(86) Date de dépôt PCT/PCT Filing Date: 2017/03/15
(87) Date publication PCT/PCT Publication Date: 2017/09/21
(85) Entrée phase nationale/National Entry: 2018/09/07
(86) N° demande PCT/PCT Application No.: US 2017/022539
(87) N° publication PCT/PCT Publication No.: 2017/161005
(30) Priorité/Priority: 2016/03/16 (US62/309,402)

(54) Titre : PRODUCTION DE PROTEINES DANS DES LABYRINTHULOMYCETES
(54) Title: PRODUCTION OF PROTEINS IN LABYRINTHULOMYCETES

(57) Abrégé/Abstract:
The present invention provides recombinant host cells that produce proteins or therapeutic proteins, and nucleic acid constructs for producing the cells. The cells have nucleic acid constructs that encode a heterologous protein, for example an antibody. The nucleic acid constructs also can have a functional signal sequence that directs the secretion of the protein from the cell. The signal sequence can be any functional signal sequence, and various signal sequences are disclosed herein. The invention also provides methods of producing the proteins.
Title: PRODUCTION OF PROTEINS IN LABYRINTHULOMYCETES

Abstract: The present invention provides recombinant host cells that produce proteins or therapeutic proteins, and nucleic acid constructs for producing the cells. The cells have nucleic acid constructs that encode a heterologous protein, for example an antibody. The nucleic acid constructs also can have a functional signal sequence that directs the secretion of the protein from the cell. The signal sequence can be any functional signal sequence, and various signal sequences are disclosed herein. The invention also provides methods of producing the proteins.
PRODUCTION OF PROTEINS IN LABYRINTHULOMYCETES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. provisional application serial number 62/309,402, filed March 16, 2016, which is hereby incorporated by reference in its entirety, including all Tables, Figures, and Claims.

INCORPORATION OF SEQUENCE LISTING

[0002] The material in the accompanying sequence listing is hereby incorporated by reference into this application. The accompanying sequence listing text file, name SGI2040_IWO_Sequence_Listing.txt, was created on March 13, 2017, and is 47 kb. The file can be assessed using Microsoft Word on a computer that uses Windows OS.

FIELD OF THE INVENTION

[0003] The invention involves the production of proteins in *Labyrinthulomycetes* cells, methods of producing the proteins, and expression cassettes and other tools useful in the methods.

BACKGROUND OF THE INVENTION

[0004] Mammalian cell cultures are very useful for the production of certain products, such as viral vaccines, interferons, recombinant therapeutic proteins, and monoclonal antibodies. However, these cell cultures also carry inherent limitations, such as a very slow growth rate, low biomass density, the requirement for a complex media, and high operational costs.

[0005] Microbial expression systems have numerous advantages for the production of useful proteins. While certain microbial systems are useful for producing simple proteins, such microbial systems would need to be improved in terms of the quality and complexity of proteins that can be produced. The improvement of microbial cell specific productivities requires complex engineering, and substantial understanding and rewiring of the underlying microbial metabolism. An ideal strain would be genetically stable, have a high specific and volumetric productivity, form no by-products, and use a well-defined medium. These characteristics would allow for downstream processing with a limited number of steps.

[0006] *Labyrinthulomycetes* are robustly fermentable eukaryotic microalgae. These heterotrophic microorganisms are recognized for their industrial ability to consume sugar and store large amounts of cellular oils as triglycerides; the most commercially important is docosahexaenoic acid (DHA), an omega-3 polyunsaturated fatty acid (PUFA) that is a major component of fish oil. These organisms produce oils that can be used in human and animal
nutritional supplements, as well as for food fortification applications. These triglyceride oils can be produced in culture using inexpensive media. Because of these desirable qualities it would be advantageous to have recombinant *Labyrinthulomycetes* cells that are able to produce a variety of proteins or therapeutic proteins, including functional antibodies that can have both heavy and light chains.

**SUMMARY OF THE INVENTION**

[0007] The present invention provides recombinant *Labyrinthulomycetes* cells that produce proteins or therapeutic proteins, nucleic acid constructs such as expression cassettes for producing the cells, and methods of producing the cells and products of the cells. The cells have nucleic acid constructs that encode a heterologous protein, for example an antibody or other therapeutic protein. The nucleic acid constructs also can have a functional signal sequence that directs the secretion of the protein from the cell. The signal sequence can be any of various signal sequences disclosed herein, such as those nucleotide and amino acid sequences disclosed as SEQ ID NOs: 9-94.

[0008] In a first aspect the invention provides a recombinant *Labyrinthulomycetes* cell having an expression cassette that encodes a heterologous protein and a functional signal sequence that directs the secretion of the heterologous protein from the cell. The signal sequence can have at least 90% sequence identity to a sequence of at least 10 contiguous amino acids selected from any one of SEQ ID NO: 9-20 or 33-47 or 63-78, or can be encoded by a nucleic acid sequence having at least 90% sequence homology to SEQ ID NO: 21-32 or 48-62 or 79-94. In some embodiments the heterologous protein is a therapeutic protein, and can be a functional antibody, which can have at least one heavy chain and at least one light chain. The antibody can have an amino acid sequence selected from, but are not limited to, eculizumab, natalizumab, cetuximab, omalizumab, ustseinumab, panitumumab, denosumab, trastuzumab and adalimumab.

[0009] In some embodiments the heterologous protein can be a bispecific T cell engager. The heterologous protein can have a binding molecule that binds to an antigen, for example anti-CD19, anti-CD3, anti-CD38, and prostate-specific membrane antigen (PSMA). The expression cassette can encode a promoter active in a *Labyrinthulomycetes* and that regulates transcription of the heterologous protein. The promoter can be, for example, tubulina promoter (tubap) or translation elongation factor (TEFp) promoter. The expression cassette can also encode a terminator, for example pgk1t or eno2t. When the heterologous protein is an antibody it can have a light chain and a heavy chain that have at least 90% sequence
identity to an amino acid sequence of at least 100 contiguous amino acids of SEQ ID NO: 5 and 6 or SEQ ID NO: 7 and 8, respectively.

[0010] In another aspect the invention provides an expression cassette having at least one nucleotide sequence encoding a heterologous protein disclosed herein positioned to be regulated by the promoter, and at least one nucleotide sequence encoding a signal sequence having at least 80% or at least 85% or at least 90% sequence identity to a nucleotide sequence of at least 120 contiguous nucleotides selected from any one of SEQ ID NO: 21-32 or 48-62 or 79-94. The expression cassette can also have a promoter sequence and a terminator sequence functional in a Labyrinthulomycetes cell. In some embodiments the expression cassette is non-contiguous. The heterologous protein can be any antibody described herein, and can have a heavy chain and a light chain.

[0011] In another aspect the invention provides an antibody having a signal sequence functional in a Labyrinthulomycete, a light chain having at least 90% sequence identity to SEQ ID NO: 5, and a heavy chain having at least 90% sequence identity to SEQ ID NO: 6; or a light chain having at least 90% sequence identity to SEQ ID NO: 7 and a heavy chain having at least 90% sequence identity to SEQ ID NO: 8. The signal sequence can be any described herein.

[0012] In another aspect the invention provides an Aurantiochytrium cell having an expression cassette that encodes an exogenous protein and a functional signal sequence that directs the secretion of the heterologous protein from the cell. The heterologous protein can be any described herein. The cell can produce at least 5 mg/L of the heterologous protein in a 24 hour period. The cell can have an expression cassette that encodes a promoter active in Aurantiochytrium; a terminator sequence active in Aurantiochytrium; the heterologous protein can be operably linked to the promoter; and a signal sequence having at least 90% sequence identity to at least 20 contiguous amino acids of a sequence selected from any one of SEQ ID NO: 9-20 or 23-47 or 63-78 or any signal sequence disclosed herein.

[0013] In another aspect the invention provides a method for producing any heterologous protein described herein. The method includes steps of cultivating a Labyrinthulomycetes cell that secretes a heterologous protein and that comprises an expression cassette encoding the heterologous protein and further encoding a functional signal sequence that directs the secretion of the heterologous protein from the cell; the signal sequence can have at least 90% sequence identity to a sequence of at least 20 contiguous amino acids selected from any one of SEQ ID NO: 9-20 or 33-47 or 63-78 or can be any described herein; and a step of harvesting the heterologous protein. The heterologous protein can be any described herein.
The expression cassette can encode a promoter active in a *Labyrinthulomycetes* and that regulates transcription of the heterologous protein. The expression cassette can be any described herein, and the promoter and terminator can be any described herein. In one embodiment the promoter is tub-alpha-p or TEFp, and the terminator is pgk1t or enol2t.

[0014] In another embodiment the invention provides a therapeutic protein having an antibody molecule linked to a signal sequence. The therapeutic protein and the signal sequence can any described herein. In some embodiments the signal sequence is selected from SEQ ID NO: 9-20 or 33-47 or 63-78, and the antibody is anti-TNF-alpha or anti-HER2.

**DETAILED DESCRIPTION OF THE DRAWINGS**

[0015] **Figure 1** provides a schematic illustration of vectors and expression cassettes of the invention, identified as pCAB vectors with different expression cassettes (designated 018-030).

[0016] **Figure 2** is a general map of pCAB having homology arms and a unique BstZ171 site that is used to linearize these cassettes is also shown.

[0017] **Figure 3** is a bar graph showing the results of an HC-capture/LC-detect ELISA assay. Illustrated are the antibody titers of various clones. Clones 1-16 secrete anti-TNF-alpha antibody and clones 17-18 secrete anti-HER2 antibody.

[0018] **Figure 4** is a bar graph showing supernatant titers of secreted anti-TNF-alpha antibody light chain, using various signal peptides.

[0019] **Figure 5** is a graph illustrating calculated concentrations of anti-TNF-alpha antibody that binds TNF-alpha from the supernatants of *Labyrinthulomycetes* (here *Aurantiochytrium sp.* ) cultures. SEQ ID NO: 19 was the signal sequence for Hu-01, 03, 04, 10-19 were used as signal sequences.

[0020] **Figure 6** is a bar graph showing the relative binding of anti-HER2 antibody to recombinant HER2. Binding is indicated by the amount of signal/absorbance (Abs.) detected by an ELISA assay. Titers used in the assay are displayed in mg/L above each antibody sample tested. i6267 is conditioned media; Hu-11 is anti-TNF-alpha supernatant (o/n culture); RC3 is protein A purified anti-TNF-alpha; RC4-RC6 are protein A purified anti-HER2.

[0021] **Figure 7** provides a graph showing the antibody-dependent cell mediated cytotoxicity (ADCC) response to BT-474 cells targeted with purified anti-HER2.

[0022] **Figure 8** provides a graph showing a concentration-dependent effect of anti-TNF-alpha on TNF-alpha-induced killing of L929 cells. The EC50 (g/ml) value is 1.3e-007.
[0023] Figure 9 provides a plasmid map for a construct used for the expression of anti-CD19/anti-CD3 bi-specific T cell engager Ab.

[0024] Figure 10a provides a plasmid map for a construct used for expression of anti-PSMA/anti-CD3 bi-specific T cell engager antibody (PSMA-prostate-specific membrane antigen). Figure 10b provides a Western blot indicating secretion of anti-PSMA/anti-CD3 bi-specific T cell engager antibody. The arrow indicates the expected size for the bi-specific T cell engager antibody.

[0025] Figure 11 provides a Western blot indicating secretion of the bispecific antibody anti-II6RC/anti-IFNγ antibody. Bands corresponding to the heavy chain (HC), kappa light chain and lambda light chains are indicated.

[0026] Figure 12 provides a Western blot indicating secretion of the hybrid bispecific anti-CD38/anti-CD3 antibody (XENP13551). Bands corresponding to heavy chain 1 (HC1), heavy chain 2 (HC2), and light chain (LC) are indicated.

[0027] Figure 13 is a graphical illustration of relative fluorescence in supernatant of cultures of strains expressing TagBFP. The background fluorescence from FM2 media was subtracted.

[0028] Figure 14 provides a Western blot indicating secretion of K3-His6, indicated by the arrow.

DETAILED DESCRIPTION OF THE INVENTION

[0029] The recombinant cells, nucleic acid constructs, expression vectors, expression cassettes, and methods of the invention provide significant advantages over other tools and methods for producing proteins and other cell products. Labyrinthulomycetes cells have a very advantageous scalability that compares favorably to biomass titers achievable in yeast. Decreased fermentation cycle times carry the advantage of lower operational costs through lower material, energy, and labor costs per unit mass.

[0030] The cells of the invention offer a high fermentative capacity, which is desirable for cell-based production systems. The present invention leverages this advantage by providing methods for genetically manipulating the organisms to create a production platform for proteins or therapeutic proteins in one of the most highly productive, cost effective, and scalable fermentative eukaryotic microorganisms presently available.

[0031] In some embodiments the recombinant host cells or organisms of the invention can be any microorganism of the class Labyrinthulomycetes. While the classification of the Thraustochytrids and Labyrinthulids has evolved over the years, for the purposes of the present application, “labyrinthulomycetes” is a comprehensive term that includes
microorganisms of the orders Thraustochoytrid and Labyrinthulid, and includes (without limitation) the genera Althornia, Aplanocythrium, Aurantiocythrium, Botyrocythrium, Corallochytrium, Diplophyrids, Diplophys, Elina, Japonocythrium, Labyrinthula, Labyrinthuloides, Oblongicythrium, Pyrrhosorus, Schizocythrium, Thraustochoytrium, and Ulkenia. In some examples the microorganism is from a genus including, but not limited to, Thraustochoytrium, Labyrinthuloides, Japonocythrium, and Schizocythrium. Alternatively, a host Labyrinthulomycetes microorganism can be from a genus including, but not limited to, Aurantiochytrium, Oblongicythrium, and Ulkenia. Examples of suitable microbial species within the genera include, but are not limited to: any Schizocythrium species, including, but not limited to, Schizocythrium aggregatum, Schizocythrium limacinum, Schizocythrium minitum, Schizocythrium mangrovei, Schizocythrium marium, Schizocythrium octosporum, and any Aurantiochytrium species, any Thraustochoytrium species (including former Ulkenia species such as U. visurgensis, U. amoeboida, U. sarkariana, U. profunda, U. radiata, U. minuta and Ulkenia sp. BP-5601), and including Thraustochoytrium striatum, Thraustochoytrium aureum, Thraustochoytrium roseum; and any Japonocythrium species. Strains of Thraustochoytriales that may be particularly suitable for the presently disclosed invention include, but are not limited to: Schizocythrium sp. (S31) (ATCC 20888); Schizocythrium sp. (S8) (ATCC 20889); Schizocythrium sp. (LC-RM) (ATCC 18915); Schizocythrium sp. (SR21); Schizocythrium aggregatum (ATCC 28209); Schizocythrium limacinum (IFO 32693); Thraustochoytrium sp. 23B ATCC 20891; Thraustochoytrium striatum ATCC 24473; Thraustochoytrium aureum ATCC 34304; Thraustochoytrium roseum (ATCC 28210; and Japonocythrium sp. LI ATCC 28207. The recombinant host cells of the invention can also be a yeast cell, such as a yeast selected from the genus Saccharomyces or Candida, Pichia, Yarrowia, and Kluyveromyces lactis.

[0032] In some embodiments the recombinant host cells of the invention can comprise at least one genetic modification. The at least one genetic modification can be any manipulation of a host cell’s genome using the techniques of molecular biology or biotechnology. The genetic modification changes the genetic makeup of the cells, including the transfer of genes within and across species boundaries to produce improved or novel organisms. The at least one genetic modification can be, but is not limited to, the addition, deletion, modification, disruption, or optimization of a gene or portion of a gene. When one or more gene(s) is added it can be a heterologous gene. Any of the recombinant host cells of the invention can produce and comprise an antibody disclosed herein, which can be a functional and/or assembled antibody. In some embodiments the host cells overexpress the
therapeutic protein (e.g., an antibody or antibody fragment), for example according to the titers disclosed herein.

Nucleic Acid Constructs

[0033] As used herein, the term "construct" is intended to mean any recombinant nucleic acid molecule such as an expression cassette, vector, plasmid, cosmid, virus, autonomously replicating polynucleotide molecule, phage, or linear or circular, single-stranded or double-stranded, DNA or RNA polynucleotide molecule, derived from any source, capable of genomic integration or autonomous replication, comprising a nucleic acid molecule where one or more nucleic acid sequences has been linked in a functionally operative manner, i.e. operably linked. Constructs can also be non-naturally occurring (i.e., non-native).

[0034] The term "expression cassette" as used herein, refers to a nucleic acid construct that encodes a functional protein or functional RNA operably linked to expression control elements, such as a promoter, and optionally, any or a combination of other nucleic acid sequences that affect the transcription or translation of the gene, such as, but not limited to, a transcriptional terminator, a ribosome binding site, a splice site or splicing recognition sequence, an intron, an enhancer, a polyadenylation signal, an internal ribosome entry site, etc. The present invention provides numerous examples of expression cassettes useful for producing the cells and proteins of the invention, and for use in the methods of the invention. The expression cassettes can be comprised in any construct operable in the host cells being utilized. Generally an expression cassette will comprise a promoter, an open reading frame (ORF) encoding the heterologous protein of interest, and a terminator. Additional features can include 3’ and 5’ homology arms from genomic DNA of the host cell. These can be useful for inserting or integrating the expression cassette at a specific locus in the genome of the cell. Any of the components or features of the expression cassette can be active in any of the host cells described herein. The functional protein can be a heterologous protein. An expression cassette can also comprise a signal sequence for secretion of the expressed protein from the cell.

[0035] As used herein, "exogenous" with respect to a nucleic acid or gene indicates that the nucleic or gene has been introduced ("transformed") into an organism, microorganism, or cell by human intervention. Typically, such an exogenous nucleic acid is introduced into a cell or organism via a recombinant nucleic acid construct. An exogenous nucleic acid can be a sequence from one species introduced into another species, i.e., a heterologous nucleic acid. An exogenous nucleic acid can also be a sequence that is homologous to an organism (i.e., the nucleic acid sequence occurs naturally in that species or encodes a polypeptide that occurs
naturally in the host species) that has been isolated and subsequently reintroduced into cells of that organism. An exogenous nucleic acid that includes a homologous sequence can often be distinguished from the naturally-occurring sequence by the presence of non-natural sequences linked to the exogenous nucleic acid, e.g., non-native regulatory sequences flanking the homologous gene sequence in a recombinant nucleic acid construct. Alternatively or in addition, a stably transformed exogenous nucleic acid can be detected and/or distinguished from a native gene by its juxtaposition to sequences in the genome where it has integrated. Further, a nucleic acid is considered exogenous if it has been introduced into a progenitor of the cell, organism, or strain under consideration.

[0036] As used herein, "expression" refers to the process of converting genetic information of a polynucleotide into RNA through transcription, which is typically catalyzed by an enzyme, RNA polymerase, and, where the RNA encodes a polypeptide, into protein, through translation of mRNA on ribosomes to produce the encoded protein.

[0037] The term "gene" is used broadly to refer to any segment of nucleic acid molecule that encodes a protein or that can be transcribed into a functional RNA. Genes may include sequences that are transcribed but are not part of a final, mature, and/or functional RNA transcript, and genes that encode proteins may further comprise sequences that are transcribed but not translated, for example, 5' untranslated regions, 3' untranslated regions, introns, etc. Further, genes may optionally further comprise regulatory sequences required for their expression, and such sequences may be, for example, sequences that are not transcribed or translated. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

[0038] The term "heterologous" when used in reference to a polynucleotide, a gene, a nucleic acid, a peptide or polypeptide, a protein, or an enzyme, refers to a polynucleotide, gene, a nucleic acid, polypeptide, protein, or an enzyme that is not derived from the host species. For example, "heterologous gene" or "heterologous nucleic acid sequence" as used herein, refers to a gene or nucleic acid sequence from a different species than the species of the host organism it is introduced into. A heterologous protein or peptide can be one encoded by a heterologous nucleic acid. When referring to a gene regulatory sequence or to an auxiliary nucleic acid sequence used for manipulating expression of a gene sequence (e.g. a 5' untranslated region, 3' untranslated region, poly A addition sequence, intron sequence, splice site, ribosome binding site, internal ribosome entry sequence, genome homology region, recombination site, etc.) or to a nucleic acid sequence encoding a protein domain or protein
localization sequence, "heterologous" means that the regulatory or auxiliary sequence or sequence encoding a protein domain or localization sequence is from a different source than the gene with which the regulatory or auxiliary nucleic acid sequence or nucleic acid sequence encoding a protein domain or localization sequence is juxtaposed in a genome, chromosome, or episome. Thus, a promoter operably linked to a gene to which it is not operably linked to in its natural state (for example, in the genome of a non-genetically engineered organism) is referred to herein as a "heterologous promoter," even though the promoter may be derived from the same species (or, in some cases, the same organism) as the gene to which it is linked. Similarly, when referring to a protein localization sequence or protein domain of an engineered protein, "heterologous" means that the localization sequence or protein domain is derived from a protein different from that into which it is incorporated by genetic engineering.

[0039] The term "native" is used herein to refer to nucleic acid sequences or amino acid sequences as they naturally occur in the host. The term "non-native" is used herein to refer to nucleic acid sequences or amino acid sequences that do not occur naturally in the host, or are not configured as they are naturally configured in the host. A nucleic acid sequence or amino acid sequence that has been removed from a host cell, subjected to laboratory manipulation, and introduced or reintroduced into a host cell is considered "non-native." Synthetic or partially synthetic genes introduced into a host cell are "non-native." Non-native genes further include genes endogenous to the host microorganism operably linked to one or more heterologous regulatory sequences that have been recombined into the host genome, or genes endogenous to the host organism that are in a locus of the genome other than that where they naturally occur.

[0040] The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably herein, and refer to both RNA and DNA molecules, including nucleic acid molecules comprising cDNA, genomic DNA, synthetic DNA, and DNA or RNA molecules containing nucleic acid analogs. Nucleic acid molecules can have any three-dimensional structure. A nucleic acid molecule can be double-stranded or single-stranded (e.g., a sense strand or an antisense strand). Non-limiting examples of nucleic acid molecules include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, siRNA, micro-RNA, tracrRNAs, crRNAs, guide RNAs, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, nucleic acid probes and nucleic acid primers. A nucleic acid molecule may contain unconventional or modified nucleotides. The terms "polynucleotide sequence" and "nucleic acid sequence" as used herein interchangeably refer
to the sequence of a polynucleotide molecule. The nomenclature for nucleotide bases as set forth in 37 CFR §1.822 is used herein.

[0041] The nucleic acid molecules of the present disclosure will preferably be "biologically active" with respect to either a structural attribute, such as the capacity of a nucleic acid molecule to hybridize to another nucleic acid molecule, or the ability of a nucleic acid sequence to be recognized and bound by a transcription factor (or to compete with another nucleic acid molecule for such binding).

[0042] Nucleic acid molecules of the present disclosure include nucleic acid sequences of any length, including nucleic acid molecules that are preferably between about 0.05 kb and about 300 kb, or for example between about 0.05 kb and about 250 kb, or between about 0.05 kb and about 150 kb, or between about 0.1 kb and about 150 kb, or for example between about 0.2 kb and about 150 kb, about 0.5 kb and about 150 kb, or about 1 kb and about 150 kb.

[0043] The term "operably linked", as used herein, denotes a functional linkage between two or more sequences. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (for example, a promoter) is functional link that allows for expression of the polynucleotide of interest. In this sense, the term "operably linked" refers to the positioning of a regulatory region and a coding sequence to be transcribed so that the regulatory region is effective for regulating transcription or translation of the coding sequence of interest. In some embodiments disclosed herein, the term "operably linked" denotes a configuration in which a regulatory sequence is placed at an appropriate position relative to a sequence that encodes a polypeptide or functional RNA such that the control sequence directs or regulates the expression or cellular localization of the mRNA encoding the polypeptide, the polypeptide, and/or the functional RNA. Thus, a promoter is in operable linkage with a nucleic acid sequence if it can mediate transcription of the nucleic acid sequence. Operably linked elements may be contiguous or non-contiguous. Further, when used to refer to the joining of two protein coding regions, by "operably linked" is intended that the coding regions are in the same reading frame. Any of the promoters, terminators or other regulatory sequences disclosed herein can be operably linked to any of the SEQ ID NOs or other polynucleotide sequences disclosed herein.

[0044] The terms "promoter", "promoter region", or "promoter sequence" refer to a nucleic acid sequence capable of binding RNA polymerase to initiate transcription of a gene in a 5' to 3' ("downstream") direction. A gene is "under the control of" or "regulated by" a promoter when the binding of RNA polymerase to the promoter is the proximate cause of said gene's
transcription. The promoter or promoter region typically provides a recognition site for RNA polymerase and other factors necessary for proper initiation of transcription. A promoter may be isolated from the 5' untranslated region (5' UTR) of a genomic copy of a gene. Alternatively, a promoter may be synthetically produced or designed by altering known DNA elements. Also considered are chimeric promoters that combine sequences of one promoter with sequences of another promoter. Promoters may be defined by their expression pattern based on, for example, metabolic, environmental, or developmental conditions. A promoter can be used as a regulatory element for modulating expression of an operably linked polynucleotide molecule such as, for example, a coding sequence of a polypeptide or a functional RNA sequence. Promoters may contain, in addition to sequences recognized by RNA polymerase and, preferably, other transcription factors, regulatory sequence elements such as cis-elements or enhancer domains that affect the transcription of operably linked genes. A "Labyrinthulomycetes promoter" as used herein refers to a native or non-native promoter that is functional in Labyrinthulomycetes cells. Any of the genes or SEQ ID NOs described herein can be under the control of or regulated by any of the promoters described herein.

[0045] The term "recombinant" or "engineered" nucleic acid molecule as used herein, refers to a nucleic acid molecule that has been altered through human intervention. As non-limiting examples, a cDNA is a recombinant DNA molecule, as is any nucleic acid molecule that has been generated by in vitro polymerase reaction(s), or to which linkers have been attached, or that has been integrated into a vector, such as a cloning vector or expression vector. As non-limiting examples, a recombinant nucleic acid molecule: 1) has been synthesized or modified in vitro, for example, using chemical or enzymatic techniques (for example, by use of chemical nucleic acid synthesis, or by use of enzymes for the replication, polymerization, exonucleolytic digestion, endonucleolytic digestion, ligation, reverse transcription, transcription, base modification (including, e.g., methylation), or recombination (including homologous and site-specific recombination)) of nucleic acid molecules; 2) includes conjoined nucleotide sequences that are not conjoined in nature, 3) has been engineered using molecular cloning techniques such that it lacks one or more nucleotides with respect to the naturally occurring nucleic acid molecule sequence, and/or 4) has been manipulated using molecular cloning techniques such that it has one or more sequence changes or rearrangements with respect to the naturally occurring nucleic acid sequence. As non-limiting examples, a cDNA is a recombinant DNA molecule, as is any nucleic acid molecule that has been generated by in vitro polymerase reaction(s), or to which linkers have
been attached, or that has been integrated into a vector, such as a cloning vector or expression vector. A recombinant cell contains a recombinant nucleic acid.

[0046] When applied to organisms, the terms "transgenic" "transformed" or "recombinant" or "engineered" or "genetically engineered" refer to organisms that have been manipulated by introduction of an exogenous or recombinant nucleic acid sequence into the organism. Non-limiting examples of such manipulations include gene knockouts, targeted mutations and gene replacement, promoter replacement, deletion, or insertion, as well as introduction of transgenes into the organism. For example, a transgenic microorganism can include an introduced exogenous regulatory sequence operably linked to an endogenous gene of the transgenic microorganism. Recombinant or genetically engineered organisms can also be organisms into which constructs for gene "knock down" have been introduced. Such constructs include, but are not limited to, RNAi, microRNA, shRNA, antisense, and ribozyme constructs. Also included are organisms whose genomes have been altered by the activity of meganucleases or zinc finger nucleases. A heterologous or recombinant nucleic acid molecule can be integrated into a genetically engineered/recombinant organism's genome or, in other instances, not integrated into a recombinant/genetically engineered organism’s genome. As used herein, "recombinant microorganism" or "recombinant host cell" includes progeny or derivatives of the recombinant microorganisms of the disclosure. Because certain modifications may occur in succeeding generations from either mutation or environmental influences, such progeny or derivatives may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0047] "Regulatory sequence", "regulatory element", or "regulatory element sequence" refers to a nucleotide sequence located upstream (5'), within, or downstream (3') of a polypeptide-encoding sequence or functional RNA-encoding sequence. Transcription of the polypeptide-encoding sequence or functional RNA-encoding sequence and/or translation of an RNA molecule resulting from transcription of the coding sequence are typically affected by the presence or absence of the regulatory sequence. These regulatory element sequences may comprise promoters, cis-elements, enhancers, terminators, or introns. Regulatory elements may be isolated or identified from untranslated regions (UTRs) from a particular polynucleotide sequence. Any of the regulatory elements described herein may be present in a chimeric or hybrid regulatory expression element. Any of the regulatory elements described herein may be present in a recombinant construct of the present disclosure.
The term "terminator" or "terminator sequence" or "transcription terminator", as used herein, refers to a regulatory section of genetic sequence that causes RNA polymerase to cease transcription.

The term "transformation", "transfection", and "transduction", as used interchangeably herein, refers to the introduction of one or more exogenous nucleic acid sequences into a host cell or organism by using one or more physical, chemical, or biological methods. Physical and chemical methods of transformation include, by way of non-limiting example, electroporation and liposome delivery. Biological methods of transformation include transfer of DNA using engineered viruses or microbes (for example, *Agrobacterium*).

**Promoters and Terminators**

The recombinant cell or organism of the invention can be any suitable organism but in some embodiments is a *Labyrinthulomycetes* cell, which can contain an expression cassette or expression vector described herein. The promoter (and/or terminator) sequences comprised on the expression cassette or vector can be any suitable promoter and/or terminator. Promoters and/or terminators can be used in any combination. For example, any promoter described herein or other promoters that may be isolated from or functional in *Labyrinthulomycetes* or derived from such sequences can be used in combination with any terminator described herein or other terminators functional in the recombinant cell or organism, or derived from such sequences. Functional promoters, terminators, and genetic elements are those able to perform their genetic purpose, for example, to initiate transcription of a DNA sequence. Functional terminators mediate transcriptional termination. For example, promoter and/or terminator sequences may be derived from organisms including, but not limited to, heterokonts (including *Labyrinthulomycetes*), fungi, microalgae, algae, yeast, and other eukaryotic organisms. In various embodiments the promoter and/or terminator is any one operable in a cell or organism that is a *Labyrinthulomycetes*, including any genus thereof. Any of the constructs can also contain one or more selection markers, as appropriate. A large number of promoters and terminators can be used with the host cells of the invention. Those described herein are examples and the person of ordinary skill with resort to this disclosure will identify other promoters useful in the invention. Examples of promoters include the alpha-tubulin promoter (tubap), the TEFp promoter, Hsp60-788 promoter, Tsp-749 promoter, Tuba738 promoter, Tuba-997 promoter, a promoter from the polyketide synthase system, and a fatty acid desaturase promoter. Examples of useful terminators include pgk1, CYC1, and eno2. Promoters and terminators can be used in any
advantageous combination and all possible combinations of promoters and terminators are disclosed as if set forth fully herein.

[0051] In some embodiments the expression cassettes of the invention comprise one or more of 1) a polynucleotide encoding one or more signal sequences described herein; 2) a polynucleotide encoding one or more promoters; 3) a polynucleotide encoding one or more terminators; and 4) a polynucleotide encoding one or more proteins heterologous and/or exogenous to the host cell; 4) optionally, a polynucleotide encoding one or more selectable markers for screening on a medium or a series of media. The sequence encoding the heterologous protein can be operably linked to the promoter, or otherwise positioned to be regulated by the promoter. The promoter and terminator can be functional in a Labyrinthulomycetes cell. These components of an expression cassette can be present in any combination, and each possible sub-combination is disclosed as if fully set forth herein. In specific embodiments the signal sequences can be any described herein, but can also be other signal sequences that can be identified with reference to the present disclosure and which are also functional in the host cells. For example, the expression cassette or vector can encode a promoter and terminator sequences functional in Aurantiochytrium, a heterologous protein operably linked to the promoter, and a signal sequence having at least 90% sequence identity to any signal sequence described herein. In more embodiments the promoter can be an alpha-tubulin promoter or TEFp. The promoters can be paired with any suitable terminator, but in specific embodiments the tubop can be paired with the pgk1t terminator. In another embodiment the TEFp promoter can be paired with the eno2 terminator, both terminators being from Saccharomyces cerevisiae and also being functional in Labyrinthulomycetes. In another embodiment the promoter can be TEFp and the terminator can be pgk1 or CYC1. In another embodiment the promoter can be Hsp60-788 and the terminator can be pgk1, CYC1, or eno2. In another embodiment the promoter can be Tsp-749 and the terminator can be pgk1, CYC1, or eno2. In another embodiment the promoter can be Tubα-738 and the terminator can be pgk1, CYC1, or eno2. In another embodiment the promoter can be Tubα-997 and the terminator can be pgk1, CYC1, or eno2. In another embodiment the promoter can be from the polyketide synthase system, and the terminator can be pgk1, CYC1, or eno2. In another embodiment the promoter can be a fatty acid desaturase promoter and the terminator can be pgk1, CYC1, or eno2. The selectable marker can be any suitable selectable marker or markers but in specific embodiments it can be nptII or hph. In one embodiment nptII can be linked to the heavy chain constructs and hph can be linked to the light chain constructs.
Figure 1 provides an illustration of specific examples of an expression cassettes useful in the invention. The tubulin-alpha promoter drives the expression of the mature light chain of anti-TNF-alpha antibody with a variety of signal peptides. pCAB030 is a version of these constructs where an ATG was introduced at the start of the ORF, but no signal peptide. All these constructs also carry the hph marker for selection on hygromycin B plates. Figure 2 is an illustration of a general map of pCAB-031-043, with each serial number relating to a different signal sequence. pCAB031-043 are versions of pCAB018-030 where sequences were introduced to allow for directed integration at the GPAT2 locus via homologous recombination. The 5’ and 3’ homology arms were amplified from chytrid genomic DNA using primer pairs having 5’ extensions that introduced sequences to allow for Gibson Assembly® cloning at the SbfI site of plasmids pCAB018-029. Other primers had 5’ extensions that introduced complementary sequences to allow for Gibson Assembly®. In addition, the primers were designed to introduce a unique BstZ17I site. The 5’ and 3’ homology arms were combined with the SbfI digested pCAB018-030 in a Gibson Assembly® reaction to yield pCAB031-043.

The host cell or organism can comprise a vector or expression cassette of the invention described herein. The vector or expression cassette can be inserted or otherwise integrated into the genome of the cell or organism, or integrated into other DNA of the cell or organism. In other embodiments the vector or expression cassette can also be present within the cell or organism without being integrated in the genome, for example present on a vector or plasmid. In some embodiments the vector or expression cassette can be multiplied in the cell or organism and/or passed during cell division.

Therapeutic Proteins and Peptides

The host cells of the invention can produce a variety of functional heterologous proteins or peptides, such as therapeutic peptides or proteins, or fragments of any of them, that are encoded on the expression vectors or cassettes of the invention. Therapeutic peptides or proteins are those peptides and proteins that are useful in the treatment or alleviation of a disease or medical condition in a human or animal patient. Such use is more than merely nutritional and performs an action on targets in the human or animal body resulting in the treatment or alleviation of the disease or medical condition and examples are provided herein. But any protein or peptide can be produced according to the invention. The proteins or therapeutic proteins produced by the host cells of the invention can be heterologous to or homologous to the host cell. In some embodiments the therapeutic peptide or protein of the invention is an antibody or portion thereof, for example the light chain, heavy chain, or both
of an antibody, or one or more single-chain variable fragment(s) (scFvs) of an antibody, or a bi-specific antibody.

[0055] Antibodies are immune system-related proteins called immunoglobulins. Each antibody consists of four polypeptides - two heavy chains and two light chains joined to form a "Y" shaped molecule. The antibodies can thus form a tetrameric quartenary structure. The amino acid sequence in the tips of the "Y" can vary greatly among different antibodies, providing the variable region. This variable region provides the antibody specificity for binding an antigen. The variable region includes the ends of the light and heavy chains. In some embodiments the antibodies of the invention are composed of a heavy chain (HC) and a light chain (LC), and thus require both genes that encode the two chains to be expressed. The antibodies of the invention can be assembled antibodies, meaning that the antibody comprises a light chain and a heavy chain, and they can be covalently bound to each other as a whole antibody. In some embodiments the antibodies of the invention are IgG antibodies. The antibodies of the invention can be composed of four peptide chains, including two identical heavy chains and two identical light chains. The antibodies can have two antigen binding sites. The two heavy chains can be linked to each other and to a light chain by disulfide bonds and have two identical halves that form the Y-like shape and contain identical antigen binding sites. The antibodies can also be multimeric antibodies, and can also have binding activity to more than one epitope. It was unexpectedly discovered that the cells of the invention engineered as described herein express a light chain and a heavy chain of an antibody molecule and can also assemble the light chain and heavy chain into a complete, assembled antibody molecule, such as any of those described herein. The cells can also secrete the light chain and heavy chain of the antibody, or the assembled antibody molecule.

[0056] In some embodiments the therapeutic protein of the invention can be an antibody, such as a functional antibody or a functional fragment of an antibody. In various embodiments the antibody can have an amino acid sequence of eculizumab, natalizumab, cetuximab, omalizumab, ustekinumab, panitumumab, denosumab, trastuzumab or adalimumab (anti-TNF-alpha antibody), or a functional fragment of any of them. The antibody can also be any anti-TNF-alpha antibody or an anti-HER2 antibody, or a functional fragment thereof. Each of these proteins is an antibody and a therapeutic protein, as well as a monoclonal antibody. The therapeutic protein can also be any IgG2 or IgG4 antibody or fragment thereof. The heterologous proteins encoded by the expression vectors or cassettes of the invention can bind to any antigen, with examples including TNF-alpha and HER2.
[0057] Trastuzumab (Herceptin®) is an anti-HER2 monoclonal antibody that interferes with the HER2/neu receptor and is useful in the treatment of certain breast cancers. The HER receptors are proteins that are embedded in the cell membrane and communicate molecular signals from outside the cell (EGFs) to inside the cell, and turn genes on and off. The HER protein, Human Epidermal Growth Factor Receptor, binds to Human Epidermal Growth Factor, and stimulates cell proliferation. In some cancers, notably certain types of breast cancer, HER2 is over-expressed, and causes cancer cells to reproduce uncontrollably. Adalimumab (Humira®) is an anti-TNF-alpha monoclonal antibody useful as a TNF inhibiting anti-inflammatory medication and used to treat various ailments such as arthritis, Crohn’s disease, colitis, among other diseases. Adalimumab binds to tumor necrosis factor-alpha (TNFα). TNFα normally binds to TNFα receptors, which leads to the inflammatory response of autoimmune diseases. By binding to TNFα, adalimumab reduces this inflammatory response. A functional antibody or antibody fragment is a molecule that is an antibody or antibody fragment that binds to a target epitope and thereby produces a desired response, for example a biological response or action, or the cessation of a response or action). The desired response can be the same as the response to a natural antibody, but the response can also be to mimic or disrupt the natural biological effects associated with ligand-receptor interactions. When the protein is a functional antibody fragment it can comprise at least a portion of the variable region of the antibody or can comprise a portion of or the entire antigen binding site of an antibody, and therefore can perform the antigen binding properties that are similar to or the same in nature and affinity to those of the complete antibodies. Any of the recombinant cells disclosed herein can comprise a functional and/or assembled antibody molecule described herein, or a functional fragment thereof. And any of the expression vectors or cassettes described herein can encode such a functional antibody or functional fragment thereof. Any type of heterologous protein or antibody or antibody fragment can be produced in the present invention. Examples include IgG2 or IgG4 or subunits of immunoglobulins, for example the kappa subunit.

[0058] When the heterologous protein is an antibody, the antibody can have a light chain and a heavy chain, either or both of which can bind to any antigen described herein, including but not limited to the HER2 receptor or TNF-alpha. The antibody can also have at least 70% or at least 80% or at least 90% or at least 95% or at least 97% or at least 98% or at least 99% or 85-99% or 90-99% or 95-99% sequence identity with an amino acid sequence of any therapeutic protein or protein fragment described herein, e.g. any of SEQ ID NO: 1-8. The antibody can have a light chain and a heavy chain having any combination of said sequence
identities to SEQ ID NO: 1 and 2; or SEQ ID NO: 3 and 4; or SEQ ID NO: 5 and 6; or SEQ ID NO: 7 and 8, as the light chain and heavy chain, respectively. For example, the antibody could have at least 90% sequence identity to SEQ ID NO: 1 and at least 95% sequence identity to SEQ ID NO: 2, and so on. The therapeutic protein can also be a functional fragment of an antibody described herein. A functional fragment is a portion of the full antibody or therapeutic protein that nevertheless has at least 85% or at least 90% or at least 95% of the relevant activity of the full antibody or therapeutic protein. It can also have at least 200 or at least 400 or at least 500 or at least 550 or at least 600 or at least 800 or at least 1000 or at least 1200 or at least 1300 contiguous amino acid residues to any of SEQ ID NO: 1-8 and also have said activity. Any of the therapeutic proteins or antibodies described herein (or functional fragment of any), or combinations thereof described herein, can also have at least one substitution modification versus SEQ ID NO: 1-8.

[0059] In some embodiments the heterologous or therapeutic protein is a bi-specific T cell engager antibody. These are bispecific antibodies that are fusion proteins having two single-chain variable fragments (scFvs) of different antibodies, or amino acid sequences from four different genes on a single peptide chain (e.g. of about 55 kDa). Bispecific T cell engagers can comprise an scFv that binds to a T cell antigen (e.g. the CD3 receptor) and the other scFvs can bind to a tumor-specific antigen molecule. As shown herein, these bispecific T cell engagers can also be encoded by the expression cassettes or vectors of the invention. In various embodiment one of the scFvs can bind to a T cell antigen and the other anti-CD19, anti-CD3, anti-CD38, or prostate-specific membrane antigen (PSMA).

[0060] The invention also provides methods of producing any heterologous or therapeutic protein described herein by cultivating a host cell of the invention in a suitable culture medium. The host cell can contain an expression cassette as described herein and can encode any functional signal sequence described herein that directs the secretion of the heterologous protein from the cell and into the growth medium. The heterologous protein can be harvested from the culture. In some embodiments the production method is a continuous production method or a continuous culture method, meaning that the protein can be harvested from the growth medium and the growth medium replenished without stopping or interrupting the culture and the production of the protein.

[0061] Growth of the host cells can be conducted in any convenient bioreactor or other vessel for growing cell cultures, including both small scale and large scale vessels. Cell cultures can be grown in any convenient vessel, system, or continuous culture device. Examples include, but are not limited to, a shake flask, a chemostat, a cytostat, an auxostat, a
A chemostat is a bioreactor to which fresh medium is continuously added, while culture liquid containing left over nutrients, metabolic end products, and microorganisms are continuously removed at the same rate to keep the culture volume constant. In any of the vessels a physiological steady state under constant culture parameters can be achieved and continuous growth maintained at a specific growth rate. By changing the rate with which medium is added to the bioreactor the specific growth rate of the host cells or culture can be easily controlled. In various embodiments culture parameters can include any one or more of culture volume, flow rate of medium into (feed) and/or out of (effluent) the vessel, dilution rate, dissolved oxygen concentration, nutrient concentration, sugar concentration, product concentration, pH, and cell density. Thus, in some embodiments a constant inward flow of fresh medium can be maintained at a rate equal to a constant outward flow of the cell culture. In this way, the culture can be continuously diluted and continuously grows (divides) where the growth (rate of cell division) of the culture is directly related to the rate of culture dilution. Thus, the rate of cell division can be adjusted by adjusting the culture dilution rate, and variants that have a growth advantage under the selective condition become enriched in the culture as poorer growing cells “wash out” through continuous dilution of the culture. In some embodiments one or more of (or all of) the culture conditions are held constant. In some embodiments culture liquid is removed from the bioreactor but biomass is retained on a filter (e.g. a retentostat).

A cytostat is similar to a chemostat, except that the culture is maintained at a cell density well below that at which nutrient limitation could occur. Dilution of the cytostat occurs based on cell density of the culture, which is monitored at regular intervals or continuously based on the determination of the cell concentration and/or the single cell property distribution of the growing cell populations. The culture can be monitored and controlled by an automated flow cytometer. The flow cytometer can be integral to the cytostat apparatus (see, for example, US Patent No. 7,901,937). Because cytostat cultures do not approach nutrient limitation nor experience any significant accumulation of fermentation products that may affect growth, cytostat cultures are considered to be in a steady state of growth and can be grown even at very low cell concentrations. The culture parameters can be precisely defined by the feed composition since products of cell growth may not be present in significant amounts, and therefore effects on cell growth of various culture parameters can be easily determined. Because the dilution rate is calibrated to the cell division rate, cells can be allowed to proliferate without excessive dilution until cells begin to achieve a certain density.
[0063] In various embodiments the bioreactor can be of any volumetric capacity. For microbial strains such as fungi and chytrids, for example, culturing can be in a bioreactor (e.g., a chemostat or cytostat) having a fermenter volumetric capacity of from about 25 mL to about 10 L, and can be from 200 mL to about 5 L, or from about 300 mL to about 2 L, or from about 400 mL to about 1 L or from 300 mL to about 3 L or from about 5 L to about 10 L or more than 10 L. The culture period can be for any length of time, such as for example, from one day to several months. Preferably, for fungi and chytrids, the culturing period in the bioreactor is for a period of time greater than one day, for example, for a period of from about 2 days to about 30 days, or from about 3 days to about 20 days, or from about 4 days to about 15 days. In some embodiments cytostat selection is preferred. In a cytostat, the culture period can in some embodiments, for example, be from about 3 days to about 10 days.

[0064] Single colonies of the microorganism can be isolated by dilution plating or flow cytometry from cultures grown for any period of time and the resulting isolate or isolates can be screened for any desirable properties. For example, an isolate can be tested for any one of or any combination of: increased growth rate, increased biomass accumulation, increased lipid, fatty acid, or protein production rate, increase antibody production rate, increased triglyceride production rate, increased total lipid accumulation, increased FAME accumulation, increased triglyceride accumulation, increased FAME production rate, increased DHA production rate, increased DHA accumulation, increased DHA as a percentage of fatty acids, increased ratio of DHA to DPA, etc. Any feasible methods for determining protein, antibody, lipid, or fatty acid amounts can be employed. The testing can be under any culture conditions, including those listed hereinabove that may be used in culturing a strain prior to testing, for example, using particular carbon or nitrogen sources or concentrations, salt concentration, temperature, pH, etc.

Signal Peptides

[0065] A signal peptide (or signal sequence) is a short peptide, usually located at the N-terminus of a newly synthesized protein. These sequences direct the newly synthesized protein (usually to the endoplasmic reticulum) for secretion from the cell and into the growth medium. Signal peptides, or a functional fragment thereof, can have at least 5 or at least 10 or at least 15 or at least 20 or at least 25 amino acids, or from 5-30 or from 5-40 or 5-50 amino acids or from 5-35 or 5-25 or 5-20 or 10-20 or 10-25 or from 25-50 or from 3-12 or 3-15 amino acids, which in some embodiments can be any consecutive amino acid sequence of said length from any one of SEQ ID NO: 9-94.
[0066] The present invention also provides signal peptides and nucleic acid molecules that encode signal peptides useful in a host organism of the invention. Any suitable signal peptide can be utilized in the invention. Signal peptides or signal sequences can direct the secretion of a protein (e.g., a heterologous protein) from a host cell, i.e., they can be functional signal peptides or signal sequences. Signal peptides useful in the invention have secretion signal activity in a host cell of the invention. Examples of signal peptides useful in the invention are provided in the amino acid sequences of SEQ ID NOs: 9-94, or a functional fragment of any of them. A functional fragment of a signal peptide is a fragment that is less than the full length peptide but nevertheless directs the secretion of a protein (e.g., the protein encoded by the vector or expression cassette) from the cell.

[0067] Signal peptides of the invention can be any peptide having at least 70% or at least 75% or at least 80% or at least 85% or at least 90% or at least 92% or at least 95% or 70-99% or 80-99% or 85-99% or 90-99% or 95-99% sequence identity with any of the amino acid sequences disclosed herein (e.g., amino acids of SEQ ID NO: 9-94), or a functional fragment thereof; signal peptides can also be encoded by a nucleic acid sequence having at least 70% or at least 75% or at least 80% or at least 85% or at least 90% or at least 92% or at least 95% or 70-99% or 80-99% or 85-99% or 90-99% or 95-99% sequence identity with any of the nucleic acid sequences disclosed herein (e.g., SEQ ID NO: 21-32 or 48-67 or 79-94), or a functional fragment thereof, which functional fragment can be any sequence of at least 100 or 120 or 150 or 200 consecutive nucleotides. Signal peptides, or a functional fragment thereof, can also have at least 5 or at least 7 or at least 10 or at least 15 or at least 20 or at least 25 or at least 35 or at least 40 or at least 45 or at least 50 contiguous amino acid residues to any of the amino acid sequences disclosed herein. Any of the signal peptides or sequences, or functional fragment thereof, can also have at least one substitution modification versus any of the sequences disclosed herein. In various other embodiments the signal sequence can have at least two or at least three or at least four or at least five or at least six or at least seven or at least eight or at least nine or at least 10 or 1-3 or 3-5 or 5-7 or 7-9 or 1-10 substitution modifications versus an amino acid sequence disclosed herein. Signal sequences can also be encoded by a nucleic acid having at least one substitution modification versus any of the nucleic acid sequences disclosed herein. In various embodiments the signal sequence can be encoded by a nucleic acid having at least two or at least three or at least four or at least five or at least six or at least seven or at least eight or at least nine or at least 10 or 1-3 or 3-5 or 5-7 or 7-9 or 1-10 or 10-20 or 1-50 or 1-75 or 1-100 substitution modifications versus a nucleic acid sequence disclosed herein. Functional fragments of these sequences are also disclosed,
and can have at least 15 or at least 20 or at least 25 or at least 50 or at least 100 nucleotides contiguous to any amino acid or nucleic acid sequence disclosed herein.

**[0068]** SEQ ID NO: 9-20 are encoded by SEQ ID NO: 21-32. SEQ ID NO: 33-47 are encoded by SEQ ID NO: 48-62. SEQ ID NO: 63-78 are encoded by SEQ ID NO: 79-94. Any of the nucleic acid sequences disclosed herein can be comprised in an expression vector or expression cassette. Expression vectors can comprise any of the expression cassettes disclosed herein. Expression vectors or expression constructs can be plasmids or other constructs and are designed for gene expression in host cells. Expression vectors or constructs can be used to introduce sequences of the invention into target cells and can utilize the cell's mechanisms for expression of proteins or peptides to produce the protein or peptide in the host cell. The expression cassettes of the invention comprising sequences of the invention can also be introduced into cells as circular or linear DNA and be incorporated into the genome of the organism and function from that location on the genome. The expression vectors or expression cassettes of the invention can also contain additional sequences for propagation and/or expression of the sequences in the chosen host cell type.

**[0069]** Non-limiting examples of a substitution modification for a signal peptide or a therapeutic protein (e.g. antibody) described herein can include a substitution, an insertion, a deletion, a rearrangement, an inversion, a replacement, a point mutation, and a suppressor mutation in the encoding DNA, which can result in a change in the encoded amino acid sequence. Methods of performing substitution modifications are known in the art and are readily available to the artisan such as, for example, site-specific mutagenesis, PCR, and gene synthesis. Non-limiting examples of substitution modification methods can also be found in Maniatis et al., (1982) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. In some embodiments the substitution modification(s) do not substantially alter the functional properties of the resulting nucleic acid fragment or protein/peptide relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the disclosure encompasses more than the specific exemplary sequences. A substitution modification can also include alterations that produce silent substitutions, insertions, deletions, etc. as above, but do not alter the properties or activities of the encoded protein or how the proteins are made.

**[0070]** The degree of amino acid or nucleotide sequence identity can be determined by various computer programs for aligning the sequences to be compared based on designated program parameters. For example, sequences can be aligned and compared using the local homology algorithm of Smith & Waterman *Adv. Appl. Math.* 2:482-89, 1981, the homology
alignment algorithm of Needleman & Wunsch *J. Mol. Biol.* 48:443-53, 1970, or the search for similarity method of Pearson & Lipman *Proc. Nat'l. Acad. Sci. USA* 85:2444-48, 1988, and can be aligned and compared based on visual inspection or can use computer programs for the analysis (for example, GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI).

[0071] The BLAST algorithm, described in Altschul et al., *J. Mol. Biol.* 215:403-10, 1990, is publicly available through software provided by the National Center for Biotechnology Information. This algorithm identifies high scoring sequence pairs (HSPS) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., 1990, supra). Initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated for nucleotides sequences using the parameters M (reward score for a pair of matching residues; always $>0$) and N (penalty score for mismatching residues; always $<0$). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. For determining the percent identity of an amino acid sequence or nucleic acid sequence, the default parameters of the BLAST programs can be used. For analysis of amino acid sequences, the BLASTP defaults are: word length (W), 3; expectation (E), 10; and the BLOSUM62 scoring matrix. For analysis of nucleic acid sequences, the BLASTN program defaults are word length (W), 11; expectation (E), 10; M=5; N=-4; and a comparison of both strands. The TBLASTN program (using a protein sequence to query nucleotide sequence databases) uses as defaults a word length (W) of 3, an expectation (E) of 10, and a BLOSUM 62 scoring matrix. See, Henikoff & Henikoff, *Proc. Nat'l. Acad. Sci. USA* 89: 10915-19, 1989.

[0072] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-87, 1993). The smallest sum probability (P(N)), provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered
similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, preferably less than about 0.01, and more preferably less than about 0.001.

Titers

[0073] Another advantage of the cells and methods of the invention is that therapeutic or heterologous proteins can be produced in host cells, such as Labyrinthulomycetes cells, in high titers. In some embodiments the heterologous protein can be produced in culture having a titer, expressed as mg/ml produced in a 24 hour period, of at least 5 mg/L of culture or at least 6 mg/L of culture or at least 7 mg/L culture or at least 9 mg/L or at least 11 mg/L or at least 12 mg/L or at least 15 mg/L or at least 17 mg/L or at least 18 mg/L or at least 19 mg/L or at least 20 mg/L or at least 25 mg/L or at least 30 mg/L or at least 35 mg/L or at least 40 mg/L or at least 45 mg/L or at least 50 mg/L or at least 55 mg/L or at least 60 mg/L or 5-15 mg/L or 5-16 mg/L or 5-17 mg/L or 5-18 mg/L or 5-20 mg/L or 15-25 mg/L or 20-30 mg/L or 30-40 mg/L or 40-50 mg/L or 50-60 mg/L or 60-80 mg/L or 80-100 mg/L of culture in a 24 hour period. The heterologous protein produced can be any described herein, and in some embodiments the heterologous protein is an antibody, which can comprise a heavy chain and a light chain.

Example 1 – Expression of anti-TNF-alpha Antibody Light Chain

[0074] This example shows the expression of the light chain of the anti-TNF-alpha antibody. An Aurantiocystium sp. host cell was utilized with SEQ ID NOs: 9-20 as example signal sequences. Each of these peptides were cloned at the 5’ end of a sequence encoding the mature anti-TNF-alpha antibody light chain in the construct of Fig. 1 (pCAB018). Control constructs had no signal peptide. Another set of constructs were made containing each of the signal peptides and also with homology arms to target integration at the GPAT2 locus (Fig. 2). This was done to avoid positional effects. These constructs were designed to introduce a unique BstZ171 site between the 5’ and 3’ homology arms such that the 5’to 3’ linearized product would be configured as follows: 5’ homology arm, the expression cassette, 3’ homology arm.

[0075] Genome editing techniques were utilized to insert the pCAB vector at the GPAT2 site. The transformants were screened for correct integration of the cassette at the GPAT2 by 1) qPCR for deletion of GPAT2, 2) colony PCR for 5’ and 3’ junctions between the cassette and the external genomic sequence, and 3) qPCR for similar copy numbers across transformants using primers targeting the light chain sequence.
[0076] Transformants for each signal peptide construct were analyzed in triplicate and examined for secretion of anti-TNF-alpha antibody light chain. Each clone was cultured overnight in 3 mL FM2 (17g/L aquarium salt, 10 g/L yeast extract, 10 g/L peptone, 20 g/L dextrose) in a 24-well plate. They were then diluted 1000x into fresh FM2 (3mL) and incubated for about 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for the presence of light chain by LC-capture/LC-detect sandwich ELISA. The results are shown in Figure 4. It was also verified that the differences in titers were not due to differences in expression levels by performing RNA extraction and analysis by PCR and comparing the results with those from LC-capture/LC-detect sandwich ELISA assay. Results indicated that the differences in titer were not due simply to differences in gene expression.

Example 2 – Expression and Secretion of anti-HER2 and anti-TNF-alpha HC and LC

[0077] Constructs containing expression vectors with the sequences of anti-TNF-alpha heavy chain (SEQ ID NO: 2) and light chain (SEQ ID NO: 1), and with anti-HER2 heavy chain (SEQ ID NO: 4) and light chain (SEQ ID NO: 3) were cloned. Signal sequences (SEQ ID NOs: 11 and/or 19) were also cloned in vectors containing each of the heavy chain or light chain sequences, with alpha-tubulin as the promoter and pgk1t as the terminator. hph and nptII were used as selection markers.

[0078] Constructs containing the sequence for the heavy chain and light chain for each antibody were linearized and then co-transformed into Aurantiochytrium sp. Constructs were selected for on paromomycin medium and then hygromycin to select for cells containing both a light chain and a heavy chain sequence. Diagnostic PCR identified clones that contained full length heavy chain and light chain expression cassettes, and numerous clones were identified. Results were confirmed by PCR on genomic DNA and RNA expression of the cassettes was also confirmed by PCR.

[0079] ELISA screening was performed on the clones transformed with expression cassettes and that utilized the tubop. Clones were grown for about 20 h at 30 °C and 225 rpm and then back diluted into fresh medium and grown for about 20 hours in the same conditions. Cells were removed by centrifugation. The ELISA method utilized an HC capture and an LC detect sandwich assay that provided a positive signal only if both the HC and LC were present as a complete antibody. Figure 3 illustrates that antibody supernatant titers in the mg/L range were obtained for clones having heavy and light chains. The first 16 clones contained heavy and light chains of anti-TNF-alpha and clones 17 and 18 secrete anti-HER2. As illustrated in the bar graph almost all of the clones making anti-TNF-alpha had a
titer of greater than 7.0 mg/L. Both of the clones for anti-HER2 had a titer of greater than 10 mg/L. It was also noted that culture OD had a positive relationship with the antibody titer with a higher OD indicating a higher antibody titer. This experiment was repeated with TEFp as promoter and a similar result was obtained with some clones giving titers as high as greater than 20 mg/L or greater than 25 mg/L or greater than 30 mg/L. It was therefore demonstrated that the cells of the invention are able to produce a heavy chain and light chain of an antibody molecule, and to assemble them into a whole antibody.

**Example 3 – Binding of anti-HER2 Antibody Produced by *Labyrinthulomycetes***

[0080] This Example shows that antibody made in a host cell (*Aurantiochytrium*) binds to a recombinant form of the receptor tyrosine-protein kinase erbB-2 (Her2). Neat samples and seven half-log serial dilutions of each sample were added to a 96 well plate that was incubated overnight with 5μg/ml of recombinant Her2 as described above. The recombinant protein consists of 397 amino acids that represent the extracellular domain of the native receptor. The assay is a solid phase ELISA based on the sandwich principle performed in a 96 well plate. Buffer was 20x TBS Tween® 20 buffer. Samples were incubated in a plate coated with recombinant Her2 (human ErbB-2). Following a wash step, an HRP conjugated secondary antibody (anti-human IgG Fc (HRP)) was introduced and bound to human IgG (Figure 5).

**Example 4 – Binding of Anti-TNF-alpha Antibody Produced by *Labyrinthulomycetes***

[0081] This assay quantitates the amount of antibody present in solution that can bind to Tumor Necrosis Factor alpha (TNFα). A solid phase ELISA “sandwich” assay was utilized. Standards and diluted samples were incubated in a plate coated with TNF-alpha, followed by a wash step and then incubation with an HRP conjugated probe.

[0082] The assay was performed at room temperature (22-26 °C) and samples were assayed in duplicate and read on a microplate reader at 450nm. Raw data is expressed in absorbance units. A standard curve was generated for each assay and the anti-TNF-alpha concentrations (mg/L) calculated from the standard curve equation. The upper limit of detection was defined as 1000 mg/L and the lower limit as less than 30 mg/L.

[0083] Figure 6 illustrates data from an ELISA using supernatants from *Labyrinthulomycetes (Aurantiochytrium)* cultures expressing anti-TNF-alpha antibody. All clones had calculated values of antigen binding antibody that fell within the range of the standard curve. The data show that the anti-TNF-alpha produced as described herein binds the target antigen, TNF-alpha.
Example 5 – Anti-HER2 Functional Antibody Assay

[0084] Agonist and antagonist activity was examined for control compounds and anti-HER2 produced in the above Examples at the human epidermal growth factor receptor 2 (Her2). HER2 (also referred to as Erbb2) is a member of the human epidermal growth factor receptor (HER/EGFR/ERBB) family and a receptor tyrosine kinase (RTK). These receptors form homo/hetero dimers upon activation, resulting in the phosphorylation of tyrosine residues within the cytoplasmic domain. This process activates a variety of signaling pathways that ultimately result in maintaining cell survival, proliferation, and differentiation.

[0085] One important mechanism used by anti-HER2 to kill targeted tumor cells is Antibody-Dependent Cell-Mediated Cytotoxicity (ADCC), a process by which antibody bound cells are targeted for destruction by components of the cell-mediated immune system, such as natural killer cells. The ADCC reporter bioassay is a bioluminescent reporter assay for quantifying biological activity on pathway activation by therapeutic antibody drugs in an ADCC mechanism of action (MOA) assay. The effector cells used in this assay is an engineered Jurkat cells stably expressing the FcγRIIIa receptor (V158 variant) that binds the Fc region of antibodies bound to target cells and an NFAT response element driving expression of firefly luciferase. Antibody biological activity in ADCC MOA is quantified through the luciferase produced as a result of NFAT pathway activation. This is a surrogate read out for Her2 mediated cytotoxicity, and can be used as a direct measurement of antibody efficacy.

[0086] Purified antibody from Aurantiocchytrium cultures was tested for ADCC using the Her2 expressing human breast cancer cell line, BT-474 (ATCC® HTB-20™), as target cells. Antibody was added to target cells at various concentrations in the presence of recombinant effector cells and incubated for 6 hrs as described in the manufacturers protocol (Promega®). Luciferase activity in the effector cells was quantified with luminescence readout (RLU). The calculated EC50 value for Anti-HER2 was 0.061mg/L. Two control antibodies, Anti-TNFα produced in Aurantiocchytrium, and a commercially available human IgG produced no response. Figure 7 shows a dose response curve demonstrating the functional binding of anti-HER2.

Example 6 – Anti-TNF-alpha functional assay

[0087] The assay shows the inhibitory effect of anti-TNF-alpha in a cell based functional assay. Anti-TNF-alpha binds to free Tumor Necrosis Factor alpha (TNF-alpha) and blocks its ability to bind to TNF-alpha receptors. Upon activation, TNF-alpha receptors trigger a downstream signaling cascade that induces release of inflammatory cytokines and/or
apoptosis. Anti-TNF-alpha is expressed in *Aurantiochytrium* and the functional action of these antibodies in their ability to block TNF-alpha activity in vitro is demonstrated.

**[0088]** Anti-TNF-alpha inhibition of action was analyzed in a mouse fibroblast cell line (L929). The cells were exposed to TNF-alpha for 0-48 hours to stimulate cytokine production or cell death in the presence or absence of anti-TNF-alpha purified from the culture supernatant of *Aurantiochytrium*. Cell cytotoxicity was measured indirectly by using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega®). Figure 9 shows a concentration dependent effect of anti-TNF-alpha on TNF-alpha induced killing of the cells.

**Example 7 – Expression of anti-CD19/anti-CD3 bi-specific T-cell Engager**

**[0089]** This example shows the expression of the anti-CD19/anti-CD3 bi-specific T-cell engager antibody. An *Aurantiochytrium sp.* host cell was utilized with SEQ ID NO: 63-78 as signal sequences. Each of these peptides were cloned at the 5’ end of a sequence encoding the mature anti-CD19/anti-CD bi-specific T-cell engager antibody in the construct of Figure 9. The bi-specific T-cell engager antibody was expressed using the TEF promoter derived from the host organism. The control construct utilized the signal peptide of SEQ ID NO: 19.

**[0090]** The host strain was transformed with these constructs as circular plasmids. Transformants were screened for integration of the cassette based on homologous recombination within the TEFp sequence by colony PCR for 5’ and 3’ junctions between the cassette and external genomic sequence. In addition, transformants were screened by qPCR for similar copy numbers across transformants using primers targeting the bi-specific T-cell engager antibody.

**[0091]** Three independent transformants for each signal peptide construct were analyzed for secretion of anti-CD19/anti-CD3 antibody. Each clone was cultured overnight in FM2. They were then diluted 1000x into fresh FM2 and incubated for approximately 24 hours. The cells were pelleted by centrifugation and the supernatants assayed for the anti-CD19/anti-CD3 antibody by western blot analysis. Western blots were probed with an anti-His antibody and detected using an HRP-conjugated anti-mouse antibody and ECL-Plus™ substrate. Chemi-fluorescence was detected and band volumes calculated using a laser scanner. Band volumes were normalized to the control (SEQ ID NO: 19) and averaged (Table 1). These results show that the sequences direct the secretion of heterologous proteins.
Table 1

<table>
<thead>
<tr>
<th>Signal Peptide</th>
<th>Band density, relative to SEQ ID NO: 19 as 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID 63</td>
<td>46%</td>
</tr>
<tr>
<td>SEQ ID 64</td>
<td>56%</td>
</tr>
<tr>
<td>SEQ ID 65</td>
<td>59%</td>
</tr>
<tr>
<td>SEQ ID 66</td>
<td>46%</td>
</tr>
<tr>
<td>SEQ ID 67</td>
<td>61%</td>
</tr>
<tr>
<td>SEQ ID 68</td>
<td>70%</td>
</tr>
<tr>
<td>SEQ ID 69</td>
<td>69%</td>
</tr>
<tr>
<td>SEQ ID 70</td>
<td>62%</td>
</tr>
<tr>
<td>SEQ ID 71</td>
<td>69%</td>
</tr>
<tr>
<td>SEQ ID 72</td>
<td>73%</td>
</tr>
<tr>
<td>SEQ ID 73</td>
<td>54%</td>
</tr>
<tr>
<td>SEQ ID 74</td>
<td>60%</td>
</tr>
<tr>
<td>SEQ ID 75</td>
<td>77%</td>
</tr>
<tr>
<td>SEQ ID 76</td>
<td>72%</td>
</tr>
<tr>
<td>SEQ ID 77</td>
<td>66%</td>
</tr>
<tr>
<td>SEQ ID 78</td>
<td>68%</td>
</tr>
</tbody>
</table>

**Example 8 - Expression of anti-PSMA/anti-CD3 BiTE antibody**

[0092] This example shows the expression of the anti-PSMA/anti-CD3 bi-specific T-cell engager antibody. The signal peptide SEQ ID NO: 19 was cloned at the 5’ end of a sequence encoding the mature anti-PSMA/anti-CD bi-specific T cell engager antibody in the construct of Fig. 10a. The bi-specific Ab was expressed using the TEF promoter derived from the host organism.

[0093] The host strain was transformed with this constructs as a circular plasmid. 18 transformants were screened for integration of the cassette based on secretion of the engager antibody. The untransformed host (#267) and the anti-CD19/anti-CD3 bi-specific Ab fused to SEQ ID NO: 19 (from the previous example) were used as controls. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. They were then diluted 1000x into fresh FM2 (2.5 mL) and incubated for approximately 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for the bi-specific Ab by Western analysis. Westerns were probed with an anti-His antibody and detected using an HRP-conjugated anti-mouse antibody and ECL-Plus. Chemifluorescence was detected using a Typhoon™ FLA 9000. The results are shown in Figure 10b. Clones 1, 4, 5, 8, 9, 11, 12 and 17 secreted the anti-PSMA/anti-CD3 bi-specific Ab.

**Example 9 – Expression of an IgG2 Antibody Denosumab (Prolia®)**

[0094] This example shows the expression of an IgG2 antibody. SEQ ID NO: 19 signal peptide was cloned at the 5’ end of sequences encoding the mature heavy chain and light
chain for the IgG2 antibody (denosumab). The heavy and light chains were expressed using the TEF promoter derived from the host organism. nptII and hph were used as selection markers.

[0095] Constructs were linearized and co-transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to both paromomycin and hygromycin to identify clones carrying both constructs. Diagnostic PCR confirmed the presence of both light chain and heavy chain sequences. Six positive clones were examined for expression of denosumab. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. They were then diluted 1000x into fresh FM2 (2.5 mL) and incubated for approximately 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for IgG2 using the Human IgG subclass profile kit. All clones produced the IgG2 antibody (Table 2).

**Table 2: Titors of IgG2 Denosumab, from *Aurantiochytrium sp.***

<table>
<thead>
<tr>
<th>Clone #</th>
<th>Titors of IgG2 (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
</tr>
</tbody>
</table>

Example 10 – Expression of the IgG4 antibody natalizumab (Tysabri®)

[0096] This example shows the expression of an IgG4 antibody. SEQ ID NO: 19 signal peptide was cloned at the 5’ end of sequences encoding the mature heavy chain and light chain for the IgG4 antibody natalizumab. The heavy and light chains were expressed using the TEF promoter derived from the host organism. nptII and hph were used as selection markers.

[0097] Constructs were linearized and co-transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to both paromomycin and hygromycin to identify clones carrying both constructs. Diagnostic PCR confirmed the presence of both light chain and heavy chain sequences. Five positive clones were examined for expression of natalizumab. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. They were then diluted 1000x into fresh FM2 (2.5 mL) and incubated for approximately 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for IgG4 using the Human IgG subclass profile kit. All clones produced the IgG4 antibody (Table 3).
Table 3: Titers of the IgG4 natalizumab from *Aurantiochytrium sp.*

<table>
<thead>
<tr>
<th>Clone #</th>
<th>Titers of IgG2 (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>14</td>
<td>54</td>
</tr>
<tr>
<td>15</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>31</td>
<td>95</td>
</tr>
</tbody>
</table>

Example 11 – Expression of kappa/lambda bi-specific antibody (anti-II6RC/anti-IFNgamma)

This example shows the expression of a Kappa/Lambda bispecific antibody. The signal peptide SEQ ID NO: 19 was cloned at the 5’ end of sequences encoding the mature heavy chain, kappa light chain and lambda light chain for the bispecific anti-II6RC/anti-IFNγ antibody. The heavy and light chains were expressed using the TEF promoter derived from the host organism. nptII, hph and nat were used as selection markers.

Constructs were linearized and co-transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to paromomycin, hygromycin and nourseothricin to identify clones carrying all three constructs. Diagnostic PCR confirmed the presence of the heavy chain and both light chain sequences. Nine positive clones were examined for expression of the bispecific anti-II6RC/anti-IFNγ antibody. The untransformed *Aurantiochytrium sp* was used as negative control. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. They were then diluted 1000x into fresh FM2 (2.5 mL) and incubated for approximately 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for the bispecific anti-II6RC/anti-IFNγ antibody by Western analysis. Western blots were probed with an anti-human Ig antibody and detected using an HRP-conjugated anti-goat antibody and ECL-Plus. Chemifluorescence was detected using a Typhoon™ FLA 9000. All clones produced the heavy chain and the two light chains (Figure 11).

Example 12 – Expression of hybrid antibody XENP13551 (anti-CD38/anti-CD3)

This example shows the expression of a hybrid antibody. The signal peptide SEQ ID NO: 19 was cloned at the 5’ end of sequences encoding the mature heavy chain 1, heavy chain 2, and light chain chain for the hybrid bispecific anti-CD38/anti-CD3 antibody (XENP13551). The heavy and light chains were expressed using the TEF promoter derived from the host organism. nptII, hph and nat were used as selection markers.
[0101] Constructs were linearized and co-transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to paromomycin, hygromycin and nourseothricin to identify clones carrying all three constructs. Diagnostic PCR confirmed the presence of the heavy chain and both light chain sequences. Nine positive clones were examined for expression of the hybrid bispecific anti-CD38/anti-CD3 antibody (XENP13551). The untransformed *Aurantiochytrium sp* (#267) was used as negative control. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. They were then diluted 1000× into fresh FM2 (2.5 mL) and incubated for approximately 24 hours. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for the hybrid bispecific anti-CD38/anti-CD3 antibody (XENP13551) by Western analysis. Western blots were probed with an anti-human Ig antibody and detected using an HRP-conjugated anti-goat antibody and ECL-Plus. Chemifluorescence was detected using a Typhoon™ FLA 9000. All clones produced heavy chain 1, heavy chain 2 and the light chain (Figure 12).
Example 13 – Expression of TagBFP-His6

[0102] This example shows the expression of TagBFP-His6 (TagBFP with a C-terminal 6x His-tag). The signal peptide SEQ ID NO: 19 was cloned at the 5′ end of sequences encoding TagBFP. TagBFP was expressed using the TEF promoter derived from the host organism. hph was used as selection marker.

[0103] This construct was linearized and transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to hygromycin to identify clones carrying TagBFP, which was confirmed by diagnostic PCR. Two positive clones were examined for expression of TagBFP. The untransformed *Aurantiochytrium sp* (#267) was used as negative control. Each clone was cultured overnight in 25 mL FM2 in a 250 mL baffled shake flask. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for TagBFP by fluorescence. 200 μL of the supernatant was placed in wells of a black round-bottom 96-well plate and fluorescence measure using a 96-well plate reader (top read, Ex=402nm, Em=457nm, Cutoff=455nm). The increase in relative fluorescence in the clones carrying the TagBFP indicates that this protein is being expressed and secreted (Figure 13).
Example 14 – Expression of K3-His6

[0104] This example shows the expression of K3-His6 (human plasminogen Kringle 3 domain with a C-terminal 6xHis tag). The signal peptide SEQ ID NO: 19 was cloned at the 5′ end of sequences encoding K3-His6. K3-His6 was expressed using the TEF promoter derived from the host organism. hph was used as selection marker.
This construct was linearized and transformed into *Aurantiochytrium sp.* Transformants were screened for resistance to hygromycin to identify clones carrying K3-His6, which was confirmed by diagnostic PCR. Three positive clones were examined for expression of K3-His6. The untransformed *Aurantiochytrium sp* (#267) was used as negative control. Each clone was cultured overnight in 2.5 mL FM2 in a 24-well plate. The cells were pelleted by centrifugation (2000g x 5min) and the supernatants assayed for K3-His6. Westerns were probed with an anti-His antibody (HIS.H8) and detected using an HRP-conjugated anti-mouse antibody and ECL-Plus. Chemiluminescence was detected using a Typhoon™ FLA 9000. The results are shown in Figure 14. All clones expressed and secreted K3-His6.

Although the invention has been described with reference to the above examples, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.
Claims

1. A recombinant *Labyrinthulomycetes* cell comprising an expression cassette that encodes a heterologous protein and a functional signal sequence that directs the secretion of the heterologous protein from the cell, wherein
   a. the signal sequence has at least 90% sequence identity to a sequence of at least 10 contiguous amino acids selected from any one of SEQ ID NO: 9-20 or 33-47 or 63-78, or
   b. the signal sequence is encoded by a nucleic acid sequence having at least 90% sequence homology to SEQ ID NO: 21-32 or 48-62 or 79-94.

2. The cell of claim 1 wherein the heterologous protein is a therapeutic protein.

3. The cell of claim 1 wherein the heterologous protein is a functional antibody.

4. The cell of claim 3 wherein the functional antibody comprises at least one heavy chain and at least one light chain.

5. The cell of claim 3 wherein the antibody has an amino acid sequence selected from the group consisting of: eculizumab, natalizumab, cetuximab, omalizumab, ustekinumab, panitumumab, denosumab, trastuzumab and adalimumab.

6. The cell of claim 1 wherein the heterologous protein comprises a bispecific T cell engager.

7. The cell of claim 1 wherein the heterologous protein comprises a binding molecule that binds to an antigen selected from the group consisting of: anti-CD19, anti-CD3, anti-CD38, and prostate-specific membrane antigen (PSMA).

8. The cell of any one of claims 1-7 wherein the expression cassette encodes a promoter active in a *Labyrinthulomycetes* and that regulates transcription of the heterologous protein.

9. The cell of any one of claims 1-8 wherein the promoter is selected from the group consisting of: tubulind promoter (tubap) or translation elongation factor (TEFp) promoter.

10. The cell of any one of claims 1-9 wherein the expression cassette further encodes a terminator selected from pgk1t or eno2t.

11. The cell of claim 3 wherein the antibody has at least 90% sequence identity to an amino acid sequence of at least 100 contiguous amino acids selected from the group consisting of SEQ ID NO: 5-8.

12. An expression cassette comprising:
   a. at least one nucleotide sequence encoding a heterologous protein positioned to be regulated by the promoter, and
b. at least one nucleotide sequence encoding a signal sequence having at least 90% sequence identity to a nucleotide sequence of at least 30 contiguous nucleotides selected from any one of SEQ ID NO: 21-32 or 48-62 or 79-94.

13. The expression cassette of claim 12 further comprising a promoter sequence and a terminal sequence functional in a Labyrinthulomycetes.

14. The expression cassette of any one of claims 12-13 wherein the heterologous protein comprises an antibody.

15. The expression cassette of claim 14 wherein the antibody comprises a heavy chain and a light chain.

16. The expression cassette of claim 15 wherein the antibody has an amino acid sequence selected from the group consisting of: eculizumab, natalizumab, cetuximab, omalizumab, ustekinumab, panitumumab, denosumab, trastuzumab and adalimumab.

17. The expression cassette of any one of claims 12-15 wherein the expression cassette is non-contiguous.

18. The expression cassette of claim 12 wherein the heterologous protein comprises a bispecific T cell engager.

19. The expression cassette of claim 12 wherein the heterologous protein comprises a binding molecule that binds to an antigen selected from the group consisting of: anti-CD19, anti-CD3, anti-CD38, and prostate-specific membrane antigen (PSMA).

20. An antibody comprising a signal sequence functional in a Labyrinthulomycete and:
   a. a light chain having at least 90% sequence identity to SEQ ID NO: 5 and a heavy chain having at least 90% sequence identity to SEQ ID NO: 6; or
   b. a light chain having at least 90% sequence identity to SEQ ID NO: 7 and a heavy chain having at least 90% sequence identity to SEQ ID NO: 8.

21. The antibody of claim 20 further comprising that the signal sequence has at least 90% sequence identity to any one of the amino acid sequences of SEQ ID NO: 9-20 or 33-47 or 63-78.

22. An Aurantiocystrium cell comprising an expression cassette that encodes an exogenous protein and a functional signal sequence that directs the secretion of the heterologous protein from the cell.

23. The cell of claim 22 wherein the heterologous protein is a functional antibody.

24. The cell of claim 23 wherein the antibody has a light chain and a heavy chain.

25. The cell of claim 24 comprising a functional antibody.
26. The cell of any one of claims 22-25 wherein the cell produces at least 5 mg/L of heterologous protein in a 24 hour period.

27. The cell of any one of claims 22-26 wherein the expression cassette encodes:
   a. a promoter active in *Aurantiochytrium*;
   b. a terminator sequence active in *Aurantiochytrium*;
   c. the heterologous protein operably linked to the promoter;
   d. a signal sequence having at least 90% sequence identity to at least 40 contiguous amino acids of a sequence selected from any one of SEQ ID NO: 9-20 or 23-47 or 63-78.

28. The cell of any one of claims 22-26 wherein the signal sequence is encoded by a nucleotide sequence having at least 90% sequence identity to a nucleotide sequence of at least 300 contiguous nucleotides selected from any one of SEQ ID NO: 21-32 or 48-62 or 79-94.

29. A method of producing a heterologous protein comprising:
   a. cultivating a *Labyrinthulomyces* cell that secretes a heterologous protein and that comprises an expression cassette encoding the heterologous protein and further encoding a functional signal sequence that directs the secretion of the heterologous protein from the cell, wherein the signal sequence has at least 90% sequence identity to a sequence of at least 100 contiguous amino acids selected from any one of SEQ ID NO: 9-20 or 33-47 or 63-78; and
   b. harvesting the heterologous protein.

30. The method of claim 29 wherein the heterologous protein is a therapeutic protein.

31. The method of claim 29 wherein the heterologous protein is an antibody.

32. The method of claim 31 wherein the antibody has an amino acid sequence selected from the group consisting of: eculizumab, natalizumab, cetuximab, omalizumab, uestinumab, panitumumab, denosumab, trastuzumab and adalimumab.

33. The method of any one of claims 29-31 wherein the expression cassette encodes a promoter active in *Labyrinthulomyces* and that regulates transcription of the heterologous protein.

34. The method of claim 33 wherein the promoter is selected from the group consisting of: tubulina promoter (tubap) or translation elongation factor (TEFp) promoter.

35. The method of any one of claims 29-34 wherein the expression cassette further encodes a terminator selected from pgk1t or eno2t.
36. The method of claim 31 wherein the antibody has at least 90% sequence identity to an amino acid sequence of at least 100 contiguous amino acids selected from the group consisting of SEQ ID NO: 5-8.

37. The method of claim 29 wherein the heterologous protein comprises a bispecific T cell engager.

38. The method of claim 29 wherein the heterologous protein comprises a binding molecule that binds to an antigen selected from the group consisting of: anti-CD19, anti-CD3, anti-CD38, and prostate-specific membrane antigen (PSMA).

39. A therapeutic protein comprising an antibody molecule linked to a signal sequence selected from the group consisting of SEQ ID NOs: 9-20 or 33-47 or 63-78.

40. The therapeutic protein of claim 39 wherein the antibody is anti-TNF-alpha or anti-HER2.
FIGURE 1
FIGURE 2
FIGURE 3
FIGURE 5
Trastuzumab Binding of Recombinant Her2 (ELISA)

FIGURE 6
ADCC Reporter Response to Anti-HER2 on BT474 Cells

FIGURE 7
Anti-TNF-Alpha Induction of Cytotoxicity in L929 Cells

FIGURE 8
FIGURE 9

pCAB-SP-bi-specific T cell engager 8,790bp

TEF Promoter

P(BLA)

AP(R)

CYC1 terminator

nptll

tubulin promoter

ENO2 terminator

Signal Peptide

anti-CD19/anti-CD3 bi-specific T cell engager
**FIGURE 11**

Kappa/lambda bispecific anti-IL6RC/anti-IFNγ antibody

<table>
<thead>
<tr>
<th>WH-06267</th>
<th>B22</th>
<th>B9</th>
<th>B5</th>
<th>B4</th>
<th>A29</th>
<th>A28</th>
<th>A9</th>
<th>A4</th>
<th>2</th>
</tr>
</thead>
</table>

HC

Kappa

lambda

**FIGURE 12**

Hybrid bispecific anti-CD38/anti-CD3 antibody

<table>
<thead>
<tr>
<th>WH-06267</th>
<th>B13</th>
<th>B3</th>
<th>A32</th>
<th>A27</th>
<th>A19</th>
<th>A10</th>
<th>A8</th>
<th>A4</th>
</tr>
</thead>
</table>

HC2

HC1

LC
FIGURE 13