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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2008 (05.06.2008)

PCT

(10) International Publication Number
WO 2008/065370 A2

(51) International Patent Classification:
C12N 9/10 (2006.01) *C07H 13/00* (2006.01)
C12N 15/82 (2006.01) *A23L 1/03* (2006.01)
A61K 31/7024 (2006.01)

(21) International Application Number: PCT/GB2007/004518

(22) International Filing Date: 26 November 2007 (26.11.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0624067.5 1 December 2006 (01.12.2006) GB
0625302.5 20 December 2006 (20.12.2006) GB

(71) Applicant (*for all designated States except US*): **THE UNIVERSITY OF YORK** [GB/GB]; Heslington, York Y01 5DD (GB).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CAPUTI, Lorenzo** [IT/GB]; Department of Biology, The University of York, York YO10 5DD (GB). **LIM, Eng-Kiat** [GB/GB]; Department of Biology, The University of York, York YO10 5DD (GB). **BOWLES, Dianna, Joy** [GB/GB]; Department of Biology, The University of York, York YO10 5DD (GB).

(74) Agent: **GILHOLM HARRISON**; Marlborough House, Westminster Place, York Business Park, Nether Poppleton, York YO26 6RW (GB).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

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

WO 2008/065370 A2

(54) Title: SESQUITERPENOID MODIFYING ENZYMES

(57) Abstract: We describe glycosyltransferase polypeptides that modify sesquiterpenoids and including pharmaceutical compositions comprising glycosylated sesquiterpenoids; methods to treat diseases, in particular cancer, bacterial and fungal infections and also flavourings and scents comprising glycosylated sesquiterpenoids.

Sesquiterpenoid Modifying Enzymes

The invention relates to glycosyltransferase polypeptides that modify sesquiterpenoids and including pharmaceutical compositions comprising glycosylated sesquiterpenoids; methods to treat diseases, in particular cancer, bacterial and fungal infections and also flavourings and scents comprising glycosylated sesquiterpenoids.

Plant terpenoids, also called isoprenoids are products derived from a five carbon isoprene unit and have diverse activities that include anti-cancer and anti-microbial activity. They are also used to flavour and/or scent a variety of commercial products. Terpenoids are classified with reference to the number of isoprene units that comprise the particular terpenoid. For example a monoterpenoid comprises two isoprene units; a sesquiterpenoid comprises three isoprene units and a di-terpenoid four isoprene units. Polyterpnoids comprise multiple isoprene units. There are many thousands of examples of terpenoids.

Artemisinin is a sesquiterpene lactone endoperoxide and is a natural product produced by the plant *Artemesia annua*. Artemisinin has long been known to have anti-malarial activity and is typically used in combination with anti-malarial therapeutics, for example lumefantrine, mefloquine, amodiaquine, sulfadoxine, chloroquine, in artemisinin combination therapies (ACT). Artemisinin is only produced by the plant under certain conditions and is isolated from leaves. The *in planta* synthesis of artemisinin is via the mevalonate pathway and is a multi-step process that results in the formation of artemisinic acid which is converted to artemisinin. In addition to its anti-malarial activity artemisinin has also been shown to have anti-cancer activity with respect to melanoma, see EP1 658 844 and US5, 219, 880.

Artemisinic acid is a biologically active molecule although not as an anti-malarial compound *per se*. It has been shown to have anti-bacterial and anti-fungal activities see Dhingra *et al* Current Science (2000) 78(6): 709, and also anti-cancer activity. Artemisinic acid is far more abundant than artemisinin in *Artemesia annua* and therefore a great deal of effort has been dedicated to the purification of artemisinic acid from plants as a precursor that can then be chemically converted to artemisinin. For example, US4, 992, 561 describes a process for the conversion of artemisinic acid to artemisinin by modification of artemisinic acid to dihydroartemisinic acid by reduction of the exocyclic methylene group. Dihydroartemisinic acid is then oxidised in two successive steps to artemisinin.

An alternative to purification of artemisinic acid from a plant source is the genetic engineering of microbial cells to reproduce the metabolic pathway for the synthesis of artemisinin intermediates. The final steps in the production of artemisinin are the conversion of farnesyl diphosphate to amorpha-4, 11-diene (amorphadeine) by amorphadeine synthase followed by conversion of amorphadeine to artemisinic acid by a cytochrome P450 monooxygenase. There then follows two photooxidation steps to the conversion of artemisinic acid to artemisinin.

An example of engineering microbial cells is provided in Martin *et al* Nature Biotechnology (2003) 21(7): 796-802, which is incorporated by reference in its entirety. This describes the engineering of the mevalonate pathway in *E.coli* for the production of *terpenoids.* The manuscript describes the transformation of a bacterial host cell with a synthetic amporpha-4, 11-diene synthase gene and the mevalonate pathway from *Saccharomyces cerevisae*. The transformed bacterial cells produce significant amounts of amorphadiene. Ro *et al* Nature (2006) 440: 940-943, the content of which is incorporated by reference in its entirety, describes the engineering of *Saccharomyces cerevisae* with *amorphadeine synthase and the nucleic acid sequence of* a cytochrome P450 monooxygenase that perform the three step conversion of amorphadeine to artemisinic acid. WO2005/033287, the content of which is incorporated by reference in its entirety, discloses much of the content of Martin *et al* and Ro *et al.* In addition WO00/12725, the content of which is incorporated by reference in its entirety, discloses an isolated nucleic acid molecule that encodes amorphadeine synthase from *Artemesia annua*.

A further example of a sesquiterpenoid is farnesol which is a flavouring added to many foods. It has also been shown to have pesticide activity, particularly to mites.

This disclosure relates to the identification of plant glycosyltransferases that glycosylate sesquiterpenoids to provide a glycosylated sesquiterpenoid ester. We describe bacterial cells that are genetically modified to include nucleic acid molecules and the glycosylation of sesquiterpenoids in a whole cell bioreactor. We also disclose the anti microbial activity of glycosylated sesquiterpenoids.

According to an aspect of the invention there is provided a microbial cell wherein the microbial cell is transformed with a vector that includes a nucleic acid molecule selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;

ii) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;and that encodes a glycosyltransferase that glycosylates a sesquiterpenoid;

iii) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;.

Hybridization of a nucleic acid molecule occurs when two complementary nucleic acid molecules undergo an amount of hydrogen bonding to each other. The stringency of hybridization can vary according to the environmental conditions surrounding the nucleic acids, the nature of the hybridization method, and the composition and length of the nucleic acid molecules used. Calculations regarding hybridization conditions required for attaining particular degrees of *stringency are discussed in Sambrook et al., Molecular* Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); and Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes Part I, Chapter 2 (Elsevier, New York, 1993). The T_m is the temperature at which 50% of a given strand of a nucleic acid molecule is hybridized to its complementary strand. The following is an exemplary set of hybridization conditions and is not limiting:

<u>Very High Stringency (allows sequences that share at least 90% identity to hybridize)</u>

Hybridization:	5x SSC at 65°C for 16 hours
Wash twice:	2x SSC at room temperature (RT) for 15 minutes each
Wash twice:	0.5x SSC at 65°C for 20 minutes each

<u>High Stringency (allows sequences that share at least 80% identity to hybridize)</u>

Hybridization:	5x-6x SSC at 65°C-70°C for 16-20 hours
Wash twice:	2x SSC at RT for 5-20 minutes each
Wash twice:	1x SSC at 55°C-70°C for 30 minutes each

Low Stringency (allows sequences that share at least 50% identity to hybridize)

Hybridization:	6x SSC at RT to 55°C for 16-20 hours
Wash at least twice:	2x-3x SSC at RT to 55°C for 20-30 minutes each.

In preferred embodiment of the invention said nucleic acid molecule consists of a nucleic acid sequence as represented in Figures 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12.

In a further preferred embodiment of the invention said microbial cell is transformed with a nucleic acid molecule that encodes a polypeptide wherein said polypeptide is an amorphadeine synthase that catalyses the conversion of farnesyl diphosphate to amorpha-4, 11-diene.

In a preferred embodiment of the invention said nucleic acid molecule comprises a nucleic acid sequence as represented in Figure 14, or a nucleic acid molecule that hybridises to the nucleic acid molecule in Figure 14 and encodes a polypeptide that is an amorphadeine synthase.

In a further preferred embodiment of the invention said microbial cell is transformed with a nucleic acid molecule that encodes a polypeptide wherein said polypeptide is a cytochrome P450 that catalyses the conversion of amorpha-4, 11-diene to artemisinic acid.

In a yet further preferred embodiment of the invention said nucleic acid molecule comprises a nucleic acid sequence as represented in Figure 15, or a nucleic acid molecule that hybridises to the nucleic acid molecule in Figure 15 and encodes a polypeptide that is a P450.

In a further preferred embodiment of the invention said nucleic acid molecule consists of a nucleic acid sequence as represented in Figures 18, 19, 20, 21, 22, 23 or 24.

In a preferred embodiment of the invention said microbial cell is a bacterial cell.

In an alternative preferred embodiment of the invention said microbial cell is a yeast cell.

In a preferred embodiment of the invention said vector is an expression vector and said nucleic acid molecule encoding said glycosyltransferase is operably linked to a promoter.

A vector including nucleic acid (s) according to the invention need not include a promoter or other regulatory sequence, particularly if the vector is to be used to introduce the nucleic acid into cells for recombination into the genome for stable transfection.

Preferably the nucleic acid in the vector is operably linked to an appropriate promoter or other regulatory elements for transcription in a host cell such as a prokaryotic, (e.g. bacterial), or eukaryotic (e.g. fungal, plant, mammalian or insect cell). The vector may be a bi-functional expression vector which functions in multiple hosts. In the example of nucleic acids encoding polypeptides according to the invention this may contain its native promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in the *host cell.*

By "promoter" is meant a nucleotide sequence upstream from the transcriptional initiation site and which contains all the regulatory regions required for transcription. Suitable promoters include constitutive, tissue-specific, inducible, developmental or other promoters for expression in cells. *Such promoters include viral, fungal, bacterial, animal* and plant-derived promoters.

"Operably linked" means joined as part of the same nucleic acid molecule, suitably positioned and oriented for transcription to be initiated from the promoter. DNA operably linked to a promoter is "under transcriptional initiation regulation" of the promoter.

In a preferred embodiment the promoter is an inducible promoter or a developmentally regulated promoter.

Alternatively, or *in addition,* said vectors *are vectors suitable for mammalian cell* transfection or yeast cell transfection. In the latter example multi-copy vectors such as 2μ episomal vectors are preferred. Alternatively yeast CEN vectors and intergrating vectors such as YIP vectors are suitable for transformation of yeast species such as *Saccharomyces cerevisiae* and *Pichia spp.*

According to a further aspect of the invention there is provided the use of a cell according to the invention in the modification of a sequiterpenoid.

In a preferred embodiment of the invention said sequiterpenoid is artemisinic acid or farnesol.

According to a further aspect of the invention there is provided a process for the glycosylation of a sequiterpenoid comprising the steps of:

i) forming a preparation that includes a cell according to the invention and a sequiterpenoid;

ii) cultivating said preparation under conditions that allow the glycosylation of said sequiterpenoid with a sugar; and optionally

iii) isolating and purifying said glycosylated sequiterpenoid from said cell and/or the surrounding cell growth medium.

In a preferred method of the invention said sequiterpenoid is artemisinic acid or farnesol.

In a preferred method of the invention said cell is a bacterial cell.

In an alternative preferred method of the invention said cell is a yeast cell.

If microorganisms are used as organisms in the process according to the invention, they are grown or cultured in the manner with which the skilled worker is familiar, depending on the host organism. As a rule, microorganisms are grown in a liquid medium comprising a carbon source, usually in the form of sugars, a nitrogen source, usually in the form of organic nitrogen sources such as yeast extract or salts such as ammonium sulfate, trace elements such as salts of iron, manganese and magnesium and, if appropriate, vitamins, at temperatures of between 0°C and 100°C, preferably between 10°C and 60°C, while gassing in oxygen.

The pH of the liquid medium can either be kept constant, that is to say regulated during the culturing period, or not. The cultures can be grown batchwise, semi-batchwise or continuously. Nutrients can be provided at the beginning of the fermentation or fed in semi-continuously or continuously. The products produced can be isolated from the organisms as described above by processes known to the skilled worker, for example by extraction, distillation, crystallization, if appropriate precipitation with salt, and/or chromatography. To this end, the organisms can advantageously be disrupted beforehand. In this process, the pH value is advantageously kept between pH 4 and 12, preferably between pH 6 and 9, especially preferably between pH 7 and 8.

An overview of known cultivation methods can be found in the textbook by Chmiel (Bioprozeßtechnik 1. Einführung in die Bioverfahrenstechnik [Bioprocess technology 1. Introduction to Bioprocess technology] (Gustav Fischer Verlag, Stuttgart, 1991)) or in the textbook by Storhas (Bioreaktoren und periphere Einrichtungen [Bioreactors and peripheral equipment] (Vieweg Verlag, Brunswick/Wiesbaden, 1994)).

The culture medium to be used must suitably meet the requirements of the strains in question. Descriptions of culture media for various microorganisms can be found in the textbook "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D.C., USA, 1981).

As described above, these media which can be employed in accordance with the invention usually comprise one or more carbon sources, nitrogen sources, inorganic salts, vitamins and/or trace elements.

Preferred carbon sources are sugars, such as mono-, di- or polysaccharides. Examples of carbon sources are glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose. Sugars can also be added to the media via complex compounds such as molasses or other by-products from sugar refining. The addition of mixtures of a variety of carbon sources may also be advantageous. Other possible carbon sources are oils and fats such as, for example, soya oil, sunflower oil, peanut oil and/or coconut fat, fatty acids such as, for example, palmitic acid, stearic acid and/or linoleic acid, alcohols and/or polyalcohols such as, for example, glycerol, methanol and/or ethanol, and/or organic acids such as, for example, acetic acid and/or lactic acid.

Nitrogen sources are usually organic or inorganic nitrogen compounds or materials comprising these compounds. Examples of nitrogen sources comprise ammonia in liquid or gaseous form or ammonium salts such as ammonium sulfate, ammonium chloride, ammonium phosphate, ammonium carbonate or ammonium nitrate, nitrates, urea, amino acids or complex nitrogen sources such as cornsteep liquor, soya meal, soya protein, yeast extract, meat extract and others. The nitrogen sources can be used individually or as a mixture.

Inorganic salt compounds which may be present in the media comprise the chloride, phosphorus and sulfate salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron.

Inorganic sulfur-containing compounds such as, for example, sulfates, sulfites, dithionites, tetrathionates, thiosulfates, sulfides, or else organic sulfur compounds such as mercaptans and thiols may be used as sources of sulfur for the production of sulfur-containing fine chemicals, in particular of methionine.

Phosphoric acid, potassium dihydrogenphosphate or dipotassium hydrogenphosphate or the corresponding sodium-containing salts may be used as sources of phosphorus.

Chelating agents may be added to the medium in order to keep the metal ions in solution. Particularly suitable chelating agents comprise dihydroxyphenols such as catechol or protocatechuate and organic acids such as citric acid.

The fermentation media used according to the invention for culturing microorganisms usually also comprise other growth factors such as vitamins or growth promoters, which include, for example, biotin, riboflavin, thiamine, folic acid, nicotinic acid, panthothenate and pyridoxine. Growth factors and salts are frequently derived from complex media components such as yeast extract, molasses, cornsteep liquor and the like. It is moreover possible to add suitable precursors to the culture medium. The exact composition of the media compounds heavily depends on the particular experiment and is decided upon individually for each specific case. Information on the optimization of media can be found in the textbook "Applied Microbiol. Physiology, A Practical Approach" (Editors P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). Growth media can also be obtained from commercial suppliers, for example Standard 1 (Merck) or BHI (brain heart infusion, DIFCO) and the like.

All media components are sterilized, either by heat (20 min at 1.5 bar and 121°C) or by filter sterilization. The components may be sterilized either together or, if required, separately. All media components may be present at the start of the cultivation or added continuously or batchwise, as desired.

The culture temperature is normally between 15°C and 45°C, preferably at from 25°C to 40°C, and may be kept constant or may be altered during the experiment. The pH of the medium should be in the range from 5 to 8.5, preferably around 7.0. The pH for cultivation can be controlled during cultivation by adding basic compounds such as sodium hydroxide, potassium hydroxide, ammonia and aqueous ammonia or acidic compounds such as phosphoric acid or sulfuric acid. Foaming can be controlled by employing antifoams such as, for example, fatty acid polyglycol esters. To maintain the stability of plasmids it is possible to add to the medium suitable substances having a selective effect, for example antibiotics. Aerobic conditions are maintained by introducing oxygen or oxygen-containing gas mixtures such as, for example, ambient air into the culture. The temperature of the culture is normally 20°C to 45°C and preferably 25°C to 40°C. The culture is continued until formation of the desired product is at a maximum. This aim is normally achieved within 10 to 160 hours.

The fermentation broths obtained in this way, in particular those comprising polyunsaturated fatty acids; usually contain a dry mass of from 7.5 to 25% by weight.

The fermentation broth can then be processed further. The biomass may, according to requirement, be removed completely or partially from the fermentation broth by separation methods such as, for example, centrifugation, filtration, decanting or a combination of these methods or be left completely in said broth. It is advantageous to process the biomass after its separation.

However, the fermentation broth can also be thickened or concentrated without separating the cells, using known methods such as, for example, with the aid of a rotary evaporator, thin-film evaporator, falling-film evaporator, by reverse osmosis or by nanofiltration.

According to a further aspect of the invention there is provided a transgenic plant wherein said plant is genetically modified by transfection with a vector that includes a nucleic acid molecule selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24;

ii) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24; and that encodes a glycosyltransferase that glycosylates artemisinic acid;

iii) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24.

In a preferred embodiment of the invention said plant is selected from the group consisting of: In a preferred embodiment of the invention said plant is selected from: corn (Zea mays), canola (Brassica napus, Brassica rapa ssp.), alfalfa (Medicago sativa), rice (Oryza sativa), rye (Secale cerale), sorghum (Sorghum bicolor, Sorghum vulgare), sunflower (helianthus annuas), wheat (Tritium aestivum), soybean (Glycine max), tobacco (Nicotiana tabacum), potato (Solanum tuberosum), peanuts (Arachis hypogaea), cotton (Gossypium hirsutum), sweet potato (Iopmoea batatus), cassava (Manihot esculenta), coffee (Cofea spp.), coconut (Cocos nucifera), pineapple (Anana comosus), citris tree (Citrus spp.) cocoa (Theobroma cacao), tea (Camellia senensis), banana (Musa spp.), avacado (Persea americana), fig (Ficus casica), guava (Psidium guajava), mango (Mangifer indica), olive (Olea europaea), papaya (Carica papaya), cashew (Anacardium occidentale), macadamia (Macadamia intergrifolia), almond (Prunus amygdalus), sugar beets (Beta vulgaris), oats, barley, vegetables and ornamentals.

Preferably, plants of the present invention are crop plants for example, cereals and pulses, maize, wheat, potatoes, tapioca, rice, sorghum, millet, cassava, barley, pea, and other root, tuber or seed crops and including peppermint and spearmint.

Important seed crops are oil-seed rape, sugar beet, maize, sunflower, soybean, *and* sorghum. Horticultural plants to which the present invention may be applied may include lettuce, endive, and vegetable brassicas including cabbage, broccoli, celery and cauliflower, and carnations and geraniums. The present invention may be applied in tobacco, cucurbits, carrot, strawberry, cherry, sunflower, tomato, pepper, and chrysanthemum.

Grain plants that provide seeds of interest include oil-seed plants and leguminous plants. Seeds of interest include grain seeds, such as corn, wheat, barley, rice, sorghum, rye, etc. Oil-seed plants include cotton, soybean, safflower, sunflower, Brassica, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava been, lentils, chick pea.

In a preferred embodiment of the invention said plant is of the genus *Artemesia spp;* preferably *Artemesia annua.*

According to a further aspect of the invention there is provided a process for the glycosylation of a sesquiterpenoid comprising the steps of:

i) providing a transgenic plant or a seed transfected with a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

a) a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;

b) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 and that encodes a glycosyltransferase that glycosylates a sesquiterpenoid;

c) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24;

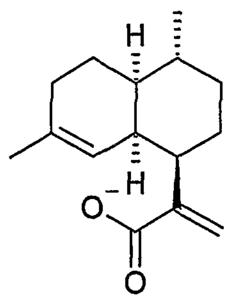
ii) cultivating said plant or seed under conditions that allow the glycosylation of a sesquiterpenoid with a sugar; and optionally

iii) isolating and purifying said glycosylated sesquiterpenoid from said plant and/or said seed.

According to an aspect of the invention there is provided an isolated sesquiterpenoid ester comprising a sugar pendant group to provide glycosylated sesquiterpenoid ester.

According to an aspect of the invention there is provided an isolated artemisinic acid ester comprising a sugar pendant group to provide glycosylated artemisinic acid.

In a preferred embodiment of the invention artemisinic acid has the structure



In a further preferred embodiment of the invention said glycosylated artemisinic acid is glycosylated with a glucose molecule.

In an alternative preferred embodiment of the invention said glycosylated artemisinic acid is glycosylated with a raffinose molecule.

In a further alternative embodiment of the invention said glycosylated artemisinic acid is glycosylated with a glucuronic acid molecule.

According to a further aspect of the invention there is provided glycosylated artemisinic acid for use as a pharmaceutical.

According to a further aspect of the invention there is provided a composition comprising a glycosylated artemisinic acid according to the invention. Preferably said composition is a pharmaceutical composition.

When administered, the compositions of the present invention are administered in pharmaceutically acceptable preparations. Such preparations may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives and compatible carriers.

The therapeutics of the invention can be administered by any conventional route, including injection or by gradual infusion over time. The administration may be, for example, oral, intravenous, intraperitoneal, intramuscular, intracavity, subcutaneous, or transdermal.

The compositions of the invention are administered in effective amounts. An "effective amount" is that amount of a composition that alone, or together with further doses, produces the desired response. In the case of treating a particular disease, such as cancer, the desired response is inhibiting the progression of the disease. This may involve only slowing the progression of the disease temporarily, although more preferably, it involves halting the progression of the disease permanently. This can be monitored by routine methods.

Such amounts will depend, of course, on the particular condition being treated, the severity of the condition, the individual patient parameters including age, physical condition, size and weight, the duration of the treatment, the nature of concurrent therapy (if any), the specific route of administration and like factors within the knowledge and expertise of the health practitioner. These factors are well known to those of ordinary skill in the art and can be addressed with no more than routine experimentation. It is generally preferred that a maximum dose of the individual components or combinations thereof be used, that is, the highest safe dose according to sound medical judgment.

The pharmaceutical compositions used in the foregoing methods preferably are sterile and contain an effective amount of glycosylated artemisinic acid for producing the desired response in a unit of weight or volume suitable for administration to a patient. The response can, for example, be measured by measuring the physiological effects of the composition, such as regression of a tumour, decrease of disease symptoms, modulation of apoptosis, etc.

The doses of glycosylated artemisinic acid administered to a subject can be chosen in accordance with different parameters, in particular in accordance with the mode of administration used and the state of the subject. Other factors include the desired period of treatment. In the event that a response in a subject is insufficient at the initial doses applied, higher doses (or effectively higher doses by a different, more localized delivery route) may be employed to the extent that patient tolerance permits.

Other protocols for the administration of glycosylated artemisinic acid will be known to one of ordinary skill in the art, in which the dose amount, schedule of injections, sites of injections, mode of administration (e.g., intra-tumoural) and the like vary from the foregoing. Administration of glycosylated artemisinic acid compositions to mammals other than humans, (e.g. for testing purposes or veterinary therapeutic purposes), is carried out under substantially the same conditions as described above. A subject, as used herein, is a mammal, preferably a human, and including a non-human primate, cow, horse, pig, sheep, goat, dog, cat or rodent.

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

Glycosylated artemisinic acid compositions may be combined, if desired, with a pharmaceutically-acceptable carrier. The term "pharmaceutically-acceptable carrier" as used herein means one or more compatible solid or liquid fillers, diluents or encapsulating substances which are suitable for administration into a human. The term "carrier" denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. The components of the pharmaceutical compositions also are capable of being co-mingled with the molecules of the present invention, and with each other, in a manner such that there is no interaction which would substantially impair the desired pharmaceutical efficacy.

The pharmaceutical compositions may contain suitable buffering agents, including: acetic acid in a salt; citric acid in a salt; boric acid in a salt; and phosphoric acid in a salt.

The pharmaceutical compositions also may contain, optionally, suitable preservatives, such as: benzalkonium chloride; chlorobutanol; parabens and thimerosal.

The pharmaceutical compositions may conveniently be presented in unit dosage form and may be prepared by any of the methods well-known in the art of pharmacy. All methods include the step of bringing the active agent into association with a carrier which constitutes one or more accessory ingredients. In general, the compositions are prepared by uniformly and intimately bringing the active compound into association with a liquid carrier, a finely divided solid carrier, or both, and then, if necessary, shaping the product.

Compositions suitable for oral administration may be presented as discrete units, such as capsules, tablets, lozenges, each containing a predetermined amount of the active compound. Other compositions include suspensions in aqueous liquids or non-aqueous liquids such as syrup, elixir or an emulsion.

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

In a preferred embodiment of the invention said pharmaceutical composition is a cream adapted for topical application.

According to a further aspect of the invention there is provided a method to treat a fungal infection comprising administering an effective amount of glycosylated artemisinic acid according to the invention to an animal; preferably a human.

According to a further aspect of the invention there is provided a method to treat a bacterial infection comprising administering an effective amount of glycosylated artemisinic acid according to the invention to an animal; preferably a human.

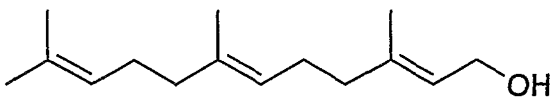
In a preferred method of the invention said treatment is the topical application of; preferably glycosylated artemisinic acid is included in a cream.

According to a further aspect of the invention there is provided a method to treat a cancer comprising administering an effective amount of glycosylated artemisinic acid according to the invention to an animal; preferably a human.

As used herein, the term "cancer" refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting, for example, lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumours, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, e.g., which include malignant tumours composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumours of mesenchymal derivation.

According to an aspect of the invention there is provided an isolated farnesol ester comprising a sugar pendant group to provide glycosylated farnesol.

In a preferred embodiment of the invention farnesol has the structure



In a further preferred embodiment of the invention said glycosylated farnesol is *glycosylated with a glucose molecule.*

In an alternative preferred embodiment of the invention said glycosylated farnesol is glycosylated with a raffinose molecule.

In a further alternative embodiment of the invention said glycosylated farnesol is glycosylated with a glucuronic acid molecule.

According to a further aspect of the invention there is provided a composition comprising a glycosylated farnesol according to the invention.

According to a further aspect of the invention there is provided the use of glycosylated farnesol according to the invention as a flavouring.

According to a further aspect of the invention there is provided the use of glycosylated farnesol according to the invention as a food additive.

According to a further aspect of the invention there is provided the use of glycosylated farnesol according to the invention as a pesticide.

Throughout the description and claims of this specification, the words "comprise" and "contain" and variations of the words, for example "comprising" and "comprises", means "including but not limited to", and is not intended to (and does not) exclude other moieties, additives, components, integers or steps.

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

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

An embodiment of the invention will now be described by example only and with reference to the following figures:

Figure 1: (A) Chemical structures of artemisinic acid and artemisinic glucose ester. (B) TLC analysis of representative GTs capable of glucosylating artemisinic acid. (C) Relative activity of GTs towards artemisinic acid;

Figure 2 Nucleic acid and amino acid sequence of UGT74B1;

Figure 3 Nucleic acid and amino acid sequence of UGT74D1;

Figure 4 Nucleic acid and amino acid sequence of UGT74E2;

Figure 5 Nucleic acid and amino acid sequence of UGT74F1;

Figure 6 Nucleic acid and amino acid sequence of UGT74F2;

Figure 7 Nucleic acid and amino acid sequence of UGT75B1;

Figure 8 Nucleic acid and amino acid sequence of UGT75B2;

Figure 9 Nucleic acid and amino acid sequence of UGT75D1;

Materials and Methods

Recombinant GTs expression and purification

Recombinant GTs were expressed as fusion proteins with glutathione-S-transferase (GST) attached to the N-terminus of the GTs. The GST gene fusion vector pGEX-2T (Amersham Biotech) containing the cDNA of GTs was transformed into *E. coli* BL21 for recombinant protein expression. The bacterial cells were grown in 75 ml of 2×YT medium containing 50 μg/ml ampicillin at 20°C until A_{600} reading reaches 1.0. The culture was then incubated with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 24 h at 20°C. Cells were harvested (5000×*g* for 5 min), resuspended (5 ml of ice-cold phosphate-buffered saline), disrupted by lysozyme (1mg/ml) and centrifuged again (40000×g for 15 min). The supernatant was mixed with 100 μl of 50% glutathione-coupled Sepharose at room temperature for 30 min. The beads were washed with phosphate buffer saline, and the absorbed proteins were eluted with 20 mM reduced-form glutathione according to the manufacturer's instructions. The protein concentration was determined using the Bradford method and bovine serum albumin as reference.

TLC analysis of the enzyme activity

Each reaction mix (20 µl) contained 100 mM TRIS-HCl (pH 7.0), 3.7 µM ^{14}C UDP-glucose (11.6 GBq/mmol, Amersham), 1 mM artemisinic acid and 300 ng of enzyme. The reaction was carried out at 30°C for 2 h. The reaction mix was stored at –20°C before TLC analysis.

The reaction mixtures were loaded on to Silica gel 60 TLC plates. The TLC analysis was carried out in a solvent system consisting of ethylacetate/acetone/dichloromethane/methanol/water (20:15:6:5:4, v/v/v/v/v). The plates were dried and exposed to phosphor-imaging screens (Molecular Dynamics) for 24 h. The screens were read using a Molecular Imager FX scanner (BioRad) supplied with Quantity One software (BioRad). The amount of ^{14}C UDP-glucose transferred by the enzymes to the substrates was calculated using a regression equation obtained by analysing ^{14}C UDP-glucose standards ranged between 0.008-0.555 kBq with the TLC method described above.

HPLC analysis of artemisinic acid and the glucoside

Reaction mix for HPLC analysis was performed in 200 µl volume containing 100 mM TRIS-HCl (pH 7.0), 2.5 mM UDP-glucose, 1 mM artemisinic acid and 1 µg of enzyme. *The reaction was incubated at 30°C for 2 h, and stored at –20°C prior to HPLC analysis.*

Reverse phase HPLC (SpectraSYSTEM HPLC systems and UV6000LP photodiode array detector, TermoQuest) was carried out using a Columbus 5-µm C18 column (250×4.6 mm, Phenomenex) at a flow rate of 1 ml/min with a linear gradient of 10-50% solvent A (methanol) against solvent B (10 mM ammonium acetate) over 10 min, followed by a linear gradient 10-50% A over 20 min against B. The column was then washed with 100% A for 5 min. The chromatography was monitored at 210 nm.

HPLC-MS analysis of glucoside

The glucoside formed in the enzymatic reaction was confirmed using an Agilent 1100 Series HPLC system (Agilente Technologies) coupled with a QStar hybrid quadrupole-TOF mass spectrometer (Applied Biosystems). The HPLC was performed with a Columbus 5 µ C_{18} column (150 × 3.2 mm, Phenomenex) at a flow rate of 0.5 ml/min following the conditions described in the previous section. The MS analysis was carried out in a positive ion mode. Total ion current and ion traces for specific $[M+H^+]$, $[M+NH_4^+]$ and $[M+Na^+]$ adducts ions were used to detect the compounds. MS-MS analysis was performed on the specific ions using different collision energies.

Claims

1. A microbial cell wherein the microbial cell is transformed with a vector that includes a nucleic acid molecule selected from the group consisting of:

   i) a nucleic acid molecule comprising a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;

   ii) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24;and that encodes a glycosyltransferase that glycosylates a sesquiterpenoid;

   iii) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;.

2. A cell according to claim 1 wherein said nucleic acid molecule consists of a nucleic acid sequence as represented in Figures 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12.

3. A cell according to claim 2 wherein said microbial cell is transformed with a nucleic acid molecule that encodes a polypeptide wherein said polypeptide is an amorphadeine synthase that catalyses the conversion of farnesyl diphosphate to amorpha-4, 11-diene.

4. A cell according to claim 3 wherein said nucleic acid molecule comprises a nucleic acid sequence as represented in Figure 14, or a nucleic acid molecule that hybridises to the nucleic acid molecule in Figure 14 and encodes a polypeptide that is an amorphadeine synthase.

5. A cell according to any of claims 2-4 wherein said microbial cell is transformed with a nucleic acid molecule that encodes a polypeptide wherein said polypeptide is a cytochrome P450 that catalyses the conversion of amorpha-4, 11-diene to artemisinic acid.

6. A cell according to claim 5 wherein said nucleic acid molecule comprises a nucleic acid sequence as represented in Figure 15, or a nucleic acid molecule that hybridises to the nucleic acid molecule in Figure 15 and encodes a polypeptide that is a P450.

7. A cell according to claim1 wherein said nucleic acid molecule consists of a nucleic acid sequence as represented in Figures 18, 19, 20, 21, 22, 23 or 24.

8. A cell according to any of claims 1-7 wherein said microbial cell is a bacterial cell.

9. A cell according to any of claims 1-7 wherein said microbial cell is a yeast cell.

10. A cell according to any of claims 1-9 wherein said vector is an expression vector and said nucleic acid molecule encoding said glycosyltransferase is operably linked to a *promoter.*

11. The use of a cell according to any of claims 1-10 for the modification of a sequiterpenoid.

12. Use according to claim 11 wherein said sequiterpenoid is artemisinic acid or farnesol.

13. A process for the glycosylation of a sequiterpenoid comprising the steps of:

   i) forming a preparation that includes a cell according to the invention and a sequiterpenoid;

   ii) cultivating said preparation under conditions that allow the glycosylation of said sequiterpenoid with a sugar; and optionally

   iii) isolating and purifying said glycosylated sequiterpenoid from said cell and/or the surrounding cell growth medium.

15. A process according to claim 14 wherein said sequiterpenoid is artemisinic acid or farnesol.

16. A process according to claim 14 or 15 wherein said cell is a bacterial cell.

17. A process according to claim 14 or 15 wherein said cell is a yeast cell.

18. A transgenic plant wherein said plant is genetically modified by transfection with a vector that includes a nucleic acid molecule selected from the group consisting of:

i) a nucleic acid molecule comprising a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24;

ii) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24; and that encodes a glycosyltransferase that glycosylates artemisinic acid;

iii) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24..

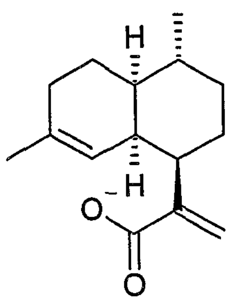
19. A plant according to claim 18 wherein said nucleic acid molecule comprises a nucleic acid sequence as represented in Figures 2, 3, 4, 5, 6; 7, 8, 9, 10, 11 or 12.

20. A plant according to claim 19 wherein said plant is of the genus *Artemesia spp.*

21. A plant according to claim 20 wherein said plant is *Artemesia annua.*

22. An isolated artemisinic acid ester comprising a sugar pendant group to provide glycosylated artemisinic acid.

23. An ester according to claim 22 wherein artemisinic acid has the structure



24. An ester according to claim 22 or 23 wherein said glycosylated artemisinic acid is glycosylated with a glucose molecule.

25. An ester according to claim 22 or 23 wherein said glycosylated artemisinic acid is glycosylated with a raffinose molecule.

26. An ester according to claim 22 or 23 wherein said glycosylated artemisinic acid is glycosylated with a glucuronic acid molecule.

27. The use of glycosylated artemisinic acid as a pharmaceutical.

28. A composition comprising a glycosylated artemisinic acid according to any of claims 22-26.

29. A composition according to claim 28 wherein said composition is a pharmaceutical composition.

30. A composition according to claim 29 wherein said pharmaceutical composition is a cream adapted for topical application.

31. A method to treat a fungal infection comprising administering an effective amount of glycosylated artemisinic acid according to any of claims 22-26 to an animal, preferably a human.

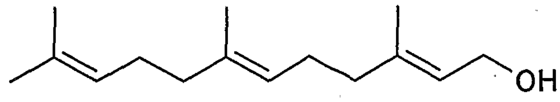
32. A method to treat a bacterial infection comprising administering an effective amount of glycosylated artemisinic acid according to any of claims 22-26 to an animal, preferably a human.

33. A method according to claim 31 or 32 wherein said treatment is the topical application of glycosylated artemisinic acid.

34. A method to treat a cancer comprising administering an effective amount of glycosylated artemisinic acid according to any of claims 22-26 to an animal, preferably a human.

35. An isolated farnesol ester comprising a sugar pendant group to provide glycosylated farnesol.

36. An ester according to claim 35 wherein farnesol has the structure



37. An ester according to claim 35 or 36 wherein said glycosylated farnesol is glycosylated with a glucose molecule.

38. An ester according to claim 35 or 36 wherein said glycosylated farnesol is glycosylated with a raffinose molecule.

39. An ester according to claim 35 or 36 wherein said glycosylated farnesol is glycosylated with a glucuronic acid molecule.

40. A composition comprising a glycosylated farnesol according to any of claims 35-39.

41. The use of glycosylated farnesol according to any of claims 35-39 as a flavouring.

42. The use of glycosylated farnesol according to any of claims 35-39 as a food additive.

43. The use of glycosylated farnesol according to any of claims 35-39 as a pesticide.

44. A process for the glycosylation of a sesquiterpenoid comprising the steps of:

i) providing a transgenic plant or a seed transfected with a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

a) a nucleic acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 ;

b) a nucleic acid molecule that hybridizes under stringent hybridization conditions to a nucleic acid molecule as represented in 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24 and that encodes a glycosyltransferase that glycosylates a sesquiterpenoid;

c) a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence as represented in Figure 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 22, 23 or 24;

ii) cultivating said plant or seed under conditions that allow the glycosylation of a sesquiterpenoid with a sugar; and optionally

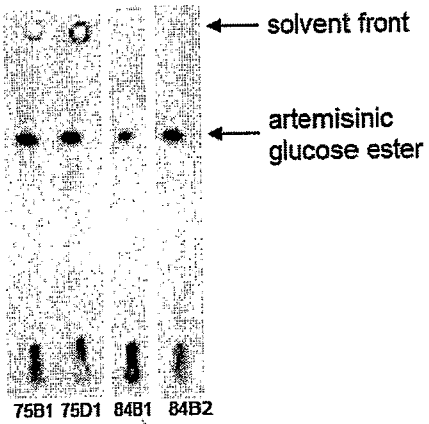
iii) isolating and purifying said glycosylated sesquiterpenoid from said plant and/or said seed.

(A)

Artemisinic acid

Artemisinic glucose ester

(B)



(C)

Enzyme	Relative activity (%)
84A3	10.60
84B2	58.32
84B1	19.06
75B2	25.56
75B1	75.51
75D1	74.72
74E2	64.39
74D1	36.09
74F1	2.74
74F2	5.51
74B1	23.53

Figure 1

UGT74B1

```
atggcggaaacaactcccaaagtgaaaggccacgtcgtaatcttaccatacccagttcaa
 M  A  E  T  T  P  K  V  K  G  H  V  V  I  L  P  Y  P  V  Q
ggccacctaaacccaatggttcaattcgctaaacgtctagtctccaaaaacgtcaaagtc
 G  H  L  N  P  M  V  Q  F  A  K  R  L  V  S  K  N  V  K  V
acaatcgccaccactacctacaccgcctcctcaatcacaacaccatcactctccgtcgaa
 T  I  A  T  T  T  Y  T  A  S  S  I  T  T  P  S  L  S  V  E
ccaatctccgatggattcgatttcatccccataggtatccccggtttcagcgtcgatact
 P  I  S  D  G  F  D  F  I  P  I  G  I  P  G  F  S  V  D  T
tactcagaatccttcaagctcaacggatccgaaaccctaactctcctaatcgagaaattc
 Y  S  E  S  F  K  L  N  G  S  E  T  L  T  L  L  I  E  K  F
aaatccacagattcaccaatcgattgcttaatctacgattcgtttcttccttggggactt
 K  S  T  D  S  P  I  D  C  L  I  Y  D  S  F  L  P  W  G  L
gaagttgctagatctatggaactttcagctgcttctttcttcactaataatctcactgtt
 E  V  A  R  S  M  E  L  S  A  A  S  F  F  T  N  N  L  T  V
tgttctgtgttgcgtaaattctctaacggtgactttcctcttcccgctgatcctaattcg
 C  S  V  L  R  K  F  S  N  G  D  F  P  L  P  A  D  P  N  S
gcgccgtttcgtatccgtggcttaccgtctttgagctacgatgagttaccttcgtttgtg
 A  P  F  R  I  R  G  L  P  S  L  S  Y  D  E  L  P  S  F  V
ggacgtcattggttgactcatcctgagcatggcagagttcttctgaatcagtttcctaac
 G  R  H  W  L  T  H  P  E  H  G  R  V  L  L  N  Q  F  P  N
catgaaaatgctgattggttattcgttaatggctttgaagggttagaagaaacacaagta
 H  E  N  A  D  W  L  F  V  N  G  F  E  G  L  E  E  T  Q  V
agagttttgattctactataaagtttgaaactttatgttacattgttgaattgaaattag
 R  V  L  I  L  L  -  S  L  K  L  Y  V  T  L  L  N  -  N  -
aactgttgttttgattaggattgtgaaaatggtgagtctgatgcaatgaaggcgacgttg
 N  C  C  F  D  -  D  C  E  N  G  E  S  D  A  M  K  A  T  L
atcggaccgatgattccatcggcttatcttgatgatcggatggaagatgataaagactat
 I  G  P  M  I  P  S  A  Y  L  D  D  R  M  E  D  D  K  D  Y
ggtgcgagtctgttgaaaccgatatcgaaggagtgtatggagtggcttgagactaagcag
 G  A  S  L  L  K  P  I  S  K  E  C  M  E  W  L  E  T  K  Q
gctcagtcagtagcatttgtttcgtttggttcgtttgggattctctttgagaagcaactt
 A  Q  S  V  A  F  V  S  F  G  S  F  G  I  L  F  E  K  Q  L
gcagaggtagctattgcgctacaagaatcggatttgaacttcttgtgggtgattaaagaa
 A  E  V  A  I  A  L  Q  E  S  D  L  N  F  L  W  V  I  K  E
gctcatatagcgaaattgcctgaagggtttgtggaatcgactaaagatagagccttgttg
 A  H  I  A  K  L  P  E  G  F  V  E  S  T  K  D  R  A  L  L
gtttcttggtgtaaccagcttgaggttttagctcatgaatcgataggttgcttttttgact
 V  S  W  C  N  Q  L  E  V  L  A  H  E  S  I  G  C  F  L  T
cattgtggttggaactctacgttggaagggttgagtttgggagttccgatggttggtgtg
 H  C  G  W  N  S  T  L  E  G  L  S  L  G  V  P  M  V  G  V
cctcagtggagtgatcagatgaatgatgctaagtttgtggaggaagtttggaaagttggg
 P  Q  W  S  D  Q  M  N  D  A  K  F  V  E  E  V  W  K  V  G
tatagagcgaaagaggaagctggggaagtaatcgtgaagagtgaagaattggtgaggtgt
 Y  R  A  K  E  E  A  G  E  V  I  V  K  S  E  E  L  V  R  C
ttgaaaggagtgatggaaggagagagtagtgtgaagattagagagagttcgaagaagtgg
 L  K  G  V  M  E  G  E  S  S  V  K  I  R  E  S  S  K  K  W
aaagatttggctgtgaaggcaatgagtgaaggaggaagctctgatcgaagcattaacgag
 K  D  L  A  V  K  A  M  S  E  G  G  S  S  D  R  S  I  N  E
tttatagagagtttagggaagtaa
 F  I  E  S  L  G  K  -
```

Figure 2

UGT74D1

```
atgggagagaaagcgaaagcaaatgtgttagtcttctcatttccgatacaaggtcacata
 M  G  E  K  A  K  A  N  V  L  V  F  S  F  P  I  Q  G  H  I
aaccctctcctccaattctcaaaacgcctactctctaaaaacgtcaacgtcacattcctc
 N  P  L  L  Q  F  S  K  R  L  L  S  K  N  V  N  V  T  F  L
accacttcctccacccacaactccatcctccgccgtgccatcaccggcggagccactgct
 T  T  S  S  T  H  N  S  I  L  R  R  A  I  T  G  G  A  T  A
cttcctctctcttttgtccccattgacgatggattcgaggaagatcacccatctacggac
 L  P  L  S  F  V  P  I  D  D  G  F  E  E  D  H  P  S  T  D
acatctcccgactacttcgcaaagttccaagaaaacgtatctcgaagcctctcagagctt
 T  S  P  D  Y  F  A  K  F  Q  E  N  V  S  R  S  L  S  E  L
atctcctcgatggacccaaaaccaaacgccgtcgtttacgactcgtgcctgccttatgtc
 I  S  S  M  D  P  K  P  N  A  V  V  Y  D  S  C  L  P  Y  V
ctcgacgtttgccggaaacatcctggcgttgctgcggcgtcgtttttcactcagtcctcc
 L  D  V  C  R  K  H  P  G  V  A  A  A  S  F  F  T  Q  S  S
accgtgaacgcgacctatattcatttcttgcgtggagagtttaaggagtttcaaaatgat
 T  V  N  A  T  Y  I  H  F  L  R  G  E  F  K  E  F  Q  N  D
gtcgttttgcctgcaatgcctccgctgaagggtaatgacttaccggtgtttctgtacgat
 V  V  L  P  A  M  P  P  L  K  G  N  D  L  P  V  F  L  Y  D
aacaatctctgccggccgttgtttgagctcattagtagccagttcgtgaatgttgacgac
 N  N  L  C  R  P  L  F  E  L  I  S  S  Q  F  V  N  V  D  D
attgacttcttcttggttaactctttcgacgaactcgaagtcgaggtgctacaatggatg
 I  D  F  F  L  V  N  S  F  D  E  L  E  V  E  V  L  Q  W  M
aaaaaccaatggccggtcaagaacataggaccgatgattccatcaatgtacttagacaaa
 K  N  Q  W  P  V  K  N  I  G  P  M  I  P  S  M  Y  L  D  K
cgattagcaggtgacaaagactacggaatcaacctcttcaatgcccaagtcaacgaatgc
 R  L  A  G  D  K  D  Y  G  I  N  L  F  N  A  Q  V  N  E  C
cttgattggcttgactcaaaaccgcccggttcagtgatctacgtgtcttttggaagcttg
 L  D  W  L  D  S  K  P  P  G  S  V  I  Y  V  S  F  G  S  L
gccgtcttaaaagacgatcaaatgatagaagtcgcggctggtctaaaacaaactggccat
 A  V  L  K  D  D  Q  M  I  E  V  A  A  G  L  K  Q  T  G  H
aacttcttatgggttgttagagaaactgaaacaaagaagcttccaagcaattacatagag
 N  F  L  W  V  V  R  E  T  E  T  K  K  L  P  S  N  Y  I  E
gacatttgtgacaagggattgatagtgaattggagtcctcaattacaagttcttgcacat
 D  I  C  D  K  G  L  I  V  N  W  S  P  Q  L  Q  V  L  A  H
aaatcaatcggttgtttcatgactcattgcgggtggaattcgactttagaggcattgagc
 K  S  I  G  C  F  M  T  H  C  G  W  N  S  T  L  E  A  L  S
ttaggagttgctttgataggaatgccggcttatagcgaccagccgactaatgctaagttt
 L  G  V  A  L  I  G  M  P  A  Y  S  D  Q  P  T  N  A  K  F
attgaagatgtgtggaaggttggggttagggttaaggcagatcaaaatgggtttgttccg
 I  E  D  V  W  K  V  G  V  R  V  K  A  D  Q  N  G  F  V  P
aaggaagagattgtgagatgtgttggagaagttatggaagatatgtcggagaaagggaag
 K  E  E  I  V  R  C  V  G  E  V  M  E  D  M  S  E  K  G  K
gagattagaaaaaatgctcggaggttgatggagtttgcaagggaagctttgtctgatgga
 E  I  R  K  N  A  R  R  L  M  E  F  A  R  E  A  L  S  D  G
ggaaattctgataagaatattgatgagtttgttgctaaaattgtgaggtaa
 G  N  S  D  K  N  I  D  E  F  V  A  K  I  V  R  -
```

Figure 3

UGT74E2

```
atgagagaaggatctcatcttatcgtcttgcctttcccaggacaaggccacataactcca
 M  R  E  G  S  H  L  I  V  L  P  F  P  G  Q  G  H  I  T  P
atgtcccagttctgcaaacgcttagcctcaaaaggtcttaagctcactctggtcctcgtc
 M  S  Q  F  C  K  R  L  A  S  K  G  L  K  L  T  L  V  L  V
tccgacaaaccctctcctccatacaaaacagagcacgactcaatcactgtcttccccatc
 S  D  K  P  S  P  P  Y  K  T  E  H  D  S  I  T  V  F  P  I
tccaacggcttccaagaaggcgaggaaccattacaagacctcgatgattacatggaaaga
 S  N  G  F  Q  E  G  E  E  P  L  Q  D  L  D  D  Y  M  E  R
gtagaaaccagcatcaaaaacaccttaccgaagttggttgaagacatgaaactgtcggga
 V  E  T  S  I  K  N  T  L  P  K  L  V  E  D  M  K  L  S  G
aatccacctagggctatcgtgtacgactccaccatgccatggcttcttgatgtagctcat
 N  P  P  R  A  I  V  Y  D  S  T  M  P  W  L  L  D  V  A  H
agttatggattgagcggtgccgtgtttttcacgcaaccttggcttgtcacagctatttac
 S  Y  G  L  S  G  A  V  F  F  T  Q  P  W  L  V  T  A  I  Y
taccatgttttcaagggttcgttctctgtaccgtctacaaagtacggtcactcgacatta
 Y  H  V  F  K  G  S  F  S  V  P  S  T  K  Y  G  H  S  T  L
gcatctttcccttcgttcccgatgctgactgcaaatgatttgccgtctttcctctgcgaa
 A  S  F  P  S  F  P  M  L  T  A  N  D  L  P  S  F  L  C  E
tcgtcctcatacccgaatatactgaggattgtggtggatcagctctcaaacattgatcga
 S  S  S  Y  P  N  I  L  R  I  V  V  D  Q  L  S  N  I  D  R
gtcgacatagtgttgtgcaacactttcgataaattggaggaaaagttgttgaaatgggtc
 V  D  I  V  L  C  N  T  F  D  K  L  E  E  K  L  L  K  W  V
caaagcttgtggccagtcttgaatattggaccaacggttccatcgatgtatttagacaaa
 Q  S  L  W  P  V  L  N  I  G  P  T  V  P  S  M  Y  L  D  K
cgactgtctgaagacaagaactacggttttagcctcttcaatgcgaaagtcgctgaatgc
 R  L  S  E  D  K  N  Y  G  F  S  L  F  N  A  K  V  A  E  C
atggagtggctaaactcaaaggagcctaattctgttgtctatttatcattcggaagtttg
 M  E  W  L  N  S  K  E  P  N  S  V  V  Y  L  S  F  G  S  L
gtgattctaaaagaagatcaaatgttggaactcgctgcgggtctgaaacagagcggacgt
 V  I  L  K  E  D  Q  M  L  E  L  A  A  G  L  K  Q  S  G  R
ttctttctgtgggttgtgagagagacagagacacacaaacttccaagaaactatgtcgag
 F  F  L  W  V  V  R  E  T  E  T  H  K  L  P  R  N  Y  V  E
gaaatcggtgaaaaaggacttattgtaagctggagtcctcagcttgacgtacttgcacat
 E  I  G  E  K  G  L  I  V  S  W  S  P  Q  L  D  V  L  A  H
aaatcaatcggttgtttcttgacacactgtggatggaactcgacgttagagggattgagt
 K  S  I  G  C  F  L  T  H  C  G  W  N  S  T  L  E  G  L  S
ttgggagttccaatgattggtatgccacactggactgatcagcccacgaatgctaagttc
 L  G  V  P  M  I  G  M  P  H  W  T  D  Q  P  T  N  A  K  F
atgcaggatgtgtggaaggttggggtaagggttaaggcagaaggtgatgggtttgtgaga
 M  Q  D  V  W  K  V  G  V  R  V  K  A  E  G  D  G  F  V  R
agagaagagattatgagaagtgtggaagaagttatggagggagagaaaggaaaagagatt
 R  E  E  I  M  R  S  V  E  E  V  M  E  G  E  K  G  K  E  I
agaaagaatgctgagaaatggaaagtgttggctcaagaggcagtttctgaaggaggtagc
 R  K  N  A  E  K  W  K  V  L  A  Q  E  A  V  S  E  G  G  S
tctgataagagcatcaatgagtttgtttctatgttttgttga
 S  D  K  S  I  N  E  F  V  S  M  F  C  -
```

Figure 4

UGT74F1

```
atgagaggacatgtattagcagtgccatttccaagccaaggacacatcaccccgattcgc
 M  R  G  H  V  L  A  V  P  F  P  S  Q  G  H  I  T  P  I  R
caattctgcaaacgacttcactccaaaggtttcaaaaccactcacactctcaccactttt
 Q  F  C  K  R  L  H  S  K  G  F  K  T  T  H  T  L  T  T  F
atcttcaacacaatccacctcgacccatctagtcctatctccatagccacaatctccgat
 I  F  N  T  I  H  L  D  P  S  S  P  I  S  I  A  T  I  S  D
ggctatgaccagggagggttctcatcagccggttctgtcccggagtacctacaaaacttc
 G  Y  D  Q  G  G  F  S  S  A  G  S  V  P  E  Y  L  Q  N  F
aaaaccttcggctccaaaaccgtcgctgatatcatccgcaaacaccagagtactgataac
 K  T  F  G  S  K  T  V  A  D  I  I  R  K  H  Q  S  T  D  N
cctattacttgtatcgtctatgattctttcatgccttgggcgcttgaccttgcaatggat
 P  I  T  C  I  V  Y  D  S  F  M  P  W  A  L  D  L  A  M  D
tttggtctagctgcggctcctttcttcacgcagtcttgcgccgttaactatatcaattat
 F  G  L  A  A  A  P  F  F  T  Q  S  C  A  V  N  Y  I  N  Y
ctttcttacataaacaatggtagcttgacacttcccatcaaggatttgcctcttcttgag
 L  S  Y  I  N  N  G  S  L  T  L  P  I  K  D  L  P  L  L  E
ctccaagatttgcctactttcgtcactcctactggttcacaccttgcttactttgagatg
 L  Q  D  L  P  T  F  V  T  P  T  G  S  H  L  A  Y  F  E  M
gtgcttcaacagttcaccaacttcgacaaagctgatttcgtactcgttaattccttccat
 V  L  Q  Q  F  T  N  F  D  K  A  D  F  V  L  V  N  S  F  H
gacctcgaccttcatgaagaggagttgttgtcgaaagtatgtcctgtgttgacaattggt
 D  L  D  L  H  E  E  E  L  L  S  K  V  C  P  V  L  T  I  G
ccaactgttccatcaatgtacttagaccaacagatcaaatcagacaacgactatgatctg
 P  T  V  P  S  M  Y  L  D  Q  Q  I  K  S  D  N  D  Y  D  L
aacctctttgacttaaaagaagctgccttatgcactgactggctagacaagaggccagaa
 N  L  F  D  L  K  E  A  A  L  C  T  D  W  L  D  K  R  P  E
ggatcggtagtatatatagcttttgggagcatggctaaactgagtagtgagcagatggaa
 G  S  V  V  Y  I  A  F  G  S  M  A  K  L  S  S  E  Q  M  E
gagattgcttcggcgataagcaacttcagctacctctgggttgtcagagcttcagaggag
 E  I  A  S  A  I  S  N  F  S  Y  L  W  V  V  R  A  S  E  E
tcaaagctcccaccagggtttcttgaaacagtggataaagacaagagcttggtcttgaag
 S  K  L  P  P  G  F  L  E  T  V  D  K  D  K  S  L  V  L  K
tggagtcctcagcttcaagttctgtcaaacaaagccatcggttgtttcatgactcactgt
 W  S  P  Q  L  Q  V  L  S  N  K  A  I  G  C  F  M  T  H  C
ggctggaactcaaccatggagggtttgagtttaggggttcccatggtggctatgcctcaa
 G  W  N  S  T  M  E  G  L  S  L  G  V  P  M  V  A  M  P  Q
tggactgatcaaccaatgaatgcaaagtatatacaagatgtatggaaggttggggttcgt
 W  T  D  Q  P  M  N  A  K  Y  I  Q  D  V  W  K  V  G  V  R
gtgaaagcagagaaagaaagtggcatttgcaaaagagaggagattgagtttagcatcaag
 V  K  A  E  K  E  S  G  I  C  K  R  E  E  I  E  F  S  I  K
gaagtgatggaaggagagaagagcaaagagatgaaagagaatgcgggaaaatggagagac
 E  V  M  E  G  E  K  S  K  E  M  K  E  N  A  G  K  W  R  D
ttggctgtgaagtcactcagtgaaggaggttctacagatatcaacattaacgaatttgta
 L  A  V  K  S  L  S  E  G  G  S  T  D  I  N  I  N  E  F  V
tcaaaaattcaaatcaaataa
 S  K  I  Q  I  K  -
```

Figure 5

UGT74F2

```
atggagcataagagaggacatgtattagcagtgccgtacccaacgcaaggacacatcaca
 M  E  H  K  R  G  H  V  L  A  V  P  Y  P  T  Q  G  H  I  T
ccattccgccaattctgcaaacgacttcacttcaaaggtctcaaaaccactctcgctctc
 P  F  R  Q  F  C  K  R  L  H  F  K  G  L  K  T  T  L  A  L
accactttcgtcttcaactccatcaatcctgacctatccggtccaatctccatagccacc
 T  T  F  V  F  N  S  I  N  P  D  L  S  G  P  I  S  I  A  T
atctccgatggctatgaccatggggggtttcgagacagctgactccatcgacgactacctc
 I  S  D  G  Y  D  H  G  G  F  E  T  A  D  S  I  D  D  Y  L
aaagactttaaaacttccggctcgaaaaccattgcagacatcatccaaaaacaccagact
 K  D  F  K  T  S  G  S  K  T  I  A  D  I  I  Q  K  H  Q  T
agtgataaccccatcacttgtatcgtctatgatgctttcctgccttgggcacttgacgtt
 S  D  N  P  I  T  C  I  V  Y  D  A  F  L  P  W  A  L  D  V
gctagagagtttggtttagttgcgactcctttctttacgcagccttgtgctgttaactat
 A  R  E  F  G  L  V  A  T  P  F  F  T  Q  P  C  A  V  N  Y
gtttattatctttcttacataaacaatggaagcttgcaacttcccattgaggaattgcct
 V  Y  Y  L  S  Y  I  N  N  G  S  L  Q  L  P  I  E  E  L  P
tttcttgagctccaagatttgccttctttcttctctgtttctggctcttatcctgcttac
 F  L  E  L  Q  D  L  P  S  F  F  S  V  S  G  S  Y  P  A  Y
tttgagatggtgcttcaacagttcataaatttcgaaaaagctgatttcgttctcgttaat
 F  E  M  V  L  Q  Q  F  I  N  F  E  K  A  D  F  V  L  V  N
agcttccaagagttggaactgcatgagaatgaattgtggtcgaaagcttgtcctgtgttg
 S  F  Q  E  L  E  L  H  E  N  E  L  W  S  K  A  C  P  V  L
acaattggtccaactattccatcaatttacttagaccaacgtatcaaatcagacaccggc
 T  I  G  P  T  I  P  S  I  Y  L  D  Q  R  I  K  S  D  T  G
tatgatcttaatctctttgaatcgaaagatgattccttctgcattaactggctcgacaca
 Y  D  L  N  L  F  E  S  K  D  D  S  F  C  I  N  W  L  D  T
aggccacaagggtcggtggtgtacgtagcattcggaagcatggctcagctgactaatgtg
 R  P  Q  G  S  V  V  Y  V  A  F  G  S  M  A  Q  L  T  N  V
cagatggaggagcttgcttcagcagtaagcaacttcagcttcctgtgggtggtcagatct
 Q  M  E  E  L  A  S  A  V  S  N  F  S  F  L  W  V  V  R  S
tcagaggaggaaaaactcccatcagggtttcttgagacagtgaataaagaaaagagcttg
 S  E  E  E  K  L  P  S  G  F  L  E  T  V  N  K  E  K  S  L
gtcttgaaatggagtcctcagcttcaagttctgtcaaacaaagccatcggttgtttcttg
 V  L  K  W  S  P  Q  L  Q  V  L  S  N  K  A  I  G  C  F  L
actcactgtggctggaactcaaccatggaggctttgaccttcggggttcccatggtggca
 T  H  C  G  W  N  S  T  M  E  A  L  T  F  G  V  P  M  V  A
atgccccaatggactgatcaaccgatgaacgcaaagtacatacaagatgtgtggaaggct
 M  P  Q  W  T  D  Q  P  M  N  A  K  Y  I  Q  D  V  W  K  A
ggagttcgtgtgaagacagagaaggagagtgggattgccaagagagaggagattgagttt
 G  V  R  V  K  T  E  K  E  S  G  I  A  K  R  E  E  I  E  F
agcattaaggaagtgatggaaggagagaggagcaaagagatgaagaagaacgtgaagaaa
 S  I  K  E  V  M  E  G  E  R  S  K  E  M  K  K  N  V  K  K
tggagagacttggctgtcaagtcactcaatgaaggaggttctacggatactaacattgat
 W  R  D  L  A  V  K  S  L  N  E  G  G  S  T  D  T  N  I  D
acatttgtatcaagggttcagagcaaatag
 T  F  V  S  R  V  Q  S  K  -
```

Figure 6

UGT75B1

```
atggcgccaccgcattttctactggtaacgtttccggcgcaaggtcacgtgaacccatct
 M  A  P  P  H  F  L  L  V  T  F  P  A  Q  G  H  V  N  P  S
ctccgttttgctcgtcggctcatcaaaagaaccggcgcacgtgtcactttcgtcacttgt
 L  R  F  A  R  R  L  I  K  R  T  G  A  R  V  T  F  V  T  C
gtctccgtcttccacaactccatgatcgcaaaccacaacaaagtcgaaaatctctcttc
 V  S  V  F  H  N  S  M  I  A  N  H  N  K  V  E  N  L  S  F
cttactttctccgacggtttcgacgatggaggcatttccacctacgaagaccgtcagaa
 L  T  F  S  D  G  F  D  D  G  G  I  S  T  Y  E  D  R  Q  K
aggtcggtgaatctcaaggttaacggcgataaggcactatcggatttcatcgaagctact
 R  S  V  N  L  K  V  N  G  D  K  A  L  S  D  F  I  E  A  T
aagaatggtgactctcccgtgacttgcttgatctacacgattcttctcaattgggctcca
 K  N  G  D  S  P  V  T  C  L  I  Y  T  I  L  L  N  W  A  P
aaagtagcacgtagatttcaacttccctccgctcttctctggatccaaccggctttggtt
 K  V  A  R  R  F  Q  L  P  S  A  L  L  W  I  Q  P  A  L  V
ttcaacatctattacactcatttcatgggaaacaagtccgtttttcgagttacctaatctg
 F  N  I  Y  Y  T  H  F  M  G  N  K  S  V  F  E  L  P  N  L
tcttctctggaaatcagagatcttccatctttcctcacaccttccaacacaaacaaaggc
 S  S  L  E  I  R  D  L  P  S  F  L  T  P  S  N  T  N  K  G
gcatacgatgcgtttcaagaaatgatggagtttctcataaaagaaaccaaaccgaaaatt
 A  Y  D  A  F  Q  E  M  M  E  F  L  I  K  E  T  K  P  K  I
ctcatcaacactttcgattcgctggaaccagaggccttaacggctttcccgaatatcgat
 L  I  N  T  F  D  S  L  E  P  E  A  L  T  A  F  P  N  I  D
atggtggcggttggtcctttacttcccacggagatttttctcaggaagcaccaacaaatca
 M  V  A  V  G  P  L  L  P  T  E  I  F  S  G  S  T  N  K  S
gttaaagatcaaagtagtagttatacactttggctagactcgaaaacagagtcctctgtt
 V  K  D  Q  S  S  S  Y  T  L  W  L  D  S  K  T  E  S  S  V
atttacgtttcctttggaacaatggttgagttgtccaagaaacagatagaggaactagcg
 I  Y  V  S  F  G  T  M  V  E  L  S  K  K  Q  I  E  E  L  A
agagcactcatagaagggaaacgaccgtttttgtgggttataactgataaatccaacaga
 R  A  L  I  E  G  K  R  P  F  L  W  V  I  T  D  K  S  N  R
gaaacgaaaacagaaggagaagaagagacagagattgagaagatagctggattcagacac
 E  T  K  T  E  G  E  E  E  T  E  I  E  K  I  A  G  F  R  H
gagcttgaagaggttgggatgattgtgtcgtggtgttcgcagatagaggttttaagtcac
 E  L  E  E  V  G  M  I  V  S  W  C  S  Q  I  E  V  L  S  H
cgagccgtaggttgttttgtgactcattgtgggtggagctcgacgctggagagtttggtt
 R  A  V  G  C  F  V  T  H  C  G  W  S  S  T  L  E  S  L  V
cttggcgttccggttgtggcgtttccgatgtggtcggatcaaccgacgaacgcgaagcta
 L  G  V  P  V  V  A  F  P  M  W  S  D  Q  P  T  N  A  K  L
ctggaagaaagttggaagactggtgtgagggtaagagagaacaaggatggtttggtggag
 L  E  E  S  W  K  T  G  V  R  V  R  E  N  K  D  G  L  V  E
agaggagagatcaggaggtgtttggaagccgtgatggaggagaagtcggtggagttgagg
 R  G  E  I  R  R  C  L  E  A  V  M  E  E  K  S  V  E  L  R
gaaaacgcaaagaaatggaagcgtttagcgatggaagcgggtagagaaggaggatcttcg
 E  N  A  K  K  W  K  R  L  A  M  E  A  G  R  E  G  G  S  S
gataagaacatggaggcttttgtggaggatatttgtggagaatctcttattcaaaacttg
 D  K  N  M  E  A  F  V  E  D  I  C  G  E  S  L  I  Q  N  L
tgtgaagcagaggaggtaaaagtaaagtaa
 C  E  A  E  E  V  K  V  K  -
```

Figure 7

UGT75B2

```
atggcgcaaccgcattttctactggtaacgtttccggcgcaaggtcacgtgaacccatct
 M  A  Q  P  H  F  L  L  V  T  F  P  A  Q  G  H  V  N  P  S
ctccgttttgctcgtcggctcatcaaaacaactggcgcacgtgtaactttcgccacgtgt
 L  R  F  A  R  R  L  I  K  T  T  G  A  R  V  T  F  A  T  C
ctctctgtcattcaccgctctatgatcccaaaccacaacaacgtcgaaaatctctctttc
 L  S  V  I  H  R  S  M  I  P  N  H  N  N  V  E  N  L  S  F
cttactttctccgacggattcgacgacggagtcatctccaacaccgacgacgtccaaaac
 L  T  F  S  D  G  F  D  D  G  V  I  S  N  T  D  D  V  Q  N
cggttggtacacttcgaacgtaatggcgataaagctctatcggatttcatcgaagctaat
 R  L  V  H  F  E  R  N  G  D  K  A  L  S  D  F  I  E  A  N
cagaatggtgactctcccgtaagttgcttgatctacacgattcttcccaactgggttcca
 Q  N  G  D  S  P  V  S  C  L  I  Y  T  I  L  P  N  W  V  P
aaagtggcgcgtagatttcatcttccctctgttcatctctggatccaaccagccttcgct
 K  V  A  R  R  F  H  L  P  S  V  H  L  W  I  Q  P  A  F  A
ttcgacatttattacaattactctacaggaaacaactccgttttcgagttcccgaatcta
 F  D  I  Y  Y  N  Y  S  T  G  N  N  S  V  F  E  F  P  N  L
ccttctctcgaaatccgcgatctgccttctttcctctcaccttccaacacgaacaaagcc
 P  S  L  E  I  R  D  L  P  S  F  L  S  P  S  N  T  N  K  A
gcacaagcagtatatcaagaactgatggattttctcaaagaagaatctaacccgaaaatt
 A  Q  A  V  Y  Q  E  L  M  D  F  L  K  E  E  S  N  P  K  I
ctcgtcaacacattcgattcgctggagccagagttcttaacagctattccgaatatagaa
 L  V  N  T  F  D  S  L  E  P  E  F  L  T  A  I  P  N  I  E
atggtggcagttggtcctttacttcctgcggagattttcactggaagcgaatcaggtaaa
 M  V  A  V  G  P  L  L  P  A  E  I  F  T  G  S  E  S  G  K
gatttatcaagagatcatcaaagtagtagttatacactttggttagactcgaaaacagag
 D  L  S  R  D  H  Q  S  S  S  Y  T  L  W  L  D  S  K  T  E
tcctctgttatttatgtttcttttggaacaatggttgagttgtcgaagaaacagatagag
 S  S  V  I  Y  V  S  F  G  T  M  V  E  L  S  K  K  Q  I  E
gaactagcgagagcactcatagaaggggggaagaccgttcttgtgggttataactgataaa
 E  L  A  R  A  L  I  E  G  G  R  P  F  L  W  V  I  T  D  K
ctcaacagagaagcgaaaatagaaggagaagaagagacagagattgagaagatagctggt
 L  N  R  E  A  K  I  E  G  E  E  E  T  E  I  E  K  I  A  G
tttagacacgagcttgaagaggttgggatgattgtctcgtggtgttcgcagatagaggtt
 F  R  H  E  L  E  E  V  G  M  I  V  S  W  C  S  Q  I  E  V
ttgagacaccgagccataggttgtttttgactcattgtgggtggagctcatcactggag
 L  R  H  R  A  I  G  C  F  L  T  H  C  G  W  S  S  S  L  E
agtttggttctcggcgttccagtggtggcgtttccgatgtggtcggatcagccagcaaat
 S  L  V  L  G  V  P  V  V  A  F  P  M  W  S  D  Q  P  A  N
gcgaagcttttggaagaaatatggaagacaggtgtgagggtgagagagaactcggaaggt
 A  K  L  L  E  E  I  W  K  T  G  V  R  V  R  E  N  S  E  G
ttagtagagagaggagagataatgcggtgtttggaagcagtgatggaggcgaaatcggtg
 L  V  E  R  G  E  I  M  R  C  L  E  A  V  M  E  A  K  S  V
gagctgagggaaaacgcagagaaatggaagcgtttagcgactgaagcgggtagagaagga
 E  L  R  E  N  A  E  K  W  K  R  L  A  T  E  A  G  R  E  G
ggatcttcggacaagaatgtggaagcttttgtgaagagtctgttttga
 G  S  S  D  K  N  V  E  A  F  V  K  S  L  F  -
```

Figure 8

UGT75D1

```
atggccaacaacaattccaactctcccaccggtccacactttctattcgtaacatttcca
 M  A  N  N  N  S  N  S  P  T  G  P  H  F  L  F  V  T  F  P
gcccaaggtcacatcaacccatctctcgagctagccaaacgcctcgccggaacaatctct
 A  Q  G  H  I  N  P  S  L  E  L  A  K  R  L  A  G  T  I  S
ggtgctcgagtcaccttcgccgcctcaatctctgcctacaaccgccgcatgttctctaca
 G  A  R  V  T  F  A  A  S  I  S  A  Y  N  R  R  M  F  S  T
gaaaacgtccccgaaaccctaatcttcgctacctactccgatggccacgacgacggtttc
 E  N  V  P  E  T  L  I  F  A  T  Y  S  D  G  H  D  D  G  F
aaatcctctgcttactccgacaaatctcgtcaagacgccactggaaacttcatgtctgag
 K  S  S  A  Y  S  D  K  S  R  Q  D  A  T  G  N  F  M  S  E
atgagacgacgtggcaaagagacactaaccgaactaatcgaagataaccggaaacaaaac
 M  R  R  R  G  K  E  T  L  T  E  L  I  E  D  N  R  K  Q  N
aggccttttacttgcgtggtttacacgattctcctcacttgggtcgctgagctagcgcgt
 R  P  F  T  C  V  V  Y  T  I  L  L  T  W  V  A  E  L  A  R
gagtttcatcttccttctgctcttctttgggtccaaccagtaacagtcttctccattttt
 E  F  H  L  P  S  A  L  L  W  V  Q  P  V  T  V  F  S  I  F
taccattacttcaatggctacgaagatgcaatctcagagatggctaatacccctctagt
 Y  H  Y  F  N  G  Y  E  D  A  I  S  E  M  A  N  T  P  S  S
tctattaaattaccttctctgccactgcttactgtccgtgatattccttctttcattgtc
 S  I  K  L  P  S  L  P  L  L  T  V  R  D  I  P  S  F  I  V
tcttccaatgtctacgcgtttcttctacccgcgtttcgagaacagattgattcactgaag
 S  S  N  V  Y  A  F  L  L  P  A  F  R  E  Q  I  D  S  L  K
gaagaaataaaccctaagatcctcatcaacactttccaagagcttgagccagaagccatg
 E  E  I  N  P  K  I  L  I  N  T  F  Q  E  L  E  P  E  A  M
agctcggttccagataatttcaagattgtccctgtcggtccgttactaacgttgagaacg
 S  S  V  P  D  N  F  K  I  V  P  V  G  P  L  L  T  L  R  T
gatttttcgagtcgcggtgaatacatagagtggttggatactaaagcggattcgtctgtg
 D  F  S  S  R  G  E  Y  I  E  W  L  D  T  K  A  D  S  S  V
ctttatgtttcgttcgggacgcttgccgtgttgagcaagaaacagcttgtggagctttgt
 L  Y  V  S  F  G  T  L  A  V  L  S  K  K  Q  L  V  E  L  C
aaagcgttgatacaaagtcggagaccattcttgtgggtgattacggataagtcgtacaga
 K  A  L  I  Q  S  R  R  P  F  L  W  V  I  T  D  K  S  Y  R
aataaagaagatgagcaagagaaggaagaagattgcataagtagtttcagagaagagctc
 N  K  E  D  E  Q  E  K  E  E  D  C  I  S  S  F  R  E  E  L
gatgagataggaatggtggtttcatggtgtgatcagtttagggttttgaatcatagatcg
 D  E  I  G  M  V  V  S  W  C  D  Q  F  R  V  L  N  H  R  S
ataggttgtttcgtgacgcattgcgggtggaactctacgctggagagcttggtttcagga
 I  G  C  F  V  T  H  C  G  W  N  S  T  L  E  S  L  V  S  G
gttccggtggtggcgtttccgcaatggaatgatcagatgatgaacgcgaagcttttagaa
 V  P  V  V  A  F  P  Q  W  N  D  Q  M  M  N  A  K  L  L  E
gattgttggaaaacaggtgtaagagtgatggagaagaaggaagaagaaggagttgtggtg
 D  C  W  K  T  G  V  R  V  M  E  K  K  E  E  E  G  V  V  V
gtggatagtgaggagatacggcggtgcattgaggaagttatggaagacaaggcggaggag
 V  D  S  E  E  I  R  R  C  I  E  E  V  M  E  D  K  A  E  E
tttagaggaaatgccacgaggtggaaggatttagcggcggaggctgtgagagaaggaggc
 F  R  G  N  A  T  R  W  K  D  L  A  A  E  A  V  R  E  G  G
tcttcctttaatcatctcaaagcttttgtcgatgagcacatgtga
 S  S  F  N  H  L  K  A  F  V  D  E  H  M  -
```

Figure 9

UGT84A3

```
atggacccgtctcgtcatactcatgtgatgctcgtatctttccccggccaaggtcacgta
 M  D  P  S  R  H  T  H  V  M  L  V  S  F  P  G  Q  G  H  V
aaccctctacttcgtctcggaaagctcatagcctctaaaggcttactcgtcacctttgtc
 N  P  L  L  R  L  G  K  L  I  A  S  K  G  L  L  V  T  F  V
accacagagaagccatggggcaagaagatgcgtcaagccaacaagattcaagacggtgtg
 T  T  E  K  P  W  G  K  K  M  R  Q  A  N  K  I  Q  D  G  V
ctcaaaccggtcggtctaggtttcatccggtttgagttcttctctgacggcttcgccgac
 L  K  P  V  G  L  G  F  I  R  F  E  F  F  S  D  G  F  A  D
gacgatgaaaaaagattcgacttcgatgccttccgaccacaccttgaagctgtcggaaaa
 D  D  E  K  R  F  D  F  D  A  F  R  P  H  L  E  A  V  G  K
caagagatcaagaatctcgttaagagatataacaaggagccggtgacgtgtctcataaac
 Q  E  I  K  N  L  V  K  R  Y  N  K  E  P  V  T  C  L  I  N
aacgcttttgtcccatgggtatgtgatgtcgccgaggagcttcacatcccttcggctgtt
 N  A  F  V  P  W  V  C  D  V  A  E  E  L  H  I  P  S  A  V
ctatgggtccagtcttgtgcttgtctcacggcttattactattaccaccaccggttagtt
 L  W  V  Q  S  C  A  C  L  T  A  Y  Y  Y  Y  H  H  R  L  V
aagttcccgaccaaaaccgagccggacatcagcgttgaaatcccttgcttgccattgtta
 K  F  P  T  K  T  E  P  D  I  S  V  E  I  P  C  L  P  L  L
aagcatgacgagatcccaagctttcttcacccttcgtctccgtatacagcttttggagat
 K  H  D  E  I  P  S  F  L  H  P  S  S  P  Y  T  A  F  G  D
atcattttagaccagttaaagagattcgaaaaccacaagtctttctatcttttcatcgac
 I  I  L  D  Q  L  K  R  F  E  N  H  K  S  F  Y  L  F  I  D
acttttcgcgaactagaaaaagacatcatggaccacatgtcacaactttgtcctcaagcc
 T  F  R  E  L  E  K  D  I  M  D  H  M  S  Q  L  C  P  Q  A
atcatcagtcctgtcggtccgctcttcaagatggctcaaaccttgagttctgacgttaag
 I  I  S  P  V  G  P  L  F  K  M  A  Q  T  L  S  S  D  V  K
ggagatatatccgagccagcgagtgactgcatggaatggcttgactcaagagaaccatcc
 G  D  I  S  E  P  A  S  D  C  M  E  W  L  D  S  R  E  P  S
tcagtcgtttacatctcctttgggactatagccaacttgaagcaagagcagatggaggag
 S  V  V  Y  I  S  F  G  T  I  A  N  L  K  Q  E  Q  M  E  E
atcgctcatggcgttttgagctctggcttgtcggtcttatgggtggttcggcctcccatg
 I  A  H  G  V  L  S  S  G  L  S  V  L  W  V  V  R  P  P  M
gaagggacatttgtagaaccacatgttttgcctcgagagctcgaagaaaagggtaaaatc
 E  G  T  F  V  E  P  H  V  L  P  R  E  L  E  E  K  G  K  I
gtggaatggtgtccccaagagagagtcttggctcatcctgcgattgcttgtttcttaagt
 V  E  W  C  P  Q  E  R  V  L  A  H  P  A  I  A  C  F  L  S
cactgcggatggaactcgacaatggaggctttaactgccggagtccccgttgtttgtttt
 H  C  G  W  N  S  T  M  E  A  L  T  A  G  V  P  V  V  C  F
ccgcaatggggagatcaagtgactgatgcggtgtacttggctgatgttttcaagacagga
 P  Q  W  G  D  Q  V  T  D  A  V  Y  L  A  D  V  F  K  T  G
gtgagactaggccgcggagccgctgaggagatgattgtttcgagggaggttgtagcagag
 V  R  L  G  R  G  A  A  E  E  M  I  V  S  R  E  V  V  A  E
aagctgcttgaggccacagttggggaaaaggcggtggagctgagagaaaacgctcggagg
 K  L  L  E  A  T  V  G  E  K  A  V  E  L  R  E  N  A  R  R
tggaaggcggaggccgaggccgccgtggcggacggtggatcatctgatatgaactttaaa
 W  K  A  E  A  E  A  A  V  A  D  G  G  S  S  D  M  N  F  K
gagtttgtggacaagttggttacgaaacatgtgacgagagaagacaacggagaacactag
 E  F  V  D  K  L  V  T  K  H  V  T  R  E  D  N  G  E  H  -
```

Figure 10

UGT84B1

atgggcagtagtgagggtcaagaaacacatgtcctaatggtaacactaccattccaaggt
M G S S E G Q E T H V L M V T L P F Q G
cacatcaatccaatgctcaaactcgcaaaacatctctcgttatcatcaaagaacctacac
H I N P M L K L A K H L S L S S K N L H
atcaatctcgccactattgagtcagcccgtgatctcctctccaccgtagaaaaacctcgt
I N L A T I E S A R D L L S T V E K P R
tatccggtggacctcgtgttcttctccgatggtctacctaaagaagatccaaaggcccct
Y P V D L V F F S D G L P K E D P K A P
gaaactcttttgaagtcattgaataaagtcggagccatgaacttgtctaaaatcatcgaa
E T L L K S L N K V G A M N L S K I I E
gaaaagagatactcttgtatcatctcttcgccttttactccatgggttccagctgttgca
E K R Y S C I I S S P F T P W V P A V A
gcctctcataacatctcttgtgcaatactttggatccaagcttgtggagcttactcggtt
A S H N I S C A I L W I Q A C G A Y S V
tattaccgttactacatgaagacaaactctttccctgatcttgaagatctgaatcaaacg
Y Y R Y Y M K T N S F P D L E D L N Q T
gtggagttaccagctttaccattgttggaagttcgagatcttccatcgtttatgttacct
V E L P A L P L L E V R D L P S F M L P
tctggtggtgctcacttctataatctaatggcggaatttgcagattgtttgaggtatgtg
S G G A H F Y N L M A E F A D C L R Y V
aaatgggttttggttaattcattctatgaactcgaatcagagataatcgaatcgatggct
K W V L V N S F Y E L E S E I I E S M A
gatttaaaacctgtaattccaattggtcctctggtttctccatttctgttgggcgatggt
D L K P V I P I G P L V S P F L L G D G
gaggaggaaaccctagacggtaaaaacctagattttgtaaatctgatgattgttgtatg
E E E T L D G K N L D F C K S D D C C M
gagtggcttgacaagcaagctaggtcttctgttgtgtacatatctttcggaagtatgctc
E W L D K Q A R S S V V Y I S F G S M L
gaaacattggagaatcaggtcgagaccatagcgaaggcgctgaagaacagaggacttcca
E T L E N Q V E T I A K A L K N R G L P
tttctttgggtgataaggccaaaggagaaagcccaaaacgttgctgttttgcaggagatg
F L W V I R P K E K A Q N V A V L Q E M
gtgaaagaaggacaaggggttgttctcgagtggagtccacaagagaagattttgagccac
V K E G Q G V V L E W S P Q E K I L S H
gaggcaatctcttgttttgtcacgcattgcggctggaactcgactatggagacggtggtg
E A I S C F V T H C G W N S T M E T V V
gctggtgttcctgtggtagcgtaccctagctggacggatcagcccattgacgcgcggttg
A G V P V V A Y P S W T D Q P I D A R L
cttgttgatgtgtttggaatcggagtaaggatgaggaatgacagtgtcgatggcgagctt
L V D V F G I G V R M R N D S V D G E L
aaggtcgaagaagtagaaagatgcattgaggccgtgacggagggacccgctgccgtggat
K V E E V E R C I E A V T E G P A A V D
ataagaaggagagcggcggagctaaagcgcgtggcgagattggcgttggcacctggtgga
I R R R A A E L K R V A R L A L A P G G
tcttcgacacggaatttagacttgttcattagtgatatcacaatcgcctaa
S S T R N L D L F I S D I T I A -

Figure 11

UGT84B2

```
atgggaagtaatgagggtcaagaaacacatgtcctaatggtagcattagcattccaaggt
 M  G  S  N  E  G  Q  E  T  H  V  L  M  V  A  L  A  F  Q  G
catctcaatccaatgctcaaattcgcaaaacatctcgcacgaaccaatctacacttcact
 H  L  N  P  M  L  K  F  A  K  H  L  A  R  T  N  L  H  F  T
ctcgccaccactgagcaagcccgtgacctcctcttccaccgctgacgaacctcataga
 L  A  T  T  E  Q  A  R  D  L  L  S  S  T  A  D  E  P  H  R
ccggtggacctcgctttcttctcagacggtctacctaaagacgatccaagagatcccgac
 P  V  D  L  A  F  F  S  D  G  L  P  K  D  D  P  R  D  P  D
actctcgcaaagtcattgaaaaaagatggagccaagaacttgtcaaaaatcatcgaagaa
 T  L  A  K  S  L  K  K  D  G  A  K  N  L  S  K  I  I  E  E
aagagatttgattgcatcatctctgtgccttttactccctgggttccagctgttgcagct
 K  R  F  D  C  I  I  S  V  P  F  T  P  W  V  P  A  V  A  A
gcacataacattccttgtgcaatcctctggatccaagcttgtggagctttttctgtttat
 A  H  N  I  P  C  A  I  L  W  I  Q  A  C  G  A  F  S  V  Y
taccgttattacatgaagacaaatcctttccccgaccttgaagatctgaatcaaacagtg
 Y  R  Y  Y  M  K  T  N  P  F  P  D  L  E  D  L  N  Q  T  V
gagttaccagctttaccattgttggaagtccgagatctcccgtcattgatgttaccttct
 E  L  P  A  L  P  L  L  E  V  R  D  L  P  S  L  M  L  P  S
caaggagctaatgtcaataccctaatggcggaatttgcagattgtttgaaagatgtgaaa
 Q  G  A  N  V  N  T  L  M  A  E  F  A  D  C  L  K  D  V  K
tgggttttggttaactcgttttacgaactcgaatcagagatcatcgagtctatgtctgat
 W  V  L  V  N  S  F  Y  E  L  E  S  E  I  I  E  S  M  S  D
ttaaaacctataatcccaattggtcctcttgtttctccattcctgttgggaaatgatgaa
 L  K  P  I  I  P  I  G  P  L  V  S  P  F  L  L  G  N  D  E
gaaaaaaccctagatatgtggaaagttgatgattattgtatggagtggcttgacaagcaa
 E  K  T  L  D  M  W  K  V  D  D  Y  C  M  E  W  L  D  K  Q
gctaggtcttcagttgtttacatatctttcggaagcatactcaaatcattggagaatcaa
 A  R  S  S  V  V  Y  I  S  F  G  S  I  L  K  S  L  E  N  Q
gttgagaccatagcaacggcattaaaaaacagaggagttccatttctttgggtgatacgg
 V  E  T  I  A  T  A  L  K  N  R  G  V  P  F  L  W  V  I  R
ccgaaggagaaaggcgaaaacgtccaggttttgcaggagatggttaaagaaggtaaaggg
 P  K  E  K  G  E  N  V  Q  V  L  Q  E  M  V  K  E  G  K  G
gttgtaactgaatggggtcaacaagaaaagatattgagccacatggcgatttcttgcttc
 V  V  T  E  W  G  Q  Q  E  K  I  L  S  H  M  A  I  S  C  F
atcacgcattgtggatggaactcgacgatcgagacggtggtgactggtgttcccgtggtg
 I  T  H  C  G  W  N  S  T  I  E  T  V  V  T  G  V  P  V  V
gcgtatccgacttggatagatcagccgcttgatgcgagactgcttgtggatgtgtttgga
 A  Y  P  T  W  I  D  Q  P  L  D  A  R  L  L  V  D  V  F  G
atcggagtaaggatgaagaacgacgctatcgatggagagcttaaggttgcagaggtggag
 I  G  V  R  M  K  N  D  A  I  D  G  E  L  K  V  A  E  V  E
agatgcattgaggccgtgacagagggacctgccgccgcggatatgaggaggagagcgacg
 R  C  I  E  A  V  T  E  G  P  A  A  A  D  M  R  R  R  A  T
gagctgaagcacgccgcaagatcggcgatgtcacctggtggatcttccgctcagaattta
 E  L  K  H  A  A  R  S  A  M  S  P  G  G  S  S  A  Q  N  L
gactcgttcattagtgatatcccaatcacttga
 D  S  F  I  S  D  I  P  I  T  -
```

Figure 12

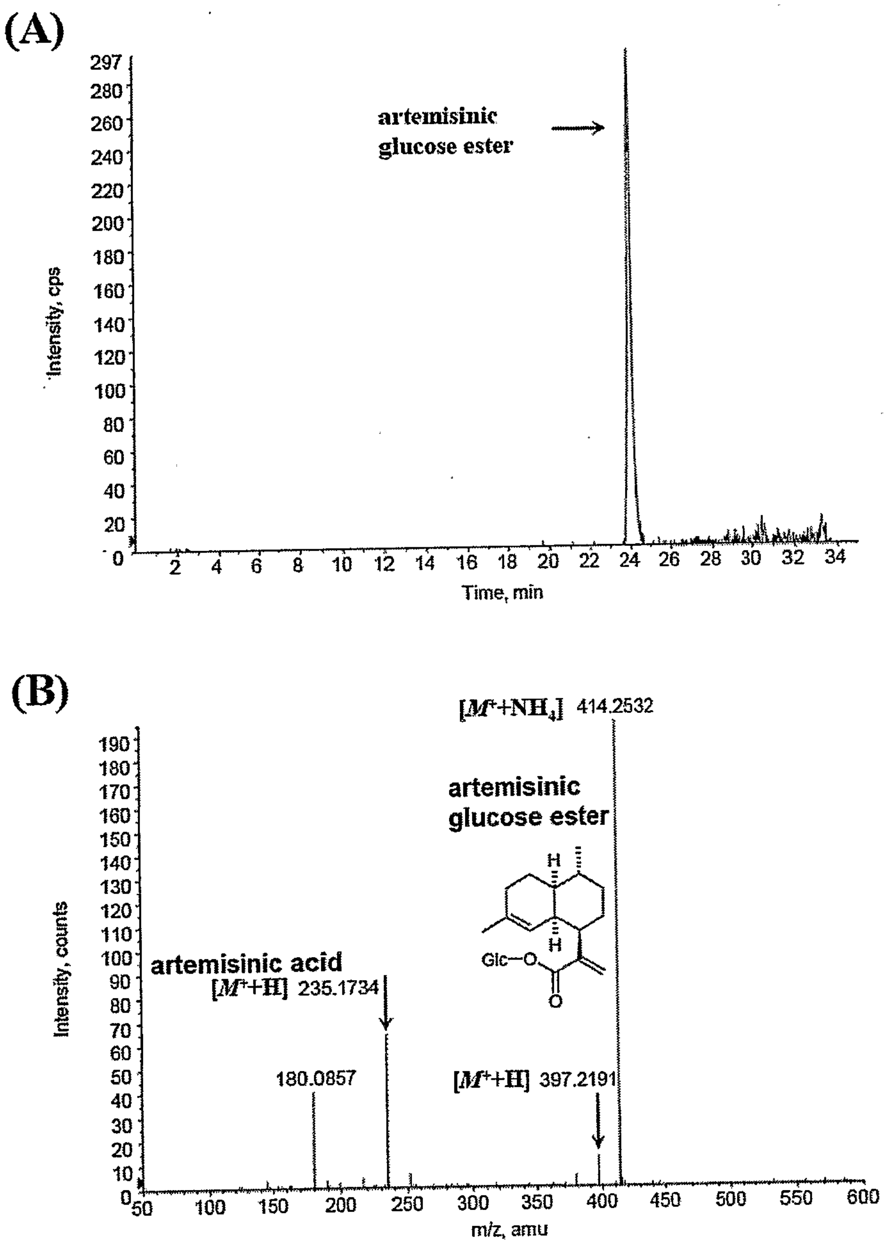

(A)
artemisinic glucose ester
Intensity, cps
Time, min
(B)
[M⁺+NH₄] 414.2532
artemisinic glucose ester
Glc—O
artemisinic acid
[M⁺+H] 235.1734
180.0857
[M⁺+H] 397.2191
Intensity, counts
m/z, amu

Figure 13

Figure 14

Amorphadiene synthase

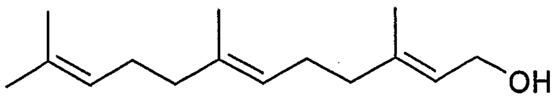
```
   1 ATGTCCCTTA CAGAAGAAAA ACCTATTCGC CCCATTGCCA ACTTTCCTCC AAGCATTTGG
  61 GGAGATCAGT CTCTCATCTA TGAAAAGCAA GTAGAGCAAG GGGTGGAACA GATAGTGAAT
 121 GATTTAAAAA AAGAAGTGCG GCAACTACTA AAAGAAGCTT TGGATATTCC TATGAAACAT
 181 GCCAATTTGT TGAAGCTGAT TGATGAAATC CAACGCCTTG GAATACCGTA TCACTTTGAA
 241 CGGGAGATTG ATCATGCATT GCAATGTATT TATGAAACAT ATGGTGATAA CTGGAATGGT
 301 GACCGCTCTT CCTTATGGTT CCGTCTTATG CGAAAGCAAG GATATTATGT TACATGTGAT
 361 GTTTTCAATA ACTATAAAGA CAAAAATGGA GCGTTCAAGC AATCGTTAGC TAATGATGTT
 421 GAAGGTTTGC TTGAGTTGTA CGAAGCAACT TCTATGAGGG TACCTGGGGA GACTATATTA
 481 GAAGATGCTC TTGGTTTTAC ACGATCTCGT CTTAGCATTA TGACAAAAGA TGCTTTTTCT
 541 ACAAACCCCG CTCTTTTTAC CGAAATACAA CGGGCACTAA AGCAACCCCT TTGGAAAAGG
 601 TTGCCAAGAA TAGAGGCGGC GCAGTACATT CCTTTCTATC AACAACAAGA TTCTCATAAC
 661 AAGACTTTAC TTAAACTTGC TAAGTTAGAG TTCAATTTGC TTCAGTCATT GCACAAGGAA
 721 GAGCTCAGCC ATGTGTGCAA ATGGTGGAAA GCTTTCGATA TCAAGAAGAA CGCACCTTGT
 781 TTAAGAGATA GAATTGTTGA ATGCTACTTT TGGGGACTAG GTTCAGGCTA TGAGCCACAG
 841 TATTCCCGGG CTAGAGTTTT CTTCACAAAA GCTGTTGCTG TTATAACTCT TATAGATGAC
 901 ACTTATGATG CGTATGGTAC TTATGAAGAA CTTAAGATCT TTACTGAAGC TGTTGAAAGG
 961 TGGTCAATTA CATGCTTAGA CACACTTCCA GAATACATGA AACCGATATA CAAATTATTC
1021 ATGGATACAT ACACAGAAAT GGAAGAATTT CTTGCAAAGG AGGGAAGAAC AGATCTATTT
1081 AACTGCGGCA AAGAATTTGT GAAAGAGTTT GTTAGAAACC TGATGGTTGA AGCAAAATGG
1141 GCAAATGAGG GACACATACC AACCACTGAA GAGCATGATC CAGTTGTAAT CATTACTGGC
1201 GGTGCTAACC TGCTTACAAC AACTTGTTAT CTTGACATGA GTGATATATT CACAAAAGAG
1261 TCTGTCGAAT GGGCTGTCTC TGCACCTCCT CTTTTTAGAT ACTCAGGTAT ACTTGGTCGA
1321 CGCCTAGATG ATCTCATGAC CCACAAGGCC GAGCAAGAAA GAAAACATAG TTCATCGAGC
1381 CTTGAAAGTT ATATGAAGGA ATATAATGTC AATGAGGAGT ATGCCCAAAC CTTGATTTAC
1441 AAGGAAGTAG AAGATGTGTG GAAAGATATA AACCGAGAGT ACCTCACAAC TAAAAACATT
1501 CCAAGGCCGT TATTGATGGC TGTGATCTAT TTGTGCCAGT TTCTTGAAGT TCAATATGCA
1561 GGAAAGGATA ACTTCACACG TATGGGAGAC GAATACAAAC ATCTCATAAA GTCTCTACTC
1621 GTTTATCCTA TGAGTATATG A
```

Figure 15

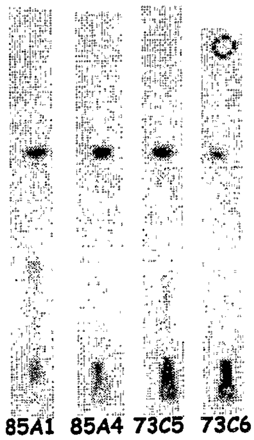
Cytochrome P450 CYP71AV1

```
     ATGAAGAGTA TACTAAAAGC AATGGCTCTC TCACTGACCA CTTCCATTGC TCTTGCAACG
  61 ATCCTTTTGT TCGTTTACAA GTTCGCTACT CGTTCCAAGT CCACCAAAAA AAGCCTTCCT
 121 GAGCCATGGC GACTTCCCAT TATTGGTCAC ATGCATCACT TGATTGGTAC AACGCCACAT
 181 CGTGGGGTTA GGGATTTAGC CAGAAAGTAT GGATCTTTGA TGCATTTACA GCTTGGTGAA
 241 GTTCCAACAA TTGTGGTGTC ATCTCCGAAA TGGGCTAAAG AGGTTTTGAC AACGTACGAC
 301 ATTACCTTTG CTAACAGGCC CGAGACTTTA ACTGGTGAGA TTGTTTTATA TCACAATACG
 361 GATGTTGTTC TTGCACCTTA TGGTGAGTAC TGGAGGCAAT TACGTAAAAT TTGCACATTG
 421 GAGCTTTTGA GTGTTAAGAA AGTAAAGTCA TTTCAGTCGC TTCGTGAAGA GGAGTGTTGG
 481 AATTTGGTTC AAGAGATTAA AGCTTCAGGT TCAGGGAGAC CGGTTAACCT TTCAGAGAAT
 541 GTTTTCAAGT TGATTGCAAC GATACTTAGT AGAGCCGCAT TTGGGAAAGG GATCAAGGAC
 601 CAGAAAGAGT TAACGGAGAT TGTGAAAGAG ATACTGAGGC AAACTGGTGG TTTTGATGTG
 661 GCAGATATCT TTCCTTCAAA GAAATTTCTT CATCATCTTT CGGGCAAGAG AGCTCGGTTA
 721 ACTAGCCTTC GCAAAAAGAT CGATAATTTA ATCGATAACC TTGTAGCTGA GCATACTGTT
 781 AACACCTCCA GTAAAACTAA CGAGACACTC CTCGATGTTC TTTTAAGGCT CAAAGACAGT
 841 GCTGAATTCC CATTAACATC TGATAACATT AAAGCCATCA TTTTGGATAT GTTTGGAGCA
 901 GGCACAGACA CTTCCTCATC CACAATCGAA TGGGCGATTC CGGAACTCAT AAAGTGTCCG
 961 AAAGCAATGG AGAAAGTACA AGCGGAATTG AGGAAAGCAT TGAACGGAAA AGAAAAGATC
1021 CATGAGGAAG ACATTCAAGA ACTAAGCTAC TTGAACATGG TAATCAAAGA AACATTGAGG
1081 TTGCACCCTC CACTACCCTT GGTTCTGCCA AGAGAGTGCC GCCAACCAGT CAATTTGGCT
1141 GGATACAACA TACCCAATAA GACCAAACTT ATTGTCAACG TCTTTGCGAT AAATAGGGAC
1201 CCTGAATATT GGAAAGACGC TGAAGCTTTC ATCCCTGAAC GATTTGAAAA TAGTTCTGCA
1261 ACTGTCATGG GTGCAGAATA CGAGTATCTT CCGTTTGGAG CTGGGAGAAG GATGTGTCCT
1321 GGAGCCGCAC TTGGTTTAGC TAACGTGCAG CTCCCGCTCG CTAATATACT ATATCATTTC
1381 AACTGGAAAC TCCCCAATGG TGTGAGCTAT GACCAGATCG ACATGACCGA GAGCTCTGGA
1441 GCCACGATGC AAAGAAAGGC TGAGTTGTTA CTCGTTCCAA GTTTCTAG
```

(A)
Farnesol



(B)

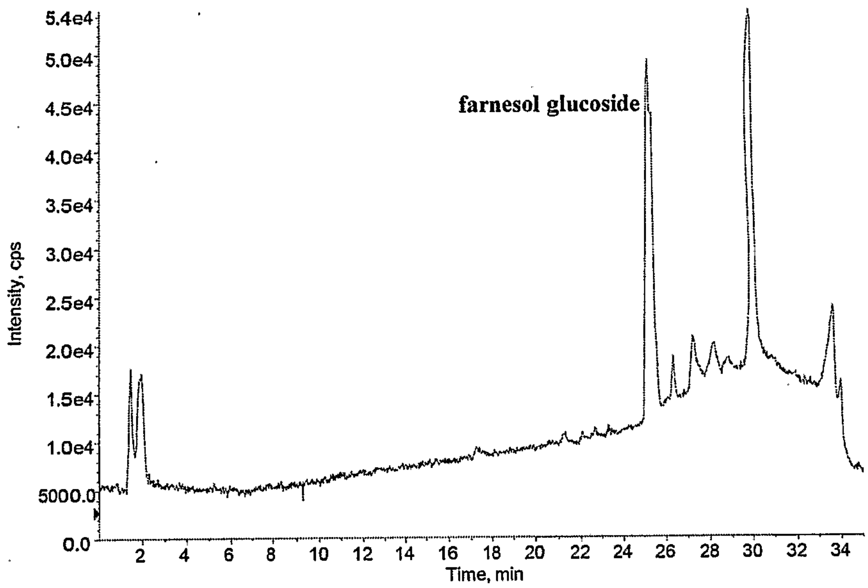


(C)

GT	Relative activity (%)
73C3	5.94
73C5	32.60
73C6	15.24
85A1	29.60
85A2	28.72
85A4	38.21
85A7	20.94

Figure 16

(A)


farnesol glucoside
Intensity, cps
5.4e4
5.0e4
4.5e4
4.0e4
3.5e4
3.0e4
2.5e4
2.0e4
1.5e4
1.0e4
5000.0
0.0
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
Time, min

(B)

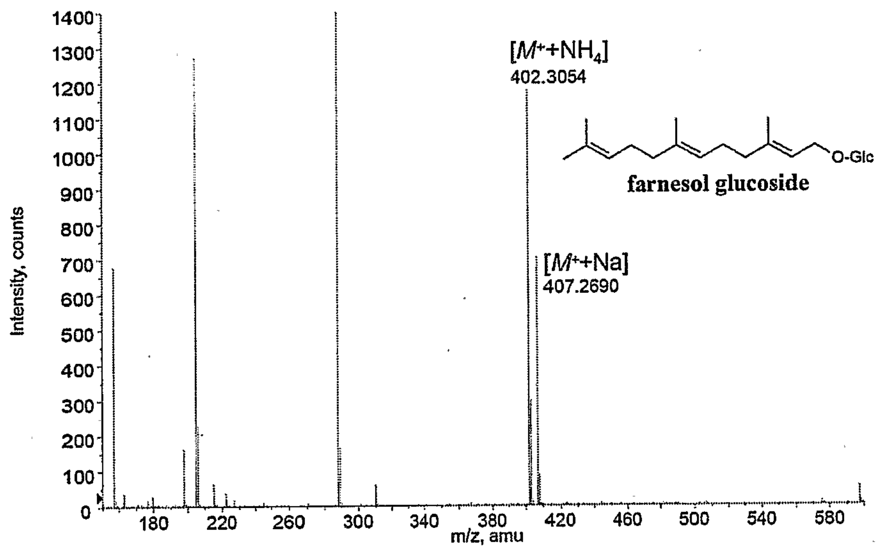

[M+·+NH4]
402.3054
[M+·+Na]
407.2690
O-Glc
farnesol glucoside
Intensity, counts
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
180 220 260 300 340 380 420 460 500 540 580
m/z, amu

Figure 17

UGT73C3

```
atggctacggaaaaaacccaccaatttcatccttctcttcactttgtcctcttccctttc
 M  A  T  E  K  T  H  Q  F  H  P  S  L  H  F  V  L  F  P  F
atggctcaaggccacatgattcccatgattgatattgcaagactcttggctcagcgtggt
 M  A  Q  G  H  M  I  P  M  I  D  I  A  R  L  L  A  Q  R  G
gtgaccataacaattgtcacgacacctcacaacgcagcaaggtttaagaatgtcctaaac
 V  T  I  T  I  V  T  T  P  H  N  A  A  R  F  K  N  V  L  N
cgagcgatcgagtctggcttggccatcaacatactgcatgtgaagtttccatatcaagag
 R  A  I  E  S  G  L  A  I  N  I  L  H  V  K  F  P  Y  Q  E
tttggtttgccagaaggaaaagagaatatagattcgttagactcaacggagttgatggta
 F  G  L  P  E  G  K  E  N  I  D  S  L  D  S  T  E  L  M  V
cctttcttcaaagcggtgaacttgcttgaagatccggtcatgaagctcatggaagagatg
 P  F  F  K  A  V  N  L  L  E  D  P  V  M  K  L  M  E  E  M
aaacctagacctagctgtctaatttctgattggtgtttgccttatacaagcataatcgcc
 K  P  R  P  S  C  L  I  S  D  W  C  L  P  Y  T  S  I  I  A
aagaacttcaatataccaaagatagttttccacggcatgggttgctttaatcttttgtgt
 K  N  F  N  I  P  K  I  V  F  H  G  M  G  C  F  N  L  L  C
atgcatgttctacgcagaaacttagagatcctagagaatgtaaagtcggatgaagagtat
 M  H  V  L  R  R  N  L  E  I  L  E  N  V  K  S  D  E  E  Y
ttcttggttcctagttttcctgatagagttgaatttacaaagcttcaacttcctgtgaaa
 F  L  V  P  S  F  P  D  R  V  E  F  T  K  L  Q  L  P  V  K
gcaaatgcaagtggagattggaaagagataatggatgaaatggtaaaagcagaatacaca
 A  N  A  S  G  D  W  K  E  I  M  D  E  M  V  K  A  E  Y  T
tcctatggtgtgatcgtcaacacatttcaggagttggagccaccttatgtcaaagactac
 S  Y  G  V  I  V  N  T  F  Q  E  L  E  P  P  Y  V  K  D  Y
aaagaggcaatggatggaaaagtatggtccattggacccgtttccttgtgtaacaaggca
 K  E  A  M  D  G  K  V  W  S  I  G  P  V  S  L  C  N  K  A
ggtgcagacaaagctgagaggggaagcaaggccgccattgatcaagatgagtgtcttcaa
 G  A  D  K  A  E  R  G  S  K  A  A  I  D  Q  D  E  C  L  Q
tggcttgattctaaagaagaaggttcggtgctctatgtttgccttggaagtatatgtaat
 W  L  D  S  K  E  E  G  S  V  L  Y  V  C  L  G  S  I  C  N
cttcctttgtctcagctcaaggagctggggctaggccttgaggaatctcgaagatctttt
 L  P  L  S  Q  L  K  E  L  G  L  G  L  E  E  S  R  R  S  F
atttgggtcataagaggttcggaaaagtataaagaactatttgagtggatgttggagagc
 I  W  V  I  R  G  S  E  K  Y  K  E  L  F  E  W  M  L  E  S
ggttttgaagaaagaatcaaagagagaggacttctcattaaagggtgggcacctcaagtc
 G  F  E  E  R  I  K  E  R  G  L  L  I  K  G  W  A  P  Q  V
cttatcctttcacatccttccgttggaggattcctgacacactgtggatggaactcgact
 L  I  L  S  H  P  S  V  G  G  F  L  T  H  C  G  W  N  S  T
ctcgaaggaatcacctcaggcattccactgatcacttggccgctgtttggagaccaattc
 L  E  G  I  T  S  G  I  P  L  I  T  W  P  L  F  G  D  Q  F
tgcaaccaaaaactggtcgttcaagtactaaaagccggtgtaagtgccggggttgaagaa
 C  N  Q  K  L  V  V  Q  V  L  K  A  G  V  S  A  G  V  E  E
gtcatgaaatggggagaagaagataaaataggagtgttagtggataaagaaggagtgaaa
 V  M  K  W  G  E  E  D  K  I  G  V  L  V  D  K  E  G  V  K
aaggctgtggaagaattgatgggtgatagtgatgatgcaaaagagaggagaagaagagtc
 K  A  V  E  E  L  M  G  D  S  D  D  A  K  E  R  R  R  R  V
aaagagcttggagaattagctcacaaagctgtggaaaaaggaggctcttctcattctaac
 K  E  L  G  E  L  A  H  K  A  V  E  K  G  G  S  S  H  S  N
atcacactcttgctacaagacataatgcaactagcacaattcaagaattga
 I  T  L  L  L  Q  D  I  M  Q  L  A  Q  F  K  N  -
```

Figure 18

UGT73C5

```
atggtttccgaaacaaccaaatcttctccacttcactttgttctcttccctttcatggct
 M  V  S  E  T  T  K  S  S  P  L  H  F  V  L  F  P  F  M  A
caaggccacatgattcccatggttgatattgcaaggctcttggctcagcgtggtgtgatc
 Q  G  H  M  I  P  M  V  D  I  A  R  L  L  A  Q  R  G  V  I
ataacaattgtcacgacgcctcacaatgcagcgaggttcaagaatgtcctaaaccgtgcc
 I  T  I  V  T  T  P  H  N  A  A  R  F  K  N  V  L  N  R  A
attgagtctggcttgcccatcaacttagtgcaagtcaagtttccatatctagaagctggt
 I  E  S  G  L  P  I  N  L  V  Q  V  K  F  P  Y  L  E  A  G
ttgcaagaaggacaagagaatatcgattctcttgacacaatggagcggatgatacctttc
 L  Q  E  G  Q  E  N  I  D  S  L  D  T  M  E  R  M  I  P  F
tttaaagcggttaactttctcgaagaaccagtccagaagctcattgaagagatgaaccct
 F  K  A  V  N  F  L  E  E  P  V  Q  K  L  I  E  E  M  N  P
cgaccaagctgtctaatttctgattttgtttgccttatacaagcaaaatcgccaagaag
 R  P  S  C  L  I  S  D  F  C  L  P  Y  T  S  K  I  A  K  K
ttcaatatcccaaagatcctcttccatggcatgggttgcttttgtcttctgtgtatgcat
 F  N  I  P  K  I  L  F  H  G  M  G  C  F  C  L  L  C  M  H
gttttacgcaagaaccgtgagatcttggacaatttaaagtcagataaggagcttttcact
 V  L  R  K  N  R  E  I  L  D  N  L  K  S  D  K  E  L  F  T
gttcctgattttcctgatagagttgaattcacaagaacgcaagttccggtagaaacatat
 V  P  D  F  P  D  R  V  E  F  T  R  T  Q  V  P  V  E  T  Y
gttccagctggagactggaaagatatctttgatggtatggtagaagcgaatgagacatct
 V  P  A  G  D  W  K  D  I  F  D  G  M  V  E  A  N  E  T  S
tatggtgtgatcgtcaactcatttcaagagctcgagcctgcttatgccaaagactacaag
 Y  G  V  I  V  N  S  F  Q  E  L  E  P  A  Y  A  K  D  Y  K
gaggtaaggtccggtaaagcatggaccattggacccgtttccttgtgcaacaaggtagga
 E  V  R  S  G  K  A  W  T  I  G  P  V  S  L  C  N  K  V  G
gccgacaaagcagagaggggaaacaaatcagacattgatcaagatgagtgccttaaatgg
 A  D  K  A  E  R  G  N  K  S  D  I  D  Q  D  E  C  L  K  W
ctcgattctaagaaacatggctcggtgctttacgtttgtcttggaagtatctgtaatctt
 L  D  S  K  K  H  G  S  V  L  Y  V  C  L  G  S  I  C  N  L
cctttgtctcaactcaaggagctgggactaggcctagaggaatcccaaagacctttcatt
 P  L  S  Q  L  K  E  L  G  L  G  L  E  E  S  Q  R  P  F  I
tgggtcataagaggttgggagaagtacaaagagttagttgagtggttctcggaaagcggc
 W  V  I  R  G  W  E  K  Y  K  E  L  V  E  W  F  S  E  S  G
tttgaagatagaatccaagatagaggacttctcatcaaaggatggtcccctcaaatgctt
 F  E  D  R  I  Q  D  R  G  L  L  I  K  G  W  S  P  Q  M  L
atcctttcacatccatcagttggagggttcctaacacactgtggttggaactcgactctt
 I  L  S  H  P  S  V  G  G  F  L  T  H  C  G  W  N  S  T  L
gaggggataactgctggtctaccgctacttacatggccgctattcgcagaccaattctgc
 E  G  I  T  A  G  L  P  L  L  T  W  P  L  F  A  D  Q  F  C
aatgagaaattggtcgttgaggtactaaaagccggtgtaagatccggggttgaacagcct
 N  E  K  L  V  V  E  V  L  K  A  G  V  R  S  G  V  E  Q  P
atgaaatggggagaagaggagaaaataggagtgttggtggataaagaaggagtgaagaag
 M  K  W  G  E  E  E  K  I  G  V  L  V  D  K  E  G  V  K  K
gcagtggaagaattaatgggtgagagtgatgatgcaaaagagagaagaagaagagccaaa
 A  V  E  E  L  M  G  E  S  D  D  A  K  E  R  R  R  R  A  K
gagcttggagattcagctcacaaggctgtggaagaaggaggctcttctcattctaacatc
 E  L  G  D  S  A  H  K  A  V  E  E  G  G  S  S  H  S  N  I
tctttcttgctacaagacataatggaactggcagaacccaataattga
 S  F  L  L  Q  D  I  M  E  L  A  E  P  N  N  -
```

Figure 19

UGT73C6

```
atggctttcgaaaaaaacaacgaaccttttcctcttcactttgttctcttccctttcatg
 M  A  F  E  K  N  N  E  P  F  P  L  H  F  V  L  F  P  F  M
gctcaaggccacatgattcccatggttgatattgcaaggctcttggctcagcgaggtgtg
 A  Q  G  H  M  I  P  M  V  D  I  A  R  L  L  A  Q  R  G  V
cttataacaattgtcacgacgcctcacaatgcagcaaggttcaagaatgtcctaaaccgt
 L  I  T  I  V  T  T  P  H  N  A  A  R  F  K  N  V  L  N  R
gccattgagtctggtttgcccatcaacctagtgcaagtcaagtttccatatcaagaagct
 A  I  E  S  G  L  P  I  N  L  V  Q  V  K  F  P  Y  Q  E  A
ggtctgcaagaaggacaagaaaatatggatttgcttaccacgatggagcagataacatct
 G  L  Q  E  G  Q  E  N  M  D  L  L  T  T  M  E  Q  I  T  S
ttctttaaagcggttaacttactcaaagaaccagtccagaaccttattgaagagatgagc
 F  F  K  A  V  N  L  L  K  E  P  V  Q  N  L  I  E  E  M  S
ccgcgaccaagctgtctaatctctgatatgtgtttgtcgtatacaagcgaaatcgccaag
 P  R  P  S  C  L  I  S  D  M  C  L  S  Y  T  S  E  I  A  K
aagttcaaaataccaaagatcctcttccatggcatgggttgcttttgtcttctgtgtgtt
 K  F  K  I  P  K  I  L  F  H  G  M  G  C  F  C  L  L  C  V
aacgttctgcgcaagaaccgtgagatcttggacaatttaaagtctgataaggagtacttc
 N  V  L  R  K  N  R  E  I  L  D  N  L  K  S  D  K  E  Y  F
attgttccttattttcctgatagagttgaattcacaagacctcaagttccggtggaaaca
 I  V  P  Y  F  P  D  R  V  E  F  T  R  P  Q  V  P  V  E  T
tatgttcctgcaggctggaaagagatcttggaggatatggtagaagcggataagacatct
 Y  V  P  A  G  W  K  E  I  L  E  D  M  V  E  A  D  K  T  S
tatggtgttatagtcaactcatttcaagagctcgaacctgcgtatgccaaagacttcaag
 Y  G  V  I  V  N  S  F  Q  E  L  E  P  A  Y  A  K  D  F  K
gaggcaaggtctggtaaagcatggaccattggacctgtttccttgtgcaacaaggtagga
 E  A  R  S  G  K  A  W  T  I  G  P  V  S  L  C  N  K  V  G
gtagacaaagcagagaggggaaacaaatcagatattgatcaagatgagtgccttgaatgg
 V  D  K  A  E  R  G  N  K  S  D  I  D  Q  D  E  C  L  E  W
ctcgattctaaggaaccgggatctgtgctctacgtttgccttggaagtatttgtaatctt
 L  D  S  K  E  P  G  S  V  L  Y  V  C  L  G  S  I  C  N  L
cctctgtctcagctccttgagctgggactaggcctagaggaatcccaaagacctttcatc
 P  L  S  Q  L  L  E  L  G  L  G  L  E  E  S  Q  R  P  F  I
tgggtcataagaggttgggagaaatacaaagagttagttgagtggttctcggaaagcggc
 W  V  I  R  G  W  E  K  Y  K  E  L  V  E  W  F  S  E  S  G
tttgaagatagaatccaagatagaggacttctcatcaaaggatggtcccctcaaatgctt
 F  E  D  R  I  Q  D  R  G  L  L  I  K  G  W  S  P  Q  M  L
atcctttcacatccttctgttggagggttcttaacgcactgcggatggaactcgactctt
 I  L  S  H  P  S  V  G  G  F  L  T  H  C  G  W  N  S  T  L
gaggggataactgctggtctaccaatgcttacatggccactatttgcagaccaattctgc
 E  G  I  T  A  G  L  P  M  L  T  W  P  L  F  A  D  Q  F  C
aacgagaaactggtcgtacaaatactaaaagtcggtgtaagtgccgaggttaaagaggtc
 N  E  K  L  V  V  Q  I  L  K  V  G  V  S  A  E  V  K  E  V
atgaaatggggagaagaagagaagataggagtgttggtggataaagaaggagtgaagaag
 M  K  W  G  E  E  E  K  I  G  V  L  V  D  K  E  G  V  K  K
gcagtggaagaactaatgggtgagagtgatgatgcaaaagagagaagaagaagagccaaa
 A  V  E  E  L  M  G  E  S  D  D  A  K  E  R  R  R  R  A  K
gagcttggagagaatcagctcacaaggctgtggaagaaggaggctcctctcattctaatatc
 E  L  G  E  S  A  H  K  A  V  E  E  G  G  S  S  H  S  N  I
actttcttgctacaagacataatgcaactagcacagtccaataattga
 T  F  L  L  Q  D  I  M  Q  L  A  Q  S  N  N  -
```

Figure 20

UGT85A1

```
atgggatctcagatcattcataactcacaaaaaccacatgtagtttgtgttccatatccg
 M  G  S  Q  I  I  H  N  S  Q  K  P  H  V  V  C  V  P  Y  P
gctcaaggccacatcaaccctatgatgagagtggctaaactcctccacgccagaggcttc
 A  Q  G  H  I  N  P  M  M  R  V  A  K  L  L  H  A  R  G  F
tacgtcaccttcgtcaacaccgtctacaaccacaatcgtttccttcgttctcgtgggtcc
 Y  V  T  F  V  N  T  V  Y  N  H  N  R  F  L  R  S  R  G  S
aatgccctagatggacttccttcgttccgatttgagtccattgctgacggtctaccagag
 N  A  L  D  G  L  P  S  F  R  F  E  S  I  A  D  G  L  P  E
acagacatggatgccacgcaggacatcacagctctttgcgagtccaccatgaagaactgt
 T  D  M  D  A  T  Q  D  I  T  A  L  C  E  S  T  M  K  N  C
ctcgctccgttcagagagcttctccagcggatcaacgctggagataatgttcctccggta
 L  A  P  F  R  E  L  L  Q  R  I  N  A  G  D  N  V  P  P  V
agctgtattgtatctgacggttgtatgagctttactcttgatgttgcggaggagcttgga
 S  C  I  V  S  D  G  C  M  S  F  T  L  D  V  A  E  E  L  G
gtcccggaggttctttttggacaaccagtggctgtgcgttcctggcttatctacacttt
 V  P  E  V  L  F  W  T  T  S  G  C  A  F  L  A  Y  L  H  F
tatctcttcatcgagaagggcttatgtccgctaaaagatgagagttacttgacgaaggag
 Y  L  F  I  E  K  G  L  C  P  L  K  D  E  S  Y  L  T  K  E
tacttagaagacacggttatagatttataccaaccatgaagaatgtgaaactaaaggat
 Y  L  E  D  T  V  I  D  F  I  P  T  M  K  N  V  K  L  K  D
attcctagcttcatacgtaccactaatcctgatgatgttatgattagtttcgccctccgc
 I  P  S  F  I  R  T  T  N  P  D  D  V  M  I  S  F  A  L  R
gagaccgagcgagccaaacgtgcttctgctatcattctaaacacatttgatgaccttgag
 E  T  E  R  A  K  R  A  S  A  I  I  L  N  T  F  D  D  L  E
catgatgttgttcatgctatgcaatctatcttacctccggtttattcagttggaccgctt
 H  D  V  V  H  A  M  Q  S  I  L  P  P  V  Y  S  V  G  P  L
catctcttagcaaaccgggagattgaagaaggtagtgagattggaatgatgagttcgaat
 H  L  L  A  N  R  E  I  E  E  G  S  E  I  G  M  M  S  S  N
ttatggaaagaggagatggagtgtttggattggcttgatactaagactcaaaatagtgtc
 L  W  K  E  E  M  E  C  L  D  W  L  D  T  K  T  Q  N  S  V
atttatatcaactttgggagcataacggttttgagtgtgaagcagcttgtggagtttgct
 I  Y  I  N  F  G  S  I  T  V  L  S  V  K  Q  L  V  E  F  A
tggggtttggcgggaagtgggaaagagtttttatgggtgatccggccagatttagtagcg
 W  G  L  A  G  S  G  K  E  F  L  W  V  I  R  P  D  L  V  A
ggagaggaggctatggttccgccggactttttaatggagactaaagaccgcagtatgcta
 G  E  E  A  M  V  P  P  D  F  L  M  E  T  K  D  R  S  M  L
gcgagttggtgtcctcaagagaaagtactttctcatcctgctattggagggtttttgacg
 A  S  W  C  P  Q  E  K  V  L  S  H  P  A  I  G  G  F  L  T
cattgcgggtggaactcgatattggaaagtctttcgtgtggagttccgatggtgtgttgg
 H  C  G  W  N  S  I  L  E  S  L  S  C  G  V  P  M  V  C  W
ccatttttgctgaccagcaaatgaattgtaagttttgttgtgacgagtgggatgttggg
 P  F  F  A  D  Q  Q  M  N  C  K  F  C  C  D  E  W  D  V  G
attgagataggtggagatgtgaagagagaggaagttgaggcggtggttagagagctcatg
 I  E  I  G  G  D  V  K  R  E  E  V  E  A  V  V  R  E  L  M
gatggagagaagggaaagaaaatgagagaaaaggcggtagagtggcagcgcttagccgag
 D  G  E  K  G  K  K  M  R  E  K  A  V  E  W  Q  R  L  A  E
aaagcgacggaacataaacttggttcttccgttatgaattttgagacggttgttagcaag
 K  A  T  E  H  K  L  G  S  S  V  M  N  F  E  T  V  V  S  K
tttcttttgggacaaaaatcacaggattaa
 F  L  L  G  Q  K  S  Q  D  -
```

Figure 21

UGT85A2

```
atgggatctcatgtcgcacaaaaacaacacgtagtttgcgttccttatccggctcaaggc
 M  G  S  H  V  A  Q  K  Q  H  V  V  C  V  P  Y  P  A  Q  G
cacatcaacccaatgatgaaagtggctaaactcctttacgccaaaggcttccatattacc
 H  I  N  P  M  M  K  V  A  K  L  L  Y  A  K  G  F  H  I  T
ttcgtcaacaccgtctacaaccacaaccgtctcctccggtcccgtgggcctaacgccgtt
 F  V  N  T  V  Y  N  H  N  R  L  L  R  S  R  G  P  N  A  V
gacgggcttccttctttccggtttgagtccatccctgacggtctacccgagactgacgtg
 D  G  L  P  S  F  R  F  E  S  I  P  D  G  L  P  E  T  D  V
gacgtcactcaggacatccctactctttgcgagtccacaatgaagcactgtctcgctcca
 D  V  T  Q  D  I  P  T  L  C  E  S  T  M  K  H  C  L  A  P
ttcaaggagcttctccggcagatcaacgcaagggatgatgttcctcctgtgagctgtatc
 F  K  E  L  L  R  Q  I  N  A  R  D  D  V  P  P  V  S  C  I
gtatccgacggttgtatgagcttcacacttgatgctgcggaggagctcggtgtcccggag
 V  S  D  G  C  M  S  F  T  L  D  A  A  E  E  L  G  V  P  E
gttctttttggacaactagtgcttgtggcttcttggcttacctttactactatcgcttc
 V  L  F  W  T  T  S  A  C  G  F  L  A  Y  L  Y  Y  Y  R  F
atcgagaagggattatcaccaataaaagatgagagttacttaaccaaggaacacttggac
 I  E  K  G  L  S  P  I  K  D  E  S  Y  L  T  K  E  H  L  D
acaaaaatagactggataccatcgatgaagaacctaagactaaaagacatccctagcttc
 T  K  I  D  W  I  P  S  M  K  N  L  R  L  K  D  I  P  S  F
atccgaacgactaatcctgacgacatcatgctcaactttatcatccgtgaggctgaccga
 I  R  T  T  N  P  D  D  I  M  L  N  F  I  I  R  E  A  D  R
gccaaacgcgcttcagctatcattctcaacacgtttgatgatctcgaacacgacgttatc
 A  K  R  A  S  A  I  I  L  N  T  F  D  D  L  E  H  D  V  I
caatctatgaaatccattgtacctccggtttattctattggaccgttacatttactagag
 Q  S  M  K  S  I  V  P  P  V  Y  S  I  G  P  L  H  L  L  E
aaacaagagagcggcgagtatagtgaaatcggacggacaggatcgaatctttggagagag
 K  Q  E  S  G  E  Y  S  E  I  G  R  T  G  S  N  L  W  R  E
gagactgagtgtctggactggctaaacacgaaagctagaaacagtgttgtgtacgttaac
 E  T  E  C  L  D  W  L  N  T  K  A  R  N  S  V  V  Y  V  N
ttcgggagtataactgttttgagcgcaaaacagcttgtggagtttgcatggggtttggct
 F  G  S  I  T  V  L  S  A  K  Q  L  V  E  F  A  W  G  L  A
gcaacggggaaagagttttgtgggtgatccggccggatttagtagccggggatgaggca
 A  T  G  K  E  F  L  W  V  I  R  P  D  L  V  A  G  D  E  A
atggttccaccggagttttaacggctacggcggaccggaggatgttggcaagttggtgt
 M  V  P  P  E  F  L  T  A  T  A  D  R  R  M  L  A  S  W  C
cctcaagagaaagtcctttctcatccggccattggagggttcttgacgcattgcgggtgg
 P  Q  E  K  V  L  S  H  P  A  I  G  G  F  L  T  H  C  G  W
aactcgacgttggaaagtctatgcggtggagttccaatggtgtgttggccgtttttttgca
 N  S  T  L  E  S  L  C  G  G  V  P  M  V  C  W  P  F  F  A
gagcaacaaactaattgtaagttttctcgtgacgaatgggaggttgggattgagattggt
 E  Q  Q  T  N  C  K  F  S  R  D  E  W  E  V  G  I  E  I  G
ggagatgtgaagagagaagaggttgaggcggtggttagggagttgatggatgaagagaag
 G  D  V  K  R  E  E  V  E  A  V  V  R  E  L  M  D  E  E  K
ggaaagaatatgagagagaaggcggaagagtggcggcgcttggcgaatgaagcgacggag
 G  K  N  M  R  E  K  A  E  E  W  R  R  L  A  N  E  A  T  E
cataagcatggttcttctaaattgaactttgagatgctcgttaataaggttcttttaggg
 H  K  H  G  S  S  K  L  N  F  E  M  L  V  N  K  V  L  L  G
gagtag
 E  -
```

Figure 22

UGT85A4

```
atggaacaacatggcggttctagctcacagaaacctcacgcaatgtgcataccttatcca
 M  E  Q  H  G  G  S  S  S  Q  K  P  H  A  M  C  I  P  Y  P
gcacaaggccacatcaacccaatgctgaaactagccaagctcctccacgctagaggcttc
 A  Q  G  H  I  N  P  M  L  K  L  A  K  L  L  H  A  R  G  F
cacgtcactttcgtcaacaccgactacaaccaccgccgtatcctccaatcacgtggccct
 H  V  T  F  V  N  T  D  Y  N  H  R  R  I  L  Q  S  R  G  P
cacgctctcaacggtctcccctcgtttcgcttcgagactatccccgacggtcttccttgg
 H  A  L  N  G  L  P  S  F  R  F  E  T  I  P  D  G  L  P  W
acagacgtcgacgctaagcaagacatgctcaagcttattgactccacaataaacaactgt
 T  D  V  D  A  K  Q  D  M  L  K  L  I  D  S  T  I  N  N  C
ttagctccattcaaagacctcatcctccggttaaactccggttctgatataccaccggtt
 L  A  P  F  K  D  L  I  L  R  L  N  S  G  S  D  I  P  P  V
agctgtatcatctccgacgcttcaatgagcttcacaattgacgcagcggaggagcttaaa
 S  C  I  I  S  D  A  S  M  S  F  T  I  D  A  A  E  E  L  K
attccggtagttctcctctggaccaacagtgctactgctttaatcttgtatctccattac
 I  P  V  V  L  L  W  T  N  S  A  T  A  L  I  L  Y  L  H  Y
caaaaactcatcgagaaagagataattcccctcaaagattcgagtgacttgaagaagcat
 Q  K  L  I  E  K  E  I  I  P  L  K  D  S  S  D  L  K  K  H
ttagagacggagattgattggataccgtcgatgaagaagattaagcttaaggattttcca
 L  E  T  E  I  D  W  I  P  S  M  K  K  I  K  L  K  D  F  P
gatttcgtcaccacgacgaatcctcaagatccgatgattagtttcatccttcatgtaacc
 D  F  V  T  T  T  N  P  Q  D  P  M  I  S  F  I  L  H  V  T
ggaagaatcaaaagagcttctgcgatcttcatcaacactttcgaaaaactcgagcataac
 G  R  I  K  R  A  S  A  I  F  I  N  T  F  E  K  L  E  H  N
gttctcttatctctgcgatctcttctccctcagatctactccgttggaccgttccagatt
 V  L  L  S  L  R  S  L  L  P  Q  I  Y  S  V  G  P  F  Q  I
ctggagaatcgcgaaatcgataagaacagcgaaatcagaaagctaggattgaatctctgg
 L  E  N  R  E  I  D  K  N  S  E  I  R  K  L  G  L  N  L  W
gaagaagagacggagtctttggattggctagatactaaagctgagaaagctgtgatttac
 E  E  E  T  E  S  L  D  W  L  D  T  K  A  E  K  A  V  I  Y
gtcaacttcgggagtctaacggttttgactagtgagcagatcttagagttcgcttggggt
 V  N  F  G  S  L  T  V  L  T  S  E  Q  I  L  E  F  A  W  G
ttagcgaggagcgggaaagagtttctctgggtggtgagatctggtatggtcgacggagat
 L  A  R  S  G  K  E  F  L  W  V  V  R  S  G  M  V  D  G  D
gattcgattcttccggcggagtttttatcggagacgaagaatcgaggaatgttaattaaa
 D  S  I  L  P  A  E  F  L  S  E  T  K  N  R  G  M  L  I  K
ggatggtgttctcaggagaaggtactttcgcatccggcgattggaggatttttgactcac
 G  W  C  S  Q  E  K  V  L  S  H  P  A  I  G  G  F  L  T  H
tgtggatggaattcgacgttggagagtttgtacgccggtgttccgatgatctgttggcca
 C  G  W  N  S  T  L  E  S  L  Y  A  G  V  P  M  I  C  W  P
ttttttgctgatcagttgacgaatcgaaagttctgttgcgaggattgggggattgggatg
 F  F  A  D  Q  L  T  N  R  K  F  C  C  E  D  W  G  I  G  M
gagatcggcgaggaggtgaagagggagagagtggagacggtggttaaagagctcatggac
 E  I  G  E  E  V  K  R  E  R  V  E  T  V  V  K  E  L  M  D
ggagagaagggaaagaggttaagagagaaggtggtggagtggcggcgcttggcggaagaa
 G  E  K  G  K  R  L  R  E  K  V  V  E  W  R  R  L  A  E  E
gcttcggcgccaccgttgggatcatcgtacgtgaattttgaaacggtggttaataaagtc
 A  S  A  P  P  L  G  S  S  Y  V  N  F  E  T  V  V  N  K  V
cttacatgtcacacgattagatcgacctaa
 L  T  C  H  T  I  R  S  T  -
```

Figure 23

UGT85A7

```
atggaatctcatgttgttcataacgcacaaaagccacacgtagtttgcgtgccttacccg
 M  E  S  H  V  V  H  N  A  Q  K  P  H  V  V  C  V  P  Y  P
gctcaaggccacatcaatccgatgctgaaagtggctaaactcctctacgctaaaggcttt
 A  Q  G  H  I  N  P  M  L  K  V  A  K  L  L  Y  A  K  G  F
cacgtcaccttcgttaacactctctacaaccacaaccgtctcctccggtcccgtggtccc
 H  V  T  F  V  N  T  L  Y  N  H  N  R  L  L  R  S  R  G  P
aacgcgctcgacgggtttccttcattccggttcgagtccatccctgacggtctaccggag
 N  A  L  D  G  F  P  S  F  R  F  E  S  I  P  D  G  L  P  E
actgatggcgataggacgcagcatactcctaccgtttgcatgtccattgagaaaaactgt
 T  D  G  D  R  T  Q  H  T  P  T  V  C  M  S  I  E  K  N  C
ctcgctccattcaaagagattctgcgccggatcaacgataaagatgatgttcctccagtg
 L  A  P  F  K  E  I  L  R  R  I  N  D  K  D  D  V  P  P  V
agttgtattgtatcggacggtgtgatgagttttactcttgacgcagccgaggaactaggt
 S  C  I  V  S  D  G  V  M  S  F  T  L  D  A  A  E  E  L  G
gtcccagaggttatttttggaccaatagtgcttgtggtttcatgactattctacacttt
 V  P  E  V  I  F  W  T  N  S  A  C  G  F  M  T  I  L  H  F
tatcttttcatcgagaagggtctatctcctttaaagacgaaagttacatgtcaaaggag
 Y  L  F  I  E  K  G  L  S  P  F  K  D  E  S  Y  M  S  K  E
catctagacacagttatagattggataccatcaatgaagaatcttaggttaaaggacatc
 H  L  D  T  V  I  D  W  I  P  S  M  K  N  L  R  L  K  D  I
cctagctatatacgtaccacaaatcctgacaacataatgcttaatttcctcattcgagaa
 P  S  Y  I  R  T  T  N  P  D  N  I  M  L  N  F  L  I  R  E
gttgagcgatctaaacgcgctagtgctatcattctcaacacgtttgatgaactcgagcat
 V  E  R  S  K  R  A  S  A  I  I  L  N  T  F  D  E  L  E  H
gatgttatccagtctatgcaatctattttacctccggtttattctattgggccactccat
 D  V  I  Q  S  M  Q  S  I  L  P  P  V  Y  S  I  G  P  L  H
ctccttgtgaaggaagaaataaacgaggctagtgaaataggacagatgggattaaatttg
 L  L  V  K  E  E  I  N  E  A  S  E  I  G  Q  M  G  L  N  L
tggagagaggagatggaatgtttggattggctcgatacaaaaactccaaacagtgttctt
 W  R  E  E  M  E  C  L  D  W  L  D  T  K  T  P  N  S  V  L
tttgttaactttggatgcataacggtgatgagtgcaaaacagcttgaagaatttgcttgg
 F  V  N  F  G  C  I  T  V  M  S  A  K  Q  L  E  E  F  A  W
ggtttggcggcaagtaggaaagagttttatgggtgatccgtcctaatttagtggtggga
 G  L  A  A  S  R  K  E  F  L  W  V  I  R  P  N  L  V  V  G
gaggcgatggtggttcttccacaagagtttttagcggagacgatagaccggagaatgtta
 E  A  M  V  V  L  P  Q  E  F  L  A  E  T  I  D  R  R  M  L
gctagttggtgtcctcaggagaaagttctttctcatcccgcgataggagggttcttgacg
 A  S  W  C  P  Q  E  K  V  L  S  H  P  A  I  G  G  F  L  T
cattgcgggtggaactcaacattggagagtctcgctggtggtgttccgatgatatgttgg
 H  C  G  W  N  S  T  L  E  S  L  A  G  G  V  P  M  I  C  W
ccatgtttttcggagcaaccgacgaattgtaagttttgttgtgatgagtggggagtgggt
 P  C  F  S  E  Q  P  T  N  C  K  F  C  C  D  E  W  G  V  G
atagagattggtaaagatgtgaagagagaggaggtcgagacggtggttagagaacttatg
 I  E  I  G  K  D  V  K  R  E  E  V  E  T  V  V  R  E  L  M
gatggagaaaaggggaaaaagctgagagaaaaggcggaagagtggcggcggttggccgag
 D  G  E  K  G  K  K  L  R  E  K  A  E  E  W  R  R  L  A  E
gaagcgacgaggtataaacatggttcgtcggtcatgaatcttgagacgcttatacataaa
 E  A  T  R  Y  K  H  G  S  S  V  M  N  L  E  T  L  I  H  K
gttttcttagaaaatcttagatga
 V  F  L  E  N  L  R  -
```

Figure 24