NOVEL BACTERIAL GENES AND PROTEINS THAT ARE ESSENTIAL FOR CELL VIABILITY AND THEIR USES

Throughout this application various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

FIELD OF THE INVENTION

The present invention relates generally to nucleotide sequences, and polypeptides encoded by the sequences, that are essential for bacterial viability, and to methods of using the nucleotide and polypeptide sequences.

BACKGROUND OF THE INVENTION

Bacterial genera, such as *Streptococcus, Staphylococcus, Pseudomonas, Yersinia, Salmonella,* and *Enterobacter*, are the cause of numerous afflictions in humans and animals. Bacterial infection can lead to serious health conditions, including pneumonia, osteomyelitis, meningitis, sinusitis, otitis, cystitis, and even food poisoning. Typically, these infections can be treated with standard antimicrobial agents such as antibiotics. However, the emergence of pathogenic bacterial strains that are resistant to antibiotics has risen alarmingly in the past two decades. This situation has created an urgent need for the development of new antimicrobial agents.

One strategy for developing new antimicrobial agents is to identify bacterial gene sequences that encode gene products that are essential for bacterial cell viability and

develop and/or identify agents which inhibit the function of the gene product. DNA sequencing technology has advanced from sequencing one gene at a time to sequencing entire genomes, the sum of all genes in an organism. With the recent arrival of bacterial genomic information, it is now possible to compare multiple bacterial genomes in an attempt to identify genes that encode conserved gene products. In this manner, one skilled in the art may identify a set of conserved bacterial genes, including a subset of genes that are essential for bacterial cell viability. The essential gene is then used as a starting point to develop therapeutic agents that inhibit or inactivate the product of the essential gene.

The availability of DNA sequence information for multiple microbial genomes is a recent development. The public release of the first complete genome, *Haemophilus influenzae* (Fleischmann, R.D., et al. 1995 *Science* 269:496-512), was followed in rapid succession by a number of public and private genome sequencing programs. Presently, some 20 completely sequenced bacterial genomes have been published, and over 100 other sequencing projects are underway (Blattner, F.R., et al., 1997 *Science* 277:1453-74; Ferretti, J.J., et al., 1997 *Adv Exp Med Biol* 418:961-963; Koonin, E.V., et al., 1996 *Methods Enzymol* 266:295-322). Analyses of these data indicate that approximately 46% of putative bacterial genes are of unknown function having no attributable function.

Others have pursued various strategies to identify bacterial genes that are essential for viability. These strategies include: identifying genes that are expressed by the bacteria when present in the infected host (Hensel, M., et al., 1995 *Science* 269:400-3), identifying essential genes by isolating temperature sensitive mutants (Schmid, M.B., et al., 1998 *Curr Opin Chem Biol* 2:529-34), and identifying genes in pathways known from prior physiological studies to be essential (Skarzynski, T. et al., 1996 *Structure* 1996 4:1465-74)

There continues to be a need to identify bacterial genes that encode gene products that are essential for cell viability, such as cell replication, growth, and survival. These genes and their encoded gene products can be used as a starting point towards identifying agents

that inhibit functions essential for cell viability, thereby causing bacterial cell stasis or death (e.g., antibacterial agents).

The present invention provides experimental identification of novel, conserved essential genes (*ceg*) from bacteria and their encoded protein products. The *ceg* genes are considered essential to cell viability because disruption of an endogenous *ceg* gene results in lethality of a bacterial cell (e.g., as determined by failure to recover viable chloramphenicol-resistant colonies, as described herein). Thus, the gene products encoded by these genes are potentially valuable targets for chemotherapeutic intervention of bacterial infections.

The *ceg* nucleotide sequences of the invention were obtained by large-scale computational comparisons of multiple genome sequences to identify conserved protein coding regions, followed by gene disruption to identify *ceg*s. The conservation of protein sequences in many cases is believed to reflect the higher level conservation of common biochemical pathways essential for bacterial function and viability.

SUMMARY OF THE INVENTION

The acronyms "CEG" and "*ceg*" stand for Conserved Essential Gene. For convenience, the italicized term *ceg* refers herein to *ceg* nucleotide sequences. The capitalized term CEG refers herein to CEG polypeptide sequences.

Embodiments of the *ceg* nucleotide sequences and the CEG polypeptide sequences are designated CFEs which stands for CEG For Expression. The CFEs are polypeptides resulting from expression of the *ceg* nucleotide sequence.

The present invention provides isolated nucleotide sequences of conserved essential genes from bacteria, designated *ceg*. The invention also provides recombinant nucleic acid molecules including the *ceg* sequences of the invention, and methods of uses thereof. Examples of nucleic acid molecules having *ceg* sequences are described in SEQ ID

NOS.: 1-113. The invention further provides isolated polypeptides and recombinant polypeptides having the CEG sequences of the invention, and methods of uses thereof. Examples of polypeptides having CEG sequences are described in SEQ ID NOS.:114-226.

The *ceg* sequences of the present invention are DNA or RNA. Further, the invention includes nucleic acid molecules that are identical or nearly identical (e.g., similar) with the *ceg* sequences of the invention. The invention additionally provides polynucleotide sequences that hybridize under stringent conditions to the *ceg* sequences of the invention. A further embodiment provides polynucleotide sequences which are complementary to the *ceg* sequences of the invention. Yet another embodiment provides *ceg* nucleic acid molecules that are labeled with a detectable marker. Another embodiment provides recombinant nucleic acid molecules, such as a vector or a fusion molecule, including the *ceg* sequences of the invention.

The present invention provides various *ceg* sequences, fragments thereof having essential gene activity, and related molecules such as antisense molecules, oligonucleotides, peptide nucleic acids (PNA), fragments, and portions thereof.

The present invention relates to the inclusion of the polynucleotides encoding CEG gene products, such as CEG polypeptides, in an expression vector which can be used to transform host cells or organisms. Such transgenic hosts are useful for the production of CEG gene products for the development of antibacterial agents such as antibiotics.

The invention further provides substantially purified CEG gene products, and uses thereof.

The invention also relates to pharmaceutical compositions comprising antisense molecules capable of disrupting expression of *ceg* sequences, agonists, antagonists or inhibitors of CEG gene products, and antibodies reactive against the CEG polypeptides.

These compositions are useful for preventing the growth or survival of bacteria, for example, in the treatment of conditions associated with bacterial infections.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: A schematic representation of the gene disruption assay, as described in Example 3, *infra*. A) A recombinant vector undergoing homologous recombination with the host genome. B) The result of homologous recombination.

Figure 2: A schematic representation of the polarity test for operons, as described in Examples 2 and 3, *infra*. A) The recombinant vector undergoing homologous recombination with the host genome. B) Case 1: one possible result of homologous recombination; the downstream Gene B has an independent promoter. C) Case 2; another possible result of homologous recombination; the downstream Gene B does not have an independent promoter.

Figure 3: Purification of 2CFE 75, as described in Example 6, *infra*. A) Fractionation profile of 2CFE 75 eluted from a Ni-NTA column. B) Gel electrophoresis of pooled fractions of CFE 75. C) Non-denaturing gel electrophoresis to determine oligo form of 2CFE 75.

Figure 4: Fractionation profile of 2CFE 3 eluted from a hydroxyapatite column, as described in Example 7, *infra*.

Figure 5: The biosynthesis pathway of Coenzyme A which starts with phosphorylation of pantothenate.

Figure 6: Circular dichroism spectra of 2CFE 101 and 103, as described in Example 10, *infra*. A) Circular dichroism spectra of 2CFE 101 and 103 at 25 degrees C. B) Circular dichroism thermal melt spectra of 2CFE 101 and 103 at a range of zero to 100 degrees C.

Figure 7: Circular dichroism spectra of aggregate and monomer pools of 2CFE 101 and 103, as described in Example 10, *infra*. A) Circular dichroism spectra of aggregate and monomer pools of 2CFE 101 and 103 at 25 degrees C. B) Circular dichroism thermal melt spectra of aggregate and monomer pools of 2CFE 101 and 103 at a range of zero to 100 degrees C.

Figure 8: Absorbance spectra of pantothenate-dependent production of ADP, as described in Example 10, *infra*.

Figure 9: The results of size exclusion chromatography and gel electrophoresis showing the oligomeric forms of 2CFE 21 and 39, as described in Example 11, *infra*. Lanes 1-6 contain 2CFE 21, lane 7 is a molecular weight marker, lanes 8-10 contain 2CFE 39.

Figure 10: Gel electrophoresis of a helicase reaction using 2CFE 21 and 39 and radiolabeled synthetic Holliday Junction template, as described in Example 11, *infra*. Lane 1 contains the synthetic Holliday Junction template; lane 2 contains the synthetic duplex; lane 3 contains a single-stranded template; lane 4 contains the helicase reaction using 2CFE 39; lane 5 contains the helicase reaction using 2CFE 21; lanes 6-8 contain the helicase reaction using 2CFE 39 and 21 at varying concentrations (e.g., 1, 2, and 3 μM each); and lane 9 contains the helicase reaction using 2 μM each 2CFE 39 and 21 in the presence of ethidium bromide.

Figure 11: A graph depicting the results of the helicase reaction which were monitored by measuring the unquenching of the Holliday Junction templates with time, as described in Example 11, *infra*.

Figure 12: Capillary electrophoresis results of 2CFE 8 with and without ssDNA, as described in Example 12, *infra*. A) Electropherogram of 2CFE 8 alone. B) Electropherogram of 2CFE 8 in the presence of a 32-nucleotide single-stranded oligomer.

Figure 13: Gel mobility shift assay of 2CFE 8, and 2CFE 8 in the presence of a single-stranded 32-mer, as described in Example 12, *infra*. A) An ethidium bromide-stained,

native, polyacrylamide gel containing 2CFE 8, and 2CFE 8 in the presence of a 32-mer. B) The same native, polyacrylamide gel stained with Coomassie.

Figure 14: The N-acetyl glucosamine pathway putatively mediated by 2CFE 3 and 2CFE 86, as described in Example 13, *infra.*

Figure 15: Capillary electrophoresis results of 2CFE 3 with and without putative substrates, as described in Example 13, *infra.*. A) Electropherogram of 2CFE 3 with and without glucosamine-1-phosphate. B) Electropherogram of 2CFE 3 with and without D-glucose-1-phosphate. C) Electropherogram of 2CFE 3 alone, 2CFE 3 and glucose-1-phosphate, and 2CFE 3 and glucose-6-phosphate. D) Electropherogram of 2CFE 3 alone or in the presence of glucosamine-1-phosphate, glucosamine-6-phosphate, D-glucose, D(+) galactose, and α-D-glucose-1-phosphate.

Figure 16: Capillary electrophoresis results of FITC-derivitized 2CFE 3 polypeptide with and without D-glucosamine-6-phosphate (substrate) to produce the product D-glucosamine-1-phosphate, using laser-induced fluorescence, as described in Example 13, *infra.* Electropherogram of D-glucosamine-6-phosphate (putative substrate), 2CFE 3 reacted with D-glucosamine-6-phosphate, and the product glucosamine-1-phhosphate.

Figure 17: Gel electrophoresis of 2CFE 86 eluted from an Ni-NTA column, as described in Example 13, *infra.*

Figure 18: HPLC analysis of a coupled reaction including 2CFE 3, 2CFE 86, and D-glucosamine-6-phosphate to produce the product, UDP-N-acetylglucosamine-1-phosphate (UDPAG), as described in Example 13, *infra.*

Figure 19: A fatty acid biosynthesis pathway.

Figure 20: Size exclusion chromatography to determine the molecular weight and oligomeric form of 2CFE 34, as described in Example 14, *infra*.. Selected eluted samples were sized by gel electrophoresis.

Figure 21: Gel electrophoresis of 2CFE 41 eluted from a Ni-NTA column, as described in Example 15, *infra*.

Figure 22: Capillary electrophoresis results of 2CFE 40, 41, and 46, as described in Example 15, *infra*.

Figure 23: Depicts a schematic diagram of a ligand which binds 2CFE 34. The ligand is 2-phenyl-N-(3 corboxyl-4hydroxyphenyl) azabicyclo [4.3.0] nona-2, 8-diene.

Figure 24: Depicts a schematic diagram of a ligand which binds 2CFE 43. The ligand is N-(3, 5-dinitrobenzyl)-7-trifluoromethyl benza diaza furanolactone.

Figure 25: Depicts a schematic diagram of a ligand which binds 2CFE 43. The ligand is 2-amino (N-para-methylphenyl sulfonamide)-3-phenylpropianic acid.

Figure 26: A nucleic acid sequence of 2CFE1 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 27: A nucleic acid sequence of 2CFE2 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 28: A nucleic acid sequence of 2CFE3 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 29: A nucleic acid sequence of 2CFE4 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 30: A nucleic acid sequence of 2CFE5 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 31: A nucleic acid sequence of 2CFE6 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 32: A nucleic acid sequence of 2CFE7 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 33: A nucleic acid sequence of 2CFE8 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 34: A nucleic acid sequence of 2CFE9 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 35: A nucleic acid sequence of 2CFE10 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 36: A nucleic acid sequence of 2CFE11 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 37: A nucleic acid sequence of 2CFE12 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 38: A nucleic acid sequence of 2CFE13 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 39: A nucleic acid sequence of 2CFE14 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 40: A nucleic acid sequence of 2CFE15 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 41: A nucleic acid sequence of 2CFE16 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 42: A nucleic acid sequence of 2CFE17 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 43: A nucleic acid sequence of 2CFE19 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 44: A nucleic acid sequence of 2CFE21 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 45: A nucleic acid sequence of 2CFE24 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 46: A nucleic acid sequence of 2CFE25 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 47: A nucleic acid sequence of 2CFE26 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 48: A nucleic acid sequence of 2CFE27 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 49: A nucleic acid sequence of 2CFE28 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 50: A nucleic acid sequence of 2CFE29 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 51: A nucleic acid sequence of 2CFE30 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 52: A nucleic acid sequence of 2CFE31 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 53: A nucleic acid sequence of 2CFE32 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 54: A nucleic acid sequence of 2CFE33 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 55: A nucleic acid sequence of 2CFE34 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 56: A nucleic acid sequence of 2CFE35 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 57: A nucleic acid sequence of 2CFE36 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 58: A nucleic acid sequence of 2CFE37 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 59: A nucleic acid sequence of 2CFE38 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 60: A nucleic acid sequence of 2CFE39 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 61: A nucleic acid sequence of 2CFE40 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 62: A nucleic acid sequence of 2CFE41 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 63: A nucleic acid sequence of 2CFE42 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 64: A nucleic acid sequence of 2CFE43 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 65: A nucleic acid sequence of 2CFE44 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 66: A nucleic acid sequence of 2CFE45 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 67: A nucleic acid sequence of 2CFE46 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 68: A nucleic acid sequence of 2CFE47 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 69: A nucleic acid sequence of 2CFE48 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 70: A nucleic acid sequence of 2CFE49 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 71: A nucleic acid sequence of 2CFE50 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 72: A nucleic acid sequence of 2CFE51 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 73: A nucleic acid sequence of 2CFE52 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 74: A nucleic acid sequence of 2CFE53 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 75: A nucleic acid sequence of 2CFE54 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 76: A nucleic acid sequence of 2CFE55 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 77: A nucleic acid sequence of 2CFE56 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 78: A nucleic acid sequence of 2CFE57 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 79: A nucleic acid sequence of 2CFE58 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 80: A nucleic acid sequence of 2CFE59 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 81: A nucleic acid sequence of 2CFE60 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 82: A nucleic acid sequence of 2CFE61 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 83: A nucleic acid sequence of 2CFE62 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 84: A nucleic acid sequence of 2CFE64 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 85: A nucleic acid sequence of 2CFE65 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 86: A nucleic acid sequence of 2CFE66 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 87: A nucleic acid sequence of 2CFE67 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 88: A nucleic acid sequence of 2CFE68 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 89: A nucleic acid sequence of 2CFE69 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 90: A nucleic acid sequence of 2CFE70 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 91: A nucleic acid sequence of 2CFE71 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 92: A nucleic acid sequence of 2CFE72 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 93: A nucleic acid sequence of 2CFE75 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 94: A nucleic acid sequence of 2CFE76 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 95: A nucleic acid sequence of 2CFE78 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 96: A nucleic acid sequence of 2CFE79 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 97: A nucleic acid sequence of 2CFE80 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 98: A nucleic acid sequence of 2CFE81 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 99: A nucleic acid sequence of 2CFE82 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 100: A nucleic acid sequence of 2CFE83 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 101: A nucleic acid sequence of 2CFE84 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 102: A nucleic acid sequence of 2CFE85 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 103: A nucleic acid sequence of 2CFE86 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 104: A nucleic acid sequence of 2CFE87 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 105: A nucleic acid sequence of 2CFE88 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 106: A nucleic acid sequence of 2CFE89 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 107: A nucleic acid sequence of 2CFE90 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 108: A nucleic acid sequence of 2CFE91 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 109: A nucleic acid sequence of 2CFE92 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 110: A nucleic acid sequence of 2CFE94 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 111: A nucleic acid sequence of 2CFE95 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 112: A nucleic acid sequence of 2CFE96 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 113: A nucleic acid sequence of 2CFE97 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 114: A nucleic acid sequence of 2CFE99 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 115: A nucleic acid sequence of 2CFE101 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 116: A nucleic acid sequence of 2CFE102 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 117: A nucleic acid sequence of 2CFE103 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 118: A nucleic acid sequence of 2CFE104 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 119: A nucleic acid sequence of 2CFE105 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 120: A nucleic acid sequence of 2CFE106 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 121: A nucleic acid sequence of 2CFE107 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 122: A nucleic acid sequence of 2CFE108 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 123: A nucleic acid sequence of 2CFE109 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 124: A nucleic acid sequence of 2CFE111 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 125: A nucleic acid sequence of 2CFE112 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 126: A nucleic acid sequence of 2CFE113 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 127: A nucleic acid sequence of 2CFE114 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 128: A nucleic acid sequence of 2CFE115 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 129: A nucleic acid sequence of 2CFE116 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 130: A nucleic acid sequence of 2CFE117 deposited with the American Type Culture Collection as ATCC designation _____ on December 20, 2000.

Figure 131: Schematic structures of alkyloids which are ligands, for example, of 2CFE42.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

All scientific and technical terms used in this application have meanings commonly used in the art unless otherwise specified. As used in this application, the following words or phrases have the meanings specified.

As used herein, a *ceg* nucleic acid molecule is said to be "isolated" when the nucleic acid molecule is substantially separated from contaminant nucleic acid molecules that encode polypeptides other than CEGs. Additionally, isolated nucleic acid molecule refers to any RNA or DNA sequence obtained from a natural source, or constructed by recombinant methods, or synthesized. A skilled artisan can readily employ nucleic acid isolation procedures to obtain an isolated nucleic acid molecule having *ceg* sequences.

The term "*ceg*" includes all isolated forms of *ceg* nucleotide and CEG amino acid sequences disclosed herein. The *ceg* sequences encode gene products that have essential biological functions in bacterial cells, such as, for example, nucleotide biosynthesis, amino acid biosynthesis, DNA replication, RNA transcription, protein translation, DNA recombination, DNA repair, biosynthesis of cofactors (e.g., Coenzyme A), biosynthesis of prosthetic groups, cellular processes (e.g., chaperones, cell division, and polypeptide secretion), energy metabolism (e.g., pentose phosphate pathway, glycolysis, gluconeogenesis), fatty acid biosynthesis, cell wall biosynthesis, and/or biosynthesis of purines, pyrimidines, nucleosides, and nucleotides. Accordingly, the gene products of the *ceg* nucleotide sequences are required for viability of bacterial cells. The term "*ceg*" also includes variants having nucleotide sequence similarity to the disclosed *ceg* sequences,

including sequences isolated from various bacterial genera and species, allelic variants, mutant variants, and *ceg* variants that encode conservative and non-conservative amino acid substitutions. The present invention also provides for all *ceg* sequences generated by recombinant DNA technology, including complementary sequences, *ceg* sequences that hybridize to the sequences of the invention at high stringency hybridization conditions, fusion genes comprising a *ceg* sequence, and codon usage variants.

The term "essential genes" refers to a nucleotide sequence that encodes a gene product having a function which is required for cell viability. The term "essential protein" refers to a polypeptide that is encoded by an essential gene and has a function that is required for cell viability. Accordingly, a mutation that disrupts the function of the essential gene or essential proteins results in a loss of viability of cells harboring the mutation.

"Non-essential genes" or "non-essential proteins" refer to genomic information or the protein(s) or RNAs encoded therefrom which, when disrupted by a mutation, do not result in a loss of viability of cells harboring said mutation under defined laboratory conditions.

As used herein, a nucleotide sequence is said to be "identical" to another reference sequence when both nucleotide sequences are exactly alike.

As used herein, a nucleotide sequence is said to be "similar" to another reference sequence when a comparison of the two sequences shows that they have a low level of sequence differences. For example, two sequences are considered to be similar to each other when the percentage of nucleotides that are shared between the two sequences is between about 70 % to 99.99% over the entire length of the two sequences.

As used herein an amino acid sequence is said to be "similar" to another reference sequence when a comparison of the two sequences shows that they have a low level of sequence differences. For example, two sequences are considered to be similar to each

other when the percentage of amino acids that are shared between the two sequences may be between about 30% to 100% identity over the entire length of the two sequences.

As used herein, an "allele" or "allelic sequence" is an alternative form of the naturally-occurring *ceg* sequence. Alleles result from a mutation, that changes the nucleotide sequence, and generally produce altered mRNAs or polypeptides whose structure or function may or may not be altered.

"Substantially purified" as used herein means a specific isolated nucleic acid or protein, or fragment thereof, in which substantially all contaminants (i.e. substances that differ from said specific molecule) have been separated from said nucleic acid or protein.

In a host cell, an "endogenous" sequence as used herein means a nucleic acid sequence that is naturally-occurring and resides within the host genome.

In a host cell, an "exogenous" sequence as used herein means an isolated nucleic acid sequence that is introduced into the host cell, using any one of a variety of introduction methods, such as transfection, electroporation, cationic lipid or salt treatment methods.

"Knockout mutant" or "knockout mutation" as used herein refers to an *in vitro* engineered disruption of a region of endogenous chromosomal DNA (e.g., disruption of the genome), typically within a protein coding region. A knockout mutation can be generated by inserting an exogenous DNA sequence into the homologous endogenous sequence. A knockout mutation occurring in a protein coding region is expected to disrupt normal expression of the protein coding region. This usually leads to loss of the function provided by the protein.

In order that the invention herein described may be more fully understood, the following description is set forth.

A) MOLECULES OF THE INVENTION

1.) CEG NUCLEIC ACID MOLECULES

The present invention provides isolated and recombinant *ceg* nucleic acid molecules and fragments thereof, and related molecules, such as sequences complementary to *ceg* sequences or a portion thereof, and those that hybridize to the nucleic acid molecules of the invention.

The *ceg* polynucleotide sequences, also referred to herein as nucleic acid molecules of the invention, are preferably in isolated form, including DNA, RNA, DNA/RNA hybrids, and related molecules, and fragments thereof. Specifically contemplated are genomic DNA, ribozymes, and antisense molecules, as well as nucleic acid molecules based on an alternative backbone or including alternative bases, whether derived from natural sources or synthesized. Embodiments of particular *ceg* polynucleotide and amino acid sequences include, but are not limited to, the sequences described in Tables I and II (e.g., SEQ ID NOS:1-113, 114-226 and SEQ ID NOS: 227-339, 340-452, respectively). The *ceg* polynucleotide and amino acid sequences were designated *cfe* which stands for CEG For Expression.

Biological samples of the 2CFE nucleic acid molecules (e.g., SEQ ID NOS: 227-331) were deposited on December 20, 2000 with the American Type Culture Collection (ATCC), 10801 University Blvd., Manassas, VA 20110-2209.

TABLE I

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	POLARITY
CFE 1	1	114	+
CFE 2	2	115	-
CFE 3	3	116	-
CFE 4	4	117	+
CFE 5	5	118	-
CFE 6	6	119	+

TABLE I

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	POLARITY
CFE 7	7	120	-
CFE 8	8	121	+
CFE 9	9	122	+
CFE 10	10	123	+
CFE 11	11	124	+
CFE 12	12	125	+
CFE 13	13	126	-
CFE 14	14	127	+
CFE 15	15	128	-
CFE 16	16	129	-
CFE 17	17	130	-
CFE 19	18	131	+
CFE 21	19	132	-
CFE 24	20	133	-
CFE 25	21	134	+
CFE 26	22	135	-
CFE 27	23	136	+
CFE 28	24	137	-
CFE 29	25	138	-
CFE 30	26	139	-
CFE 31	27	140	+
CFE 32	28	141	+
CFE 33	29	142	-
CFE 34	30	143	+
CFE 35	31	144	+
CFE 36	32	145	+
CFE 37	33	146	-
CFE 38	34	147	+
CFE 39	35	148	-
CFE 40	36	149	-
CFE 41	37	150	-
CFE 42	38	151	-
CFE 43	39	152	-
CFE 44	40	153	+
CFE 45	41	154	-
CFE 46	42	155	-

TABLE I

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	POLARITY
CFE 47	43	156	-
CFE 48	44	157	
CFE 49	45	158	+
CFE 50	46	159	+
CFE 51	47	160	+
CFE 52	48	161	-
CFE 53	49	162	+
CFE 54	50	163	+
CFE 55	51	164	+
CFE 56	52	165	+
CFE 57	53	166	+
CFE 58	54	167	+
CFE 59	55	168	-
CFE 60	56	169	+
CFE 61	57	170	+
CFE 62	58	171	
CFE 63	59	172	
CFE 64	60	173	+
CFE 65	61	174	+
CFE 66	62	175	+
CFE 67	63	176	+
CFE 68	64	177	-
CFE 69	65	178	+
CFE 70	66	179	+
CFE 71	67	180	-
CFE 72	68	181	-
CFE 73	69	182	+
CFE 74	70	183	-
CFE 75	71	184	-
CFE 76	72	185	+
CFE 77	73	186	
CFE 78	74	187	+
CFE 79	75	188	-
CFE 80	76	189	-
CFE 81	77	190	+
CFE 82	78	191	

TABLE I

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	POLARITY
CFE 83	79	192	-
CFE 84	80	193	-
CFE 85	81	194	-
CFE 86	82	195	-
CFE 87	83	196	-
CFE 88	84	197	-
CFE 89	85	198	+
CFE 90	86	199	+
CFE 91	87	200	-
CFE 92	88	201	-
CFE 93	89	202	+
CFE 94	90	203	+
CFE 95	91	204	+
CFE 96	92	205	+
CFE 97	93	206	-
CFE 98	94	207	
CFE 99	95	208	+
CFE 100	96	209	
CFE 101	97	210	-
CFE 102	98	211	+
CFE 103	99	212	-
CFE 104	100	213	+
CFE 105	101	214	-
CFE 106	102	215	-
CFE 107	103	216	-
CFE 108	104	217	+
CFE 109	105	218	-
CFE110	106	219	-
CFE 111	107	220	-
CFE 112	108	221	-
CFE 113	109	222	-
CFE 114	110	223	-
CFE 115	111	224	-
CFE 116	112	225	-
CFE 117	113	226	-

TABLE II

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	FIGURE
2CFE 1			26
2CFE 2			27
2CFE 3			28
2CFE 4			29
2CFE 5			30
2CFE 6			31
2CFE 7			32
2CFE 8			33
2CFE 9			34
2CFE 10			35
2CFE 11			36
2CFE 12			37
2CFE 13			38
2CFE 14			39
2CFE 15			40
2CFE 16			41
2CFE 17			42
2CFE 19			43
2CFE 21			44
2CFE 24			45
2CFE 25			46
2CFE 26			47
2CFE 27			48
2CFE 28			49
2CFE 29			50
2CFE 30			51
2CFE 31			52
2CFE 32			53
2CFE 33			54
2CFE 34			55
2CFE 35			56
2CFE 36			57
2CFE 37			58
2CFE 38			59
2CFE 39			60

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	FIGURE
2CFE 40			61
2CFE 41			62
2CFE 42			63
2CFE 43			64
2CFE 44			65
2CFE 45			66
2CFE 46			67
2CFE 47			68
2CFE 48			69
2CFE 49			70
2CFE 50			71
2CFE 51			72
2CFE 52			73
2CFE 53			74
2CFE 54			75
2CFE 55			76
2CFE 56			77
2CFE 57			78
2CFE 58			79
2CFE 59			80
2CFE 60			81
2CFE 61			82
2CFE 62			83
2CFE 64			84
2CFE 65			85
2CFE 66			86
2CFE 67			87
2CFE 68			88
2CFE 69			89
2CFE 70			90
2CFE 71			91
2CFE 72			92
2CFE 75			93
2CFE 76			94
2CFE 78			95
2CFE 79			96
2CFE 80			97

CFE Designation	SEQ. ID NO. (Nucleotide)	SEQ. ID NO. (Polypeptide)	FIGURE
2CFE 81			98
2CFE 82			99
2CFE 83			100
2CFE 84			101
2CFE 85			102
2CFE 86			103
2CFE 87			104
2CFE 88			105
2CFE 89			106
2CFE 90			107
2CFE 91			108
2CFE 92			109
2CFE 94			110
2CFE 95			111
2CFE 96			112
2CFE 97			113
2CFE 99			114
2CFE 101			115
2CFE 102			116
2CFE 103			117
2CFE 104			118
2CFE 105			119
2CFE 106			120
2CFE 107			121
2CFE 108			122
2CFE 109			123
2CFE 111			124
2CFE 112			125
2CFE 113			126
2CFE 114			127
2CFE 115			128
2CFE 116			129
2CFE 117			130

a) Variant *ceg* Nucleotide Sequences

The present invention also provides nucleic acid molecules having a nucleotide sequence substantially identical or similar to the *ceg* sequences (SEQ ID NOS: 1-113, 227-331) disclosed herein.

The present invention provides nucleotide sequences which are similar to SEQ ID NOS:1-113 and/or SEQ ID NOS:227-331. The present invention provides nucleotide sequences which vary from SEQ ID NOS:1-113 or 227-331 by a range of about 1% to about 70%.

The present invention encompasses variations in polynucleotide sequences resulting from mutations and/or from transfer of genetic material from one cell to another (e.g., horizontal gene transfer or horizontal gene exchange).

The present invention also provides for variants of the polynucleotide *ceg* sequences disclosed herein, including variants isolated from naturally-occurring sources, those generated by recombinant DNA technology or other in vitro synthesis methodologies (e.g., PCR). The variant polynucleotide sequences of the invention encode polypeptides that exhibit the biological activity of naturally-occurring CEG polypeptides, such as activity required for bacterial cell viability.

In general, for example, a variant of *ceg* polynucleotide sequences may encode a polypeptide that differs by one or more amino acid substitutions. The variant may have conservative changes, wherein a substituted amino acid has similar structural or chemical properties, eg, replacement of leucine with isoleucine.

A polynucleotide sequence can encode conservative amino acid substitutions without altering either the conformation or the function of the polypeptide. Such changes include substituting any of isoleucine (I), valine (V), and leucine (L) for any other of these

hydrophobic amino acids; aspartic acid (D) for glutamic acid (E) and vice versa; glutamine (Q) for asparagine (N) and vice versa; and serine (S) for threonine (T) and vice versa. Other substitutions can also be considered conservative, depending on the environment of the particular amino acid and its role in the three-dimensional structure of the protein. For example, glycine (G) and alanine (A) can frequently be interchangeable, as can alanine (A) and valine (V). Methionine (M), which is relatively hydrophobic, can frequently be interchanged with leucine and isoleucine, and sometimes with valine. Lysine (K) and arginine (R) are frequently interchangeable in locations in which the significant feature of the amino acid residue is its charge and the differing pK's of these two amino acid residues are not significant. Still other changes can be considered "conservative" in particular environments.

A variant may also have nonconservative changes, eg, replacement of a glycine with a tryptophan. Other variations may also include amino acid deletions or insertions, or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

Another type of *ceg* sequence variant includes naturally-occurring allelic variants of *ceg* which share significant similarity (e.g., between about 30- 99%) to the disclosed CEG polypeptide sequence. Allelic variants of the *ceg* sequences can encode conservative or non-conservative amino acid substitutions of the CEG polypeptide sequence herein described.

An example of allelic variants of *ceg* are mutant alleles of *ceg* polynucleotide sequences that encode a polypeptide having one or more changes in the polypeptide sequence, such as amino acid substitutions, deletions, insertions, frame shifts, or truncations. The mutant alleles of *ceg* may or may not encode a CEG polypeptide having the same biological functions as wild-type CEG proteins.

Variations in the bacterial genomic sequences can also arise from transfer of genetic material to another bacterial cell. The transfer of gene sequences can occur intraspecies or interspecies. Gene transfer can occur between bacterial cells which are members of the same or different populations. A population includes, but is not limited to, a serotype isolate, a clinical isolate, a naturally-occurring isolate, a strain, and a species. The transfer of genetic material can occur between cells within a population; for example transfer between serotype A to serotype A, or between *S. pneumoniae* and *S. pneumoniae*. The transfer of genetic material can occur between cells of different populations; for example, between serotype A to serotype B or *S. pneumoniae* and *S. mutans*.

Gene transfer can give rise to mutant or polymorphic variant genes sequences. In rare cases, gene transfer introduces new gene sequences that confer a new phenotype, such as antibiotic resistance. The transfer of genetic material includes transfer of large regions of genomic sequences which include partial gene sequences, whole single gene sequences, or multiple gene sequences. This mode of transfer can give rise to replacement of native whole gene sequences or introduction of new sequences in the recipient cell. This mode of transfer gives rise to mosaic gene sequences in the recipient cell.

The variation of genomic sequences resulting from gene transfer can be examined using molecular techniques, including: multilocus enzyme electrophoresis (Selander. R. K., et al., 1986 *Appl. Environ. Microbiol.* 51:837-884); and restriction endonuclease cleavage electrophoretic profiling (Coffey, T. J., et al., 1991 *Mol. Microbio.* 5:2255-2260); pulse-field gel electrophoresis fingerprinting (Bygraves, J. A. and Maiden, M. C. J. 1992 *J. Gen. Microbiol.* 138:523-531); and ribotyping (Stull, T. L., et al., 1988 *J. Infect. Dis.* 157:280-286). The degree of variation can vary greatly, and ranges from little or no variation as exemplified by gene sequences of *E. coli* (Caugant, d. A., et al., 1981 *Genetics* 98:467-490; Whittam, T. S., et al., 1983 *Mol. Biol. Evol.* 1:67-83; Souza, V., et al., 1992 *Proc. Natl. Acad. Sci. USA* 89:8389-8393) and *Salmonella* (Selander, R. K., et al., 1990 *Infect. Immun.* 58:2262-2275; Selander, R.K. and Smith, N. H. 1990 *Rev. Med. Microbiol.* 1:219-228; Smith, J. M., et al., 1993 *Proc. Natl. Acad. Sci. USA* 90:4384-

4388), to extensive gene transfer in *Neisseria gonorrhoeae* (Smith, J. M., et al., 1993 *Proc. Natl. Acad. Sci. USA* 90:4384-4388).

Gene transfer can be examined between various isolates of a particular microbial species which are antibiotic-sensitive or antibiotic-resistent (Coffey, T. J., et al., 1991 *Molec. Microbiol.* 5:2255-2260). Molecular biology techniques can be utilized to study the degree of transfer between populations, such as, for example, the degree of gene transfer between serotypes, isolates, strains, or species . The degree of transfer can be examined by comparing, for example, the penicillin binding proteins and numerous different loci which encode metabolic enzymes or capsular biosynthesis enzymes.

For example, intra-species, inter-serotype, gene transfer is possible (Coffey, T. J., et al., 1991 *supra*). Additionally, intraspecies gene transfer in *S. pneumoniae* (Coffey, T. J., et al., 1998 *Mol. Microbiol.* 27:73-83), *Vibrio cholerae* (Bik, E. M., et al., 1995 *EMBO J.* 14:209-216), and *Haemophilus influenzae* (Kroll, J. S. and Moxon, E. R. 1990 *J. Bacteriol.* 172: 1374-1379) are possible.

Interspecies gene transfer is also possible (Dowson, C. G., et al., 1989 *Proc. Natl. Acad. Sci. USA* 86:8842-8846; Laibl, G., et al., 1991 *Mol. Microbiol.* 5:1993-2002; Bourgoin, F., et al., 1999 *Gene* 233:151-161).

Variant gene sequences arising from gene transfer can be continually generated in transformable bacteria (e.g., transformation competent), such as *S. pneumoniae*. For example, the worldwide spread of varying degrees of antibiotic resistance has been documented and reviewed (Dowson, C. G., et al., 1994 *Trends Microbiol.* 2:361-366; Spratt, B. G. in *Bacterial Cell Wall*, eds Ghuysen J-M. and Hakenbeck, R. 1994 pp. 517-534; and reviewed in Maiden, M. C. J. 1998 *Clinic. Infect. Dis.* 27 (Supplement 1) S12-S20). For example, variant gene sequence arising from gene transfer can be tracked using a marker gene such as the gene which encodes the penicillin binding protein (Barcus, V. A., et al., 1995 *FEMS Microbiol. Lett.* 126:299-303). At the nucleotide level, gene sequences encoding the penicillin binding proteins in susceptible and resistant

strains differ by about 14% to 23% (Hakenbeck, R. 1995 *Biochem. Pharmacol.* 50:1121-1127; Spratt, B. G. in *Bacterial Cell Wall*, eds Ghuysen J-M. and Hakenbeck, R. 1994 pp. 517-534; Spratt, B. G., et al., 1991 *Neisseria meningitidis* and *Streptococcus pneumoniae* eds. Camisi, J., et al., pp. 73-83; Coffey, T. J., et al., 1995 *Micro. Drug Resist.* 1:29-34).

The *ceg* nucleotide sequences can be isolated from various species of *Streptococcus* including *Streptococcus pneumoniae*. Additionally, the *ceg* sequences can be isolated from other Steptococcal species, including *S. mutans, S. pyogenes,* and *S. thermophila,* The *ceg* polynucleotide sequences can also be isolated from strains of other bacterial genera including, but not limited to, *Streptococcus, Escherichia, Bacillus, Pseudomonas, Yersinia, Salmonella,* and *Haemophilus*.

The present invention additionally provides isolated codon-usage variants that differ from the disclosed *ceg* nucleotide sequences, yet do not alter the predicted CEG polypeptide sequence or function. The codon-usage variants may be generated by recombinant DNA technology. Codons may be selected to optimize the level of production of the *ceg* transcript or CEG polypeptide in a particular prokaryotic or eukaryotic expression host, in accordance with the frequency of codon utilized by the host cell. Alternative reasons for altering the nucleotide sequence encoding a CEG polypeptide include the production of RNA transcripts having more desirable properties, such as an extended half-life or increased stability. A multitude of variant *ceg* nucleotide sequences that encode the respective CEG polypeptide may be isolated, as a result of the degeneracy of the genetic code. Accordingly, the present invention contemplates selecting every possible triplet codon to generate every possible combination of nucleotide sequences that encode the disclosed CEG polypeptides. This particular embodiment provides isolated nucleotide sequences that vary from the sequences as described in SEQ ID NOs.: 1-113 or 227-331, such that each variant nucleotide sequence encodes a polypeptide having sequence identity with the amino acid sequences, as described in SEQ ID NOs.:114-226 or 332-436, respectively.

b) Complementary Sequences

The present invention includes polynucleotide sequences that are complementary to the sequences disclosed herein. The term "complementary" as used herein refers to the capacity of purine and/or pyrimidine nucleotides to associate through hydrogen bonding to form double stranded nucleic acid molecules. The following base pairs are related by complementarity: guanine and cytosine; adenine and thymine; and adenine and uracil. Complementary applies to all base pairs comprising at least two single-stranded nucleic acid molecules.

c) Sequences Capable of Hybridizing

Another embodiment provides nucleic acid molecules that will hybridize to *ceg* sequences under hybridization conditions. It is readily apparent to one skilled in the art that the stringency of the hybridization condition selected will depend upon the characteristics of the nucleic acid molecule to be hybridized, such as, the length, the degree of complementarity (e.g., exact or non-exact complementarity), the percent A/T content, and the objective of the hybridization experiment.

The hybridization procedure may by performed in low stringency hybridization conditions. Low stringency hybridization conditions will permit hybridization between two nucleic acid molecules that differ from exact complementarity by about 25% to 70%. Hybridization under standard high stringency conditions will occur between two complementary nucleic acid molecules (e.g., 100% exact complementarity) or two complementary nucleic acid molecules that differ from exact complementarity by about 1% to about 70%.

The high stringency hybridization conditions that disfavor non-homologous base pairing are well known in the art. Typically, high stringency hybridization conditions, includes but is not limited to, hybridizing at 50 °C to 65 °C in 5X SSPE, and washing at 50 °C to

65 °C in 0.5X SSPE. Typically, low stringency conditions, includes but is not limited to, hybridizing at 35 °C to 37 °C in 5X SSPE and 40% to 45% formamide and washing at 42 °C in 1-2X SSPE. The conditions and formulas for high stringency hybridization methods are well known in the art and can be readily obtained in *Molecular Cloning; A Laboratory Manual* (2nd edition, Sambrook, Fritch, and Maniatis 1989, Cold Spring Harbor Press) or in *Short Protocols in Molecular Biology* (Ausubel, F. M., et al., 1989, John Wiley & Sons).

d) Fragments of *ceg* Sequences

The invention further provides nucleic acid molecules having fragments of the *ceg* sequences, such as a portion of the *ceg* sequence (e.g., SEQ ID NOS:1-113, 227-331) disclosed herein. The size of the fragment will be determined by its intended use. For example, the length of the fragment to be used as a nucleic acid probe or PCR primer is chosen to obtain a relatively small number of false positives during probing or priming. Alternatively, a fragment of the *ceg* sequence may be used to construct a recombinant fusion gene having a *ceg* sequence fused to a non-*ceg* sequence.

The nucleic acid molecules, fragments thereof, and probes and primers of the present invention are useful for a variety of molecular biology techniques including, for example, hybridization screens of libraries, or detection and quantification of mRNA transcripts as a means for analysis of gene transcription and/or expression. Preferably, the probes and primers are DNA. A probe or primer length of at least 15 base pairs is suggested by theoretical and practical considerations (Wallace, B. and Miyada, G. 1987 "Oligonucleotide Probes for the Screening of Recombinant DNA Libraries" in: *Methods in Enzymology*, 152:432-442, Academic Press). Other lengths of fragments, probes, or primers are possible and routine to determine.

The probes and primers of this invention can be prepared by methods well known to those skilled in the art (Sambrook, et al. *supra*). In a preferred embodiment the probes

and primers are synthesized by chemical synthesis methods (ed: Gait, M. J. 1984 *Oligonucleotide* Synthesis, IRL Press, Oxford, England).

One embodiment of the present invention provides nucleic acid primers that are complementary to *ceg* sequences, which allow the specific amplification of nucleic acid molecules of the invention or of any specific parts thereof. Another embodiment provides nucleic acid probes that are complementary for selectively or specifically hybridizing to the *ceg* sequences or to any part thereof.

e) Derivative Nucleic Acid Molecules

The nucleic acid molecules of the invention include peptide nucleic acids (PNAs), or derivative molecules such as phosphorothioate, phosphotriester, phosphoramidate, and methylphosphonate, that specifically bind to single-stranded DNA or RNA in a base pair-dependent manner (Zamecnik, P. C., et al., 1978 *Proc. Natl. Acad. Sci.* 75:280284; Goodchild, P. C., et al., 1986 *Proc. Natl. Acad. Sci.* 83:4143-4146).

PNA molecules comprise a nucleic acid oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen, P. E., et al., 1993 *Anticancer Drug Des* 8:53-63). For example, reviews of methods for synthesis of DNA, RNA, and their analogues can be found in: *Oligonucleotides and Analogues,* eds. F. Eckstein, 1991, IRL Press, New York; *Oligonucleotide* Synthesis, ed. M. J. Gait, 1984, IRL Press, Oxford, England. Additionally, methods for antisense RNA technology are described in U. S. patents 5,194,428 and 5,110,802. A skilled artisan can readily obtain these classes of nucleic acid molecules using the herein described *ceg* polynucleotide sequences, see for example *Innovative and Perspectives in Solid Phase Synthesis* (1992) Egholm, et al. pp 325-328 or U. S. Patent No. 5,539,082.

f) RNA Molecules

The present invention provides RNA molecules that encode the predicted ceg gene products. In particular, the RNA molecules of the invention may be isolated full-length or partial mRNA molecules or RNA oligomers that encode CEG gene products. The RNA molecules of the invention include the nucleotide sequences encoding all or portions of CEGs.

The RNA molecules of the invention also include antisense RNA molecules, peptide nucleic acids (PNAs), or non-nucleic acid molecules such as phosphorothioate derivatives, that specifically bind to the sense strand of DNA or RNA in a base pair-dependent manner. A skilled artisan can readily obtain these classes of nucleic acid molecules using the herein described *ceg* sequences.

g) Labeled Nucleic Acid Molecules

The nucleic acid molecules having *ceg* sequences can be labeled with a detectable marker. Examples of a detectable marker include, but are not limited to, a radioisotope, a fluorescent compound, a bioluminescent compound, a chemiluminescent compound, a metal chelator or an enzyme. Technologies for generating labeled DNA and RNA probes are well known in the art (See e.g. Sambrook et al., *supra*).

2.) RECOMBINANT NUCLEIC ACID MOLECULES

Also provided are recombinant nucleic acid molecules, such as recombinant DNA molecules (rDNAs) that comprise *ceg* sequences or fragments thereof. As used herein, a recombinant DNA molecule is a DNA molecule that has been subjected to molecular manipulation *in vitro*. Methods for generating rDNA molecules are well known in the art, for example, see Sambrook et al., *Molecular Cloning* (1989), *supra*.

a) Vectors

The nucleic acid molecules of the invention may be recombinant molecules each comprising the sequence, or portions thereof, of a *ceg* sequence linked to a non-*ceg* sequence. For example, the *ceg* sequence may be fused operatively to a vector to generate a recombinant molecule. The term vector includes, but is not limited to, plasmids, cosmids, and phagemids. A preferred vector includes an autonomously replicating vector comprising a replicon that directs the replication of the rDNA within the appropriate host cell. The preferred vectors can also include an expression control element, such as a promoter sequence, which enables transcription of the inserted *ceg* sequences and can be used for regulating the expression (e.g., transcription and/or translation) of an operably linked *ceg* sequence in an appropriate host cell such as *Escherichia coli.* Expression control elements are known in the art and include, but are not limited to, inducible promoters, constitutive promoters, secretion signals, enhancers, transcription terminators, and other transcriptional regulatory elements. Other expression control elements that are involved in translation are known in the art, and include the Shine-Dalgarno sequence, and initiation and termination codons. The preferred vector also includes at least one selectable marker gene that encodes a gene product that confers drug resistance such as resistance to ampicillin or tetracyline. The vector also comprises multiple endonuclease restriction sites that enable convenient insertion of exogenous DNA sequences.

The preferred vectors for generating *ceg* transcripts and/or the encoded CEG polypeptides are expression vectors which are compatible with prokaryotic host cells. Prokaryotic cell expression vectors are well known in the art and are available from several commercial sources. For example, a pET vectors (e.g., pET-21, Novagen Corp.) may be used to express CEG polypeptides in bacterial host cells.

b) Recombinant Vectors for Integration

The present invention provides recombinant vectors that may be used to integrate exogenously provided sequences into the genome of a host cell. The recombinant integration vectors of the present invention include a gene that encodes a selectable marker and *ceg* sequences, or fragments thereof. The integration vectors are used to integrate the *ceg* sequence into a target gene sequence that resides within the bacterial host genome (e.g., endogenous sequence), thereby disrupting the function of the target gene sequence within the bacterial cells. These integration vectors may be used in a gene disruption assay to screen candidate *ceg* nucleotide sequences, in order to identify the candidate sequences that encode a gene product that is required for bacterial cell viability.

Accordingly, these recombinant integration vectors include candidate *ceg* sequences that will be screened to determine if the candidate *ceg* sequences encode a gene product that is required for cell viability. The candidate *ceg* sequence that is included as part of the recombinant integration vector is the "exogenous" *ceg* sequence that is employed as the "disrupting" sequence in a gene disruption assay. The *ceg* sequence that resides within the host genome is the "endogenous" or "target" *ceg* sequence.

The integration event rarely occurs, for example, by non-homologous recombination in which a recombinant vector, that includes the exogenous *ceg* sequence, inserts the exogenous *ceg* sequence into a random location within the host genome. In a more preferred embodiment, the integration event inserts the exogenous *ceg* sequence into a specific target site within the host genome. The targeted integration event can involve homologous recombination in which the integration vector, that includes the exogenous *ceg* sequence, inserts the exogenous *ceg* sequence into its homologous target *ceg* sequence that resides within the host's genome (e.g., the endogenous *ceg* sequence) (Figure 1). Further, the exogenous *ceg* sequence can be used as a disrupting sequence whereby the homologous recombination event integrates the exogenous *ceg* sequence into the endogenous target *ceg* sequence resulting in disruption of the function of the

endogenous *ceg* sequence. For example, disrupting the function of the endogenous *ceg* sequence may result in the loss of bacterial cell viability.

An example of a recombinant vector that can be used as an integration vector in *S. pneumoniae* is the pEVP-3 vector (Jean-Pierre Claverys, et al. 1995 *Gene* 164: 123-128). The pEVP-3 vector integrates an exogenous sequence by homologous recombination involving a Campbell-type event (S. Adhya and A. Campbell 1970 *J. Mol. Biol.* 50:481-490). The pEVP-3 vector includes a replicon that functions only in gram-negative bacteria, such as *E. coli.* Therefore, the pEVP-3 vector cannot replicate in *S. pneumoniae*. This vector also contains multiple cloning sites, and confers resistance to chloramphenicol in both a gram-negative and gram-positive bacteria, such as *S. pneumoniae*.

c) Fusion Gene Sequences

A fusion *ceg* gene is another example of a recombinant molecule of the invention. A fusion gene includes a *ceg* sequence operatively fused (e.g., linked) to a non-*ceg* sequence such as, for example, a tag sequence to facilitate isolation and/or purification of the expressed CEG gene product (Kroll, D.J., et al., 1993 *DNA Cell Biol* 12:441-53).

Alternatively, a recombinant fusion molecule has a *ceg* sequence of the invention fused to a *ceg* sequence isolated from a different microbial source. For example, the disclosed *ceg* sequences isolated from *S. pneumoniae* can be fused to a *ceg* sequence isolated from a different bacterial species.

3.) CEG PROTEINS AND POLYPEPTIDE MOLECULES

The invention additionally provides CEG proteins and peptide fragments thereof that are isolated or substantially purified. Embodiments of particular CEG amino acid sequences are disclosed in Tables I and II (SEQ ID NOS:114-226 and SEQ ID NOS:332-436, respectively).

The present invention also includes polypeptides having sequence variations from the predicted CEG polypeptide sequences disclosed herein, including mutant variants, conservative substitution variants, and similar CEG polypeptides from other prokaryotic organisms. For convenience, such proteins are referred to herein as "CEG proteins", "CEG polypeptides", or "proteins of the invention".

As used herein, CEG protein refers to a polypeptide having amino acid sequence identity or similarity to any one of the predicted amino acid sequences, as provided in SEQ ID NO.: 114-226 or 332-436. The variant CEG polypeptides can be allelic forms of CEG, such as mutant forms of CEG polypeptides. The present invention also provides conservative substitution-mutants of the CEG proteins that maintain functional activity of wild-type CEG (e.g., the CEG polypeptide is required for bacterial cell viability).

The CEG protein may be isolated from any source whether natural, synthetic, semi-synthetic, or recombinant. As used herein, "natural" refers to a polypeptide which is found in nature. Accordingly, the CEG proteins may be isolated from a prokaryotic organism, such as a bacterial strain including, but not limited to, *Streptococcus, Escherichia, Bacillus, Pseudomonas, Yersinia, Salmonella,* and *Streptomyces.* The CEG proteins of the invention, and fragments thereof, can also be generated by recombinant methods or chemical synthesis methods.

The CEG polypeptides of the invention are essential for the viability of a bacterial cell. Further, the CEG polypeptides can exhibit at least any one of the following functions: a pantothenate kinase, a Holliday Junction branch migration protein, a single stranded DNA binding protein, a phosphoglucosamine mutase, an acetyltransferase, an uridylyltransferase, a malonyl CoenzymeA:ACP transcylase, a 3-oxoacyl-ACP synthase II, a 3-oxoacyl-ACP reductase, a phosphomethylpyrimidine (HMP-P) kinase, a GTP binding protein, a ATP binding protein, or a 4-aminoimidazole carboxylase. Putative functions can include, but are not limited to, sugar transferase, techoic acid biosynthesis, ribosome recycling factor, response regulator, nicotinate phosphoribosyltransferase,

nitropropane dioxygenase, (3R)-hydroxymyristol acyl carrier protein dehydrase, sugar dehydrogenase, murein biosynthesis, cobalimin biosynthesis, ABC transporter, tRNA modification enzyme, arylsulfatase, 16S processing enzyme, tRNA methyl transferase, elongation factor P, signal recognition particle, protein export, undecaprenol kinase, SRP docking domain, diacyl glycerol kinase, dihydopicilinate reductase, HU-DNA binding protein, thiamine biosynthase, GreA transcription elongation factor, dTDP-L-rhamnose synthase, ATP-binding motif, ribose-5-p-3-epimerase-like activity, GTP pyrophosphokinase, acetyl-CoA carboxylase, O-sialoglycoprotein endopeptidase, glucosamine-fructose-6-phosphase aminotransferase, Strpn adhesion-associated ABC-permease, GTP pyrophosphokinase RelA, IMP dehydrogenase, DNA gyrase subunit B, acetyl-CoA carboxylase subunit AccD, phosphoglycerol kinase, acetyl-CoA carboxylase carbonyl transferase, phosphopanthetheine adenylyltransferase, oligopeptide transport permease subunit, translocation protein, perM permease, DNA pol III gamma and tau subunits, DNA pol III delta subunit, signal peptidase I, acetyl-coA carboxylase biotin carboxyl carrier protein, protein chain release factor-1, replicative DNA helicase, topoisomerase, pentapeptide-transferase, elongation factor G, spore coat polysaccharide biosynthesis protein C, protein release factor B, DNA polymerase III alpha subunit, phosphoprotein phosphatase, chaparonin, UDP-N-acetylmuramoylalanyl-D-glutamate-2, 6-diaminopimelate ligase, techuronic acid biosynthesis, UDP-glucose lipid carrier transferase, transcription termination factor, chromosome segregation factor, amino acid biosynthesis, HMG-CoA reductase, hypoxanthine-guanine phosphoribosyltransferase.

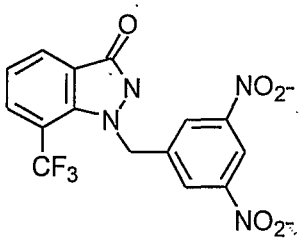
a) MODULATORS OF CEG POLYPEPTIDES

The invention provides compounds that modulate (e.g., activate or inhibit) the function of a CEG polypeptide. Such compounds can provide lead-compounds for developing drugs for diagnosing and/or treating conditions associated with bacterial infections. The modulator is a compound that may alter the function of the CEG polypeptide, such as activating or inhibiting the function of a CEG polypeptide. For example, the compound can act as agonist, antagonist, partial agonist, partial antagonist, cytotoxic agents,

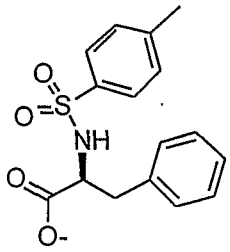
inhibitors of cell proliferation, and cell proliferation-promoting agents. The activity of the compound may be known, unknown or partially known.

Suitable ligands include, but are not limited to, diazalactones, *N*-protected amino acid, azabicyclodiene, and alkaloids.

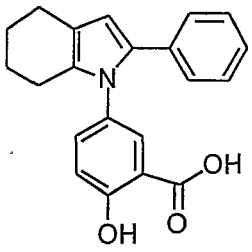
An example of a diazalactone is:



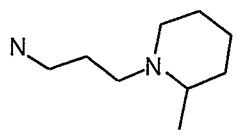
An example of a *N*-protected amino acid is:

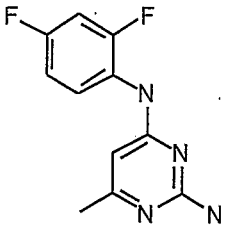


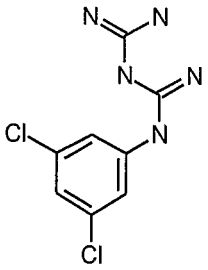
An example of an azabicyclodiene is:

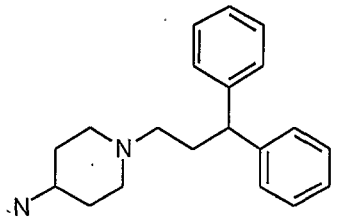


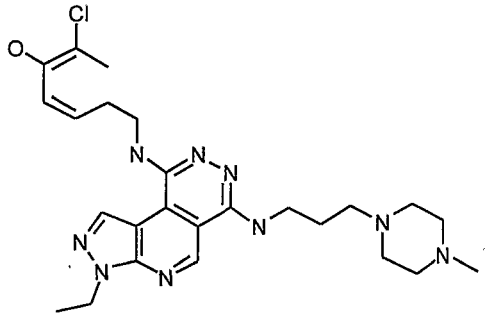
Examples of alkaloids are:











B) METHODS FOR MAKING THE CEG PROTEINS AND POLYPEPTIDES

Recombinant methods are preferred if a high yield is desired. Recombinant methods involve expressing the cloned gene in a suitable host cell. For example, a host cell is introduced with an expression vector having the CEG sequence, then the host cell is cultured under conditions that permit *in vivo* production of the CEG protein. The recombinant vector can integrate the CEG sequence into the host genome. Alternatively, the CEG sequence can be maintained extra-chromosomally, as part of an autonomously replicating vector.

1. HOST-VECTOR SYSTEMS

The invention further provides a host-vector system comprising the vector, plasmid, phagemid, or cosmid comprising a *ceg* nucleotide sequence, or a fragment thereof, introduced into a suitable host cell. The host-vector system can be used to produce the

CEG polypeptides encoded by the *ceg* nucleotide sequences. The host cell can be prokaryotic or eukaryotic. Examples of suitable prokaryotic host cells include bacteria strains from genera such as *Escherichia, Bacillus, Pseudomonas, Streptococcus,* and *Streptomyces*. Examples of suitable eukaryotic host cells include a yeast cell, a plant cell, or an animal cell, such as a mammalian cell. A preferred embodiment provides a host-vector system comprising the pET21 vector having a *ceg* sequence introduced into an *E. coli* λDE3 lysogen which is useful, for example for the production of the CEG protein, herein designated CFE polypeptides and CFE proteins.

Introduction of the rDNA molecules of the present invention into an appropriate cell host is accomplished by well known methods that typically depend on the type of vector used and host system employed. For example, transformation of prokaryotic host cells by electroporation and salt treatment methods are typically employed, see for example, Cohen et al., 1972 *Proc Acad Sci USA* 69:2110; Maniatis, T., et al., 1989 *Molecular Cloning, A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of vertebrate cells with vectors containing rDNAs, electroporation, cationic lipid or salt treatment methods are typically employed, see, for example, Graham *et al.,* 1973 *Virol* 52:456; Wigler et al., 1979 *Proc Natl Acad Sci USA* 76:1373-76.

Successfully transformed cells, i.e., cells that contain a rDNA molecule of the present invention, can be identified by well known techniques. For example, cells resulting from the introduction of a rDNA of the present invention can be selected and cloned to produce single colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a method such as that described by Southern, *J Mol Biol* (1975) 98:503, or Berent et al., *Biotech* (1985) 3:208, or the proteins produced from the cell assayed via a biochemical assay or immunological method.

Procaryotes are generally used as host cells for cloning and producing the products of exogenous DNA sequences. For example, the *Escherichia coli* K12 BL21 (λDE3) (Novagen) is particularly useful for expression of foreign proteins. Other strains of *E. coli*, and bacilli such as *Bacillus subtilis*, Enterobacteriaceae such as *Salmonella*

typhimurium or *Serratia marcescans*, various *Pseudomonas, Streptococcus,* and *Streptomyces* species may also be employed as host cells in cloning and expressing the recombinant proteins of this invention.

In general terms, the production of recombinant CEG proteins may involve using a host/vector system, or other methods may be used. The host/vector system may employ the following steps.

A nucleic acid molecule is obtained that encodes a CEG protein or a fragment thereof, such as any one of the polynucleotides disclosed in SEQ ID NOs.: 1-113 or 227-331. The CEG-encoding nucleic acid molecule is preferably inserted into an expression vector in operable linkage with suitable expression control sequences, to generate an expression vector including the CEG-encoding sequence. The expression vector is introduced into a suitable host, by standard transformation methods, and the resulting transformed host is cultured under conditions that allow the production of the CEG protein. For example, if expression of the CEG gene is under the control of an inducible promoter, then suitable growth conditions would include the appropriate inducer. The CEG protein (e.g., designated a CFE polypeptide or protein), so produced, is isolated from the growth medium or directly from the cells; recovery and purification of the protein may not be necessary in some instances where some impurities may be tolerated. A skilled artisan can readily adapt an appropriate host/expression system known in the art for use with CEG-encoding sequences to produce a CEG protein (Cohen, et al., *supra*; Maniatis et al., *supra*).

Host cells harboring the nucleic acids disclosed herein are also provided by the present invention. A preferred host is *E. coli* strain BL21(λDE3) transfected or transformed with a vector comprising a nucleic acid of the present invention. The invention also provides a host cell capable of expressing the *ceg* sequences described herein. The preferred host cell is any strain of *E. coli* that can accommodate high level expression of an exogenously introduced gene.

The proteins of the present invention can also be made by chemical synthesis. The principles of solid phase chemical synthesis of polypeptides are well known in the art and may be found in general texts relating to this area (Dugas, H. and Penney, C. 1981 *Bioorganic Chemistry*, pp 54-92, Springer-Verlag, New York). CEG polypeptides may be synthesized by solid-phase methodology utilizing an Applied Biosystems 430A peptide synthesizer (Applied Biosystems, Foster City, Calif.) and synthesis cycles supplied by Applied Biosystems. Protected amino acids, such as t-butoxycarbonyl-protected amino acids, and other reagents are commercially available from many chemical supply houses.

The polypeptides of the invention exhibit properties of a CEG protein, such as, for example, the ability to elicit the generation of antibodies that specifically bind an epitope associated with CEG polypeptides. Accordingly, the CEG polypeptide, or any oligopeptide thereof, is capable of inducing a specific immune response in appropriate animals or cells and binding with specific antibodies.

C) ANTIBODIES THAT RECOGNIZE AND BIND THE PROTEINS AND POLYPEPTIDES OF THE INVENTION

The invention further provides antibodies (e.g., polyclonal, monoclonal, chimeric, humanized, and human antibodies) that bind a CEG polypeptide. The most preferred antibodies will selectively bind a CEG polypeptide and will not bind (or will bind weakly) a non-CEG polypeptide. Antibodies that are particularly contemplated include monoclonal and polyclonal antibodies, as well as fragments thereof (e.g., recombinant proteins) which include the antigen binding domain and/or one or more complement determining regions of these antibodies. These antibodies can be from any source, for example, rabbit, sheep, rat, dog, cat, pig, horse, mouse, and human.

The invention encompasses antibody fragments that specifically recognize a CEG polypeptide. As used herein, an antibody fragment is defined as at least a portion of the variable region of the immunoglobulin molecule that binds to its target, i.e., the antigen binding region. Some of the constant region of the immunoglobulin may be included.

As will be understood by those skilled in the art, the regions or epitopes of a CEG polypeptide to which an antibody is directed may vary with the intended application. For example, antibodies intended for use in an immunoassay for the detection of membrane-bound CEG proteins on viable bacterial cells should be directed to an accessible epitope on membrane-bound CEG proteins. Antibodies that recognize other epitopes may be useful for the identification of CEG protein within damaged or dying cells, for the detection of secreted CEG protein or fragments thereof.

Various methods for the preparation of antibodies are well known in the art. For example, antibodies may be prepared by immunizing a suitable mammalian host using a CEG protein, peptide, or fragment, in isolated or immunoconjugated form (Harlow, 1989 *Antibodies*, Cold Spring Harbor Press, NY). In addition, fusion proteins comprising CEG polypeptides may also be used, such as a CEG protein/GST-fusion protein. Cells expressing or overexpressing a CEG polypeptide may also be used for immunizations. Similarly, any cell engineered to express CEG protein may be used. This strategy may result in the production of monoclonal antibodies with enhanced capacities for recognizing endogenous CEG protein.

The present invention contemplates chimeric antibodies that comprise a human and non-human immunoglobin portion. The antigen combining region (variable region) of a chimeric antibody can be derived from a prokaryotic source (e.g., bacteria) and the constant region of the chimeric antibody which confers biological effector function to the immunoglobulin can be derived from a eukaryotic source (e.g., human). The chimeric antibody should have the antigen binding specificity of the prokaryotic antibody molecule and the effector function conferred by the eukaryotic antibody molecule.

In one example, the procedure used to produce chimeric antibodies can involve the following steps:

a) Identifying and cloning the correct immunoglobin gene segment encoding the antigen binding portion of the antibody molecule. This gene segment is known as the VDJ, variable, diversity and joining regions for heavy chains or VJ, variable,

joining regions for light chains or simply as the V or variable region. This gene regions may be in either the cDNA or genomic form;

b) Cloning the gene segments encoding the constant region or desired part thereof;

c) Ligating the variable region with the constant region so that the complete chimeric antibody is encoded in a form that can be transcribed and translated;

d) Ligating this construct into a vector containing a selectable marker and gene control regions such as promoters, enhancers and poly(A) addition signals;

e) Amplifying this construct in bacteria;

f) Introducing this DNA into eukaryotic cells (transfection) most often mammalian lymphocytes;

g) Selecting for cells expressing the selectable marker;

h) Screening for cells expressing the desired chimeric antibody; and

k) Testing the antibody for appropriate binding specificity and effector functions.

Chimeric antibodies of several distinct antigen binding specificities have been produced by protocols well known in the art, including anti-TNP antibodies (Boulianne et al., 1984 *Nature* 312:643); and anti-tumor antigen antibodies (Sahagan et al., 1986 *J. Immunol.* 137:1066). Likewise, several different effector functions have been achieved by linking new sequences to those encoding the antigen binding region. Examples of these include enzymes (Neuberger et al., 1984 *Nature* 312:604); immunoglobulin constant regions from another species and constant regions of another immunoglobulin chain (Sharon et al., 1984 *Nature* 309:364; Tan et al., 1985 *J. Immunol.* 135:3565-3567). Additionally, procedures for modifying antibody molecules and for producing chimeric antibody molecules using homologous recombination to target gene modification have been described (Fell et al., 1989 *Proc. Natl. Acad. Sci. USA* 86:8507-8511).

The predicted amino acid sequence of a CEG protein may be used to select specific regions of the CEG protein for generating antibodies. For example, hydrophobicity and hydrophilicity analyses of a CEG polypeptide may be used to identify hydrophobic and hydrophilic regions in the CEG protein. Regions of the CEG protein that show immunogenic structure, as well as other regions and domains, can readily be identified using

various other methods known in the art, such as Chou-Fasman, Garnier-Robson , Kyte-Doolittle, Eisenberg, Karplus-Schult or Jameson-Wolf analysis. Fragments that include the immunogenic regions are particularly suited for generating specific classes of antibodies.

Methods for preparing a protein for use as an immunogen and for preparing immunogenic conjugates of a protein with a carrier such as BSA, KLH, or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents may be used; in other instances linking reagents such as those supplied by Pierce Chemical Co., Rockford, IL, may be effective. Administration of a CEG immunogen is conducted generally by injection over a suitable time period and with use of a suitable adjuvant, as is generally understood in the art. During the immunization schedule, titers of antibodies can be taken to determine adequacy of polyclonal antibody formation.

While the polyclonal antisera produced in this way may be satisfactory for some applications, for pharmaceutical compositions, monoclonal antibody preparations are preferred. Immortalized cell lines which secrete a desired monoclonal antibody may be prepared using the standard method of Kohler and Milstein (*Nature* 256: 495-497) or other techniques as described in *Monoclonal Antibodies; A Manual of Techniques,* CRC press, Inc., Boca Raton, Fla. (1987) ed. Zola. The immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the CEG polypeptide having binding activity, or a fragment thereof. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production in ascites fluid.

The desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal antibodies of the invention or the polyclonal antisera (e.g., Fab, $F(ab')_2$, Fv fragments, fusion proteins) which contain the immunologically significant portion (i.e., a portion that recognizes and binds a CEG protein) can be used as antagonists, as well as the intact antibodies. Humanized antibodies directed against a CEG polypeptide are also useful. The advantage of using humanized antibodies is that they are less immunogenic in humans. As used herein, a humanized antibody is an

immunoglobulin molecule which is capable of binding to a CEG polypeptide and which comprises a FR region having substantially the amino acid sequence of a human immunoglobulin and a CDR having substantially the amino acid sequence of non-human immunoglobulin or a sequence engineered to bind a CEG protein. Methods for humanizing murine and other non-human antibodies by substituting one or more of the non-human antibody CDRs for corresponding human antibody sequences are well known (Jones et al., 1986 *Nature* 321: 522-525; Riechmnan et al., 1988 *Nature* 332: 323-327; Verhoeyen et al., 1988 *Science* 239: 1534-1536; Carter et al., 1993 *Proc. Natl. Acad. Sci. USA* 89: 4285; and Sims et al., 1993 *J. Immunol.* 151: 2296).

Use of immunologically reactive fragments, such as the Fab, Fab', or $F(ab')_2$ fragments is often preferable, especially in a therapeutic context, as these fragments are generally less immunogenic than the whole immunoglobulin. Further, bi-specific antibodies specific for two or more epitopes may be generated using methods generally known in the art. Further, antibody effector functions may be modified so as to enhance the therapeutic effect of the antibodies of the invention. For example, cysteine residues may be engineered into the Fc region, permitting the formation of interchain disulfide bonds and the generation of homodimers which may have enhanced capacities for internalization, ADCC and/or complement-mediated cell killing (Caron et al., 1992 *J. Exp. Med.* 176: 1191-1195; Shopes, 1992 *J. Immunol.* 148: 2918-2922). Homodimeric antibodies may also be generated by cross-linking techniques known in the art (Wolff et al., *Cancer Res.* 53: 2560-2565). The invention also provides pharmaceutical compositions having the monoclonal antibodies or anti-idiotypic monoclonal antibodies of the invention.

The antibodies or fragments may also be produced, using current technology, by recombinant means. Regions that bind specifically to the desired regions of the CEG protein can also be produced in the context of chimeric or CDR grafted antibodies of multiple species origin. The invention includes an antibody, e.g., a monoclonal antibody which competitively inhibits the immunospecific binding of any of the monoclonal antibodies of the invention to a CEG protein.

Alternatively, methods for producing fully human monoclonal antibodies, include phage display and transgenic methods, are known and may be used for the generation of human monoclonal antibodies (reviewed in: Vaughan et al., 1998 *Nature Biotechnology* 16: 535-539). For example, fully human monoclonal antibodies may be generated using cloning technologies employing large human Ig gene combinatorial libraries (i.e., phage display) (Griffiths and Hoogenboom, "Building an *in vitro* immune system: human antibodies from phage display libraries", in: *Protein Engineering of Antibody Molecules for Prophylactic and Therapeutic Applications in Man,* Clark, M. (Ed.), Nottingham Academic, pp 45-64 (1993); Burton and Barbas, "Human Antibodies from combinatorial libraries" *Id.*, pp 65-82). Fully human monoclonal antibodies may also be produced using transgenic mice engineered to contain human immunoglobulin gene loci as described in PCT Patent Application WO98/24893, Jakobovits et al., published December 3, 1997 (see also, Jakobovits, 1998 *Exp. Opin. Invest. Drugs* 7: 607-614). This method avoids the *in vitro* manipulation required with phage display technology and efficiently produces high affinity, authentic human antibodies.

The antibody or fragment thereof of the invention may be labeled with a detectable marker or conjugated to a second molecule, such as a therapeutic agent (e.g., a cytotoxic agent) thereby resulting in an immunoconjugate. For example, the therapeutic agent includes, but is not limited to, an anti-tumor drug, a toxin, a radioactive agent, a cytokine, a second antibody or an enzyme. Further, the invention provides an embodiment wherein the antibody of the invention is linked to an enzyme that converts a prodrug into a cytotoxic drug.

Examples of cytotoxic agents include, but are not limited to ricin, ricin A-chain, doxorubicin, daunorubicin, taxol, ethiduim bromide, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicine, dihydroxy anthracin dione, actinomycin D, diphteria toxin, *Pseudomonas* exotoxin (PE) A, PE40, abrin, arbrin A chain, modeccin A chain, alpha-sarcin, gelonin, mitogellin, retstrictocin, phenomycin, enomycin, curicin, crotin, calicheamicin, sapaonaria officinalis inhibitor, and glucocorticoid and other chemotherapeutic agents, as well as radioisotopes such as ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Suitable detectable markers for diagnostic used include, but are not limited to, a radioisotope, a fluorescent compound, a bioluminescent compound, chemiluminescent compound, a metal chelator or an enzyme. Antibodies may also be conjugated to an anti-cancer pro-drug activating enzyme capable of converting the pro-drug to its active form. See, for example, U.S. Patent Nos. 4,952,394 and 5,716,990.

Additionally, a recombinant protein of the invention comprising the antigen-binding region of any of the monoclonal antibodies of the invention can be made. In such a situation, the antigen-binding region of the recombinant protein is joined to at least a functionally active portion of a second protein having therapeutic activity. The second protein can include, but is not limited to, an enzyme, lymphokine, oncostatin or toxin. Suitable toxins include those described above.

Techniques for conjugating or joining therapeutic agents to antibodies are well known (Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in: *Monoclonal Antibodies And Cancer Therapy,* Reisfeld et al. (eds.), pp. 243-56, Alan R. Liss, Inc. 1985; Hellstrom et al., "Antibodies For Drug Delivery", in: *Controlled Drug Delivery* (2nd Ed.), Robinson et al. (eds.), pp. 623-53, Marcel Dekker, Inc. 1987; Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in: *Monoclonal Antibodies '84: Biological And Clinical Applications,* Pinchera et al. (eds.), pp. 475-506 (1985); and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", in: *Immunol. Rev.,* 62:119-58 (1982)). Techniques for joining detectable markers to antibodies are also known.

D) PHARMACEUTICAL COMPOSITIONS OF THE INVENTION

The invention includes pharmaceutical compositions for use in the treatment of microbial infections comprising a pharmaceutically effective amount of an anti-CEG antibody or a CEG polypeptide.

In one embodiment, the pharmaceutical compositions may comprise a CEG antibody, either unmodified, conjugated to a therapeutic agent (e.g., drug, toxin, enzyme or second antibody) or in a recombinant form (e.g., chimeric or bispecific). The compositions may additionally include other antibodies or conjugates (e.g., an antibody cocktail).

The pharmaceutical compositions also preferably include suitable carriers and adjuvants which include any material which when combined with the molecule of the invention (e.g., an anti-CEG antibody or a CEG protein) retains the molecule's activity and is non-reactive with the subject's immune systems. Examples of suitable carriers and adjuvants include, but are not limited to, human serum albumin, ion exchangers, alumina, lecithin, buffer substances such as phosphates, glycine, sorbic acid, potassium sorbate, and salts or electrolytes such as protamine sulfate. Other examples include any of the standard pharmaceutical carriers such as a phosphate buffered saline solution, water, emulsions such as oil/water emulsion, and various types of wetting agents. Other carriers may also include sterile solutions, tablets including coated tablets and capsules. Typically such carriers contain excipients such as starch, milk, sugar, certain types of clay, gelatin, stearic acid or salts thereof, magnesium or calcium stearate, talc, vegetable fats or oils, gums, glycols, or other known excipients. Such carriers may also include flavor and color additives or other ingredients. Compositions comprising such carriers are formulated by well known conventional methods. Such compositions may also be formulated within various lipid compositions, such as, for example, liposomes as well as in various polymeric compositions, such as polymer microspheres.

The pharmaceutical compositions of the invention can be administered using conventional modes of administration including, but not limited to, intravenous, intraperitoneal, oral, intralymphatic or administration directly into the tumor. Intravenous administration is preferred.

The pharmaceutical compositions of the invention may be in a variety of dosage forms which include, but are not limited to, liquid solutions or suspensions, tablets, pills, powders, suppositories, polymeric microcapsules or microvesicles, liposomes, and

injectable or infusible solutions. The preferred form depends upon the mode of administration and the therapeutic application.

The CEG polypeptides and proteins of this invention are found in common pathogenic bacterial species such as *Streptococcus pneumoniae*. This organism causes upper respiratory tract infections. Thus, the peptides and proteins of this invention can be used as immunogens in subunit vaccines for vaccination against a pathogenic bacteria such as *Streptococcus pneumoniae*. Additionally, the *ceg* sequences of the invention can be used as DNA vaccines (U.S. Patent No. 5,736,524 and U.S. Patent No. 5,989,553).

The polypeptides and proteins of this invention can be formulated as univalent and multivalent vaccines. The protein can be mixed, conjugated or fused with other antigens, including B or T cell epitopes of other antigens.

Further, when a haptenic peptide of the proteins of the invention is used, (i.e., a peptide which reacts with cognate antibodies, but cannot itself elicit an immune response), it can be conjugated to an immunogenic carrier molecule. Conjugation to an immunogenic carrier can render the oligopeptide immunogenic. Examples of carrier molecules are tetanus toxin or toxoid, diphtheria toxin or toxoid and any mutant forms of these proteins such as CRM.sub.197. Others include exotoxin A of *Pseudomonas*, the heat labile toxin of *E. coli* and rotaviral particles (including rotavirus and VP6 particles). Alternatively, a fragment or epitope of the carrier protein or other immunogenic protein can be used. For example, the happen can be coupled to a T cell epitope of a bacterial toxin.

In formulating the vaccine compositions with the CEG polypeptides or proteins of the invention, alone or in the various combinations described, the immunogen is adjusted to an appropriate concentration and formulated with any suitable vaccine adjuvant. Suitable adjuvants include, but are not limited to: surface active substances, e.g., hexadecylamine, octadecylamine, octadecyl amino acid esters, lysolecithin, dimethyl-dioctadecylammonium bromide), methoxyhexadecylgylcerol, and pluronic polyols; polyamines, e.g., pyran, dextransulfate, poly IC, carbopol; peptides, e.g., muramyl

dipeptide, dimethylglycine, tuftsin; oil emulsions; and mineral gels, e.g., aluminum hydroxide, aluminum phosphate, etc. and immune stimulating complexes. The immunogen may also be incorporated into liposomes, or conjugated to polysaccharides and/or other polymers.

The vaccines can be administered to a human or animal in a variety of ways. These include intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, oral and intranasal routes of administration. Further, the vaccines can be live or inactivated vaccines.

The most effective mode of administration and dosage regimen for the compositions of this invention depends upon the severity and course of the disease, the patient's health and response to treatment and the judgment of the treating physician. Accordingly, the dosages of the compositions should be titrated to the individual patient.

E) USES OF THE MOLECULES OF THE INVENTION

1) MOLECULAR WEIGHT MARKERS

The nucleic acid molecules of the invention and their encoded proteins may be employed as molecular weight markers. For example, the molecular weight of each of the nucleic acid molecules having *ceg* sequences and their predicted polypeptides can be determined and can be used to compare against other gene sequences and proteins whose molecular weights are unknown.

2) DIAGNOSTICS

The nucleic acid molecules of the invention may be employed in diagnostic embodiments. For example, the presence of nucleotide sequences which are identical or similar to the *ceg* sequences of the invention may be detected within a biological sample.

The biological sample may include blood, serum or a swab from nose, ear or throat, may be determined by means of a nucleic acid detection assay.

Nucleic acid probes or primers having sequences complementary to *ceg* sequences may be used in a hybridization assay to detect the presence of the sequences which are identical or similar to the *ceg* sequences of the invention in the biological samples. Typically, nucleic acids molecules obtained from a suitable biological sample are hybridized with labeled probes or primers. The resulting hybridized molecules are detected and resolved by methods well known in the art , such as Northern or Southern blotting, micro-array technology, or amplifying with PCR technology. Other hybridization techniques and systems are known that can be used in connection with the detection aspects of the invention, including diagnostic assays such as those described in Falkow et al., U.S. Pat. No. 4,358,535.

Examples of the PCR technology are disclosed in U.S. Patent Nos. 4,683,202 and 4,965,188 (incorporated herein by reference). Generally, nucleic acid molecules are obtained from a suitable biological source and contacted with two primers corresponding to the *ceg* sequences disclosed herein, under conditions which allow for hybridization and polymerization to occur. A pair of probes, one corresponding to the 5' flanking region and the other corresponding to the 3' flanking region, would be sufficient to detect the nucleic acid molecules of the invention in a biological sample and may be used to indicate the amount of bacteria present.

Alternative methods of detecting nucleic acid molecules include, for example, in situ hybridization techniques, where a *ceg* probe is used to detect homologous sequences within one or more cells, such as cells within a clinical sample or even cells grown in tissue culture. As is well known in the art, the cells are prepared for hybridization by fixation, e.g. chemical fixation, and placed in conditions that allow for the hybridization of a detectable probe with nucleic acids located within the fixed cell.

The amount of *ceg* sequences present in a biological sample can be quantified and compared to the levels in a normal or "healthy" sample. For example, *ceg* sequences present in either increased or decreased levels, compared to the levels found in the control sample may indicate the presence of bacteria. This information is useful for diagnosis of a bacterial infection that requires treatment with an antibacterial agent.

Alternatively, the amount of CEG polypeptides present in a biological sample may be determined by means of an immunoassay. For example, labeled antibodies reactive against CEG polypeptides may be used in an immuno-reactive assay to detect the presence of CEG polypeptides in the biological samples.

3) SCREENING CANDIDATE CEG SEQUENCES

a) Gene Disruption Assay

The *ceg* nucleotide sequences of the invention can be used to identify nucleotide sequences which are identical or similar to the *ceg* sequences that are required for bacterial cell viability. For example, the *ceg* sequences can be used in a bacterial gene disruption assay to screen candidate nucleotide sequences to identify sequences required for bacterial cell viability.

The disruption assay can involve: introducing into a host cell a recombinant vector that is capable of integration into the host genome, where the recombinant vector includes a candidate sequence that putatively encodes a cell-viability gene product (e.g., the exogenous *ceg* sequence); the vector integrates the candidate sequence into a target sequence within the host's genome (e.g., the endogenous *ceg* sequence); and the host cell, so introduced, is screened for viability. The recombinant vector preferably includes a selectable marker so that the introduced host cell can be screened for viability in the presence of a selectable agent.

For example, Figure 1 shows a schematic representation of a gene disruption assay, within a bacterial host cell. In Figure 1A, the recombinant vector, pEVP3, includes the CAT gene (e.g., the selectable marker chloramphenicol acetyl transferase) and an internal region of the *ceg* disrupting sequence; the internal region excludes the 5' and 3' ends of the *ceg* sequence. The "X" in Figure 1 indicates the recombinant pEVP3 vector undergoing homologous recombination with the target sequence (e.g., within the host genome). In Figure 1B, the resolved pEVP3 vector that is integrated into the host genome, is shown. Left to right are the following elements: the native promoter of the target gene; a 5' partial copy of the target gene; the body of the integrated pEVP3 vector including the disrupting gene and CAT; and, a 3' partial copy of the target gene. Thus, integration of the pEVP3 vector via homologous recombination results in two partial gene duplications flanking the integrated vector. If the target gene is not essential for survival, it is possible to recover chloramphenicol-resistant colonies of *S. pneumoniae*. Failure to recover chloramphenicol resistant colonies, in the presence of the proper controls as described below, indicates that the target gene may be essential for cell viability.

More particularly, the gene disruption assay for screening candidate *ceg* sequences can involve the following steps. The recombinant pEVP-3 vector encoding CAT resistance and having a fragment of a candidate *ceg* sequence, can be introduced into transformation-competent *S. pneumoniae* cells by methods that are well-known in the art (Lee, M.S., et al., 1998 *Appl. Environ. Microbiol.* 64:4796-4802). The preferred size of the *ceg* fragment can be between about 200 to about 500 bp in length. It is advantageous that the candidate *ceg* sequence does not include the 5' and 3' ends that encode the N- and C-terminal ends of the CEG polypeptide. This insures that the inserted *ceg* fragment and the disrupted endogenous *ceg* gene sequence are not capable of expression of a full-length, functional *ceg* gene product. The transformation-competent cells can be obtained by performing the transformation step in the presence of a heptadecapeptide that induces competence for transformation of *S. pneumoniae* (Havarstein, L. S., et al., 1995 *Proc. Nat'l. Acad. Sci.* 92:11140-11144), such as the CSP-1 peptide. The CSP-1 can be naturally-derived or synthetic. Additionally, the transformation step can be optimized by performing the transformation when the cells have reached a density which is optimal for

transformation (e.g., 3 X 10^7 cells per ml.) (Havarstein, L. S. et al. *supra*). The recombinant vector can be introduced into the competent pneumococci and may undergo homologous recombination, whereby the candidate *ceg* fragment recombines with the corresponding endogenous *ceg* sequence, resulting in targeted integration of the vector into the pneumococcal genome and disruption of the endogenous *ceg*.

The transformed cells can be plated on or cultured in chloramphenicol-containing growth medium. The cells can be cultured under standard conditions, such as 37° C in 5% CO_2 for approximately 40 to 48 hours, for the purpose of selecting cells that carry the integrated vector.

Additionally, control samples can be run in parallel with the gene disruption assay, in order to determine whether the gene disruption procedure is working properly. For example, the control samples can be used to calibrate the gene disruption experiment so that disruption of a known non-essential bacterial gene results in an approximate number of colonies per plate. Similarly, the disruption of a known essential gene can be calibrated to yield only zero or one colony per plate. The appearance of one colony is due to the rare illegitimate recombination into a non-homologous sequence. In particular, a known non-essential gene such as the *lytA* gene (Tomasz, A., et al., 1988 *J. Bacteriol.* 170:5931-5934) can be used so that between about 70 to 100 chloramphenicol-resistant colonies will grow per plate. Similarly, the *ftsZ* gene (Lutkenhaus, J. F., et al., 1980 *J. Bacteriol.* 143:1281-1288), a known essential gene, can be used to yield zero or, rarely, one colony per plate. As is well known in the art, specific parameters that are involved in any given gene disruption assay can be adjusted to calibrate the desired number of plated cells in the control samples. Experimental parameters that can be adjusted include, but are not limited to, the *E. coli* strain used to propagate the vector/insert, the fragment length of the sequence to be integrated, the amount of recombinant integration vector used to transform the cells, use of transformation-competent cells, and plating density of the transformed cells.

The transformed cells carrying the recombinant integration vector that disrupts expression of an endogenous essential gene (e.g., the target *ceg* gene) can be identified, based on a selectable phenotype such as non-viability. For example, the cells that carry a disrupted non-essential gene will be viable and, due to the integration of pEVP3, will grow on chloramphenicol-containing medium. In contrast, cells that carry a disrupted essential gene will not grow (e.g., non-viable) on the chloramphenicol-containing medium. Thus, the transformed cells that do not grow under these selective conditions carry an endogenous gene sequence that is essential for cell viability which has been disrupted by an exogenous candidate fragment, thereby identifying a *ceg* sequence. Steps one through three may be repeated in order to confirm that the *ceg* sequences, so identified, are essential for cell viability.

b) Autolysin Assay

It is advantageous to perform additional steps to determine whether the homologous recombination events result in disruption of the intended target gene sequence. The *lytA* transformation control can be used to confirm that the transformation system is functioning properly. For example, a phenotypic test for autolysin activity (*lytA* gene product) can be performed to determine that the exogenous *lytA* fragment is correctly integrated into the *lytA* site within the host genome. This typically involves flooding the culture plates containing transformants carrying the integrated *lytA* control vector with a solution of detergent, such as 0.1% deoxycholate, which triggers cell lysis in *lytA*-intact cells (e.g., the cells that have not undergone homologous recombination). After about 5-10 minutes the colonies with intact *lytA* will appear ghost-like due to cell lysis, and the colonies with a disrupted *lytA* gene will appear intact.

c) Polarity Analysis

The *ceg* sequences that are confirmed to be essential for cell viability can be examined further by performing a polarity analysis to determine if the corresponding endogenous *ceg* sequence is organized in an operon. Polarity is an effect unique to prokaryotes and is

the result of the operon organization of bacterial genomes. Many bacterial genes are arranged in operons in which multiple genes are under the control of a single regulatory sequence (e.g., a promoter) and are transcribed into a single mRNA transcript. With respect to the orientation of multiple genes within an operon, the genes that are proximal to the regulatory sequence are said to be "upstream" genes and the genes that are distal are said to be "downstream" genes. For example, many operons contain genes encoding different proteins that catalyze discrete steps of a common biochemical pathway. Thus, any of the proteins that catalyze the steps of the pathway may be essential for cell viability.

The presence of operons in a bacterial host genome may influence the interpretations of the gene disruption results. For example, disruption of an upstream gene may be erroneously interpreted as affecting the expression of the disrupted gene but may, in fact, have expression affects on the intact downstream genes. Therefore, it is advantageous to perform a polarity analysis to determine if a *ceg* sequence is part of an operon.

A polarity analysis can involve performing an *in vivo* gene disruption procedure using, as the disrupting sequence, a *ceg* sequence that includes the entire *ceg* coding sequence region but lacking expression regulatory sequences. This differs from the gene disruption assay, which involves the central region of the *ceg* sequence. The polarity analysis involves gene duplication via homologous recombination. For example, the pEVP-3 vector having the entire coding region of a *ceg* sequence can be used for the polarity analysis (Figure 2 A). The polarity analysis will yield different results depending on the organization of the endogenous target sequence within the host genome.

For example, Figure 2 shows a schematic representation of the polarity test for operons, within a bacterial host cell. In Figure 2A, the recombinant vector, pEVP3, includes the CAT gene and the entire coding region of the *ceg* disrupting sequence. The "X" in Figure 2 indicates the recombinant pEVP3 vector undergoing homologous recombination with the target sequence. Two of the possible results of homologous recombination are shown in Figures 2 B and C.

In Figure 2 B, case 1, if the endogenous target sequence is not organized in an operon, the integration event may yield: a functional target sequence (e.g., it is capable of expression); a duplicate non-functional target sequence that lacks a promoter; and a functional downstream gene (e.g., Gene B) that is controlled by its own promoter. The cells carrying this type of integrated target sequence can be recovered as viable cells that grow in the presence of chloramphenicol; this condition is termed "polarity negative".

In Figure 2 C, case 2, if the target sequence is organized in an operon, then the integration event may yield an integration site that is similar to that described for case 1, including: a functional target sequence; and a duplicate non-functional target sequence which is not functional. However, this integration event may also yield a non-functional downstream gene (e.g., Gene B) because expression of this downstream gene is controlled by a promoter located upstream of the insertion site. The cells that carry this type of integrated target sequence will be non-viable; this condition is termed "polarity positive". Thus, the polarity analysis provides a method to determine whether integration of a recombinant vector into a target *ceg* sequence effects expression of downstream genes.

The *ceg* sequences disclosed herein (SEQ ID NOs.: 1-113, 227-331) encode gene products that are essential for viability in *S. pneumoniae*. Furthermore, many of these *ceg* sequences have been analyzed for the polarity effect and the results are presented in Table I. One subset of *ceg* sequences is classified as polarity negative (-), since the homologous recombination event did not effect the expression of downstream genes. Another subset of *ceg* sequences is classified as polarity positive (+), since the homologous recombination event did affect the expression of downstream genes. The *ceg* sequences that have not yet been classified as polarity positive or negative are indicated in Table I as a blank. For the *ceg* sequences that are classified as polarity positive, the genes downstream of the disrupted endogenous *ceg* sequences may or may not also be essential.

4) ASSAYS FOR IDENTIFYING CEG LIGANDS AND OTHER BINDING AGENTS

The present invention provides screening methods for identifying agents that interact and/or bind to the CEG proteins of the invention, such as a ligand. An agent can be, for example, a natural product, a derived or synthetic chemical molecule, a polypeptide, a nucleic acid molecule, or a metal. The agents that interact with CEG proteins may cause bacterial cell death by disrupting the functions of CEG proteins, including, but not limited to, nucleotide biosynthesis, DNA replication, RNA transcription, protein translation, and/or cell wall biosynthesis. Accordingly, the present invention provides screening methods for identifying agents having antibacterial activity, such as agents that cause bacterial cell death by interacting with the CEG proteins. These antibacterial agents are useful for treating diseases and afflictions associated with bacterial infections.

Various methods can be used to discover agents having antibacterial activity, as determined by the ability of the binding agent to bind to a CEG protein and disrupt the function of the CEG protein. These screening methods include whole cell *in vivo* assays as well as *in vitro* assays with cellular components.

An *in vivo* screening method for identifying ligands that bind CEG polypeptides can be performed in a whole cell assay. A typical method may be the use of whole bacterial cells to assess the antibacterial properties based on cell growth or viability. These methods can include methods for measuring cell growth and/or viability, for example, by optical density or zones of growth (Koch, A. L. et al., 1970 *Anal. Biochem.* 38:252-259; Biemer, J. J. et al., 1973 *Ann. Clin. Lab. Sci.* 2:135-140; *Manual of Clinical Microbiology*, 7th edition, Murray, P. R. (ed), ASM Press), by growth inhibition in an agar assay (Murray, P. R., *supra*), or other means of detecting cell metabolism (Mychajlonka, M. et al., 1980 *Antimicrob. Agents Chemother.* 17:572-582), and are well known to those skilled in the art. In addition, there are molecular biology-based detection methods for use with whole bacterial cells, such as gene reporter assays, to monitor the effect of the ligand on specific targets (Slauch, J. M., et al., 1991 *Methods Enzymol.* 204:213-248). Examples of the reporter genes include, but are not limited to, beta-

galactosidase, alkaline phosphatase, luciferase, and green fluorescent protein. For example, one embodiment provides a reporter system that monitors inhibition of DNA synthesis by fusing a reporter such as beta-galactosidase (*lacZ*) to genes known to be upregulated by the cessation of DNA synthesis as a result of the binding of ligands to the DNA synthetic apparatus. (Shurvinton, C. E., et al., 1982 *Mol. Gen. Genetics* 185:352-355; Rosato, A., et al., 1998 *Antimicrob. Agents Chemother.* 42:1392-1396).

Alternatively, the yeast two-hybrid system (Fields, S. and Song, O. 1989, *Nature* 340:245-246) may be adapted to screen for ligands that bind CEG polypeptides. Generally, the yeast two-hybrid system is performed in a yeast host cell carrying a reporter gene, and is based on the modular nature of the GAL transcription factor which has a DNA binding domain and a transcriptional activation domain. The yeast two-hybrid system relies on the physical interaction between a recombinant polypeptide that comprises the GAL DNA binding domain and another recombinant polypeptide that comprises the GAL transcriptional activation domain. The physical interaction between the two recombinant polypeptides reconstitutes the transcriptional activity of the transcription factor, thereby causing expression of the reporter gene. Either of the recombinant polypeptides used in the two-hybrid system can be generated to include a CEG polypeptide sequence to screen for binding partners of CEG.

Another method uses the bacterial CEG proteins as the basis for *in vitro* assay systems to detect binding agents. Typically, the *in vitro* screening method comprises: a) generating the CEG protein of the invention, or membranes enriched in the CEG protein; b) exposing the CEG protein or membranes to a candidate agent; and c) detecting the interaction of the CEG protein with the agent by any suitable means. Additionally, the screening methods may be adapted to automated high-throughput procedures, such as PANDEX.RTM Baxter-Dade Diagnostics, allowing for efficient high-volume screening of candidate agents.

An alternative method for screening potential ligands involves an *in vitro* binding procedure. Typically, the CEG proteins of the invention can be produced using

recombinant DNA technology and host-vector systems as described herein. A candidate agent is introduced into a reaction vessel containing the CEG protein, or fragment thereof; the candidate agents may be detectable by methods such as, but not limited to, radioisotope or chemical labeling. Binding of the CEG protein by a candidate agent can be determined by any suitable means, including, for example, quantifying bound label versus unbound label using any suitable method. Binding of a candidate agent may also be detected by methods similar to an alternative physical method disclosed in U.S. Patent No. 5,585,277. In this method, binding of a candidate agent to a protein is assessed by monitoring the ratio of folded protein to unfolded protein, for example by monitoring sensitivity of the protein to a protease, or amenability to binding of the protein by a specific antibody against the folded state of the protein, or binding to chaperone protein, or by binding to any suitable surface.

The invention provides methods of identifying compounds that modulate (e.g., activate or inhibit) the function of a CEG polypeptide. Essentially any compound can be used in the assays of the invention. The preferred compounds are those that are soluble in aqueous or organic solutions. It will be appreciated by those of skill in the art that there are many commercial suppliers of chemical compounds that can be used in the methods of the invention, including Sigma Chemical Co. (St. Louis, Mo.), Aldrich Chemical Co. (St. Louis, Mo.), Sigma-Aldrich (St. Louis, Mo.), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland), and the like.

The present invention provides methods for detecting compounds which are identified as modulators of CEG function. The methods of the invention can be performed using isolated CEG polypeptides, or use whole cells expressing the CEG polypeptide. The steps of the method using isolated CEG polypeptides include: contacting the isolated CEG polypeptide with a candidate compound; and determining whether the function of the CEG polypeptide is altered. The steps of the method using whole cells include: contacting the whole cells with a candidate compound; and determining whether the cell dies, indicating the compound inhibited the function of a CEG polypeptide.

The preferred methods of the invention provide high-throughput screening assays for identifying compounds which modulate the function of a CEG polypeptide. The high throughput methods permit screening of large libraries of compounds. For example the high throughput methods can use automated assay steps. The assays can be performed in parallel on a solid support, as microtiter formats on microtiter plates in robotic assays are well known. A preferred embodiment of the methods includes adapting the methods to use microtiter plates or pico- nano- or micro-liter arrays. In high throughput assays it is desirable to run positive controls to ensure that the components of the assays are working properly.

The high throughput screening methods of the invention include providing a combinatorial library containing a large number of compounds (candidate modulator compounds) (Borman, S, *C. & E. News*, 1999, 70(10), 33-48). Such combinatorial chemical libraries can be screened in one or more assays to identify library members (particular chemical species or subclasses) that exhibit the ability to modulate the function of the CEG polypeptide (Borman, S., *supra*; Dagani, R. *C. & E. News*, 1999, 70(10), 51-60). The compounds, so identified, can serve as lead-compounds or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by using either chemical synthesis or biological synthesis, to combine a number of chemical building blocks, such as reagents. For example, a linear combinatorial chemical library, such as a polypeptide library, is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka, *Int. J. Pept. Prot. Res.*, 1991, 37:487-493 and

Houghton, et al., *Nature*, 1991, 354, 84-88). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to, peptoids (PCT Publication No. WO 91/19735); encoded peptides (PCT Publication WO 93/20242); random bio-oligomers (PCT Publication No. WO 92/00091); benzodiazepines (U.S. Pat. No. 5,288,514); diversomers, such as hydantoins, benzodiazepines and dipeptides (Hobbs, et al., *Proc. Nat. Acad. Sci. USA*, 1993, 90, 6909-6913); vinylogous polypeptides (Hagihara, et al., *J. Amer. Chem. Soc.* 1992, 114, 6568); nonpeptidal peptidomimetics with *beta*-D-glucose scaffolding (Hirschmann, et al., *J. Amer. Chem. Soc.*, 1992, 114, 9217-9218); analogous organic syntheses of small compound libraries (Chen, et al., *J. Amer. Chem. Soc.*, 1994, 116, 2661; Armstrong, et al. *Acc. Chem. Res.*, 1996, 29, 123-131); or small organic molecule libraries (see, e.g., benzodiazepines, Baum *C&E News*, 1993, Jan. 18, page 33,); oligocarbamates (Cho, et al., *Science*, 1993, 261, 1303); and/or peptidyl phosphonates (Campbell, et al., *J. Org. Chem.* 1994, 59, 658); nucleic acid libraries (see, Seliger, H et al., *Nucleosides & Nucleotides*, 1997, 16, 703-710); peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083); antibody libraries (see, e.g., Vaughn, et al., *Nature Biotechnology*, 1996, 14(3), 309-314 and PCT/US96/10287); carbohydrate libraries (see, e.g., Liang, et al., *Science*, 1996, 274, 1520-1522 and U.S. Pat. No. 5,593,853, Nilsson, UJ, et al., *Combinatorial Chemistry & High Throughput Screening*, 1999 2, 335-352; Schweizer, F; Hindsgaul, O. *Current Opinion In Chemical Biology*, 1999 3, 291-298); isoprenoids (U.S. Pat. No. 5,569,588); thiazolidinones and metathiazanones (U.S. Pat. No. 5,549,974); pyrrolidines (U.S. Pat. Nos. 5,525,735 and 5,519,134); morpholino compounds (U.S. Pat. No. 5,506,337); benzodiazepines (U.S. Pat. No. 5,288,514); and other similar art.

Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem. Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, Mo., ChemStar, Ltd., Moscow, RU, 3D Pharmaceuticals, Exton, Pa., Martek Bio sciences, Columbia, Md., etc.).

In the high throughput methods of the invention, several thousand different candidate compounds can be screened in a relatively short period of time. For example, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay many different plates per day; assay screens for up to about 6,000-20,000, and even up to about 100,000-1,000,000 different candidate modulator compounds are possible using the methods of the invention.

The following examples are presented to illustrate the present invention and to assist one of ordinary skill in making and using the same. The examples are not intended in any way to otherwise limit the scope of the invention.

EXAMPLE 1

The following provides a general description of how a list of candidate *ceg* sequences was generated. The list was generated by selecting candidate *ceg* gene sequences from a Concordance web engine using the method described in: Bruccoleri, R.E., Dougherty, T.J., Davison, D.B. (1998) "Concordance analysis of microbial genomes" in: *Nucleic Acids Res* 26:4482-4486.

Microbial Genomics CEG Discovery Process Summary.

Microbial Concordance Analysis

The entire genomic sequence data of various bacteria was acquired from several public and proprietary sequence database sources, including GTC (Genome Therapeutics Corporation), and TIGR (The Institute for Genomic Research).

Predicted ORFs from the genomic data were identified, translated, and stored. The desirable ORFs were at least 90 amino acid residues in length. Concordance analysis was performed among bacteria and various parameters were used to filter out genes with high similarity to eukaryotes.

Concordance Analysis

The entire genomic sequence of various Eubacteria was acquired from several public and private sources. The proprietary PathoGenome System from Genome Therapeutics Corporation, Waltham, MA, USA contributed data. Public data was obtained from GenBank (http://ncbi.nlm.nih.gov), The Institute for Genomic Research (TIGR), the Yeast Proteome Database, from Proteome, Inc. of Beverly, MA, and the Sanger Center of the Medical Research Council of the United Kingdom (http://www.sanger.ac.uk). Additionally, the non-microbial sequence data used as a basis for comparison and data subtraction was obtained from a proprietary database, including the LifeSeq Database from Incyte Pharmaceuticals, Palo Alto, CA.

Where required, Incyte nucleotide sequences were translated into protein sequences in all six possible reading frames. GTC supplied predicted protein sequences with their data. In the case of other eubacterial nucleotide sequences, the program CRITICA (Badger, J. and Olsen, G., 1999 "CRITICA: coding region identification tool invoking comparative analysis" in: *Molecular Biology and Evolution* 16:512-524). The sequences were stored in flat files on a Unix computer system. Each predicted amino acid sequence had to be greater than 90 amino acids.

Each predicted protein sequence was compared to every other sequence (an "all-against-all" comparison). The program used was FASTA (Pearson, W.R., "Flexible sequence similarity searching with the FASTA3 program package." *Methods in Molecular Biology 2000* 132:185-219.) The parameters used were ktup=2, and all scores above the default cutoff were kept. The output was processed and stored in a PostGres 95 database (http://www.postgresql.org). Graphical user interfaces, using web browser technology, were constructed to query the database.

A Concordance Analysis was performed on the data. The question used to generate the dataset was show all *Streptococcus pneumoniae* open reading frames with a similarity greater than or equal to 30% overall protein sequence identity to both selected gram-positive and/or gram-negative bacteria in the database. The data was further required not to match yeast or human sequences at greater than 30% overall protein sequence similarity. The resulting dataset included a list of more than 400 conserved amino acid sequences having known or unknown function. The amino acid sequences having unknown functions formed the basis of a list designated Conserved Unknown Reading Frames, or CURFs which is a subset of the total list of CEGs (e.g., CURFs includes known and unknown).

The resulting list of conserved genes (e.g., more than 400 sequences) was used as a basis for selecting and screening bacterial gene sequences that are essential for cell viability. The Concordance system was designed to permit high-throughput identification of conserved gene sequences in the database. (Bruccoleri, R, Dougherty, T, and Davison, D. 1998 "Concordance analysis of microbial genomes" *Nucleic Acids Res.* 26:4482-4486.)

Data Curation And Analysis

Exact N-terminal and C-terminal translational start sites of genes were identified by pairwise similarity searches, multiple sequence alignments. Ribosome binding sites, terminators, nearby genes, operons were identified.

The resulting list of conserved genes was used as a basis for selecting and screening bacterial gene sequences that are essential for cell viability. This Concordance system was designed to permit high throughput use of the conserved gene sequences contained on the list. A set of Knockout PCR primers were generated, based on the list of conserved genes, for the purpose of use in the gene disruption procedure described below. The PCR primers were designed to amplify a central 300-500 bp region of the *ceg* (to prevent generation of a functional copy of the *ceg* gene following integration),

ordered electronically, the primers were placed in a 96-well format, and used in the gene disruption procedure as described below.

EXAMPLE 2

The following provides a description of the procedure to generate recombinant vectors of pEVP-3 having inserts of candidate *ceg* nucleotide sequences. The Knockout primers generated by the method described in Example 1 above were used to generate DNA fragments comprising candidate *ceg* sequences.

Genomic PCR Knockout Target Fragment Generation

96-well plate format were set up (36 µl H_2O , 5 µl 10× Vent™ buffer, 1 µl gene specific, knockout forward primer (0.5 µg/µl), 1 µl gene specific knockout reverse primer (0.5 µg/µl), 0.5 µl Vent™ DNA polymerase (2000 U/ml New England Biolabs, Beverly, MA), 1.5 µl each dNTPs (10mM; 6.0 µl total), 0.5 µl *S. pneumoniae* chromosomal DNA (0.5 µg/µl), 50 µl total volume/reaction).

The nucleotide sequences of the forward and reverse knockout primer pairs were generated from the nucleotide sequence information obtained from the Genomic Therapeutics Corporation database for *Streptococcus pneumoniae.* The primer pairs were each used in a PCR reaction to generate a unique internal (e.g., central region) fragment of the candidate gene targeted for knockout.

The PCR program was set in the PCR machine (Initial 95 °C - 5 minutes: 30 Cycles of: 95 °C - 1 minute, 58 °C - 1 minute, 72 °C - 30 seconds; Final, 72 °C - 10 minutes, 4 °C – hold indefinitely). 5 µl of each reaction was run on an 0.8% agarose gel after purifying fragment over PCR purification kit (Qiagen) to visualize the fragments then ligation reactions were performed.

Ligation Reactions proceeded (set up in 96-well plate format (10.0 µl genomic PCR fragment (generated from step 2 above), 1.0 µl pEPV-3 SmaI-cut vector (1: 10 dilution of vector DNA at 50-100 ng/µl), 1.5 µl 10× ligation buffer (New England Biolabs™), 1.0 µl T4 DNA Ligase (New England Biolabs™ 400,000 U/ml), 1.5 µl ddH_2O, 15.0 µl total reaction volume).

Reactions were allowed to incubate in 96-well plate at 14 °C overnight in the PCR machine. Transformations into E. coli for in vivo amplification were proceeded the following day.

The nucleotide sequences of the forward and reverse primer pairs used for the polarity test were generated in a similar manner, from the nucleotide sequence information obtained from the Genomic Therapeutics Corporation database for *Streptococcus pneumoniae*. The primer pairs were each used in a PCR reaction to generate a unique fragment of the candidate gene targeted for the polarity test. The fragment generated for the polarity test included the entire *ceg* coding sequence region but lacking the expression regulatory sequences.

Transformation into *E. coli* (strain LE392):

The next day, 3 µl of above ligation mix was used per transformation reaction plus 50 µl LE392 competent cells. Reactions were set up in 96-well plate format; incubated on ice for 30 minutes; heat-shocked at 42° C for 90 seconds; and incubated on ice 2 minutes; 100 µl SOC media (Gibco BRL) was added; then incubated at 37° C on platform shaker for 1 hour; plated on LB/chloramphenicol (13.0 µg/ml) agar plates for constructs over night at 37° C with plates inverted and proceeded with colony PCR to confirm constructs. The universal primers flanking the insert site in pEVP-3 were used for PCR amplification.

The colony PCR involved the following. 96-well plate format was set up (36.5 µl H_2O, 0.5 µl pEPV3 forward primer (0.25 µg/µl), 0.5 µl pEPV3 reverse primer (0.25 µg/µl), 1.5

µl each (6.0 µl total) dNTPs (10 mM), 0.5 µl Vent™ DNA polymerase, 5 µl 10× Vent™ buffer, 1 µl of a 1:50 cell dilution, 50 µl total volume).

pEPV3 forward primer: 5' CATCAAGCTTATCGATACCGTCG 3' (SEQ ID NO:437)
p EPV3 reverse primer: 5' CACAGTAGTTCACCACCTTTTCCC 3' (SEQ ID NO:438)

Colonies of *E. coli* LE392 were picked onto a master plate of LB + 13 µg/ml chloramphenicol (incubate throughout the day at 37° C) and then into 50 µl H_20 which has been placed into a 96-well plate. 1 µl of this dilution was used in above PCR reaction (if the 96-well dilution plate is kept you will not need to prepare a master plate). Cultures for minipreps of plasmid candidates may be prepared directly from the cell dilutions.

The PCR program was run (95 °C - 5 minutes, 30 Cycles of: 95 °C - 1 minute, 58 °C - 1 minute, 72 °C - 30 seconds, 72 °C - 10 minutes, 4 °C - hold).

A 10 µl/ reaction was run on a 1.0 % TBE gel. A gel designed for 96 well plates and a multichannel pipettor were used to ease loading of the sample rows. The gel was run and stained with ethidium bromide. The positive clones were identified with appropriate molecular size insert(s), amplified by the flanking pEVP-3 primers.

Minipreps Of Plasmids To Identify Cells Carrying The Pevp-3 Vector With An Insert

The constructs that carried an insert were identified. The constructs having an insert were inoculated into a 5 ml LB/Cm culture, and incubated over night at 37 °C with aeration. Miniprep plasmid DNA was prepared by a standard procedure. The miniprep DNA was digested with appropriate restriction enzymes to confirm the presence of the insert (enzymes flank SmaI site in pEVP-3) (10 µl miniprep DNA, 2 µl 10 × buffer, 1 µl XbaI, 1 µl XhoI, 6 µl ddH20, 20 µl total volume for digest).

To confirm the presence of an insert, the digest reactions were electrophoresed on an agarose gel and the gel was stained with ethidium bromide. The positive clones were used for the *S. pneumoniae* KNOCKOUTs procedure.

The confirmatory PCR reactions, using knock out-specific primers (quality control step) involved 35.5 µl H_2O, 5 µl 10 × Vent™ buffer, 1 µl knockout forward primer (0.5 µg/µl), 1 µl knockout reverse primer (0.5 µg/µl), 0.5 µl Vent™ (6.0 µl total) DNA Polymerase (2000 U/ml), 1.5 µl each dNTPs (10mM, 6.0 µl total), 1.0 µl miniprep DNA from test clone, 50 µl total reaction volume. The PCR program was as follows: 95 °C for 5 minutes, 30 Cycles of: 95 °C for 1 minute, 60 °C for 1 minute, 72 °C for 30 seconds, 72 °C for 10 minutes, hold at 4 °C. The presence of the correct-sized insert was confirmed by agarose gel electrophoresis and ethidium bromide staining. The confirmed clones were used for the *S. pneumoniae* gene KNOCKOUT procedure. Glycerol stocks were made of all positive *E. coli* LE392 constructs and frozen at – 80 degrees C.

EXAMPLE 3

The following provides a description of the high throughput gene disruption procedure used in *S. pneunomiae* strain (e.g., gene knockout procedure). The candidate *ceg* fragments that were generated by the method described in Example 2 were used in the gene disruption procedure in order to identify *ceg* nucleotide sequences that are required for cell viability.

Reactions were set up in a 1.5 ml eppendorf tubes or 96 well plate (1 µg total of miniprep pEVP-3 + insert DNA (usually 10 µl of Qiagen miniprep DNA); then 200 µl of *S. pneumoniae* (strain Rx-1) competent cells diluted 1:10 in competence media was added (1 ml of competence media = 980 µl Todd Hewitt (Difco Laboratories) with 0.5% yeast extract, 20 µl 10% BSA, 1 µl 10 % CaCl2, and 0.5 µl (200 µg/ml) Csp-1 competence peptide).

Controls were run with each KNOCKOUT experiment and involved 1 μg pEPV3 Lyt A construct = positive control (non-essential), or 1 μg pEPV3 Fts Z construct = negative control (essential). Then the 96 well plates and controls were incubated at 37 °C for 2.5 to 3 hours in 37 °C room without shaking. The 200 μl of the samples were plated on Todd Hewitt agar plates with 0.5% yeast extract and 2 μg/ml chloramphenicol.

The samples were incubate over night at 37 °C in 5% CO_2 incubator. Control plates were checked for presence of colonies (pEVP-3::lytA) and no growth (pEVP-3::ftsZ). Plates were examined for growth (ca. 70-150 colonies) designating nonessentials and zero colonies designating essential genes.

The polarity test was performed in a similar manner, using the polarity fragments described in Example 3.

EXAMPLE 4

The following provides a description of the autolysin procedure used to determine that the non-essential control samples of *S pneumoniae* contain a disrupted *lytA* gene.

Phenotypic Autolysin Test

The culture plates containing transformants carrying the *lytA* control vector were flooded with 0.1% deoxycholate in H_2O. The plates were observed after 5-10 minutes. Plates with "ghosts" indicated intact *lytA* gene, or plates without "ghosts" indicated a disrupted *lytA* gene. The "ghost" phenomenon is due to detergent triggered autolysis of the cells, causing a gradual fading of the colonies.

The detergent treatment triggers the autolysin in *lytA* intact cells; it cannot trigger the autolysin (*lytA* gene product) in *lytA* disrupted cells. Colonies with intact *lytA* "ghost" in 5-10 minutes due to massive pneumococcal cell lysis.

EXAMPLE 5

The following provides a description of the procedure used to express the CEG proteins (e.g., designated CFE proteins) in *E. coli* cells.

CEG Protein Production

Full-length *ceg* gene were inserted into pET-21 expression vector using the E. coli BL21 λDE3 expression system using the following method:

For each *ceg*, custom primers were used to insert N- and C- termini into vectors such that the 5' end (N-terminus of the CEG) is positioned properly for expression behind the T7 promoter and optimally placed with regard to the pET ribosome binding site. The pET vectors contain an *NdeI* site which allows positioning of ATG start site in the vector. In cases where the *ceg* sequence contains an internal *NdeI* site, blunt ligation of the *ceg* PCR fragment into the vector is accomplished via Klenow fill-in of the *NdeI* site. In many cases, primers were also designed such that the *ceg* 3' (C-terminus of the expressed protein) will contain an in-frame extension of 6X-histidine residues, encoded in the vector sequence of pET-21. The individual *ceg*s were PCR amplified via custom designed primers as described above. Both *ceg* PCR and vector DNA were digested with appropriate restriction enzymes. The full-length *ceg* were ligated into the pET expression vector. The ligation mixture was transformed into competant *E. coli* BL21 λDE3 cells and selected for transformants on LB agar with 50 μg/ml ampicillin. Positive insert bearing clones were screened via minipreps of the plasmids and size analysis on 0.8% agarose gels, with detection by ethidium bromide staining, as above.

Protein Production

The proper reading frame of each *ceg* inserted into pET-21 is verified by DNA sequencing.

A small (2-5 ml) test culture of *E. coli* BL21 λDE3 with the insert-bearing plasmid is tested for protein expression by IPTG induction of the expression vector for 1-2 hours. The expression is verified by SDS-Polyacrylamide Gel Electrophoresis analysis of a whole cell extract (SDS extract of 0.5-1 ml of cells treated at 100 °C for 5 minutes) to determine whether the protein is over-expressed and migrates at the correct predicted molecular weight.

The protein is overproduced and purified via the following method. A large scale (500-1000ml) culture of *E. coli* is grown to early logarithmic phase in broth (e.g., LB broth) and protein expression induced for 2 hours with IPTG (isopropyl-D-thiogalactoside). The cells are harvested by centrifugation (8000 X G; 15 minutes) and the cell pellets resuspended in 20 ml. of buffer. The cells are lysed by sonication, and the supernatant fluid centrifuged at low speed (5000 X G, 15 min.) to remove unbroken cells. The supernatant fluid, containing the over-expressed protein is subjected to Ni- NTA affinity column chromatography (Quiagen, Inc., Chatsworth, CA). The 6X-histidine residues linked at the C-terminal end of the CEG proteins permit rapid protein purification via selective binding to a Ni-NTA resin column. The protein-bound Ni-NTA resin was to remove contaminants, and the bound proteins subsequently eluted with imidazole and recovered. It is possible to upscale this procedure to larger volumes for higher yields of proteins.

EXAMPLE 6

The following provides a description of the methods used to purify all 2CEG polypeptides (e.g., 2CFE polypeptides #19-117; SEQ ID NOS:349-436) having a histidine tag at their C-terminal ends. The 2CEG polypeptides having the his-tags were produced by the methods described in Example 5, *supra.* As an example, results of purification of 2CFE 75 polypeptide are presented.

Production Of The CFE Polypeptides

The BL21λDE3 cells harboring recombinant pET-21 vectors carrying a 2CFE nucleotide sequence (SEQ ID NOS:244-331) were cultured in LB broth containing ampicillin. When the A_{600} reached approximately 0.6, protein production was induced by adding 1.0 mM of IPTG, the cells were cultured for an additional 2 hours. The cell pellet was collected by centrifugation, and the collected cell pellet was sonicated in Solution A (50 mM $NaPO_4$; 300 mM NaCl, pH 8.0). The sonicated cells were centrifuged at 10,000 RPM to remove the debris.

Purification Of The CFE Polypeptide

The supernatant was diluted with Solution A, loaded onto a Ni-NTA column (Quiagen) equilibrated with Solution A; the column bed size was 2.5 x 25 cm, and the flow rate was approximately 3.0 ml/minute. The 2CFE protein was eluted using a linear gradient of imidazole, using 0-250 mM in 450 ml, flow rate approximately 3.0 ml/minute. The eluted samples were collected as 22 ml fractions per tube and the eluted samples were monitored using spectrophotometry. The amount of protein in the eluted fractions was estimated using the Bradford method (Bradford, M. M., 1976 *Anal. Biochem.* 72:248) and the samples were run on an SDS-PAGE gel (Novex EC6008) (Figure 3 A). Fractions were selected for pooling based on the results of the SDS-PAGE gel. The pooled fractions were concentrated using a 10,000 MW Centricon (Amicon) to approximately 5 ml.

The 2CFE 75 polypeptide, a precipitate formed and was redissolved upon increasing the sample volume and removing the imidazole by repeated concentration in 50 mM Tris, 100 mM NaCl, pH 7.5. Varying amounts of the 2CFE 75 polypeptide were diluted in either 20 mM Tris, 20 mM KCl, pH 7.5 or 20 mM Tris, 20 mM $MgCl_2$, pH 7.5 at concentrations of 12, 24, or 36 ug/ml. The diluted samples were electrophoresed on an SDS-PAGE gel under non-reducing conditions (Figure 3 B). The results of Figure 3 B suggests that 2CFE 75 forms a multimer.

EXAMPLE 7

The following provides a description of the methods used to purify CEG polypeptides that lack a histidine tag (e.g., 2CFE polypeptides #1-17; SEQ ID NOS:332-348). As an example, the results of purification of CFE 3 polypeptide are presented.

Purification of the CFE 3 Polypeptide

The 2CFE 3 polypeptide was produced using the large scale IPTG-induced method described in Example 5, *supra*. The 2CFE 3 (SEQ ID NO:334) polypeptide lacks a C-terminal histidine tag. The 2CFE 3 polypeptide was purified using a 2-column procedure. The 2CFE 3 polypeptide preparation was eluted from a 26/10 Q Sepharose column (Pharmacia) using a 0-1.0 M NaCl gradient, 2 ml/minute flow rate, and the gradient size was 1 liter. Then the 2CFE 3 polypeptide was eluted from a hydroxyapatite Bio-gel column (Bio-Rad) using a 5-200 mM potassium phosphate (pH 8.0) gradient, the flow rate was 0.3 ml/minute, and the gradient size was 300 ml. A sample of the 2CFE 3 preparation was run on a polyacrylamide gel (Figure 4).

EXAMPLE 8

The following provides a description of the size exclusion chromatography methods used to estimate the molecular weight and determine whether the CEG polypeptides oligomerize. The CFE polypeptide may olimerize to form monomers, dimers, tetramers, hexameric rings, or other oligomeric forms.

Size exclusion chromatography was performed on all isolated 2CFE polypeptides #s 1-117 (e.g., SEQ ID NOS:332-436). This method was performed using various types of columns, depending on the particular 2CFE polypepeptide tested.

The Biosil SEC-125 HPLC Gel Filtration column (BioRad Laboratories, Inc) was used, for example, to characterize CFE 8. The mobile phase was 0.2 M KH_2PO_4, 0.9% NaCl pH 6.8.

The Phenomenex 600 x 7.5 mm Biosep SECS 3000 column was used, for example to characterize 2CFE 21 and 39. The mobile phase for size exclusion was 50 mM Na_2HPO_4, pH 7.0 and 150 mM NaCl run at 1 ml/minute in a Gilson HPLC system, with protein detection at 280 nm.

EXAMPLE 9

The following provides a description of the computer-aided methods used to search for similarities between the amino acid sequences of the CEG polypeptides and sequences available through public and proprietary databases. In many cases, the function of the CEG polypeptides was suggested by the results of the similarity searches. The function of some of these CEG polypeptides has been confirmed by performing additional analyses. Table V provides a list of the suggested and confirmed functions of CEG polypeptides designated CFEs #1-117.

The putative function of the CFE polypeptides were determined using computer-aided bioinformatic approaches, including distant homologies, motif searching, or predictions based on statistical rules. For example, the distant homology approach involved pairwise or multiple sequence alignments, employing tools such as FASTA, and Psi-BLAST. The motif searching approach involved using sophisticated hidden Markov models. The approach based upon predictions of statistical rules involved prediction of transmembrane regions, coiled-coil, and other structural motifs. These approaches have been reviewed in *Computational Methods In Molecular Biology* 1998, eds. Salxber, S.L., Searls, D.B. Searls, and Kasif, S. , Elsevier, and in *Bioinformatics: A Practical Guide To The Analysis Of Genes And Proteins* 1998 eds Baxevanis, A. D. and Francis Ouellete, B.F. , Wiley-Interscience.

Global sequence similarity searches were performed using the amino acid sequences of all the conserved essential gene sequences (e.g., CFEs 1-117; SEQ ID NOS:114-226) to search against a non-redundant protein database using the BLAST2 algorithm (Altschul S.F., et al., 1997 *Nucleic Acids Res.* 25(17):3389-3402). In a similar search, similar sequences were identified in the Concordance database using the "Neighbor" function (Bruccoleri R. E., Dougherty T.J., Davison D.B. 1998 *Nucleic Acids Res.* 26(19):4482-4486). To determine if the predicted amino acid sequences were full length and in the proper reading frame, BLAST-type searching and CLUSTAL multiple sequence alignments (Higgins D.G., et al., 1996 *Methods Enzymol.* 266:383-402) were used. Local sequence similarity searches were performed, by searching for Prosite (Hofmann K., et al., 1999 *Nucleic Acids Res.* 27(1):215-219) and Pfam motifs (Bateman A., et al., 2000 *Nucleic Acids Res.* 28(1):263-266). Additionally, the amino acid sequences of the CFEs were analyzed by performing protein threading analyses using the ProCeryon fold recognition program (Sippl, et al., 1992 *Proteins* 13:258-271; Sippl, J. 1993 *J. Comp. Aided Mol. Design* 7:473-501; www.proceryon.com) and Geneformatics.

In bacteria, many operons include genes encoding different proteins that catalyze discrete steps of a common biochemical pathway. Therefore, the operon structures in *S. pneumoniae* was compared with that in other bacteria in order to predict the function of CFE polypeptides.

Additionally, analysis of bacterial metabolic pathways were performed using Pathway Tools from DoubleTwist, based on the EcoCyc system (Karp P.D., et al., 1999 *Nucleic Acids Res.* 1999 27(1):55-58). This analysis was used to predict which CFEs mediate various steps of the pathways.

When the sequence identity between a CFE polypeptide and the annotated database (e.g., SwissProt, Genbank) was low (e.g., sequence identity less than about 30%), a Protein Threading (e.g., fold recognition) method was used to predict similarities in the folded protein structure of CFE polypeptides in the absence of a high level of sequence similarity with proteins in the databases (review by Teichmann, et al., 1999 *Current Opinion in*

Structural Biology 9:390-399). The Protein Threading method predicts the compatibility of a query sequence (e.g., CFE polypeptide sequences) with each of the folds in a library of known protein structures. The library of known protein structures as developed, maintained, and updated throughout the search process.

A list of potential structural folds, onto which each query was compatible, was generated for all CFE polypeptides (e.g, SEQ ID NOS:114-226). The fold assignments for each query were used to generate pairwise sequence alignments. The pairwise sequence alignments were used to generate protein models of the query polypeptide (e.g., CFE polypeptides).

The pairwise sequence alignments were also used to compare the position of critical residues of the structural template with the query polypeptide. The list of critical residues was generated by using multiple sequence alignments derived from a structural classification of proteins to generate a conservation profile which provided sequence-specific positions conserved across a homologous family of protein folds. Comparative modeling was used to search the model of the query polypeptide for the critical residues and determine whether the structural and functional motifs are conserved in the query protein. Conservation of structural and functional motifs permitted assignment of putative structure and function to a query polypeptide sequence.

The Protein Threading method was used to search for putative folded structure and function for all CFE polypeptides (SEQ ID NOS:114-226). The CFE polypeptides having significant sequence identity (e.g., more than 30%) to known proteins were assigned putative functions with a high level of confidence.

EXAMPLE 10

The following provides a description of the methods used to characterize purified, CFE 101 polypeptide. The 2CFE 101 polypeptide mediates the conversion of pantothenate to 4' phosphophantothenate, and is predicted to be a pantothenate kinase.

Computer-Aided Comparison

The computer-aided comparison, as described in Example 9 *supra*, suggests that the amino acid sequence of the CFE 101 polypeptide (SEQ ID NO:210) is 42% similar to the amino acid sequence of the coaA protein of *E. coli.* Thus, CFE 101 may be a pantothenate kinase, which mediates the conversion of pantothenate to 4' phosphophantothenate (Figure 5).

Circular Dichroism and Circular Dichroism Thermal Melt Analysis

Circular dichroism and circular dichroism melt methods were used to determine the folded structure of the expressed and isolated 2CFE polypeptides. For example, this method was used to characterize the folded structure of isolated 2CFE 101 (SEQ ID NO:421).

The starting concentration of the 2CFE 101 polypeptide was such that OD_{205} was approximately 1.5, and the OD_{280} was approximately 0.05 (e.g., 0.05 to 0.1 mg/ml). The starting concentration of 2CFE 101 was approximately 344 μM in 50% glycerol, 50 mM Tris, 100 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA, at pH 7.5. The polypeptide was diluted to a final concentration of 7 μM, as determined by absorbance at A_{280}, in 20 mM Na-phosphate, 100 mM KCl, at pH 7.0. The circular dichroism analysis was performed using quartz cuvettes, the instrumentation was from JASCO (Model J-720), the readings were performed at 25 degrees C (Figure 6 A). The band width was 1 nm, the sensitivity was 20 mdeg, the response was 0.25 seconds, the scan speed was 50 nm/minute, and the step was 0.5. The circular dichroism thermal melt analysis was performed at a range of between 0 and 100 degrees C (Figure 6 B). Additionally, the circular dichroism was performed comparing monomer and aggregate pools of 2CFE 101.

Size Exclusion Analyses

Size exclusion chromatography methods were performed using the Biosil SEC column, as described in Example 8 *supra*. The results suggest that the 2CFE 101 polypeptide forms monomer (40,200 Da) and oligomers (194,000 Da). The specific activity of the monomer and oligomeric forms of 2CFE 101 were determined, as described below.

Biochemical Assays

The biochemical assays of the 2CFE 101 polypeptide was based on the PK/LDH coupled enzyme assays described by Vallari, D. S., et al. (1987 *J. Biol. Chem.* 262:2468-2471) and Song, W. –J., et al., (1994 *J. Biol. Chem.* 269:27051-27058).

Briefly, the assay was performed as follows. The reaction included: 885 µl of 0.1 M Tris-HCl (pH 7.6), 25 µl NADH (14.1 mM), 20 µl ATP (10.7 mM), 50 µl phospho-enol-pyruvate (56 mM), 5 µl LDH/PK (lactose dehydrogenase/PK; Sigma, catalog # P-0294, 60 U/ ml PK, 1050 U/ml LDH), 5 µl of the 2CFE 101 polypeptide (9 mg/ml in 50 mM Tris-HCl, pH 7.5, 100 mM NaCl which was diluted to 4.5 mg/ml in 50% glycerol). The reaction was started by adding 10 µl pantothenate (100 mM; Sigma, catalog # P2250). The production of ADP in the reaction was monitored by measuring the absorbance a 340 nm. The results in Figure 8 show that the 2CFE 101 polypeptide mediates ADP production in the presence of pantothenate and ATP. The K_m of pantothenate (n=4) was 144 (±16.5) µM, the V_{max} of the 2CFE 101 polypeptide (n=4) was 2.04 (±0.25) µM min^{-1} mg^{-1}. The monomer form has a specific activity of approximately 1.7 µM min^{-1} mg^{-1}. The oligomeric form has a specific activity of 0.26 µM min^{-1} mg^{-1}.

Alternatively, the 2CFE 101 polypeptide can be tested in an assay that monitors the conversion of pantothenate to 4'-phosphopantothenate. The same reaction described above can be used, except ^{14}C-labeled pantothenate is used. The reaction can be monitored by measuring the amount of ^{14}C-labeled 4'-phosphopantothanate produced.

EXAMPLE 11

The following provides a description of the methods used to characterize purified, CFE 39 and CFE 21 polypeptides, carrying a C-terminal histidine 6-tag. The methods include helicase reactions, in which synthetic Holliday Junction templates are resolved into duplex structures. In one method, helicase reaction was monitored using radiolabeled templates. In another method, the helicase assay was adapted for use in a high throughput assay employing fluorescence labeled templates.

Computer-Aided Comparison

The computer-aided comparison, as described in Example 9 *supra*, suggests that the CFE 39 polypeptide (SEQ ID NO: 148) is an RuvA homologue. The comparison also suggests that CFE 21 (SEQ ID NO:132) is an RuvB homologue.

Previous studies by Parsons and others have shown that RuvA and RuvB proteins, in *E. coli,* promote branch migration or movement of Holliday Junctions during genetic recombination and DNA repair (Parsons, C. A., et al., 1992 *Proc. Natl., Acad. Sci. USA* 89:5452-5456; Tsaneva, I. R., et al., 1993 *Proc. Natl., Acad. Sci. USA* 90:1315-1319; Muller, B., et al., 1993 *J. Biol. Chem.* 268:17179-17184; Mitchell, A. H. and S. C. West 1996 *J. Biol. Chem.* 271:19497-19502; Parsons, C. A. and S. C. West 1993 *J. Molec. Biol.* 232:397-405; Tsaneva, I. R., et al., 1992 *Molec. Gen. Genet.* 235:1-10; Mitchell, A. H. and S. C. West 1994 *J. Molec. Biol.* 1994 243:208-215).

Size Exclusion Chromatography

Size exclusion chromatography was performed on 2CFE 39 (SEQ ID NO:366) and 2CFE 21 (SEQ ID NO:350) using the Phenomenex 600 x 7.5 mm Biosep SECS 3000 column, as described in Example 8 *supra*. Protein standards (BioRad) were used to calibrate the column, including thyroglobulin (670,000 Da), gamma globulin (158,000 Da), ovalbumin (44,00 Da), myoglobin (17,00 Da), and B-12 (1350 Da).

The results indicate that 2CFE 39 (RuvA) forms tetramers and 2CFE 21 (RuvB) forms a hexameric ring structure. Selected eluted samples were electrophoresed on a polyacrylamide gel (Novagen) (Figure 9).

The Holliday Junction Analysis Using Radiolabeled Templates

The Holliday Junction analysis was performed using radiolabeled, synthetic, asymmetrical, Holliday Junction templates, as described in Hiom, K. and S. C. West 1995 *Cell* 80:787-793. The Holliday Junction templates were produced by annealing together four separate, single-stranded, oligonucleotide strands to form four-stranded structures (e.g., the Holliday Junction template). The Holliday Junction templates were reacted with the 2CFE 39 and 2CFE 21 polypeptides, in a helicase reaction, to test their ability to generate two duplex structures.

Producing the Synthetic Holliday Junction Templates

The asymmetrical Holliday Junction templates were produced by annealing the following oligonucleotide sequences:

Oligonucleotide strand 1:
5'-CCAGTGATCACATACGCTTTGCTAGGACATCTTGATATCAGCCCACGTT
CACCCGCCTACCAGTGCCACGTTGTATGCCCACGTTGACC-3' (SEQ ID NO:438)

Oligonucleotide strand 2:
5'-GGGTCAACGTGGGCATACAACGTGGCACTGGTAGGCGGGTGAACGTGGG
CTGATATCAAGATGTCCATCTGTCCGTTCATCTATGACGT-3' (SEQ ID NO:439)

Oligonucleotide strand 3:
5'-AACGTCATAGATGAACGGACAGATCATGGTGCTTTTAAAGTCTAGAGAC
TATCGAGCATTAGTACCAGTATCGAATCCGTCTTGTCAA-3' (SEQ ID NO:440)

Oligonucleotide strand 4:

5'-TTTGACAAGACGGATTCGATACTGGTACTAATGCTCGATAGTCTCTAGAC TTTAAAAGCACCATGTAGCAAAGCGTATGTGATCACTG-3' (SEQ ID NO:441)

Oligonucleotide strand 3 was labeled at the 5' end using approximately 300 ng of oligonucleotide strand 3, 1 µl 10x Phosphate Buffer, 5 µl ^{32}P ATP, 1 µl T4 polynuclotide kinase (Gibco-BRL)), in a 10 µl volume, and the reaction was performed at 37 degrees C for 30 minutes. The reaction was loaded onto a G50 column to remove the unincorporated radiolabel. The final concentration of the radiolabeled oligonucleotide strand 3 was approximately 15 ng per µl.

Approximately equimolar amounts of the four oligonucleotide strands were annealed (e.g., hybridized). The annealing reaction included: 5 µl Annealing Buffer (200 mM Tris-Cl pH 8.0, 100 mM $MgCl_2$, 1 M NaCl, 10 mM DTT); 450 ng of radiolabeled oligonucleotide strand 3; and 1000 ng each of oligonucleotide strands 1, 2, and 4; in 50 µl total reaction volume. The control annealing reaction included: 5 µl Annealing Buffer, 60 ng radiolabeled oligonucleotide strand 3; 1000 ng oligonucleotide strand 4; in 50 µl total reaction volume. Annealing was performed at 95 degrees C for 5 minutes, 65 degrees C for 30 minutes, 42 degrees C for 30 minutes, and room temperature (e.g., between about 23 to 27 degrees C) for 30 minutes to generate the synthetic Holliday Junction templates. The synthetic Holliday Junction templates were gel or column-purified to remove the duplex and non-annealed products. As a control, oligonucleotide strands 3 and 4 were annealed to form duplex structures. The synthetic Holliday Junction templates and duplex structures were stored at –20 degrees C.

CFE 39 and CFE 21: The Helicase Reaction Using Radiolabeled Templates

The helicase reaction was performed to determine whether 2CFE 39 and 2CFE 21 resolved the synthetic Holliday Junction templates into duplex structures. The helicase reaction was performed as follows. A 50 µl total reaction volume included: 25 µl of 2x

Reaction Buffer (50 mM Tris-Cl pH8.0, 30 mM $MgCl_2$, 2 mM ATP); 1 µl synthetic Holliday Junction template (36 ng); 2 µl 2CFE 39 (1 µM); and 2 µl 2CFE 21 (1 µM). The reaction was incubated at 37 degrees for 30 minutes. The reaction was stopped by adding 5 µl Stop Buffer (100 mM Tris-Cl pH 7.5, 5 mg/ml Proteinase-K, 5% SDS). The stopped reaction was returned to 37 degrees C for 5 minutes. The helicase reaction was loaded onto and run on a non-denaturing, 12% PAGE, Tris-glycine gel.

The results shown in Figure 10, lanes 6, 7 and 8, indicate that the 2CFE 39 and 2CFE 21 polypeptides resolved the synthetic Holliday Junction templates into duplex structures.

CFE 39: The Helicase Reaction

It has been previously shown that *E. coli* RuvA binds to Holliday Junction templates (Parsons, C. A., et al., 1992 *Proc. Natl., Acad. Sci. USA* 89:5452-5456). The ability of *S. pneumoniae* CFE 39 to bind to a Holliday Junction template can be tested by employing the helicase assay described herein. The results of the helicase assay can be monitored by performing a gel shift assay and/or capillary electrophoresis. The presence of a Holliday Junction template bound to 2CFE 39, which migrates more slowly than the Holliday Junction template alone, would indicate that *S. pneumoniae* 2CFE 39 binds to Holliday Junction templates.

CFE 39 and CFE 21: Holliday Junction Analysis Using Fluorescent-Labeled Templates

The helicase reaction described herein was performed using Holliday Junction templates having one oligonucleotide strand labeled with a fluorescent agent and another strand labeled with a quenching agent. The 5' fluorescent end and the 3' quenching end of the strands that make up the Holliday Junction templates are in proximity to each other, resulting in a non-fluorescent template. When the Holliday Junction templates are resolved into duplex structures, the fluorescent and quench ends are not in proximity to each other, resulting in fluorescence.

The Holliday Junction templates used to perform this experiment comprised the following: the 5' end of oligonucleotide strand 1 was labeled with a fluorescein (e.g., the fluorescent agent), and the 3' end of oligonucleotide strand 4 was labeled with DABCYL (e.g., the quenching agent). The oligonucleotide strand 1 labeled with fluorescein and the oligonucleotide strand 4 labeled with DABCYL were custom synthesized (Gibco-BRL Life Technologies, Inc.).

The fluorescein and DABCYL labled oligonucleotides were annealed in a reaction, as described above, to generate synthetic Holliday Junction templates. The helicase reaction was performed as described above. The results of the helicase reaction were monitored by measuring the unquenching of the Holliday Junction templates with time (Figure 11).

The helicase assay using Holliday Junction templates labeled with fluorescent-quenching agents can be adapted for use in high throughput analyses to test 2CFE 39, 2CFE 21, and other polypeptides for their ability to resolve the templates into duplex structures.

EXAMPLE 12

The following provides a description of the methods used to characterize purified, CFE 8 polypeptide, which lacks a histidine tag. The CFE 8 is a putative DNA single-stranded binding protein.

Computer-Aided Comparison

The computer-aided comparison, as described in Example 9 *supra*, suggests that the CFE 8 polypeptide (SEQ ID NO:121) may be a single stand binding protein homologue, such as SSB.

Size Exclusion Chromatography

The 2CFE 8 polypeptide (SEQ ID NO:339) was characterized by size exclusion chromatography, using the Biosil SEC-125 HPLC Gel Filtration column as described in Example 8 *supra*. The chromatogram showed one peak corresponding to a molecular weight of approximately 89 kDa. Based on the nucleotide sequence, the predicted molecular weight of 2CFE 8 is 17,351 Da. In non-denaturing conditions, 2CFE 8 forms a multimer.

Binding Reaction

The 2CFE 8 polypeptide was reacted with a single-stranded oligonucleotide A. Briefly, the binding reaction included: 50 μM of 2CFE 8 polypeptide, 50 μM oligo strand A, 20 mM Tris/20 mM KCl pH 7.5. The binding reaction was performed at 37 degrees C, for 2 hours.

Oligonucleotide strand A:
5'-TTAGGGCCCGGGCTATCTTACAATCTCGTT-3' (SEQ ID NO:442)

Capillary Electrophoresis

The results of the binding reaction was monitored by capillary electrophoresis, following the methods described in "Handbook of Capillary Electrophoresis" 2nd Edition, 1997, ed. J. Landers.

Separation was performed using an uncoated capillary tube (360 μm o.d., 50 μm i.d., with a 50 cm effective separation length; Watrex International, Inc., Pittsford, NY) and 50 mM borate pH 9.3 as the mobile phase, at 25 kVolts, 20 minutes separation time.

The results indicate that 2CFE 8 alone elutes as a sharp peak, indicating little adsorption to the uncoated capillary wall (Figure 12 A). The shape of the peak and peak retention

time changed with 2CFE 8 in the presence of all oligonucleotides tested (Figure 12 B). As a negative control, MurB polypeptide (Pucci, M. J., L. F. Discotto, and T. J. Dougherty 1992 "Cloning and Identification of the Escherichia coli murB DNA sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine reductase" *J. Bacteriol.* 174:1690-1693) was reacted with the same oligonucleotides. MurB reacted with or with out the oligonucleotides showed no change in peak shape or retention time.

After capillary electrophoresis analyses, the 2CFE8 alone and 2CFE plus oligonucleotide samples were run on native polyacrylamide gels to determine whether the polypeptide was intact. The results indicate that in all cases, 2CFE 8 was intact and had not degraded with time or storage.

Mobility Shift Assays

The ability of 2CFE 8 polypeptide to bind oligonucleotide strand A was tested in a mobility shift assay.

The results indicate that 2CFE 8 binds single stranded oligonucleotides (Figure 13 A and B). In Figure 13 A, the gel was stained with ethidium bromide. The unbound oligonucleotides appear near the bottom of the gel, while the bound oligonucleotides appear near the middle. The same gel was stained with Coomassie (Figure 13 B), revealing that 2CFE 8 polypeptide bound to the oligonucleotide migrated further than unbound 2CFE 8, due to the change in charge carried by the oligonucleotide. Various ratios of 2CFE8:oligo were tested. The optimal binding ratio was 2:1.

The Effect of $MgCl_2$

The 2CFE 8 polypeptide precipitated in the presence of 5 mM $MgCl_2$. The precipitation was reversible by the addition of 1 μM of the oligonucleotides tested. The observation indicates specific binding between 2CFE 8 polypeptide and the oligonucleotides tested.

Scintillation Proximity Assay

Scintillation proximity assay (SPA) methods can be used in a high throughput screening procedure to monitor, for example, a binding reaction. SPA utilizes beads (Amersham) which are coated on the surface with a particular compound or molecule. For example, the SPA bead may be coated with avidin to facilitate binding with any molecule having a biotin tag.

The binding reaction of the 2CFE 8 polypeptide and the oligonucleotide strand A can be monitored using SPA beads and a scintillation counter. The beads can be coated with avidin, the 2CFE 8 polypeptide can be tagged with biotin, and the oligonucleotide strand A can be radiolabeled.

EXAMPLE 13

The following provides a description of the methods used to characterize purified, 2CFE 3 (SEQ ID NO:334) and 2CFE 86 (SEQ ID NO:409) polypeptides.

The 2CFE 3 polypeptide catalyzes the conversion of D-glucosamine-6-phosphate to D-glucosamine-1-phosphate, indicating that 2CFE 3 mediates amino-sugar biosynthesis through the N-acetyl glucosamine pathway (Figure 14).

The 2CFE 86 polypeptide catalyzes the conversion of D-glucosamine-1-phosphate to N-acetylglucosamine-1-phosphate, and the conversion of N-acetylglucosamine-1-phosphate to UDP-N-acetylglucosamine-1-phosphate, which indicates that 2CFE 86 also mediates amino-sugar biosynthesis through the N-acetyl glucosamine pathway (Figure 14).

Computer-Aided Comparisons Of CFE 3

The computer-aided comparison, as described in Example 9 *supra*, suggested that the CFE 3 polypeptide (SEQ ID NO:116) is a phosphoglucosamine mutase, such as GlmM.

Purification of the CFE 3 Polypeptide

The 2CFE 3 polypeptide was produced using the large scale IPTG-induced method described in Example 5, *supra*. The 2CFE 3 polypeptide lacks a C-terminal histidine tag. The 2CFE 3 polypeptide was purified using a 2-column procedure. The 2CFE 3 polypeptide preparation was eluted from a 26/10 Q Sepharose column (Pharmacia) using a 0-1.0 M NaCl gradient, 2 ml/minute flow rate, and the gradient size was 1 liter. Then the 2CFE 3 polypeptide was eluted from a hydroxyapatite Bio-gel column (Bio-Rad) using a 5-200 mM potassium phosphate (pH 8.0) gradient, the flow rate was 0.3 ml/minute, and the gradient size was 300 ml. A sample of the 2CFE 3 preparation was electrophoresed on an SDS polyacrylamide gel (Figure 4).

Affinity Capillary Electrophoresis of CFE 3

Affinity capillary electrophoresis methods were used to determine whether the 2CFE 3 polypeptide binds to various glucose derivatives. Binding was performed under equilibrium conditions, in which the sugars were dissolved in the running buffer and reacts with 2CFE 3 during separation in the column. The affinity capillary electrophoresis method used to analyze 2CFE 3 follows the methods described in "Handbook of Capillary Electrophoresis" 2nd Edition, 1997, ed. J. Landers.

Briefly, 2CFE 3 polypeptide was reacted with increasing amounts of various glucose derivatives (e.g., substrate) at 25, 30 and 37 degrees C. The glucose derivatives included UDP-glucose, glucose-1-phosphate, glucose-6-phosphate, glucosamine-1-phosphate, and glucosamine-6-phosphate. The reaction included: 2CFE 3 polypeptide (2.0 mg/ml), separation buffer (25 mM Tris; 192 mM Glycine, pH 8.0; BupH Tris-Glycine Buffer Packs, Pierce). Separation was performed at 25 kVolts, separation time was 15 or 20 minutes.

The results shown in Figure 15 A indicate that at 25 degrees C, 2CFE 3 binds to D-glucose-1-phosphate in a dose-dependent manner, as the peak shape and/or the retention

time for 2CFE 3 changes in the presence of 100 and 500 μM D-glucose-1-phosphate compared to unreacted 2CFE 3.

The results shown in Figure 15 B indicate that at 25 degrees C, 2CFE 3 binds to D-glucosamine-6-phosphate in a dose-dependent manner, as the peak shape and/or the retention time for 2CFE 3 changes in the presence of 100 and 500 μM D-glucosamine-6-phosphate compared to unreacted 2CFE 3.

The results shown in Figure 15 C indicate that at 25 degrees C, the 2CFE 3 polypeptide also binds to glucose-6-phosphate.

A comparison of 2CFE 3 reacted with various glucose derivatives, at 30 degrees C, is shown in Figure 15 D. The results indicate that D-glucosamine-6-phosphate is a putative substrate for 2CFE 3, as this reaction exhibits the greatest change in peak shape and/or retention time.

CFE 3: Capillary Electrophoresis and Laser-Induced Fluorescence

In a further analysis of 2CFE 3 polypeptide, capillary electrophoresis was performed with laser-induced fluorescence in order to separate and detect interaction between the substrate (e.g., D-glucosamine-6-phosphate) and the product (e.g., D-glucosamine-1-phosphate) in a one dose, one time-point procedure.

The 2CFE 3 polypeptide was derivitized by reacting 10 mM FITC (fluorescein isothiocyanate dissolved in methanol; Calbiochem, San Diego, CA) with D-glucosamine-6-phosphate, at ambient temperature, in the dark, overnight. The FITC-derivatized 2CFE 3 polypeptide (2.0 mg/ml) was reacted with the substrate (D-glucosamine-6-phosphate and D-glucosamine-1-phosphate) for one hour.

Separation was performed using an uncoated capillary (360 μm o.d., 50 μm i.d., with a 50 cm effective separation length) and 50 mM borate (pH 9.3) as the mobile phase. The

argon-ion laser had an excitation wavelength of 488 nm and an emission filter of 520 nm (Beckman, Fullerton, CA). The results shown in Figure 16 indicate that 2CFE 3 binds and catalyzes the conversion of D-glucosamine-6-phosphate to D-glucosamine-1-phosphate.

Computer-Aided Comparison Of CFE 86

The comparison results, as described in Example 9 *supra*, suggested that the CFE 86 polypeptide (SEQ ID NO:195) is an acetyltransferase, such as GlmU which is a bifunctional enzyme in *E. coli*. It has been previously shown that, in *E* coli, GlmU is a bifunctional protein having both the acetyltransferase and uridylyltransferase active sites (Mengin-Lecreulx, D. and J. van Heijennort 1994 *J. Bacteriol.* 176:5788-5795; Gehring, Al., et al., 1996 *Biochemistry* 35:579-585). The bifunctional enzyme catalyzes the conversion of D-glucosamine-1-phosphate to N-acetylglucosamine-1-phosphate (acetyltransferase), and catalyzes the conversion of N-acetylglucosamine-1-phosphate to UDP-N-acetylglucosmine-1-phosphate (uridylyltransferase). The Km of the acetyltransferase and uridylyltransferase reactions has been previously calculated (Mengin-Lecreulx, D. and J. van Heijennort 1994 *supra*). Additionally, the crystal structure of GlmU from *E. coli* is known (Brown, K., et al., 1999 *EMBO J.* 18:4096-4107).

Purification of the CFE 86 Polypeptide

The 2CFE 86 polypeptide (SEQ ID NO:409) has a C-terminal histidine tag. The 2CFE 86 polypeptide was produced using the large scale IPTG-induced method described in Example 5, *supra*. The 2CFE 86 polypeptide was purified using the Ni-NTA affinity column method described in Example 6, *supra*. The eluted 2CFE 86 polypeptide was dialyzed against 50 mM Tris-Cl, 100 mM NaCl, 25% glycerol, pH 8.0. Samples of the purified 2CFE 86 polypeptide were electrophoresed on a polyacrylamide gel (Figure 17).

Coupling CFE 3 and CFE 86 to Produce UDPAG

A biochemical assay was performed, to determine whether 2CFE 3 and 2CFE 86 convert D-glucosamine-6-phosphate to UDP-N-acetylglucosamine-1-phosphate (e.g., UDPAG). The 2CFE 3 and 2CFE 86 polypeptides were used in a coupled reaction based on the assays described in Jolly, L. P., et al., 1999 *Eur. J. Biochem.* 262:202-210.

A time-dependent and dose-dependent assay were performed. Briefly, the assay was performed in 96-well plates, each well including 100 µl volume. The assay included: 1 mM D-glucosamine-6-phosphate (Sigma); 0.7 mM D-glucosamine-1,6-diphosphate (Sigma); 1.2 mM acetyl-Coenzyme A (Sigma); and 5 mM uridine-5'-phosphate (Sigma); 3 mM $MgCl_2$ (Sigma); 50 mM Tris-Cl, pH 8.0 (Life Technologies). The reaction was started by adding 1 µg of 2CFE 3; and 10 µg of 2CFE 86. The reaction was performed at room temperature. The reaction was stopped at 0, 15, 30, and 65 minutes, by filtering out the 2CFE polypeptides.

The results of the assay was monitored by HPLC (high pressure liquid chromatography) using an Optisil 10µ SAX column (250 x 4.6 mm), measuring at 262 nm, the mobile phase was 150 mM KH_2PO_4 (pH 3.5), and 1.5 ml/minute flow rate. The results shown in Figure 18 show the time-dependent assay and indicate that HPLC detected the presence of UDPAG.

CFE 86: The Uridylyltransferase Reaction

The 2CFE 86 polypeptide was tested in a uridylyltransferase reaction, in which N-acetyl-D-glucosamine-1-phosphate and UTP produce UDP-N-acetylglucosamine. The uridylyltransferase reaction was monitored using a malachite green/inorganic pyrophosphatase assay (e.g., malachite green-IPPAse assay) and/or monitored using HPLC. The malachite green-IPPAse assay was used to measure orthophosphate production from digestion of the pyrophosphate liberated in the uridylyltransferase reaction.

The malachite green reagent was prepared as follows. A 0.045 % solution of malachite green (Sigma; M9636) was prepared in water. A 4.2 % solution of ammonium molybdate (Mallinckrodt) was prepared in 4N HCl. The malachite green and ammonium molybdate were mixed in a 3:1 ratio, and stirred for about 20 minutes. The mixture was filtered, and stored at 4 degrees C. The inorganic pyrophosphatase (Sigma; I-2267) was diluted to 0.1 U/μl in 50 mM Tris/3mM $MgCl_2$ ph 8.0, and stored at 4 degrees C.

The uridylyltransferase reaction was performed in 96-well plates. The coupled reaction described herein was performed, in the presence of 2CFE 3 alone or 2CFE 3 and 2CFE 86, and included the addition of 0.5 U/well of the diluted inorganic pyrophosphate. The reaction was mixed for 5 minutes at room temperature. The reaction was stopped by the addition of 240 μl/well of the malachite green reagent and 30 μl/well of 34% sodium citrate, and the reaction was mixed. The results of the uridylyltransferase reaction was monitored by spectrophotometry at 660 nm.

The results of separate uridylyltransferase reactions were monitored by HPLC, using a Phenosphere-NEXT C18 column (250 x 4.6 mm). The mobile phases included A and B as follows: A) methanol/10 mM potassium phosphate pH 6.5 (0:100); and B) methanol/10 mM potassium phosphate pH 6.5 (40:60). The mobile phases were run under the following conditions: 100% mobile phase A for 5 minutes, to 100% mobile phase B in 3 minutes; and hold 100% mobile phase B for 9 minutes. The retention time for the UDPAG product is approximately 5.75 to 6.0 minutes.

The results three uridylyltransferase reactions, monitored by HPLC are summarized in Table III below.

TABLE III

Purified CFE 86:	Specific Activity (nmol/min/μg):
2CFE 86-1	3.1
2CFE 86-2	3.4
2CFE 86-3	3.1

The results of the uridylyltransferase reactions, monitored by HPLC or HPLC and Malachite Green IPPAse assays are summarized in Table IV below.

TABLE IV

Reaction:	Km (μM):	Method:
Acetyltransferase reaction:		
Glucosamine-1-P	94	HPLC
Acetyl-coA	150	HPLC
Uridylytransferase reaction:		
N-acetylglucosamine-1-P	48	HPLC and MG/IPPAse
UTP	79	HPLC

EXAMPLE 14

The following provides a description of the methods used to characterize various 2CFE polypeptides, including CFE 21, 34, 35, 39, and 90. The molecular weight of these 2CFE polypeptides were analyzed by size exclusion chromatography and gel electrophoresis. The 2CFE 34, 35, and 90 polypeptides putatively mediate fatty acid biosynthesis.

Computer-Aided Comparison

The computer-aided comparison, as described in Example 9 *supra*, suggests that CFE 34 (SEQ ID NO:143), CFE 35 (SEQ ID NO:144), and 90 (SEQ ID NO:199) are polypeptides which mediate a fatty acid biosynthesis pathway (Figure 19)

The comparison suggests that CFE 34 is a malonyl CoA:ACP transcylase, which catalyzes the reaction in which malonyl CoA and acyl carrier protein (ACP) are converted to malonyl-ACP and CoA. Thus, the CFE 34 polypeptide may be a homologue of *E. coli* FabD.

The comparison suggests that CFE 90 is a 3-oxoacyl-ACP synthase II (beta ketoacyl-ACP synthase II) which catalyzes the reaction in which malonyl-ACP is converted to beta aceto acetyl-ACP. Thus, the CFE 90 polypeptide may be a homologue of *E. coli* FabF.

The comparison suggests that CFE 35 is a 3-oxoacyl-ACP reductase (beta aceto acetyl ACP reductase) which catalyzes the reaction in which beta-keto-acetyl-ACP is converted to beta-hydroxy-acetyl-ACP. Thus, the CFE 35 polypeptide may be a homologue of *E. coli* FabG.

Size Exclusion Chromatography

The estimated molecular weights of 2CFE 34 (SEQ ID NO:361), 2CFE 35 (SEQ ID NO:362), and 2CFE 90 (SEQ ID NO:413) were determined using the Biosil SEC-125 HPLC Gel Filtration column as described in Example 8, *supra.*

The results suggest that 2CFE 34 polypeptide is a monomeric protein (33,093 Da), 2CFE 35 is a trimeric protein (25,758 Da; approximately 85%), and 2CFE 90 is a dimeric protein (43,930 Da). Selected eluted samples of 2CFE 34 were electrophoresed on a polyacrylamide gel (Figure 20).

Biochemical Assay: CFE 34

The function of 2CFE 34 was determined by performing various biochemical reactions. To determine whether 2CFE 34 catalyzes the convertion of malonyl-CoA to malonyl and CoA, the following reaction was performed.

The biochemical reaction was performed in the presence of acyl carrier protein. The reaction included the following: 10 μM ^{14}C labeled malonyl-CoA, 20 μM ACP, 30 μM 2CFE 34 (e.g., FabD) in 20 mM Tris-Cl, pH 8.0 and 5 mM DTT in 300 μl volume. The reaction was performed at room temperature (e.g., approximately 24 degrees C) for 30 minutes. The reaction was terminated with the addition of 45μl of 0.5% TFA. The labeled reaction was injected onto a MonoQ 5/5 column on a Gilson HPLC. Detection was performed by monitoring the radioactivity of the continuous flow-through of the HPLC effluent. Chromatography was performed using a buffer gradient for column elution. Buffer A included 20 mM Tris-Cl, pH 8.3. Buffer B was the same as Buffer A and included 1 M NaCl. The program was held at 90% A, 10% B for 10 minutes followed by a linear ramp to a final mix of 50% of each Buffer A and B over 10 minutes.

The substrate (e.g., ^{14}C malonyl-CoA) eluted at 9.9 minutes, the product (e.g., ^{14}C malonyl-ACP) eluted at 14.3 minutes. The results indicate that CFE 34 catalyzes the conversion of malonyl-CoA and acyl carrier protein (ACP) to malonyl-ACP and CoA.

EXAMPLE 15

The following provides a description of the methods used to characterize CFE polypeptides 40, 41, and 46.

Computer-Aided Comparison

The computer-aided comparison, as described in Example 9 *supra*, suggests that the CFE 40 polypeptide (SEQ ID NO:149) is a phosphomethylpyrimidine (HMP-P) kinase involved in thiamine biosynthesis.

The comparison, as described in Example 9 *supra*, suggests that the CFE 41 polypeptide (SEQ ID NO:150) has a GTP-binding motif and may be a protease.

The comparison, as described in Example 9 *supra*, suggests that the CFE 46 polypeptide (SEQ ID NO:155) has an ATP-binding motif.

Affinity Purification of CFE 41

The large-scale method described in Example 5 *supra* (e.g., IPTG-induced protein production) was used to prepare a sample of 2CFE 41 polypeptide (SEQ ID NO:368). The sample was affinity purified using the Ni-NTA method described in Example 6, *supra*. The eluted fractions were loaded onto and run on a 12% SDS-PAGE gel (Novex) (Figure 21).

Circular Dichroism and Circular Dichroism Thermal Melt Analysis

Circular dichroism and circular dichroism thermal melt methods were performed using JASCO instrumentation. The concentration of the isolated 2CFE 40 (SEQ ID NO:367) was approximately 21 μM, in a 0.1 cm pathlength cell at 210 nm. The circular dichroism spectrum suggests that this preparation of 2CFE 40 had mixed alpha and beta secondary structure. The circular dichroism thermal melt spectrum suggests that 2CFE 40 has a T_m of approximately 67 degrees C. The 2CFE 40 polypeptide precipitates at approximately the T_m.

The concentration of the isolated 2CFE 41 (SEQ ID NO:368) was approximately 70 μM, in a 0.02 cm pathlength cell. The circular dichroism spectrum suggests that this preparation of 2CFE 41 had mixed alpha and beta secondary structure, with a greater percentage of alpha structures. The circular dichroism thermal melt spectrum suggests that 2CFE 41 has a T_m of approximately 38 degrees C. The 2CFE 41 polypeptide precipitates at approximately the T_m.

The concentration of the isolated 2CFE 46 (SEQ ID NO:373) was approximately 23 μM, in a 0.1 cm pathlength cell at 280 nm. The circular dichroism spectrum suggests that this preparation of 2CFE 46 had mixed alpha and beta secondary structure. The circular dichroism thermal melt spectrum suggests that 2CFE 46 is highly stable at elevated temperatures. At 90 degrees C, the 2CFE 46 polypeptide exhibited only a 27% loss in signal and the polypeptide remained soluble.

Capillary Electrophoresis

Capillary electrophoresis was performed on samples of purified 2CFE 40, 41 and 46. The electropherograms of 2CFE 40, 41, and 46 are shown in Figure 22.

EXAMPLE 16

The following provides a description of methods that can be used to characterize CEG polypeptides (e.g., CFE polypeptides).

Computer-Aided Compilation

Computer-aided compilation of bacterial metabolic pathways may be analyzed using Pathway Tools from Doubletwist, based on the EcoCyc system (Karp P.D., et al., 1999 *Nucleic Acids Res.* 1999 27(1):55-58). This analysis may be used to predict which CFEs mediate various steps of the pathways. This information may be used in combination

with the results of a binding reaction which identifies a ligand or substrate that binds with a CFE polypeptide.

Identifying the Function of a CFE Polypeptide

The function of a CFE polypeptide may be identified by identifying a ligand or substrate which binds with the CFE polypeptide. The ligand or substrate may be identified using fractionation and affinity capillary electrophoresis methods. The following method is based upon the assumption that the bacterial cell lysate includes the ligand or substrate.

A bacterial host cells carrying an endogenous (e.g. native) CFE gene or carrying a recombinant vector which includes a CFE gene may be cultured so that the CFE polypeptide is produced by the cell. The cells may be ruptured in order to obtain the cell lysate. The cell lysate may be fractionated using HPLC technology. The HPLC fractions may be reacted with a CFE polypeptide in a binding reaction, and the binding reaction may be analyzed by affinity capillary electrophoresis methods. The ligand or substrate which reacts with the CFE polypeptide may be identified using mass spectrophotometry methods (in "Mass Spectrometry" 1990 eds. McCloskey, J. A., in *Methods in Enzymology* volume 193; Henion, J., et al., 1993 "Mass Spectrometric Investigations of Drug-Receptor Interactions" *Ther. Drug Monit.* 15:563-569; Loo, J. A., et al., 1999 "Application of Mass Spectrometry for Target Identification and Characterization" *Med. Res. Rev.* 19:307-319; Nguyen, D. N., et al., 1995 "Protein Mass Spectometry: Applications to Analytical Biotechnology *J. Chromatogr.*705:21-45).

EXAMPLE 17

The following provides a description of nuclear magnetic resonance (NMR) spectroscopy methods that were used to characterize CFE polypeptides.

High resolution NMR spectroscopy was applied to ^{15}N-labled, $^{13}C/^{15}N$-labeled, $^{2}H/^{13}C/^{15}N$-labeled, and type-specifically isotopically labeled CFE polypeptide samples

in the solution state for the following purposes: to assess various aspects of the structural state, e.g., foldedness, structural integrity; to refine a previously determined experimental structure of a close sequence homologue; to refine a homology-modeled structure; to assess the potential for a CFE polypeptide to bind small molecules; and to identify small-molecule pharmacophoric fragments that bind specifically to the CFE polypeptide ("Nuclear Magnetic Resonance" 1994 eds. James, T. L. in *Methods in Enzymology* volume 239).

The NMR analysis includes screening both a compound deck of approximately 4,500 commercially available, structurally and chemically diverse compounds (the small-molecule pharmacophore deck) and a compound deck of proprietary, known, anti-microbial compounds (anti-microbial deck) against the CFE polypeptides (i.e., target polypeptides) to determine, either based upon perturbations to the chemical shifts of the amide proton and/or nitrogen resonances, as measured from a two-dimensional proton-nitrogen heteronuclear single-quantum correlation spectrum (2D screening method), or based upon increases in the linewidth of the compound's proton resonance(s), as measured by a one-dimensional $T_{1\rho}$ spin-lock difference spectrum (1D screening method), both whether a compound binds to a CFE polypeptide and, in the case of the 2D screening method, where the compound binds on the CFE polypeptide.

Isotopic Labeling of CFE Polypeptides

BL21-DE3 *E. coli* bacteria are transformed with the CFE expression vectors. Expression takes place between 20°C and 37°C in minimal media containing [^{15}N]-ammonium sulfate as the sole nitrogen source and either glucose, [^{2}H]$_{13}$-glucose, or [^{13}C]$_{6}$-glucose as the sole carbon source. Glucose is used for preparing uniformly ^{15}N-labeled and ^{2}H/^{15}N-labeled CFE polypeptides. [^{2}H]$_{13}$-glucose is used for preparing type-specifically ^{1}H/^{13}C-labeled, uniformly ^{15}N-labeled CFE polypeptides. [^{13}C]$_{6}$-glucose is used for preparing ^{13}C/^{15}N-labeled CFE polypeptides. The minimal media is prepared in 100% H_2O for expressing both uniformly ^{15}N-labeled and uniformly ^{13}C/^{15}N-labeled CFE polypeptides; the minimal media is prepared in 95% D_2O (deuterium oxide) and 5% H_2O for expressing

both type-specifically $^{1}H/^{13}C$-labeled, uniformly ^{15}N-labeled and just uniformly $^{2}H/^{15}N$-labeled CFE polypeptides. In the case of type-specifically $^{1}H/^{13}C$-labeled, uniformly ^{15}N-labeled CFE polypeptides, 40 mg/L of protonated and uniformly $^{13}C/^{15}N$-labeled isoleucine, valine and leucine amino acids are added to the minimal media.

NMR Screening

Compounds in the anti-microbial deck are pre-dissolved to a target concentration of 16 mM in deuterated DMSO (dimethylsulfoxide) with each deck well containing only one compound. Compounds in the small-molecule, pharmacophore deck are pre-dissolved in deuterated dmso to a target concentration of 50 mM in groups of 8, i.e., each deck well contains 8 unique compounds with each compound at a target concentration of 50 mM.

3.5 µl of compound is placed at the bottom of a well in a 96-well, screening plate. This well will be referred to as the compound screening well. Each compound screening well contains solution from only one deck well. 166.5 µl of buffer is added to each compound screening well. 170 µl of a CFE polypeptide solution, initially at a concentration ranging from 200-300 µM, is added to each compound screening well; the contents of that well are then thoroughly mixed. The control screening well contains only 3.5 µl of deuterated dmso. The screening plate is then centrifuged in a bucket rotor for 15 minutes at 3,500 rpm to insure that all particulate matter is at the bottom of the well.

The 2D screening method requires a single control screening well in which the compound solution consists only of deuterated DMSO. The 1D screening method requires a control screening well for each compound screening well. In the case of the 1D screening method, the control screening well is prepared identically to the compound screening well except that the 170 µl of a CFE polypeptide solution is replaced by 170 µl of buffer.

The screening plate is covered with aluminum foil and placed onto a rack of a Gilson liquid handler. The Gilson liquid handler, under computer control by the NMR host/data-acquisition software, is responsible for removing each sample from the screening plate,

injecting the sample into a high-resolution, $^{1}H/^{15}N$ double-resonance NMR flow-probe, removing the sample from the flow-probe, and dispensing it back into the screening plate well from which the sample was originally removed. NMR data are collected on the sample while the sample resides in the NMR flow-probe. The type of NMR data collected depends upon whether the 2D or 1D screening method is being used.

Determining Structural Characteristics of a CFE Polypeptide

In assessing various aspects of the structural state of a CFE polypeptide, NMR was used to provide the following information. The proton 1D spectra and proton-nitrogen 2D correlation NMR spectra were used to assess the overall foldedness of a CFE polypeptide without actually describing in detail that folded state. Unfolded and substantially misfolded proteins produced distinct signatures in these two types of NMR spectra.

The chemical shift of most protein nuclei in either the set $\{H_N, H_\alpha, H_\beta, C', C_\alpha, C_\beta, N\}$ or the set $\{H_N, C', C_\alpha, C_\beta, N\}$ for perdeuterated (e.g., ^{2}H-labeled) proteins were determined by procedures well known in the art that involve collecting up to 10 triple-resonance NMR data sets. The protein secondary structure was delineated as either helical, turn or extended (e.g., β-sheet) by measuring $\Delta(\delta_{C\alpha} - \delta_{C\beta})$, $\Delta\delta C'$, and $\Delta\delta_{H\alpha}$ where δ refers to the chemical-shift value and Δ refers to the difference between chemical-shift values measured in this protein and those measured for the same residue type in a random-coil (unstructured), tetrameric peptide.

This secondary-structure profile was generated in approximately 2-3 weeks per protein. The secondary-structure profile was used to confirm the functional identity of a protein. It was also used to refine the list of possible functional identities of folds, predicted by various computational techniques including fold recognition which is associated with a protein or polypeptide.

NMR was used to generate folds of proteins or polypeptides for which both no structure was known of a sequence homologue and no structural homologue was discernible in the PDB by fold recognition techniques.

Refining a Structural Model

Nuclear Overhauser (NOE) data were used to refine both homology-modeled structured and previously determined experimental structures of close sequence homologues. This process took approximately 2-3 weeks per structure.

The CFE 88 polypeptide was characterized by NMR analysis to establish its secondary structure. The NMR data was used to filter the computer-aided threading analysis. The NMR-determined secondary structure for CFE 88 suggested that CFE 88 is structurally similar to 4-aminoimidazole carboxylase.

The characteristics of other CFE polypeptides were analyzed by NMR methods. A computer-aided threading analysis revealed that the N-terminal domain of the protein EGA, which both binds and hydrolyzes GTP, was both structurally similar and sufficiently similar in sequence to CFE 52 to suggest that CFE 52 had a similar function.

The NMR data of CFE 103 suggests that this polypeptide is unfolded. Circular dichroism spectra, as a function of temperature, also indicated that CFE103 was unfolded.

The CFEs 2, 42, 43, 68 and 88 polypeptides were tested for their ability to bind potential inhibitor molecules by screening both the anti-microbial deck and the small-molecule, pharmacophore deck. CFE 34 was tested for its ability to bind potential inhibitor molecules by screening the anti-microbial deck.

Characterizing Small-Molecule Binding

NMR-based screening was used to measure binding against both the small-molecule, pharmacophore deck and the anti-microbial deck. Binding data from these screens allowed assessment of the propensity of a protein to bind small molecules. The binding data was also used to identify sites on the protein which are capable of binding small molecules. The binding data was also used to identify common pharmacophores among the compounds which bind.

Reverse screening refers to a process whereby known anti-microbial compounds, the microbial target of which is unknown, are screened by a general method, e.g., binding as assessed by NMR, to find a physical interaction with polypeptide targets previously determined to be essential to the bacteria (i.e., the CFEs). The reverse screening method was used to determine which CFE polypeptides bind to which compounds in the anti-microbial deck. The reverse screening method included the following. The compounds in a proprietary compound deck were screened for Minimal Inhibitory Concentration (e.g., MIC). The compounds exhibiting antimicrobial activity were designated active compounds. The CFE polypeptides were screened to determine which polypeptide bind to which active compounds. The CFE polypeptides which bound to the active compound(s) were confirmed, where possible, i.e., in cases where an in-vitro assay was possible to construct, as being inhibited in their function as a polypeptide by the active compound(s) by examination of the inhibition profile of the compound(s) against the CFE polypeptides. For additional confirmation, the effect of the compound on the microorganism harboring the CFE polypeptide was monitored (e.g., whole cell assays). The structure of the active compound was used as a basis to generate chemically-related compounds by iterative synthesis. The chemically-related compounds were tested in a screening assay for binding with CFE polypeptides. The active compounds and the chemically-related compounds of interest were the compounds which exhibited an increase in binding affinity for a CFE polypeptide and/or exhibited drug-like properties.

The results of the reverse screening are as follows. 127 compounds from the proprietary compound deck exhibited anti-microbial activity. 94 of these active compounds were selected based upon both lack of cytotoxicity and lack of excessive hydrophobicity. These 94 compounds were soluble to 16 mM in deuterated DMSO; these compounds were also deemed to be sufficiently soluble in aqueous buffer for both the 2D and 1D NMR screening methods.

This subset of 94 compounds was used in an NMR-based screen to determine which compound binds to which CFE polypeptide. The CFE 42 polypeptide bound two different compounds with K_d's in the range of 0.2 to 1 mM; the CFE 43 polypeptide bound one compound with K_d ~ 30-50 μM; the CFE 34 polypeptide bound 13 compounds, one of which inhibited the polypeptide function with IC_{50} < 10 μM.

The enzyme assay used to confirm the NMR results which suggested CFE 34 interaction with the compounds included the following: 10 μM ^{14}C-labeled malonyl CoA; 20 μM ACP, 30 pM CFE 34; 20 mM Tris-Cl, pH 8.0; 5 mM DTT; in the presence of absence of 50 μM of a compound solubilized at 40 mM in 100% DMSO and dilute 100-fold into 10% DMSO and further diluted 8-fold for a final concentration of 50 μM in 1.25% DMSO. The reaction was performed at room temperature, the reaction was stopped with the addition of TFA. Two hundred μl of the reaction was injected onto a Mono Q 5/5 column. The chromatography conditions included: A) 20 mM Tris-Cl, pH 8.3; B) 20 mM Tris-Cl, pH 8.3, 1 M NaCl. Hold 10% B for 5 minutes, linear gradient from 10% B to 50%B in 10 minutes, back to 10% B in 1 minute, hold for 14 minutes to re-equilibrate. The reaction substrate (^{14}C- malonyl CoA) eluted at 9.9 minutes, the reaction product (^{14}C-malonyl ACP) eluted at 14.3 minutes.

What is claimed is:

1. An isolated nucleic acid molecule encoding a polypeptide which is (1) essential for the viability of a bacterial cell and (2) has at least any one of the functions of a pantothenate kinase, a Holliday Junction branch migration protein, a single stranded DNA binding protein, a phosphoglucosamine mutase, an acetyltransferase, an uridylyltransferase, a malonyl CoenzymeA:ACP transcylase, a 3-oxoacyl-ACP synthase II, a 3-oxoacyl-ACP reductase, a phosphomethylpyrimidine (HMP-P) kinase, a GTP binding protein, a ATP binding protein, or a 4-aminoimidazole carboxylase.

2. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:97 or Figure 115 and wherein the polypeptide is a pantothenate kinase.

3. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:35, Figure 60, SEQ ID NO:19, or Figure 44,and wherein the polypeptide is a Holliday Junction branch migration protein.

4. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:8 or Figure 33 and wherein the polypeptide is a single stranded DNA binding protein.

5. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:3 or Figure 28 and wherein the polypeptide is a phosphoglucosamine mutase.

6. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:82 or Figure 103 and wherein the polypeptide is a acetyltransferase.

7. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:82 or Figure 103 and wherein the polypeptide is a uridylyltransferase.

8. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:30 or Figure 55 and wherein the polypeptide is a malonyl CoenzymeA:ACP transcylase.

9. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:86 or Figure 107 and wherein the polypeptide is a 3-oxoacyl-ACP synthase II.

10. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:31 or Figure 56 and wherein the polypeptide is a 3-oxoacyl-ACP reductase.

11. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:36 or Figure 61 and wherein the polypeptide is a phosphomethylpyrimidine (HMP-P) kinase.

12. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:37, Figure 62, SEQ ID NO:48, or Figure 73, and wherein the polypeptide is a GTP binding protein.

13. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:42 or Figure 67 and wherein the polypeptide is a ATP binding protein.

14. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:84 or Figure 105 and wherein the polypeptide is a 4-aminoimidazole carboxylase.

15. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid molecule is shown in SEQ ID NO:48 or Figure 73 and wherein the polypeptide is a GTP binding protein.

16. An isolated nucleic acid molecule encoding a polypeptide which is essential for the viability of a bacterial cell, the nucleic acid molecule comprising a sequence shown in any one of SEQ ID NOS:1-113.

17. An isolated nucleic acid molecule encoding a polypeptide which is essential for the viability of a bacterial cell, the nucleic acid molecule comprising a sequence shown in any one of Figures 26-130.

18. An isolated nucleic acid molecule encoding any one of a polypeptide designated CFE 1-117 having the amino acid sequence shown in SEQ ID NO:114-226.

19. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of claim 1, 16, 17 or 18.

20. The isolated nucleic acid molecule of claim 1, 16, 17 or 18 which is DNA or RNA.

21. The isolated nucleic acid molecule of claim 20, which is labeled with a detectable marker.

22. The isolated nucleic acid molecule of claim 21, wherein the detectable marker is selected from the group consisting of a radioisotope, a fluorescent compound, a

bioluminescent compound, a chemiluminescent compound, a metal chelator and an enzyme.

23. A vector comprising the nucleotide sequence of claim 1, 16, 17, or 18.

24. A host-vector system comprising the vector of claim 23, in a suitable host cell.

25. The host-vector system of claim 24, wherein the suitable host cell is selected from a group consisting of a yeast cell, a plant cell, and an animal cell.

26. The host-vector system of claim 24, wherein the suitable host cell is selected from a group consisting of an *Escherichia* cell, a *Bacillus* cell, a *Pseudomonas* cell, a *Streptococcus* cell, and a *Streptomyces* cell.

27. An isolated polypeptide which is essential for the viability of a bacterial cell comprising the amino acid sequence as shown in any one of SEQ. ID NOS: 114-226.

28. An isolated polypeptide which is essential for the viability of a bacterial cell encoded by the isolated nucleic acid molecule of claim 1, 16, 17, or 18.

29. The isolated polypeptide of claim 27 or 28 which is a fusion polypeptide.

30. A method for producing a polypeptide having the amino acid sequence of any one of SEQ ID NOS: 114-226 or a polypeptide encoded by the polynucleotide sequence as shown in any one of Figures 26-130, comprising:

a) culturing the host-vector system of claim 24 under suitable conditions so as to produce the polypeptide; and

b) recovering the polypeptide so produced.

31. A polypeptide produced by the method of claim 30.

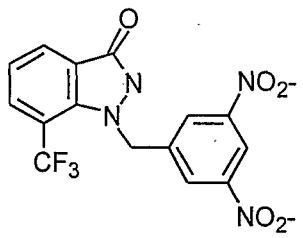
32. A ligand which binds the polypeptide of claim 27 or 28.

33. The ligand of claim 32 which is an antibody or an immunologically active fragment thereof.

34. The ligand of claim 33, wherein the antibody is a monoclonal antibody.

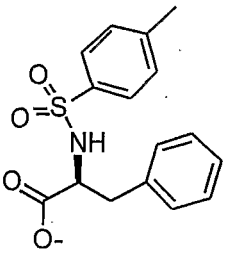
35. The ligand of claim 32 which is a diazalactone.

36. The ligand of claim 35, wherein the diazalactone comprises the structure:



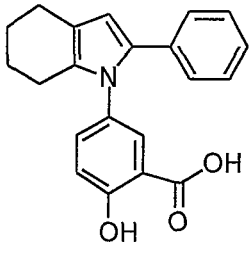
37. The ligand of claim 32 which is a *N*-protected amino acid.

38. The ligand of claim 37, wherein the *N*-protected amino acid comprises the structure:



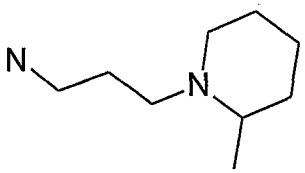
39. The ligand of claim 32 which is an azabicyclodiene.

40. The ligand of claim 39, wherein the azabicyclodiene comprises the structure:

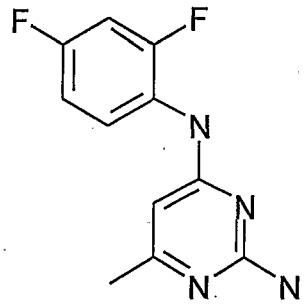


41. The ligand of claim 32 which is an alkaloid.

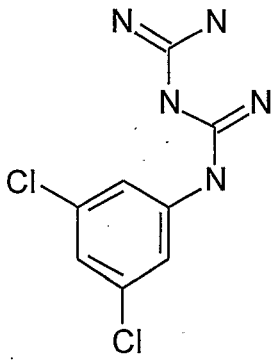
42. The ligand of claim 41, wherein the alkaloid comprises the structure:



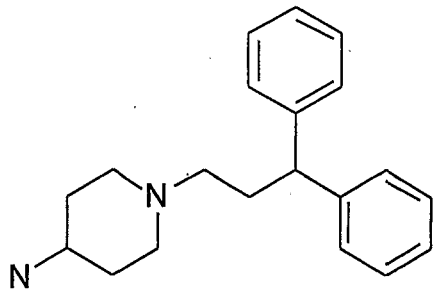
43. The ligand of claim 41, wherein the alkaloid comprises the structure:



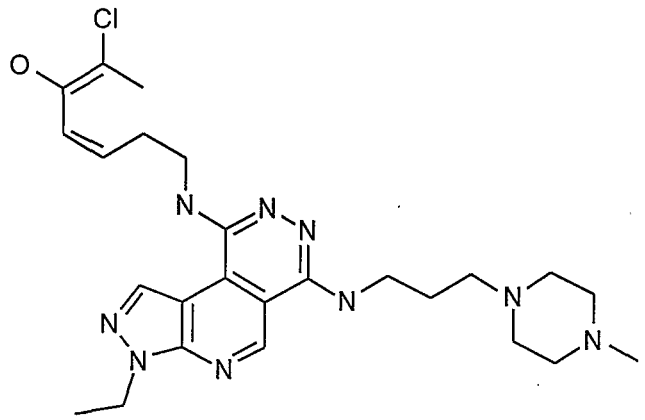
44. The ligand of claim 41, wherein the alkaloid comprises the structure:



45. The ligand of claim 41, wherein the alkaloid comprises the structure:



46. The ligand of claim 41, wherein the alkaloid comprises the structure:



47. A method for detecting the presence of the polypeptide of claim 27 or 28 in a sample, comprising contacting the sample with a ligand which binds the polypeptide and detecting the binding of the polypeptide with the ligand in the sample.

48. The method of claim 47, wherein the detecting comprises:

a) contacting the sample with the ligand; and

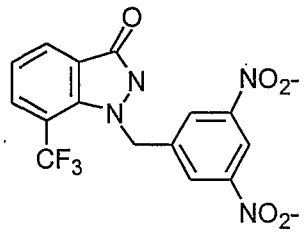
b) determining whether a polypeptide-ligand complex is so formed.

49. The method of claim 47, wherein the sample is a cell, a tissue, or a biological fluid.

50. The method of claim 47, wherein the sample is blood, serum, a swab from nose, a swab from ear, or a swab from throat.

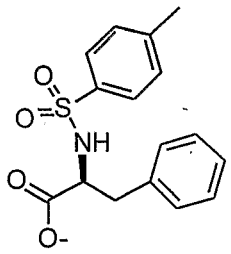
51. The method of claim 47, wherein the ligand is a diazalactone.

52. The method of claim 51, wherein the diazalactone comprises the structure:



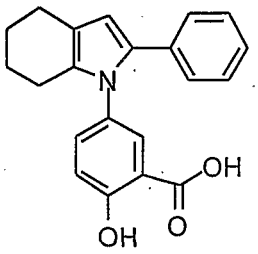
53. The method of claim 47, wherein the ligand is a *N*-protected amino acid.

54. The method of claim 53, wherein the *N*-protected amino acid comprises the structure:



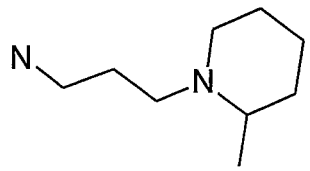
55. The method of claim 47, wherein the ligand is an azabicyclodiene.

56. The method of claim 55, wherein the azabicyclodiene comprises the structure:

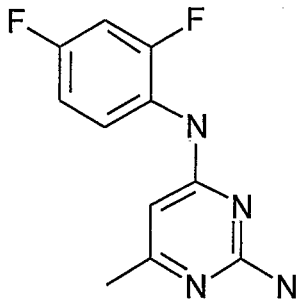


57. The ligand of claim 47 which is an alkaloid.

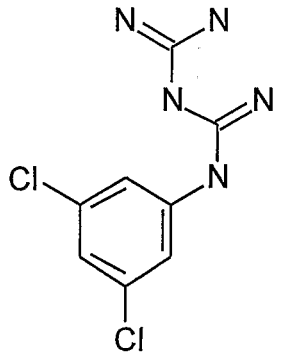
58. The ligand of claim 57, wherein the alkaloid comprises the structure:



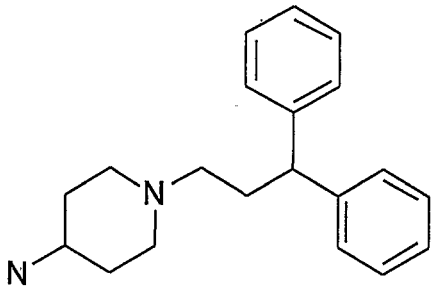
59. The ligand of claim 57, wherein the alkaloid comprises the structure:



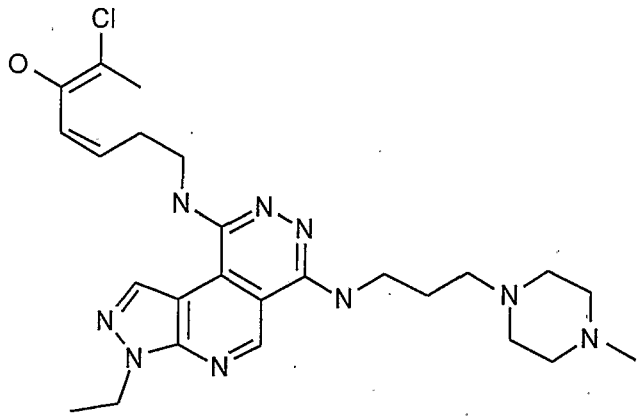
60. The ligand of claim 57, wherein the alkaloid comprises the structure:



61. The ligand of claim 57, wherein the alkaloid comprises the structure:



62. The ligand of claim 57, wherein the alkaloid comprises the structure:



63. A method for detecting the presence of a target nucleic acid molecule as shown in any one of SEQ ID NOS:1-113 in a sample, comprising contacting the sample with the complementary nucleic acid molecule of claim 19 and detecting the binding of the target nucleic acid molecule with the complementary nucleic acid molecule in the sample.

64. The method of claim 63, wherein the detecting comprises:

a) contacting the sample with the complementary nucleic acid molecule; and

b) determining whether a complex comprising the target nucleic acid molecule and the complementary nucleic acid molecule is so formed.

65. The method of claim 63, wherein the sample is a cell, a tissue, or a biological fluid.

66. The method of claim 63, wherein the sample is blood, serum, a swab from nose, a swab from ear, or a swab from throat.

67. A pharmaceutical composition comprising the nucleic acid molecule of claim 1, 16, 17, or 18.

68. A pharmaceutical composition comprising the polypeptide of claim 27 or 28.

69. A pharmaceutical composition comprising the ligand of claim 32.

70. A method for determining whether a genomic nucleotide sequence of interest is essential for viability of a bacterial cell, comprising

a. integrating an exogenous nucleotide sequence into the genomic nucleotide sequence of interest, wherein the exogenous nucleotide sequence comprises a portion of an open reading frame of the genomic nucleotide sequence of interest, and

b. determining whether the cell having the genomic nucleotide sequence of interest so integrated is viable.

71. The method of claim 70, wherein the portion of the open reading frame comprises about 200 to 500 base pairs in length.

72. The method of claim 70, wherein the exogenous nucleotide sequence further comprises a nucleotide sequence conferring a selectable phenotype to the cell having the genome so integrated.

73. The method of claim 70, wherein determining comprises selecting the cell having the genome so integrated in the presence of a selection agent.

74. The method of claim 73, wherein the selection agent is chloramphenicol.

75. A nucleotide sequence of interest which is essential for viability of a bacterial cell isolated by the method of claim 70.

76. A bacterial cell comprising an exogenous nucleotide sequence integrated into the genomic nucleotide sequence of interest, generated by the method of claim 70.

77. A method for determining whether a genomic nucleotide sequence of interest resides within an operon, comprising

a) integrating an exogenous nucleotide sequence into the genomic nucleotide sequence of interest; and

b) determining whether the cell having the genomic nucleotide sequence of interest so integrated is viable, and wherein the exogenous nucleotide sequence lacks an expression regulatory sequence.

78. The method of claim 77, wherein the exogenous nucleotide sequence further comprises a nucleotide sequence conferring a selectable phenotype to the cell having the genome so integrated.

79. The method of claim 77, wherein determining comprises selecting the cell having the genome so integrated in the presence of a selection agent.

80. The method of claim 79, wherein the selection agent is chloramphenicol.

81. A method for inhibiting a function of a CEG polypeptide which is essential for viability of a bacterial cell, the method comprising contacting the CEG polypeptide with the ligand of claim 32 under suitable conditions thereby inhibiting the function of the CEG polypeptide.

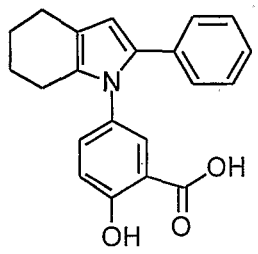
82. The method of claim 81, wherein the function of the CEG polypeptide is selected from a group consisting of a pantothenate kinase, a Holliday Junction branch migration protein, a single stranded DNA binding protein, a phosphoglucosamine mutase, an acetyltransferase, an uridylyltransferase, a malonyl CoenzymeA:ACP transcylase, a 3-oxoacyl-ACP synthase II, a 3-oxoacyl-ACP reductase, a phosphomethylpyrimidine (HMP-P) kinase, a GTP binding protein, a ATP binding protein, or a 4-aminoimidazole carboxylase.

83. The method of claim 81, wherein the CEG polypeptide is selected from a group consisting of CFE1-113.

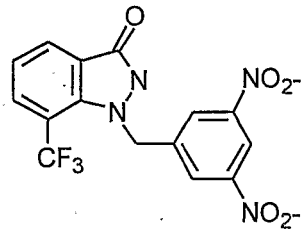
84. The method of claim 81, wherein the CEG polypeptide is 2CFE 34 shown in Figure 55.

85. The method of claim 81, wherein the CEG polypeptide is 2CFE 43 shown in Figure 64.

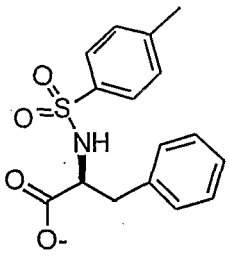
86. The method of claim 81, wherein the CEG polypeptide is 2CFE 34 shown in Figure 55 and the ligand is:



87. The method of claim 81, wherein the CEG polypeptide is 2CFE 43 shown in Figure 64 and the ligand is:



88. The method of claim 81, wherein the CEG polypeptide is 2CFE 43 shown in Figure 64 and the ligand is:



89. A method for identifying a ligand in a sample which specifically binds a CEG polypeptide, the method comprising:

a) contacting the CEG polypeptide with the sample under suitable conditions so that a complex having the CEG polypeptide and the ligand is formed;

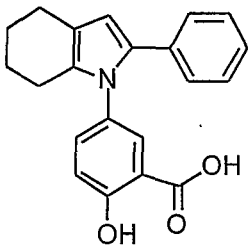
b) recovering the complex so formed ; and

c) separating the CEG polypeptide from the ligand in the complex and identifying the ligand so separated.

90. The method of claim 89, wherein the sample is a tissue or biological fluid.

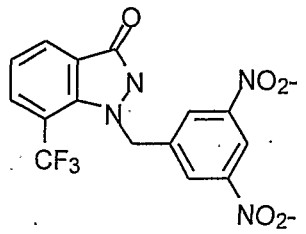
91. The method of claim 89, wherein the ligand is an azabicyclodiene.

92. The method of claim 91, wherein the azabicyclodiene comprises the structure:



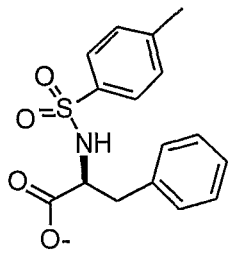
93. The method of claim 89, wherein the ligand is a diazalactone.

94. The method of claim 93, wherein the diazalactone comprises the structure:



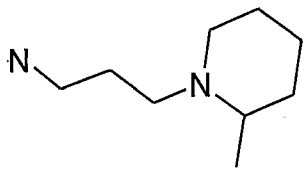
95. The method of claim 89, wherein the ligand is a *N*-protected amino acid.

96. The method of claim 95, wherein the *N*-protected amino acid comprises the structure:

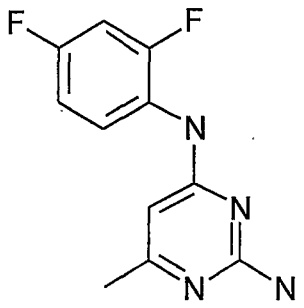


97. The method of claim 89, wherein the ligand is an alkoloid.

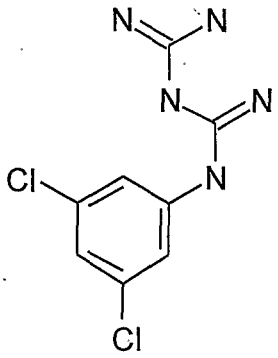
98. The ligand of claim 97, wherein the alkaloid comprises the structure:



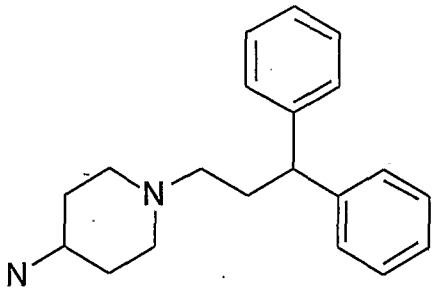
99. The ligand of claim 97, wherein the alkaloid comprises the structure:



100. The ligand of claim 97, wherein the alkaloid comprises the structure:



101. The ligand of claim 97, wherein the alkaloid comprises the structure:



102. The ligand of claim 97, wherein the alkaloid comprises the structure:

