



US 2007011230A1

(19) **United States**(12) **Patent Application Publication**
Pompejus et al.(10) **Pub. No.: US 2007/011230 A1**(43) **Pub. Date: May 17, 2007**(54) **CORYNEBACTERIUM GLUTAMICUM**
GENES ENCODING PROTEINS INVOLVED
IN MEMBRANE SYNTHESIS AND
MEMBRANE TRANSPORT(75) Inventors: **Markus Pompejus**, Freinsheim (DE);
Burkhard Kroger, Limburgerhof (DE);
Hartwig Schroder, Nussloch (DE);
Oskar Zelder, Speyer (DE); **Gregor**
Haberhauer, Limburgerhof (DE)Correspondence Address:
LAHIVE & COCKFIELD, LLP
ONE POST OFFICE SQUARE
BOSTON, MA 02109-2127 (US)(73) Assignee: **BASF Aktiengesellschaft**, Ludwigshafen
(DE)(21) Appl. No.: **11/507,094**(22) Filed: **Aug. 18, 2006****Related U.S. Application Data**(60) Division of application No. 10/627,476, filed on Jul.
25, 2003, which is a continuation of application No.
09/602,787, filed on Jun. 23, 2000, now Pat. No.
6,696,561.(60) Provisional application No. 60/141,031, filed on Jun.
25, 1999.(30) **Foreign Application Priority Data**Jul. 8, 1999 (DE)..... 19931454.3
Jul. 8, 1999 (DE)..... 19931478.0
Jul. 8, 1999 (DE)..... 19931563.9
Jul. 9, 1999 (DE)..... 19932122.1
Jul. 9, 1999 (DE)..... 19932124.8
Jul. 9, 1999 (DE)..... 19932125.6
Jul. 9, 1999 (DE)..... 19932128.0
Jul. 9, 1999 (DE)..... 19932180.9
Jul. 9, 1999 (DE)..... 19932182.5
Jul. 9, 1999 (DE)..... 19932190.6
Jul. 9, 1999 (DE)..... 19932191.4
Jul. 9, 1999 (DE)..... 19932209.0
Jul. 9, 1999 (DE)..... 19932212.0
Jul. 9, 1999 (DE)..... 19932227.9Jul. 9, 1999 (DE)..... 19932228.7
Jul. 9, 1999 (DE)..... 19932229.5
Jul. 9, 1999 (DE)..... 19932230.9
Jul. 14, 1999 (DE)..... 19932927.3
Jul. 14, 1999 (DE)..... 19933005.0
Jul. 14, 1999 (DE)..... 19933006.9
Aug. 27, 1999 (DE)..... 19940764.9
Aug. 27, 1999 (DE)..... 19940765.7
Aug. 27, 1999 (DE)..... 19940766.5
Aug. 27, 1999 (DE)..... 19940830.0
Aug. 27, 1999 (DE)..... 19940831.9
Aug. 27, 1999 (DE)..... 19940832.7
Aug. 27, 1999 (DE)..... 19940833.5
Aug. 31, 1999 (DE)..... 19941378.9
Aug. 31, 1999 (DE)..... 19941379.7
Aug. 31, 1999 (DE)..... 19941395.9
Sep. 3, 1999 (DE)..... 19942077.7
Sep. 3, 1999 (DE)..... 19942078.5
Sep. 3, 1999 (DE)..... 19942079.3
Sep. 3, 1999 (DE)..... 19942088.2**Publication Classification**(51) **Int. Cl.****C12Q 1/68** (2006.01)
G01N 33/554 (2006.01)
C07H 21/04 (2006.01)
C12N 15/74 (2006.01)
G01N 33/569 (2006.01)
C07K 14/195 (2006.01)(52) **U.S. Cl.** **435/6**; 536/23.7; 435/7.32;
435/69.3; 435/252.3; 435/471;
530/350(57) **ABSTRACT**

Isolated nucleic acid molecules, designated MCT nucleic acid molecules, which encode novel MCT proteins from *Corynebacterium glutamicum* are described. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing MCT nucleic acid molecules, and host cells into which the expression vectors have been introduced. The invention still further provides isolated MCT proteins, mutated MCT proteins, fusion proteins, antigenic peptides and methods for the improvement of production of a desired compound from *C. glutamicum* based on genetic engineering of MCT genes in this organism.

**CORYNEBACTERIUM GLUTAMICUM GENES
ENCODING PROTEINS INVOLVED IN
MEMBRANE SYNTHESIS AND MEMBRANE
TRANSPORT**

RELATED APPLICATIONS

[0001] This application is a divisional of U.S. application Ser. No. 10/627,476, filed Jul. 25, 2003, which is a continuation of U.S. application Ser. No. 09/602,787, filed Jun. 23, 2000, which, in turn, claims priority to prior filed U.S. Provisional Patent Application Ser. No. 60/141031, filed Jun. 25, 1999. This application also claims priority to German Patent Application No. 19931454.3, filed Jul. 8, 1999, German Patent Application No. 19931478.0, filed Jul. 8, 1999, German Patent Application No. 19931563.9, filed Jul. 8, 1999, German Patent Application No. 19932122.1, filed Jul. 9, 1999, German Patent Application No. 19932124.8, filed Jul. 9, 1999, German Patent Application No. 19932125.6, filed Jul. 9, 1999, German Patent Application No. 19932128.0, filed Jul. 9, 1999, German Patent Application No. 19932180.9, filed Jul. 9, 1999, German Patent Application No. 19932182.5, filed Jul. 9, 1999, German Patent Application No. 19932190.6, filed Jul. 9, 1999, German Patent Application No. 19932191.4, filed Jul. 9, 1999, German Patent Application No. 19932209.0, filed Jul. 9, 1999, German Patent Application No. 19932212.0, filed Jul. 9, 1999, German Patent Application No. 19932227.9, filed Jul. 9, 1999, German Patent Application No. 19932228.7, filed Jul. 9, 1999, German Patent Application No. 19932229.5, filed Jul. 9, 1999, German Patent Application No. 19932230.9, filed Jul. 9, 1999, German Patent Application No. 19932927.3, filed Jul. 14, 1999, German Patent Application No. 19933005.0, filed Jul. 14, 1999, German Patent Application No. 19933006.9, filed Jul. 14, 1999, German Patent Application No. 19940764.9, filed Aug. 27, 1999, German Patent Application No. 19940765.7, filed Aug. 27, 1999, German Patent Application No. 19940766.5, filed Aug. 27, 1999, German Patent Application No. 19940830.0, filed Aug. 27, 1999, German Patent Application No. 19940831.9, filed Aug. 27, 1999, German Patent Application No. 19940832.7, filed Aug. 27, 1999, German Patent Application No. 19940833.5, filed Aug. 27, 1999, German Patent Application No. 19941378.9, filed Aug. 31, 1999, German Patent Application No. 19941379.7, filed Aug. 31, 1999, German Patent Application No. 19941395.9, filed Aug. 31, 1999, German Patent Application No. 19942077.7, filed Sep. 3, 1999, German Patent Application No. 19942078.5, filed September 3, 1999, German Patent Application No. 19942079.3, filed Sep. 3, 1999, and German Patent Application No. 19942088.2, filed Sep. 3, 1999. The entire contents of all of the above referenced applications are hereby expressly incorporated herein by this reference.

**INCORPORATION OF MATERIAL SUBMITTED
ON COMPACT DISCS**

[0002] This application incorporates herein by reference the material contained on the compact discs submitted herewith as part of this application. Specifically, the file "seqlistcorr2" (2.43 MB) contained on each of Copy 1, Copy 2 and the CRF copy of the Sequence Listing is hereby incorporated herein by reference. This file was created on Jul. 31, 2006. In addition, the files "Appendix A" (399 KB) and "Appendix B" (140 KB) contained on each of the

compact disks entitled "Appendices Copy 1" and "Appendices Copy 2" are hereby incorporated herein by reference. Each of these files were created on Jul. 31, 2006.

BACKGROUND OF THE INVENTION

[0003] Certain products and by-products of naturally-occurring metabolic processes in cells have utility in a wide array of industries, including the food, feed, cosmetics, and pharmaceutical industries. These molecules, collectively termed 'fine chemicals', include organic acids, both proteinogenic and non-proteinogenic amino acids, nucleotides and nucleosides, lipids and fatty acids, diols, carbohydrates, aromatic compounds, vitamins and cofactors, and enzymes. Their production is most conveniently performed through the large-scale culture of bacteria developed to produce and secrete large quantities of one or more desired molecules. One particularly useful organism for this purpose is *Corynebacterium glutamicum*, a gram positive, nonpathogenic bacterium. Through strain selection, a number of mutant strains have been developed which produce an array of desirable compounds. However, selection of strains improved for the production of a particular molecule is a time-consuming and difficult process.

SUMMARY OF THE INVENTION

[0004] The invention provides novel bacterial nucleic acid molecules which have a variety of uses. These uses include the identification of microorganisms which can be used to produce fine chemicals, the modulation of fine chemical production in *C. glutamicum* or related bacteria, the typing or identification of *C. glutamicum* or related bacteria, as reference points for mapping the *C. glutamicum* genome, and as markers for transformation. These novel nucleic acid molecules encode proteins, referred to herein as membrane construction and membrane transport (MCT) proteins.

[0005] *C. glutamicum* is a gram positive, aerobic bacterium which is commonly used in industry for the large-scale production of a variety of fine chemicals, and also for the degradation of hydrocarbons (such as in petroleum spills) and for the oxidation of terpenoids. The MCT nucleic acid molecules of the invention, therefore, can be used to identify microorganisms which can be used to produce fine chemicals, e.g., by fermentation processes. Modulation of the expression of the MCT nucleic acids of the invention, or modification of the sequence of the MCT nucleic acid molecules of the invention, can be used to modulate the production of one or more fine chemicals from a microorganism (e.g., to improve the yield or production of one or more fine chemicals from a *Corynebacterium* or *Brevibacterium* species).

[0006] The MCT nucleic acids of the invention may also be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof, or to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present. Although *Corynebacterium glutamicum* itself is nonpathogenic, it is

related to species pathogenic in humans, such as *Corynebacterium diphtheriae* (the causative agent of diphtheria); the detection of such organisms is of significant clinical relevance.

[0007] The MCT nucleic acid molecules of the invention may also serve as reference points for mapping of the *C. glutamicum* genome, or of genomes of related organisms. Similarly, these molecules, or variants or portions thereof, may serve as markers for genetically engineered *Corynebacterium* or *Brevibacterium* species. e.g.e.g. The MCT proteins encoded by the novel nucleic acid molecules of the invention are capable of, for example, performing a function involved in the metabolism (e.g., the biosynthesis or degradation) of compounds necessary for membrane biosynthesis, or of assisting in the transmembrane transport of one or more compounds either into or out of the cell. Given the availability of cloning vectors for use in *Corynebacterium glutamicum*, such as those disclosed in Sinskey et al., U.S. Pat. No. 4,649,119, and techniques for genetic manipulation of *C. glutamicum* and the related *Brevibacterium* species (e.g., *lactofermentum*) (Yoshihama et al, *J. Bacteriol.* 162: 591-597 (1985); Katsumata et al., *J. Bacteriol.* 159: 306-311 (1984); and Santamaria et al., *J. Gen. Microbiol.* 130: 2237-2246 (1984)), the nucleic acid molecules of the invention may be utilized in the genetic engineering of this organism to make it a better or more efficient producer of one or more fine chemicals. This improved production or efficiency of production of a fine chemical may be due to a direct effect of manipulation of a gene of the invention, or it may be due to an indirect effect of such manipulation.

[0008] There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursors, cofactors, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

[0009] The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, MCT proteins of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the

cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from *C. glutamicum* in large-scale fermentative culture.

[0010] The invention provides novel nucleic acid molecules which encode proteins, referred to herein as MCT proteins, which are capable of, for example, participating in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Nucleic acid molecules encoding an MCT protein are referred to herein as MCT nucleic acid molecules. In a preferred embodiment, the MCT protein participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of such proteins include those encoded by the genes set forth in Table 1.

[0011] Accordingly, one aspect of the invention pertains to isolated nucleic acid molecules (e.g., cDNAs, DNAs, or RNAs) comprising a nucleotide sequence encoding an MCT protein or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection or amplification of MCT-encoding nucleic acid (e.g., DNA or mRNA). In particularly preferred embodiments, the isolated nucleic acid molecule comprises one of the nucleotide sequences set forth in Appendix A or the coding region or a complement thereof of one of these nucleotide sequences. In other particularly preferred embodiments, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes to or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80% or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence set forth in Appendix A, or a portion thereof. In other preferred embodiments, the isolated nucleic acid molecule encodes one of the amino acid sequences set forth in Appendix B. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein.

[0012] In another embodiment, the isolated nucleic acid molecule encodes a protein or portion thereof wherein the protein or portion thereof includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B, e.g., sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains an MCT activity. Preferably, the protein or

portion thereof encoded by the nucleic acid molecule maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In one embodiment, the protein encoded by the nucleic acid molecule is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90% and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an amino acid sequence of Appendix B (e.g., an entire amino acid sequence selected from those sequences set forth in Appendix B). In another preferred embodiment, the protein is a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

[0013] In another preferred embodiment, the isolated nucleic acid molecule is derived from *C. glutamicum* and encodes a protein (e.g., an MCT fusion protein) which includes a biologically active domain which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1, and which also includes heterologous nucleic acid sequences encoding a heterologous polypeptide or regulatory regions.

[0014] In another embodiment, the isolated nucleic acid molecule is at least 15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising a nucleotide sequence of Appendix A. Preferably, the isolated nucleic acid molecule corresponds to a naturally-occurring nucleic acid molecule. More preferably, the isolated nucleic acid encodes a naturally-occurring *C. glutamicum* MCT protein, or a biologically active portion thereof.

[0015] Another aspect of the invention pertains to vectors, e.g., recombinant expression vectors, containing the nucleic acid molecules of the invention, and host cells into which such vectors have been introduced. In one embodiment, such a host cell is used to produce an MCT protein by culturing the host cell in a suitable medium. The MCT protein can be then isolated from the medium or the host cell.

[0016] Yet another aspect of the invention pertains to a genetically altered microorganism in which an MCT gene has been introduced or altered. In one embodiment, the genome of the microorganism has been altered by introduction of a nucleic acid molecule of the invention encoding wild-type or mutated MCT sequence as a transgene. In another embodiment, an endogenous MCT gene within the genome of the microorganism has been altered, e.g., functionally disrupted, by homologous recombination with an altered MCT gene. In another embodiment, an endogenous or introduced MCT gene in a microorganism has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MCT protein. In still another embodiment, one or more of the regulatory regions (e.g., a promoter, repressor, or inducer) of an MCT gene in a microorganism has been altered (e.g., by deletion, truncation, inversion, or point mutation) such that the expression of the MCT gene is modulated. In a preferred embodiment, the microorganism belongs to the genus *Corynebacterium* or

Brevibacterium, with *Corynebacterium glutamicum* being particularly preferred. In a preferred embodiment, the microorganism is also utilized for the production of a desired compound, such as an amino acid, with lysine being particularly preferred.

[0017] In another aspect, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (e.g., the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject.

[0018] Still another aspect of the invention pertains to an isolated MCT protein or a portion, e.g., a biologically active portion, thereof. In a preferred embodiment, the isolated MCT protein or portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In another preferred embodiment, the isolated MCT protein or portion thereof is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes.

[0019] The invention also provides an isolated preparation of an MCT protein. In preferred embodiments, the MCT protein comprises an amino acid sequence of Appendix B. In another preferred embodiment, the invention pertains to an isolated full length protein which is substantially homologous to an entire amino acid sequence of Appendix B (encoded by an open reading frame set forth in Appendix A). In yet another embodiment, the protein is at least about 50%, preferably at least about 60%, and more preferably at least about 70%, 80%, or 90%, and most preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous to an entire amino acid sequence of Appendix B. In other embodiments, the isolated MCT protein comprises an amino acid sequence which is at least about 50% or more homologous to one of the amino acid sequences of Appendix B and is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1.

[0020] Alternatively, the isolated MCT protein can comprise an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, or is at least about 50%, preferably at least about 60%, more preferably at least about 70%, 80%, or 90%, and even more preferably at least about 95%, 96%, 97%, 98%, or 99% or more homologous, to a nucleotide sequence of Appendix B. It is also preferred that the preferred forms of MCT proteins also have one or more of the MCT bioactivities described herein.

[0021] The MCT polypeptide, or a biologically active portion thereof, can be operatively linked to a non-MCT polypeptide to form a fusion protein. In preferred embodiments, this fusion protein has an activity which differs from that of the MCT protein alone. In other preferred embodiments, this fusion protein participate in the metabolism of

compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In particularly preferred embodiments, integration of this fusion protein into a host cell modulates production of a desired compound from the cell.

[0022] In another aspect, the invention provides methods for screening molecules which modulate the activity of an MCT protein, either by interacting with the protein itself or a substrate or binding partner of the MCT protein, or by modulating the transcription or translation of an MCT nucleic acid molecule of the invention.

[0023] Another aspect of the invention pertains to a method for producing a fine chemical. This method involves the culturing of a cell containing a vector directing the expression of an MCT nucleic acid molecule of the invention, such that a fine chemical is produced. In a preferred embodiment, this method further includes the step of obtaining a cell containing such a vector, in which a cell is transfected with a vector directing the expression of an MCT nucleic acid. In another preferred embodiment, this method further includes the step of recovering the fine chemical from the culture. In a particularly preferred embodiment, the cell is from the genus *Corynebacterium* or *Brevibacterium*, or is selected from those strains set forth in Table 3.

[0024] Another aspect of the invention pertains to methods for modulating production of a molecule from a microorganism. Such methods include contacting the cell with an agent which modulates MCT protein activity or MCT nucleic acid expression such that a cell associated activity is altered relative to this same activity in the absence of the agent. In a preferred embodiment, the cell is modulated for one or more *C. glutamicum* metabolic pathways for cell membrane components or is modulated for the transport of compounds across such membranes, such that the yields or rate of production of a desired fine chemical by this microorganism is improved. The agent which modulates MCT protein activity can be an agent which stimulates MCT protein activity or MCT nucleic acid expression. Examples of agents which stimulate MCT protein activity or MCT nucleic acid expression include small molecules, active MCT proteins, and nucleic acids encoding MCT proteins that have been introduced into the cell. Examples of agents which inhibit MCT activity or expression include small molecules and antisense MCT nucleic acid molecules.

[0025] Another aspect of the invention pertains to methods for modulating yields of a desired compound from a cell, involving the introduction of a wild-type or mutant MCT gene into a cell, either maintained on a separate plasmid or integrated into the genome of the host cell. If integrated into the genome, such integration can be random, or it can take place by homologous recombination such that the native gene is replaced by the introduced copy, causing the production of the desired compound from the cell to be modulated. In a preferred embodiment, said yields are increased. In another preferred embodiment, said chemical is a fine chemical. In a particularly preferred embodiment, said fine chemical is an amino acid. In especially preferred embodiments, said amino acid is L-lysine.

DETAILED DESCRIPTION OF THE INVENTION

[0026] The present invention provides MCT nucleic acid and protein molecules which are involved in the metabolism

of cellular membrane components in *C. glutamicum* or in the transport of compounds across such membranes. The molecules of the invention may be utilized in the modulation of production of fine chemicals from microorganisms, such as *C. glutamicum*, either directly (e.g., where overexpression or optimization of a fatty acid biosynthesis protein has a direct impact on the yield, production, and/or efficiency of production of the fatty acid from modified *C. glutamicum*), or may have an indirect impact which nonetheless results in an increase of yield, production, and/or efficiency of production of the desired compound (e.g., where modulation of the metabolism of cell membrane components results in alterations in the yield, production, and/or efficiency of production or the composition of the cell membrane, which in turn may impact the production of one or more fine chemicals). Aspects of the invention are further explicated below.

I. Fine Chemicals

[0027] The term 'fine chemical' is art-recognized and includes molecules produced by an organism which have applications in various industries, such as, but not limited to, the pharmaceutical, agriculture, and cosmetics industries. Such compounds include organic acids, such as tartaric acid, itaconic acid, and diaminopimelic acid, both proteinogenic and non-proteinogenic amino acids, purine and pyrimidine bases, nucleosides, and nucleotides (as described e.g. in Kuninaka, A. (1996) Nucleotides and related compounds, p. 561-612, in Biotechnology vol. 6, Rehm et al., eds. VCH: Weinheim, and references contained therein), lipids, both saturated and unsaturated fatty acids (e.g., arachidonic acid), diols (e.g., propane diol, and butane diol), carbohydrates (e.g. hyaluronic acid and trehalose), aromatic compounds (e.g. aromatic amines, vanillin, and indigo), vitamins and cofactors (as described in Ullmann's Encyclopedia of Industrial Chemistry, vol. A27, "Vitamins", p. 443-613 (1996) VCH: Weinheim and references therein; and Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research—Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press, (1995)), enzymes, polyketides (Cane et al. (1998) *Science* 282: 63-68), and all other chemicals described in Gutcho (1983) Chemicals by Fermentation, Noyes Data Corporation, ISBN: 0818805086 and references therein. The metabolism and uses of certain of these fine chemicals are further explicated below.

A. Amino Acid Metabolism and Uses

[0028] Amino acids comprise the basic structural units of all proteins, and as such are essential for normal cellular functioning in all organisms. The term "amino acid" is art-recognized. The proteinogenic amino acids, of which there are 20 species, serve as structural units for proteins, in which they are linked by peptide bonds, while the nonproteinogenic amino acids (hundreds of which are known) are not normally found in proteins (see Ullmann's Encyclopedia of Industrial Chemistry, vol. A2, p. 57-97 VCH: Weinheim (1985)). Amino acids may be in the D- or L-optical configuration, though L-amino acids are generally the only type found in naturally-occurring proteins. Biosynthetic and degradative pathways of each of the 20 proteinogenic amino acids have been well characterized in both prokaryotic and eukaryotic cells (see, for example, Stryer, L. Biochemistry, 3rd edition, pages 578-590 (1988)). The 'essential' amino

acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine), so named because they are generally a nutritional requirement due to the complexity of their biosyntheses, are readily converted by simple biosynthetic pathways to the remaining 11 'non-essential' amino acids (alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, proline, serine, and tyrosine). Higher animals do retain the ability to synthesize some of these amino acids, but the essential amino acids must be supplied from the diet in order for normal protein synthesis to occur.

[0029] Aside from their function in protein biosynthesis, these amino acids are interesting chemicals in their own right, and many have been found to have various applications in the food, feed, chemical, cosmetics, agriculture, and pharmaceutical industries. Lysine is an important amino acid in the nutrition not only of humans, but also of monogastric animals such as poultry and swine. Glutamate is most commonly used as a flavor additive (mono-sodium glutamate, MSG) and is widely used throughout the food industry, as are aspartate, phenylalanine, glycine, and cysteine. Glycine, L-methionine and tryptophan are all utilized in the pharmaceutical industry. Glutamine, valine, leucine, isoleucine, histidine, arginine, proline, serine and alanine are of use in both the pharmaceutical and cosmetics industries. Threonine, tryptophan, and D/ L-methionine are common feed additives. (Leuchtenberger, W. (1996) Amino acids—technical production and use, p. 466-502 in Rehm et al. (eds.) *Biotechnology* vol. 6, chapter 14a, VCH: Weinheim). Additionally, these amino acids have been found to be useful as precursors for the synthesis of synthetic amino acids and proteins, such as N-acetylcysteine, S-carboxymethyl-L-cysteine, (S)-5-hydroxytryptophan, and others described in *Ullman's Encyclopedia of Industrial Chemistry*, vol. A2, p. 57-97, VCH: Weinheim, 1985.

[0030] The biosynthesis of these natural amino acids in organisms capable of producing them, such as bacteria, has been well characterized (for review of bacterial amino acid biosynthesis and regulation thereof, see Umbarger, H. E. (1978) *Ann. Rev. Biochem.* 47: 533-606). Glutamate is synthesized by the reductive amination of α -ketoglutarate, an intermediate in the citric acid cycle. Glutamine, proline, and arginine are each subsequently produced from glutamate. The biosynthesis of serine is a three-step process beginning with 3-phosphoglycerate (an intermediate in glycolysis), and resulting in this amino acid after oxidation, transamination, and hydrolysis steps. Both cysteine and glycine are produced from serine; the former by the condensation of homocysteine with serine, and the latter by the transfer of the side-chain β -carbon atom to tetrahydrofolate, in a reaction catalyzed by serine transhydroxymethylase. Phenylalanine, and tyrosine are synthesized from the glycolytic and pentose phosphate pathway precursors erythrose 4-phosphate and phosphoenolpyruvate in a 9-step biosynthetic pathway that differ only at the final two steps after synthesis of prephenate. Tryptophan is also produced from these two initial molecules, but its synthesis is an 11-step pathway. Tyrosine may also be synthesized from phenylalanine, in a reaction catalyzed by phenylalanine hydroxylase. Alanine, valine, and leucine are all biosynthetic products of pyruvate, the final product of glycolysis. Aspartate is formed from oxaloacetate, an intermediate of the citric acid cycle. Asparagine, methionine, threonine, and lysine are each produced by the conversion of aspartate. Isoleucine is

formed from threonine. A complex 9-step pathway results in the production of histidine from 5-phosphoribosyl-1-pyrophosphate, an activated sugar.

[0031] Amino acids in excess of the protein synthesis needs of the cell cannot be stored, and are instead degraded to provide intermediates for the major metabolic pathways of the cell (for review see Stryer, L. *Biochemistry* 3rd ed. Ch. 21 "Amino Acid Degradation and the Urea Cycle" p. 495-516 (1988)). Although the cell is able to convert unwanted amino acids into useful metabolic intermediates, amino acid production is costly in terms of energy, precursor molecules, and the enzymes necessary to synthesize them. Thus it is not surprising that amino acid biosynthesis is regulated by feedback inhibition, in which the presence of a particular amino acid serves to slow or entirely stop its own production (for overview of feedback mechanisms in amino acid biosynthetic pathways, see Stryer, L. *Biochemistry*, 3rd ed. Ch. 24: "Biosynthesis of Amino Acids and Heme" p. 575-600 (1988)). Thus, the output of any particular amino acid is limited by the amount of that amino acid present in the cell.

B. Vitamin, Cofactor, and Nutraceutical Metabolism and Uses

[0032] Vitamins, cofactors, and nutraceuticals comprise another group of molecules which the higher animals have lost the ability to synthesize and so must ingest, although they are readily synthesized by other organisms such as bacteria. These molecules are either bioactive substances themselves, or are precursors of biologically active substances which may serve as electron carriers or intermediates in a variety of metabolic pathways. Aside from their nutritive value, these compounds also have significant industrial value as coloring agents, antioxidants, and catalysts or other processing aids. (For an overview of the structure, activity, and industrial applications of these compounds, see, for example, *Ullman's Encyclopedia of Industrial Chemistry*, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996.) The term "vitamin" is art-recognized, and includes nutrients which are required by an organism for normal functioning, but which that organism cannot synthesize by itself. The group of vitamins may encompass cofactors and nutraceutical compounds. The language "cofactor" includes nonproteinaceous compounds required for a normal enzymatic activity to occur. Such compounds may be organic or inorganic; the cofactor molecules of the invention are preferably organic. The term "nutraceutical" includes dietary supplements having health benefits in plants and animals, particularly humans. Examples of such molecules are vitamins, antioxidants, and also certain lipids (e.g., polyunsaturated fatty acids).

[0033] The biosynthesis of these molecules in organisms capable of producing them, such as bacteria, has been largely characterized (*Ullman's Encyclopedia of Industrial Chemistry*, "Vitamins" vol. A27, p. 443-613, VCH: Weinheim, 1996; Michal, G. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, John Wiley & Sons; Ong, A.S., Niki, E. & Packer, L. (1995) "Nutrition, Lipids, Health, and Disease" Proceedings of the UNESCO/Confederation of Scientific and Technological Associations in Malaysia, and the Society for Free Radical Research—Asia, held Sept. 1-3, 1994 at Penang, Malaysia, AOCS Press: Champaign, IL X, 374 S).

[0034] Thiamin (vitamin B₁) is produced by the chemical coupling of pyrimidine and thiazole moieties. Riboflavin

(vitamin B₂) is synthesized from guanosine-5'-triphosphate (GTP) and ribose-5'-phosphate. Riboflavin, in turn, is utilized for the synthesis of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The family of compounds collectively termed 'vitamin B₆' (e.g., pyridoxine, pyridoxamine, pyridoxa-5'-phosphate, and the commercially used pyridoxin hydrochloride) are all derivatives of the common structural unit, 5-hydroxy-6-methylpyridine. Pantothenate (pantothenic acid, (R)-(+)-N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine) can be produced either by chemical synthesis or by fermentation. The final steps in pantothenate biosynthesis consist of the ATP-driven condensation of β-alanine and pantoic acid. The enzymes responsible for the biosynthesis steps for the conversion to pantoic acid, to β-alanine and for the condensation to pantothenic acid are known. The metabolically active form of pantothenate is Coenzyme A, for which the biosynthesis proceeds in 5 enzymatic steps. Pantothenate, pyridoxal-5'-phosphate, cysteine and ATP are the precursors of Coenzyme A. These enzymes not only catalyze the formation of pantothenate, but also the production of (R)-pantoic acid, (R)-pantolacton, (R)-panthenol (provitamin B₅), pantotheine (and its derivatives) and coenzyme A.

[0035] Biotin biosynthesis from the precursor molecule pimeloyl-CoA in microorganisms has been studied in detail and several of the genes involved have been identified. Many of the corresponding proteins have been found to also be involved in Fe-cluster synthesis and are members of the nifS class of proteins. Lipoic acid is derived from octanoic acid, and serves as a coenzyme in energy metabolism, where it becomes part of the pyruvate dehydrogenase complex and the α-ketoglutarate dehydrogenase complex. The folates are a group of substances which are all derivatives of folic acid, which is turn is derived from L-glutamic acid, p-amino-benzoic acid and 6-methylpterin. The biosynthesis of folic acid and its derivatives, starting from the metabolism intermediates guanosine-5'-triphosphate (GTP), L-glutamic acid and p-amino-benzoic acid has been studied in detail in certain microorganisms.

[0036] Corrinoids (such as the cobalamines and particularly vitamin B₁₂) and porphyrines belong to a group of chemicals characterized by a tetrapyrrole ring system. The biosynthesis of vitamin B₁₂ is sufficiently complex that it has not yet been completely characterized, but many of the enzymes and substrates involved are now known. Nicotinic acid (nicotinate), and nicotinamide are pyridine derivatives which are also termed 'niacin'. Niacin is the precursor of the important coenzymes NAD (nicotinamide adenine dinucleotide) and NADP (nicotinamide adenine dinucleotide phosphate) and their reduced forms.

[0037] The large-scale production of these compounds has largely relied on cell-free chemical syntheses, though some of these chemicals have also been produced by large-scale culture of microorganisms, such as riboflavin, Vitamin B₆, pantothenate, and biotin. Only Vitamin B₁₂ is produced solely by fermentation, due to the complexity of its synthesis. In vitro methodologies require significant inputs of materials and time, often at great cost.

C. Purine, Pyrimidine, Nucleoside and Nucleotide Metabolism and Uses

[0038] Purine and pyrimidine metabolism genes and their corresponding proteins are important targets for the therapy

of tumor diseases and viral infections. The language "purine" or "pyrimidine" includes the nitrogenous bases which are constituents of nucleic acids, co-enzymes, and nucleotides. The term "nucleotide" includes the basic structural units of nucleic acid molecules, which are comprised of a nitrogenous base, a pentose sugar (in the case of RNA, the sugar is ribose; in the case of DNA, the sugar is D-deoxy-ribose), and phosphoric acid. The language "nucleoside" includes molecules which serve as precursors to nucleotides, but which are lacking the phosphoric acid moiety that nucleotides possess. By inhibiting the biosynthesis of these molecules, or their mobilization to form nucleic acid molecules, it is possible to inhibit RNA and DNA synthesis; by inhibiting this activity in a fashion targeted to cancerous cells, the ability of tumor cells to divide and replicate may be inhibited. Additionally, there are nucleotides which do not form nucleic acid molecules, but rather serve as energy stores (i.e., AMP) or as coenzymes (i.e., FAD and NAD).

[0039] Several publications have described the use of these chemicals for these medical indications, by influencing purine and/or pyrimidine metabolism (e.g. Christopherson, R. I. and Lyons, S. D. (1990) "Potent inhibitors of de novo pyrimidine and purine biosynthesis as chemotherapeutic agents." *Med. Res. Reviews* 10: 505-548). Studies of enzymes involved in purine and pyrimidine metabolism have been focused on the development of new drugs which can be used, for example, as immunosuppressants or anti-proliferants (Smith, J. L., (1995) "Enzymes in nucleotide synthesis." *Curr. Opin. Struct. Biol.* 5: 752-757; (1995) *Biochem Soc. Transact.* 23: 877-902). However, purine and pyrimidine bases, nucleosides and nucleotides have other utilities: as intermediates in the biosynthesis of several fine chemicals (e.g., thiamine, S-adenosyl-methionine, folates, or riboflavin), as energy carriers for the cell (e.g., ATP or GTP), and for chemicals themselves, commonly used as flavor enhancers (e.g., IMP or GMP) or for several medicinal applications (see, for example, Kuninaka, A. (1996) *Nucleotides and Related Compounds in Biotechnology* vol. 6, Rehm et al., eds. VCH: Weinheim, p. 561-612). Also, enzymes involved in purine, pyrimidine, nucleoside, or nucleotide metabolism are increasingly serving as targets against which chemicals for crop protection, including fungicides, herbicides and insecticides, are developed.

[0040] The metabolism of these compounds in bacteria has been characterized (for reviews see, for example, Zalkin, H. and Dixon, J. E. (1992) "de novo purine nucleotide biosynthesis", in: *Progress in Nucleic Acid Research and Molecular Biology*, vol. 42, Academic Press:, p. 259-287; and Michal, G. (1999) "Nucleotides and Nucleosides", Chapter 8 in: *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, Wiley: New York). Purine metabolism has been the subject of intensive research, and is essential to the normal functioning of the cell. Impaired purine metabolism in higher animals can cause severe disease, such as gout. Purine nucleotides are synthesized from ribose-5-phosphate, in a series of steps through the intermediate compound inosine-5'-phosphate (IMP), resulting in the production of guanosine-5'-monophosphate (GMP) or adenosine-5'-monophosphate (AMP), from which the triphosphate forms utilized as nucleotides are readily formed. These compounds are also utilized as energy stores, so their degradation provides energy for many different biochemical processes in the cell. Pyrimidine biosynthesis proceeds by the formation of uridine-5'-monophosphate

(UMP) from ribose-5-phosphate. UMP, in turn, is converted to cytidine-5'-triphosphate (CTP). The deoxy- forms of all of these nucleotides are produced in a one step reduction reaction from the diphosphate ribose form of the nucleotide to the diphosphate deoxyribose form of the nucleotide. Upon phosphorylation, these molecules are able to participate in DNA synthesis.

D. Trehalose Metabolism and Uses

[0041] Trehalose consists of two glucose molecules, bound in α , α -1,1 linkage. It is commonly used in the food industry as a sweetener, an additive for dried or frozen foods, and in beverages. However, it also has applications in the pharmaceutical, cosmetics and biotechnology industries (see, for example, Nishimoto et al., (1998) U.S. Pat. No. 5,759,610; Singer, M. A. and Lindquist, S. (1998) *Trends Biotech.* 16: 460-467; Paiva, C. L. A. and Panek, A. D. (1996) *Biotech. Ann. Rev.* 2: 293-314; and Shiosaka, M. (1997) *J. Japan* 172: 97-102). Trehalose is produced by enzymes from many microorganisms and is naturally released into the surrounding medium, from which it can be collected using methods known in the art.

II. Membrane Biosynthesis and Transmembrane Transport

[0042] Cellular membranes serve a variety of functions in a cell. First and foremost, a membrane differentiates the contents of a cell from the surrounding environment, thus giving integrity to the cell. Membranes may also serve as barriers to the influx of hazardous or unwanted compounds, and also to the efflux of desired compounds. Cellular membranes are by nature impervious to the unfacilitated diffusion of hydrophilic compounds such as proteins, water molecules and ions due to their structure: a bilayer of lipid molecules in which the polar head groups face outwards (towards the exterior and interior of the cell, respectively) and the non-polar tails face inwards at the center of the bilayer, forming a hydrophobic core (for a general review of membrane structure and function, see Gennis, R. B. (1989) *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg). This barrier enables cells to maintain a relatively higher concentration of desired compounds and a relatively lower concentration of undesired compounds than are contained within the surrounding medium, since the diffusion of these compounds is effectively blocked by the membrane. However, the membrane also presents an effective barrier to the import of desired compounds and the export of waste molecules. To overcome this difficulty, cellular membranes incorporate many kinds of transporter proteins which are able to facilitate the transmembrane transport of different kinds of compounds. There are two general classes of these transport proteins: pores or channels and transporters. The former are integral membrane proteins, sometimes complexes of proteins, which form a regulated hole through the membrane. This regulation, or 'gating' is generally specific to the molecules to be transported by the pore or channel, rendering these transmembrane constructs selectively permeable to a specific class of substrates; for example, a potassium channel is constructed such that only ions having a like charge and size to that of potassium may pass through. Channel and pore proteins tend to have discrete hydrophobic and hydrophilic domains, such that the hydrophobic face of the protein may associate with the interior of the membrane while the hydrophilic face lines the interior of the channel, thus providing a sheltered hydrophilic environment through

which the selected hydrophilic molecule may pass. Many such pores/channels are known in the art, including those for potassium, calcium, sodium, and chloride ions.

[0043] This pore and channel-mediated system of facilitated diffusion is limited to very small molecules, such as ions, because pores or channels large enough to permit the passage of whole proteins by facilitated diffusion would be unable to prevent the passage of smaller hydrophilic molecules as well. Transport of molecules by this process is sometimes termed 'facilitated diffusion' since the driving force of a concentration gradient is required for the transport to occur. Permeases also permit facilitated diffusion of larger molecules, such as glucose or other sugars, into the cell when the concentration of these molecules on one side of the membrane is greater than that on the other (also called 'uniport'). In contrast to pores or channels, these integral membrane proteins (often having between 6-14 membrane-spanning β -helices) do not form open channels through the membrane, but rather bind to the target molecule at the surface of the membrane and then undergo a conformational shift such that the target molecule is released on the opposite side of the membrane.

[0044] However, cells frequently require the import or export of molecules against the existing concentration gradient ('active transport'), a situation in which facilitated diffusion cannot occur. There are two general mechanisms used by cells for such membrane transport: symport or antiport, and energy-coupled transport such as that mediated by the ABC transporters. Symport and antiport systems couple the movement of two different molecules across the membrane (via permeases having two separate binding sites for the two different molecules); in symport, both molecules are transported in the same direction, while in antiport, one molecule is imported while the other is exported. This is possible energetically because one of the two molecules moves in accordance with a concentration gradient, and this energetically favorable event is permitted only upon concomitant movement of a desired compound against the prevailing concentration gradient. Single molecules may be transported across the membrane against the concentration gradient in an energy-driven process, such as that utilized by the ABC transporters. In this system, the transport protein located in the membrane has an ATP-binding cassette; upon binding of the target molecule, the ATP is converted to ADP+Pi, and the resulting release of energy is used to drive the movement of the target molecule to the opposite face of the membrane, facilitated by the transporter. For more detailed descriptions of all of these transport systems, see: Bamberg, E. et al., (1993) "Charge transport of ion pumps on lipid bilayer membranes", *Q. Rev. Biophys.* 26: 1-25; Findlay, J. B. C. (1991) "Structure and function in membrane transport systems", *Curr. Opin. Struct. Biol.* 1:804-810; Higgins, C. F. (1992) "ABC transporters from microorganisms to man", *Ann. Rev. Cell Biol.* 8: 67-113; Gennis, R. B. (1989) "Pores, Channels and Transporters", in: *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg, p. 270-322; and Nikaido, H. and Saier, H. (1992) "Transport proteins in bacteria: common themes in their design", *Science* 258: 936-942, and references contained within each of these references.

[0045] The synthesis of membranes is a well-characterized process involving a number of components, the most important of which are lipid molecules. Lipid synthesis may

be divided into two parts: the synthesis of fatty acids and their attachment to sn-glycerol-3-phosphate, and the addition or modification of a polar head group. Typical lipids utilized in bacterial membranes include phospholipids, glycolipids, sphingolipids, and phosphoglycerides. Fatty acid synthesis begins with the conversion of acetyl CoA either to malonyl CoA by acetyl CoA carboxylase, or to acetyl-ACP by acetyltransacylase. Following a condensation reaction, these two product molecules together form acetoacetyl-ACP, which is converted by a series of condensation, reduction and dehydration reactions to yield a saturated fatty acid molecule having a desired chain length. The production of unsaturated fatty acids from such molecules is catalyzed by specific desaturases either aerobically, with the help of molecular oxygen, or anaerobically (for reference on fatty acid synthesis, see F. C. Neidhardt et al. (1996) *E. coli* and *Salmonella*. ASM Press: Washington, D.C., p. 612-636 and references contained therein; Lengeler et al. (eds) (1999) *Biology of Prokaryotes*. Thieme: Stuttgart, New York, and references contained therein; and Magnuson, K. et al., (1993) *Microbiological Reviews* 57: 522-542, and references contained therein). The cyclopropane fatty acids (CFA) are synthesized by a specific CFA-synthase using SAM as a cosubstrate. Branched chain fatty acids are synthesized from branched chain amino acids that are deaminated to yield branched chain 2-oxo-acids (see Lengeler et al., eds. (1999) *Biology of Prokaryotes*. Thieme: Stuttgart, New York, and references contained therein). Another essential step in lipid synthesis is the transfer of fatty acids onto the polar head groups by, for example, glycerol-phosphate-acyltransferases. The combination of various precursor molecules and biosynthetic enzymes results in the production of different fatty acid molecules, which has a profound effect on the composition of the membrane.

III. Elements and Methods of the Invention

[0046] The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as MCT nucleic acid and protein molecules, which control the production of cellular membranes in *C. glutamicum* and govern the movement of molecules across such membranes. In one embodiment, the MCT molecules participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. In a preferred embodiment, the activity of the MCT molecules of the present invention to regulate membrane component production and membrane transport has an impact on the production of a desired fine chemical by this organism. In a particularly preferred embodiment, the MCT molecules of the invention are modulated in activity, such that the *C. glutamicum* metabolic pathways which the MCT proteins of the invention regulate are modulated in yield, production, and/or efficiency of production and the transport of compounds through the membranes is altered in efficiency, which either directly or indirectly modulates the yield, production, and/or efficiency of production of a desired fine chemical by *C. glutamicum*.

[0047] The language, "MCT protein" or "MCT polypeptide" includes proteins which participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Examples of MCT proteins include those encoded by the MCT genes set forth in Table

1 and Appendix A. The terms "MCT gene" or "MCT nucleic acid sequence" include nucleic acid sequences encoding an MCT protein, which consist of a coding region and also corresponding untranslated 5' and 3' sequence regions. Examples of MCT genes include those set forth in Table 1. The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, the desired fine chemical) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical). The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules, or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased. The terms "biosynthesis" or a "biosynthetic pathway" are art-recognized and include the synthesis of a compound, preferably an organic compound, by a cell from intermediate compounds in what may be a multistep and highly regulated process. The terms "degradation" or a "degradation pathway" are art-recognized and include the breakdown of a compound, preferably an organic compound, by a cell to degradation products (generally speaking, smaller or less complex molecules) in what may be a multistep and highly regulated process. The language "metabolism" is art-recognized and includes the totality of the biochemical reactions that take place in an organism. The metabolism of a particular compound, then, (e.g., the metabolism of an amino acid such as glycine) comprises the overall biosynthetic, modification, and degradation pathways in the cell related to this compound.

[0048] In another embodiment, the MCT molecules of the invention are capable of modulating the production of a desired molecule, such as a fine chemical, in a microorganism such as *C. glutamicum*. There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Those MCT proteins involved in the export of fine chemical molecules from the cell may be increased in number or activity such that greater quantities of these compounds are secreted to the extracellular medium, from which they are more readily recovered. Similarly, those MCT proteins involved in the import of nutrients necessary for the biosynthesis of one or more fine chemicals (e.g., phosphate, sulfate, nitrogen compounds, etc.) may be increased in number or activity such that these precursor, cofactor, or intermediate compounds are increased in concentration within the cell. Further, fatty acids and lipids themselves are desirable fine chemicals; by optimizing the activity or increasing the number of one or more MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

[0049] The mutagenesis of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, MCT proteins of the invention involved in the export of waste products may be increased in number or activity such that the normal metabolic wastes of the cell (possibly increased in quantity due to the overproduction of the desired fine chemical) are efficiently exported before they are able to damage nucleotides and proteins within the cell (which would decrease the viability of the cell) or to interfere with fine chemical biosynthetic pathways (which would decrease the yield, production, or efficiency of production of the desired fine chemical). Further, the relatively large intracellular quantities of the desired fine chemical may in itself be toxic to the cell, so by increasing the activity or number of transporters able to export this compound from the cell, one may increase the viability of the cell in culture, in turn leading to a greater number of cells in the culture producing the desired fine chemical. The MCT proteins of the invention may also be manipulated such that the relative amounts of different lipid and fatty acid molecules are produced. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, as well as the integrity of the cell, both of which have a profound effect on the production of fine chemicals from *C. glutamicum* in large-scale fermentative culture.

[0050] The isolated nucleic acid sequences of the invention are contained within the genome of a *Corynebacterium glutamicum* strain available through the American Type Culture Collection, given designation ATCC 13032. The nucleotide sequence of the isolated *C. glutamicum* MCT DNAs and the predicted amino acid sequences of the *C. glutamicum* MCT proteins are shown in Appendices A and B, respectively. Computational analyses were performed which classified and/or identified these nucleotide sequences as sequences which encode proteins involved in the metabolism of cellular membrane components or proteins involved in the transport of compounds across such membranes.

[0051] The present invention also pertains to proteins which have an amino acid sequence which is substantially homologous to an amino acid sequence of Appendix B. As used herein, a protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence is least about 50% homologous to the selected amino acid sequence, e.g., the entire selected amino acid sequence. A protein which has an amino acid sequence which is substantially homologous to a selected amino acid sequence can also be least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, or 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to the selected amino acid sequence.

[0052] The MCT protein or a biologically active portion or fragment thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or have one or more of the activities set forth in Table 1.

[0053] Various aspects of the invention are described in further detail in the following subsections:

A. Isolated Nucleic Acid Molecules

[0054] One aspect of the invention pertains to isolated nucleic acid molecules that encode MCT polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes or primers for the identification or amplification of MCT-encoding nucleic acid (e.g., MCT DNA). As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. This term also encompasses untranslated sequence located at both the 3' and 5' ends of the coding region of the gene: at least about 100 nucleotides of sequence upstream from the 5' end of the coding region and at least about 20 nucleotides of sequence downstream from the 3' end of the coding region of the gene. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MCT nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (e.g., a *C. glutamicum* cell). Moreover, an "isolated" nucleic acid molecule, such as a DNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized.

[0055] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having a nucleotide sequence of Appendix A, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, a *C. glutamicum* MCT DNA can be isolated from a *C. glutamicum* library using all or portion of one of the sequences of Appendix A as a hybridization probe and standard hybridization techniques (e.g., as described in Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). Moreover, a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this sequence (e.g., a nucleic acid molecule encompassing all or a portion of one of the sequences of Appendix A can be isolated by the polymerase chain reaction using oligonucleotide primers designed based upon this same sequence of Appendix A). For example, mRNA can be isolated from normal endothelial cells (e.g., by the guanidinium-thiocyanate extraction procedure of Chirgwin et al. (1979) *Biochemistry* 18: 5294-5299) and DNA can be prepared using reverse transcriptase (e.g., Moloney MLV reverse transcriptase, available from Gibco/BRL, Bethesda, Md.; or AMV reverse transcriptase, avail-

able from Seikagaku America, Inc., St. Petersburg, Fla.). Synthetic oligonucleotide primers for polymerase chain reaction amplification can be designed based upon one of the nucleotide sequences shown in Appendix A. A nucleic acid of the invention can be amplified using cDNA or, alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to an MCT nucleotide sequence can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

[0056] In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises one of the nucleotide sequences shown in Appendix A. The sequences of Appendix A correspond to the *Corynebacterium glutamicum* MCT DNAs of the invention. This DNA comprises sequences encoding MCT proteins (i.e., the “coding region”, indicated in each sequence in Appendix A), as well as 5' untranslated sequences and 3' untranslated sequences, also indicated in Appendix A. Alternatively, the nucleic acid molecule can comprise only the coding region of any of the sequences in Appendix A.

[0057] For the purposes of this application, it will be understood that each of the sequences set forth in Appendix A has an identifying RXA, RXN, RXS, or RXC number having the designation “RXA”, “RXN”, “RXS” or “RXC” followed by 5 digits (i.e., RXA02099, RXN03097, RXS00148, or RXC01748). Each of these sequences comprises up to three parts: a 5' upstream region, a coding region, and a downstream region. Each of these three regions is identified by the same RXA, RXN, RXS, or RXC designation to eliminate confusion. The recitation “one of the sequences in Appendix A”, then, refers to any of the sequences in Appendix A, which may be distinguished by their differing RXA, RXN, RXS, or RXC designations. The coding region of each of these sequences is translated into a corresponding amino acid sequence, which is set forth in Appendix B. The sequences of Appendix B are identified by the same RXA, RXN, RXS, or RXC designations as Appendix A, such that they can be readily correlated. For example, the amino acid sequences in Appendix B designated RXA02099, RXN03097, RXS00148, and RXC01748 are translations of the coding region of the nucleotide sequences of nucleic acid molecules RXA02099, RXN03097, RXS00148, and RXC01748, respectively, in Appendix A. Each of the RXA, RXN, RXS, and RXC nucleotide and amino acid sequences of the invention has also been assigned a SEQ ID NO, as indicated in Table 1. For example, as set forth in Table 1, the nucleotide sequence of RXA00104 is SEQ ID NO:5, and the amino acid sequence of RXA00104 is SEQ ID NO:6.

[0058] Several of the genes of the invention are “F-designated genes”. An F-designated gene includes those genes set forth in Table 1 which have an ‘F’ in front of the RXA, RXN, RXS, or RXC designation. For example, SEQ ID NO:11, designated, as indicated on Table 1, as “F RXA02581”, is an F-designated gene, as are SEQ ID NOs: 31, 33, and 43 (designated on Table 1 as “F RXA02487”, “F RXA02490”, and “F RXA02809”, respectively).

[0059] In one embodiment, the nucleic acid molecules of the present invention are not intended to include those

compiled in Table 2. In the case of the *dapD* gene, a sequence for this gene was published in Wehrmann, A., et al. (1998) *J. Bacteriol.* 180(12): 3159-3165. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

[0060] In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of one of the nucleotide sequences shown in Appendix A, or a portion thereof. A nucleic acid molecule which is complementary to one of the nucleotide sequences shown in Appendix A is one which is sufficiently complementary to one of the nucleotide sequences shown in Appendix A such that it can hybridize to one of the nucleotide sequences shown in Appendix A, thereby forming a stable duplex.

[0061] In still another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to a nucleotide sequence shown in Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited ranges, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In an additional preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to one of the nucleotide sequences shown in Appendix A, or a portion thereof.

[0062] Moreover, the nucleic acid molecule of the invention can comprise only a portion of the coding region of one of the sequences in Appendix A, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of an MCT protein. The nucleotide sequences determined from the cloning of the MCT genes from *C. glutamicum* allows for the generation of probes and primers designed for use in identifying and/or cloning MCT homologues in other cell types and organisms, as well as MCT homologues from other *Corynebacteria* or related species. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of a sense strand of one of the sequences set forth in Appendix A, an anti-sense sequence of one of the sequences set forth in Appendix A, or naturally occurring mutants thereof. Primers based on a nucleotide sequence of Appendix A can be used in PCR reactions to clone MCT homologues. Probes based on the MCT nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous

proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g. the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells which misexpress an MCT protein, such as by measuring a level of an MCT-encoding nucleic acid in a sample of cells, e.g., detecting MCT mRNA levels or determining whether a genomic MCT gene has been mutated or deleted.

[0063] In one embodiment, the nucleic acid molecule of the invention encodes a protein or portion thereof which includes an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. As used herein, the language “sufficiently homologous” refers to proteins or portions thereof which have amino acid sequences which include a minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain as an amino acid residue in one of the sequences of Appendix B) amino acid residues to an amino acid sequence of Appendix B such that the protein or portion thereof is able to participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. Protein members of such membrane component metabolic pathways or membrane transport systems, as described herein, may play a role in the production and secretion of one or more fine chemicals. Examples of such activities are also described herein. Thus, “the function of an MCT protein” contributes either directly or indirectly to the yield, production, and/or efficiency of production of one or more fine chemicals. Examples of MCT protein activities are set forth in Table 1.

[0064] In another embodiment, the protein is at least about 50-60%, preferably at least about 60-70%, and more preferably at least about 70-80%, 80-90%, 90-95%, and most preferably at least about 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B.

[0065] Portions of proteins encoded by the MCT nucleic acid molecules of the invention are preferably biologically active portions of one of the MCT proteins. As used herein, the term “biologically active portion of an MCT protein” is intended to include a portion, e.g., a domain/motif, of an MCT protein that participates in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has an activity as set forth in Table 1. To determine whether an MCT protein or a biologically active portion thereof can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, an assay of enzymatic activity may be performed. Such assay methods are well known to those of ordinary skill in the art, as detailed in Example 8 of the Exemplification.

[0066] Additional nucleic acid fragments encoding biologically active portions of an MCT protein can be prepared by isolating a portion of one of the sequences in Appendix

B, expressing the encoded portion of the MCT protein or peptide (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of the MCT protein or peptide.

[0067] The invention further encompasses nucleic acid molecules that differ from one of the nucleotide sequences shown in Appendix A (and portions thereof) due to degeneracy of the genetic code and thus encode the same MCT protein as that encoded by the nucleotide sequences shown in Appendix A. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in Appendix B. In a still further embodiment, the nucleic acid molecule of the invention encodes a full length *C. glutamicum* protein which is substantially homologous to an amino acid sequence of Appendix B (encoded by an open reading frame shown in Appendix A).

[0068] It will be understood by one of ordinary skill in the art that in one embodiment the sequences of the invention are not meant to include the sequences of the prior art, such as those Genbank sequences set forth in Tables 2 or 4 which were available prior to the present invention. In one embodiment, the invention includes nucleotide and amino acid sequences having a percent identity to a nucleotide or amino acid sequence of the invention which is greater than that of a sequence of the prior art (e.g., a Genbank sequence (or the protein encoded by such a sequence) set forth in Tables 2 or 4). For example, the invention includes a nucleotide sequence which is greater than and/or at least 38% identical to the nucleotide sequence designated RXA01420 (SEQ ID NO:7), a nucleotide sequence which is greater than and/or at least 43% identical to the nucleotide sequence designated RXA00104 (SEQ ID NO:5), and a nucleotide sequence which is greater than and/or at least 45% identical to the nucleotide sequence designated RXA02173 (SEQ ID NO:25). One of ordinary skill in the art would be able to calculate the lower threshold of percent identity for any given sequence of the invention by examining the GAP-calculated percent identity scores set forth in Table 4 for each of the three top hits for the given sequence, and by subtracting the highest GAP-calculated percent identity from 100 percent. One of ordinary skill in the art will also appreciate that nucleic acid and amino acid sequences having percent identities greater than the lower threshold so calculated (e.g., at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more identical) are also encompassed by the invention.

[0069] In addition to the *C. glutamicum* MCT nucleotide sequences shown in Appendix A, it will be appreciated by one of ordinary skill in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of MCT proteins may exist within a population (e.g., the *C. glutamicum* population). Such genetic polymorphism in the MCT gene may exist among individuals within a population due to natural variation. As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules comprising an open reading frame encoding an MCT pro-

tein, preferably a *C. glutamicum* MCT protein. Such natural variations can typically result in 1-5% variance in the nucleotide sequence of the MCT gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in MCT that are the result of natural variation and that do not alter the functional activity of MCT proteins are intended to be within the scope of the invention.

[0070] Nucleic acid molecules corresponding to natural variants and non-*C. glutamicum* homologues of the *C. glutamicum* MCT DNA of the invention can be isolated based on their homology to the *C. glutamicum* MCT nucleic acid disclosed herein using the *C. glutamicum* DNA, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising a nucleotide sequence of Appendix A. In other embodiments, the nucleic acid is at least 30, 50, 100, 250 or more nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those of ordinary skill in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 50-65° C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of Appendix A corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). In one embodiment, the nucleic acid encodes a natural *C. glutamicum* MCT protein.

[0071] In addition to naturally-occurring variants of the MCT sequence that may exist in the population, one of ordinary skill in the art will further appreciate that changes can be introduced by mutation into a nucleotide sequence of Appendix A, thereby leading to changes in the amino acid sequence of the encoded MCT protein, without altering the functional ability of the MCT protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in a sequence of Appendix A. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of one of the MCT proteins (Appendix B) without altering the activity of said MCT protein, whereas an "essential" amino acid residue is required for MCT protein activity. Other amino acid residues, however, (e.g., those that are not conserved or only semi-conserved in the domain having MCT activity) may not be essential for activity and thus are likely to be amenable to alteration without altering MCT activity.

[0072] Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding MCT proteins that contain changes in amino acid residues that are not essential for MCT activity. Such MCT proteins differ in amino acid sequence from a sequence contained in Appendix B yet retain at least one of the MCT activities described herein. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 50% homologous to an amino acid sequence of Appendix B and is capable of participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more activities set forth in Table 1. Preferably, the protein encoded by the nucleic acid molecule is at least about 50-60% homologous to one of the sequences in Appendix B, more preferably at least about 60-70% homologous to one of the sequences in Appendix B, even more preferably at least about 70-80%, 80-90%, 90-95% homologous to one of the sequences in Appendix B, and most preferably at least about 96%, 97%, 98%, or 99% homologous to one of the sequences in Appendix B.

[0073] To determine the percent homology of two amino acid sequences (e.g., one of the sequences of Appendix B and a mutant form thereof) or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence (e.g., one of the sequences of Appendix B) is occupied by the same amino acid residue or nucleotide as the corresponding position in the other sequence (e.g., a mutant form of the sequence selected from Appendix B), then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = # of identical positions/total# of positions×100).

[0074] An isolated nucleic acid molecule encoding an MCT protein homologous to a protein sequence of Appendix B can be created by introducing one or more nucleotide substitutions, additions or deletions into a nucleotide sequence of Appendix A such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into one of the sequences of Appendix A by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine,

phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in an MCT protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MCT coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for an MCT activity described herein to identify mutants that retain MCT activity. Following mutagenesis of one of the sequences of Appendix A, the encoded protein can be expressed recombinantly and the activity of the protein can be determined using, for example, assays described herein (see Example 8 of the Exemplification).

[0075] In addition to the nucleic acid molecules encoding MCT proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire MCT coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MCT protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the entire coding region of SEQ ID NO:5 (RXA00104 in Appendix A) comprises nucleotides 1 to 756). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding MCT. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

[0076] Given the coding strand sequences encoding MCT disclosed herein (e.g., the sequences set forth in Appendix A), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MCT mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of MCT mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MCT mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-

thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0077] The antisense nucleic acid molecules of the invention are typically administered to a cell or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MCT protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. The antisense molecule can be modified such that it specifically binds to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecule to a peptide or an antibody which binds to a cell surface receptor or antigen. The antisense nucleic acid molecule can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong prokaryotic, viral, or eukaryotic promoter are preferred.

[0078] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

[0079] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave MCT mRNA transcripts to thereby inhibit translation of MCT mRNA. A ribozyme having specificity for an MCT-encoding nucleic acid can be designed based upon the nucleotide sequence of an MCT

DNA disclosed herein (i.e., SEQ ID NO. 5 (RXA00104) in Appendix A)). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MCT-encoding mRNA. See, e.g., Cech et al. U.S. Pat. No. 4,987,071 and Cech et al. U.S. Pat. No. 5,116,742. Alternatively, MCT mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J. W. (1993) *Science* 261:1411-1418.

[0080] Alternatively, MCT gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of an MCT nucleotide sequence (e.g., an MCT promoter and/or enhancers) to form triple helical structures that prevent transcription of an MCT gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L. J. (1992) *Bioassays* 14(12):807-15.

B. Recombinant Expression Vectors and Host Cells

[0081] Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MCT protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

[0082] The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements

(e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells. Preferred regulatory sequences are, for example, promoters such as cos-, tac-, trp-, tet-, trp-tet-, lpp-, lac-, lpp-lac-, lacI^q-, T7-, T5-, T3-, gal-, trc-, ara-, SP6-, amy, SPO2, λ -P_R-or λ P_L, which are used preferably in bacteria. Additional regulatory sequences are, for example, promoters from yeasts and fungi, such as ADC1, MF α , AC, P-60, CYC1, GAPDH, TEF, rp28, ADH, promoters from plants such as CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or ubiquitin- or phaseolin-promoters. It is also possible to use artificial promoters. It will be appreciated by one of ordinary skill in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., MCT proteins, mutant forms of MCT proteins, fusion proteins, etc.).

[0083] The recombinant expression vectors of the invention can be designed for expression of MCT proteins in prokaryotic or eukaryotic cells. For example, MCT genes can be expressed in bacterial cells such as *C. glutamicum*, insect cells (using baculovirus expression vectors), yeast and other fungal cells (see Romanos, M. A. et al. (1992) "Foreign gene expression in yeast: a review", *Yeast* 8: 423-488; van den Hondel, C.A.M.J.J. et al. (1991) "Heterologous gene expression in filamentous fungi" in: *More Gene Manipulations in Fungi*, J. W. Bennet & L. L. Lasure, eds., p. 396-428; Academic Press: San Diego; and van den Hondel, C.A.M.J.J. & Punt, P. J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, Peberdy, J. F. et al., eds., p. 1-28, Cambridge University Press: Cambridge), algae and multicellular plant cells (see Schmidt, R. and Willmitzer, L. (1988) High efficiency *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* leaf and cotyledon explants" *Plant Cell Rep.*: 583-586), or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0084] Expression of proteins in prokaryotes is most often carried out with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein but also to the C-terminus or fused within suitable regions in the proteins. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable sepa-

ration of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.

[0085] Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D. B. and Johnson, K. S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. In one embodiment, the coding sequence of the MCT protein is cloned into a pGEX expression vector to create a vector encoding a fusion protein comprising, from the N-terminus to the C-terminus, GST-thrombin cleavage site-X protein. The fusion protein can be purified by affinity chromatography using glutathione-agarose resin. Recombinant MCT protein unfused to GST can be recovered by cleavage of the fusion protein with thrombin.

[0086] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315), pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pLc236, pMBL24, pLG200, pUR290, pIN-III113-B1, λ gt11, pBdC1, and pET 11d (Studier et al, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990) 60-89 ; and Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: New York ISBN 0 444 904018). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gnl0-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gnl). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gnl gene under the transcriptional control of the lacUV 5 promoter. For transformation of other varieties of bacteria, appropriate vectors may be selected. For example, the plasmids pIJ101, pIJ364, pIJ702 and pIJ361 are known to be useful in transforming *Streptomyces*, while plasmids pUB110, pC194, or pBD214 are suited for transformation of *Bacillus* species. Several plasmids of use in the transfer of genetic information into *Corynebacterium* include pHM1519, pBL1, pSA77, or pAJ667 (Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: N.Y. ISBN 0 444 904018).

[0087] One strategy to maximize recombinant protein expression is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the bacterium chosen for expression, such as *C. glutamicum* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0088] In another embodiment, the MCT protein expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), 2 μ , pAG-1, Yep6, Yep13, pEMBLYe23, pMFa (Kurjan and Herskowitz,

(1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, Calif.). Vectors and methods for the construction of vectors appropriate for use in other fungi, such as the filamentous fungi, include those detailed in: van den Hondel, C.A.M.J.J. & Punt, P. J. (1991) "Gene transfer systems and vector development for filamentous fungi, in: *Applied Molecular Genetics of Fungi*, J. F. Peberdy, et al., eds., p. 1-28, Cambridge University Press: Cambridge, and Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: N.Y. (ISBN 0 444 904018).

[0089] Alternatively, the MCT proteins of the invention can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

[0090] In another embodiment, the MCT proteins of the invention may be expressed in unicellular plant cells (such as algae) or in plant cells from higher plants (e.g., the spermatophytes, such as crop plants). Examples of plant expression vectors include those detailed in: Becker, D., Kemper, E., Schell, J. and Masterson, R. (1992) "New plant binary vectors with selectable markers located proximal to the left border", *Plant Mol. Biol.* 20: 1195-1197; and Bevan, M. W. (1984) "Binary *Agrobacterium* vectors for plant transformation", *Nuc. Acid Res.* 12: 8711-8721, and include pLGV23, pGHlac+, pBIN19, pAK2004, and pDH51 (Pouwels et al., eds. (1985) *Cloning Vectors*. Elsevier: N.Y. ISBN 0 444 904018).

[0091] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0092] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and

mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[0093] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to MCT mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., *Antisense RNA as a molecular tool for genetic analysis, Reviews—Trends in Genetics, Vol. 1(1)* 1986.

[0094] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms “host cell” and “recombinant host cell” are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0095] A host cell can be any prokaryotic or eukaryotic cell. For example, an MCT protein can be expressed in bacterial cells such as *C. glutamicum*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to one of ordinary skill in the art. Microorganisms related to *Corynebacterium glutamicum* which may be conveniently used as host cells for the nucleic acid and protein molecules of the invention are set forth in Table 3.

[0096] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection”, “conjugation” and “transduction” are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., linear DNA or RNA (e.g., a linearized vector or a gene construct alone without a vector) or nucleic acid in the form of a vector (e.g., a plasmid, phage, phasmid, phagemid, transposon or other DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, natural competence, chemical-me-

diated transfer, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

[0097] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding an MCT protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by, for example, drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0098] To create a homologous recombinant microorganism, a vector is prepared which contains at least a portion of an MCT gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the MCT gene. Preferably, this MCT gene is a *Corynebacterium glutamicum* MCT gene, but it can be a homologue from a related bacterium or even from a mammalian, yeast, or insect source. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous MCT gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a “knock out” vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous MCT gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous MCT protein). In the homologous recombination vector, the altered portion of the MCT gene is flanked at its 5' and 3' ends by additional nucleic acid of the MCT gene to allow for homologous recombination to occur between the exogenous MCT gene carried by the vector and an endogenous MCT gene in a microorganism. The additional flanking MCT nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K. R., and Capecchi, M. R. (1987) *Cell* 51: 503 for a description of homologous recombination vectors). The vector is introduced into a microorganism (e.g., by electroporation) and cells in which the introduced MCT gene has homologously recombined with the endogenous MCT gene are selected, using art-known techniques.

[0099] In another embodiment, recombinant microorganisms can be produced which contain selected systems which allow for regulated expression of the introduced gene. For example, inclusion of an MCT gene on a vector placing it under control of the lac operon permits expression of the MCT gene only in the presence of IPTG. Such regulatory systems are well known in the art.

[0100] In another embodiment, an endogenous MCT gene in a host cell is disrupted (e.g., by homologous recombina-

tion or other genetic means known in the art) such that expression of its protein product does not occur. In another embodiment, an endogenous or introduced MCT gene in a host cell has been altered by one or more point mutations, deletions, or inversions, but still encodes a functional MCT protein. In still another embodiment, one or more of the regulatory regions (e.g., a promoter, repressor, or inducer) of an MCT gene in a microorganism has been altered (e.g., by deletion, truncation, inversion, or point mutation) such that the expression of the MCT gene is modulated. One of ordinary skill in the art will appreciate that host cells containing more than one of the described MCT gene and protein modifications may be readily produced using the methods of the invention, and are meant to be included in the present invention.

[0101] A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) an MCT protein. Accordingly, the invention further provides methods for producing MCT proteins using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding an MCT protein has been introduced, or into which genome has been introduced a gene encoding a wild-type or altered MCT protein) in a suitable medium until MCT protein is produced. In another embodiment, the method further comprises isolating MCT proteins from the medium or the host cell.

C. Isolated MCT Proteins

[0102] Another aspect of the invention pertains to isolated MCT proteins, and biologically active portions thereof. An “isolated” or “purified” protein or biologically active portion thereof is substantially free of cellular material when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. The language “substantially free of cellular material” includes preparations of MCT protein in which the protein is separated from cellular components of the cells in which it is naturally or recombinantly produced. In one embodiment, the language “substantially free of cellular material” includes preparations of MCT protein having less than about 30% (by dry weight) of non-MCT protein (also referred to herein as a “contaminating protein”), more preferably less than about 20% of non-MCT protein, still more preferably less than about 10% of non-MCT protein, and most preferably less than about 5% non-MCT protein. When the MCT protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The language “substantially free of chemical precursors or other chemicals” includes preparations of MCT protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language “substantially free of chemical precursors or other chemicals” includes preparations of MCT protein having less than about 30% (by dry weight) of chemical precursors or non-MCT chemicals, more preferably less than about 20% chemical precursors or non-MCT chemicals, still more preferably less than about 10% chemical precursors or non-MCT chemicals, and most preferably less than about 5% chemical

precursors or non-MCT chemicals. In preferred embodiments, isolated proteins or biologically active portions thereof lack contaminating proteins from the same organism from which the MCT protein is derived. Typically, such proteins are produced by recombinant expression of, for example, a *C. glutamicum* MCT protein in a microorganism such as *C. glutamicum*.

[0103] An isolated MCT protein or a portion thereof of the invention can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or has one or more of the activities set forth in Table 1. In preferred embodiments, the protein or portion thereof comprises an amino acid sequence which is sufficiently homologous to an amino acid sequence of Appendix B such that the protein or portion thereof maintains the ability participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes. The portion of the protein is preferably a biologically active portion as described herein. In another preferred embodiment, an MCT protein of the invention has an amino acid sequence shown in Appendix B. In yet another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A. In still another preferred embodiment, the MCT protein has an amino acid sequence which is encoded by a nucleotide sequence that is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%, preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to one of the nucleic acid sequences of Appendix A, or a portion thereof. Ranges and identity values intermediate to the above-recited values, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. The preferred MCT proteins of the present invention also preferably possess at least one of the MCT activities described herein. For example, a preferred MCT protein of the present invention includes an amino acid sequence encoded by a nucleotide sequence which hybridizes, e.g., hybridizes under stringent conditions, to a nucleotide sequence of Appendix A, and which can participate in the metabolism of compounds necessary for the construction of cellular membranes in *C. glutamicum*, or in the transport of molecules across these membranes, or which has one or more of the activities set forth in Table 1.

[0104] In other embodiments, the MCT protein is substantially homologous to an amino acid sequence of Appendix B and retains the functional activity of the protein of one of the sequences of Appendix B yet differs in amino acid sequence due to natural variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the MCT protein is a protein which comprises an amino acid sequence which is at least about 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, or 60%,

preferably at least about 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, or 70%, more preferably at least about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, or 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, or 90%, or 91%, 92%, 93%, 94%, and even more preferably at least about 95%, 96%, 97%, 98%, 99% or more homologous to an entire amino acid sequence of Appendix B and which has at least one of the MCT activities described herein. Ranges and identity values intermediate to the above-recited values, (e.g., 70-90% identical or 80-95% identical) are also intended to be encompassed by the present invention. For example, ranges of identity values using a combination of any of the above values recited as upper and/or lower limits are intended to be included. In another embodiment, the invention pertains to a full length *C. glutamicum* protein which is substantially homologous to an entire amino acid sequence of Appendix B.

[0105] Biologically active portions of an MCT protein include peptides comprising amino acid sequences derived from the amino acid sequence of an MCT protein, e.g., the amino acid sequence shown in Appendix B or the amino acid sequence of a protein homologous to an MCT protein, which include fewer amino acids than a full length MCT protein or the full length protein which is homologous to an MCT protein, and exhibit at least one activity of an MCT protein. Typically, biologically active portions (peptides, e.g., peptides which are, for example, 5, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) comprise a domain or motif with at least one activity of an MCT protein. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the activities described herein. Preferably, the biologically active portions of an MCT protein include one or more selected domains/motifs or portions thereof having biological activity.

[0106] MCT proteins are preferably produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the protein is cloned into an expression vector (as described above), the expression vector is introduced into a host cell (as described above) and the MCT protein is expressed in the host cell. The MCT protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Alternative to recombinant expression, an MCT protein, polypeptide, or peptide can be synthesized chemically using standard peptide synthesis techniques. Moreover, native MCT protein can be isolated from cells (e.g., endothelial cells), for example using an anti-MCT antibody, which can be produced by standard techniques utilizing an MCT protein or fragment thereof of this invention.

[0107] The invention also provides MCT chimeric or fusion proteins. As used herein, an MCT "chimeric protein" or "fusion protein" comprises an MCT polypeptide operatively linked to a non-MCT polypeptide. An "MCT polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an MCT protein, whereas a "non-MCT polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the MCT protein, e.g., a protein which is different from the MCT protein and which is derived from the same or a different organism. Within the fusion protein, the term "operatively linked" is intended to indicate that the

MCT polypeptide and the non-MCT polypeptide are fused in-frame to each other. The non-MCT polypeptide can be fused to the N-terminus or C-terminus of the MCT polypeptide. For example, in one embodiment the fusion protein is a GST-MCT fusion protein in which the MCT sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant MCT proteins. In another embodiment, the fusion protein is an MCT protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of an MCT protein can be increased through use of a heterologous signal sequence.

[0108] Preferably, an MCT chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An MCT-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MCT protein.

[0109] Homologues of the MCT protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the MCT protein. As used herein, the term "homologue" refers to a variant form of the MCT protein which acts as an agonist or antagonist of the activity of the MCT protein. An agonist of the MCT protein can retain substantially the same, or a subset, of the biological activities of the MCT protein. An antagonist of the MCT protein can inhibit one or more of the activities of the naturally occurring form of the MCT protein, by, for example, competitively binding to a downstream or upstream member of the cell membrane component metabolic cascade which includes the MCT protein, or by binding to an MCT protein which mediates transport of compounds across such membranes, thereby preventing translocation from taking place.

[0110] In an alternative embodiment, homologues of the MCT protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the MCT protein for MCT protein agonist or antagonist activity. In one embodiment, a variegated library of MCT variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MCT variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MCT sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of MCT

sequences therein. There are a variety of methods which can be used to produce libraries of potential MCT homologues from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential MCT sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S. A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

[0111] In addition, libraries of fragments of the MCT protein coding can be used to generate a variegated population of MCT fragments for screening and subsequent selection of homologues of an MCT protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MCT coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/anti-sense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the MCT protein.

[0112] Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of MCT homologues. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MCT homologues (Arkin and Yourvan (1992) *PNAS* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

[0113] In another embodiment, cell based assays can be exploited to analyze a variegated MCT library, using methods well known in the art.

D. Uses and Methods of the Invention

[0114] The nucleic acid molecules, proteins, protein homologues, fusion proteins, primers, vectors, and host cells described herein can be used in one or more of the following methods: identification of *C. glutamicum* and related organisms; mapping of genomes of organisms related to *C. glutamicum*; identification and localization of *C. glutamicum* sequences of interest; evolutionary studies; determination of MCT protein regions required for function; modulation of an MCT protein activity; modulation of the

metabolism of one or more cell membrane components; modulation of the transmembrane transport of one or more compounds; and modulation of cellular production of a desired compound, such as a fine chemical.

[0115] The MCT nucleic acid molecules of the invention have a variety of uses. First, they may be used to identify an organism as being *Corynebacterium glutamicum* or a close relative thereof. Also, they may be used to identify the presence of *C. glutamicum* or a relative thereof in a mixed population of microorganisms. The invention provides the nucleic acid sequences of a number of *C. glutamicum* genes; by probing the extracted genomic DNA of a culture of a unique or mixed population of microorganisms under stringent conditions with a probe spanning a region of a *C. glutamicum* gene which is unique to this organism, one can ascertain whether this organism is present.

[0116] Although *Corynebacterium glutamicum* itself is nonpathogenic, it is related to pathogenic species, such as *Corynebacterium diphtheriae*. *Corynebacterium diphtheriae* is the causative agent of diphtheria, a rapidly developing, acute, febrile infection which involves both local and systemic pathology. In this disease, a local lesion develops in the upper respiratory tract and involves necrotic injury to epithelial cells; the bacilli secrete toxin which is disseminated through this lesion to distal susceptible tissues of the body. Degenerative changes brought about by the inhibition of protein synthesis in these tissues, which include heart, muscle, peripheral nerves, adrenals, kidneys, liver and spleen, result in the systemic pathology of the disease. Diphtheria continues to have high incidence in many parts of the world, including Africa, Asia, Eastern Europe and the independent states of the former Soviet Union. An ongoing epidemic of diphtheria in the latter two regions has resulted in at least 5,000 deaths since 1990. In one embodiment, the invention provides a method of identifying the presence or activity of *Corynebacterium diphtheriae* in a subject. This method includes detection of one or more of the nucleic acid or amino acid sequences of the invention (e.g., the sequences set forth in Appendix A or Appendix B) in a subject, thereby detecting the presence or activity of *Corynebacterium diphtheriae* in the subject. *C. glutamicum* and *C. diphtheriae* are related bacteria, and many of the nucleic acid and protein molecules in *C. glutamicum* are homologous to *C. diphtheriae* nucleic acid and protein molecules, and can therefore be used to detect *C. diphtheriae* in a subject.

[0117] The nucleic acid and protein molecules of the invention may also serve as markers for specific regions of the genome. This has utility not only in the mapping of the genome, but also for functional studies of *C. glutamicum* proteins. For example, to identify the region of the genome to which a particular *C. glutamicum* DNA-binding protein binds, the *C. glutamicum* genome could be digested, and the fragments incubated with the DNA-binding protein. Those which bind the protein may be additionally probed with the nucleic acid molecules of the invention, preferably with readily detectable labels; binding of such a nucleic acid molecule to the genome fragment enables the localization of the fragment to the genome map of *C. glutamicum*, and, when performed multiple times with different enzymes, facilitates a rapid determination of the nucleic acid sequence to which the protein binds. Further, the nucleic acid molecules of the invention may be sufficiently homologous to the sequences of related species such that these nucleic acid

molecules may serve as markers for the construction of a genomic map in related bacteria, such as *Brevibacterium lactofermentum*.

[0118] The MCT nucleic acid molecules of the invention are also useful for evolutionary and protein structural studies. The metabolic and transport processes in which the molecules of the invention participate are utilized by a wide variety of prokaryotic and eukaryotic cells; by comparing the sequences of the nucleic acid molecules of the present invention to those encoding similar enzymes from other organisms, the evolutionary relatedness of the organisms can be assessed. Similarly, such a comparison permits an assessment of which regions of the sequence are conserved and which are not, which may aid in determining those regions of the protein which are essential for the functioning of the enzyme. This type of determination is of value for protein engineering studies and may give an indication of what the protein can tolerate in terms of mutagenesis without losing function.

[0119] Manipulation of the MCT nucleic acid molecules of the invention may result in the production of MCT proteins having functional differences from the wild-type MCT proteins. These proteins may be improved in efficiency or activity, may be present in greater numbers in the cell than is usual, or may be decreased in efficiency or activity.

[0120] The invention provides methods for screening molecules which modulate the activity of an MCT protein, either by interacting with the protein itself or a substrate or binding partner of the MCT protein, or by modulating the transcription or translation of an MCT nucleic acid molecule of the invention. In such methods, a microorganism expressing one or more MCT proteins of the invention is contacted with one or more test compounds, and the effect of each test compound on the activity or level of expression of the MCT protein is assessed.

[0121] There are a number of mechanisms by which the alteration of an MCT protein of the invention may directly affect the yield, production, and/or efficiency of production of a fine chemical from a *C. glutamicum* strain incorporating such an altered protein. Recovery of fine chemical compounds from large-scale cultures of *C. glutamicum* is significantly improved if *C. glutamicum* secretes the desired compounds, since such compounds may be readily purified from the culture medium (as opposed to extracted from the mass of *C. glutamicum* cells). By either increasing the number or the activity of transporter molecules which export fine chemicals from the cell, it may be possible to increase the amount of the produced fine chemical which is present in the extracellular medium, thus permitting greater ease of harvesting and purification. Conversely, in order to efficiently overproduce one or more fine chemicals, increased amounts of the cofactors, precursor molecules, and intermediate compounds for the appropriate biosynthetic pathways are required. Therefore, by increasing the number and/or activity of transporter proteins involved in the import of nutrients, such as carbon sources (i.e., sugars), nitrogen sources (i.e., amino acids, ammonium salts), phosphate, and sulfur, it may be possible to improve the production of a fine chemical, due to the removal of any nutrient supply limitations on the biosynthetic process. Further, fatty acids and lipids are themselves desirable fine chemicals, so by optimizing the activity or increasing the number of one or more

MCT proteins of the invention which participate in the biosynthesis of these compounds, or by impairing the activity of one or more MCT proteins which are involved in the degradation of these compounds, it may be possible to increase the yield, production, and/or efficiency of production of fatty acid and lipid molecules from *C. glutamicum*.

[0122] The engineering of one or more MCT genes of the invention may also result in MCT proteins having altered activities which indirectly impact the production of one or more desired fine chemicals from *C. glutamicum*. For example, the normal biochemical processes of metabolism result in the production of a variety of waste products (e.g., hydrogen peroxide and other reactive oxygen species) which may actively interfere with these same metabolic processes (for example, peroxynitrite is known to nitrate tyrosine side chains, thereby inactivating some enzymes having tyrosine in the active site (Groves, J. T. (1999) *Curr. Opin. Chem. Biol.* 3(2): 226-235). While these waste products are typically excreted, the *C. glutamicum* strains utilized for large-scale fermentative production are optimized for the overproduction of one or more fine chemicals, and thus may produce more waste products than is typical for a wild-type *C. glutamicum*. By optimizing the activity of one or more MCT proteins of the invention which are involved in the export of waste molecules, it may be possible to improve the viability of the cell and to maintain efficient metabolic activity. Also, the presence of high intracellular levels of the desired fine chemical may actually be toxic to the cell, so by increasing the ability of the cell to secrete these compounds, one may improve the viability of the cell.

[0123] Further, the MCT proteins of the invention may be manipulated such that the relative amounts of various lipid and fatty acid molecules produced are altered. This may have a profound effect on the lipid composition of the membrane of the cell. Since each type of lipid has different physical properties, an alteration in the lipid composition of a membrane may significantly alter membrane fluidity. Changes in membrane fluidity can impact the transport of molecules across the membrane, which, as previously explicated, may modify the export of waste products or the produced fine chemical or the import of necessary nutrients. Such membrane fluidity changes may also profoundly affect the integrity of the cell; cells with relatively weaker membranes are more vulnerable in the large-scale fermentor environment to mechanical stresses which may damage or kill the cell. By manipulating MCT proteins involved in the production of fatty acids and lipids for membrane construction such that the resulting membrane has a membrane composition more amenable to the environmental conditions extant in the cultures utilized to produce fine chemicals, a greater proportion of the *C. glutamicum* cells should survive and multiply. Greater numbers of *C. glutamicum* cells in a culture should translate into greater yields, production, or efficiency of production of the fine chemical from the culture.

[0124] The aforementioned mutagenesis strategies for MCT proteins to result in increased yields of a fine chemical from *C. glutamicum* are not meant to be limiting; variations on these strategies will be readily apparent to one of ordinary skill in the art. Using such strategies, and incorporating the mechanisms disclosed herein, the nucleic acid and protein molecules of the invention may be utilized to generate *C. glutamicum* or related strains of bacteria expressing mutated

MCT nucleic acid and protein molecules such that the yield, production, and/or efficiency of production of a desired compound is improved. This desired compound may be any natural product of *C. glutamicum*, which includes the final products of biosynthesis pathways and intermediates of naturally-occurring metabolic pathways, as well as molecules which do not naturally occur in the metabolism of *C. glutamicum*, but which are produced by a *C. glutamicum* strain of the invention.

[0125] This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patent applications, patents, published patent applications, Tables, Appendices, and the sequence listing cited throughout this application are hereby incorporated by reference.

Exemplification

EXAMPLE 1

Preparation of Total Genomic DNA of *Corynebacterium glutamicum* ATCC 13032

[0126] A culture of *Corynebacterium glutamicum* (ATCC 13032) was grown overnight at 30° C. with vigorous shaking in BHI medium (Difco). The cells were harvested by centrifugation, the supernatant was discarded and the cells were resuspended in 5 ml buffer-I (5% of the original volume of the culture—all indicated volumes have been calculated for 100 ml of culture volume). Composition of buffer-I: 140.34 g/l sucrose, 2.46 g/l $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 10 ml/l KH_2PO_4 solution (100 g/l, adjusted to pH 6.7 with KOH), 50 ml/l M12 concentrate (10 g/l $(\text{NH}_4)_2\text{SO}_4$, 1 g/l NaCl, 2 g/l $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 0.2 g/l CaCl_2 , 0.5 g/l yeast extract (Difco), 10 ml/l trace-elements-mix (200 mg/l $\text{FeSO}_4 \times \text{H}_2\text{O}$, 10 mg/l $\text{ZnSO}_4 \times 7\text{H}_2\text{O}$, 3 mg/l $\text{MnCl}_2 \times 4\text{H}_2\text{O}$, 30 mg/l H_3BO_3 , 20 mg/l $\text{CoCl}_2 \times 6\text{H}_2\text{O}$, 1 mg/l $\text{NiCl}_2 \times 6\text{H}_2\text{O}$, 3 mg/l $\text{Na}_2\text{MoO}_4 \times 2\text{H}_2\text{O}$, 500 mg/l complexing agent (EDTA or citric acid), 100 ml/l vitamins-mix (0.2 mg/l biotin, 0.2 mg/l folic acid, 20 mg/l p-amino benzoic acid, 20 mg/l riboflavin, 40 mg/l ca-pantothenate, 140 mg/l nicotinic acid, 40 mg/l pyridoxole hydrochloride, 200 mg/l myo-inositol). Lysozyme was added to the suspension to a final concentration of 2.5 mg/ml. After an approximately 4 h incubation at 37° C., the cell wall was degraded and the resulting protoplasts are harvested by centrifugation. The pellet was washed once with 5 ml buffer-I and once with 5 ml TE-buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8). The pellet was resuspended in 4 ml TE-buffer and 0.5 ml SDS solution (10%) and 0.5 ml NaCl solution (5 M) are added. After adding of proteinase K to a final concentration of 200 µg/ml, the suspension is incubated for ca. 18 h at 37° C. The DNA was purified by extraction with phenol, phenol-chloroform-isoamylalcohol and chloroform-isoamylalcohol using standard procedures. Then, the DNA was precipitated by adding 1/50 volume of 3 M sodium acetate and 2 volumes of ethanol, followed by a 30 min incubation at -20° C. and a 30 min centrifugation at 12,000 rpm in a high speed centrifuge using a SS34 rotor (Sorvall). The DNA was dissolved in 1 ml TE-buffer containing 20 µg/ml RNaseA and dialysed at 4° C. against 1000 ml TE-buffer for at least 3 hours. During this time, the buffer was exchanged 3 times. To aliquots of 0.4 ml of the dialysed DNA solution, 0.4 ml of 2 M LiCl and 0.8 ml of ethanol are added. After a 30 min incubation at -20° C., the DNA was collected by centrifugation (13,000 rpm,

Biofuge Fresco, Heraeus, Hanau, Germany). The DNA pellet was dissolved in TE-buffer. DNA prepared by this procedure could be used for all purposes, including southern blotting or construction of genomic libraries.

EXAMPLE 2

Construction of Genomic Libraries in *Escherichia coli* of *Corynebacterium glutamicum* ATCC13032.

[0127] Using DNA prepared as described in Example 1, cosmid and plasmid libraries were constructed according to known and well established methods (see e.g., Sambrook, J. et al. (1989) "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, or Ausubel, F. M. et al. (1994) "Current Protocols in Molecular Biology", John Wiley & Sons.)

[0128] Any plasmid or cosmid could be used. Of particular use were the plasmids pBR322 (Sutcliffe, J. G. (1979) *Proc. Natl. Acad. Sci. USA*, 75:3737-3741); pACYC177 (Change & Cohen (1978) *J. Bacteriol* 134:1141-1156), plasmids of the pBS series (pBSSK+, pBSSK- and others; Stratagene, LaJolla, USA), or cosmids as SuperCos1 (Stratagene, LaJolla, USA) or Lorist6 (Gibson, T. J., Rosenthal A. and Waterson, R. H. (1987) *Gene* 53:283-286. Gene libraries specifically for use in *C. glutamicum* may be constructed using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

EXAMPLE 3

DNA Sequencing and Computational Functional Analysis

[0129] Genomic libraries as described in Example 2 were used for DNA sequencing according to standard methods, in particular by the chain termination method using ABI377 sequencing machines (see e.g., Fleischman, R. D. et al. (1995) "Whole-genome Random Sequencing and Assembly of Haemophilus Influenzae Rd.", *Science*, 269:496-512). Sequencing primers with the following nucleotide sequences were used: 5'-GGAAACAGTATGACCATG-3' or 5'-GTAAAACGACGGCCAGT-3'.

EXAMPLE 4

In vivo Mutagenesis

[0130] In vivo mutagenesis of *Corynebacterium glutamicum* can be performed by passage of plasmid (or other vector) DNA through *E. coli* or other microorganisms (e.g. *Bacillus* spp. or yeasts such as *Saccharomyces cerevisiae*) which are impaired in their capabilities to maintain the integrity of their genetic information. Typical mutator strains have mutations in the genes for the DNA repair system (e.g., mutHLS, mutD, mutT, etc.; for reference, see Rupp, W. D. (1996) DNA repair mechanisms, in: *Escherichia coli* and *Salmonella*, p. 2277-2294, ASM: Washington.) Such strains are well known to those of ordinary skill in the art. The use of such strains is illustrated, for example, in Greener, A. and Callahan, M. (1994) *Strategies* 7: 32-34.

EXAMPLE 5

DNA Transfer Between *Escherichia coli* and *Corynebacterium glutamicum*

[0131] Several *Corynebacterium* and *Brevibacterium* species contain endogenous plasmids (as e.g., pHM1519 or

pBL1) which replicate autonomously (for review see, e.g., Martin, J. F. et al. (1987) *Biotechnology*, 5:137-146). Shuttle vectors for *Escherichia coli* and *Corynebacterium glutamicum* can be readily constructed by using standard vectors for *E. coli* (Sambrook, J. et al. (1989), "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press or Ausubel, F. M. et al. (1994) "Current Protocols in Molecular Biology", John Wiley & Sons) to which a origin or replication for and a suitable marker from *Corynebacterium glutamicum* is added. Such origins of replication are preferably taken from endogenous plasmids isolated from *Corynebacterium* and *Brevibacterium* species. Of particular use as transformation markers for these species are genes for kanamycin resistance (such as those derived from the Tn5 or Tn903 transposons) or chloramphenicol (Winnacker, E. L. (1987) "From Genes to Clones—Introduction to Gene Technology, VCH, Weinheim). There are numerous examples in the literature of the construction of a wide variety of shuttle vectors which replicate in both *E. coli* and *C. glutamicum*, and which can be used for several purposes, including gene over-expression (for reference, see e.g., Yoshihama, M. et al. (1985) *J. Bacteriol.* 162:591-597, Martin J. F. et al. (1987) *Biotechnology*, 5:137-146 and Eikmanns, B. J. et al. (1991) *Gene*, 102:93-98).

[0132] Using standard methods, it is possible to clone a gene of interest into one of the shuttle vectors described above and to introduce such a hybrid vectors into strains of *Corynebacterium glutamicum*. Transformation of *C. glutamicum* can be achieved by protoplast transformation (Kastsumata, R. et al. (1984) *J. Bacteriol.* 159:306-311), electroporation (Liebl, E. et al. (1989) *FEMS Microbiol. Letters*, 53:399-303) and in cases where special vectors are used, also by conjugation (as described e.g. in Schäfer, A et al. (1990) *J. Bacteriol.* 172:1663-1666). It is also possible to transfer the shuttle vectors for *C. glutamicum* to *E. coli* by preparing plasmid DNA from *C. glutamicum* (using standard methods well-known in the art) and transforming it into *E. coli*. This transformation step can be performed using standard methods, but it is advantageous to use an Mcr-deficient *E. coli* strain, such as NM522 (Gough & Murray (1983) *J. Mol. Biol.* 166:1-19).

[0133] Genes may be overexpressed in *C. glutamicum* strains using plasmids which comprise pCG1 (U.S. Pat. No. 4,617,267) or fragments thereof, and optionally the gene for kanamycin resistance from TN903 (Grindley, N. D. and Joyce, C. M. (1980) *Proc. Natl. Acad. Sci. USA* 77(12): 7176-7180). In addition, genes may be overexpressed in *C. glutamicum* strains using plasmid pSL109 (Lee, H.-S. and A. J. Sinskey (1994) *J. Microbiol. Biotechnol.* 4: 256-263).

[0134] Aside from the use of replicative plasmids, gene overexpression can also be achieved by integration into the genome. Genomic integration in *C. glutamicum* or other *Corynebacterium* or *Brevibacterium* species may be accomplished by well-known methods, such as homologous recombination with genomic region(s), restriction endonuclease mediated integration (REMI) (see, e.g., DE Patent 19823834), or through the use of transposons. It is also possible to modulate the activity of a gene of interest by modifying the regulatory regions (e.g., a promoter, a repressor, and/or an enhancer) by sequence modification, insertion, or deletion using site-directed methods (such as homologous recombination) or methods based on random events (such as transposon mutagenesis or REMI). Nucleic acid sequences

which function as transcriptional terminators may also be inserted 3' to the coding region of one or more genes of the invention; such terminators are well-known in the art and are described, for example, in Winnacker, E. L. (1987) *From Genes to Clones—Introduction to Gene Technology*. VCH: Weinheim.

EXAMPLE 6

Assessment of the Expression of the Mutant Protein

[0135] Observations of the activity of a mutated protein in a transformed host cell rely on the fact that the mutant protein is expressed in a similar fashion and in a similar quantity to that of the wild-type protein. A useful method to ascertain the level of transcription of the mutant gene (an indicator of the amount of mRNA available for translation to the gene product) is to perform a Northern blot (for reference see, for example, Ausubel et al. (1988) *Current Protocols in Molecular Biology*, Wiley: N.Y.), in which a primer designed to bind to the gene of interest is labeled with a detectable tag (usually radioactive or chemiluminescent), such that when the total RNA of a culture of the organism is extracted, run on gel, transferred to a stable matrix and incubated with this probe, the binding and quantity of binding of the probe indicates the presence and also the quantity of mRNA for this gene. This information is evidence of the degree of transcription of the mutant gene. Total cellular RNA can be prepared from *Corynebacterium glutamicum* by several methods, all well-known in the art, such as that described in Bormann, E. R. et al. (1992) *Mol. Microbiol.* 6: 317-326.

[0136] To assess the presence or relative quantity of protein translated from this mRNA, standard techniques, such as a Western blot, may be employed (see, for example, Ausubel et al. (1988) *Current Protocols in Molecular Biology*, Wiley: N.Y.). In this process, total cellular proteins are extracted, separated by gel electrophoresis, transferred to a matrix such as nitrocellulose, and incubated with a probe, such as an antibody, which specifically binds to the desired protein. This probe is generally tagged with a chemiluminescent or colorimetric label which may be readily detected. The presence and quantity of label observed indicates the presence and quantity of the desired mutant protein present in the cell.

EXAMPLE 7

Growth of Genetically Modified *Corynebacterium glutamicum*—Media and Culture Conditions

[0137] Genetically modified *Corynebacteria* are cultured in synthetic or natural growth media. A number of different growth media for *Corynebacteria* are both well-known and readily available (Lieb et al. (1989) *Appl. Microbiol. Biotechnol.*, 32:205-210; von der Osten et al. (1998) *Biotechnology Letters*, 11:11-16; Patent DE 4,120,867; Liebl (1992) "The Genus *Corynebacterium*, in: *The Prokaryotes, Volume II*, Balows, A. et al., eds. Springer-Verlag). These media consist of one or more carbon sources, nitrogen sources, inorganic salts, vitamins and trace elements. Preferred carbon sources are sugars, such as mono-, di-, or polysaccharides. For example, glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose,

raffinose, starch or cellulose serve as very good carbon sources. It is also possible to supply sugar to the media via complex compounds such as molasses or other by-products from sugar refinement. It can also be advantageous to supply mixtures of different carbon sources. Other possible carbon sources are alcohols and organic acids, such as methanol, ethanol, acetic acid or lactic acid. Nitrogen sources are usually organic or inorganic nitrogen compounds, or materials which contain these compounds. Exemplary nitrogen sources include ammonia gas or ammonia salts, such as NH_4Cl or $(\text{NH}_4)_2\text{SO}_4$, NH_4OH , nitrates, urea, amino acids or complex nitrogen sources like corn steep liquor, soy bean flour, soy bean protein, yeast extract, meat extract and others.

[0138] Inorganic salt compounds which may be included in the media include the chloride-, phosphorous- or sulfate-salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper and iron. Chelating compounds can be added to the medium to keep the metal ions in solution. Particularly useful chelating compounds include dihydroxyphenols, like catechol or protocatechuate, or organic acids, such as citric acid. It is typical for the media to also contain other growth factors, such as vitamins or growth promoters, examples of which include biotin, riboflavin, thiamin, folic acid, nicotinic acid, pantothenate and pyridoxin. Growth factors and salts frequently originate from complex media components such as yeast extract, molasses, corn steep liquor and others. The exact composition of the media compounds depends strongly on the immediate experiment and is individually decided for each specific case. Information about media optimization is available in the textbook "Applied Microbiol. Physiology, A Practical Approach (eds. P. M. Rhodes, P. F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). It is also possible to select growth media from commercial suppliers, like standard 1 (Merck) or BHI (grain heart infusion, DIFCO) or others.

[0139] All medium components are sterilized, either by heat (20 minutes at 1.5 bar and 121° C.) or by sterile filtration. The components can either be sterilized together or, if necessary, separately. All media components can be present at the beginning of growth, or they can optionally be added continuously or batchwise.

[0140] Culture conditions are defined separately for each experiment. The temperature should be in a range between 15° C. and 45° C. The temperature can be kept constant or can be altered during the experiment. The pH of the medium should be in the range of 5 to 8.5, preferably around 7.0, and can be maintained by the addition of buffers to the media. An exemplary buffer for this purpose is a potassium phosphate buffer. Synthetic buffers such as MOPS, HEPES, ACES and others can alternatively or simultaneously be used. It is also possible to maintain a constant culture pH through the addition of NaOH or NH_4OH during growth. If complex medium components such as yeast extract are utilized, the necessity for additional buffers may be reduced, due to the fact that many complex compounds have high buffer capacities. If a fermentor is utilized for culturing the microorganisms, the pH can also be controlled using gaseous ammonia.

[0141] The incubation time is usually in a range from several hours to several days. This time is selected in order

to permit the maximal amount of product to accumulate in the broth. The disclosed growth experiments can be carried out in a variety of vessels, such as microtiter plates, glass tubes, glass flasks or glass or metal fermentors of different sizes. For screening a large number of clones, the microorganisms should be cultured in microtiter plates, glass tubes or shake flasks, either with or without baffles. Preferably 100 ml shake flasks are used, filled with 10% (by volume) of the required growth medium. The flasks should be shaken on a rotary shaker (amplitude 25 mm) using a speed-range of 100-300 rpm. Evaporation losses can be diminished by the maintenance of a humid atmosphere; alternatively, a mathematical correction for evaporation losses should be performed.

[0142] If genetically modified clones are tested, an unmodified control clone or a control clone containing the basic plasmid without any insert should also be tested. The medium is inoculated to an OD_{600} of 0.5-1.5 using cells grown on agar plates, such as CM plates (10 g/l glucose, 2.5 g/l NaCl , 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l NaCl , 2 g/l urea, 10 g/l polypeptone, 5 g/l yeast extract, 5 g/l meat extract, 22 g/l agar, pH 6.8 with 2M NaOH) that had been incubated at 30° C. Inoculation of the media is accomplished by either introduction of a saline suspension of *C. glutamicum* cells from CM plates or addition of a liquid preculture of this bacterium.

EXAMPLE 8

In vitro Analysis of the Function of Mutant Proteins

[0143] The determination of activities and kinetic parameters of enzymes is well established in the art. Experiments to determine the activity of any given altered enzyme must be tailored to the specific activity of the wild-type enzyme, which is well within the ability of one of ordinary skill in the art. Overviews about enzymes in general, as well as specific details concerning structure, kinetics, principles, methods, applications and examples for the determination of many enzyme activities may be found, for example, in the following references: Dixon, M., and Webb, E. C., (1979) *Enzymes*. Longmans: London; Fersht, (1985) *Enzyme Structure and Mechanism*. Freeman: New York; Walsh, (1979) *Enzymatic Reaction Mechanisms*. Freeman: San Francisco; Price, N. C., Stevens, L. (1982) *Fundamentals of Enzymology*. Oxford Univ. Press: Oxford; Boyer, P. D., ed. (1983) *The Enzymes*, 3rd ed. Academic Press: New York; Bisswanger, H., (1994) *Enzymkinetik*, 2nd ed. VCH: Weinheim (ISBN 3527300325); Bergmeyer, H. U., Bergmeyer, J., Gräßl, M., eds. (1983-1986) *Methods of Enzymatic Analysis*, 3rd ed., vol. I-XII, Verlag Chemie: Weinheim; and Ullmann's Encyclopedia of Industrial Chemistry (1987) vol. A9, "Enzymes". VCH: Weinheim, p. 352-363.

[0144] The activity of proteins which bind to DNA can be measured by several well-established methods, such as DNA band-shift assays (also called gel retardation assays). The effect of such proteins on the expression of other molecules can be measured using reporter gene assays (such as that described in Kolmar, H. et al. (1995) *EMBO J.* 14: 3895-3904 and references cited therein). Reporter gene test systems are well known and established for applications in both pro- and eukaryotic cells, using enzymes such as beta-galactosidase, green fluorescent protein, and several others.

[0145] The determination of activity of membrane-transport proteins can be performed according to techniques such as those described in Gennis, R. B. (1989) "Pores, Channels and Transporters", in *Biomembranes, Molecular Structure and Function*, Springer: Heidelberg, p. 85-137; 199-234; and 270-322.

EXAMPLE 9

Analysis of Impact of Mutant Protein on the Production of the Desired Product

[0146] The effect of the genetic modification in *C. glutamicum* on production of a desired compound (such as an amino acid) can be assessed by growing the modified microorganism under suitable conditions (such as those described above) and analyzing the medium and/or the cellular component for increased production of the desired product (i.e., an amino acid). Such analysis techniques are well known to one of ordinary skill in the art, and include spectroscopy, thin layer chromatography, staining methods of various kinds, enzymatic and microbiological methods, and analytical chromatography such as high performance liquid chromatography (see, for example, Ullman, *Encyclopedia of Industrial Chemistry*, vol. A2, p. 89-90 and p. 443-613, VCH: Weinheim (1985); Fallon, A. et al., (1987) "Applications of HPLC in Biochemistry" in: *Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 17; Rehm et al. (1993) *Biotechnology*, vol. 3, Chapter III: "Product recovery and purification", page 469-714, VCH: Weinheim; Belter, P. A. et al. (1988) *Bioseparations: downstream processing for biotechnology*, John Wiley and Sons; Kennedy, J. F. and Cabral, J. M. S. (1992) *Recovery processes for biological materials*, John Wiley and Sons; Shaeiwitz, J. A. and Henry, J. D. (1988) *Biochemical separations*, in: *Ullmann's Encyclopedia of Industrial Chemistry*, vol. B3, Chapter 11, page 1-27, VCH: Weinheim; and Dechow, F. J. (1989) *Separation and purification techniques in biotechnology*, Noyes Publications.)

[0147] In addition to the measurement of the final product of fermentation, it is also possible to analyze other components of the metabolic pathways utilized for the production of the desired compound, such as intermediates and side-products, to determine the overall efficiency of production of the compound. Analysis methods include measurements of nutrient levels in the medium (e.g., sugars, hydrocarbons, nitrogen sources, phosphate, and other ions), measurements of biomass composition and growth, analysis of the production of common metabolites of biosynthetic pathways, and measurement of gasses produced during fermentation. Standard methods for these measurements are outlined in *Applied Microbial Physiology, A Practical Approach*, P. M. Rhodes and P. F. Stanbury, eds., IRL Press, p. 103-129; 131-163; and 165-192 (ISBN: 0199635773) and references cited therein.

EXAMPLE 10

Purification of the Desired Product from *C. glutamicum* Culture

[0148] Recovery of the desired product from the *C. glutamicum* cells or supernatant of the above-described culture can be performed by various methods well known in the art. If the desired product is not secreted from the cells,

the cells can be harvested from the culture by low-speed centrifugation, the cells can be lysed by standard techniques, such as mechanical force or sonication. The cellular debris is removed by centrifugation, and the supernatant fraction containing the soluble proteins is retained for further purification of the desired compound. If the product is secreted from the *C. glutamicum* cells, then the cells are removed from the culture by low-speed centrifugation, and the supernate fraction is retained for further purification.

[0149] The supernatant fraction from either purification method is subjected to chromatography with a suitable resin, in which the desired molecule is either retained on a chromatography resin while many of the impurities in the sample are not, or where the impurities are retained by the resin while the sample is not. Such chromatography steps may be repeated as necessary, using the same or different chromatography resins. One of ordinary skill in the art would be well-versed in the selection of appropriate chromatography resins and in their most efficacious application for a particular molecule to be purified. The purified product may be concentrated by filtration or ultrafiltration, and stored at a temperature at which the stability of the product is maximized.

[0150] There are a wide array of purification methods known to the art and the preceding method of purification is not meant to be limiting. Such purification techniques are described, for example, in Bailey, J. E. & Ollis, D. F. *Biochemical Engineering Fundamentals*, McGraw-Hill: N.Y. (1986).

[0151] The identity and purity of the isolated compounds may be assessed by techniques standard in the art. These include high-performance liquid chromatography (HPLC), spectroscopic methods, staining methods, thin layer chromatography, NIRS, enzymatic assay, or microbiologically. Such analysis methods are reviewed in: Patek et al. (1994) *Appl. Environ. Microbiol.* 60: 133-140; Malakhova et al. (1996) *Biotehnologiya* 11: 27-32; and Schmidt et al. (1998) *Bioprocess Engineer.* 19: 67-70. *Ullmann's Encyclopedia of Industrial Chemistry*, (1996) vol. A27, VCH: Weinheim, p. 89-90, p. 521-540, p. 540-547, p. 559-566, 575-581 and p. 581-587; Michal, G. (1999) *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, John Wiley and Sons; Fallon, A. et al. (1987) *Applications of HPLC in Biochemistry in: Laboratory Techniques in Biochemistry and Molecular Biology*, vol. 17.

EXAMPLE 11

Analysis of the Gene Sequences of the Invention

[0152] The comparison of sequences and determination of percent homology between two sequences are art-known techniques, and can be accomplished using a mathematical algorithm, such as the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-68, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-77. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to MCT nucleic acid molecules of the invention. BLAST protein searches can be performed with the

XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to MCT protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, one of ordinary skill in the art will know how to optimize the parameters of the program (e.g., XBLAST and NBLAST) for the specific sequence being analyzed.

[0153] Another example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Meyers and Miller ((1988) *Comput. Appl. Biosci.* 4: 11-17). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM 120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art, and include ADVANCE and ADAM, described in Torelli and Robotti (1994) *Comput. Appl. Biosci.* 10:3-5; and FASTA, described in Pearson and Lipman (1988) *P.N.A.S.* 85:2444-8.

[0154] The percent homology between two amino acid sequences can also be accomplished using the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. The percent homology between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using standard parameters, such as a gap weight of 50 and a length weight of 3.

[0155] A comparative analysis of the gene sequences of the invention with those present in Genbank has been performed using techniques known in the art (see, e.g., Bexevanis and Ouellette, eds. (1998) *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. John Wiley and Sons: New York). The gene sequences of the invention were compared to genes present in Genbank in a three-step process. In a first step, a BLASTN analysis (e.g., a local alignment analysis) was performed for each of the sequences of the invention against the nucleotide sequences present in Genbank, and the top 500 hits were retained for further analysis. A subsequent FASTA search (e.g., a combined local and global alignment analysis, in which limited regions of the sequences are aligned) was performed on these 500 hits. Each gene sequence of the invention was subsequently globally aligned to each of the top three FASTA hits, using the GAP program in the GCG software package (using standard parameters). In order to obtain correct results, the length of the sequences extracted from Genbank were adjusted to the length of the query sequences by methods well-known in the art. The results of this analysis are set forth in Table 4. The resulting data is identical to that which would have been obtained had a GAP (global) analysis alone been performed on each of the genes of the invention in comparison with each of the references in Genbank, but required significantly reduced computational time as compared to such a database-wide GAP

(global) analysis. Sequences of the invention for which no alignments above the cutoff values were obtained are indicated on Table 4 by the absence of alignment information. It will further be understood by one of ordinary skill in the art that the GAP alignment homology percentages set forth in Table 4 under the heading “% homology (GAP)” are listed in the European numerical format, wherein a ‘,’ represents a decimal point. For example, a value of “40,345” in this column represents “40.345%”.

EXAMPLE 12

Construction and Operation of DNA Microarrays

[0156] The sequences of the invention may additionally be used in the construction and application of DNA microarrays (the design, methodology, and uses of DNA arrays are well known in the art, and are described, for example, in Schena, M. et al. (1995) *Science* 270: 467-470; Wodicka, L. et al. (1997) *Nature Biotechnology* 15: 1359-1367; DeSaizieu, A. et al. (1998) *Nature Biotechnology* 16: 45-48; and DeRisi, J. L. et al. (1997) *Science* 278: 680-686).

[0157] DNA microarrays are solid or flexible supports consisting of nitrocellulose, nylon, glass, silicone, or other materials. Nucleic acid molecules may be attached to the surface in an ordered manner. After appropriate labeling, other nucleic acids or nucleic acid mixtures can be hybridized to the immobilized nucleic acid molecules, and the label may be used to monitor and measure the individual signal intensities of the hybridized molecules at defined regions. This methodology allows the simultaneous quantification of the relative or absolute amount of all or selected nucleic acids in the applied nucleic acid sample or mixture. DNA microarrays, therefore, permit an analysis of the expression of multiple (as many as 6800 or more) nucleic acids in parallel (see, e.g., Schena, M. (1996) *BioEssays* 18(5): 427-431).

[0158] The sequences of the invention may be used to design oligonucleotide primers which are able to amplify defined regions of one or more *C. glutamicum* genes by a nucleic acid amplification reaction such as the polymerase chain reaction. The choice and design of the 5' or 3' oligonucleotide primers or of appropriate linkers allows the covalent attachment of the resulting PCR products to the surface of a support medium described above (and also described, for example, Schena, M. et al. (1995) *Science* 270: 467-470).

[0159] Nucleic acid microarrays may also be constructed by in situ oligonucleotide synthesis as described by Wodicka, L. et al. (1997) *Nature Biotechnology* 15: 1359-1367. By photolithographic methods, precisely defined regions of the matrix are exposed to light. Protective groups which are photolabile are thereby activated and undergo nucleotide addition, whereas regions that are masked from light do not undergo any modification. Subsequent cycles of protection and light activation permit the synthesis of different oligonucleotides at defined positions. Small, defined regions of the genes of the invention may be synthesized on microarrays by solid phase oligonucleotide synthesis.

[0160] The nucleic acid molecules of the invention present in a sample or mixture of nucleotides may be hybridized to the microarrays. These nucleic acid molecules can be labeled according to standard methods. In brief, nucleic acid molecules (e.g., mRNA molecules or DNA molecules) are labeled by the incorporation of isotopically or fluorescently labeled nucleotides, e.g., during reverse transcription or DNA synthesis. Hybridization of labeled nucleic acids to microarrays is described (e.g., in Schena, M. et al. (1995) *supra*; Wodicka, L. et al. (1997), *supra*; and DeSaizieu A. et al. (1998), *supra*). The detection and quantification of the hybridized molecule are tailored to the specific incorporated label. Radioactive labels can be detected, for example, as described in Schena, M. et al. (1995) *supra* and fluorescent labels may be detected, for example, by the method of Shalon et al. (1996) *Genome Research* 6: 639-645).

[0161] The application of the sequences of the invention to DNA microarray technology, as described above, permits comparative analyses of different strains of *C. glutamicum* or other *Corynebacteria*. For example, studies of inter-strain variations based on individual transcript profiles and the identification of genes that are important for specific and/or desired strain properties such as pathogenicity, productivity and stress tolerance are facilitated by nucleic acid array methodologies. Also, comparisons of the profile of expression of genes of the invention during the course of a fermentation reaction are possible using nucleic acid array technology.

EXAMPLE 13

Analysis of the Dynamics of Cellular Protein Populations (Proteomics)

[0162] The genes, compositions, and methods of the invention may be applied to study the interactions and dynamics of populations of proteins, termed 'proteomics'. Protein populations of interest include, but are not limited to, the total protein population of *C. glutamicum* (e.g., in comparison with the protein populations of other organisms), those proteins which are active under specific environmental or metabolic conditions (e.g., during fermentation, at high or low temperature, or at high or low pH), or those proteins which are active during specific phases of growth and development.

[0163] Protein populations can be analyzed by various well-known techniques, such as gel electrophoresis. Cellular proteins may be obtained, for example, by lysis or extraction, and may be separated from one another using a variety of electrophoretic techniques. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) separates proteins largely on the basis of their molecular weight. Isoelectric focusing polyacrylamide gel electrophoresis (IEF-PAGE) separates proteins by their isoelectric point (which reflects not only the amino acid sequence but also posttranslational modifications of the protein). Another, more preferred method of protein analysis is the consecutive combination of both IEF-PAGE and SDS-PAGE, known as 2-D-gel electrophoresis (described, for example, in Hermann et

al. (1998) *Electrophoresis* 19: 3217-3221; Fountoulakis et al. (1998) *Electrophoresis* 19: 1193-1202; Langen et al. (1997) *Electrophoresis* 18: 1184-1192; Antelmann et al. (1997) *Electrophoresis* 18: 1451-1463). Other separation techniques may also be utilized for protein separation, such as capillary gel electrophoresis; such techniques are well known in the art.

[0164] Proteins separated by these methodologies can be visualized by standard techniques, such as by staining or labeling. Suitable stains are known in the art, and include Coomassie Brilliant Blue, silver stain, or fluorescent dyes such as Sypro Ruby (Molecular Probes). The inclusion of radioactively labeled amino acids or other protein precursors (e.g., ^{35}S -methionine, ^{35}S -cysteine, ^{14}C -labelled amino acids, ^{15}N -amino acids, $^{15}\text{NO}_3$ or $^{15}\text{NH}_4^+$ or ^{13}C -labelled amino acids) in the medium of *C. glutamicum* permits the labeling of proteins from these cells prior to their separation. Similarly, fluorescent labels may be employed. These labeled proteins can be extracted, isolated and separated according to the previously described techniques.

[0165] Proteins visualized by these techniques can be further analyzed by measuring the amount of dye or label used. The amount of a given protein can be determined quantitatively using, for example, optical methods and can be compared to the amount of other proteins in the same gel or in other gels. Comparisons of proteins on gels can be made, for example, by optical comparison, by spectroscopy, by image scanning and analysis of gels, or through the use of photographic films and screens. Such techniques are well-known in the art.

[0166] To determine the identity of any given protein, direct sequencing or other standard techniques may be employed. For example, N- and/or C-terminal amino acid sequencing (such as Edman degradation) may be used, as may mass spectrometry (in particular MALDI or ESI techniques (see, e.g., Langen et al. (1997) *Electrophoresis* 18: 1184-1192)). The protein sequences provided herein can be used for the identification of *C. glutamicum* proteins by these techniques.

[0167] The information obtained by these methods can be used to compare patterns of protein presence, activity, or modification between different samples from various biological conditions (e.g., different organisms, time points of fermentation, media conditions, or different biotopes, among others). Data obtained from such experiments alone, or in combination with other techniques, can be used for various applications, such as to compare the behavior of various organisms in a given (e.g., metabolic) situation, to increase the productivity of strains which produce fine chemicals or to increase the efficiency of the production of fine chemicals.

Equivalents

[0168] Those of ordinary skill in the art will recognize, or will be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

TABLE 1

GENES IN THE APPLICATION						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
1	2	RXN03097	VV0062	3	557	AMMONIUM TRANSPORT SYSTEM
3	4	RXA02099	GR00630	6198	6470	AMMONIUM TRANSPORT SYSTEM
5	6	RXA00104	GR00014	15895	16650	CYSQ PROTEIN, ammonium transport protein
Polyketide Synthesis						
7	8	RXA01420	GR00416	775	17	4"-MYCAROSYL ISOVALERYL-COA TRANSFERASE (EC 2.—.—.—)
9	10	RXN02581	VV0098	30482	28623	POLYKETIDE SYNTHASE
11	12	F RXA02581	GR00741	1	1527	POLYKETIDE SYNTHASE
13	14	RXA02582	GR00741	1890	6719	PROBABLE POLYKETIDE SYNTHASE CY338.20
15	16	RXA01138	GR00318	1656	2072	ACTINORHODIN POLYKETIDE DIMERASE (EC —.—.—.—)
17	18	RXA01980	GR00573	1470	838	POLYKETIDE CYCLASE
19	20	RXN01007	VV0021	2572	866	FRNA
21	22	RXN00784	VV0103	27531	28265	FRNE
Fatty acid and lipid synthesis						
23	24	RXA02335	GR00672	550	2322	BIOTIN CARBOXYLASE (EC 6.3.4.14)
25	26	RXA02173	GR00641	7473	8924	ACETYL-COENZYME A CARBOXYLASE CARBOXYL TRANSFERASE SUBUNIT BETA (EC 6.4.1.2)
27	28	RXA01764	GR00500	2178	3110	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
29	30	RXN02487	VV0007	6367	4664	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
31	32	F RXA02487	GR00718	4937	4650	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
33	34	F RXA02490	GR00720	817	5	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
35	36	RXA01467	GR00422	920	1210	ACYL CARRIER PROTEIN
37	38	RXA00796	GR00212	202	5	Acyl carrier protein phosphodiesterase
39	40	RXA01897	GR00544	617	1159	Acyl carrier protein phosphodiesterase
41	42	RXN02809	VV0342	380	6	Acyl carrier protein phosphodiesterase
43	44	F RXA02809	GR00790	277	5	Acyl carrier protein phosphodiesterase
45	46	RXN00113	VV0129	103	5724	FATTY ACID SYNTHASE (EC 2.3.1.85) [INCLUDES: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41;
47	48	F RXA00113	GR00017	2	3295	FATTY-ACID SYNTHASE (EC 2.3.1.85)
49	50	RXN03111	VV0084	6040	5	FATTY ACID SYNTHASE (EC 2.3.1.85) [INCLUDES: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
51	52	F RXA00158	GR00024	2088	4	FATTY ACID SYNTHASE (EC 2.3.1.85)
53	54	F RXA00572	GR00155	2	3832	FATTY ACID SYNTHASE (EC 2.3.1.85)
55	56	RXA02582	GR00741	1890	6719	PROBABLE POLYKETIDE SYNTHASE CY338.20
57	58	RXA02691	GR00754	15347	14541	FATTY ACYL RESPONSIVE REGULATOR
59	60	RXA00880	GR00242	6213	8057	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
61	62	RXA01060	GR00296	9566	10489	OMEGA-3 FATTY ACID DESATURASE (EC 1.14.99.—)
63	64	RXN01722	VV0036	2938	1214	MEDIUM-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.—)
65	66	F RXA01722	GR00488	5746	4022	MEDIUM-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.—)
67	68	RXA01644	GR00456	9854	8577	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1.79)
69	70	RXA02029	GR00618	356	1669	CYCLOPROPANE-FATTY-ACYL-PHOSPHOLIPID SYNTHASE (EC 2.1.1.79)
71	72	RXA01801	GR00509	3396	2380	ENOYL-COA HYDRATASE (EC 4.2.1.17)
73	74	RXN02512	VV0171	16147	15185	LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE (EC 2.3.1.—)
75	76	F RXA02512	GR00721	3303	4259	LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE (EC 2.3.1.—)
77	78	RXA00899	GR00245	1599	2864	CARDIOLIPIN SYNTHETASE (EC 2.7.8.—)
79	80	RXN00819	VV0054	18127	19455	ACYL-COA DEHYDROGENASE (EC 1.3.99.—)
81	82	F RXA00819	GR00221	18	1007	ACYL-COA DEHYDROGENASE (EC 1.3.99.—)
83	84	F RXA01766	GR00500	4081	4371	ACYL-COA DEHYDROGENASE (EC 1.3.99.—)
85	86	RXN01762	VV0054	15318	13783	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
87	88	F RXA01762	GR00500	1272	10	LONG-CHAIN-FATTY-ACID—COA LIGASE (EC 6.2.1.3)
89	90	RXA00681	GR00179	3405	2662	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
91	92	RXA00802	GR00214	3803	4516	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
93	94	RXA02133	GR00639	3	308	3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100)
95	96	RXN01114	VV0182	9118	10341	3-KETOACYL-COA THIOLASE (EC 2.3.1.16)
97	98	F RXA01114	GR00308	2	793	3-KETOACYL-COA THIOLASE (EC 2.3.1.16)
99	100	RXA01894	GR00542	1622	2476	PHOSPHATIDATE CYTIDYLYLTRANSFERASE (EC 2.7.7.41)
101	102	RXA02599	GR00742	3179	3655	PHOSPHATIDYLGLYCEROPHOSPHATASE B (EC 3.1.3.27)
103	104	RXN02638	VV0098	54531	53656	1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.51)
105	106	F RXA02638	GR00749	8	511	1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.51)
107	108	RXA00856	GR00232	720	1256	CDP-DIACYLGLYCEROL—GLYCEROL-3-PHOSPHATE 3- PHOSPHATIDYLTRANSFERASE (EC 2.7.8.5)
109	110	RXA02511	GR00721	2621	3277	CDP-DIACYLGLYCEROL—GLYCEROL-3-PHOSPHATE 3- PHOSPHATIDYLTRANSFERASE (EC 2.7.8.5)
111	112	RXN02836	VV0102	32818	33372	KETOACYL REDUCTASE HETN (EC 1.3.1.—)
113	114	F RXA02836	GR00827	106	411	KETOACYL REDUCTASE HETN (EC 1.3.1.—)
115	116	RXA02578	GR00740	2438	3541	PUTATIVE ACYLTRANSFERASE

TABLE 1-continued

GENES IN THE APPLICATION						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
117	118	RXA02150	GR00639	18858	19658	1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (EC 2.3.1.51)
119	120	RXA00607	GR00160	1869	2249	POLY(3-HYDROXYALKANOATE) POLYMERASE (EC 2.3.1.—)
121	122	RXA02397	GR00698	1688	2683	POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.—)
123	124	RXN03110	VV0083	16568	17929	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
125	126	F RXA00660	GR00171	1027	5	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
127	128	RXA00801	GR00214	3138	3770	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
129	130	RXA00821	GR00221	1469	2311	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
131	132	RXN02966	VV0143	12056	13462	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
133	134	F RXA01833	GR00517	1666	260	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
135	136	RXA01853	GR00525	5561	5010	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
137	138	RXN02424	VV0116	10570	11169	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
139	140	F RXA02424	GR00706	808	428	HYDROXYACYLGLUTATHIONE HYDROLASE (EC 3.1.2.6)
141	142	RXN00419	VV0112	1024	266	ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
143	144	F RXA00419	GR00095	3	464	ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
145	146	F RXA00421	GR00096	565	723	ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
147	148	RXN02923	VV0088	3301	2564	ACETOACETYL-COA REDUCTASE (EC 1.1.1.36)
149	150	RXN02922	VV0321	11407	10328	ACYL-COA DEHYDROGENASE, SHORT-CHAIN SPECIFIC (EC 1.3.99.2)
151	152	RXN03065	VV0038	6237	6629	HOLO-[ACYL-CARRIER PROTEIN] SYNTHASE (EC 2.7.8.7)
153	154	RXN03132	VV0127	39053	39472	POLY-BETA-HYDROXYBUTYRATE POLYMERASE (EC 2.3.1.—)
155	156	RXN03157	VV0188	1607	1170	LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS PROTEIN KDTB
157	158	RXN00934	VV0171	15181	14099	(AE000805) LPS biosynthesis RfbU related protein [Methanobacterium thermoautotrophicum]
159	160	RXN00792	VV0321	10328	9132	ACYL-COA DEHYDROGENASE, SHORT-CHAIN SPECIFIC (EC 1.3.99.2)
161	162	RXN00931	VV0171	13011	12166	ACYL-COA THIOESTERASE II (EC 3.1.2.—)
163	164	F RXA00931	GR00253	4959	4114	thioesterase II
165	166	RXN01421	VV0122	16024	15638	ACYLTRANSFERASE (EC 2.3.1.—)
167	168	RXN02342	VV0078	3460	4266	BIOTIN—[ACETYL-COA-CARBOXYLASE] SYNTHETASE (EC 6.3.4.15)
169	170	RXN00563	VV0038	1	2739	FATTY ACID SYNTHASE (EC 2.3.1.85) [INCLUDES: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
171	172	RXN02168	VV0100	2894	81	FATTY ACID SYNTHASE (EC 2.3.1.85) [INCLUDES: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14]
173	174	RXN01090	VV0155	6483	5686	KETOACYL REDUCTASE HETN (EC 1.3.1.—)
175	176	RXN02062	VV0222	3159	1990	Lipopolysaccharide N-acetylglucosaminyltransferase
177	178	RXN02148	VV0300	16561	17703	Lipopolysaccharide N-acetylglucosaminyltransferase
179	180	RXN02595	VV0098	11098	9935	Lipopolysaccharide N-acetylglucosaminyltransferase
181	182	RXS00148	VV0167	9849	12059	METHYLMALONYL-COA MUTASE ALPHA-SUBUNIT (EC 5.4.99.2)
183	184	RXS00149	VV0167	7995	9842	METHYLMALONYL-COA MUTASE BETA-SUBUNIT (EC 5.4.99.2)
185	186	RXS02106	VV0123	22649	21594	LIPOATE-PROTEIN LIGASE A (EC 6.—.—.—)
187	188	RXS01746	VV0185	934	1686	LIPOATE-PROTEIN LIGASE B (EC 6.—.—.—)
189	190	RXS01747	VV0185	1826	2869	LIPOIC ACID SYNTHETASE
191	192	RXC01748	VV0185	3001	3780	protein involved in lipid metabolism
193	194	RXC00354	VV0135	33604	32792	Cytosolic Protein involved in lipid metabolism
195	196	RXC01749	VV0185	3953	5569	Membrane Spanning Protein involved in lipid metabolism
Fatty acid degradation						
197	198	RXA02268	GR00655	2182	3081	LIPASE (EC 3.1.1.3)
199	200	RXA02269	GR00655	3094	4065	LIPASE (EC 3.1.1.3)
201	202	RXA01614	GR00449	8219	7197	LYSOPHOSPHOLIPASE L2 (EC 3.1.1.5)
203	204	RXA01983	GR00573	3559	3053	LIPASE (EC 3.1.1.3)
205	206	RXN02947	VV0078	1319	6	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
207	208	F RXA02320	GR00667	593	6	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
209	210	F RXA02851	GR00851	524	6	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
211	212	RXN02321	VV0078	3291	1663	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
213	214	F RXA02321	GR00667	1380	937	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
215	216	F RXA02343	GR00675	1403	1816	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
217	218	F RXA02850	GR00850	2	493	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
219	220	RXA02583	GR00741	6743	8290	PROPIONYL-COA CARBOXYLASE BETA CHAIN (EC 6.4.1.3)
221	222	RXA00870	GR00239	809	2320	METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING) (EC 1.2.1.27) 2-Methyl-3-oxopropanoate: NAD+ oxidoreductase (CoA-propanoylating)
223	224	RXA01260	GR00367	2381	1200	LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
225	226	RXA01261	GR00367	2607	2437	LIPOAMIDE DEHYDROGENASE COMPONENT (E3) OF BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COMPLEX (EC 1.8.1.4)
227	228	RXA01136	GR00318	685	1116	ISOVALERYL-COA DEHYDROGENASE (EC 1.3.99.10)
229	230	RXN00559	VV0103	7568	6552	PROTEIN VLDL
231	232	F RXA00559	GR00149	218	6	PROTEIN VLDL

TABLE 1-continued

GENES IN THE APPLICATION						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
233	234	RXA01580	GR00440	707	6	Glycerophosphoryl diester phosphodiesterase
235	236	RXA02677	GR00754	3119	3877	GLYCEROPHOSPHORYL DIESTER PHOSPHODIESTERASE (EC 3.1.4.46)
237	238	RXS01166	VV0117	18142	16838	EXTRACELLULAR LIPASE PRECURSOR (EC 3.1.1.3)
Terpenoid biosynthesis						
239	240	RXA00875	GR00241	2423	1857	ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (EC 5.3.3.2)
241	242	RXA01292	GR00373	1204	2388	PHYTOENE DEHYDROGENASE (EC 1.3.—.—)
243	244	RXA01293	GR00373	2370	2696	PHYTOENE DEHYDROGENASE (EC 1.3.—.—)
245	246	RXA02310	GR00665	1132	2394	GERANYLGERANYL HYDROGENASE
247	248	RXA02718	GR00758	18539	19585	GERANYLGERANYL PYROPHOSPHATE SYNTHASE (EC 2.5.1.1)
249	250	RXA01067	GR00298	1453	2181	undecaprenyl-diphosphate synthase (EC 2.5.1.31)
251	252	RXA01269	GR00367	20334	19894	UNDECAPRENYL-PHOSPHATE GALACTOSEPHOSPHOTRANSFERASE (EC 2.7.8.6)
253	254	RXA01205	GR00346	3	533	PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE (EC 2.4.1.—)
255	256	RXA01576	GR00438	8053	8811	DOLICHYL-PHOSPHATE BETA-GLUCOSYLTRANSFERASE (EC 2.4.1.117)
257	258	RXN02309	VV0025	28493	29542	OCTAPRENYL-DIPHOSPHATE SYNTHASE (EC 2.5.1.—)
259	260	F RXA02309	GR00665	978	4	OCTAPRENYL-DIPHOSPHATE SYNTHASE (EC 2.5.1.—)
261	262	RXN00477	VV0086	38905	37262	PHYTOENE DEHYDROGENASE (EC 1.3.—.—)
263	264	F RXA00477	GR00119	13187	11544	PHYTOENE DEHYDROGENASE (EC 1.3.—.—)
265	266	RXA00478	GR00119	14020	13190	PHYTOENE SYNTHASE (EC 2.5.1.—)
267	268	RXA01291	GR00373	345	1277	PHYTOENE SYNTHASE (EC 2.5.1.—)
269	270	RXA00480	GR00119	17444	16329	FARNESYL DIPHOSPHATE SYNTHASE (EC 2.5.1.1) (EC 2.5.1.10)
271	272	RXS01879	VV0105	1505	573	isopentenyl-phosphate kinase (EC 2.7.4.—)
273	274	RXS02023	VV0160	3234	4001	P450 cytochrome, isopentenyltransf, ferridox
275	276	RXS00948	VV0107	4266	5384	12-oxophytodienoate reductase (EC 1.3.1.42)
277	278	RXS02228	VV0068	1876	2778	TRNA DELTA(2)-ISOPENTENYLPYROPHOSPHATE TRANSFERASE (EC 2.5.1.8)
279	280	RXC01971	VV0105	4545	3715	Metal-Dependent Hydrolase involved in metabolism of terpenoids
281	282	RXC02697	VV0017	31257	32783	membrane protein involved in metabolism of terpenoids
ABC-Transporter						
283	284	RXN01946	VV0228	2	1276	Hypothetical ABC Transporter ATP-Binding Protein
285	286	F RXA01946	GR00559	1849	575	(AL021184) ABC transporter ATP binding protein [Mycobacterium tuberculosis]
287	288	RXN00164	VV0232	1782	94	Hypothetical ABC Transporter ATP-Binding Protein
289	290	F RXA00164	GR00025	1782	94	, P, G, R ATPase subunits of ABC transporters
291	292	RXN00243	VV0057	28915	27899	, P, G, R ATPase subunits of ABC transporters
293	294	F RXA00243	GR00037	930	4	, P, G, R ATPase subunits of ABC transporters
295	296	RXA00259	GR00039	8469	6268	, P, G, R ATPase subunits of ABC transporters
297	298	RXN00410	VV0086	51988	51323	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
299	300	F RXA00410	GR00092	829	164	, P, G, R ATPase subunits of ABC transporters
301	302	RXN00456	VV0076	6780	8156	, P, G, R ATPase subunits of ABC transporters
303	304	F RXA00456	GR00114	316	5	, P, G, R ATPase subunits of ABC transporters
305	306	F RXA00459	GR00115	1231	245	, P, G, R ATPase subunits of ABC transporters
307	308	RXN01604	VV0137	8117	7470	, P, G, R ATPase subunits of ABC transporters
309	310	F RXA01604	GR00448	2	607	, P, G, R ATPase subunits of ABC transporters
311	312	RXN02547	VV0057	27726	25588	, P, G, R ATPase subunits of ABC transporters
313	314	F RXA02547	GR00726	22055	19932	, P, G, R ATPase subunits of ABC transporters
315	316	RXN02571	VV0101	12331	13359	MALTOSE/MALTODEXTRIN TRANSPORT ATP-BINDING PROTEIN MALK
317	318	F RXA02571	GR00736	1469	2497	, P, G, R ATPase subunits of ABC transporters
319	320	RXN02074	VV0318	12775	11153	TRANSPORT ATP-BINDING PROTEIN CYDD
321	322	F RXA02074	GR00628	5798	4176	, P, G, R ATPase subunits of ABC transporters
323	324	RXA02095	GR00629	14071	15474	, P, G, R ATPase subunits of ABC transporters
325	326	RXA02225	GR00652	3156	2275	, P, G, R ATPase subunits of ABC transporters
327	328	RXA02253	GR00654	20480	21406	, P, G, R ATPase subunits of ABC transporters
329	330	RXN01881	VV0105	529	95	Hypothetical ABC Transporter ATP-Binding Protein
331	332	F RXA01881	GR00537	3092	3532	ATPase components of ABC transporters with duplicated ATPase domains
333	334	RXA00526	GR00136	1353	664	Hypothetical ABC Transporter ATP-Binding Protein
335	336	RXN00733	VV0132	1647	2531	Hypothetical ABC Transporter ATP-Binding Protein
337	338	F RXA00733	GR00197	411	4	Hypothetical ABC Transporter ATP-Binding Protein
339	340	RXA00735	GR00198	849	181	Hypothetical ABC Transporter ATP-Binding Protein
341	342	RXA00878	GR00242	3733	1871	Hypothetical ABC Transporter ATP-Binding Protein
343	344	RXN01191	VV0169	10478	12067	Hypothetical ABC Transporter ATP-Binding Protein
345	346	F RXA01191	GR00341	1571	165	Hypothetical ABC Transporter ATP-Binding Protein
347	348	RXN01212	VV0169	3284	4207	Hypothetical ABC Transporter ATP-Binding Protein
349	350	F RXA01212	GR00350	1	813	Hypothetical ABC Transporter ATP-Binding Protein
351	352	RXA02749	GR00764	4153	5028	Hypothetical ABC Transporter ATP-Binding Protein
353	354	RXA02224	GR00652	2271	475	Hypothetical ABC Transporter ATP-Binding Protein

TABLE 1-continued

GENES IN THE APPLICATION						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
355	356	RXN01602	VV0229	1109	2638	Hypothetical ABC Transporter ATP-Binding Protein
357	358	RXN02515	VV0087	962	1717	Hypothetical ABC Transporter ATP-Binding Protein
359	360	RXN00525	VV0079	26304	27566	Hypothetical ABC Transporter Permease Protein
361	362	RXN02096	VV0126	20444	22135	Hypothetical ABC Transporter Permease Protein
363	364	RXN00412	VV0086	53923	52844	Hypothetical Amino Acid ABC Transporter ATP-Binding Protein
365	366	RXN00411	VV0086	52844	52170	Hypothetical Amino Acid ABC Transporter Permease Protein
367	368	RXN02614	VV0313	5964	5236	TAURINE TRANSPORT ATP-BINDING PROTEIN TAUB
369	370	RXN02613	VV0313	5223	4267	TAURINE-BINDING PERIPLASMIC PROTEIN PRECURSOR
371	372	RXN00368	VV0226	2300	726	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
373	374	F RXA00368	GR00076	1	579	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
375	376	F RXA00370	GR00077	6	803	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA
377	378	RXN01285	VV0215	1780	1055	FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
379	380	RXN00523	VV0194	1363	338	FERRIC ENTEROBACTIN TRANSPORT PROTEIN FEPC
381	382	RXN01142	VV0077	5805	6302	NITRATE TRANSPORT ATP-BINDING PROTEIN NRTD
383	384	RXN01141	VV0077	4644	5468	NITRATE TRANSPORT PROTEIN NRTA
385	386	RXN01002	VV0106	8858	8055	PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNC
387	388	RXN01000	VV0106	7252	6407	PHOSPHONATES TRANSPORT SYSTEM PERMEASE PROTEIN PHNE
389	390	RXN01732	VV0106	9944	8895	PHOSPHONATES-BINDING PERIPLASMIC PROTEIN PRECURSOR
391	392	RXN03080	VV0045	1670	2449	FERRIC ENTEROBACTIN TRANSPORT ATP-BINDING PROTEIN FEPC
393	394	RXN03081	VV0045	2476	2934	FERRIENTEROBACTIN-BINDING PERIPLASMIC PROTEIN PRECURSOR
395	396	RXN03082	VV0045	3131	3451	FERRIENTEROBACTIN-BINDING PERIPLASMIC PROTEIN PRECURSOR
Other transporters						
397	398	RXA02261	GR00654	30936	32291	AMMONIUM TRANSPORT SYSTEM
399	400	RXA02020	GR00613	1015	5	AROMATIC AMINO ACID TRANSPORT PROTEIN AROP
401	402	RXA00281	GR00043	4721	5404	BACITRACIN TRANSPORT ATP-BINDING PROTEIN BCRA
403	404	RXN00570	VV0147	855	4	BENZOATE MEMBRANE TRANSPORT PROTEIN
405	406	F RXA00570	GR00153	1	498	BENZOATE MEMBRANE TRANSPORT PROTEIN
407	408	RXN00571	VV173	1298	42	BENZOATE MEMBRANE TRANSPORT PROTEIN
409	410	F RXA00571	GR00154	2	1186	BENZOATE MEMBRANE TRANSPORT PROTEIN
411	412	RXA00962	GR00268	2	667	BENZOATE MEMBRANE TRANSPORT PROTEIN
413	414	RXA02811	GR00792	177	560	BENZOATE MEMBRANE TRANSPORT PROTEIN
415	416	RXA02115	GR00635	2	1198	BENZOATE MEMBRANE TRANSPORT PROTEIN
417	418	RXN00590	VV0178	5043	6230	BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II
419	420	F RXA00590	GR00157	178	564	BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II
421	422	F RXA01538	GR00427	5040	5429	BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II
423	424	RXA01727	GR00489	1471	194	BRANCHED-CHAIN AMINO ACID TRANSPORT SYSTEM
425	426	RXA00623	GR00163	6525	7862	C4-DICARBOXYLATE TRANSPORT PROTEIN
427	428	RXA01584	GR00441	55	597	CHROMATE TRANSPORT PROTEIN
429	430	RXA00852	GR00231	3137	2448	COBALT TRANSPORT ATP-BINDING PROTEIN CBIQ
431	432	RXA00690	GR00181	1213	68	COBALT TRANSPORT PROTEIN CBIQ
433	434	RXA00827	GR00223	1319	567	COBALT TRANSPORT PROTEIN CBIQ
435	436	RXA00851	GR00231	2448	1840	COBALT TRANSPORT PROTEIN CBIQ
437	438	RXS03220				D-XYLOSE-PROTON SYMPORT
439	440	F RXA02762	GR00768	346	630	D-XYLOSE PROTON-SYMPORTER
441	442	RXN00092	VV0129	27509	26844	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
443	444	F RXA00092	GR00014	1	204	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
445	446	RXN03060	VV0030	6227	5376	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
447	448	F RXA02618	GR00745	1914	2351	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
449	450	F RXA02900	GR10040	2979	2128	GLUTAMINE TRANSPORT ATP-BINDING PROTEIN GLNQ
451	452	RXS03212				GLYCINE BETAINE TRANSPORTER BETP
453	454	F RXA01591	GR00446	3	947	GLYCINE BETAINE TRANSPORTER BETP
455	456	RXN00201	VV0096	197	6	HIGH AFFINITY RIBOSE TRANSPORT PROTEIN RBSD
457	458	F RXA00201	GR00032	191	6	HIGH AFFINITY RIBOSE TRANSPORT PROTEIN RBSD
459	460	RXA01221	GR00354	2108	2833	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID
461	462	RXA01222	GR00354	2844	3542	TRANSPORT ATP-BINDING PROTEIN BRAG
463	464	RXA01219	GR00354	151	1032	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID
465	466	RXA01220	GR00354	1032	2108	TRANSPORT PERMEASE PROTEIN LIVH
467	468	RXA00091	GR00013	7762	8514	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID
469	470	RXA00228	GR00032	29232	28642	TRANSPORT PERMEASE PROTEIN LIVM
						IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE
						IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE

TABLE 1-continued

GENES IN THE APPLICATION							
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function	
471	472	RXA00346	GR00064	1054	1743	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE	
473	474	RXA00524	GR00135	779	1111	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE	
475	476	RXA01823	GR00516	591	1367	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE	
477	478	RXA02767	GR00770	1032	1814	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE	
479	480	RXA02792	GR00777	8581	7829	IRON(III) DICITRATE TRANSPORT ATP-BINDING PROTEIN FECE	
481	482	RXN02929	VV0090	36837	37874	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
483	484	F RXA01235	GR00358	1165	194	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
485	486	RXN02794	VV0134	10625	9552	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
487	488	F RXA01419	GR00415	888	1151	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
489	490	F RXA02794	GR00777	10172	9552	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
491	492	RXN03079	VV0045	644	1660	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
493	494	F RXA02865	GR10007	3832	2816	IRON(III) DICITRATE TRANSPORT SYSTEM PERMEASE PROTEIN FECD	
495	496	RXA00181	GR00028	3954	2383	PROLINE TRANSPORT SYSTEM	
497	498	RXA00591	GR00158	229	1581	PROLINE/BETAINE TRANSPORTER	
499	500	RXA01629	GR00453	3476	1965	PROLINE/BETAINE TRANSPORTER	
501	502	RXA02030	GR00618	3072	1687	PROLINE/BETAINE TRANSPORTER	
503	504	RXA00186	GR00028	12242	12988	SHORT-CHAIN FATTY ACIDS TRANSPORTER	
505	506	RXA00187	GR00028	13097	13447	SHORT-CHAIN FATTY ACIDS TRANSPORTER	
507	508	RXA01667	GR00464	703	1908	SODIUM/GLUTAMATE SYMPORT CARRIER PROTEIN	
509	510	RXA02171	GR00641	6571	4919	SODIUM/PROLINE SYMPORTER	
511	512	RXA00902	GR00245	4643	5875	SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN	
513	514	RXA00941	GR00257	1999	683	sodium-dependent phosphate transport protein	
515	516	RXN00449	VV0112	30992	32572	Sodium-Dicarboxylate Symport Protein	
517	518	F RXA00449	GR00109	2040	1036	Sodium-Dicarboxylate Symport Protein	
519	520	F RXA01755	GR00498	352	5	Sodium-Dicarboxylate Symport Protein	
521	522	RXA00269	GR00041	1826	1038	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA	
523	524	RXA00369	GR00076	583	1299	SPERMIDINE/PUTRESCINE TRANSPORT ATP-BINDING PROTEIN POTA	
525	526	RXA02073	GR00628	4176	2647	TRANSPORT ATP-BINDING PROTEIN CYDC	
527	528	RXA01399	GR00409	1	1119	TRANSPORT ATP-BINDING PROTEIN CYDD	
529	530	RXA01339	GR00389	8408	7164	TYROSINE-SPECIFIC TRANSPORT PROTEIN	
531	532	RXA02527	GR00725	5519	6847	2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR	
533	534	RXN00298	VV0176	40228	42072	HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN	
535	536	F RXA00298	GR00048	4459	6303	Ectoine/Proline/Glycine betaine carrier ectP	
537	538	RXA00596	GR00159	335	787	potassium efflux system protein phaE	
539	540	RXA02364	GR00686	841	215	C4-DICARBOXYLATE-BINDING PERIPLASMIC PROTEIN PRECURSOR, transport protein	
541	542	RXN01411	VV0050	26015	26779	SHIKIMATE TRANSPORTER	
543	544	RXN00960	VV0075	1139	105	PROTON/SODIUM-GLUTAMATE SYMPORT PROTEIN	
545	546	RXN02447	VV0107	14297	13203	GALACTOSE-PROTON SYMPORT	
547	548	RXN02395	VV0176	16747	14858	GLYCINE BETAINE TRANSPORTER BETP	
549	550	RXN02348	VV0078	6027	7910	KUP SYSTEM POTASSIUM UPTAKE PROTEIN	
551	552	RXN00297	VV0176	38630	39541	Hypothetical Malonate Transporter	
553	554	RXN03103	VV0070	845	1087	GLUTAMATE-BINDING PROTEIN PRECURSOR	
555	556	RXN02993	VV0071	736	65	GLUTAMINE-BINDING PROTEIN	
557	558	RXN00349	VV0135	35187	36653	Hypothetical Trehalose Transport Protein	
559	560	RXN03095	VV0057	4056	4424	CADMIUM EFFLUX SYSTEM ACCESSORY PROTEIN HOMOLOG	
561	562	RXN03160	VV0189	5150	5617	CHROMATE TRANSPORT PROTEIN	
563	564	RXN02955	VV0176	8666	9187	DICARBOXYLATE TRANSPORTER	
565	566	RXN03109	VV0082	659	6	HEMIN TRANSPORT SYSTEM PERMEASE PROTEIN HMUU	
567	568	RXN02979	VV0149	2150	2383	MERCURIC TRANSPORT PROTEIN PERIPLASMIC COMPONENT PRECURSOR	
569	570	RXN02987	VV0234	527	294	MERCURIC TRANSPORT PROTEIN PERIPLASMIC COMPONENT PRECURSOR	
571	572	RXN03084	VV0048	900	1817	IRON(III) DICITRATE-BINDING PERIPLASMIC PROTEIN PRECURSOR	
573	574	RXN03183	VV0372	1	417	TREHALOSE/MALTOSE BINDING PROTEIN	
575	576	RXN01139	VV0077	2776	1823	CATION EFFLUX SYSTEM PROTEIN CZCD	
577	578	RXN00378	VV0223	8027	5418	Cation transport ATPases	
579	580	RXN01338	VV0032	2	1903	CATION-TRANSPORTING ATPASE PACS (EC 3.6.1.—)	
581	582	RXN00980	VV0149	2635	4428	CATION-TRANSPORTING P-TYPE ATPASE B (EC 3.6.1.—)	
583	584	RXN00099	VV0129	18876	17704	CYANATE TRANSPORT PROTEIN CYNX	
585	586	RXN02662	VV0315	1461	1724	DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC	
587	588	RXN02442	VV0217	5970	6818	zinc transport system membrane protein	
589	590	RXN02443	VV0217	6818	7771	zinc-binding periplasmic protein precursor	
591	592	RXN00842	VV0138	8686	7487	BRANCHED CHAIN AMINO ACID TRANSPORT SYSTEM II CARRIER PROTEIN	
593	594	F RXA00842	GR00228	3208	2009	Permeases	
595	596	RXN00832	VV0180	3133	4182	CALCIUM/PROTON ANTIporter	
597	598	RXN00466	VV0086	63271	64266	Ferrichrome transport proteins	

TABLE 1-continued

GENES IN THE APPLICATION						
Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO	Identification Code	Contig.	NT Start	NT Stop	Function
599	600	RXN01936	VV0127	40116	41387	MACROLIDE-EFFLUX PROTEIN
601	602	RXN01995	VV0182	2139	3476	PUTATIVE 3-(3-HYDROXYPHENYL) PROPIONATE TRANSPORT PROTEIN
603	604	RXN00661	VV0142	9718	9029	PNUC PROTEIN
Permeases						
605	606	RXN02566	VV0154	11823	13031	NUCLEOSIDE PERMEASE NUPG
607	608	F RXA02561	GR00732	664	5	NUCLEOSIDE PERMEASE NUPG
609	610	F RXA02566	GR00733	782	345	NUCLEOSIDE PERMEASE NUPG
611	612	RXA00051	GR00008	5770	7173	PROLINE-SPECIFIC PERMEASE PROY
613	614	RXA01172	GR00334	2687	4141	SULFATE PERMEASE
615	616	RXA02128	GR00637	2906	4600	SULFATE PERMEASE
617	618	RXA02634	GR00748	6045	7655	SULFATE PERMEASE
619	620	RXN02233	VV0068	6856	8142	URACIL PERMEASE
621	622	F RXA02233	GR00653	6856	8067	URACIL PERMEASE
623	624	RXN02372	VV0213	9311	11197	XANTHINE PERMEASE
625	626	F RXA02372	GR00688	6	560	XANTHINE PERMEASE
627	628	F RXA02377	GR00689	3336	4526	XANTHINE PERMEASE
629	630	RXA02676	GR00754	2697	1309	GLUCONATE PERMEASE
631	632	RXN00432	VV0112	14751	13267	NA(+)-LINKED D-ALANINE GLYCINE PERMEASE
633	634	F RXA00432	GR00100	1	891	NA(+)-LINKED D-ALANINE GLYCINE PERMEASE
635	636	F RXA00436	GR00101	45	569	NA(+)-LINKED D-ALANINE GLYCINE PERMEASE
637	638	RXA00847	GR00230	1829	381	OLIGOPEPTIDE-BINDING PROTEIN APPA PRECURSOR (permease)
639	640	RXN01382	VV0119	8670	9761	OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR
641	642	F RXA01382	GR00405	1067	6	OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)
643	644	RXA02659	GR00753	2	313	OLIGOPEPTIDE-BINDING PROTEIN OPPA PRECURSOR (permease)
645	646	RXN02933	VV0176	30042	29233	DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPC
647	648	RXN02991	VV0072	618	4	GLUTAMINE TRANSPORT SYSTEM PERMEASE PROTEIN GLNP
649	650	RXN02992	VV0072	842	621	GLUTAMINE TRANSPORT SYSTEM PERMEASE PROTEIN GLNP
651	652	RXN02996	VV0069	1980	2648	HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PERMEASE PROTEIN LIVH
653	654	RXN03126	VV0112	9894	9001	TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN TAGG
655	656	RXN00443	VV0112	21572	20769	MOLYBDATE-BINDING PERIPLASMIC PROTEIN PRECURSOR
657	658	RXN00444	VV0112	20785	19949	MOLYBDENUM TRANSPORT SYSTEM PERMEASE PROTEIN MODB
659	660	RXN00193	VV0371	1	594	POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
661	662	RXN01298	VV0116	2071	1142	POTENTIAL STARCH DEGRADATION PRODUCTS TRANSPORT SYSTEM PERMEASE PROTEIN AMYD
Channel Proteins						
663	664	RXA01737	GR00493	2913	3971	POTASSIUM CHANNEL PROTEIN
665	666	RXN02348	VV0078	6027	7910	KUP SYSTEM POTASSIUM UPTAKE PROTEIN
667	668	RXA02426	GR00707	2165	633	PROBABLE NA(+)/H(+) ANTIporter
669	670	RXN03164	VV0277	1586	2455	POTASSIUM CHANNEL BETA SUBUNIT
671	672	RXN00024	VV0127	64219	63275	POTASSIUM CHANNEL BETA SUBUNIT
Lipoprotein and Lipopolysaccharide synthesis						
673	674	RXN01164	VV0117	15894	14260	DOLICHOL-PHOSPHATE MANNOsylTRANSFERASE (EC 2.4.1.83)/ APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.—)
675	676	RXN01168	VV0117	14224	13415	DOLICHOL-PHOSPHATE MANNOsylTRANSFERASE (EC 2.4.1.83)/ APOLIPOPROTEIN N-ACYLTRANSFERASE (EC 2.3.1.—)

TABLE 2

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
A09073	ppg	Phosphoenol pyruvate carboxylase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvate carboxylase, recombinant DNA carrying said fragment, strains carrying the recombinant DNA and method for producing L-aminino acids using said strains," Patent: EP 0358940-A 3 Mar. 21, 1990
A45579, A45581, A45583, A45585 A45587		Threonine dehydratase	Moeckel, B. et al. "Production of L-isoleucine by means of recombinant micro-organisms with deregulated threonine dehydratase," Patent: WO 9519442-A 5 Jul. 20, 1995
AB003132	murC; ftsQ; ftsZ		Kobayashi, M. et al. "Cloning, sequencing, and characterization of the ftsZ gene from <i>Corynebacterium</i> bacteria," Biochem. Biophys. Res. Commun., 236(2): 383-388 (1997)
AB015023	murC; ftsQ		Wachi, M. et al. "A murC gene from <i>Corynebacterium</i> bacteria," Appl. Microbiol. Biotechnol., 51(2): 223-228 (1999)
AB018530	ftsR		Kinura, E. et al. "Molecular cloning of a novel gene, ftsR, which rescues the detergent sensitivity of a mutant derived from <i>Brevibacterium lactofermentum</i> ," Biosci. Biotechnol. Biochem., 60(10): 1565-1570 (1996)
AB018531	ftsR1; ftsR2		
AB020624	murI	D-glutamate racemase	
AB023377	tkt	transketolase	
AB024708	glbB; glbD	Glutamine 2-oxoglutarate aminotransferase large and small subunits	
AB025424	acn	aconitase	
AB027714	rep	Replication protein	
AB027715	rep; aad	Replication protein; aminoglycoside adenyltransferase	
AF005242	argC	N-acetylglutamate-5-semialdehyde dehydrogenase	
AF005635	glnA	Glutamine synthetase	
AF030405	hisF	cyclase	
AF030520	argG	Argininosuccinate synthetase	
AF031518	argI	Ornithine carbamoyltransferase	
AF036932	aroD	3-dehydroquinate dehydratase	
AF038548	pyc	Pyruvate carboxylase	
AF038651	dcIAE; apt; rel	Dipeptide-binding protein; adenine phosphoribosyltransferase; GDP pyrophosphokinase	Wehmeier, L. et al. "The role of the <i>Corynebacterium glutamicum</i> rel gene in (p)ppGpp metabolism," Microbiology, 144: 1853-1862 (1998)
AF041436	argR	Arginine repressor	
AF045998	impA	Inositol monophosphate phosphatase	
AF048764	argH	Argininosuccinate lyase	
AF049897	argC; argI; argB; argD; argF; argR; argG; argH	N-acetylglutamylphosphate reductase; ornithine acetyltransferase; N-acetylglutamate kinase; acetylornithine transaminase; ornithine carbamoyltransferase; arginine repressor; argininosuccinate synthase; argininosuccinate lyase	

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
AF050109	inhA	Enoyl-acyl carrier protein reductase	Park, S. et al. "Isolation and analysis of metA, a methionine biosynthetic gene encoding homoserine acetyltransferase in <i>Corynebacterium glutamicum</i> ," Mol. Cells, 8(3): 286–294 (1998)
AF050166	hisG	ATP phosphoribosyltransferase	
AF051846	hisA	Phosphoribosylformimino-5-amino-1-phosphoribosyl-4-imidazolecarboxamide isomerase	
AF052652	metA	Homoserine O-acetyltransferase	
AF053071	aroB	Dehydroquininate synthetase	Dusch, N. et al. "Expression of the <i>Corynebacterium glutamicum</i> panD gene encoding L-aspartate-alpha-decarboxylase leads to pantothenate overproduction in <i>Escherichia coli</i> ," Appl. Environ. Microbiol., 65(4)1530–1539 (1999)
AF060558	hisH	Glutamine amidotransferase	
AF086704	hisE	Phosphoribosyl-ATP-pyrophosphohydrolase	
AF114233	aroA	5-enolpyruvylshikimate 3-phosphate synthase	
AF116184	panD	L-aspartate-alpha-decarboxylase precursor	Peter, H. et al. " <i>Corynebacterium glutamicum</i> is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP," J. Bacteriol., 180(22): 6005–6012 (1998)
AF124518	aroD; aroE	3-dehydroquinase; shikimate dehydrogenase	
AF124600	aroC; aroK; aroB; pepQ	Chorismate synthase; shikimate kinase; 3-dehydroquininate synthase; putative cytoplasmic peptidase	
AF145897	inhA	Transport of ectoine, glycine betaine, proline	
AF145898	inhA		
AJ001436	ectP	Tetrahydrodipicolinate succinylase (incomplete)	Wehmann, A. et al. "Different modes of diaminopimelate synthesis and their role in cell wall integrity: A study with <i>Corynebacterium glutamicum</i> ," J. Bacteriol., 180(12): 3159–3165 (1998)
AJ004934	dapD	Phosphoenolpyruvate-carboxylase; ?, high affinity ammonium uptake protein; putative ornithine-cyclodecarboxylase; sarcosine oxidase	
AJ007732	ppc; secG; amt; ocd; soxA	Involved in cell division; PII protein; uridylyltransferase (uridylyl-removing enzyme); signal recognition particle; low affinity ammonium uptake protein	
AJ010319	ftsY, gltB, gltD; srp; amtP	Chloramphenicol acetyl transferase L-malate: quinone oxidoreductase	
AJ132968	cat	NADH dehydrogenase	Jakoby, M. et al. "Nitrogen regulation in <i>Corynebacterium glutamicum</i> ; Isolation of genes involved in biochemical characterization of corresponding proteins," FEMS Microbiol., 173(2): 303–310 (1999)
AJ224946	mgo		
AJ238250	ndh	Porin	Molenaar, D. et al. "Biochemical and genetic characterization of the membrane-associated malate dehydrogenase (acceptor) from <i>Corynebacterium glutamicum</i> ," Eur. J. Biochem., 254(2): 395–403 (1998)
AJ238703	porA		
			Lichtinger, T. et al. "Biochemical and biophysical characterization of the cell wall porin of <i>Corynebacterium glutamicum</i> : The channel is formed by a low molecular mass polypeptide," Biochemistry, 37(43): 15024–15032 (1998)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			Reference
GenBank™ Accession No.	Gene Name	Gene Function	
D17429		Transposable element IS1831	Vertes, A. A. et al. "Isolation and characterization of IS1831, a transposable element from <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 11(4): 739-746 (1994)
D84102	odhA	2-oxoglutarate dehydrogenase	Usuda, Y. et al. "Molecular cloning of the <i>Corynebacterium glutamicum</i> (<i>Brevibacterium lactofermentum</i> AJ12036) odhA gene encoding a novel type of 2-oxoglutarate dehydrogenase," Microbiology, 142: 3347-3354 (1996)
E01358	hdh; hk	Homoserine dehydrogenase; homoserine kinase	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 1 Oct. 12, 1987
E01359		Upstream of the start codon of homoserine kinase gene	Katsumata, R. et al. "Production of L-threonine and L-isoleucine," Patent: JP 1987232392-A 2 Oct. 12, 1987
E01375	trpL; trpE	Tryptophan operon	
E01376		Leader peptide; anthranilate synthase	
E01377		Promoter and operator regions of tryptophan operon	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 Oct. 24, 1987
E03937		Biotin-synthase	Matsui, K. et al. "Tryptophan operon, peptide and protein coded thereby, utilization of tryptophan operon gene expression and production of tryptophan," Patent: JP 1987244382-A 1 Oct. 24, 1987
E04040		Diamino pelargonic acid aminotransferase	Hatakeyama, K. et al. "DNA fragment containing gene capable of coding biotin synthetase and its utilization," Patent: JP 1992278088-A 1 Oct. 02, 1992
E04041		Desthiobiotinsynthetase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 Nov. 18, 1992
E04307		Flavum aspartase	Kohama, K. et al. "Gene coding diaminopelargonic acid aminotransferase and desthiobiotin synthetase and its utilization," Patent: JP 1992330284-A 1 Nov. 18, 1992
E04376		Isocitric acid lyase	Kurusu, Y. et al. "Gene DNA coding aspartase and utilization thereof," Patent: JP 1993030977-A 1 Feb. 09, 1993
A09073	ppg	Phosphoenol pyruvate carboxylase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 Mar. 09, 1993
E04377		Isocitric acid lyase N-terminal fragment	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvate
E04484		Prephenate dehydratase	Katsumata, R. et al. "Gene manifestation controlling DNA," Patent: JP 1993056782-A 3 Mar. 09, 1993
E05108		Aspartokinase	Sotouchi, N. et al. "Production of L-phenylalanine by fermentation," Patent: JP 1993076352-A 2 Mar. 30, 1993
E05112		Dihydro-dipicolinate synthetase	Fugono, N. et al. "Gene DNA coding Aspartokinase and its use," Patent: JP 1993184366-A 1 Jul. 27, 1993
E05776		Diaminopimelic acid dehydrogenase	Hatakeyama, K. et al. "Gene DNA coding dihydrodipicolinic acid synthetase and its use," Patent: JP 1993184371-A 1 Jul. 27, 1993
E05779		Threonine synthase	Kobayashi, M. et al. "Gene DNA coding Diaminopimelic acid dehydrogenase and its use," Patent: JP 1993284970-A 1 Nov. 02, 1993
E06110		Prephenate dehydratase	Kohama, K. et al. "Gene DNA coding threonine synthase and its use," Patent: JP 1993284972-A 1 Nov. 02, 1993
E06111		Mutated Prephenate dehydratase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 Dec. 27, 1993
E06146		Acetohydroxy acid synthetase	Kikuchi, T. et al. "Production of L-phenylalanine by fermentation method," Patent: JP 1993344881-A 1 Dec. 27, 1993
			Inui, M. et al. "Gene capable of coding Acetohydroxy acid synthetase and its use," Patent: JP 1993344893-A 1 Dec. 27, 1993

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK		
GenBank™ Accession No.	Gene Name	Gene Function Reference
E06825		Aspartokinase Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E06826		Mutated aspartokinase alpha subunit Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E06827		Mutated aspartokinase alpha subunit Sugimoto, M. et al. "Mutant aspartokinase gene," patent: JP 1994062866-A 1 Mar. 08, 1994
E07701	secY	Honno, N. et al. "Gene DNA participating in integration of membraneous protein to membrane," Patent: JP 1994169780-A 1 Jun. 21, 1994
E08177		Aspartokinase Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 Sep. 20, 1994
E08178, E08179, E08180, E08181, E08182 E08232		Sato, Y. et al. "Genetic DNA capable of coding Aspartokinase released from feedback inhibition and its utilization," Patent: JP 1994261766-A 1 Sep. 20, 1994
E08234	secE	Inui, M. et al. "Gene DNA coding acetohydroxy acid isomeroreductase," Patent: JP 1994277067-A 1 Oct. 04, 1994
E08643		Asai, Y. et al. "Gene DNA coding for translocation machinery of protein," Patent: JP 1994277073-A 1 Oct. 04, 1994
E08646		Hatakeyama, K. et al. "DNA fragment having promoter function in <i>Corynebacterium</i> ," Patent: JP 1995031476-A 1 Feb. 03, 1995
E08649		Hatakeyama, K. et al. "DNA fragment having promoter function in <i>Corynebacterium</i> ," Patent: JP 1995031476-A 1 Feb. 03, 1995
E08900		Kohama, K. et al. "DNA fragment having promoter function in <i>Corynebacterium</i> ," Patent: JP 1995031478-A 1 Feb. 03, 1995
E08901		Madori, M. et al. "DNA fragment containing gene coding Dihydrodipicolinate acid reductase and utilization thereof," Patent: JP 1995075578-A 1 Mar. 20, 1995
E12594		Madori, M. et al. "DNA fragment containing gene coding Diaminopimelic acid decarboxylase and utilization thereof," Patent: JP 1995075579-A 1 Mar. 20, 1995
E12760, E12759, E12758 E12764		Hatakeyama, K. et al. "Production of L-tryptophan," Patent: JP 1997028391-A 1 Feb. 4, 1997
E12767		Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12770		Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E12773		Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
E13655		Moriya, M. et al. "Amplification of gene using artificial transposon," Patent: JP 1997070291-A Mar. 18, 1997
L01508	lvaA	Hatakeyama, K. et al. "Glucose-6-phosphate dehydrogenase and DNA capable of coding the same," Patent: JP 1997224661-A 1 Sep. 02, 1997 Moeckel, B. et al. "Functional and structural analysis of the threonine dehydratase of <i>Corynebacterium glutamicum</i> ," J. Bacteriol., 174: 8065-8072 (1992)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
L07603	EC 4.2.1.15	3-deoxy-D-arabinoheptulosonate-7-phosphate synthase	Chen, C. et al. "The cloning and nucleotide sequence of <i>Corynebacterium glutamicum</i> 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase gene," FEMS Microbiol. Lett., 107: 223-230 (1993)
L09232	ilvB; ilvN; ilvC	Acetohydroxy acid synthase large subunit; Acetohydroxy acid synthase small subunit;	Keilhauer, C. et al. "Isoleucine synthesis in <i>Corynebacterium glutamicum</i> : molecular analysis of the ilvB-ilvN-ilvC operon," J. Bacteriol., 175(17): 5595-5603 (1993)
L18874	PtsM	Acetohydroxy acid isomeroreductase Phosphoenolpyruvate sugar phosphotransferase	Fouet, A. et al. " <i>Bacillus subtilis</i> sucrose-specific enzyme II of the phosphotransferase system: expression in <i>Escherichia coli</i> and homology to enzymes II from enteric bacteria," PNAS USA, 84(24): 8773-8777 (1987); Lee, J. K. et al. "Nucleotide sequence of the gene encoding the <i>Corynebacterium glutamicum</i> mannose enzyme II and analyses of the deduced protein sequence," FEMS Microbiol. Lett., 119(1-2): 137-145 (1994)
L27123	aceB	Malate synthase	Lee, H-S. et al. "Molecular characterization of aceB, a gene encoding malate synthase in <i>Corynebacterium glutamicum</i> ," J. Microbiol. Biotechnol., 4(4): 256-263 (1994)
L27126		Pyruvate kinase	Jetten, M. S. et al. "Structural and functional analysis of pyruvate kinase from <i>Corynebacterium glutamicum</i> ," Appl. Environ. Microbiol., 60(7): 2501-2507 (1994)
L28760	aceA	Isocitrate lyase	Oguiza, J. A. et al. "Molecular cloning, DNA sequence analysis, and characterization of the <i>Corynebacterium diptheriae</i> dtxR from <i>Brevibacterium lactofermentum</i> ," J. Bacteriol., 177(2): 465-467 (1995)
L35906	dtxR	Diphtheria toxin repressor	Follettie, M. T. et al. "Molecular cloning and nucleotide sequence of the <i>Corynebacterium glutamicum</i> pheA gene," J. Bacteriol., 167: 695-702 (1986)
M13774		Prephenate dehydratase	Park, Y-H. et al. "Phylogenetic analysis of the <i>coryneform</i> bacteria by 5S rRNA sequences," J. Bacteriol., 169: 1801-1806 (1987)
M16175	5S rRNA		Sano, K. et al. "Structure and function of the trp operon control regions of <i>Brevibacterium lactofermentum</i> , a glutamic-acid-producing bacterium," Gene, 52: 191-200 (1987)
M16663	trpE	Anthranilate synthase, 5' end	Sano, K. et al. "Structure and function of the trp operon control regions of <i>Brevibacterium lactofermentum</i> , a glutamic-acid-producing bacterium," Gene, 52: 191-200 (1987)
M16664	trpA	Tryptophan synthase, 3'end	O'Regan, M. et al. "Cloning and nucleotide sequence of the phosphoenolpyruvate carboxylase-coding gene of <i>Corynebacterium glutamicum</i> ATCC13032," Gene, 77(2): 237-251 (1989)
M25819		Phosphoenolpyruvate carboxylase	Roller, C. et al. "Gram-positive bacteria with a high DNA G + C content are characterized by a common insertion within their 23S rRNA genes," J. Gen. Microbiol., 138: 1167-1175 (1992)
M85106		23S rRNA gene insertion sequence	Roller, C. et al. "Gram-positive bacteria with a high DNA G + C content are characterized by a common insertion within their 23S rRNA genes," J. Gen. Microbiol., 138: 1167-1175 (1992)
M85107, M85108		23S rRNA gene insertion sequence	Rosoli, I. et al. "The <i>Corynebacterium glutamicum</i> aceD gene encodes a C-S lyase with alpha, beta-elimination activity that degrades aminoethylcysteine," J. Bacteriol., 174(9): 2968-2977 (1992); Tauch, A. et al. "Isoleucine uptake in <i>Corynebacterium glutamicum</i> ATCC 13032 is directed by the brnQ gene product," Arch. Microbiol., 169(4): 303-312 (1998)
M89931	aceD; brnQ; yhbW	Beta C-S lyase; branched-chain amino acid uptake carrier; hypothetical protein yhbW	

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
S59299	trp	Leader gene (promoter)	Herry, D. M. et al. "Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of <i>Corynebacterium glutamicum</i> ; identification of a mutation in the trp leader sequence," Appl. Environ. Microbiol., 59(3): 791-799 (1993)
U11545	trpD	Anthranilate phosphoribosyltransferase	O'Gara, J. P. and Dunican, L. K. (1994) Complete nucleotide sequence of the <i>Corynebacterium glutamicum</i> ATCC 21850 trpD gene," Thesis, Microbiology Department, University College Galway, Ireland.
U13922	cgIM; cgIR; cglIR	Putative type II 5-cytosine methyltransferase; putative type II restriction endonuclease; putative type I or type III restriction endonuclease	Schafer, A. et al. "Cloning and characterization of a DNA region encoding a stress-sensitive restriction system from <i>Corynebacterium glutamicum</i> ATCC 13032 and analysis of its role in intergeneric conjugation with <i>Escherichia coli</i> ," J. Bacteriol., 176(23): 7309-7319 (1994); Schafer, A. et al. "The <i>Corynebacterium glutamicum</i> cglIM gene encoding a 5-cytosine in an MerBC-deficient <i>Escherichia coli</i> strain," Gene, 203(2): 95-101 (1997)
U14965	recA		Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," J. Bacteriol., 178(15): 4412-4419 (1996)
U31224	ppx		Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," J. Bacteriol., 178(15): 4412-4419 (1996)
U31225	proC	L-proline: NADP+ 5-oxidoreductase	Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," J. Bacteriol., 178(15): 4412-4419 (1996)
U31230	obg; proB; unkdh	?; gamma glutaryl kinase; similar to D-isomer specific 2-hydroxyacid dehydrogenases	Ankri, S. et al. "Mutations in the <i>Corynebacterium glutamicum</i> proline biosynthetic pathway: A natural bypass of the proA step," J. Bacteriol., 178(15): 4412-4419 (1996)
U31281	bioB	Biotin synthase	Serebriskii, I. G., "Two new members of the bio B superfamily: Cloning, sequencing and expression of bio B genes of <i>Methylobacillus flagellatum</i> and <i>Corynebacterium glutamicum</i> ," Gene, 175: 15-22 (1996)
U35023	thrR; accBC	Thiosulfate sulfurtransferase; acyl CoA carboxylase	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene encoding a two-domain protein similar to biotin carboxylases and biotin-carboxyl-carrier proteins," Arch. Microbiol., 166(2): 76-82 (1996)
U43535	cmr	Multidrug resistance protein	Jager, W. et al. "A <i>Corynebacterium glutamicum</i> gene conferring multidrug resistance in the heterologous host <i>Escherichia coli</i> ," J. Bacteriol., 179(7): 2449-2451 (1997)
U43536	clpB	Heat shock ATP-binding protein	
U53587	aphA-3	3'S'-aminoglycoside phosphotransferase	
U89648		<i>Corynebacterium glutamicum</i> unidentified sequence involved in histidine biosynthesis, partial sequence	
X04960	trpA; trpB; trpC; trpD; trpE; trpG; trpL	Tryptophan operon	Matsui, K. et al. "Complete nucleotide and deduced amino acid sequences of the <i>Brevibacterium lactofermentum</i> tryptophan operon," Nucleic Acids Res., 14(24): 10113-10114 (1986)
X07563	lys A	DAP decarboxylase (meso-diaminopimelate decarboxylase, EC 4.1.1.20)	Yeh, P. et al. "Nucleic sequence of the lysA gene of <i>Corynebacterium glutamicum</i> and possible mechanisms for modulation of its expression," Mol. Gen. Genet., 212(1): 112-119 (1988)
X14234	EC 4.1.1.31	Phosphoenolpyruvate carboxylase	Eikmanns, B. J. et al. "The Phosphoenolpyruvate carboxylase gene of <i>Corynebacterium glutamicum</i> : Molecular cloning, nucleotide sequence, and expression," Mol. Gen. Genet., 218(2): 330-339 (1989); Lepiniec, L. et al. "Sorghum Phosphoenolpyruvate carboxylase gene family: structure, function and molecular evolution," Plant. Mol. Biol., 21 (3): 487-502 (1993)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
X17313	fdA	Fructose-bisphosphate aldolase	Von der Osten, C. H. et al. "Molecular cloning, nucleotide sequence and fine-structural analysis of the <i>Corynebacterium glutamicum</i> fdA gene: structural comparison of <i>C. glutamicum</i> fructose-1,6-bisphosphate aldolase to class I and class II aldolases," Mol. Microbiol.,
X53993	dapA	L-2,3-dihydrodipicolinate synthetase (EC 4.2.1.52)	Bonnasse, S. et al. "Nucleic sequence of the dapA gene from <i>Corynebacterium glutamicum</i> ," Nucleic Acids Res., 18(21): 6421 (1990)
X54223		AttB-related site	Ciancioffo, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diptheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of lambdaCorynephage," FEMS. Microbiol. Lett., 66: 299-302 (1990)
X54740	argS; lysA	Arginyl-tRNA synthetase; Diaminopimelate decarboxylase	Marcel, T. et al. "Nucleotide sequence and organization of the upstream region of the <i>Corynebacterium glutamicum</i> lysA gene," Mol. Microbiol., 4(11): 1819-1830 (1990)
X55994	trpL; trpE	Putative leader peptide; anthranilate synthase component 1	Heery, D. M. et al. "Nucleotide sequence of the <i>Corynebacterium glutamicum</i> trpE gene," Nucleic Acids Res., 18(23): 7138 (1990)
X56037	thrC	Threonine synthase	Han, K. S. et al. "The molecular structure of the <i>Corynebacterium glutamicum</i> threonine synthase gene," Mol. Microbiol., 4(10): 1693-1702 (1990)
X56075	attB-related site	Attachment site	Ciancioffo, N. et al. "DNA sequence homology between att B-related sites of <i>Corynebacterium diptheriae</i> , <i>Corynebacterium ulcerans</i> , <i>Corynebacterium glutamicum</i> , and the attP site of lambdaCorynephage," FEMS. Microbiol. Lett., 66: 299-302 (1990)
X57226	lysC-alpha; lysC-beta; asd	Aspartokinase-alpha subunit; Aspartokinase-beta subunit; aspartate beta semialdehyde dehydrogenase	Kalinowski, J. et al. "Genetic and biochemical analysis of the Aspartokinase from <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 5(5): 1197-1204 (1991); Kalinowski, J. et al. "Aspartokinase genes lysC alpha and lysC beta overlap and are adjacent to the aspartate beta-semialdehyde dehydrogenase gene asd in <i>Corynebacterium glutamicum</i> ," Mol. Gen. Genet., 224(3): 317-324 (1990)
X59403	gap; pgk; tpi	Glyceraldehyde-3-phosphate; phosphoglycerate kinase; triosephosphate isomerase	Eikmanns, B. J. "Identification, sequence analysis, and expression of a <i>Corynebacterium glutamicum</i> gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate kinase, and triosephosphate isomerase," J. Bacteriol., 174(19): 6076-6086 (1992)
X59404	gdh	Glutamate dehydrogenase	Bormann, E. R. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> gdh gene encoding glutamate dehydrogenase," Mol. Microbiol., 6(3): 317-326 (1992)
X60312	lysI	L-lysine permease	Seep-Feldhaus, A. H. et al. "Molecular analysis of the <i>Corynebacterium glutamicum</i> lysI gene involved in lysine uptake," Mol. Microbiol., 5(12): 2995-3005 (1991)
X66078	cop1	Ps1 protein	Joliff, G. et al. "Cloning and nucleotide sequence of the csp1 gene encoding PS1, one of the two major secreted proteins of <i>Corynebacterium glutamicum</i> : The deduced N-terminal region of PS1 is similar to the Mycobacterium antigen 85 complex," Mol. Microbiol., 6(16): 2349-2362 (1992)
X66112	glt	Citrate synthase	Eikmanns, B. J. et al. "Cloning sequence, expression and transcriptional analysis of the <i>Corynebacterium glutamicum</i> gltA gene encoding citrate synthase," Microbiol., 140: 1817-1828 (1994)
X67737	dapB	Dihydrodipicolinate reductase	Peyret, J. L. et al. "Characterization of the cspB gene encoding PS2, an ordered surface-layer protein in <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 9(1): 97-109 (1993)
X69103	csp2	Surface layer protein PS2	

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
X69104		IS3 related insertion element	Bonamy, C. et al. "Identification of IS1206, a <i>Corynebacterium glutamicum</i> IS3-related insertion sequence and phylogenetic analysis," Mol. Microbiol., 14(3): 571-581 (1994)
X70959	leuA	Isopropylmalate synthase	Patek, M. et al. "Leucine synthesis in <i>Corynebacterium glutamicum</i> : enzyme activities, structure of leuA, and effect of leuA inactivation on lysine synthesis," Appl. Environ. Microbiol., 60(1): 133-140 (1994)
X71489	icd	Isocitrate dehydrogenase (NADP+)	Eikmanns, B. J. et al. "Cloning sequence analysis, expression, and inactivation of the <i>Corynebacterium glutamicum</i> icd gene encoding isocitrate dehydrogenase and biochemical characterization of the enzyme," J. Bacteriol., 177(3): 774-782 (1995)
X72855	GDHA	Glutamate dehydrogenase (NADP+)	Heery, D. M. et al. "A sequence from a tryptophan-hyperproducing strain of <i>Corynebacterium glutamicum</i> encoding resistance to 5-methyltryptophan," Biochem. Biophys. Res. Commun., 201(3): 1255-1262 (1994)
X75083, X70584	mtrA	5-methyltryptophan resistance	Fitzpatrick, R. et al. "Construction and characterization of recA mutant strains of <i>Corynebacterium glutamicum</i> and <i>Brevibacterium lactofermentum</i> ," Appl. Microbiol. Biotechnol., 42(4): 575-580 (1994)
X75085	recA		Reinisch, D. J. et al. "Characterization of the isocitrate lyase gene from <i>Corynebacterium glutamicum</i> and biochemical analysis of the enzyme," J. Bacteriol., 176(12): 3474-3483 (1994)
X75504	aceA; thiX	Partial Isocitrate lyase; ?	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," Antonie Van Leeuwenhoek, 64: 285-305 (1993)
X76875		ATPase beta-subunit	Ludwig, W. et al. "Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes," Antonie Van Leeuwenhoek, 64: 285-305 (1993)
X77034	tuf	Elongation factor Tu	Billman-Jacobe, H. "Nucleotide sequence of a recA gene from <i>Corynebacterium glutamicum</i> ," DNA Seq., 4(6): 403-404 (1994)
X77384	recA		Reinisch, D. J. et al. "Malate synthase from <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase: sequence analysis," Microbiology, 140: 3099-3108 (1994)
X78491	aceB	Malate synthase	Rainey, F. A. et al. "Phylogenetic analysis of the genera Rhodococcus and Norcardia and evidence for the evolutionary origin of the genus Norcardia from within the radiation of Rhodococcus species," Microbiol., 141: 523-528 (1995)
X80629	16S rDNA	16S ribosomal RNA	Kronmeyer, W. et al. "Structure of the gluABCD cluster encoding the glutamate uptake system of <i>Corynebacterium glutamicum</i> ," J. Bacteriol., 177(5): 1152-1158 (1995)
X81191	gluA; gluB; gluC; gluD	Glutamate uptake system	Wehrmann, A. et al. "Analysis of different DNA fragments of <i>Corynebacterium glutamicum</i> complementing dapE of <i>Escherichia coli</i> ," Microbiology, 40: 3349-36 (1994)
X81379	dapE	Succinyl/diaminopimelate desuccinylase	Ruimy, R. et al. "Phylogeny of the genus <i>Corynebacterium</i> deduced from analyses of small-subunit ribosomal DNA sequences," Int. J. Syst. Bacteriol., 45(4): 740-746 (1995)
X82061	16S rDNA	16S ribosomal RNA	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous proA in proA mutants," J. Bacteriol., 177(24): 7255-7260 (1995)
X82928	asd; lysC	Aspartate-semialdehyde dehydrogenase; ?	

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
X82929	pta	Gamma-glutamyl phosphate reductase	Serebrijski, I. et al. "Multicopy suppression by asd gene and osmotic stress-dependent complementation by heterologous pta in proA mutants," J. Bacteriol., 177(24): 7255-7260 (1995)
X84257	16S rDNA	16S ribosomal RNA	Pascual, C. et al. "Phylogenetic analysis of the genus <i>Corynebacterium</i> based on 16S rRNA gene sequences," Int. J. Syst. Bacteriol., 45(4): 724-728 (1995)
X85965	aroP; dapE	Aromatic amino acid permease; ?	Wehmann, A. et al. "Functional analysis of sequences adjacent to dapE of <i>Corynebacterium glutamicum</i> proline reveals the presence of aroP, which encodes the aromatic amino acid transporter," J. Bacteriol., 177(20): 5991-5993 (1995)
X86157	argB; argC; argD; argF; argJ	Acetylglutamate kinase; N-acetyl-gamma-glutamyl-phosphate reductase; acetylornithine aminotransferase; ornithine carbamoyltransferase; glutamate N-acetyltransferase	Sakanyan, V. et al. "Genes and enzymes of the acetyl cycle of arginine biosynthesis in <i>Corynebacterium glutamicum</i> : enzyme evolution in the early steps of the arginine pathway," Microbiology, 142: 99-108 (1996)
X89084	pta; ackA	Phosphate acetyltransferase; acetate kinase	Reinscheid, D. J. et al. "Cloning, sequence analysis, expression and inactivation of the <i>Corynebacterium glutamicum</i> pta-ack operon encoding phosphotransacetylase and acetate kinase," Microbiology, 145: 503-513 (1999)
X89850	attB	Attachment site	Le Marrec, C. et al. "Genetic characterization of site-specific integration functions of phi AAU2 infecting "Arthrobacter aureus C70," J. Bacteriol., 178(7): 1996-2004 (1996)
X90356	Promoter fragment F1	Promoter fragment F1	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90357	Promoter fragment F2	Promoter fragment F2	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90358	Promoter fragment F10	Promoter fragment F10	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90359	Promoter fragment F13	Promoter fragment F13	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90360	Promoter fragment F22	Promoter fragment F22	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90361	Promoter fragment F34	Promoter fragment F34	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90362	Promoter fragment F37	Promoter fragment F37	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90363	Promoter fragment F45	Promoter fragment F45	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90364	Promoter fragment F64	Promoter fragment F64	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
X90365		Promoter fragment F75	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90366		Promoter fragment PF101	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90367		Promoter fragment PF104	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X90368		Promoter fragment PF109	Patek, M. et al. "Promoters from <i>Corynebacterium glutamicum</i> : cloning, molecular analysis and search for a consensus motif," Microbiology, 142: 1297-1309 (1996)
X93513	amt	Ammonium transport system	Siewe, R. M. et al. "Functional and genetic characterization of the (methyl) ammonium uptake carrier of <i>Corynebacterium glutamicum</i> ," J. Biol. Chem., 271(10): 5398-5403 (1996)
X93514	betP	Glycine betaine transport system	Peter, H. et al. "Isolation, characterization, and expression of the <i>Corynebacterium glutamicum</i> betP gene, encoding the transport system for the compatible solute glycine betaine," J. Bacteriol., 178(17): 5229-5234 (1996)
X95649	orf4		Patek, M. et al. "Identification and transcriptional analysis of the dapB-ORF2-dapA-ORF4 operon of <i>Corynebacterium glutamicum</i> , encoding two enzymes involved in L-lysine synthesis," Biotechnol. Lett., 19: 1113-1117 (1997)
X96471	lysE; lysG	Lysine exporter protein; Lysine export regulator protein	Vrijic, M. et al. "A new type of transporter with a new type of cellular function: L-lysine export from <i>Corynebacterium glutamicum</i> ," Mol. Microbiol., 22(5): 815-826 (1996)
X96580	panB; panC; xylB	3-methyl-2-oxobutanoate hydroxymethyltransferase; pantoate-beta-alanine ligase; xylulokinase	Sahm, H. et al. "D-pantothenate synthesis in <i>Corynebacterium glutamicum</i> and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction," Appl. Environ. Microbiol., 65(5): 1973-1979 (1999)
X96962		Insertion sequence IS1207 and transposase	
X99289		Elongation factor P	Ramos, A. et al. "Cloning, sequencing and expression of the gene encoding elongation factor P in the amino-acid producer <i>Brevibacterium lactofermentum</i> (<i>Corynebacterium glutamicum</i> ATCC 13869)," Gene, 198: 217-222 (1997)
Y00140	thrB	Homoserine kinase	Mateos, L. M. et al. "Nucleotide sequence of the homoserine kinase (thrB) gene of the <i>Brevibacterium lactofermentum</i> ," Nucleic Acids Res., 15(9): 3922 (1987)
Y00151	ddl	Meso-diaminopimelate D-dehydrogenase (EC 1.4.1.16)	Ishino, S. et al. "Nucleotide sequence of the meso-diaminopimelate D-dehydrogenase gene from <i>Corynebacterium glutamicum</i> ," Nucleic Acids Res., 15(9): 3917 (1987)
Y00476	thrA	Homoserine dehydrogenase	Mateos, L. M. et al. "Nucleotide sequence of the homoserine dehydrogenase (thrA) gene of the <i>Brevibacterium lactofermentum</i> ," Nucleic Acids Res., 15(24): 10598 (1987)
Y00546	hom; thrB	Homoserine dehydrogenase; homoserine kinase	Peoples, O. P. et al. "Nucleotide sequence and fine structural analysis of the <i>Corynebacterium glutamicum</i> hom-thrB operon," Mol. Microbiol., 2(1): 63-72 (1988)
Y08964	murC; ftsQ/divD; ftsZ	UPD-N-acetylmuramate-alanine ligase; division initiation protein or cell division protein; cell division protein	Honrubia, M. P. et al. "Identification, characterization, and chromosomal organization of the ftsZ gene from <i>Brevibacterium lactofermentum</i> ," Mol. Gen. Genet., 259(1): 97-104 (1998)
Y09163	putP	High affinity proline transport system	Peter, H. et al. "Isolation of the putP gene of <i>Corynebacterium glutamicum</i> and characterization of a low-affinity uptake system for compatible solutes," Arch. Microbiol., 168(2): 143-151 (1997)

TABLE 2-continued

GENES IDENTIFIED FROM GENBANK			
GenBank™ Accession No.	Gene Name	Gene Function	Reference
Y09548	pyc	Pyruvate carboxylase	Peters-Weidisch, P. G. et al. "Pyruvate carboxylase from <i>Corynebacterium glutamicum</i> : characterization, expression and inactivation of the pyc gene," Microbiology, 144: 915-927 (1998)
Y09578	lenB	3-isopropylmalate dehydrogenase	Patek, M. et al. "Analysis of the lenB gene from <i>Corynebacterium glutamicum</i> ," Appl. Microbiol. Biotechnol., 50(1): 42-47 (1998)
Y12472		Attachment site bacteriophage Phi-16	Moreau, S. et al. "Site-specific integration of coryneophage Phi-16: The construction of an integration vector," Microbiol., 145: 539-548 (1999)
Y12537	proP	Proline/ectoine uptake system protein	Peter, H. et al. " <i>Corynebacterium glutamicum</i> is equipped with four secondary carriers for compatible solutes: Identification, sequencing, and characterization of the proline/ectoine uptake system, proP, and the ectoine/proline/glycine betaine carrier, EctP," J. Bacteriol., 180(22): 6005-6012 (1998)
Y13221	glnA	Glutamine synthetase I	Jakoby, M. et al. "Isolation of <i>Corynebacterium glutamicum</i> glnA gene encoding glutamine synthetase I," FEMS Microbiol. Lett., 154(1): 81-88 (1997)
Y16642	lpd	Dihydroliipoamide dehydrogenase	Moreau, S. et al. "Analysis of the integration functions of ϕ 304L: An integrase module among corynephages," Virology, 255(1): 150-159 (1999)
Y18059		Attachment site Corynephage 304L	Oguiza, J. A. et al. "A gene encoding arginyl-tRNA synthetase is located in the upstream region of the lysA gene in <i>Brevibacterium lactofermentum</i> : Regulation of argS-lysA cluster expression by arginine," J. Bacteriol., 175(22): 7356-7362 (1993)
Z21501	argS; lysA	Arginyl-tRNA synthetase; diaminiopimelate decarboxylase (partial)	Pisabarro, A. et al. "A cluster of three genes (dapA, orf2, and dapB) of <i>Brevibacterium lactofermentum</i> encodes dihydrodipicolinate reductase, and a third polypeptide of unknown function," J. Bacteriol., 175(9): 2743-2749 (1993)
Z21502	dapA; dapB	Dihydrodipicolinate synthase; dihydrodipicolinate reductase	Malumbres, M. et al. "Analysis and expression of the thrC gene of the encoded threonine synthase," Appl. Environ. Microbiol., 60(7): 2209-2219 (1994)
Z29563	thrC	Threonine synthase	Bachmann, B. et al. "DNA fragment coding for phosphoenolpyruvate
Z46753	16S rDNA	Gene for 16S ribosomal RNA	Oguiza, J. A. et al. "Multiple sigma factor genes in <i>Brevibacterium lactofermentum</i> : Characterization of sigA and sigB," J. Bacteriol., 178(2): 550-553 (1996)
A09073	ppg	Phosphoenol pyruvate carboxylase	Oguiza, J. A. et al. "The galE gene encoding the UDP-galactose 4-epimerase of <i>Brevibacterium lactofermentum</i> is coupled transcriptionally to the dmndR gene," Gene, 177: 103-107 (1996)
Z49822	sigA	SigA sigma factor	Oguiza, J. A. et al. "Multiple sigma factor genes in <i>Brevibacterium lactofermentum</i> : Characterization of sigA and sigB," J. Bacteriol., 178(2): 550-553 (1996)
Z49823	galE; dtxR	Catalytic activity UDP-galactose 4-epimerase; diphtheria toxin regulatory protein	Correia, A. et al. "Cloning and characterization of an IS-like element present in the genome of <i>Brevibacterium lactofermentum</i> ATCC 13869," Gene, 170(1): 91-94 (1996)
Z49824	orf1; sigB	?; SigB sigma factor	
Z66534		Transposase	

¹A sequence for this gene was published in the indicated reference. However, the sequence obtained by the inventors of the present application is significantly longer than the published version. It is believed that the published version relied on an incorrect start codon, and thus represents only a fragment of the actual coding region.

[0170]

TABLE 3

<i>Corynebacterium</i> and <i>Brevibacterium</i> Strains Which May be Used in the Practice of the Invention									
Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21054							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19350							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19351							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19352							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19353							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19354							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19355							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	19356							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21055							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21077							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21553							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	21580							
<i>Brevibacterium</i>	<i>ammoniagenes</i>	39101							
<i>Brevibacterium</i>	<i>butanicum</i>	21196							
<i>Brevibacterium</i>	<i>divaricatum</i>	21792	P928						
<i>Brevibacterium</i>	<i>flavum</i>	21474							
<i>Brevibacterium</i>	<i>flavum</i>	21129							
<i>Brevibacterium</i>	<i>flavum</i>	21518							
<i>Brevibacterium</i>	<i>flavum</i>			B11474					
<i>Brevibacterium</i>	<i>flavum</i>			B11472					
<i>Brevibacterium</i>	<i>flavum</i>	21127							
<i>Brevibacterium</i>	<i>flavum</i>	21128							
<i>Brevibacterium</i>	<i>flavum</i>	21427							
<i>Brevibacterium</i>	<i>flavum</i>	21475							
<i>Brevibacterium</i>	<i>flavum</i>	21517							
<i>Brevibacterium</i>	<i>flavum</i>	21528							
<i>Brevibacterium</i>	<i>flavum</i>	21529							
<i>Brevibacterium</i>	<i>flavum</i>			B11477					
<i>Brevibacterium</i>	<i>flavum</i>			B11478					
<i>Brevibacterium</i>	<i>flavum</i>	21127							
<i>Brevibacterium</i>	<i>flavum</i>			B11474					
<i>Brevibacterium</i>	<i>healii</i>	15527							
<i>Brevibacterium</i>	<i>ketoglutamicum</i>	21004							
<i>Brevibacterium</i>	<i>ketoglutamicum</i>	21089							
<i>Brevibacterium</i>	<i>ketosoreductum</i>	21914							
<i>Brevibacterium</i>	<i>lactofermentum</i>				70				
<i>Brevibacterium</i>	<i>lactofermentum</i>				74				
<i>Brevibacterium</i>	<i>lactofermentum</i>				77				
<i>Brevibacterium</i>	<i>lactofermentum</i>	21798							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21799							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21800							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21801							
<i>Brevibacterium</i>	<i>lactofermentum</i>			B11470					
<i>Brevibacterium</i>	<i>lactofermentum</i>			B11471					
<i>Brevibacterium</i>	<i>lactofermentum</i>	21086							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21420							
<i>Brevibacterium</i>	<i>lactofermentum</i>	21086							
<i>Brevibacterium</i>	<i>lactofermentum</i>	31269							
<i>Brevibacterium</i>	<i>linens</i>	9174							
<i>Brevibacterium</i>	<i>linens</i>	19391							
<i>Brevibacterium</i>	<i>linens</i>	8377							
<i>Brevibacterium</i>	<i>paraffinolyticum</i>					11160			
<i>Brevibacterium</i>	spec.						717.73		
<i>Brevibacterium</i>	spec.						717.73		
<i>Brevibacterium</i>	spec.	14604							
<i>Brevibacterium</i>	spec.	21860							
<i>Brevibacterium</i>	spec.	21864							
<i>Brevibacterium</i>	spec.	21865							
<i>Brevibacterium</i>	spec.	21866							
<i>Brevibacterium</i>	spec.	19240							
<i>Corynebacterium</i>	<i>acetoacidophilum</i>	21476							
<i>Corynebacterium</i>	<i>acetoacidophilum</i>	13870							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>			B11473					
<i>Corynebacterium</i>	<i>acetoglutamicum</i>			B11475					
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	15806							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	21491							
<i>Corynebacterium</i>	<i>acetoglutamicum</i>	31270							
<i>Corynebacterium</i>	<i>acetophilum</i>			B3671					
<i>Corynebacterium</i>	<i>ammoniagenes</i>	6872						2399	
<i>Corynebacterium</i>	<i>ammoniagenes</i>	15511							

TABLE 3-continued

<i>Corynebacterium</i> and <i>Brevibacterium</i> Strains Which May be Used in the Practice of the Invention									
Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Corynebacterium</i>	<i>fujikense</i>	21496							
<i>Corynebacterium</i>	<i>glutamicum</i>	14067							
<i>Corynebacterium</i>	<i>glutamicum</i>	39137							
<i>Corynebacterium</i>	<i>glutamicum</i>	21254							
<i>Corynebacterium</i>	<i>glutamicum</i>	21255							
<i>Corynebacterium</i>	<i>glutamicum</i>	31830							
<i>Corynebacterium</i>	<i>glutamicum</i>	13032							
<i>Corynebacterium</i>	<i>glutamicum</i>	14305							
<i>Corynebacterium</i>	<i>glutamicum</i>	15455							
<i>Corynebacterium</i>	<i>glutamicum</i>	13058							
<i>Corynebacterium</i>	<i>glutamicum</i>	13059							
<i>Corynebacterium</i>	<i>glutamicum</i>	13060							
<i>Corynebacterium</i>	<i>glutamicum</i>	21492							
<i>Corynebacterium</i>	<i>glutamicum</i>	21513							
<i>Corynebacterium</i>	<i>glutamicum</i>	21526							
<i>Corynebacterium</i>	<i>glutamicum</i>	21543							
<i>Corynebacterium</i>	<i>glutamicum</i>	13287							
<i>Corynebacterium</i>	<i>glutamicum</i>	21851							
<i>Corynebacterium</i>	<i>glutamicum</i>	21253							
<i>Corynebacterium</i>	<i>glutamicum</i>	21514							
<i>Corynebacterium</i>	<i>glutamicum</i>	21516							
<i>Corynebacterium</i>	<i>glutamicum</i>	21299							
<i>Corynebacterium</i>	<i>glutamicum</i>	21300							
<i>Corynebacterium</i>	<i>glutamicum</i>	39684							
<i>Corynebacterium</i>	<i>glutamicum</i>	21488							
<i>Corynebacterium</i>	<i>glutamicum</i>	21649							
<i>Corynebacterium</i>	<i>glutamicum</i>	21650							
<i>Corynebacterium</i>	<i>glutamicum</i>	19223							
<i>Corynebacterium</i>	<i>glutamicum</i>	13869							
<i>Corynebacterium</i>	<i>glutamicum</i>	21157							
<i>Corynebacterium</i>	<i>glutamicum</i>	21158							
<i>Corynebacterium</i>	<i>glutamicum</i>	21159							
<i>Corynebacterium</i>	<i>glutamicum</i>	21355							
<i>Corynebacterium</i>	<i>glutamicum</i>	31808							
<i>Corynebacterium</i>	<i>glutamicum</i>	21674							
<i>Corynebacterium</i>	<i>glutamicum</i>	21562							
<i>Corynebacterium</i>	<i>glutamicum</i>	21563							
<i>Corynebacterium</i>	<i>glutamicum</i>	21564							
<i>Corynebacterium</i>	<i>glutamicum</i>	21565							
<i>Corynebacterium</i>	<i>glutamicum</i>	21566							
<i>Corynebacterium</i>	<i>glutamicum</i>	21567							
<i>Corynebacterium</i>	<i>glutamicum</i>	21568							
<i>Corynebacterium</i>	<i>glutamicum</i>	21569							
<i>Corynebacterium</i>	<i>glutamicum</i>	21570							
<i>Corynebacterium</i>	<i>glutamicum</i>	21571							
<i>Corynebacterium</i>	<i>glutamicum</i>	21572							
<i>Corynebacterium</i>	<i>glutamicum</i>	21573							
<i>Corynebacterium</i>	<i>glutamicum</i>	21579							
<i>Corynebacterium</i>	<i>glutamicum</i>	19049							
<i>Corynebacterium</i>	<i>glutamicum</i>	19050							
<i>Corynebacterium</i>	<i>glutamicum</i>	19051							
<i>Corynebacterium</i>	<i>glutamicum</i>	19052							
<i>Corynebacterium</i>	<i>glutamicum</i>	19053							
<i>Corynebacterium</i>	<i>glutamicum</i>	19054							
<i>Corynebacterium</i>	<i>glutamicum</i>	19055							
<i>Corynebacterium</i>	<i>glutamicum</i>	19056							
<i>Corynebacterium</i>	<i>glutamicum</i>	19057							
<i>Corynebacterium</i>	<i>glutamicum</i>	19058							
<i>Corynebacterium</i>	<i>glutamicum</i>	19059							
<i>Corynebacterium</i>	<i>glutamicum</i>	19060							
<i>Corynebacterium</i>	<i>glutamicum</i>	19185							
<i>Corynebacterium</i>	<i>glutamicum</i>	13286							
<i>Corynebacterium</i>	<i>glutamicum</i>	21515							
<i>Corynebacterium</i>	<i>glutamicum</i>	21527							
<i>Corynebacterium</i>	<i>glutamicum</i>	21544							
<i>Corynebacterium</i>	<i>glutamicum</i>	21492							
<i>Corynebacterium</i>	<i>glutamicum</i>			B8183					
<i>Corynebacterium</i>	<i>glutamicum</i>			B8182					
<i>Corynebacterium</i>	<i>glutamicum</i>			B12416					
<i>Corynebacterium</i>	<i>glutamicum</i>			B12417					
<i>Corynebacterium</i>	<i>glutamicum</i>			B12418					
<i>Corynebacterium</i>	<i>glutamicum</i>			B11476					

TABLE 3-continued

<i>Corynebacterium</i> and <i>Brevibacterium</i> Strains Which May be Used in the Practice of the Invention									
Genus	species	ATCC	FERM	NRRL	CECT	NCIMB	CBS	NCTC	DSMZ
<i>Corynebacterium</i>	<i>glutamicum</i>	21608							
<i>Corynebacterium</i>	<i>lilium</i>		P973						
<i>Corynebacterium</i>	<i>nitrilophilus</i>	21419				11594			
<i>Corynebacterium</i>	spec.		P4445						
<i>Corynebacterium</i>	spec.		P4446						
<i>Corynebacterium</i>	spec.	31088							
<i>Corynebacterium</i>	spec.	31089							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	31090							
<i>Corynebacterium</i>	spec.	15954							20145
<i>Corynebacterium</i>	spec.	21857							
<i>Corynebacterium</i>	spec.	21862							
<i>Corynebacterium</i>	spec.	21863							

ATCC: American Type Culture Collection, Rockville, MD, USA

FERM: Fermentation Research Institute, Chiba, Japan

NRRL: ARS Culture Collection, Northern Regional Research Laboratory, Peoria, IL, USA

CECT: Coleccion Espanola de Cultivos Tipo, Valencia, Spain

NCIMB: National Collection of Industrial and Marine Bacteria Ltd., Aberdeen, UK

CBS: Centraalbureau voor Schimmelcultures, Baam, NL

NCTC: National Collection of Type Cultures, London, UK

DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany

For reference see Sugawara, H. et al. (1993) World directory of collections of cultures of microorganisms: Bacteria, fungi and yeasts (4th edn), World federation for culture collections world data center on microorganisms, Saimata, Japan.

[0171]

TABLE 4

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxax00051	1527	GB_HTG3: AC009685	210031	AC009685	<i>Homo sapiens</i> chromosome 15 clone 91_E_13 map 15, *** SEQUENCING IN PROGRESS ***, 27 unordered pieces.	<i>Homo sapiens</i>	34,247	29-Sep-99	
		GB_HTG3: AC009685	210031	AC009685	<i>Homo sapiens</i> chromosome 15 clone 91_E_13 map 15, *** SEQUENCING IN PROGRESS ***, 27 unordered pieces.	<i>Homo sapiens</i>	34,247	29-Sep-99	
		GB_HTG7: AC009511	271896	AC009511	<i>Homo sapiens</i> clone RP11-860B13, *** SEQUENCING IN PROGRESS ***, 59 unordered pieces.	<i>Homo sapiens</i>	35,033	09-DEC-1999	
	876	GB_BA1: D50453	146191	D50453	<i>Bacillus subtilis</i> DNA for 25-36 degree region containing the amyE-srfA region, complete cds.	<i>Bacillus subtilis</i>	54,452	10-Feb-99	
rxax00092		GB_BA1: SCI51	40745	AL109848	<i>Streptomyces coelicolor</i> cosmid I51.	<i>Streptomyces coelicolor</i> A3(2)	36,806	16-Aug-99	
		GB_BA1: ECOUW93	338534	U14003	<i>Escherichia coli</i> K-12 chromosomal region from 92.8 to 00.1 minutes.	<i>Escherichia coli</i>	38,642	17-Apr-96	
	789	GB_BA1: SCH35	45396	AL078610	<i>Streptomyces coelicolor</i> cosmid H35.	<i>Streptomyces coelicolor</i>	49,934	4-Jun-99	
		GB_HTG3: AC011498_0	312343	AC011498	<i>Homo sapiens</i> chromosome 19 clone CIT978SKB_50L17, *** SEQUENCING IN PROGRESS ***, 190 unordered pieces.	<i>Homo sapiens</i>	37,117	13-Dec-99	
rxax00104		GB_HTG3: AC011498_0	312343	AC011498	<i>Homo sapiens</i> chromosome 19 clone CIT978SKB_50L17, *** SEQUENCING IN PROGRESS ***, 190 unordered pieces.	<i>Homo sapiens</i>	37,117	13-Dec-99	
	879	GB_BA1: MTCY270	37586	Z95388	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162.	<i>Mycobacterium tuberculosis</i>	36,732	10-Feb-99	
		GB_PL2: T24M8	68251	AF077409	<i>Arabidopsis thaliana</i> BAC T24M8.	<i>Arabidopsis thaliana</i>	37,150	3-Aug-98	
		GB_BA1: MTCY270	37586	Z95388	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 96/162.	<i>Mycobacterium tuberculosis</i>	42,874	10-Feb-99	
rxax00113	5745	GB_BA1: MAFASGEN	10520	X87822	<i>B. ammoniagenes</i> FAS gene.	<i>Corynebacterium ammoniagenes</i>	68,381	03-OCT-1996	
		GB_BA1: BAFASAA	10549	X64795	<i>B. ammoniagenes</i> FAS gene.	<i>Corynebacterium ammoniagenes</i>	57,259	14-OCT-1997	
		GB_BA1: MTCY159	33818	Z83863	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 111/162.	<i>Corynebacterium tuberculosis</i>	39,870	17-Jun-98	
	1812	GB_HTG2: HSI1153D9	118360	AL109806	<i>Homo sapiens</i> chromosome 20 clone RP5-1153D9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	35,714	03-DEC-1999	
rxax00164		GB_HTG2: HSI1153D9	118360	AL109806	<i>Homo sapiens</i> chromosome 20 clone RP5-1153D9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	35,714	03-DEC-1999	
		GB_HTG2: HSI1153D9	118360	AL109806	<i>Homo sapiens</i> chromosome 20 clone RP5-1153D9, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	35,334	03-DEC-1999	
	1695	GB_BA1: CGPUTP	3791	Y09163	<i>C. glutamicum</i> putP gene.	<i>Corynebacterium glutamicum</i>	100,000	8-Sep-97	
		GB_BA2: U32814	10393	U32814	<i>Haemophilus influenzae</i> Rd section 129 of 163 of the complete genome.	<i>Haemophilus influenzae</i> Rd	36,347	29-MAY-1998	
rxax00186		GB_BA1: CGPUTP	3791	Y09163	<i>C. glutamicum</i> putP gene.	<i>Corynebacterium glutamicum</i>	37,454	8-Sep-97	
	870	GB_PR3: AC004843	136655	AC004843	<i>Homo sapiens</i> PAC clone DJ0612F12 from 7p12-p14, complete sequence.	<i>Homo sapiens</i>	37,315	5-Nov-98	
		GB_HTG2: HS745114	133309	AL033532	<i>Homo sapiens</i> chromosome 1 clone RP4-745114 map q23.1-24.3, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	38,129	03-DEC-1999	
		GB_HTG2: HS745114	133309	AL033532	<i>Homo sapiens</i> chromosome 1 clone RP4-745114 map q23.1-24.3, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	38,129	03-DEC-1999	

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rx00187	474	GB_GSS10: AQ184082	506	AQ184082	HS_3216_A1_G08_T7 CIT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 3216 Col = 15 Row = M, genomic survey sequence. <i>Drosophila melanogaster</i> genome survey sequence T7 end of BAC # BACR18L01 of RSC198 library from <i>Drosophila melanogaster</i> (fruit fly), genomic survey sequence.	<i>Homo sapiens</i>	37,297	1-Nov-98	
		GB_GSS1: CNS008ZZ	1101	AL052951	HS_3216_A1_G08_T7 CIT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 3216 Col = 15 Row = M, genomic survey sequence. Human DNA sequence from clone 824F16 on chromosome 20, complete sequence. <i>R. capsulatus</i> DNA for secA gene.	<i>Drosophila melanogaster</i>	34,120	3-Jun-99	
rx00201	292	GB_PR3: HSJ824F16	139330	AL050325	HS_3216_A1_G08_T7 CIT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 3216 Col = 15 Row = M, genomic survey sequence.	<i>Homo sapiens</i>	39,655	1-Nov-98	
		GB_BA1: RCSECA	2724	X89411	AV122904 <i>Mus musculus</i> C57BL/6J 10-day embryo <i>Mus musculus</i> cDNA clone 2610529H07, mRNA sequence.	<i>Homo sapiens</i>	34,520	23-Nov-99	
		GB_EST34: AV122904	242	AV122904	ab40c08.r1 Stratagene HeLa cell s3 937216 <i>Homo sapiens</i> cDNA clone IMAGE: 843278 5', mRNA sequence.	<i>Rhodobacter capsulatus</i>	38,163	6-Jan-96	
rx00228	714	GB_EST15: AA486042	515	AA486042	IMAGE: 843278 5', mRNA sequence.	<i>Mus musculus</i>	38,889	1-Jul-99	
		GB_EST15: AA486042	515	AA486042	IMAGE: 843278 5', mRNA sequence.	<i>Homo sapiens</i>	37,500	06-MAR-1998	
rx00243	1140	GB_PR2: CNS01DS5	101584	AL121655	BAC sequence from the SPC4 candidate region at 2p21-2p22, complete sequence. <i>Homo sapiens</i> clone C17978SKB_65D22, *** SEQUENCING IN PROGRESS ***	<i>Homo sapiens</i>	38,816	06-MAR-1998	
		GB_HTG3: AC011408	79332	AC011408	<i>Homo sapiens</i> clone C17978SKB_65D22, *** SEQUENCING IN PROGRESS ***	<i>Homo sapiens</i>	37,001	29-Sep-99	
		GB_HTG3: AC011408	79332	AC011408	<i>Homo sapiens</i> clone C17978SKB_65D22, *** SEQUENCING IN PROGRESS ***	<i>Homo sapiens</i>	38,040	06-OCT-1999	
rx00259	2325	GB_HTG1: CEY62E10	254217	AL031580	<i>Caenorhabditis elegans</i> chromosome IV clone Y62E10, *** SEQUENCING IN PROGRESS ***	<i>Caenorhabditis elegans</i>	36,776	6-Sep-99	
		GB_HTG1: CEY62E10	254217	AL031580	<i>Caenorhabditis elegans</i> chromosome IV clone Y62E10, *** SEQUENCING IN PROGRESS ***	<i>Caenorhabditis elegans</i>	36,776	6-Sep-99	
		GB_PL2: YSCCHROMI	41988	L22015	<i>Saccharomyces cerevisiae</i> chromosome I centromere and right arm sequence.	<i>Saccharomyces cerevisiae</i>	39,260	05-MAR-1998	
rx00269	912	GB_HTG4: AC009974	219565	AC009974	<i>Homo sapiens</i> chromosome unknown clone NH0459119, WORKING DRAFT SEQUENCE, in unordered pieces.	<i>Homo sapiens</i>	37,358	29-OCT-1999	
		GB_HTG4: AC009974	219565	AC009974	<i>Homo sapiens</i> chromosome unknown clone NH0459119, WORKING DRAFT SEQUENCE, in unordered pieces.	<i>Homo sapiens</i>	37,358	29-OCT-1999	
rx00281	766	GB_BA1: AB017508	32050	AB017508	<i>Bacillus halodurans</i> C-125 genomic DNA, 32 kb fragment, complete cds.	<i>Bacillus halodurans</i>	44,622	14-Apr-99	
		GB_BA1: SCF8	24700	AL035654	<i>Streptomyces coelicolor</i> cosmid E8.	<i>Streptomyces coelicolor</i>	36,328	11-MAR-1999	
		GB_BA1: SCU51332	3216	U51332	<i>Streptomyces coelicolor</i> histidine kinase homolog (absA1) and response regulator homolog (absA2) genes, complete cds.	<i>Streptomyces coelicolor</i>	39,089	14-Sep-96	
		GB_HTG4: AC011122	187123	AC011122	<i>Homo sapiens</i> chromosome 8 clone 23_D_19 map 8, *** SEQUENCING IN PROGRESS ***	<i>Homo sapiens</i>	38,658	14-OCT-1999	
rx00298	1968	GB_BA1: CGECTP	2719	AJ001436	<i>Corynebacterium glutamicum</i> ctp gene.	<i>Corynebacterium glutamicum</i>	100,000	20-Nov-98	
		GB_BA1: CGECTP	2719	AJ001436	<i>Corynebacterium glutamicum</i> ctp gene.	<i>Corynebacterium glutamicum</i>	100,000	20-Nov-98	
		GB_EST24: AL234006	432	AL234006	EST230694 Normalized rat lung, Beito Soares <i>Rattus</i> sp. cDNA clone RLUCU01 3' end, mRNA sequence.	<i>Rattus</i> sp.	46,552	31-Jan-99	
rx00346	813	GB_BA1: SC2E9	20850	AL021530	<i>Streptomyces coelicolor</i> cosmid 2E9.	<i>Streptomyces coelicolor</i>	43,267	28-Jan-98	
		GB_BA1: SC9B1	24800	AL049727	<i>Streptomyces coelicolor</i> cosmid 9B1.	<i>Streptomyces coelicolor</i>	44,613	27-Apr-99	
		GB_BA1: ECU70214	123171	U70214	<i>Escherichia coli</i> chromosome minutes 4-6.	<i>Escherichia coli</i>	39,490	21-Sep-96	

TABLE 4-continued

ALIGNMENT RESULTS					% homology		
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	Date of Deposit
rx00368	1698	GB_BA2: AF065159	35209	AF065159	<i>Bradyrhizobium japonicum</i> putative arylsulfatase (arsA), putative soluble lytic transglycosylase precursor (sltA), dihydripyridine synthase (dapA), MscL (mscL), SmpB (smpB), BcpB (bcpB), RnpO (rnpO), RelA/SpoT homolog (relA), PdxJ (pdxJ), and acyl carrier protein synthase AcpS (acpS) genes, complete cds; prokaryotic type I signal peptidase SipF (sipF) gene, sipF-sipS allele, complete cds; RNase III (rnc) gene, complete cds; GTP-binding protein Era (era) gene, partial cds; and unknown genes.	<i>Bradyrhizobium japonicum</i>	40,409 27-OCT-1999
rx00369	817	GB_BA1: AEOCHIT1	6861	D63139	<i>Aeromonas</i> sp. gene for chitinase, complete and partial cds.	<i>Aeromonas</i> sp. 10S-24	38,577 13-Feb-99
		GB_EST4: D62996	314	D62996	HUM347G01B Clontech human aorta polyA+ mRNA (#6572) <i>Homo sapiens</i> cDNA clone GEN-347G01 5', mRNA sequence.	<i>Homo sapiens</i>	41,613 29-Aug-95
		GB_BA1: YP102KB	119443	AL031866	<i>Yersinia pestis</i> 102 kbases unstable region: from 1 to 119443.	<i>Yersinia pestis</i>	35,396 4-Jan-99
rx00410	789	GB_GSS8: AQ012142	501	AQ012142	8750H1A037010398 Cosmid library of chromosome II <i>Rhodobacter sphaeroides</i> genomic clone 8750H1A037010398, genomic survey sequence.	<i>Rhodobacter sphaeroides</i>	54,800 4-Jun-98
		GB_HITG2: AC005081	180096	AC005081	<i>Homo sapiens</i> clone RG270D13, *** SEQUENCING IN PROGRESS ***, 18 unordered pieces.	<i>Homo sapiens</i>	45,786 12-Jun-98
		GB_BA1: ATPLOCC	8870	Z30328	<i>A. tumefaciens</i> Ti plasmid pTiAch5 genes for OccR, OccQ, OccM, OccP, OccT, OoxB, OoxA and ornithine cyclodeaminase.	<i>Agrobacterium tumefaciens</i>	46,490 10-OCT-1994
rx00419	882	GB_BA2: U67591	9829	U67591	<i>Methanococcus jannaschii</i> section 133 of 150 of the complete genome.	<i>Methanococcus jannaschii</i>	45,677 28-Jan-98
		GB_BA1: TIPOCCQMPJ	4350	M80607	Plasmid pTiA6 (from <i>Agrobacterium tumefaciens</i>) periplasmic-type octopine pemecase (occR, occQ, occM, occP, and occL) and lysR-type regulatory protein (occR) genes, complete cds.	Plasmid pTiA6	46,490 24-Apr-96
		GB_BA2: MSU46844	16951	U46844	<i>Mycobacterium smegmatis</i> catalase-peroxidase (katG), putative arabinosyl transferase (embC, embA, embB), genes complete cds and putative propionyl-coA carboxylase beta chain (pccB) genes, partial cds.	<i>Mycobacterium smegmatis</i>	57,029 12-MAY-1997
rx00432	1608	GB_EST28: AI513245	471	AI513245	GH13311.3prime GH <i>Drosophila melanogaster</i> head pOT2 <i>Drosophila melanogaster</i> cDNA clone GH13311 3prime, mRNA sequence.	<i>Drosophila melanogaster</i>	37,696 16-MAR-1999
		GB_HITG4: AC010066	187240	AC010066	<i>Drosophila melanogaster</i> chromosome 3L/72A4 clone RPC198-2501, *** SEQUENCING IN PROGRESS ***, 70 unordered pieces.	<i>Drosophila melanogaster</i>	39,607 16-OCT-1999
		GB_BA1: BSUB0015	218410	Z99118	<i>Bacillus subtilis</i> complete genome (section 15 of 21): from 2795131 to 3013540.	<i>Bacillus subtilis</i>	49,810 26-Nov-97
rx00449	1704	GB_PL1: CAC35A5	42565	AL033396	<i>C. albicans</i> cosmid Ca35A5.	<i>Candida albicans</i>	35,041 5-Nov-98
		GB_EST13: AA336266	378	AA336266	EST40981 Endometrial tumor <i>Homo sapiens</i> cDNA 5' end, mRNA sequence.	<i>Homo sapiens</i>	39,733 21-Apr-97
		GB_HITG2: AC008199	124050	AC008199	<i>Drosophila melanogaster</i> chromosome 3 clone BACR01K08 (D756) RPCL-98 01.K.8 map 94D-94D strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 83 unordered pieces.	<i>Drosophila melanogaster</i>	38,392 2-Aug-99
rx00456	1500	GB_HITG2: AC008199	124050	AC008199	<i>Drosophila melanogaster</i> chromosome 3 clone BACR01K08 (D756) RPCL-98 01.K.8 map 94D-94D strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 83 unordered pieces.	<i>Drosophila melanogaster</i>	38,392 2-Aug-99
		GB_RO: RATLNKP2	177	M22337	Rat link protein gene, exon 2.	<i>Rattus</i> sp.	40,678 27-Apr-93
		GB_GSS1: FR0030597	476	AL026966	<i>Fugu rubripes</i> GSS sequence, clone 091C22aF9, genomic survey sequence.	<i>Fugu rubripes</i>	47,407 25-Jun-98
rx00456	1500	GB_GSS5: AQ786587	556	AQ786587	HS_3086_B1_H05_MR CIT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 3086 Col = 9 Row = P, genomic survey sequence.	<i>Homo sapiens</i>	38,406 3-Aug-99
		GB_GSS14: AQ526586	434	AQ526586	HS_5198_B1_B03_SP6E RPCL-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 774 Col = 5 Row = D, genomic survey sequence.	<i>Homo sapiens</i>	36,951 11-MAY-1999

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	homology (GAP) % Date of Deposit
rxa00477	1767	GB_EST17: AA610489	407	AA610489	np93c05.s1 NCL_CGAP_Thy1 <i>Homo sapiens</i> cDNA clone IMAGE: 1133888 similar to gb: M11353 HISTONE H3.3 (HUMAN); mRNA sequence.	41,791 09-DEC-1997
		GB_PRI: HSH33C4	1015	X05857	Human H3.3 gene exon 4.	38,182 24-Jan-96
		GB_EST30: A1637667	579	A1637667	tt10g11.x1 NCL_CGAP_GC6 <i>Homo sapiens</i> cDNA clone IMAGE: 2240420 3'; mRNA sequence.	35,417 27-Apr-99
rxa00478	954	GB_HTG3: AC008708	83932	AC008708	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_78F1, *** SEQUENCING IN PROGRESS ***; 12 unordered pieces.	38,769 3-Aug-99
		GB_HTG3: AC008708	83932	AC008708	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_78F1, *** SEQUENCING IN PROGRESS ***; 12 unordered pieces.	38,769 3-Aug-99
		GB_HTG3: AC008708	83932	AC008708	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_78F1, *** SEQUENCING IN PROGRESS ***; 12 unordered pieces.	36,797 3-Aug-99
rxa00480	1239	GB_HTG1: HSI575L21	94715	AL096841	<i>Homo sapiens</i> chromosome 1 clone RP4-575L21, *** SEQUENCING IN PROGRESS ***; In unordered pieces.	38,138 23-Nov-99
		GB_HTG1: HSI575L21	94715	AL096841	<i>Homo sapiens</i> chromosome 1 clone RP4-575L21, *** SEQUENCING IN PROGRESS ***; In unordered pieces.	38,138 23-Nov-99
		GB_RO: AC005960	158414	AC005960	<i>Mus musculus</i> chromosome 17 BAC clb20h22 from the MHC region, complete sequence.	38,712 01-DEC-1998
rxa00524	433	GB_BA1: SCI51	40745	AL109848	<i>Streptomyces coelicolor</i> cosmid I51.	40,284 16-Aug-99
		GB_BA2: AF082879	3434	AF082879	<i>Yersinia enterocolitica</i> ABC transporter enterochelin/enterobactin gene cluster, complete sequence.	55,634 20-OCT-1999
		GB_BA1: BSP132617	5192	AJ132617	<i>Burkholderia</i> sp. P-transporter operon and flanking genes.	40,793 13-Jul-99
rxa00526	813	GB_BA1: BSUB0008	208230	Z99111	<i>Bacillus subtilis</i> complete genome (section 8 of 21); from 1394791 to 1603020.	54,534 26-Nov-97
		GB_BA2: AF012285	46864	AF012285	<i>Bacillus subtilis</i> mobA-nprE gene region.	54,534 1-Jul-98
		GB_BA1: D90725	13796	D90725	<i>Escherichia coli</i> genomic DNA. (19.7-20.0 min).	51,481 7-Feb-99
rxa00559	1140	GB_BA2: CAU77910	3385	U77910	<i>Corynebacterium ammoniagenes</i> sequence upstream of the 5-phosphoribosyl-1-pyrophosphate amidotransferase (purF) gene.	39,007 1-Jan-98
		GB_EST4: H34952	382	H34952	EST108261 Rat PC-12 cells, untreated <i>Rattus</i> sp. cDNA clone RPCK07 similar to NADH-ubiquinone oxidoreductase complex I 23 kDa precursor (iron-sulfur protein), mRNA sequence.	39,267 2-Apr-98
		GB_BA2: AE000963	22014	AE000963	<i>Archaeoglobus fulgidus</i> section 144 of 172 of the complete genome.	38,338 15-DEC-1997
rxa00570	852	GB_GSSI2: AQ422451	563	AQ422451	RPCL11-185C3.TV RPCL11 <i>Homo sapiens</i> genomic clone RPCL11-185C3, genomic survey sequence.	38,767 23-MAR-1999
		GB_EST28: AI504741	568	AI504741	v116c01.x1 Striatagene mouse Tcell 937311 <i>Mus musculus</i> cDNA clone IMAGE: 972384 3' similar to gb: Z14044 <i>M. musculus</i> mRNA for valosin-containing protein (MOUSE); mRNA sequence.	37,900 11-MAR-1999
		GB_EST18: AA712043	68	AA712043	v129f10.r1 Barstead mouse myotubes MPLRB5 <i>Mus musculus</i> cDNA clone IMAGE: 1182091 5' similar to gb: L05093 60S RIBOSOMAL PROTEIN L18A (HUMAN); mRNA sequence.	42,647 24-DEC-1997
rxa00571	1280	GB_BA1: MTCY78	33818	Z77165	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 145/162.	38,468 17-Jun-98
		GB_PR3: AC005788	36224	AC005788	<i>Homo sapiens</i> chromosome 19, cosmid R26652, complete sequence.	36,911 06-OCT-1998
		GB_PR3: AC005338	34541	AC005338	<i>Homo sapiens</i> chromosome 19, cosmid R31646, complete sequence.	36,911 30-Jul-98
rxa00590	1288	GB_HTG6: AC010932	203273	AC010932	<i>Homo sapiens</i> chromosome 15 clone RP11-296E22 map 15, *** SEQUENCING IN PROGRESS ***; 36 unordered pieces.	37,242 30-Nov-99

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% (GAP)	homology	Date of Deposit
rxa00591	1476	GB_HTG6: AC010932	203273	AC010932	<i>Homo sapiens</i> chromosome 15 clone RP11-296E22 map 15, *** SEQUENCING IN PROGRESS ***, 36 unordered pieces.	<i>Homo sapiens</i>	36,485	30-Nov-99	
		GB_BA1: MSGB26CS	37040	L78816	<i>Mycobacterium leprae</i> cosmid B26 DNA sequence.	<i>Mycobacterium leprae</i>	39,272	15-Jun-96	
		GB_IN1: CEK09E9	30098	Z79602	<i>Caenorhabditis elegans</i> cosmid K09E9, complete sequence.	<i>Caenorhabditis elegans</i>	34,092	2-Sep-99	
		GB_PR4: AF135802	4965	AF135802	<i>Homo sapiens</i> thyroid hormone receptor-associated protein complex component TRAP170 mRNA, complete cds.	<i>Homo sapiens</i>	36,310	9-Apr-99	
rxa00596	576	GB_PR4: AF104256	4365	AF104256	<i>Homo sapiens</i> transcriptional co-activator CRSP150 (CRSP150) mRNA, complete cds.	<i>Homo sapiens</i>	36,617	4-Feb-99	
		GB_PR3: AC004659	129577	AC004659	<i>Homo sapiens</i> chromosome 19, CIT-HSP-87m17 BAC clone, complete sequence.	<i>Homo sapiens</i>	34,321	02-MAY-1998	
		GB_PR3: AC004659	129577	AC004659	<i>Homo sapiens</i> chromosome 19, CIT-HSP-87m17 BAC clone, complete sequence.	<i>Homo sapiens</i>	35,739	02-MAY-1998	
		GB_PR1: HUMCBP2	2047	D83174	Human mRNA for collagen binding protein 2, complete cds.	<i>Homo sapiens</i>	40,404	6-Feb-99	
rxa00607	504	GB_BA1: MTV010	3400	AL021186	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 119/162.	<i>Mycobacterium tuberculosis</i>	40,862	23-Jun-99	
		GB_BA1: MTV010	3400	AL021186	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 119/162.	<i>Mycobacterium tuberculosis</i>	38,833	23-Jun-99	
rxa00623	1461	GB_BA1: MTCV428	26914	Z81451	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 107/162.	<i>Mycobacterium tuberculosis</i>	60,552	17-Jun-98	
		GB_BA1: RSPNGR234	34010	Z68203	<i>Rhizobium</i> sp. plasmid NGR234a DNA.	<i>Rhizobium</i> sp.	51,992	8-Aug-96	
rxa00681	1269	GB_BA2: AE000101	10057	AE000101	<i>Rhizobium</i> sp. NGR234 plasmid pNGR234a, section 38 of 46 of the complete plasmid sequence.	<i>Rhizobium</i> sp. NGR234	51,992	12-DEC-1997	
		GB_HTG5: AC008338	136685	AC008338	<i>Drosophila melanogaster</i> chromosome X clone BACR30104 (D908) RPCL98 30J.4 map 19C-19E strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 93 unordered pieces.	<i>Drosophila melanogaster</i>	35,341	15-Nov-99	
rxa00733	1008	GB_HTG4: AC009766	170502	AC009766	<i>Homo sapiens</i> chromosome 11 clone 404_A_03 map 11, *** SEQUENCING IN PROGRESS ***, 27 unordered pieces.	<i>Homo sapiens</i>	37,984	19-OCT-1 999	
		GB_HTG4: AC009766	170502	AC009766	<i>Homo sapiens</i> chromosome 11 clone 404_A_03 map 11, *** SEQUENCING IN PROGRESS ***, 27 unordered pieces.	<i>Homo sapiens</i>	37,984	19-OCT-1999	
		GB_EST30: AU054038	245	AU054038	AU054038 <i>Dictyostelium discoideum</i> SL (<i>H. Urushihara</i>) <i>Dictyostelium discoideum</i> cDNA clone SLK472, mRNA sequence.	<i>Dictyostelium discoideum</i>	43,265	28-Apr-99	
		GB_EST30: AU054038	245	AU054038	AU054038 <i>Dictyostelium discoideum</i> SL (<i>H. Urushihara</i>) <i>Dictyostelium discoideum</i> cDNA clone SLK472, mRNA sequence.	<i>Dictyostelium discoideum</i>	43,265	28-Apr-99	
rxa00735	692	GB_BA1: MTCV50	36030	Z77137	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 55/162.	<i>Mycobacterium tuberculosis</i>	36,819	17-Jun-98	
		GB_BA1: D90904	150894	D90904	<i>Synechocystis</i> sp. PCC6803 complete genome, 6/27, 630555-781448.	<i>Synechocystis</i> sp.	52,585	7-Feb-99	
rxa00796	298	GB_BA1: D90904	150894	D90904	<i>Synechocystis</i> sp. PCC6803 complete genome, 6/27, 630555-781448.	<i>Synechocystis</i> sp.	39,699	7-Feb-99	
		GB_GSS14: AQ579838	651	AQ579838	T135342b shotgun sub-library of BAC clone 31P06 Medicago truncatula genomic clone 31-P-06-C-054, genomic survey sequence.	<i>Medicago truncatula</i>	37,153	27-Sep-99	
rxa00801	756	GB_PR4: AC007625	174701	AC007625	Genomic sequence of <i>Homo sapiens</i> clone 2314F2 from chromosome 18, complete sequence.	<i>Homo sapiens</i>	38,014	30-Jun-99	
		GB_EST14: AA427576	580	AA427576	zw54b04.s1 Soares_total_fetus_Nb2HF8_9w <i>Homo sapiens</i> cDNA clone IMAGE: 773839 3' similar to gb: M86852 PEROXISOME ASSEMBLY FACTOR-1 (HUMAN); mRNA sequence.	<i>Homo sapiens</i>	42,731	16-OCT-1997	
		GB_BA1: MTV022	13025	AL021925	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 100/162.	<i>Mycobacterium tuberculosis</i>	59,350	17-Jun-98	

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxa00802	837	GB_RO: AC002109	160048	AC002109	Genomic sequence from Mouse 9, complete sequence.	<i>Mus musculus</i>	39,398	9-Sep-97	
		GB_BA1: MTU022	13025	AL021925	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 100/162.	<i>Mycobacterium tuberculosis</i>	36,842	17-Jun-98	
		GB_GSS14: AQ563349	642	AQ563349	HS_5335_B2_A09_T7A RPCL11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 911 Col = 18 Row = B; genomic survey sequence.	<i>Homo sapiens</i>	37,649	29-MAY-1999	
		GB_BA1: DIHCLPBA	2441	M32229	<i>B. nodosus</i> clpB gene encoding a regulatory subunit of ATP-dependent protease.	<i>Dichelobacter nodosus</i>	41,140	26-Apr-93	
rxa00819	1452	GB_GSS3: B61538	698	B61538	T17M17TR TAMU <i>Arabidopsis thaliana</i> genomic clone T17M17, genomic survey sequence.	<i>Arabidopsis thaliana</i>	36,946	21-Nov-97	
		GB_HTG3: AC008691_1	110000	AC008691	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_63A22, *** SEQUENCING IN PROGRESS ***, 253 unordered pieces.	<i>Homo sapiens</i>	38,270	3-Aug-99	
		GB_HTG3: AC008691_1	110000	AC008691	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_63A22, *** SEQUENCING IN PROGRESS ***, 253 unordered pieces.	<i>Homo sapiens</i>	38,270	3-Aug-99	
		GB_HTG3: AC009127	186591	AC009127	<i>Homo sapiens</i> chromosome 16 clone RPCL11_498D10, *** SEQUENCING IN PROGRESS ***, 49 unordered pieces.	<i>Homo sapiens</i>	38,947	3-Aug-99	
rxa00821	966	GB_HTG1: HS32B1	271488	AL023693	<i>Homo sapiens</i> chromosome 6 clone RP1-32B1, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	36,565	23-Nov-99	
		GB_HTG1: HS32B1	271488	AL023693	<i>Homo sapiens</i> chromosome 6 clone RP1-32B1, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	36,565	23-Nov-99	
		GB_PR3: AC004919	75547	AC004919	<i>Homo sapiens</i> PAC clone D10895B23 from UL, complete sequence.	<i>Homo sapiens</i>	34,346	19-Sep-98	
		GB_EST6: W06539	300	W06539	T2367 MVA14 bloodstream form of serodeme WRA1at1.1 Trypanosoma brucei rhodesiense cDNA 5', mRNA sequence.	<i>Trypanosoma brucei rhodesiense</i>	40,000	12-Aug-96	
rxa00842	1323	GB_EST18: AA710415	181745	AC008179	<i>Homo sapiens</i> clone NH0576F01, complete sequence.	<i>Homo sapiens</i>	35,903	28-Sep-99	
		GB_PR2: AC002379	533	AA710415	IMAGE: 1166823 5', mRNA sequence.	<i>Mus musculus</i>	41,562	24-DEC-1997	
		GB_PR2: AC002379	118595	AC002379	Human BAC clone GSI65104 from 7q21, complete sequence.	<i>Homo sapiens</i>	36,321	23-Jul-97	
		GB_IN1: CEF02D8	118595	AC002379	Human BAC clone GSI65104 from 7q21, complete sequence.	<i>Homo sapiens</i>	37,284	23-Jul-97	
rxa00847	1572	GB_OV: XELRDS38A	31624	Z78411	<i>Caenorhabditis elegans</i> cosmid F02D8, complete sequence.	<i>Caenorhabditis elegans</i>	38,163	23-Nov-98	
		GB_HTG4: AC007920	1209	L79915	<i>Xenopus laevis</i> rds/peripherin (rds38) mRNA, complete cds.	<i>Xenopus laevis</i>	36,044	30-Jul-97	
		GB_HTG4: AC007920	234529	AC007920	<i>Homo sapiens</i> chromosome 3q27 clone RPCL11-208N14, *** SEQUENCING IN PROGRESS ***, 51 unordered pieces.	<i>Homo sapiens</i>	33,742	21-OCT-1999	
		GB_HTG4: AC007920	234529	AC007920	<i>Homo sapiens</i> chromosome 3q27 clone RPCL11-208N14, *** SEQUENCING IN PROGRESS ***, 51 unordered pieces.	<i>Homo sapiens</i>	33,742	21-OCT-1999	
rxa00851	732	GB_HTG2: AC004064	185000	AC004064	<i>Homo sapiens</i> chromosome 4, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.	<i>Homo sapiens</i>	39,833	9-Jul-98	
		GB_HTG2: AC004064	185000	AC004064	<i>Homo sapiens</i> chromosome 4, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.	<i>Homo sapiens</i>	39,833	9-Jul-98	
		GB_PR3: HSJ824F16	139330	AL050325	Human DNA sequence from clone 824F16 on chromosome 20, complete sequence.	<i>Homo sapiens</i>	39,833	23-Nov-99	
		GB_HTG3: AC010120	121582	AC010120	<i>Drosophila melanogaster</i> chromosome 3 clone BACR22N13 (D1061) RPCL98 22.N.13 map 96F-96F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 83 unordered pieces.	<i>Drosophila melanogaster</i>	36,855	24-Sep-99	
rxa00852	813	GB_HTG3: AC010120	121582	AC010120	<i>Drosophila melanogaster</i> chromosome 3 clone BACR22N13 (D1061) RPCL98 22.N.13 map 96F-96F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 83 unordered pieces.	<i>Drosophila melanogaster</i>	36,855	24-Sep-99	
		GB_HTG2: AC006898	299308	AC006898	<i>Caenorhabditis elegans</i> clone Y73B6x, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.	<i>Caenorhabditis elegans</i>	36,768	24-Feb-99	

TABLE 4-continued

ALIGNMENT RESULTS							
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	% homology (GAP) Date of Deposit	
rx00856 rx00870	1635	GB_BA1: STMMSDA	3986	L48550	<i>Streptomyces coelicolor</i> methylmalonic acid semialdehyde dehydrogenase (msdA) gene, complete cds.	<i>Streptomyces coelicolor</i>	63,743 09-MAY-1996
		GB_PAT: I92043	713	I92043	Sequence 10 from patent U.S. Pat. No. 5726299.	Unknown.	38,850 01-DEC-1998
		GB_PAT: I78754	713	I78754	Sequence 10 from patent U.S. Pat. No. 5693781.	Unknown.	38,850 3-APR-98
	690	GB_BA2: AF119715	549	AF119715	<i>Escherichia coli</i> isopenitl diphosphate isomerase (idi) gene, complete cds.	<i>Escherichia coli</i>	54,827 22-APR-99
		GB_BA2: AE000372	12144	AE000372	<i>Escherichia coli</i> K-12 MG1655 section 262 of 400 of the complete genome.	<i>Escherichia coli</i>	51,416 12-NOV-98
rx00875		GB_BA1: ECU28375	55175	U28375	<i>Escherichia coli</i> K-12 genome; approximately 64 to 65 minutes.	<i>Escherichia coli</i>	51,416 08-DEC-1995
	1986	GB_HTG2: AC007472	114003	AC007472	<i>Drosophila melanogaster</i> chromosome 2 clone BACR30D19 (D587) RPCL-98 30.D.19 map 49E-49F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 79 unordered pieces.	<i>Drosophila melanogaster</i>	36,592 2-AUG-99
		GB_HTG2: AC007472	114003	AC007472	<i>Drosophila melanogaster</i> chromosome 2 clone BACR30D19 (D587) RPCL-98 30.D.19 map 49E-49F strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 79 unordered pieces.	<i>Drosophila melanogaster</i>	36,592 2-AUG-99
		GB_HTG2: AC006798	207370	AC006798	<i>Caenorhabditis elegans</i> clone Y51F8, *** SEQUENCING IN PROGRESS ***, 30 unordered pieces.	<i>Caenorhabditis elegans</i>	36,699 25-FEB-99
	1968	GB_EST4: H22888	468	H22888	ym54e12.r1 Soares infant brain IN1B <i>Homo sapiens</i> cDNA clone IMAGE: 52158 5', mRNA sequence.	<i>Homo sapiens</i>	37,179 6-JUL-95
rx00880		GB_GSSI3: AQ426858	516	AQ426858	CITBI-E1-2578F1.TF CITBI-E1 <i>Homo sapiens</i> genomic clone 2578F1, genomic survey sequence.	<i>Homo sapiens</i>	38,447 24-MAR-1999
		GB_PRI: AB002335	6289	AB002335	Human mRNA for KIAA0337 gene, complete cds.	<i>Homo sapiens</i>	35,799 13-FEB-99
	1389	GB_BA1: NGU58849	2401	U58849	<i>Neisseria gonorrhoeae</i> pilS6 silent pilus locus.	<i>Neisseria gonorrhoeae</i>	40,623 20-JUN-96
		GB_BA1: PLPDHOS	3119	L06822	Plasmid pSa (from <i>Escherichia coli</i>) dihydropteroate synthase gene, 3' end.	<i>Plasmid pSa</i>	38,966 20-MAR-1996
		GB_BA1: PDGINTORE	6747	L06418	Integron In7 (from Plasmid pDGO100 from <i>Escherichia coli</i>) integrase (int), aminoglycoside adenyltransferase (aad), quaternary ammonium compound-resistance protein, dihydrofolate reductase (dhfrX), and dihydropteroate synthase (sulI) genes.	<i>Plasmid pDGO100</i>	38,966 20-MAR-1996
rx00902	1333	GB_GSSI5: AQ606873	581	AQ606873	HS_5404_B2_H05_T7A RPCL-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 980 Col = 10 Row = P, genomic survey sequence.	<i>Homo sapiens</i>	37,900 10-JUN-99
		GB_GSS9: AQ163442	658	AQ163442	nbxb0007A07f CUGI Rice BAC Library <i>Oryza sativa</i> genomic clone nbxb0007A07f, genomic survey sequence.	<i>Oryza sativa</i>	41,885 12-SEP-98
		GB_PL1: PSST70	4974	X69213	<i>P. sativum</i> Pss70 gene for heat-shock protein.	<i>Pisum sativum</i>	36,866 3-JUL-96
	969	GB_GSSI: FR0025208	612	AL018047	<i>F. rubripes</i> GSS sequence, clone 145D10aA8, genomic survey sequence.	<i>Fugu rubripes</i>	37,815 10-DEC-1997
		GB_GSSI: FR0021844	252	AL014715	<i>F. rubripes</i> GSS sequence, clone 069K22aG5, genomic survey sequence.	<i>Fugu rubripes</i>	37,698 10-DEC-1997
rx00941		GB_GSSI2: AQ403344	593	AQ403344	HS_2257_B1_B03_T7C CTT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 2257 Col = 5 Row = D, genomic survey sequence.	<i>Homo sapiens</i>	31,552 13-MAR-1999
	1440	GB_BA1: MTCY180	44201	Z97193	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 85/162.	<i>Mycobacterium tuberculosis</i>	37,902 17-JUN-98
		GB_BA1: MTCY180	44201	Z97193	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 85/162.	<i>Mycobacterium tuberculosis</i>	39,140 17-JUN-98
		GB_BA2: MSGKATG	1745	L14268	<i>Mycobacterium tuberculosis</i> ethyl methane sulphonate resistance protein (katG) gene, 3'end.	<i>Mycobacterium tuberculosis</i>	42,517 26-AUG-99
	689	GB_HTG6: AC010998	144338	AC010998	<i>Homo sapiens</i> clone RP11-95116, *** SEQUENCING IN PROGRESS ***, 17 unordered pieces.	<i>Mycobacterium tuberculosis</i>	39,497 08-DEC-1999
rx00962		GB_GSSI: GGA340111	990	AI232089	<i>Gallus gallus</i> anonymous sequence from Cosmid mapping to chromosome 2 (Cosmid 34 - Contig 15), genomic survey sequence.	<i>Gallus gallus</i>	37,970 25-AUG-98

TABLE 4-continued

ALIGNMENT RESULTS								
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxax01060	1047	GB_HTG6: AC010998	144338	AC010998	<i>Homo sapiens</i> clone RP11-95116, *** SEQUENCING IN PROGRESS ***, 17 unordered pieces.	<i>Homo sapiens</i>	38,226	08-DEC-1999
		GB_BA1: ECTTN7	2280	AJ001816	<i>Escherichia coli</i> left end of transposon Tn7 including type 2 Integron.	<i>Escherichia coli</i>	38,822	4-Nov-97
		GB_IN2: AF176377	8220	AF176377	<i>Caenorhabditis briggsae</i> CES-1 (ces-1) gene, complete cds; and CPN-1 (cpn-1) gene, partial cds.	<i>Caenorhabditis briggsae</i>	39,921	09-DEC-1999
		GB_GSSI0: AQ196728	429	AQ196728	CIT-HSP-2381F4.TR CIT-HSP <i>Homo sapiens</i> genomic clone 2381F4, genomic survey sequence.	<i>Homo sapiens</i>	39,019	16-Sep-98
rxax01067	852	GB_BA1: U00016	42931	U00016	<i>Mycobacterium leprae</i> cosmid B1937.	<i>Mycobacterium leprae</i>	58,303	01-MAR-1994
		GB_BA1: SYCGROESL	3256	D12677	<i>Synechocystis</i> sp. groES and groEL genes.	<i>Synechocystis</i> sp.	34,593	3-Feb-99
		GB_BA1: D90905	139467	D90905	<i>Synechocystis</i> sp. PCC6803 complete genome, 7/27, 781,449-920,915.	<i>Synechocystis</i> sp.	34,593	7-Feb-99
		GB_BA1: PSEFAOAB	3480	D10390	<i>P. fragi</i> faoA and faoB genes, complete cds.	<i>Pseudomonas fragi</i>	51,919	2-Feb-99
rxax01114	1347	GB_BA1: AB014757	6057	AB014757	<i>Pseudomonas</i> sp. 61-3 genes for PhbR, acetoacetyl-CoA reductase, beta-ketothiolase and PHB synthase, complete cds.	<i>Pseudomonas</i> sp. 61-3	50,573	26-DEC-1998
		GB_BA1: SC8D9	38681	AL035569	<i>Streptomyces coelicolor</i> cosmid 8D9.	<i>Streptomyces coelicolor</i>	42,200	26-Feb-99
		GB_EST11: AA244557	379	AA244557	mx07.a01.r1 Soares mouse NML <i>Mus musculus</i> cDNA clone IMAGE: 679464 5', mRNA sequence.	<i>Mus musculus</i>	39,050	10-MAR-1997
		GB_EST14: AA407673	306	AA407673	EST01834 Mouse 7.5 dpc embryo ectoplacental cone cDNA library <i>Mus musculus</i> cDNA clone C0014F02 3', mRNA sequence.	<i>Mus musculus</i>	38,562	26-Aug-98
rxax01138	540	GB_EST26: AI390328	604	AI390328	mx07.a01.y1 Soares mouse NML <i>Mus musculus</i> cDNA clone IMAGE: 679464 5', mRNA sequence.	<i>Mus musculus</i>	33,136	2-Feb-99
		GB_OV: XLXINT1	1278	X13138	<i>Xenopus laevis</i> int-1 mRNA for int-1 protein.	<i>Xenopus laevis</i>	40,038	31-MAR-1995
		GB_PR4: AC006054	143738	AC006054	<i>Homo sapiens</i> Xq28 BAC RPC111-382P7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	<i>Homo sapiens</i>	37,996	1-Apr-99
		GB_PR4: AC006054	143738	AC006054	<i>Homo sapiens</i> Xq28 BAC RPC111-382P7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	<i>Homo sapiens</i>	36,053	1-Apr-99
rxax01172	1578	GB_BA1: SCE39	23550	AL049573	<i>Streptomyces coelicolor</i> cosmid E39.	<i>Streptomyces coelicolor</i>	62,357	31-MAR-1999
		GB_BA1: MSU50335	5193	U50335	<i>Mycobacterium smegmatis</i> phage resistance (mpr) gene, complete cds.	<i>Mycobacterium smegmatis</i>	37,853	1-Feb-97
		GB_BA1: BACTHTRNA	15467	D84213	<i>Bacillus subtilis</i> genome, trnI-feuABC region.	<i>Bacillus subtilis</i>	53,807	6-Feb-99
		GB_PR2: HSI191B2	60828	AL022237	Human DNA sequence from clone 1191B2 on chromosome 22q13.2-13.3. Contains part of the BIK (NBK, BP4, BIP1) gene for BCL2-interacting killer (apoptosis-inducing), a 40S Ribosomal Protein S25 pseudogene and part of an alternatively spliced novel Aeyl Transferase gene similar to <i>C. elegans</i> C50D2.7. Contains ESTs, STSs, GSSs, two putative CpG islands and genomic marker D22S1151, complete sequence.	<i>Homo sapiens</i>	38,366	23-Nov-99
rxax01191	1713	GB_PR2: HSI191B2	60828	AL022237	Human DNA sequence from clone 1191B2 on chromosome 22q13.2-13.3. Contains part of the BIK (NBK, BP4, BIP1) gene for BCL2-interacting killer (apoptosis-inducing), a 40S Ribosomal Protein S25 pseudogene and part of an alternatively spliced novel Aeyl Transferase gene similar to <i>C. elegans</i> C50D2.7. Contains ESTs, STSs, GSSs, two putative CpG islands and genomic marker D22S1151, complete sequence.	<i>Homo sapiens</i>	39,595	23-Nov-99
		GB_BA1: MTCY373	35516	Z73419	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 57/162.	<i>Mycobacterium tuberculosis</i>	57,762	17-Jun-98
		GB_PL1: ATY12776	38483	Y12776	<i>Arabidopsis thaliana</i> DNA, 40 kb surrounding ACS1 locus.	<i>Arabidopsis thaliana</i>	32,971	7-Sep-98
		GB_PL2: ATT6K21	99643	AL021889	<i>Arabidopsis thaliana</i> DNA chromosome 4, BAC clone T6K21 (ESSA project).	<i>Arabidopsis thaliana</i>	35,273	16-Aug-99
rxax01205	554	GB_BA1: MTCY373	35516	Z73419	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 57/162.	<i>Mycobacterium tuberculosis</i>	57,762	17-Jun-98

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	% homology (GAP)
rxa01212	1047	GB_BA2: SCD25	41622	AL118514	<i>Streptomyces coelicolor</i> cosmid D25.	39,654 21-Sep-99
rxs01219	1005	GB_BA1: SLGLYUB	2576	X65556	<i>S. lividans</i> tRNA-GlyU beta gene.	54,493 20-DEC-1993
		GB_BA1: SCH10	39524	AL049754	<i>Streptomyces coelicolor</i> cosmid H10.	44,638 04-MAY-1999
		GB_PAT: A68024	520	A68024	Sequence 19 from Patent WO9743409.	42,553 05-MAY-1999
		GB_PAT: A68025	193	A68025	Sequence 20 from Patent WO9743409.	43,229 05-MAY-1999
		GB_PAT: A68027	193	A68027	Sequence 22 from Patent WO9743409.	38,342 05-MAY-1999
rxs01220	1200	GB_PR3: H8512B11	64356	AL031058	Human DNA sequence from clone 512B11 on chromosome 6p24-25. Contains the Desmoplakin I (DPI) gene, ESTs, STSs and GSSs, complete sequence.	35,478 23-Nov-99
		GB_EST6: N99239	424	N99239	z676h1.s1 Soares_senescent_fibroblasts_NBHSF <i>Homo sapiens</i> cDNA clone IMAGE: 309573 3', mRNA sequence.	39,623 20-Aug-96
		GB_EST16: AA554268	400	AA554268	nk36c09.s1 NCL CGAP_GC2 <i>Homo sapiens</i> cDNA clone IMAGE: 1015600 3' similar to gb: X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, LIVER (HUMAN); mRNA sequence.	36,111 8-Sep-97
rxs01221	849	GB_PR4: AF179633	96371	AF179633	<i>Homo sapiens</i> chromosome 16 map 16q23.3-q24.1 sequence.	40,199 5-Sep-99
rxs01222	822	GB_VI: EHVVU20824	184427	U20824	Equine herpesvirus 2, complete genome.	37,001 2-Feb-96
		GB_BA2: AE000407	10601	AE000407	<i>Escherichia coli</i> K-12 MG1655 section 297 of 400 of the complete genome.	39,471 12-Nov-98
		GB_PAT: AR068625	28804	AR068625	Sequence 1 from patent U.S. Pat. No. 5854034.	40,574 29-Sep-99
		GB_BA2: SSU51197	28804	U51197	<i>Sphingomonas</i> S88 sphingian polysaccharide synthesis (spsG), (spsS), (spsR), glycosyl transferase (spsQ), (spsI), glycosyl transferase (spsK), glycosyl transferase (spsL), (spsF), (spsD), (spsC), (spsE), Urf 32, Urf 26, ATP-binding cassette transporter (atrD), ATP-binding cassette transporter (atrB), glucosyl-isoprenylphosphate transferase (spsB), glucose-1-phosphate thymidyl transferase (thsA), dTDP-6-deoxy-D-glucose-3,5-epimerase (rhcC) dTDP-D-glucose-4,6-dehydratase (rhcB), dTDP-6-deoxy-L-mannose-dehydrogenase (rhcD), Urf 31, and Urf 34 genes, complete cds.	40,574 16-MAY-1996
rxs01260	1305	GB_IN1: BBU44918	2791	U44918	<i>Babesia bovis</i> ATP-binding protein (babo) mRNA, complete cds.	39,228 9-Aug-97
		GB_BA1: CGLPD	1800	Y16642	<i>Corynebacterium glutamicum</i> lpd gene, complete CDS.	99,923 1-Feb-99
rxs01261	294	GB_BA1: MTV038	16094	AL021933	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 24/162.	59,056 17-Jun-98
		GB_PR3: AC005618	176714	AC005618	<i>Homo sapiens</i> chromosome 5, BAC clone 249i5 (LBNL H149), complete sequence.	36,270 5-Sep-98
		GB_BA1: CGLPD	1800	Y16642	<i>Corynebacterium glutamicum</i> lpd gene, complete CDS.	100,000 1-Feb-99
rxs01269	564	GB_HITG4: AC010045	164829	AC010045	<i>Drosophila melanogaster</i> chromosome 3L/75A1 clone RPC198-17C17, *** SEQUENCING IN PROGRESS ***, 50 unordered pieces.	50,512 16-OCT-1999
		GB_HITG4: AC010045	164829	AC010045	<i>Drosophila melanogaster</i> chromosome 3L/75A1 clone RPC198-17C17, *** SEQUENCING IN PROGRESS ***, 50 unordered pieces.	50,512 16-OCT-1999
		GB_BA2: AFI25164	26443	AFI25164	<i>Bacteroides fragilis</i> 638R polysaccharide B (PS B2) biosynthesis locus, complete sequence; and unknown genes.	56,071 01-DEC-1999
		GB_BA1: AB002668	24907	AB002668	<i>Actinobacillus actinomycetemcomitans</i> DNA for glycosyltransferase, lytic transglycosylase, dTDP-4-rhamnose reductase, complete cds.	46,679 21-Feb-98

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxa01291	1056	GB_BA1: AB010415	23112	AB010415	<i>Actinobacillus actinomycetemcomitans</i> gene cluster for 6-deoxy-L-talan synthesis, complete cds.	<i>Actinobacillus actinomycetemcomitans</i>	46,679	13-Feb-99	
		GB_STS: AU027820	238	AU027820	<i>Rattus norvegicus</i> OTSUKA clone, OT78.02/918b07, microsatellite sequence, sequence tagged site.	<i>Rattus norvegicus</i>	34,874	02-MAR-1999	
		GB_STS: AU027820	238	AU027820	<i>Rattus norvegicus</i> OTSUKA clone, OT78.02/918b07, microsatellite sequence, sequence tagged site.	<i>Rattus norvegicus</i>	34,874	02-MAR-1999	
		GB_HTG3: AC006445	174547	AC006445	<i>Homo sapiens</i> chromosome 4, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	<i>Homo sapiens</i>	34,812	15-Sep-99	
rxa01292	1308	GB_BA1: BSUB0017	217420	Z99120	<i>Bacillus subtilis</i> complete genome (section 17 of 21): from 3197001 to 3414420.	<i>Bacillus subtilis</i>	37,802	26-Nov-97	
		GB_HTG3: AC010580	121119	AC010580	<i>Drosophila melanogaster</i> chromosome 3 clone BACR48J06 (D1102) RPCI-98 48J.1.6 map 96F-96F strain Y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 71 unordered pieces.	<i>Drosophila melanogaster</i>	35,637	01-OCT-1999	
		GB_HTG3: AC010580	121119	AC010580	<i>Drosophila melanogaster</i> chromosome 3 clone BACR48J06 (D1102) RPCI-98 48J.1.6 map 96F-96F strain Y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 71 unordered pieces.	<i>Drosophila melanogaster</i>	35,637	01-OCT-1999	
		GB_GSS8: AQ001809	705	AQ001809	CIT-HSP-2290D17.1F CIT-HSP <i>Homo sapiens</i> genomic clone 2290D17, genomic survey sequence.	<i>Homo sapiens</i>	42,021	26-Jun-98	
rxa01339	1111	GB_GSS8: AQ001809	705	AQ001809	CIT-HSP-2290D17.1F CIT-HSP <i>Homo sapiens</i> genomic clone 2290D17, genomic survey sequence.	<i>Homo sapiens</i>	40,323	26-Jun-98	
		GB_PL1: MGU60290	4614	U60290	<i>Magnaporthe grisea</i> nitrogen regulatory protein (NUT1) gene, complete cds.	<i>Magnaporthe grisea</i>	38,707	3-Jul-96	
		GB_HTG3: AC011371	189187	AC011371	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_107C20, *** SEQUENCING IN PROGRESS ***, 31 unordered pieces.	<i>Homo sapiens</i>	39,741	06-OCT-1999	
		GB_HTG3: AC011371	189187	AC011371	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_107C20, *** SEQUENCING IN PROGRESS ***, 31 unordered pieces.	<i>Homo sapiens</i>	39,741	06-OCT-1999	
rxa01382	1192	GB_HTG4: AC009892	138122	AC009892	<i>Homo sapiens</i> chromosome 19 clone CIT978SKB_83J4, *** SEQUENCING IN PROGRESS ***, 6 ordered pieces.	<i>Homo sapiens</i>	40,154	31-OCT-1999	
		GB_HTG4: AC009892	138122	AC009892	<i>Homo sapiens</i> chromosome 19 clone CIT978SKB_83J4, *** SEQUENCING IN PROGRESS ***, 6 ordered pieces.	<i>Homo sapiens</i>	40,154	31-OCT-1999	
		GB_PR3: AC002416	128915	AC002416	Human Chromosome X, complete sequence.	<i>Homo sapiens</i>	37,521	29-Jan-98	
		GB_EST9: AA096601	524	AA096601	mo3b09.r1 Stratagene mouse lung 937302 <i>Mus musculus</i> cDNA clone IMAGE: 552473 5' similar to gb: L06505 60S RIBOSOMAL PROTEIN L12 (HUMAN); gb: L04280 <i>Mus musculus</i> ribosomal protein (MOUSE); mRNA sequence.	<i>Mus musculus</i>	40,525	15-Feb-97	
rxa01420	1065	GB_EST37: AI982114	626	AI982114	pat.pk0074.e9.f chicken activated T cell cDNA <i>Gallus gallus</i> cDNA clone pat.pk0074.e9.f 5' similar to H-ATPase B subunit, mRNA sequence.	<i>Gallus gallus</i>	37,785	15-Sep-99	
		GB_OV: GGU20766	1645	U20766	<i>Gallus gallus</i> vacuolar H+-ATPase B subunit gene, complete cds.	<i>Gallus gallus</i>	38,244	07-DEC-1995	
		GB_HTG2: AC005690	193424	AC005690	<i>Homo sapiens</i> chromosome 4, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	<i>Homo sapiens</i>	37,464	11-Apr-99	
		GB_HTG2: AC005690	193424	AC005690	<i>Homo sapiens</i> chromosome 4, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	<i>Homo sapiens</i>	37,464	11-Apr-99	
rxa01420	1065	GB_HTG2: AC006637	22092	AC006637	<i>Caenorhabditis elegans</i> clone F41B4, *** SEQUENCING IN PROGRESS ***, 1 unordered pieces.	<i>Caenorhabditis elegans</i>	37,488	23-Feb-99	

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxaxa01467	414	GB_HTG1: CEY102G3_21	10000	AL020985	<i>Caenorhabditis elegans</i> chromosome V clone Y102G3, *** SEQUENCING IN	<i>Caenorhabditis elegans</i>	35,437	3-Dec-98	
		GB_HTG1: CEY102G3_21	10000	AL020985	<i>Caenorhabditis elegans</i> chromosome V clone Y102G3, *** SEQUENCING IN	<i>Caenorhabditis elegans</i>	35,437	3-Dec-98	
		GB_HTG1: CEY113G7_41	10000	AL031113	<i>Caenorhabditis elegans</i> chromosome V clone Y113G7, *** SEQUENCING IN	<i>Caenorhabditis elegans</i>	35,437	12-Jan-99	
	882	GB_BA2: AF030975	2511	AF030975	<i>Aeromonas salmonicida</i> chaperonin GroES and chaperonin GroEL genes, complete cds.	<i>Aeromonas salmonicida</i>	41,516	2-Apr-98	
rxaxa01576		GB_BA2: AF030975	2511	AF030975	<i>Aeromonas salmonicida</i> chaperonin GroES and chaperonin GroEL genes, complete cds.	<i>Aeromonas salmonicida</i>	38,171	2-Apr-98	
		GB_EST22: AI068560	965	AI068560	mgae0003aC11f <i>Magnaporthe grisea</i> Appressorium Stage cDNA Library Pyricularia grisea cDNA clone mgae0003aC11f, mRNA sequence.	<i>Pyricularia grisea</i>	40,073	09-DEC-1999	
	840	GB_GSS14: AQ554460	681	AQ554460	RPC1-11-419F2.TV RPC1-11 <i>Homo sapiens</i> genomic clone RPC1-11-419F2, genomic survey sequence.	<i>Homo sapiens</i>	36,522	28-MAY-1999	
		GB_IN2: AC005449	85518	AC005449	<i>Drosophila melanogaster</i> , chromosome 2R, region 44C4-44C5, P1 clone DS06765, complete sequence.	<i>Drosophila melanogaster</i>	36,609	23-DEC-1998	
rxaxa01584 rxaxa01604		GB_IN2: AC005449	85518	AC005449	<i>Drosophila melanogaster</i> , chromosome 2R, region 44C4-44C5, P1 clone DS06765, complete sequence.	<i>Drosophila melanogaster</i>	33,612	23-DEC-1998	
	771	GB_HTG3: AC011352	160167	AC011352	<i>Homo sapiens</i> chromosome 5 clone CIT-HSPC_327F10, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.	<i>Homo sapiens</i>	33,688	06-OCT-1999	
		GB_HTG3: AC011352	160167	AC011352	<i>Homo sapiens</i> chromosome 5 clone CIT-HSPC_327F10, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.	<i>Homo sapiens</i>	33,688	06-OCT-1999	
		GB_HTG3: AC011402	168868	AC011402	<i>Homo sapiens</i> chromosome 5 clone CIT978SKB_38B5, *** SEQUENCING IN PROGRESS ***, 7 unordered pieces.	<i>Homo sapiens</i>	33,688	06-OCT-1999	
rxaxa01614	1146	GB_BA1: CGA224946	2408	AJ224946	<i>Corynebacterium glutamicum</i> DNA for L-Malate:quinone oxidoreductase.	<i>Corynebacterium glutamicum</i>	42,284	11-Aug-98	
		GB_EST17: AA608825	439	AA608825	af03g07.s1 Soares_testis_NHT <i>Homo sapiens</i> cDNA clone IMAGE: 1030620 3' similar to TR: G976083 G976083 HISTONE H2A RELATED.; mRNA sequence.	<i>Homo sapiens</i>	40,092	02-MAR-1998	
		GB_PR4: AC005377	102311	AC005377	<i>Homo sapiens</i> PAC clone DJ1136G02 from 7q32-q34, complete sequence.	<i>Homo sapiens</i>	37,811	28-Apr-99	
	1635	GB_BA1: CGPROPGEN	2936	Y12537	<i>C. glutamicum</i> proP gene.	<i>Corynebacterium glutamicum</i>	100,000	17-Nov-98	
rxaxa01629		GB_BA1: CGPROPGEN	2936	Y12537	<i>C. glutamicum</i> proP gene.	<i>Corynebacterium glutamicum</i>	100,000	17-Nov-98	
		GB_PR4: AF191071	88481	AF191071	<i>Homo sapiens</i> chromosome 8 clone BAC 388D06, complete sequence.	<i>Homo sapiens</i>	35,612	11-OCT-1999	
	1401	GB_BA1: MSGB577COS	37770	L01263	<i>M. leprae</i> genomic dna sequence, cosmid b577.	<i>Mycobacterium leprae</i>	55,604	14-Jun-96	
		GB_BA1: MLCB2407	35615	AL023596	<i>Mycobacterium leprae</i> cosmid B2407.	<i>Mycobacterium leprae</i>	36,416	27-Aug-99	
rxaxa01644		GB_BA1: MTV025	121125	AL022121	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 155/162.	<i>Mycobacterium tuberculosis</i>	55,844	24-Jun-99	
	1329	GB_BA1: CGU43536	3464	U43536	<i>Corynebacterium glutamicum</i> heat shock, ATP-binding protein (elpB) gene, complete cds.	<i>Corynebacterium glutamicum</i>	100,000	13-MAR-1997	
		GB_HTG4: AC009841	164434	AC009841	<i>Drosophila melanogaster</i> chromosome 3L/77E1 clone RPC198-13F11, *** SEQUENCING IN PROGRESS ***, 70 unordered pieces.	<i>Drosophila melanogaster</i>	33,205	16-OCT-1999	
		GB_HTG4: AC009841	164434	AC009841	<i>Drosophila melanogaster</i> chromosome 3L/77E1 clone RPC198-13F11, *** SEQUENCING IN PROGRESS ***, 70 unordered pieces.	<i>Drosophila melanogaster</i>	33,205	16-OCT-1999	

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rx01722	1848	GB_GSSI1: FR0022586	522	AL015452	<i>F. rubripes</i> GSS sequence, clone 077P23aB10, genomic survey sequence.	<i>Fugu rubripes</i>	40,192	10-DEC-1997	
		GB_GSSI1: FR0022584	485	AL015450	<i>F. rubripes</i> GSS sequence, clone 077P23aB11, genomic survey sequence.	<i>Fugu rubripes</i>	35,876	10-DEC-1997	
		GB_IN1: CET26H2	37569	Z82055	<i>Caenorhabditis elegans</i> cosmid T26H2, complete sequence.	<i>Caenorhabditis elegans</i>	34,759	19-Nov-99	
		GB_BA2: CORCSLYS	2821	M89931	<i>Corynebacterium glutamicum</i> beta C-S lyase (acd) and branched-chain amino acid uptake carrier (brnQ) genes, complete cds, and hypothetical protein Ybbw (ybbw) gene, partial cds.	<i>Corynebacterium glutamicum</i>	99,929	4-Jun-98	
rx01727		GB_HTG6: AC011037	167849	AC011037	<i>Homo sapiens</i> clone RP11-7F18, WORKING DRAFT SEQUENCE, 19 unordered pieces.	<i>Homo sapiens</i>	36,903	30-Nov-99	
		GB_HTG6: AC011037	167849	AC011037	<i>Homo sapiens</i> clone RP11-7F18, WORKING DRAFT SEQUENCE, 19 unordered pieces.	<i>Homo sapiens</i>	35,642	30-Nov-99	
rx01737	1182	GB_BA1: SCGD3	33779	AL096822	<i>Streptomyces coelicolor</i> cosmid GD3.	<i>Streptomyces coelicolor</i>	38,054	8-Jul-99	
		GB_HTG1: CNS01DSB	222193	AL121768	<i>Homo sapiens</i> chromosome 14 clone R-976B16, *** SEQUENCING IN PROGRESS ***, in ordered pieces.	<i>Homo sapiens</i>	35,147	05-OCT-1999	
rx01762	1659	GB_HTG1: CNS01DSB	222193	AL121768	<i>Homo sapiens</i> chromosome 14 clone R-976B16, *** SEQUENCING IN PROGRESS ***, in ordered pieces.	<i>Homo sapiens</i>	35,147	05-OCT-1999	
		GB_BA1: MTC128	36300	Z97050	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 10/162.	<i>Mycobacterium tuberculosis</i>	49,574	23-Jun-98	
rx01764	1056	GB_BA1: SC6G10	36734	AL049497	<i>Streptomyces coelicolor</i> cosmid 6G10.	<i>Streptomyces coelicolor</i>	44,049	24-MAR-1999	
		GB_BA1: SCE29	26477	AL035707	<i>Streptomyces coelicolor</i> cosmid E29.	<i>Streptomyces coelicolor</i>	40,246	12-MAR-1999	
		GB_PL2: SPAC343	42947	AL109739	<i>S. pombe</i> chromosome I cosmid c343.	<i>Schizosaccharomyces pombe</i>	37,084	6-Sep-99	
		GB_PL2: SPAC343	42947	AL109739	<i>S. pombe</i> chromosome I cosmid c343.	<i>Schizosaccharomyces pombe</i>	34,890	6-Sep-99	
rx01801	1140	GB_EST38: AW066306	334	AW066306	687009D03.y1 687 Early embryo from Delaware <i>Zea mays</i> cDNA, mRNA sequence.	<i>Zea mays</i>	46,108	12-OCT-1999	
rx01823	900	GB_GSSI3: AQ484750	375	AQ484750	RPC1-11-248N4.TV RPC1-11 <i>Homo sapiens</i> genomic clone RPC1-11-248N4, genomic survey sequence.	<i>Homo sapiens</i>	32,000	24-Apr-99	
		GB_GSSI3: AQ489971	252	AQ489971	RPC1-11-247N23.TV RPC1-11 <i>Homo sapiens</i> genomic clone RPC1-11-247N23, genomic survey sequence.	<i>Homo sapiens</i>	36,111	24-Apr-99	
		GB_BA1: SCI51	40745	AL109848	<i>Streptomyces coelicolor</i> cosmid I51.	<i>Streptomyces coelicolor</i> A3(2)	35,779	16-Aug-99	
		GB_BA1: ECU82598	136742	U82598	<i>Escherichia coli</i> genomic sequence of minutes 9 to 12.	<i>Escherichia coli</i>	39,211	15-Jan-97	
rx01853	675	GB_BA1: BSUB0018	209510	Z99121	<i>Bacillus subtilis</i> complete genome (section 18 of 21): from 3399551 to 3609060.	<i>Bacillus subtilis</i>	36,999	26-Nov-97	
		GB_BA1: MTCY227	35946	Z77724	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 114/162.	<i>Mycobacterium tuberculosis</i>	37,612	17-Jun-98	
		GB_HTG3: AC010189	265962	AC010189	<i>Homo sapiens</i> clone RPC11-296K13, *** SEQUENCING IN PROGRESS ***; 80 unordered pieces.	<i>Homo sapiens</i>	39,006	16-Sep-99	
		GB_HTG3: AC010189	265962	AC010189	<i>Homo sapiens</i> clone RPC11-296K13, *** SEQUENCING IN PROGRESS ***; 80 unordered pieces.	<i>Homo sapiens</i>	39,006	16-Sep-99	
rx01881	558	GB_HTG4: AC011117	148447	AC011117	<i>Homo sapiens</i> chromosome 4 clone 173_C_09 map 4, *** SEQUENCING IN PROGRESS ***; 10 ordered pieces.	<i>Homo sapiens</i>	39,130	14-OCT-1999	

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	% homology (GAP) Date of Deposit
rxa01894	978	GB_HTG4; AC011117	148447	AC011117	<i>Homo sapiens</i> chromosome 4 clone 173_C_09 map 4, *** SEQUENCING IN PROGRESS ***, 10 ordered pieces.	39,130 14-OCT-1999
		GB_BA1; MTCY2B12	20431	Z81011	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 61/162.	37,893 18-Jun-98
		GB_BA1; MTCY274	39991	Z74024	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 126/162.	37,229 19-Jun-98
		GB_IN1; CELF46H5 GB_HTG3; AC009204	38886 115633	U41543 AC009204	<i>Caenorhabditis elegans</i> cosmid F46H5. <i>Drosophila melanogaster</i> chromosome 2 clone BACR03E19 (D1033) RPC1-98 03.E.19 map 36E-37C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 94 unordered pieces.	38,525 29-Nov-96 31,579 18-Aug-99
rxa01897	666	GB_HTG1; CEY48B6	293827	AL021151	<i>Caenorhabditis elegans</i> chromosome II clone Y48B6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	34,703 1-Apr-99
		GB_HTG1; CEY48B6	293827	AL021151	<i>Caenorhabditis elegans</i> chromosome II clone Y48B6, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	34,703 1-Apr-99
		GB_HTG1; CEY53F4_2	110000	Z92860	<i>Caenorhabditis elegans</i> chromosome II clone Y53F4, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	33,333 15-Oct-99
		GB_BA1; MTV007	32806	AL021184	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 64/162.	65,560 17-Jun-98
rxa01980	756	GB_BA1; SC5F2A GB_BA1; SCARD1GN GB_PL2; AC008262	40105 2321 99698	AL049587 X84374 AC008262	<i>Streptomyces coelicolor</i> cosmid 5F2A. <i>S. capreolus</i> ard1 gene. Genomic sequence for <i>Arabidopsis thaliana</i> BAC F4N2 from chromosome I, complete sequence.	50,648 24-MAY-1999 44,973 23-Aug-95 35,310 21-Aug-99
		GB_PL1; AB013388	73428	AB013388	<i>Arabidopsis thaliana</i> genomic DNA, chromosome 5, TAC clone: K19E1, complete sequence.	35,505 20-Nov-99
		GB_PL1; AB013388	73428	AB013388	<i>Arabidopsis thaliana</i> genomic DNA, chromosome 5, TAC clone: K19E1, complete sequence.	39,973 20-Nov-99
		GB_HTG4; AC006467	175695	AC006467	<i>Drosophila melanogaster</i> chromosome 2 clone BACR03L08 (D532) RPCL98 03.L.8 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.	36,672 27-OCT-1999
rxa01983	630	GB_HTG4; AC006467	175695	AC006467	<i>Drosophila melanogaster</i> chromosome 2 clone BACR03L08 (D532) RPCL98 03.L.8 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.	36,672 27-OCT-1999
		GB_HTG4; AC006467	175695	AC006467	<i>Drosophila melanogaster</i> chromosome 2 clone BACR03L08 (D532) RPCL98 03.L.8 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN PROGRESSSo ***, 9 unordered pieces.	32,367 27-OCT-1999
		GB_HTG4; AC006467	175695	AC006467	<i>Drosophila melanogaster</i> chromosome 2 clone BACR03L08 (D532) RPCL98 03.L.8 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.	32,367 27-OCT-1999
		GB_BA1; CGDNAAROP	2612	X85965	<i>C. glutamicum</i> ORF3 and aroP gene.	100,000 30-Nov-97
rxa02020	1111	GB_PAT; A58887 GB_BA1; STYCARABA GB_HTG2; AC003023	1612 4378 104768	A58887 M95047 AC003023	Sequence 1 from Patent WO9701637. <i>Salmonella typhimurium</i> transport protein, complete cds, and transfer RNA-Arg. <i>Homo sapiens</i> chromosome 11 clone pDJ363p2, *** SEQUENCING IN PROGRESS ***, 22 unordered pieces.	100,000 06-MAR-1998 50,547 13-MAR-1996 35,820 21-OCT-1997
		GB_HTG2; AC003023	104768	AC003023	<i>Homo sapiens</i> chromosome 11 clone pDJ363p2, *** SEQUENCING IN PROGRESS ***, 22 unordered pieces.	35,820 21-OCT-1997

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
txa02030	1509	GB_HTG2: HSI18B18	104729	AL034344	<i>Homo sapiens</i> chromosome 6 clone RP1-118B18 map p24.1-25.3, *** SEQUENCING IN PROGRESS ***, in unordered pieces.	<i>Homo sapiens</i>	34,355	03-DEC-1999	
		GB_PR4: AC007695	63247	AC007695	<i>Homo sapiens</i> 12q24 BAC RPC111-124N23 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	<i>Homo sapiens</i>	38,681	1-Sep-99	
	1653	GB_PR4: AC006464	99908	AC006464	<i>Homo sapiens</i> BAC clone NH0436C12 from 2, complete sequence.	<i>Homo sapiens</i>	35,445	22-OCT-1999	
		GB_PR4: AC006464	99908	AC006464	<i>Homo sapiens</i> BAC clone NH0436C12 from 2, complete sequence.	<i>Homo sapiens</i>	35,968	22-OCT-1999	
txa02073		GB_BA1: CGGDHA	2037	X72855	<i>C. glutamicum</i> GDHA gene.	<i>Corynebacterium glutamicum</i>	39,655	24-MAY-1993	
txa02074	1527	GB_BA1: CGGDH	2037	X59404	<i>Corynebacterium glutamicum</i> , gdh gen for glutamate dehydrogenase.	<i>Corynebacterium glutamicum</i>	44,444	30-Jul-99	
		GB_BA2: SC2H4	25970	AL031514	<i>Streptomyces coelicolor</i> cosmid 2H4.	<i>Streptomyces coelicolor</i> A3(2)	38,452	19-OCT-1999	
		GB_EST18: AA703380	471	AA703380	zj12b06.s1 Soares_fetal_liver_spleen_INFLS_S1 <i>Homo sapiens</i> cDNA clone IMAGE: 450035 3' similar to contains LTR5.13 LTR5 repetitive element; mRNA sequence.	<i>Homo sapiens</i>	36,518	24-DEC-1997	
		GB_HTG6: AC009769	122911	AC009769	<i>Homo sapiens</i> chromosome 8 clone RP11-202112 map 8, LOW-PASS SEQUENCE SAMPLING.	<i>Homo sapiens</i>	35,473	07-DEC-1999	
txa02099	373	GB_EST7: W70175	436	W70175	z452c02.r1 Soares_fetal_heart_NbHH19W <i>Homo sapiens</i> cDNA clone IMAGE: 344258 5' similar to contains LTR5.b2 LTR5 repetitive element; mRNA sequence.	<i>Homo sapiens</i>	34,174	16-OCT-1996	
		GB_BA1: CAJ10319	5368	AJ010319	<i>Corynebacterium glutamicum</i> amtP, gltB, gltD genes and partial fisY and srp genes.	<i>Corynebacterium glutamicum</i>	100,000	14-MAY-1999	
		GB_HTG3: AC011509	111353	AC011509	<i>Homo sapiens</i> chromosome 19 clone CITB-H1_2189E23, *** SEQUENCING IN PROGRESS ***, 35 unordered pieces.	<i>Homo sapiens</i>	33,423	07-OCT-1999	
		GB_HTG3: AC011509	111353	AC011509	<i>Homo sapiens</i> chromosome 19 clone CITB-H1_2189E23, *** SEQUENCING IN PROGRESS ***, 35 unordered pieces.	<i>Homo sapiens</i>	33,423	07-OCT-1999	
txa02115	1197	GB_HTG5: AC010126	175986	AC010126	<i>Homo sapiens</i> clone GS502B02, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	<i>Homo sapiens</i>	36,717	13-Nov-99	
txa02128		GB_HTG5: AC010126	175986	AC010126	<i>Homo sapiens</i> clone GS502B02, *** SEQUENCING IN PROGRESS ***, 3 unordered pieces.	<i>Homo sapiens</i>	36,092	13-Nov-99	
		GB_PRI: HUMHM145	2214	D10925	Human mRNA for HMI45.	<i>Homo sapiens</i>	39,171	3-Feb-99	
		GB_BA1: MTCY190	34150	Z70283	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 98/162.	<i>Mycobacterium tuberculosis</i>	38,682	17-Jun-98	
		GB_BA1: MTCY190	34150	Z70283	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 98/162.	<i>Mycobacterium tuberculosis</i>	35,746	17-Jun-98	
txa02133		GB_GSS10: AQ161109	738	AQ161109	nbxb006D03r CUGI Rice BAC Library <i>Oryza sativa</i> genomic clone nbxb006D03r, genomic survey sequence.	<i>Oryza sativa</i>	38,482	12-Sep-98	
		GB_BA2: MPAE000058	28530	AE000058	<i>Mycoplasma pneumoniae</i> section 58 of 63 of the complete genome.	<i>Mycoplasma pneumoniae</i>	32,317	18-Nov-96	
		GB_HTG4: AC008308	151373	AC008308	<i>Drosophila melanogaster</i> chromosome 3 clone BACR10M16 (D743) RPC1-98 10.M16 map 93C-93D strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 186 unordered pieces.	<i>Drosophila melanogaster</i>	34,579	20-OCT-1999	

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	% homology (GAP) Date of Deposit
rxa02150	924	GB_HTG4; AC008308	151373	AC008308	<i>Drosophila melanogaster</i> chromosome 3 clone BACR10M16 (D743) RPC1-98 10.M16 map 93C-93D strain Y; cn bw sp. *** SEQUENCING IN PROGRESS ***; 186 unordered pieces.	34,579 20-OCT-1999
		GB_EST37; AW012260	358	AW012260	um06e09.y1 Sugano mouse kidney mRna <i>Mus musculus</i> cDNA clone IMAGE: 2182312 5' similar to SW: AMPL_BOVIN P00727 CYTOSOL AMINOPEPTIDASE; mRNA sequence.	39,385 10-Sep-99
		GB_GSS3; B87734	389	B87734	RPC11-30D24.TP RPC1-11 <i>Homo sapiens</i> genomic clone RPC1-11-30D24, genomic survey sequence.	37,629 9-Apr-99
		GB_PR4; AC005042	192218	AC005042	<i>Homo sapiens</i> clone NH0552E01, complete sequence.	36,901 14-Jan-99
rxa02171	1776	GB_BA2; AF010496	189370	AF010496	<i>Rhodobacter capsulatus</i> strain SBI003, partial genome. EST216450 Normalized rat lung, Bento Soares <i>Rattus</i> sp. cDNA clone RLUCO75 3'	53,714 12-MAY-1998
		GB_EST24; A1170522	367	A1170522	end, mRNA sequence.	44,186 20-Jan-99
rxa02173	1575	GB_PL1; PHVDLECA	1441	K03288	<i>P. vulgaris</i> phytohemagglutinin gene encoding erythroagglutinating phytohemagglutinin (PHA-E), complete cds.	39,103 27-Apr-93
		GB_BA1; CGGLTG	3013	X66112	<i>C. glutamicum</i> glt gene for citrate synthase and ORF.	44,118 17-Feb-95
		GB_BA1; CGGLTG	3013	X66112	<i>C. glutamicum</i> glt gene for citrate synthase and ORF.	36,189 17-Feb-95
rxa02224	1920	GB_BA2; AE000104	10146	AE000104	<i>Rhizobium</i> sp. NGR234 plasmid pNGR234a, section 41 of 46 of the complete plasmid sequence.	38,487 12-DEC-1997
		GB_BA2; CXU21300	8990	U21300	<i>Corynebacterium striatum</i> hypothetical protein YhbB gene, partial cds. ABC transporter TetB (tetB), ABC transporter TetA (tetA), transposase, 23S rRNA methyltransferase, and transposase genes, complete cds; and unknown genes.	37,264 9-Apr-99
rxa02225	905	GB_HTG3; AC009185	87184	AC009185	<i>Homo sapiens</i> chromosome 5 clone CIT-HSPC_248019, *** SEQUENCING IN PROGRESS ***; 2 ordered pieces.	36,459 07-OCT-1999
		GB_HTG3; AC009185	87184	AC009185	<i>Homo sapiens</i> chromosome 5 clone CIT-HSPC_248019, *** SEQUENCING IN PROGRESS ***; 2 ordered pieces.	36,459 07-OCT-1999
		GB_BA2; MPAE000058	28530	AE000058	<i>Mycoplasma pneumoniae</i> section 58 of 63 of the complete genome.	35,498 18-Nov-96
		GB_EST26; A1337275	618	A1337275	tb96h11.x1 NCL CGAP_Co16 <i>Homo sapiens</i> cDNA clone IMAGE: 2062245 3' similar to TR: Q15392 Q15392 ORF, COMPLETE CDS.;, mRNA sequence.	35,589 18-MAR-1999
rxa02233	1410	GB_EST26; A1337275	618	A1337275	tb96h11.x1 NCL CGAP_Co16 <i>Homo sapiens</i> cDNA clone IMAGE: 2062245 3' similar to TR: Q15392 Q15392 ORF, COMPLETE CDS.;, mRNA sequence.	42,786 18-MAR-1999
		GB_BA1; ERWPNLB	1291	M65057	<i>Erwinia carotovora</i> pectin lyase (pnl) gene, complete cds.	37,780 26-Apr-93
		GB_EST30; AV021947	313	AV021947	AV021947 <i>Mus musculus</i> 18-day embryo C57BL/6J <i>Mus musculus</i> cDNA clone 1190024M23, mRNA sequence.	39,423 28-Aug-99
rxa02253	1050	GB_EST33; AV087117	251	AV087117	AV087117 <i>Mus musculus</i> tongue C57BL/6J adult <i>Mus musculus</i> cDNA clone 2310028C15, mRNA sequence.	47,410 25-Jun-99
		GB_EST11; AA250210	532	AA250210	mx79g10.r1 Soares mouse NML <i>Mus musculus</i> cDNA clone IMAGE: 692610 5' similar to TR: E236517 E236517 F44G4.1.; mRNA sequence.	36,136 12-MAR-1997

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxa02261		GB_EST11: AA250210	532	AA250210	mx79g10.r1 Soares mouse NML <i>Mus musculus</i> cDNA clone IMAGE: 692610 5' similar to TR: E236517 F44G4.1.; mRNA sequence.	<i>Mus musculus</i>	36,202	12-MAR-1997	
	1479	GB_BA1: CGL007732	4460	AJ007732	<i>Corynebacterium glutamicum</i> 3' ppc gene, secG gene, amt gene, ocd gene and 5' soxA gene.	<i>Corynebacterium glutamicum</i>	100,000	7-Jan-99	
		GB_BA1: CGAMITGENE	2028	X93513	<i>C. glutamicum</i> amt gene.	<i>Corynebacterium glutamicum</i>	100,000	29-MAY-1996	
		GB_BA1: CORPEPC	4885	M25819	<i>C. glutamicum</i> phosphoenolpyruvate carboxylase gene, complete cds.	<i>Corynebacterium glutamicum</i>	100,000	15-DEC-1995	
rxa02268	1023	GB_PL2: AF087130	3478	AF087130	<i>Neurospora crassa</i> siderophore regulation protein (src) gene, complete cds.	<i>Neurospora crassa</i>	39,268	22-OCT-1998	
		GB_EST30: AI663709	408	AI663709	IMAGE: 1449010 5' similar to TR: O75585 O75585 MITOGEN- AND STRESS-ACTIVATED PROTEIN KINASE-2.; mRNA sequence.	<i>Mus musculus</i>	41,523	10-MAY-1999	
		GB_RO: AF074714	3120	AF074714	<i>Mus musculus</i> mitogen- and stress-activated protein kinase-2 (mMSK2) mRNA, complete cds.	<i>Mus musculus</i>	38,347	24-OCT-1998	
	1095	GB_GSS4: AQ742825	847	AQ742825	HS_5482_B2_A04_T7A RPC1-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 1058 Col = 8 Row = B, genomic survey sequence.	<i>Homo sapiens</i>	37,703	16-Jul-99	
rxa02309		GB_HTG3: AC009293	162944	AC009293	<i>Homo sapiens</i> chromosome 18 clone 53_I_06 map 18, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.	<i>Homo sapiens</i>	37,006	13-Aug-99	
		GB_HTG3: AC009293	162944	AC009293	<i>Homo sapiens</i> chromosome 18 clone 53_I_06 map 18, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.	<i>Homo sapiens</i>	37,006	13-Aug-99	
	1173	GB_BA1: MTY25D10	40838	Z95558	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 28/162.	<i>Mycobacterium tuberculosis</i>	52,344	17-Jun-98	
		GB_BA1: MSGY224	40051	AD000004	<i>Mycobacterium tuberculosis</i> sequence from clone y224.	<i>Mycobacterium tuberculosis</i>	52,344	03-DEC-1996	
rxa02310		GB_HTG2: AC007163	186618	AC007163	<i>Homo sapiens</i> clone NH0091M05, *** SEQUENCING IN PROGRESS ***, 1 unordered pieces.	<i>Homo sapiens</i>	37,263	23-Apr-99	
		GB_BA1: MTY25D10	40838	Z95558	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 28/162.	<i>Mycobacterium tuberculosis</i>	36,861	17-Jun-98	
		GB_BA1: MSGY224	40051	AD000004	<i>Mycobacterium tuberculosis</i> sequence from clone y224.	<i>Mycobacterium tuberculosis</i>	36,861	03-DEC-1996	
	1752	GB_PR3: HS279N11	169998	Z98255	Human DNA sequence from PAC 279N11 on chromosome Xq11.2-13.3.	<i>Homo sapiens</i>	34,516	23-Nov-99	
rxa02321		GB_BA1: AB018531	4961	AB018531	<i>Corynebacterium glutamicum</i> dsrR1 and dsrR2 genes, complete cds.	<i>Corynebacterium glutamicum</i>	99,030	19-OCT-1998	
		GB_PAT: EI7019	4961	EI7019	<i>Brevibacterium lactofermentum</i> dsrR and dsrR2 genes.	<i>Corynebacterium glutamicum</i>	98,973	28-Jul-99	
		GB_BA1: AB018530	2855	AB018530	<i>Corynebacterium glutamicum</i> dsrR gene, complete cds.	<i>Corynebacterium glutamicum</i>	99,030	19-OCT-1998	
	1896	GB_BA1: CGU35023	3195	U35023	<i>Corynebacterium glutamicum</i> thiosulfate sulfurtransferase (thtR) gene, partial cds, acyl CoA carboxylase (accBC) gene, complete cds.	<i>Corynebacterium glutamicum</i>	99,947	16-Jan-97	
rxa02335		GB_BA1: U00012	33312	U00012	<i>Mycobacterium leprae</i> cosmid B1308.	<i>Mycobacterium leprae</i>	40,247	30-Jan-96	
		GB_BA1: MTCY71	42729	Z92771	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 141/162.	<i>Mycobacterium tuberculosis</i>	67,568	10-Feb-99	

TABLE 4-continued

ALIGNMENT RESULTS								
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxsa02364	750	GB_BA1: AP000006	319000	AP000006	<i>Pyrococcus horikoshii</i> OT3 genomic DNA, 1166001-1485000 nt. position (6/7).	<i>Pyrococcus horikoshii</i>	36,130	8-Feb-99
	2010	GB_BA1: AP000006	319000	AP000006	<i>Pyrococcus horikoshii</i> OT3 genomic DNA, 1166001-1485000 nt. position (6/7).	<i>Pyrococcus horikoshii</i>	34,543	8-Feb-99
		GB_HTG3: AC011461	100974	AC011461	<i>Homo sapiens</i> chromosome 19 clone CIT-HSPC_429L19, *** SEQUENCING IN PROGRESS ***; 4 ordered pieces.	<i>Homo sapiens</i>	36,138	07-OCT-1999
		GB_HTG3: AC011461	100974	AC011461	<i>Homo sapiens</i> chromosome 19 clone CIT-HSPC_429L19, *** SEQUENCING IN PROGRESS ***; 4 ordered pieces.	<i>Homo sapiens</i>	36,138	07-OCT-1999
rxsa02397	1119	GB_EST21: AA992021	279	AA992021	ot36c01.s1 Soares_testis_NHT <i>Homo sapiens</i> cDNA clone IMAGE: 1618848 3', mRNA sequence.	<i>Homo sapiens</i>	41,219	3-Jun-98
		GB_HTG4: AC009273	76175	AC009273	<i>Arabidopsis thaliana</i> chromosome 1 clone TIN6, *** SEQUENCING IN PROGRESS	<i>Arabidopsis thaliana</i>	38,566	12-OCT-1999
		GB_HTG4: AC009273	76175	AC009273	***, 2 ordered pieces.	<i>Arabidopsis thaliana</i>	38,566	12-OCT-1999
		GB_BA1: D90826	19493	D90826	<i>E. coli</i> genomic DNA, Kohara clone #335(40.9-41.3 min.).	<i>Escherichia coli</i>	39,600	21-MAR-1997
rxsa02424	723	GB_EST13: AA334108	275	AA334108	EST38262 Embryo, 9 week <i>Homo sapiens</i> cDNA 5' end, mRNA sequence.	<i>Homo sapiens</i>	38,603	21-Apr-97
	1656	GB_PR3: AC005224	166687	AC005224	<i>Homo sapiens</i> chromosome 17, clone hRPK.214_O_1, complete sequence.	<i>Homo sapiens</i>	36,111	14-Aug-98
		GB_PR3: AC005224	166687	AC005224	<i>Homo sapiens</i> chromosome 17, clone hRPK.214_O_1, complete sequence.	<i>Homo sapiens</i>	33,427	14-Aug-98
		GB_PAT: A06664	1350	A06664	<i>B. stearothermophilus</i> lct gene.	<i>Bacillus</i>	39,936	29-Jul-93
rxsa02487	1827	GB_PAT: A04115	1361	A04115	<i>B. stearothermophilus</i> recombinant lct gene.	<i>stearothermophilus</i>	40,042	17-Feb-97
		GB_BA1: BACLDHL	1361	M14788	<i>B. stearothermophilus</i> lct gene encoding L-lactate dehydrogenase, complete cds.	<i>synthetic construct</i>	40,338	26-Apr-93
		GB_BA2: AF007101	32870	AF007101	<i>Streptomyces hygroscopicus</i> putative pteridine-dependent dioxygenase, PKS modules 1, 2, 3 and 4, and putative regulatory protein genes, complete cds and putative hydroxylase gene, partial cds.	<i>stearothermophilus</i>	43,298	13-Jan-98
		GB_BA1: MTCI364	29540	Z93777	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 52/162.	<i>hygroscopicus</i>	44,352	17-Jun-98
rxsa02511	780	GB_BA2: AF119621	15986	AF119621	<i>Pseudomonas abietaniphila</i> BKME-9 DltI (dltI), dioxygenase DltA oxygenase component small subunit (dltA2), dioxygenase DltA oxygenase component large subunit (dltA1), DltH (dltH), DltG (dltG), DltF (dltF), DltR (dltR), DltE (dltE), DltD (dltD), aromatic diterpenoid extradiol ring-cleavage dioxygenase (dltC), DltB (dltB), and dioxygenase DltA ferredoxin component (dltA3) genes, complete cds; and unknown genes.	<i>Mycobacterium tuberculosis</i>	43,611	28-Apr-99
		GB_PR4: AC002470	235395	AC002470	<i>Homo sapiens</i> Chromosome 22q11.2 BAC Clone b13516 in BCRL2-GGT Region, complete sequence.	<i>Homo sapiens</i>	37,971	30-Nov-99
		GB_PR4: AC002472	147100	AC002472	<i>Homo sapiens</i> Chromosome 22q11.2 PAC Clone p_n5 in BCRL2-GGT Region, complete sequence.	<i>Homo sapiens</i>	38,239	13-Sep-99
		GB_EST34: AI806938	118	AI806938	wf24b07.x1 Soares_NFL_T_GBC_S1 <i>Homo sapiens</i> cDNA clone IMAGE: 2356501 3' similar to SW: PLZF_HUMAN Q05516 ZINC FINGER PROTEIN PLZF; mRNA sequence.	<i>Homo sapiens</i>	38,983	7-Jul-99
rxsa02512	1086	GB_BA1: MTCY1A10	25949	Z95387	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 117/162.	<i>Mycobacterium tuberculosis</i>	37,407	17-Jun-98

TABLE 4-continued

ALIGNMENT RESULTS								
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit
rxsa02527	1452	GB_BA1: MLC1581	36225	Z96801	<i>Mycobacterium leprae</i> cosmid L581.	<i>Mycobacterium leprae</i>	43,193	24-Jun-97
		GB_OV: GGU43396	2738	U43396	<i>Gallus gallus</i> tropomyosin receptor kinase A (trkA) mRNA, complete cds.	<i>Gallus gallus</i>	38,789	18-Jan-96
		GB_BA2: AF008220	220060	AF008220	<i>Bacillus subtilis</i> rrmB-dnaB genomic region.	<i>Bacillus subtilis</i>	37,395	4-Feb-98
		GB_BA2: AF008220	220060	AF008220	<i>Bacillus subtilis</i> rrmB-dnaB genomic region.	<i>Bacillus subtilis</i>	36,218	4-Feb-98
		GB_HTG2: AC005861	112369	AC005861	<i>Arabidopsis thaliana</i> clone F23B24, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.	<i>Arabidopsis thaliana</i>	38,407	29-Apr-99
rxsa02547	2262	GB_PL1: AB006530	7344	AB006530	<i>Citrusillus lanatus</i> Sat gene for serine acetyltransferase, complete cds and 5'-flanking region.	<i>Citrusillus lanatus</i>	35,449	20-Aug-97
rxsa02566	1332	GB_PL1: CNASA	5729	D85624	<i>Citrusillus vulgaris</i> serine acetyltransferase (Sat) DNA, complete cds.	<i>Citrusillus lanatus</i>	35,449	6-Feb-99
		GB_PL1: AB006530	7344	AB006530	<i>Citrusillus lanatus</i> Sat gene for serine acetyltransferase, complete cds and 5'-flanking region.	<i>Citrusillus lanatus</i>	34,646	20-Aug-97
		GB_EST32: A1727189	619	A1727189	BNLGH17498 Six-day Cotton fiber <i>Gossypium hirsutum</i> cDNA 5' similar to (AB020715) KIAA0908 protein [<i>Homo sapiens</i>], mRNA sequence.	<i>Gossypium hirsutum</i>	35,099	11-Jun-99
		GB_BA1: CGPUTP	3791	Y09163	<i>C. glutamicum</i> putP gene.	<i>Corynebacterium glutamicum</i>	38,562	8-Sep-97
		GB_PL2: SPAC13G6	33481	Z54308	<i>S. pombe</i> chromosome I cosmid c13G6.	<i>Schizosaccharomyces pombe</i>	35,774	18-OCT-1999
rxsa02571	1152	GB_BA1: CGU43535	2531	U43535	<i>Corynebacterium glutamicum</i> multidrug resistance protein (cmr) gene, complete cds.	<i>Corynebacterium glutamicum</i>	41,872	9-Apr-97
rxsa02578	1227	GB_EST35: A1857385	488	A1857385	w155e03.x1 NCL_CGAP_Brn25 <i>Homo sapiens</i> cDNA clone IMAGE: 2428828 3', mRNA sequence.	<i>Homo sapiens</i>	39,139	26-Aug-99
		GB_BA1: CGU43535	2531	U43535	<i>Corynebacterium glutamicum</i> multidrug resistance protein (cmr) gene, complete cds.	<i>Corynebacterium glutamicum</i>	38,552	9-Apr-97
		GB_PL1: AB016871	79109	AB016871	<i>Arabidopsis thaliana</i> genomic DNA, chromosome 5, TAC clone: K16122, complete sequence.	<i>Arabidopsis thaliana</i>	34,213	20-Nov-99
		GB_PL1: AB025602	55790	AB025602	<i>Arabidopsis thaliana</i> genomic DNA, chromosome 5, BAC clone: F14A1, complete sequence.	<i>Arabidopsis thaliana</i>	36,461	20-Nov-99
		GB_IN1: CELF36H9	35985	AF016668	<i>Caenorhabditis elegans</i> cosmid F36H9.	<i>Caenorhabditis elegans</i>	35,977	8-Aug-97
rxsa02581	1983	GB_BA1: MTV005	37840	AL010186	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 51/162.	<i>Mycobacterium tuberculosis</i>	38,517	17-Jun-98
rxsa02582	4953	GB_BA1: MTV005	37840	AL010186	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 51/162.	<i>Mycobacterium tuberculosis</i>	39,173	17-Jun-98
		GB_BA1: MTV026	23740	AL022076	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 157/162.	<i>Mycobacterium tuberculosis</i>	38,548	24-Jun-99
		GB_BA1: MTCY338	29372	Z74697	<i>Mycobacterium tuberculosis</i> H37Rv complete genome; segment 127/162.	<i>Mycobacterium tuberculosis</i>	46,263	17-Jun-98
		GB_BA1: SEERYABS	20444	X62569	<i>S. erythraea</i> eryA gene for 6-deoxyerythronolide B synthase II & III.	<i>Saccharopolyspora erythraea</i>	45,053	28-Feb-92
		GB_BA2: AF113605	1593	AF113605	<i>Streptomyces coelicolor</i> propionyl-CoA carboxylase complex B subunit (pcbB) gene, complete cds.	<i>Streptomyces coelicolor</i>	58,397	08-DEC-1999
rxsa02583	1671	GB_BA1: SC1C2	42210	AL031124	<i>Streptomyces coelicolor</i> cosmid 1C2.	<i>Streptomyces coelicolor</i>	52,916	15-Jan-99
rxsa02599	600	GB_BA1: AB018531	4961	AB018531	<i>Corynebacterium glutamicum</i> dsrR1 and dsrR2 genes, complete cds.	<i>Corynebacterium glutamicum</i>	58,809	19-OCT-1998
		GB_BA1: AEMML	2585	X99639	<i>Ralstonia eutropha</i> mmH, mmI & mmJ genes.	<i>Ralstonia eutropha</i>	35,264	22-Jan-98
		GB_EST15: AA508926	422	AA508926	MBAFCW1C08T3 <i>Brugia malayi</i> adult female cDNA (SAW96MLW-BmAF) <i>Brugia malayi</i> cDNA clone AFCW1C08 5', mRNA sequence.	<i>Brugia malayi</i>	43,377	8-Jul-97

TABLE 4-continued

ALIGNMENT RESULTS									
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit	% homology (GAP)	Date of Deposit	
rxa02634	1734	GB_BA1: AEMML	2585	X99639	<i>Ralstonia eutropha</i> mmH, mmL & mmJ genes.	<i>Ralstonia eutropha</i>	41,148	22-Jan-98	
		GB_BA1: SYNPOO	1964	X17439	<i>Synechocystis</i> ndhC, psbG genes for NDH-C, PSII-G and ORF157.	<i>Synechocystis</i> PCC6803	38,145	10-Feb-99	
		GB_GSS9: AQ101527	184	AQ101527	HS_2265_A1_E11 MF CIT Approved Human Genomic Sperm Library D <i>Homo sapiens</i> genomic clone Plate = 2265 Col = 21 Row = 1, genomic survey sequence.	<i>Homo sapiens</i>	38,798	27-Aug-98	
rxa02638	999	GB_IN1: MNE133341	399	AJ133341	<i>Melarthaphe neritoides</i> partial caM gene, exons 1-2.	<i>Melarthaphe neritoides</i>	39,098	2-Jun-99	
		GB_BA2: AE001756	10938	AE001756	<i>Thermotoga maritima</i> section 68 of 136 of the complete genome.	<i>Thermotoga maritima</i>	40,104	2-Jun-99	
		GB_GSS12: AQ423878	689	AQ423878	CITBI-EL-2575E20.TF CITBI-E1 <i>Homo sapiens</i> genomic clone 2575E20, genomic survey sequence.	<i>Homo sapiens</i>	36,451	23-MAR-1999	
rxa02659	335	GB_HTG2: AC006765	274498	AC006765	<i>Caenorhabditis elegans</i> clone Y43H11, *** SEQUENCING IN PROGRESS***, 7 unordered pieces.	<i>Caenorhabditis elegans</i>	39,072	23-Feb-99	
		GB_EST36: AI900317	436	AI900317	sg04a02.y1 Gm-cl012 <i>Glycine max</i> cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl012-1155 5' similar to SW: PRS6_SOLITU P54778 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.; mRNA sequence.	<i>Glycine max</i>	41,566	06-DEC-1999	
		GB_GSS12: AQ342831	683	AQ342831	RPC11-122K17.TJ RPC1-11 <i>Homo sapiens</i> genomic clone RPC1-11-122K17, genomic survey sequence.	<i>Homo sapiens</i>	34,762	07-MAY-1999	
rxa02676	1512	GB_EST36: AI900856	779	AI900856	sb95c11.y1 Gm-cl012 <i>Glycine max</i> cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl012-429 5' similar to SW: PRS6_SOLITU P54778 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.; mRNA sequence.	<i>Glycine max</i>	39,063	06-DEC-1999	
		GB_IN2: CELB0213	39134	AF039050	<i>Caenorhabditis elegans</i> cosmid B0213.	<i>Caenorhabditis elegans</i>	35,814	2-Jun-99	
		GB_GSS1: CNS00PZB	364	AL085157	<i>Arabidopsis thaliana</i> genome survey sequence SP6 end of BAC F10D11 of IGF library from strain Columbia of <i>Arabidopsis thaliana</i> , genomic survey sequence.	<i>Arabidopsis thaliana</i>	38,462	28-Jun-99	
rxa02677	882	GB_RO: RNITPR2R	10708	X61677	Rat ITPR2 gene for type 2 inositol triphosphate receptor.	<i>Rattus norvegicus</i>	37,543	21-OCT-1991	
		GB_RO: D89728	5002	D89728	<i>Mus musculus</i> mRNA for LOK, complete cds.	<i>Mus musculus</i>	38,829	7-Feb-99	
		GB_GSS8: AQ062004	362	AQ062004	CIT-HSP-2346014, TR CIT-HSP <i>Homo sapiens</i> genomic clone 2346014, genomic survey sequence.	<i>Homo sapiens</i>	36,565	31-Jul-98	
rxa02691	930	GB_GSS14: AQ555818	462	AQ555818	HS_5230_B1_G06_SP6E RPC1-11 Human Male BAC Library <i>Homo sapiens</i> genomic clone Plate = 806 Col = 11 Row = N, genomic survey sequence.	<i>Homo sapiens</i>	36,534	29-MAY-1999	
		GB_IN1: DME9736	7411	AJ009736	<i>Drosophila melanogaster</i> idexif retroelement: gag, pol and env genes, partial.	<i>Drosophila melanogaster</i>	36,522	19-Jan-99	
		GB_PR4: AC004801	193561	AC004801	<i>Homo sapiens</i> 12q13.1 PAC RPC11-228P16 (Roswell Park Cancer Institute Human PAC Library) complete sequence.	<i>Homo sapiens</i>	39,341	2-Feb-99	
rxa02718	1170	GB_PR4: AC004801	193561	AC004801	<i>Homo sapiens</i> 12q13.1 PAC RPC11-228P16 (Roswell Park Cancer Institute Human PAC Library) complete sequence.	<i>Homo sapiens</i>	37,037	2-Feb-99	
		GB_EST34: AV132028	258	AV132028	AV132028 <i>Mus musculus</i> C57BL/6J 11-day embryo <i>Mus musculus</i> cDNA clone 2700087F01, mRNA sequence.	<i>Mus musculus</i>	43,529	1-Jul-99	
		GB_GSS10: AQ240654	452	AQ240654	CIT-HSP-2385D24.TFB.1 CIT-HSP <i>Homo sapiens</i> genomic clone 2385D24, genomic survey sequence.	<i>Homo sapiens</i>	40,044	30-Sep-98	
rxa02749	999	GB_GSS11: AQ309500	576	AQ309500	CIT-HSP-2384D24.TFD CIT-HSP <i>Homo sapiens</i> genomic clone 2384D24, genomic survey sequence.	<i>Homo sapiens</i>	38,869	22-DEC-1998	
		GB_BA2: AF086791	37867	AF086791	<i>Zymomonas mobilis</i> strain ZM4 clone 67E10 carbamoylphosphate synthetase small subunit (carA), carbamoylphosphate synthetase large subunit (carB), transcription elongation factor (greA), enolase (eno), pyruvate dehydrogenase alpha subunit (pdhA), pyruvate dehydrogenase beta subunit (pdhB), ribonuclease H (mh), homoserine kinase homolog, alcohol dehydrogenase II (adhB), and excinuclease ABC subunit A (uvrA) genes, complete cds; and unknown genes.	<i>Zymomonas mobilis</i>	39,024	4-Nov-98	

TABLE 4-continued

ALIGNMENT RESULTS						
ID #	length (NT)	Genbank Hit	Length	Accession	Name of Genbank Hit	Source of Genbank Hit
rxs02767	906	GB_BA1: SYCSLRB	146271	D64000	<i>Synechocystis</i> sp. PCC6803 complete genome, 19/27, 2392729-2538999.	<i>Synechocystis</i> sp.
		GB_BA2: AE001306	13316	AE001306	<i>Chlamydia trachomatis</i> section 33 of 87 of the complete genome.	<i>Chlamydia trachomatis</i>
		GB_BA2: AF126953	1638	AF126953	<i>Corynebacterium glutamicum</i> cystathionine gamma-synthase (metB) gene, complete cds.	<i>Corynebacterium glutamicum</i>
rxs02792	876	GB_BA1: SC15	6661	AL079332	<i>Streptomyces coelicolor</i> cosmid I5.	<i>Streptomyces coelicolor</i>
		GB_PR3: HS90L6	190837	Z97353	Human DNA sequence from clone 90L.6 on chromosome 22q11.21-11.23. Contains an RPL15 (60S Ribosomal Protein L15) pseudogene, ESTs, STSs and GSSs, complete sequence.	<i>Homo sapiens</i>
		GB_BA2: AF099015	5000	AF099015	<i>Streptomyces coelicolor</i> strain A3(2) integrase (int), Fe-containing superoxide dismutase II (sofF2), Fe uptake system permease (ftrF), and Fe uptake system integral membrane protein (ftrD) genes, complete cds.	<i>Streptomyces coelicolor</i>
rxs02794	1197	GB_BA1: ECOUW93	338534	U14003	<i>Escherichia coli</i> K-12 chromosomal region from 92.8 to 90.1 minutes.	<i>Escherichia coli</i>
		GB_HTG3: AC011361	186148	AC011361	<i>Homo sapiens</i> chromosome 5 clone CIT-HSPC_482N19, *** SEQUENCING IN PROGRESS ***. 69 unordered pieces.	<i>Homo sapiens</i>
		GB_PR4: AC005998	96556	AC005998	<i>Homo sapiens</i> clone DJ0622E21, complete sequence.	<i>Homo sapiens</i>
rxs02809	375	GB_PR4: AC006008	57554	AC006008	<i>Homo sapiens</i> clone DJ0820A21, complete sequence.	<i>Homo sapiens</i>
		GB_PR3: HSDJ73H14	95556	AL080272	Human DNA sequence from clone 73H14 on chromosome Xq26.3-28, complete sequence.	<i>Homo sapiens</i>
		GB_RO: MUSSPCTLT	3172	M22527	Mouse cytotoxic T lymphocyte-specific serine protease CCP11 gene, complete cds.	<i>Mus musculus</i>
rxs02811	484	GB_RO: MUSGRC	894	M18459	Mouse granzyme C serine esterase mRNA, complete cds.	<i>Mus musculus</i>
		GB_RO: RNU57062	880	U57062	<i>Rattus norvegicus</i> natural killer cell protease 4 (RNKP-4) mRNA, complete cds.	<i>Rattus norvegicus</i>
		GB_GSS6: AQ832862	476	AQ832862	HS_5261_A2_E10_SP6E RPC1-11 Human Male BAC Library Homo sapiens genomic clone Plate = 837 Col = 20 Row = I, genomic survey sequence.	<i>Homo sapiens</i>
rxs02836	678	GB_GSS5: AQ784593	515	AQ784593	HS_3248_A2_F02_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate = 3248 Col = 4 Row = K, genomic survey sequence.	<i>Homo sapiens</i>
		GB_GSS13: AQ473140	397	AQ473140	CITBI-EI-2589G6.TF CITBI-EI <i>Homo sapiens</i> genomic clone 2589G6, genomic survey sequence.	<i>Homo sapiens</i>
		GB_EST18: AA696785	316	AA696785	GM08392.5prime GM <i>Drosophila melanogaster</i> ovary BlueScript <i>Drosophila melanogaster</i> cDNA clone GM08392 5prime, mRNA sequence.	<i>Drosophila melanogaster</i>
rxs03212	1452	GB_EST18: AA696785	316	AA696785	GM08392.5prime GM <i>Drosophila melanogaster</i> ovary BlueScript <i>Drosophila melanogaster</i> cDNA clone GM08392 5prime, mRNA sequence.	<i>Drosophila melanogaster</i>
		GB_BA1: CGBETPGEN	2339	X93514	<i>C. glutamicum</i> betP gene.	<i>Corynebacterium glutamicum</i>
		GB_BA1: SC5F2A	40105	AL049587	<i>Streptomyces coelicolor</i> cosmid 5F2A.	<i>Streptomyces coelicolor</i>
rxs03220	725	GB_BA2: AF008220	220060	AF008220	<i>Bacillus subtilis</i> rmb-dnaB genomic region.	<i>Bacillus subtilis</i>
		GB_PL1: CKHUP2	2353	X66855	<i>C. kessleri</i> HUP2 mRNA.	<i>Chlorella kessleri</i>
		GB_EST18: AW048153	383	AW048153	UI-M-BH1-alq-h-05-0-UI.s1 NIH_BMAP_M_S2 <i>Mus musculus</i> cDNA clone UI-M-BH1-alq-h-05-0-UI 3', mRNA sequence.	<i>Mus musculus</i>
rxs03220	725	GB_PL1: CKHUP2	2353	X66855	<i>C. kessleri</i> HUP2 mRNA.	<i>Chlorella kessleri</i>
		GB_PL1: CKHUP2	2353	X66855	<i>C. kessleri</i> HUP2 mRNA.	<i>Chlorella kessleri</i>
		GB_PL1: CKHUP2	2353	X66855	<i>C. kessleri</i> HUP2 mRNA.	<i>Chlorella kessleri</i>

[0172]

SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (<http://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20070111230A1>). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of

- a) an isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:77, or a complement thereof;
- b) an isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:78, or a complement thereof;
- c) an isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:78, or a complement thereof;
- d) an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 50% identical to the entire nucleotide sequence of SEQ ID NO:77, or a complement thereof; and
- e) an isolated nucleic acid molecule comprising a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:77, or a complement thereof.

2. An isolated nucleic acid molecule comprising the nucleic acid molecule of claim 1 and a nucleotide sequence encoding a heterologous polypeptide.

3. A vector comprising the nucleic acid molecule of claim 1.

4. The vector of claim 3, which is an expression vector.

5. A host cell transfected with the expression vector of claim 4.

6. The host cell of claim 5, wherein said cell is a microorganism.

7. The host cell of claim 6, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

8. A method of producing a polypeptide comprising culturing the host cell of claim 5 in an appropriate culture medium to, thereby, produce the polypeptide.

9. A method for producing a fine chemical, comprising culturing the cell of claim 5 such that the fine chemical is produced.

10. The method of claim 9, wherein said method further comprises the step of recovering the fine chemical from said culture.

11. The method of claim 9, wherein said cell belongs to the genus *Corynebacterium* or *Brevibacterium*.

12. The method of claim 9, wherein said cell is selected from the group consisting of *Corynebacterium glutamicum*, *Corynebacterium herculis*, *Corynebacterium lilium*,

Corynebacterium acetoacidophilum, *Corynebacterium acetoglutamicum*, *Corynebacterium acetophilum*, *Corynebacterium ammoniagenes*, *Corynebacterium fujiokense*, *Corynebacterium nitrilophilus*, *Brevibacterium ammoniagenes*, *Brevibacterium butanicum*, *Brevibacterium divaricatum*, *Brevibacterium flavum*, *Brevibacterium heilii*, *Brevibacterium ketoglutamicum*, *Brevibacterium ketosoreductum*, *Brevibacterium lactofermentum*, *Brevibacterium linens*, *Brevibacterium paraffinolyticum*, and those strains set forth in Table 3.

13. The method of claim 9, wherein expression of the nucleic acid molecule from said vector results in modulation of production of said fine chemical.

14. The method of claim 9, wherein said fine chemical is selected from the group consisting of organic acids, proteinogenic and nonproteinogenic amino acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids, saturated and unsaturated fatty acids, diols, carbohydrates, aromatic compounds, vitamins, cofactors, polyketides, and enzymes.

15. The method of claim 9, wherein said fine chemical is an amino acid selected from the group consisting of lysine, glutamate, glutamine, alanine, aspartate, glycine, serine, threonine, methionine, cysteine, valine, leucine, isoleucine, arginine, proline, histidine, tyrosine, phenylalanine, and tryptophan.

16. An isolated polypeptide selected from the group consisting of

- a) an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:78;
- b) an isolated polypeptide comprising a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:78;
- c) an isolated polypeptide which is encoded by a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:77;
- d) an isolated polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% identical to the entire nucleotide sequence of SEQ ID NO:77;
- e) an isolated polypeptide comprising an amino acid sequence which is at least 50% identical to the entire amino acid sequence of SEQ ID NO:78; and
- f) an isolated polypeptide comprising a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:78, wherein said polypeptide fragment

maintains a biological activity of the polypeptide comprising the amino sequence.

17. The isolated polypeptide of claim 16, further comprising heterologous amino acid sequences.

18. A method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject, comprising detecting the presence of at least one of the nucleic acid molecules of claim 1, thereby diagnosing the presence or activity of *Corynebacterium diphtheriae* in the subject.

19. A method for diagnosing the presence or activity of *Corynebacterium diphtheriae* in a subject, comprising detecting the presence of at least one of the polypeptide molecules of claim 16, thereby diagnosing the presence or activity of *Corynebacterium diphtheriae* in the subject.

20. A host cell comprising a nucleic acid molecule selected from the group consisting of

- a) the nucleic acid molecule of SEQ ID NO:77, wherein the nucleic acid molecule is disrupted by at least one technique selected from the group consisting of a point

mutation, a truncation, an inversion, a deletion, an addition, a substitution and homologous recombination;

- b) the nucleic acid molecule of SEQ ID NO:77, wherein the nucleic acid molecule comprises one or more nucleic acid modifications as compared to the sequence of SEQ ID NO:77, wherein the modification is selected from the group consisting of a point mutation, a truncation, an inversion, a deletion, an addition and a substitution; and

- c) the nucleic acid molecule of SEQ ID NO:77, wherein the regulatory region of the nucleic acid molecule is modified relative to the wild-type regulatory region of the molecule by at least one technique selected from the group consisting of a point mutation, a truncation, an inversion, a deletion, an addition, a substitution and homologous recombination.

* * * * *