Q  S  H  A  G ·
1051  AGAGCTGTTA GCTAATACTT CTATTTTGGT TTCTGACAGA CTCATGCACG
       P  F  Q  P  M  Y  C  C  S  G  C  G  E  *  Q  A  H ·
      · F  S  T  Y  V  L  L  F  W  L  R  *  I  A  G  S
      · L  F  N  L  C  I  V  V  L  V  A  V  N  S  R  L  I ·
1101  CCTTTTCAAC CTATGTATTG TTGTTCTGGT TGCGGTGAAT AGCAGGCTCA
      · Y  *  E  P  D  E  G  S  C  H  T  F  S  I  E  Q  I ·
       L  L  R  T  *  *  R  F  L  S  Y  F  F  D  *  A  N ·
      · I  E  N  L  M  K  V  L  V  I  L  F  R  L  S  K
1151  TTATTCACAA CCTGATCAAG GTTCTTGTCA TACTTTTTCG ATTGAGCAAA
      · D  A  I  V  L  V  P  F  T  Q  H  I  *  A  V  L
      · *  C  H  C  S  C  A  F  H  S  T  Y  L  S  C  S  F ·
       L  M  P  L  F  L  C  L  S  L  N  I  F  E  L  F  F ·
1201  TTGATGCCAT TGTTCTTGTG CCTTTCACTC AACATATTTG AGCTGTTCTT
       F  *  C  L  L  Q  Y  G  L  L  I  R  S  G  F  W  F ·
      · L  M  P  F  A  V  W  L  I  N  K  I  W  L  L  V
      · S  D  A  F  C  S  M  A  Y  *  *  D  L  A  F  G  L ·
1251  TTCTGATGCC TTTTGCAGTA TGGCTTATTA ATAAGATCTG GCTTTTGCTT
      · N  A  T  S  L  R  D  W  P  L  L  M  C  W  *  *  S ·
       *  C  Y  I  I  A  R  L  A  T  A  N  V  L  V  I  V ·
      · M  L  H  H  C  E  T  G  H  C  *  C  V  G  N  S
1301  TAATGCTACA TCATTGCGAC ACTGGCCACT GCTAATGTGT TGGTAATAGT
      · I  L  L  Y  H  P  L  L  Q  W  L  L  N  S  E  T
      · N  T  A  L  S  S  T  S  S  M  A  S  *  L  *  N  L ·
       Q  Y  C  S  I  I  H  F  F  N  G  F  L  T  L  K  L ·
1351  CAATACTGCT CTATCATCCA CTTCTTCAAT GGCTTCTTAA CTCTGAAACT
       *  S  A  N  L  Y  F  Q  P  *  S  T  H  I  S  P  W ·
      · I  S  *  F  I  F  S  A  L  V  Y  P  Y  F  P  L
      · N  Q  L  I  Y  I  F  S  L  S  L  P  I  F  P  L  G ·
1401  TAATCAGCTA ATTTATATTT TCAGCCTTAG TCTACCCATA TTTCCCCTTG
      · C  I  C  S  R  K  V  G  I  Q  Q  S  H  *  *  S  C ·
       V  H  L  Q  S  K  S  W  H  S  T  I  S  L  V  I  L ·
      · A  F  A  V  E  K  L  A  F  N  N  L  I  S  D  P
1451  GTGCATTTGC AGTCGAAAAG TTGGCATTCA ACAATCTCAT TAGTGATCCT
      · K  R  T  P  P  S  Y  K  A  Y  F  F  W  C  P  G
      · *  A  H  T  S  F  L  *  S  L  L  F  L  V  S  R  F ·
       V  S  A  H  L  L  P  I  K  L  T  F  S  G  V  Q  V ·
1501  GTAAGCGCAC ACCTCCTTCC TATAAAGCTT ACTTTTTCTG GTGTCCAGGT
       S  H  *  Q  F  M  H  F  P  Y  C  G  W  T  A  N  Y ·
      · S  L  T  V  Y  A  F  S  I  L  W  M  D  G  E  L
      · L  I  N  S  L  C  I  F  H  I  V  D  G  R  R  T  I ·
1551  TCTCATTAAC AGTTTATGCA TTTTCCATAT TGTGGATGGA CGGCGAACTA
      · *  F  H  L  S  K  Q  L  *  D  T  F  C  L  Y  C  F ·
       L  V  S  F  I  K  T  V  V  R  H  L  L  F  V  L  F ·
      · S  F  I  Y  Q  N  S  C  E  T  P  F  V  C  I  V
1601  TTAGTTTCAT TTATCAAAAC AGTTGTGAGA CACCTTTTGT TTGTATTGTT
      · S  K  F  C  L  *  T  I  D  L  W  F  F  P  F  S
      · L  Q  I  L  L  I  N  H  R  F  M  V  L  S  F  F  R ·
       S  P  N  F  A  Y  K  P  S  I  Y  G  S  F  L  F  Q ·
1651  TCTCCAAATT TTGCTTATAA ACCATCGATT TATGGTTCTT TCCTTTTTCA
       G  Y  Y  L  F  S  H  P  F  Y  N  I  *  N  C  I  S ·
      · L  L  P  V  F  T  S  F  L  Q  H  L  K  L  Y  I
      · A  T  T  C  F  H  I  L  F  T  T  F  E  I  V  Y  P ·
1701  GGCTACTACC TGTTTTCACA TCCTTTTTAC AACATTTGAA ATTGTATATC
```

```
      · S  A  R  D  S  *  V  S  I  C  N  N  L  *  I  Y  D ·
       Q  C  S  *  F  L  S  K  Y  M  *  *  P  L  N  I  * ·
      · V  L  V  I  L  K  *  V  Y  V  I  T  F  K  Y  M
1751  CAGTGCTCGT GATTCTTAAG TAAGTATATG TAATAACCTT TAAATATATG
      · *  L  C  T  S  L  C  V  K  I  L  P  C  C  L  S
      · L  I  M  H  F  P  L  R  Q  D  I  A  M  L  F  I  K ·
       I  D  Y  A  L  P  S  A  S  R  Y  C  H  V  V  Y  Q ·
1801  ATTGATTATG CACTTCCCTC TGCGTCAAGA TATTGCCATG TTGTTTATCA
       R  K  Q  I  *  E  *  N  T  C  A  S  M  L  F  S  S ·
      · K  A  N  L  R  I  K  H  M  C  Q  Y  A  V  *  F
      · E  S  K  F  E  N  K  T  H  V  P  V  C  C  L  V  H ·
1851  AGAAAGCAAA TTTGAGAATA AAACACATGT GCCAGTATGC TGTTTAGTTC
      · Y  E  S  T  K  I  F  F  E  C  *  I  A  F  V  T  F ·
       I  *  I  Y  Q  N  F  L  R  M  L  N  R  I  C  N  F ·
      · M  N  L  P  K  F  S  S  N  V  K  S  H  L  *  L
1901  ATATGAATCT ACCAAAATTT TCTTCGAATG TTAAATCGCA TTTGTAACTT
      · G  *  Y  R  F  R  T  N  I  T  P  F  C  L  V  L
      · W  I  I  P  F  S  Y  K  Y  H  A  I  L  P  C  V  N ·
       L  D  N  T  V  F  V  Q  I  S  R  H  F  A  L  C  * ·
1951  TTGGATAATA CCGTTTTCGT ACAAATATCA CGCCATTTTG CCTTGTGTTA
       T  P  S  S  V  T  G  F  R  C  D  S  A  V  L  S  G ·
      · T  K  F  C  N  W  V  Q  V  *  F  C  S  F  I  R
      · H  Q  V  L  *  L  G  S  G  V  I  L  Q  F  Y  Q  A ·
2001  ACACCAAGTT CTGTAACTGG GTTCAGGTGT GATTCTGCAG TTTTATCAGG
      · F  V  L  M  F  I  A  C  I  V  W  L  K  L  V  S  F ·
       L  C  V  D  V  Y  C  L  H  C  L  A  E  A  C  I  F ·
      · L  C  *  C  L  L  P  A  L  F  G  *  S  L  Y  L
2051  CTTTGTGTTC ATGTTTATTG CCTGCATTGT TTGGCTGAAG CTTGTATCTT
      · A  H  T  N  H  D  I  R  K  L  I  T  S  G  K  K
      · C  T  Y  K  P  *  Y  K  K  T  D  H  K  R  Q  E  G ·
       L  H  I  Q  T  M  I  *  E  N  *  S  Q  A  A  R  R ·
2101  TTGCACATAC AAACCATGAT ATAAGAAAAC TGATCACAAG CGGCAAGAAG
       V  H  C  V  L  Y  F  F  L  L  R  L  F  Y  I  C  F ·
      · T  L  C  A  I  F  F  P  S  P  F  V  L  Y  L  L
      · Y  I  V  C  Y  I  F  S  F  S  V  C  F  I  F  A  L ·
2151  GTACATTGTG TGCTATATTT TTTCCTTCTC CGTTTGTTTT ATATTTGCTT
      · K  A  L  I  P  I  M  *  L  V  D  Q  K  E  A  N  I ·
       *  S  S  Y  S  N  Y  V  I  S  *  S  E  G  S  K  Y ·
      · K  L  L  F  Q  L  C  N  *  L  I  R  R  K  Q  I
2201  TAAAGCTCTT ATTCCAATTA TGTAATTAGT TGATCAGAAG GAAGCAAATA
      · Q  S  S  F  F  F  C  Y  I  N  T  S  E  M  P  S
      · S  V  L  V  F  F  L  L  Y  K  Y  I  *  N  A  I  Y ·
       F  S  P  R  F  F  S  V  I  *  I  H  L  K  C  H  L ·
2251  TTCAGTCCTC GTTTTTTTTC TGTTATATAA ATACATCTGA AATGCCATCT
       M  G  I  R  I  L  L  I  C  L  N  R  S  Q  I  F  V ·
      · G  D  *  N  P  F  D  L  F  E  *  V  T  D  F  R
      · W  G  L  E  S  F  *  F  V  *  I  G  H  R  F  S  F ·
2301  ATGGGGATTA GAATCCTTTT GATTTGTTTG AATAGGTCAC AGATTTTCGT
      · C  K  S  P  Y  G  R  R  H  Q  H  L  S  F  W  L  S ·
       L  *  E  S  L  W  K  K  A  S  A  F  I  L  L  V  I ·
      · V  R  V  L  M  E  E  G  I  S  I  Y  P  F  G  Y
2351  TTGTAAGAGT CCTTATGGAA GAAGGCATCA GCATTTATCC TTTTGGTTAT
      · R  L  L  S  I  *  F  D  *  I  E  Q  A  L  R  Y
      · P  F  V  V  H  I  I  *  L  N  *  T  S  T  *  V  L ·
       P  V  C  C  P  Y  N  L  I  E  L  N  K  H  L  G  T ·
2401  CCCGTTTGTT GTCCATATAA TTTGATTGAA TTGAACAAGC ACTTAGGTAC
       L  G  P  R  A  C  L  E  A  P  S  F  *  E  T  G  L ·
      · R  A  *  G  L  F  G  S  T  *  F  L  R  N  W  F
      · *  G  L  G  P  V  W  K  H  L  V  F  K  K  L  V  Y ·
2451  TTAGGGCCTA GGGCCTGTTT GGAAGCACCT AGTTTTTAAG AAACTGGTTT
      · *  K  L  R  W  F  Q  T  Y  *  F  M  S  *  F  I  E ·
       M  K  T  E  V  V  P  N  I  L  V  Y  V  L  V  Y  R ·
      · E  N  *  G  G  S  K  H  T  S  L  C  P  S  L  *
2501  ATGAAAACTG AGGTGGTTCC AAACATACTA GTTTATGTCC TAGTTTATAG
      · T  G  F  S  I  S  *  K  P  R  S  *  P  P  L  A
      · N  W  I  L  N  F  L  K  T  K  K  L  A  S  P  S  * ·
       K  L  D  S  Q  F  L  K  N  Q  E  A  S  L  P  *  L ·
2551  AAACTGGATT CTCAATTTCT TAAAAACCAA GAAGCTAGCC TCCCCTAGCT
       K  A  S  L  *  K  Q  W  W  F  Q  T  T  L  T  S  F ·
      · S  Q  F  M  K  T  V  V  V  P  N  N  I  N  Q  F
      · K  P  V  Y  E  N  S  G  G  S  K  Q  H  *  P  V  F ·
2601  AAAGCCAGTT TATGAAAACA GTGGTGGTTC CAAACAACAT TAACCAGTTT
```

```
      · F  S  S  *  I  S  L  *  K  L  D  S  *  K  L  G  C ·
       F  F  L  I  N  Q  S  L  E  T  G  F  L  K  T  R  L ·
      · F  P  H  K  S  V  S  R  N  W  I  L  K  N  *  V
2651  TTTTTCCTCA TAAATCAGTC TCTAGAAACT GGATTCTTAA AAACTAGGTT
      · F  Q  T  G  P  *  Q  T  F  *  S  T  L  V  I  L
      · L  P  N  R  A  L  A  N  V  L  I  Y  S  S  D  S  S ·
       A  S  K  Q  G  L  S  K  R  F  N  L  L  *  *  F  F ·
2701  GCTTCCAAAC AGGGCCTTAG CAAACGTTTT AATCTACTCT AGTGATTCTT
       L  L  L  C  Q  N  S  F  G  V  V  T  L  N  A  A  K ·
      · L  T  L  S  E  F  F  R  S  C  D  T  *  C  S  K
      · S  Y  S  V  R  I  L  S  E  L  *  H  L  M  Q  Q  R ·
2751  CTCTTACTCT GTCAGAATTC TTTCGGAGTT GTGACACTTA ATGCAGCAAA
      · G  F  *  H  N  K  Y  L  H  L  T  L  I  C  C  W  T ·
       G  I  L  T  *  *  I  F  T  F  N  I  D  M  L  L  D ·
      · D  F  N  I  I  N  I  Y  I  *  H  *  Y  V  A  G
2801  GGGATTTTAA CATAATAAAT ATTTACATTT AACATTGATA TGTTGCTGGA
      · C  C  F  S  S  L  G  F  P  P  S  P  D  L  T  L
      · L  L  L  L  I  T  G  F  S  S  L  S  *  L  N  P  C ·
       P  A  A  S  H  H  W  V  F  L  P  L  L  T  *  P  L ·
2851  CCTGCTGCTT CTCATCACTG GGTTTTCCTC CCTCTCCTGA CTTAACCCTT
       V  L  Y  Y  F  *  R  C  S  *  *  S  S  L  L  W  P ·
      · T  L  L  L  L  E  M  F  I  V  V  Q  P  S  L  A
      · Y  S  I  T  F  R  D  V  H  S  S  P  A  F  S  G  P ·
2901  GTACTCTATT ACTTTTAGAG ATGTTCATAG TAGTCCAGCC TTCTCTGGCC
      · K  V  L  G  L  A  W  S  W  A  C  K  *  N  A  K  I ·
       Q  G  A  G  L  S  L  V  V  G  L  *  M  K  C  K  N ·
      · R  C  W  A  *  L  G  R  G  L  V  N  E  M  Q  K
2951  CAAGGTGCTG GGCTTAGCTT GGTCGTGGGC TTGTAAATGA AATGCAAAAA
      · C  R  Y  L  Y  M  L  C  L  R  L  L  L  L  K  M
      · M  Q  V  S  V  Y  V  M  L  K  T  I  V  T  K  N  V ·
       Y  A  G  I  C  I  C  Y  A  *  D  Y  C  Y  *  K  C ·
3001  TATGCAGGTA TCTGTATATG TTATGCTTAA GACTATTGTT ACTAAAAATG
       *  Y  I  L  F  V  L  I  F  L  K  K  N  I  *  K  Y ·
      · I  Y  P  I  R  V  N  I  F  K  E  K  Y  I  K  I
      · N  I  S  Y  S  C  *  Y  F  *  R  K  I  Y  K  N  I ·
3051  TAATATATCC TATTCGTGTT AATATTTTTA AAGAAAAATA TATAAAAATA
      · *  Y  I  G  L  Q  V  R  S  *  E  F  T  G  P  T  W ·
       L  I  Y  W  A  S  G  E  E  L  R  V  Y  G  S  N  L ·
      · N  I  L  G  F  R  *  G  V  E  S  L  R  V  Q  P
3101  TTAATATATT GGGCTTCAGG TGAGGAGTTG AGAGTTTACG GGTCCAACCT
      · S  G  A  *  S  P  Q  L  L  P  F  M  F  Y  M  Y
      · V  R  C  M  I  T  S  I  T  T  I  Y  V  L  Y  V  F ·
       G  P  V  H  D  H  L  N  Y  Y  H  L  C  F  I  C  I ·
3151  GGTCCGGTGC ATGATCACCT CAATTACTAC CATTTATGTT TTATATGTAT
       L  T  V  V  S  L  L  H  K  L  I  Y  R  L  Y  *  V ·
      · N  C  C  *  F  V  T  *  V  N  I  P  F  I  L  G
      · *  L  L  L  V  C  Y  I  S  *  Y  T  V  Y  I  R  L ·
3201  TTAACTGTTG TTAGTTTGTT ACATAAGTTA ATATACCGTT TATATTAGGT
      · D  N  E  L  T  A  A  G  I  D  N  L  Q  A  P  T  L ·
       *  *  *  T  D  R  G  W  H  R  *  F  T  S  S  N  S ·
      · I  M  N  *  P  R  L  A  *  I  I  Y  K  L  Q  L
3251  TGATAATGAA CTGACCGCGG CTGGCATAGA TAATTTACAA GCTCCAACTC
      · G  S  L  T  Y  F  M  M  A  P  T  L  C  Y  Q  V
      · W  E  S  N  I  L  H  D  G  S  D  T  L  L  S  G  N ·
       L  G  V  *  H  T  S  *  W  L  R  H  S  V  I  R  * ·
3301  TTGGGAGTCT AACATACTTC ATGATGGCTC CGACACTCTG TTATCAGGTA
       I  F  R  S  C  M  L  I  *  S  R  Y  M  I  F  F  I ·
      · F  P  I  M  Y  V  D  L  K  Q  I  Y  D  F  F  Y
      · F  S  D  H  V  C  *  S  K  A  D  I  *  F  F  L  L ·
3351  ATTTTCCGAT CATGTATGTT GATCTAAAGC AGATATATGA TTTTTTTTAT
      · T  F  R  L  A  I  S  D  *  Q  Q  K  V  C  W  Q  G ·
       Y  L  S  T  G  H  F  *  L  A  A  E  S  V  L  A  R ·
      · P  F  D  W  P  F  L  I  S  S  R  K  C  A  G  K
3401  TACCTTTCGA CTGGCCATTT CTGATTAGCA GCAGAAAGTG TGCTGGCAAG
      · E  T  F  *  K  R  T  T  F  V  L  H  A  A  K  L
      · *  N  I  L  E  T  Y  Y  F  C  F  A  C  S  Q  V  I ·
       V  K  H  F  R  N  V  L  L  L  F  C  M  Q  P  S  Y ·
3451  GTGAAACATT TTAGAAACGT ACTACTTTTG TTTTGCATGC AGCCAAGTTA
       S  S  N  T  L  C  *  K  R  L  A  G  P  S  S  Y  S ·
      · L  E  H  L  M  L  E  K  V  G  W  S  V  K  L  F
      · P  R  T  P  Y  V  R  K  G  W  L  V  R  Q  V  I  L ·
3501  TCCTCGAACA CCTTATGTTA GAAAAGGTTG GCTGGTCCGT CAAGTTATTC
```

· I L D I Y W S P R I H Y * A S K L ·
Y T * Y L L V S K D S L L S K * A ·
· Y L I F T G L Q G F I I E Q V S
3551 TATACTTGAT ATTTACTGGT CTCCAAGCAT TCATTATTGA GCAAGTAAGC
· F I C L H I Y I L * N I L I H F
· V Y M S P Y I Y L V K Y P H T F L ·
C L Y V S I Y I S C K I S S Y I S ·
3601 TGTTTATATG TCTCCATATA TATATCTTGT AAAATATCCT CATACATTTC
L T F Y * F Y N L A V H K S Y C C ·
· N L L L I L Q S C S T * I L L L
· * P F I D F T I L Q Y I N P I V V ·
3651 TTAACCTTTT ATTGATTTTA CAATCTTGCA GTACATAAAT CCTATTGTTG
· E L S T S I D G R I T E C C R D C ·
* T L N I H * W E D Y * M L * R L ·
· N S Q H P L M G G L L N A V E T
3701 TGAACTCTCA ACATCCATTG ATCGCAGGAT TACTGAATGC TGTACAGACT
· F E A L I T K C L P V A L H V L
· F * S S H Y Q M S T C G F A C F I ·
V L K L S L P N V Y L W L C M F Y ·
3751 CTTTTGAAGC TCTCATTACC AAATCTCTAC CTGTGCCTTT CCATCTTTTA
L P F P S V V V N Y I F Y L F I Q ·
· A F S I C G S K L Y F L S I H S
· C L F H L W * * I I F S I Y S F K ·
3801 TTGCCTTTTC CATCTGTGGT AGTAAATTAT ATTTTCTATC TATTCATTCA
· I Q I H D R S * Y Y S N I S H R C ·
N S N T * S I L I L F * Y Q S S M ·
· F K Y M I D P D I I L I S V I D
3851 AATTCAAATA CATGATCGAT CCTGATATTA TTCTAATATC AGTCATCGAT
· F F S E H A C S R R K L F P F M
· L L L R T C M L S E K I I S F Y A ·
A S S P N M H A L G E N Y F L L C ·
3901 GCTTCTTCTC CGAACATGCA TGCTCTCGGA GAAAATTATT TCCTTTTATG
L F Q I C I F A V S N L Y S S I Q ·
· L S N L H I C S F E L I F I Y T
· S F K F A Y L Q F R T Y I H L Y R ·
3951 CTCTTTCAAA TTTGCATATT TGCAGTTTCG AACTTATATT CATCTATACA
· V K H T C * D S S I W * P R I L Q ·
G * T Y L L R F F D L V T E N S T ·
· L N I L A E I L R F G D R E F Y
4001 GGTTAAACAT ACTTGCTGAC ATTCTTCGAT TTGGTGACCG AGAATTCTAC
· R L V E C K D N * * G E I L P C
· K T G G M Q R Q L M R * N T T V F ·
K D W W N A K T I D E V K Y Y R V ·
4051 AAAGACTGGT GCAATGCAAA CACAATTGAT GAGGTGAAAT ACTACCGTGT
L L Q P S I T S F F N * Q K E S R ·
· V A T * Y N F I F Q L T K G I *
· C C N L V * L H F S T N K R N L G ·
4101 TTGTTGCAAC CTAGTATAAC TTCATTTTTC AACTAACAAA AGGAATCTAG
· * A * L M L C M L F A V V A C V A ·
V S I A D V V Y V V C C C C L C C ·
· K H S * C C V C C L L L L L V L
4151 GTAAGCATAG CTGATGTTGT GTATGTTGTT TGCTGTTGTT GCTTGTGTTG
· V L G W D Q Y W R K W N M V C L
· C A W M G S V L E K M E H G M P F ·
L C L D G I S T G E N G T W Y A F ·
4201 CTGTGCTTGG ATGGGATCAG TACTGGAGAA AATCGAACAT GGTATGCCTT
F F L G N N N S S K N V L P D M V ·
· L S W Q * * F * Q E C T A R H G
· S F L A I I I L A R M Y C Q T W Y ·
4251 TTCTTTCTTG GCAATAATAA TTCTAGCAAG AATGTACTGC CAGACATGGT
· Y A V V V R C N H L K A T A * * K ·
I R S C G * M * P S * S N C I I K ·
· T Q L W L D V T I L K Q L H N K
4301 ATACGCAGTT GTGGTTAGAT GTAACCATCT TAAAGCAACT GCATAATAAA
· W Q V T V R L L R T M K S G L L
· M A S H C K I V E N Y E I W T L G ·
N G K S L * D C * E L * N L D S W ·
4351 AATGGCAAGT CACTGTAAGA TTGTTGAGAA CTATGAAATC TGGACTCTTG
V H R T I L I C M F Q V V * I T P ·
· T Q N Y S N L Y V P S S L N N T
· Y T E L F * F V C S K * F E * H P ·
4401 GTACACAGAA CTATTCTAAT TTGTATGTTC CAAGTAGTTT GAATAACACC

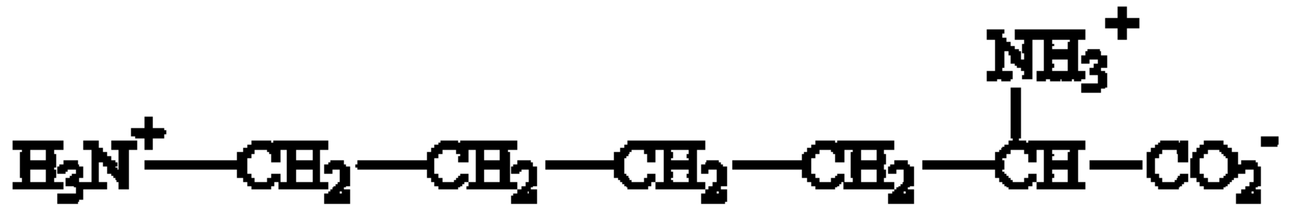
```
      · R  I  *  W  I  Y  S  S  P  M  I  F  V  G  P  Y  * ·
       Q  N  L  M  D  L  F  F  S  N  D  L  C  W  S  I  L ·
      · E  F  D  G  F  I  L  L  Q  *  S  L  L  V  H  I
4451  CAGAATTTGA TGGATTTATT CTTCTCCAAT GATCTTTGTT GGTCCATATT
      · F  H  L  L  K  Y  Y  D  *  *  F  A  H  F  D  S
      · I  S  F  A  E  I  L  *  L  I  I  C  T  F  *  Q  P ·
       N  F  I  C  *  N  I  M  T  N  N  L  H  I  L  T  A ·
4501  AATTTCATTT GCTGAAATAT TATGACTAAT AATTTGCACA TTTTGACAGG
       L  C  I  N  G  L  F  V  I  Y  I  S  L  A  C  E  M ·
      · V  H  K  W  I  V  R  H  I  Y  F  P  C  M  R  N
      · C  A  *  M  D  C  S  S  Y  I  F  P  L  H  A  K  W ·
4551  CTGTGCATAA ATGGATTGTT CGTCATATAT ATTTCCCTTC CATGCGAAAT
      · V  Y  Q  R  *  N  N  I  C  V  D  V  M  L  R  V  A ·
       G  I  S  K  V  K  *  Y  L  C  G  C  Y  A  Q  G  C ·
      · Y  I  K  G  K  I  I  F  V  W  M  L  C  S  G  L
4601  CGTATATCAA AGGTAAAATA ATATTTGTGT GGATGTTATG CTCAGGGTTG
      · S  S  A  Y  K  T  K  L  Y  L  H  Y  S  L  C  L
      · F  Q  C  I  *  D  *  T  V  P  S  L  Q  F  V  P  C ·
       L  P  V  H  I  R  L  N  C  T  F  T  T  V  C  A  L ·
4651  CTTCCAGTGC ATATAAGACT AAACTGTACC TTCACTACAG TTTGTGCCTT
       V  N  P  L  T  S  Q  C  L  L  V  I  S  N  F  S  K ·
      · K  S  L  N  Q  S  V  L  A  C  N  F  *  F  F  K
      · *  I  P  *  P  V  S  A  C  L  *  F  L  I  F  Q  S ·
4701  GTAAATCCCT TAACCAGTCA GTGCTTGCTT GTAATTTCTA ATTTTTCAAA
      · *  L  *  E  L  L  P  C  C  L  G  Q  S  N  *  F  T ·
       V  T  I  R  I  T  S  M  L  F  R  A  I  *  L  V  H ·
      · N  Y  K  N  Y  F  H  A  V  *  G  N  L  I  S  S
4751  GTAACTATAA GAATTACTTC CATGCTGTTT AGGGCAATCT AATTAGTTCA
      · S  E  T  K  V  H  V  L  I  K  I  C  K  F  N  S
      · Q  *  N  *  S  S  C  T  H  K  N  L  *  V  Q  Q  * ·
       P  V  K  L  K  F  M  Y  S  *  K  S  V  S  S  T  V ·
4801  CCAGTGAAAC TAAAGTTCAT GTACTCATAA AAATCTGTAA GTTCAACAGT
       K  S  I  L  L  S  L  L  S  N  K  N  V  F  N  F  V ·
      · K  Y  S  F  E  F  I  E  Q  *  E  C  V  *  F  C
      · K  V  F  F  *  V  Y  *  A  I  R  M  C  L  I  L  Y ·
4851  AAAAGTATTC TTTTGAGTTT ATTGAGCAAT AAGAATGTGT TTAATTTTGT
      · W  I  Y  C  P  P  V  *  L  L  Y  *  R  C  Y  I  L ·
       M  D  L  L  S  S  S  L  A  P  I  L  K  V  L  Y  F ·
      · G  S  T  V  L  Q  S  S  S  Y  T  E  G  A  I  F
4901  ATGGATCTAC TGTCCTCCAG TCTAGCTCCT ATACTGAAGG TGCTATATTT
      · L  S  S  L  L  C  C  Y  Y  Y  C  L  K  L  F  T
      · V  I  Q  S  S  V  L  L  L  L  L  F  K  V  I  H  L ·
       C  Y  P  V  F  C  A  V  I  I  T  V  *  S  Y  S  L ·
4951  TGTTATCCAG TCTTCTGTGC TGTTATTATT ACTGTTTAAA GTTATTCACT
       Y  L  C  *  F  V  T  G  S  C  C  F  Y  I  V  L  C ·
      · F  V  L  V  C  Y  R  K  L  L  F  L  Y  R  S  L
      · I  C  A  S  L  L  Q  E  V  A  V  F  I  S  F  F  V ·
5001  TATTTGTGCT AGTTTGTTAC AGCAAGTTGC TCTTTTTATA TCCTTCTTTG
      · F  C  C  T  S  *  G  N  L  F  T  F  S  L  F  I  C ·
       F  L  L  Y  F  M  R  *  L  I  Y  L  F  T  L  H  L ·
      · S  A  V  L  H  E  V  T  Y  L  P  F  H  S  S  S
5051  TTTCTGCTGT ACTTCATGAG GTAACTTATT TACCTTTTCA CTCTTCATCT
      · Y  I  N  Y  I  V  L  Y  F  Q  M  C  P  F  E  F
      · L  Y  *  L  Y  S  S  L  F  S  N  V  S  F  R  V  S ·
       A  I  L  I  I  *  F  S  I  F  K  C  V  L  S  S  F ·
5101  GCTATATTAA TTATATAGTT CTCTATTTTC AAATGTGTCC TTTCGAGTTT
       R  H  A  F  V  Q  T  Y  Q  L  Q  I  T  W  M  K  C ·
      · T  C  F  C  S  N  L  P  A  A  D  Y  L  D  E  V
      · D  M  L  L  F  K  L  T  S  C  R  L  L  G  *  S  A ·
5151  CGACATGCTT TTGTTCAAAC TTACCAGCTG CAGATTACTT GGATGAAGTG
      · S  I  *  N  *  I  F  H  N  P  V  P  F  E  K  I  M ·
       L  Y  I  K  L  N  I  S  *  S  S  P  F  R  E  N  Y ·
      · L  Y  K  I  K  Y  F  I  I  Q  S  L  S  R  K  L
5201  CTCTATATAA AATTAAATAT TTCATAATCC AGTCCCTTTC GAGAAAATTA
      · I  H  F  V  C  N  C  T  P  V  M  C  C  C  S  L
      · D  T  F  C  L  Q  L  Y  T  S  Y  V  L  L  F  P  A ·
       *  Y  I  L  F  A  I  V  H  Q  L  C  V  A  V  P  C ·
5251  TGATACATTT TGTTTGCAAT TGTACACCAG TTATCTCTTG CTGTTCCCTG
       P  H  T  Q  V  L  G  F  L  R  N  H  A  S  G  K  Y ·
      · T  Y  S  S  S  G  L  S  *  E  S  C  F  R  *  I
      · H  I  L  K  F  W  A  F  L  G  I  M  L  Q  V  N  I ·
5301  CCACATACTC AAGTTCTGGG CTTTCTTAGG AATCATGCTT CAGGTAAATA
```

```
      · N  T  T  P  D  A  V  D  Q  Y  P  T  F  M  F  F  F ·
       *  Y  N  A  R  C  C  G  S  I  S  N  I  Y  V  F  L ·
      · I  Q  R  Q  M  L  W  I  N  I  Q  H  L  C  F  S
5351  TAATACAACG CCAGATGCTG TGGATCAATA TCCAACATTT ATGTTTTTCT
      · T  H  F  F  L  Y  *  F  F  F  H  S  L  *  S  E
      ·Y  S  F  F  P  L  L  V  L  F  S  F  A  I  V  *  G ·
       L  L  I  F  S  F  I  S  S  F  F  I  R  Y  S  L  R ·
5401  TTACTCATTT TTTCCTTTAT TAGTTCTTTT TTCATTCGCT ATAGTCTGAG
       V  L  *  A  P  E  G  N  S  H  L  I  W  F  H  V  F ·
      · L  M  S  T  R  R  Q  Q  S  S  N  M  V  P  C  I
      ·S  Y  E  H  Q  K  A  T  V  I  *  Y  G  S  M  Y  L ·
5451  GTCTTATGAG CACCAGAAGG CAACAGTCAT CTAATATGGT TCCATGTATT
      · M  L  S  P  K  Q  K  T  V  L  T  P  Y  F  F  S  S ·
       Y  V  V  S  *  T  E  N  C  S  D  T  I  F  L  F  L ·
      · C  C  L  L  N  R  K  L  F  *  H  H  I  S  F  P
5501  TATGTTGTCT CCTAAACAGA AAACTGTTCT GACACCATAT TTCTTTTCCT
      · F  Y  Q  I  P  L  I  I  L  T  S  Y  L  K  N  K
      ·F  L  P  D  S  P  H  H  I  D  I  I  P  Q  K  *  I ·
       L  F  T  R  F  P  S  S  Y  *  H  H  T  S  K  I  N ·
5551  CTTTTTACCA GATTGCCCTC ATCATATTGA CATCATACCT CAAAAATAAA
       F  S  D  T  M  V  S  H  I  F  L  S  *  Y  L  L  G ·
      · Q  *  H  N  G  E  P  Y  F  S  *  L  I  L  A  R
      · S  V  T  Q  W  *  A  I  F  F  L  V  D  T  C  *  D ·
5601  TTCAGTGACA CAATGGTGAG CCATATTTTT CTTAGTTGAT ACTTGCTAGG
      · S  F  F  M  G  T  C  L  I  C  G  I  F  N  R  L  A ·
       I  V  F  Y  G  N  L  S  N  L  W  Y  I  Q  *  V  G ·
      · R  F  L  W  E  P  V  *  F  V  V  Y  S  I  G  W
5651  ATCGTTTTTT ATGGGAACCT GTCTAATTTG TGGTATATTC AATAGCTTGG
      · I  *  S  F  G  F  F  S  A  Y  T  G  S  Q  C  V
      ·N  M  I  F  W  F  F  F  C  I  Y  G  Q  P  M  C  V ·
       Q  Y  D  L  L  V  F  F  L  H  I  R  A  A  N  V  C ·
5701  CAATATGATC TTTTGGTTTT TTTTCTGCAT ATACGGGCAG CCAATCTGTG
       F  Y  C  I  T  M  M  *  *  T  G  L  R  R  Q  N  N ·
      · L  L  Y  Y  H  D  V  M  N  R  T  E  K  A  K  *
      · S  I  V  L  P  *  C  D  E  P  D  *  E  G  K  I  T ·
5751  TTCTATTGTA TTACCATGAT GTGATGAACC GGACTGAGAA GGCAAAATAA
      · H  L  *  I  F  F  G  V  S  F  L  P  S  W  K  L  K ·
       P  S  V  D  L  F  W  C  F  I  S  S  I  M  E  T  E ·
      · I  C  R  S  F  L  V  F  H  F  F  H  H  G  N  *
5801  CCATCTGTAG ATCTTTTTTG GTGTTTCATT TCTTCCATCA TGGAAACTGA
      · A  I  I  C  A  H  S  K  P  A  S  C  L  P  V  F
      · S  N  N  L  C  T  Q  *  T  S  I  V  S  S  S  F
       K  Q  *  S  V  H  T  V  N  Q  H  R  V  F  Q  F
5851  AAGCAATAAT CTGTGCACAC AGTAAACCAG CATCGTGTCT TCCAGTTTTT (SEQ ID NO:118)
```

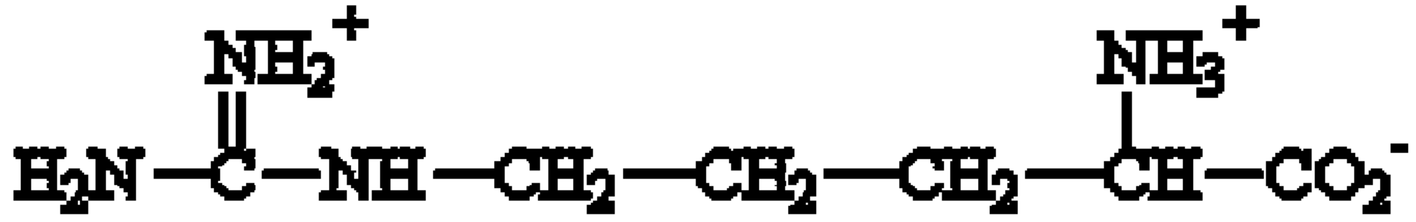
## Figure 4

AtDGAT1- MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRS----DSSNGLLLSGSDNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGR--GNADATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMK
BjDGAT1- MAILDSGGVAVPPTENG--VADLDRLHRRKSRS----DSSNGLLS----DTSPSDDVGAAAAERDRVDSAAEEEAQGTANLA-----GGDAETRESAGG--------DVRFTYRPSVPAHRRTRESPLSSDAIFKQSHAGLFNLCVVVLVAVNSRPIIENLMK
BnDGAT1-AF MAILDSGGVAVPPTENG--VADLDRLHRRKSSS----DSSNGLLS----DTSPSDDVGAAAAERDRVDSAAEEEAQGTANLA-----GGDAETRESAGG--------DVRFTYRPSVPAHRRTRESPLSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMK
BjDGAT1- MAILDSG-TVTMATENG--VADLDMLRRRKSRS----DSSNGLLS----ETSPSDDAGAPADVEDRVDS-A---AQGTANLA------GDTETRESGGGGGG--GNGEVRFTYRPSVPAHRRTRESPLSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMK
TmajusDGAT1 MAVAESSQNTTTMSGHG--DSDLNNFRRRKPSSSVIEPSSSGFTS-----TNGVPATGHVAENRDQDRVGAMENATGSVNLIG---NGGG----VVIGNEEKQVGETDIRFTYRPSFPAHRRVRESPLSSDAIFKQSHAGLFNLCIVVLIAVNSRLIIENLMK
EpDGAT1- MTIWESPEIISSDEAAAALRRR----------GGAKEVAEQRLDSEEEKKKEEE-NGK----------LKYTYRASAPAHRRIKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
VgDGAT1- MALLDTPQIGEITTTATTTIRRRTTVKPDAGIGDG----LFDSSSSSKTNSSFEDGDSLNGDFNDKFKEQIGAGDESKDDSKGNGQKIDHGGVK--------KGRE-----T--TVVHYAYRPSSPAHRRIKESPLSSDAIFKQSHAGLFNLCIVVLVAVNGRLIIENLMK
NtDGAT1 MVIMELPESVEMTTTTTTSGIENLNSDLNHSVRR------RRGSNGFEAASAINSSDANMSEDRRDVCGSGAGLETVNERSKSVGESSDVIRKEDDRNDNVANGEESKSTETTTTPFKFAYRASAPAHRRIKESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMK
PfDGAT1 MAILDSPEILDTTSSSADNGAAHHTTLRRRQSARSVPPLLLDSDSNSLEAESAINDSENVRNDANLIENLRGGAVESENEKQESYGKEEGAKVKE--------NGETSNGNGTDVMAVKFTFRPAAPAHRKNKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
ZmL ----------MADSEDAP----PAVHRR-PPRPARG-AAAAQGFAAALRRRLRSGAAVAARASFAADSGDESGPGEPSSS------RRR---DNSGGASSAAGG---RAGAGDFSAFTFRAAAPVHRKAKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
SbDGAT1 ----------MADTDDAPPA--PAVHRR-PPRPARG-AAAAQGFAAKLRRRLSSGAAAAARASFAADSGDESGPGEPSSS------RRR---DNGGDASSAADGG--RGGAGDFSAFTFRAAAPVHRKAKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
OsL MVGSDGDGDGGGGEAHAPAA--PAHHHRRPPRPRGGSGAIVEGFAAALRRRIRSGAAAAARASFGGDSGDEAASGEPSSSSSSSPSRRRGGDSNGAEASSAAGGGGGRGGGGDFSAFTFRAAAPVHRKAKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
OsS MAPPPSLAPDRGGGEP------DDALRLRARAAAAAGDAPAP-----QQQQEQR------HQEQQQQ--LLWYRASAPAHRRVRESPLSSDAIFRQSHAGLLNLCIVVLVAVNSRLIIENLMK
SbDGAT1- MAPPPSMAAASDRAVPGADATEASSLRLRRAPSADAGDLADDSSGDRRENGEPQ-----PPQEQQQQHEMLYYRASAPAHRRVKESPLSSDAIFRQSHAGLLNLCIVVLIAVNSRLIIENLMK
Zms- MAPPPSMPAASDRAGPGRDAGDSSSLRLRRAPSADAGDLAGDSSGGLRENGEPQSPTNPPPQEQQQ-HEMLYYRASAPAHRRVKESPLSSDAIFRQSHAGLLNLCIVVLIAVNSRLIIENLMK
PpDGAT1- MPVKSSNLAGERAATSHINANTKFDLRGCTPAHRVRRESPLSSDAIFHQSHAGLFNLCIVVLIAVNSRLIIENLMK
SmDGAT1- ------------------------MRPSLPAHRRSKESPLSSDAIFTQSHAGLFNLCIVVLVAVNSRLIIENLMK
EaDGAT1- ----------MAANLNEASDLNFSLRRRTGGISSTTVPDSSSETSSSEADYLDGGKGAADVKDRGDGAVEFQNSMKNVERIEKHESRVGLDSRFTYRPSVPAHRTIKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
VvDGAT1- MAICNSPVSVTTSSSSSHADSDLDFSIRKRFGGKGKAVADSSLETETEAAAAAVLEAEKSVGEVGSGGDRGESGSQVVRNGENGVAEVAAK----FAYRPCAPAHRKVKESPLSSDAIFRQSHAGLFNLCIVVLVAVNSRLIIENLMK
GmDGAT1- MAISDEPETVATA----------LNHSSLRRRP---TAAGLFNSPETT--TDSSGDDLAKDSGSDDSISSDAAN--------SQPQQ---KQDTDFSVLK----------FAYRPSVPAHRKVKESPLSSDTIFRQSHAGLFNLCIVVLVAVNSRLIIENLMK
GmDGAT1- MAISDEPESVATA----------LNHSSLRRRPSATSTAGLFNSPETT--TDSSGDDLAKDSGSDDSINSDDAA--------VNSQQQNEKQDTDFSVLK----------FAYRPSVPAHRKVKESPLSSDTIFRQSHAGLFNLCIVVLVAVNSRLIIENLMK
LjDGAT1- MAISEDSESLFAAAAASSVIQ----SG-SSVRRRPSAISAVATVEDE-----SSSEEPVPVRDSGSDVDDSVSSEQ-----HVSPATANREKNQVHDISATK----------FAYRPSAPAHRRVKESPLSSDNIFRH-HAGLFNLCIVVLVAVNSRLIIENLMK
MtDGAT1- MAISDTPETTATATATVTTIETDTDLKRSSLRRRPSATSTAGGLFD------AESAAADAVRDSGSDDSLNGKINNEEEVKDRKTDHAEGIVDDDDDNAVKKNGGNDVINDRENVAVDFKFTYRPSVPAHRRSKESPLSSGNIFRQSHAGLFNLCIVVLVAVNSRLIIENLMK
JcDGAT1- MTILET----TTSGGDGVAESSSDLNVSLRRRRKGTSSDGALPELTSNIVELE----SESGG-QVMMDPGMVTEPETEKINGKDCGGDKDKIDNRENRGRSDIKFTYRPSVPAHRALRESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
VfDGAT1- MTIPETPDNSTDATTSGGAESSSDLNLSLRRRRTASNSDGAVAELASKIDELE----SDAGGGQVIKDPGAEMDSGTLKSNGKDCGTVKDRIENRENRGGSDVKFTYRPSVPAHRALKESPLSSDNIFKQSHAGLFNLCIVVLVAVNSRLIIENIMK
RcDGAT1- MTILETP----ETLGVISSSATSDLNLSLRRRRTSNDSDGALADLASKFDDDDDVRSEDSAENIIEDPVAAVTELATAKSNGKDCVANSNKDKIDSHGGSSDFKLAYRPSVPAHRRSLKESPLSSDLIFKQSHAGLFNLCIVVLVAVNSRLIIENLMK
PtDGAT1- ----------MMESEDLKSNGKECDKVTNENRSDIKFNYRPSMPAHRGVRESPLSSDAIFKQSHAGLFNLCIVVLVAINSRLIIENIIK
PtDGAT1- MAESESPEN-RIAAMESTSSSTSDLNFSIRRRSTVMDSASTEMMGSEGLKSSGKACDKVKIEKQSDMKFNYRPSMPAHSGVRESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIENLIK

Figure 5



lysine (K)



arginine (R)