SYSTEMS AND METHODS FOR THE SECRETION OF RECOMBINANT PROTEINS IN GRAM NEGATIVE BACTERIA

Applicants: WISCONSIN ALUMNI RESEARCH FOUNDATION, Madison, WI (US); CORNELL UNIVERSITY, Ithaca, NY (US)

Inventors: Sydor T. Withers, III, Madison, WI (US); Miguel A. Domínguez, Madison, WI (US); Matthew P. DeLisa, Ithaca, NY (US); Charles H. Haitjema, Ithaca, NY (US)

Assignees: WISCONSIN ALUMNI RESEARCH FOUNDATION, Madison, WI (US); CORNELL UNIVERSITY, Ithaca, NY (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

Appl. No.: 15/211,632
Filed: Jul. 15, 2016

Prior Publication Data

Related U.S. Application Data
Division of application No. 13/192,058, filed on Jul. 27, 2011, now Pat. No. 9,410,157.

Provisional application No. 61/369,188, filed on Jul. 30, 2010.

References Cited
U.S. PATENT DOCUMENTS

OTHER PUBLICATIONS

Primary Examiner — Nancy Treptow
Attorney, Agent, or Firm — Casimir Jones, S.C.

ABSTRACT
Disclosed herein are systems and methods for producing recombinant proteins utilizing mutant *E. coli* strains containing expression vectors carrying nucleic acids encoding the proteins, and secretory signal sequences to direct the secretion of the proteins to the culture medium. Host cells transformed with the expression vectors are also provided.

4 Claims, 12 Drawing Sheets
FIG. 3

FlAsH Fluorescence of Secretion Mutants

Fluorescence (mean value ± SD)
FIG. 4

PLASMID MAPS

pTRC99a-YebF-Cel5B (SEQ ID NO: 3)

FEATURES  Location/Qualifiers
source  1..6123
/organism="Cloning vector pTRC99A"
mol_type="genomic DNA"
/db_keyref="takon:40992"
/lab_host="Escherichia coli"
/note="derived from pK233-2"
misc_feature 1..17
/note="derived from cloning vector pBR322"
promoter 18..263
/note="trc promoter from pK233-2"
s/citation=[1]
misc_feature 264..370
/note="HincI/EcoRI linker"
misc_feature 2273..2697
/note="5' RNA, T1, T2, trnE"
misc_feature 2698..4876
/note="derived from cloning vector pBR322"
misc_feature 4877..4884
/note="BglII linker"
misc_feature 4877..4882
/note="BglII linker"
misc_feature 4883..4889
/note="EcoRI linker"
misc_feature 4890..4920
/note="derived from plasmid RP4"
misc_feature 4921..6107
/note="lacI-q region"
misc_feature 6108..6114
/note="EcoRI linker"
misc_feature 6115..6120
/note="BglII linker"
misc_feature 2244..2251
/note="GlyHis" (SEQ ID NO: 6)
gene 391..644
/note="YebF"
gene 651..2243
/note="Cel5B"

ORIGIN
1 gtgtgccttg tatactatgtg cggcacgtgc aechoagta tcttggtggtc ggcggcgtc
gagaaggtgt gatagctctg gcaagctctg aactacagca taattgcgtg cgctccaggc
taacaggtc ttcgtcagtt gtttttcgcc gcacatctt cacggtttgc aatattattc
tttcggatcg tgtctacatt tttctgctgc ctgattaggc tgtgggtggc
tatatcggg gacacagatt gacgagcggt gactgtgctg cagagcaggt gagacgttgt
361 gagagtcgg cagagcccgg aagcaggtgc gacggtgtct gacggtctct cagaggttgt
391 tccgctctcg gcttgcctct aagtacctac cggcttctgt gcgttttcgt tgtctctctg
421 cgtttctctgc tctcattcag cgttcttctc tcttctctctct ctctctctctct ctctctctctct
451 caaattgcct gctgctctct tctgctctct cgcagctctc gtcagctctc gtcagctctc
gatccattggt ctttctctct ctctctctct ctctctctctct ctctctctctct ctctctctctct
541 tctacctctgt cctgatattc ctttctctct ccttctctctct ctctctctctct ctctctctctct
661 ctgcagcttgc gctgctctct tcttctctct ctctctctctct ctctctctctct ctctctctctct
691 gattgatctt gctgctctct tcttctctctct ctctctctctct ctctctctctct ctctctctctct
721 tctctctctct tcttctctctct ctctctctctct ctctctctctct ctctctctctct ctctctctctct
751 gctgctctctct ctctctctctct ctctctctctct ctctctctctct ctctctctctct ctctctctctct
781 tctctctctctct ctctctctctct ctctctctctct ctctctctctct ctctctctctct ctctctctctct
811 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
841 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
871 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
901 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
931 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt
961 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt

FIG. 4, con’t.

1021 AGCGTGATCTA TGAAATCTAC AATGAGGCTTTC TGCGGCGTAA CTTATTCAAAC
1081 GCGATGCGGA AGCGGGGTTT TTGGCGATCC CGCGGACTTC CGGCGATTTG CCGGAGGACT
1141 TGCGGATGCTAA GGCGGCTTAC CAGGGAGTGG CAGGGGACTC ATGCAGGACTT
1201 AGCGAGCACG TGGCGATTTTT CGCGGATATT TCGCGGATATT CCGGAGGACTT
1261 ATAGCGGACGA GACGCGACTTC CGCGGATATT TCGCGGATATT CCGGAGGACTT
1321 TTAACGCTTAT CTTGGCGCTAT GTCGCTGTTT CTTGGCGCTAT GTCGCTGTTT
1381 TGAAGACCAAG TCCACCTCTTC AACGCGAGAT CGCGCGTCGAA ATCGGATAAT
1441 GGCGATGCTGA CGCGGATATT TCGCGGATATT CCGGAGGACTT
1501 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1561 GCGGAGGACTT TCGCGGATATT CCGGAGGACTT
1621 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1681 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1741 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1801 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1861 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1921 CGCGGATATT TCGCGGATATT CCGGAGGACTT
1981 CGCGGATATT TCGCGGATATT CCGGAGGACTT
2041 CGCGGATATT TCGCGGATATT CCGGAGGACTT
2101 CGCGGATATT TCGCGGATATT CCGGAGGACTT
2161 CGCGGATATT TCGCGGATATT CCGGAGGACTT
2221 TGAGCGGCAA CGCGGATATT CCGGAGGACTT
2281 GCGGATATT TCGCGGATATT CCGGAGGACTT
2341 GCGGATATT TCGCGGATATT CCGGAGGACTT
2401 GCGGATATT TCGCGGATATT CCGGAGGACTT
2461 GCGGATATT TCGCGGATATT CCGGAGGACTT
2521 GCGGATATT TCGCGGATATT CCGGAGGACTT
2581 GCGGATATT TCGCGGATATT CCGGAGGACTT
2641 GCGGATATT TCGCGGATATT CCGGAGGACTT
2701 GCGGATATT TCGCGGATATT CCGGAGGACTT
2761 GCGGATATT TCGCGGATATT CCGGAGGACTT
2821 GCGGATATT TCGCGGATATT CCGGAGGACTT
2881 GCGGATATT TCGCGGATATT CCGGAGGACTT
2941 GCGGATATT TCGCGGATATT CCGGAGGACTT
3001 GCGGATATT TCGCGGATATT CCGGAGGACTT
3061 GCGGATATT TCGCGGATATT CCGGAGGACTT
3121 GCGGATATT TCGCGGATATT CCGGAGGACTT
3181 GCGGATATT TCGCGGATATT CCGGAGGACTT
3241 GCGGATATT TCGCGGATATT CCGGAGGACTT
3301 GCGGATATT TCGCGGATATT CCGGAGGACTT
3361 GCGGATATT TCGCGGATATT CCGGAGGACTT
3421 GCGGATATT TCGCGGATATT CCGGAGGACTT
3481 GCGGATATT TCGCGGATATT CCGGAGGACTT
3541 GCGGATATT TCGCGGATATT CCGGAGGACTT
3601 GCGGATATT TCGCGGATATT CCGGAGGACTT
3661 GCGGATATT TCGCGGATATT CCGGAGGACTT
3721 GCGGATATT TCGCGGATATT CCGGAGGACTT
3781 GCGGATATT TCGCGGATATT CCGGAGGACTT
3841 GCGGATATT TCGCGGATATT CCGGAGGACTT
3901 GCGGATATT TCGCGGATATT CCGGAGGACTT
3961 GCGGATATT TCGCGGATATT CCGGAGGACTT
4021 GCGGATATT TCGCGGATATT CCGGAGGACTT
4081 GCGGATATT TCGCGGATATT CCGGAGGACTT
4141 GCGGATATT TCGCGGATATT CCGGAGGACTT
4201 GCGGATATT TCGCGGATATT CCGGAGGACTT
4261 GCGGATATT TCGCGGATATT CCGGAGGACTT
4321 GCGGATATT TCGCGGATATT CCGGAGGACTT
FIG. 4, con’t.

PTRC599p(Cm)-YeKF-Flash-MH1a (SEQ ID NO: 6) (SEQ ID NO: 4)

FEAT URES      | Location/Qualifiers                   
---           | -------------------------------------
promoter     | 153..266                              
             | /label=trc_promoter                   
             | /ApEinfo_fwdcolor="#804000"          
             | /ApEinfo_revcolor="#804000"          
misc_feature | 239..239                               
             | /label=sl3_pUC_rev_primer            
             | /ApEinfo_fwdcolor="#80ff80"          
             | /ApEinfo_revcolor="#80ff80"          
misc_feature | complement(764..781)                   
             | /label=pBAD_rev_primer                
             | /ApEinfo_fwdcolor="#80ff80"          
             | /ApEinfo_revcolor="#80ff80"          
misc_feature | complement(764..781)                   
             | /label=pTrcMH1a_rev_primer            
             | /ApEinfo_fwdcolor="#80ff80"          
             | /ApEinfo_revcolor="#80ff80"          
terminator   | 814..971                              
             | /label=rzMB_terminator                
             | /ApEinfo_fwdcolor="#ff0080"          
             | /ApEinfo_revcolor="#ff0080"          
terminator   | 917..980                              
             | /label=rzMB_Tl_terminator             
             | /ApEinfo_fwdcolor="#ff0080"          
             | /ApEinfo_revcolor="#ff0080"          
terminator   | 1112..1139                            
             | /label=rzMB_T2_terminator             
             | /ApEinfo_fwdcolor="#ff0080"          
             | /ApEinfo_revcolor="#ff0080"          
promoter     | 1181..1269                            
             | /label=Ampl promoter                  
             | /ApEinfo_fwdcolor="#804000"          
             | /ApEinfo_revcolor="#804000"          

CDS          | 1251..1816                            
             | /gene="Ampicillin"                   
             | /note="CRAF frames 3"                 
             | /translation="NGGCHFP/ALIIFPAACLPLVFAHPETLVKVQAGDQLARVGY 
             | IEILDLNHGHIWFPFPDPEERFPHMNSTTFKVILCOAVLRVRDAQEQEGLRGRSHFGMDVE 
             | YQFVTQEH/LIDGMVNRELCSAAITHSDNTANLTTTIGGFFELTAPLRHRNGCHVTPL 
             | DNWPELNA1PNSERDT7HTPTAMATTJEMCLLYLTLNASQQLIQG08MBDFVAGPL 
             | LGVLALACNFA1DFAFSGASER5SCQGI1AALGPDGRKRFIRYVITTGQSADMERHQQA 
             | E1GAUL1SMN*" (SEQ ID NO: 8)           
             | /label=Ampicillin                     
             | /ApEinfo_fwdcolor="#c0c0c0"          
             | /ApEinfo_revcolor="#c0c0c0"          
gene         | 1251..1816                            
             | /gene="Ampicillin"                   
             | /label=Ampicillin(1)                  
             | /ApEinfo_label="Ampicillin"          
             | /ApEinfo_fwdcolor="#ff8040"          
             | /ApEinfo_revcolor="#ff8040"          
rep_origin   | 3679..4298                            
             | /label=pBR322_origin                  
             | /ApEinfo_fwdcolor="#ff8000"          
             | /ApEinfo_revcolor="#ff8000"          
miso_feature | 4699..4717                            
             | /label=pSEX_3_primer                  
             | /ApEinfo_fwdcolor="#80ff80"          


FIG. 4, con't.

901 ctcccccatcgc gaaatgctagg aactgcacagc ctaaaatataa aacgaaaggcc tcaagtccagaa
961 cacctggcccct ttgcttttat ctgtgatttg ctgctgtaagcc ctctccctgcgtg tagqacaaat
1021 ccgccccgagc cggatgtaaac gcctttaacgc caaacggcgcc cgggtgtgcgg gcggaggagcc
1081 cccgctcaaaa tctgggaggc ccaaaatataa aacgaaaggcc tcaagtccagaa
t1201 cgggtttctctt tacaaatataa aacgaaaggcc tcaagtccagaa
t1261 acacaggaacc gcgtgagctaa ctgacaggtgctaatagc gacagagaaatctagtgacagct
1321 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
1381 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
1441 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
1521 gaggttaacctggtctaggtctaatagc gacagagaaatctagtgacagct
1581 agattgttttcagc atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
1641 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
1721 gaggttaacctggtctaggtctaatagc gacagagaaatctagtgacagct
1781 ccacagagtaa gcgtgagctaatagc gacagagaaatctagtgacagct
1841 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
1921 gaggttaacctggtctaggtctaatagc gacagagaaatctagtgacagct
1981 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2041 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
2101 gaggttaacctggtctaggtctaatagc gacagagaaatctagtgacagct
2161 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
2221 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2281 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
2361 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2441 ttagatgtgattc gctaatatctgtatactgtaatagc gacagagaaatctagtgacagct
2521 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2601 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2681 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2761 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2841 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
2921 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3001 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3081 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3161 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3241 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3321 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3401 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3481 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3561 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3641 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3721 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3801 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3881 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
3961 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4041 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4121 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4201 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4281 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4361 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
4441 atctgctttaac gcgtgagctaatagc gacagagaaatctagtgacagct
| 4501 | ccgggtttttc acaacgcaata tgggtgacac tcagttcagat cgattctgtat gcccgcataagt |
| 4561 | taagccacgta tacatcccgct tatcgctacgc tgaactgggct atggtgcggc ccgcaaccc |
| 4621 | gtcacaacccc gctggagcgcg ccggcaaggg cttggctgcgc ccgcaaccc |
| 4681 | agcttggtgac gcctggcgcgg cctggcatctg tcaggagttg cctggctgcgc ccgcaaccc |
| 4741 | cgccgagcgcc agcgtcagtt ggcggggcgg ggcgggggag catgcagattt cgtggcggcc |
| 4801 | atcgaatgttg gcaaacaccc tcggcgattgc gcattgatatc gcgcggaaag gacgtaatcc |
| 4861 | aggtggtggtga atgggacccc atgaagttta tgcgatgctgg cagagtctcg ccggtctctt |
| 4921 | tacagggcgg tttcggcggt ggtgacccag gcccagccacgt tctcgtcgca gacgctgggg |
| 4981 | aaagttgggaag cggcggatgg ccagggtgaaat caccaatccca aacgggtggc ccacaaactg |
| 5041 | cggggcacaac agcttggtgtcg gattggccgt gcagccctcc gctggccccc gcagggcgccg |
| 5101 | tcgaatattc tgtggggtatt taattttcgc ggagtcacaa taggtgcggag cgtgttggagt |
| 5161 | tcaagtgttag aaccgagcgcc gttggaaacgg cggtaaacag tctctcggcg |
| 5221 | cagcgcgtcc ggctggcgttg ccattcacta cctgtggagt gcacgattgc catgcgtctcg |
| 5281 | gaaggtgctgc gcacataattc ttcctgtgatt tttcctgcga gacacccccc |
| 5341 | aacagttattt ttctcttccca tgcgaacgtg tccggactcg gctgggacgc tctggggtca |
| 5401 | tgtgggctacgc aacatctccac tgggttaagcc gcgcaattaa gttctggctgc gcggtcgtctg |
| 5461 | cgtcggggtcgc gttggcataa ccattctgca ccggaggtcag cggagggaaggg |
| 5521 | gaaggcgcatt ggtgggtgcat cttcggttttt caacacccca tgcgaattgc gataagggcc |
| 5581 | tcttgtccca ccgggtcgtcg ggggtggcag ccagcgattg cggctggcgc gtgcgctccg |
| 5641 | attacggagt ctggggtgctgc gctgtgtggcg gattcctcgcc tgggtgggact ccacgatacc |
| 5701 | gaagagacact tggattatat cccggggtca caacactttta aacgagttttt tctgctgcttg |
| 5761 | gcgggggacaag cgggtgacacgt tctgtgctca ttcctcggag gcacgggggt gaagggaact |
| 5821 | cagctgtgcgtc gcctttgtcg gcggagttta atggctgttgg ccagcaaggg ttcgcggactg |
| 5881 | gtcctcttccc ggctgggtgaca ggtgttcata gttggctgag ccaagacggg ccggcagcagc |
| 5941 | gaaagggggc cagtyaggtgca acgcaattta tgtgagcttgt gcggagaattga tcggtcgctgcgtg |
FIG. 4, con't.

pTKC99a-YebF-FlaE-5His (SEQ ID NO: 6) (SEQ ID NO: 5)

FEATURES Location/Qualifiers
source 1..4583
/organism="Cloning vector pTKC99a"
/mc1_type="genomic DNA"
/db_xref="taxon:40592"
/loc_host="Escherichia coli"
(note="derived from pPR233-2"

misc_feature 1..17
/loc="derived from cloning vector pPR233-2"

promoter 18..263
/loc="trc promoter from pPR233-2"
/citation=[1]

misc_feature 264..266
/loc="EcoI/EcoRI linker"

misc_feature 734..1108
/loc="SS RNA, T1, T2, rnb"

misc_feature 1159..3337
/loc="derived from cloning vector pPR233-2"

misc_feature 3338..3345
/loc="BglII linker"

misc_feature 3382..3383
/loc="BglII linker"

misc_feature 3344..3350
/loc="EcoRI linker"

misc_feature 3551..3581
/loc="derived from plasmid KP4"

misc_feature 3582..4568
/loc="lacI-q region"

misc_feature 4569..4575
/loc="EcoRI linker"

misc_feature 4575..4581
/loc="BglII linker"

misc_feature 681..698
/loc="5XHis" (SEQ ID NO: 6)

misc_binding 645..680
/loc="FlaE"

gene 291..644
/loc="YebF"

ORIGIN
1 gtttgacgct ttatcatcga ctggcagcttg caccatgtct tctggcgctca ggcagccgtcgc
61 ggagacagtg tggagctcttg gacagtctga aatcaagtct aatttctgtt cggttcaagggc
121 gcctctgggct tctggtattt gttttttttc gccgtagaat cacgatattt cacaatattcc
181 ttggagctaga tggagcatat taactattcg gttctaatga tggagtatgg tggagttggta
241 taagaatattt taataattttt cagacggttt cagtaattttt cagtaattttt cagtaattttt
301 GGAGCTCGAG TTAAGGATCA TTTTGAGTGC CATGCAGGCT ATGGAGTGAG GCTGCTTGCC
361 ATGACACTCAG CGAAGTCGCGG ACTCTCCTCC AGTGGGTGAG CAGCTGTTAT TGTATCGATGC
421 CCGGTGCAGC AAAAACTCAGT TACACCAAAA ATCCTGGCTGC GCTCTGTTAT GCAGCTGTTAT
481 GAAATGATGCA TGCCGCGCTG CCTGCGGCTG CGTACGCGG CGTACGCGG CGTACGCGG
541 ATGATATAGG CCGCTCTGCG CTTGCTGCGT TGTGTTTCTG TGGGCGTGTTG TGGGCGTGTTG
601 TCACGCTGCGC CAGGAGCGAG CGGAGAGCAG AAGTGGCGC AAGTGGCGC AAGTGGCGC
661 CGGTCCCGTG AGTCGACGAG CAATCGAGCC ATCAGTCCAT ATCGATGAC CAACATGAC
721 cgcgaacagct tcggtgttctt gcgcgagagcaga aagataaattt cacaatatttt cacaatatttt
781 cgcgaacagct tcggtgttctt gcgcgagagcaga aagataaattt cacaatatttt cacaatatttt
841 cgcgaacagct tcggtgttctt gcgcgagagcaga aagataaattt cacaatatttt cacaatatttt
901 cgcgaacagct tcggtgttctt gcgcgagagcaga aagataaattt cacaatatttt cacaatatttt
1021 cgcgaacagct tcggtgttctt gcgcgagagcaga aagataaattt cacaatatttt cacaatatttt
SYSTEMS AND METHODS FOR THE SECRECTION OF RECOMBINANT PROTEINS IN GRAM NEGATIVE BACTERIA

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a divisional of U.S. application Ser. No. 13/192,058, filed Jul. 27, 2011, which claims priority to U.S. Provisional Application Ser. No. 61/369,188, filed Jul. 30, 2010, the entire disclosures of which are hereby incorporated by reference.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with government support under DE-FC02-07ER64494 awarded by the U.S. Department of Energy. The government has certain rights in the invention.

TECHNICAL FIELD

The present disclosure relates generally to systems and methods for producing recombinant proteins by secreting the recombinant proteins to the extracellular growth medium of a gram-negative bacteria.

BACKGROUND

The following discussion of the background is merely provided to aid the reader in understanding the invention and is not admitted to describe or constitute prior art.

Prokaryotes have been widely used for the production of recombinant proteins. Controlled expression of the desired polypeptide or protein is accomplished by coupling the gene encoding the protein through recombinant DNA techniques behind a promoter, the activity of which can be regulated by external factors. This expression construct is carried on a vector, most often a plasmid. Introduction of the plasmid carrying the expression construct into a host bacterium and culturing that organism in the presence of compounds which activate the promoter results in expression of the desired protein. In this way, large quantities of the desired protein can be produced.

E. coli is the most commonly used prokaryote for protein production. A variety of plasmid vectors have been developed for use in E. coli, which employ several different types of promoters, selectable markers, and origins of replication. In the most common arrangement, the expressed protein accumulates in the cytoplasm. While this approach is useful for some proteins, not all proteins can be accumulated in the cytoplasm in an active state. Often, when the desired protein is produced at high levels, it may be toxic to the host cell, or accumulate as an insoluble particle known as an inclusion body. Proteins which accumulate as inclusion bodies are difficult to recover in an active form. In such cases, it may be desirable to engineer the protein so that it is secreted from the cell.

E. coli and other gram-negative bacteria are generally considered poor hosts for secretory protein production. There are no well-understood secretory pathways in E. coli to transport heterologous proteins to the extracellular environment. The recent discovery of YebF-mediated secretion (Nat Biotechnol. 2006. 24(1):100-4) is the first report of a native E. coli system capable of secreting both the native protein, YebF, and translational fusions to YebF. However, the expression level of YebF fusion proteins is typically low.

SUMMARY

The present disclosure is based on the discovery of E. coli mutations that substantially increase the amount of recombinant protein secreted from cells compared to wild-type E. coli.

In one aspect, the present disclosure provides a recombinant bacterium comprising a mutant bacterium that has been transformed with a recombinant vector comprising a first DNA sequence encoding a signal peptide or secretory protein operatively linked to a second DNA sequence encoding a heterologous protein, wherein the mutant bacterium comprises mutations in at least one gene selected from the group consisting of: ompR, envZ, nlpD, emC, emE, yebE, yihF, yebG, mzzA, ftsK, tuaA, ompC, and ompF or homologs thereof.

In one embodiment, the bacterium is a gram negative bacterium. In one embodiment, the bacterium is selected from the group consisting of Escherichia, Salmonella, Yersinia, and Shigella. In one embodiment, both the NlpD and EnvZ gene products are not expressed or are rendered non-functional. In one embodiment, both the NlpD and OmpR gene products are not expressed or are rendered non-functional. In one embodiment, the NlpD and YihF gene products are not expressed or are rendered non-functional. In one embodiment, the secretory protein is YebF.

In one aspect, the present disclosure provides an expression system for secreting a recombinant protein into a culture medium, the system comprising: (a) a mutant E. coli bacterium, wherein at least one gene product selected from the group consisting of OmpR, EnvZ, NlpD, EmC, EmE, YebE, YihF, YebG, MzzA, FtsK, TuaA, OmpC, and OmpF is not expressed or is rendered non-functional; and (b) a recombinant vector comprising a first DNA sequence encoding a signal peptide or secretory protein operatively linked to a second DNA sequence encoding a heterologous protein.

In one embodiment, both the NlpD and EnvZ gene products are not expressed or are rendered non-functional. In one embodiment, both the NlpD and OmpR gene products are not expressed or are rendered non-functional. In one embodiment, both the NlpD and OmpR gene products are not expressed or are rendered non-functional. In one embodiment, at least one gene product is not expressed or is rendered non-functional by deleting all or part of the gene encoding the gene product. In one embodiment, the at least one gene product is not expressed or is rendered non-functional by way of alteration of a promoter control sequence. In one embodiment, the promoter control sequence is altered by incorporation of an inducible promoter sequence element. In one embodiment, the promoter control sequence is altered by incorporation of a repressor promoter sequence element. In one embodiment, the promoter control sequence is altered so as to provide a non-functional promoter control sequence.

In one embodiment, the secretory protein is YebF. In one embodiment, the signal peptide is capable of mediating transport of a protein to the periplasmic space. In one embodiment, the signal peptide is associated with the SEC, TAT, or SRP export pathway.

In one embodiment, the heterologous protein that is secreted is biologically active. In one embodiment, the heterologous protein is selected from the group consisting of a cellulase, a protease, a lipase, a cutinase, an amylase, a galactosidase, pullulanase, a gluco isomerase, a protein disulfide isomerase, a cycloledin glucosetransferase, a phytase, a glucose oxidase, a glucoyl transferase, laccase, bilirubin oxidase, a xylanase, an antigenic microbial or
protozoan protein, a bacterial protein toxin, a viral protein, and a pharmaceutical. In one embodiment, the heterologous protein is selected from the group consisting of an immunoglobulin light chain, an immunoglobulin heavy chain, an immunoglobulin light chain fragment or an immunoglobulin heavy chain fragment.

In one embodiment, the expression of both DNA sequences is under the control of an inducible promoter. In one embodiment, the inducible promoter is a lac promoter. In one embodiment, the at least one gene product selected from the group consisting of OmpR, EnvZ, NlpD, EntC, EntE, YebE, YihF, YebG, MzaA, FtsK, TnaA, OmpC, and OmpF is not expressed or is rendered non-functional by substitution, deletion, or insertion of one or more nucleotides in the gene encoding the at least one gene product.

In another aspect, the present disclosure provides a method for producing a recombinant protein comprising: (a) culturing an E. coli bacterium under conditions in which the bacterium secretes a heterologous protein into a culture medium, wherein the E. coli bacterium comprises: (i) a mutant E. coli bacterium, wherein at least one gene product selected from the group consisting of OmpR, EnvZ, NlpD, EntC, EntE, YebE, YebG, YihF, YebG, MzaA, FtsK, TnaA, OmpC, and OmpF is not expressed or is rendered non-functional; and (ii) a recombinant vector comprising a first DNA sequence encoding a signal peptide or carrier protein operatively linked to a second DNA sequence encoding a heterologous protein, and (b) isolating the secreted protein from the culture medium. In one embodiment, the method further comprises the step of purifying the secreted protein. In another aspect, the present disclosure provides a method for producing a heterologous protein comprising: (a) transforming a host cell with a recombinant vector, wherein the host cell is a mutant E. coli bacterium, wherein at least one gene product selected from the group consisting of OmpR, EnvZ, NlpD, EntC, EntE, YebE, YihF, YebG, MzaA, FtsK, TnaA, OmpC, and OmpF is not expressed or is rendered non-functional, and wherein the recombinant vector comprises a first DNA sequence encoding a signal peptide or carrier protein operatively linked to a second DNA sequence encoding a heterologous protein; (b) culturing the host cell under conditions in which the bacterium secretes the heterologous protein into the culture medium; and (c) isolating the secreted protein from the culture medium.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 is a graph showing improved protein secretion in mutant strains. The K602 ho strain indicates what gene has been deleted from strain BW25113 AdsB-A. WT indicates the wild-type background (E. coli BW25113 AdsB-A). The upper graph shows the relative fluorescence from FlAsH-tagged YebF. Below are the results of Western blots of secreted YebF-6xHis-cellulase (6xHis disclosed as SEQ ID NO: 6) fusion proteins.

FIG. 2 is a graph showing the relative fluorescence from FlAsH-tagged YebF in E. coli having single- and double-mutations in YebF-related genes.

FIG. 3 is a graph showing the relative fluorescence from FlAsH-tagged YebF in E. coli having single- and multiple-mutations in YebF-related genes.

FIG. 4 shows the plasmid maps of the plasmids described in the examples.

DETAILED DESCRIPTION

The present disclosure relates inter alia to a recombinant bacterium that has been mutated in one or more genes that affect a YebF-mediated protein secretory pathway. The mutants exhibit increased secretion of YebF fusion proteins compared to wild-type E. coli. The mutants include bacteria containing mutations in at least one gene selected from the group consisting of: ompr, envZ, nlpD, entC, entE, YebF, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF or homologs thereof.


As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the content clearly dictates otherwise. For example, reference to “a cell” includes a combination of two or more cells, and the like.

As used herein, the term “expression vector” refers to a recombinant DNA molecule containing the appropriate control nucleotide sequences (e.g., promoters, enhancers, repressors, operator sequences and ribosome binding sites) necessary for the expression of an operably linked nucleotide sequence in a particular host cell. By “operably linked/linking” or “in operable combination” is meant that the nucleotide sequence is positioned relative to the control nucleotide sequences to initiate, regulate or otherwise direct transcription and/or the synthesis of the desired protein molecule. The expression vector may be self-replicating, such as a plasmid, and may therefore carry a replication site, or it may be a vector that integrates into a host chromosome either randomly or at a targeted site. The expression vector may contain a gene as a selectable marker for providing phenotypic selection in transformed cells. The expression vector may also contain sequences that are useful for the control of translation.

As used herein, a “fusion” protein is a recombinant protein comprising regions derived from at least two different proteins. The term “fusion protein” as used herein refers to a protein molecule in which a heterologous protein of interest is fused to secretory protein or a signal peptide, such as YebF. “Fused”, in one context means that nucleic acid encoding the secretory protein or signal peptide is joined in frame. In another context, “fused” may also be a reference to the joining of a recombinant protein of interest to the secretory protein or signal peptide, such as YebF.

As used herein, “heterologous” refers to DNA, RNA, or protein that does not occur naturally as part of the organism in which it is present or which is found in a location or locations in the genome that differ from that in which it occurs in nature. It is DNA, RNA, or protein that is not endogenous to the cell and has been artificially introduced into the cell. Examples of heterologous DNA include, but are not limited to, DNA that encodes a cellulase. The
heterologous DNA need not be expressed and may be introduced in a manner such that it is integrated into the host cell genome or is maintained episomally.

As used herein, the term “homolog” refers to any gene that is related to a reference gene by descent from a common ancestral DNA sequence. The term “ortholog” refers to homologs in different species that evolved from a common ancestral gene by speciation. Typically, orthologs retain the same or similar function despite differences in their primary structure (mutations). The term “paralog” refers to homologs in the same species that evolved by genetic duplication of a common ancestral gene. In many cases, paralogs exhibit related (but not always identical functions). As used herein, the term homolog encompasses both orthologs and paralogs. To the extent that a particular species has evolved multiple related genes from an ancestral DNA sequence shared with another species, the term ortholog can encompass the term paralog.

As used herein, the terms “identical” or percent “identity,” when used in the context of two or more nucleic acids or polypeptide sequences, refers to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., at least 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI web site). Such sequences are then said to be “substantially identical.” This term also refers to, or can be applied to, the complement of a test sequence. The term also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Suitably, identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

As used herein, the term “mutant” of a gene refers to a gene which has been altered, either naturally or artificially, changing the base sequence of the gene. The change in the base sequence may be of several different types, including changes of one or more bases for different bases, deletions, and/or insertions, such as by a transposon. By contrast, a normal form of a gene (wild type) is a form commonly found in natural populations of an organism. Commonly a single form of a gene will predominate in natural populations. In some embodiments, a mutant gene will be altered such that the product of that gene is not expressed, expressed at reduced or increased levels compared to wild type, or is rendered non-functional.

As used herein, “periplasm” refers to a gel-like region between the outer surface of the cytoplasmic membrane and the inner surface of the lipopolysaccharide layer of gram-negative bacteria.

As used herein, the term “polynucleotide” or “nucleic acid” means any RNA or DNA, which may be unmodified or modified RNA or DNA. Polynucleotides include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, RNA that is mixture of single- and double-stranded regions, and hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons.

As used herein, the terms “polypeptide”, “peptide” and “protein” are used interchangeably herein to mean a polymer comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. Polypeptide refers to both short chains, commonly referred to as peptides, glycopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. Polypeptides include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art.

As used herein, a “promoter” or “promoter region” refers to a portion of DNA that controls transcription of the DNA to which it is operatively linked. The promoter region includes specific sequences of DNA that are sufficient for mRNA polymerase recognition, binding and transcription initiation. This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of the RNA polymerase. These sequences may be cis acting or may be responsive to trans acting factors. Promoters, depending upon the nature of the regulation, may be constitutive or regulated.

As used herein, the term “recombinant” when used with reference, e.g., to a cell, or a nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the material is derived from a cell so modified. Thus, e.g., recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

As used herein, “secretion” refers to the excretion of the recombinant protein that is expressed in a bacterium to the periplasm or extracellular medium.

As used herein, “YebF” refers to an extracellular protein of E. coli with no known function having the amino acid sequence of SEQ ID NO:1 or biologically-active variants thereof. “yebF” is a reference to a nucleic acid or nucleotide sequence encoding SEQ ID NO: 1 or biologically-active variants thereof. In one embodiment, yebF has the sequence of SEQ ID NO:2.

Bacterial Strains and Mutants

Disclosed herein are modified bacteria useful for the production of secreted proteins. Modified bacteria may include bacteria with an improved (increased) ability to secrete proteins into the culture media, as compared to the similar, but non-modified (non-mutated) bacteria. An increase in the ability to secrete proteins includes, in various embodiments, about a 5%, 10%, 20%, 50%, 75%, 90%, 100%, 125%, or more increase in the amount of protein secreted into the medium compared to a similar, but non-modified (non-mutated) bacteria.

In one aspect, the present disclosure relates to genetically-modified E. coli bacteria containing a mutation in at least one gene which inhibits the YebF secretion pathway. In some embodiments, the mutation is in one or more genes selected from ompR, envZ, nlpD, enC, enE, yebE, yhfF, yebG, mrrA, ftsK, tuaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains a single mutation in the ompR, envZ, nlpD, enC, enE, yebE, yhfF, yebG, mrrA, ftsK, tuaA, ompC, or ompF gene. In one
embodiment, the genetically modified bacterium contains a single mutation in the nlpD gene. In one embodiment, the genetically modified bacterium is a double mutant containing mutations in two genes selected from ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium is a double mutant containing mutations in the nlpD and ompr genes. In one embodiment, the genetically modified bacterium is a double mutant containing mutations in the nlpD and envZ genes. In one embodiment, the genetically modified bacterium is a triple mutant containing mutations in three genes selected from ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in four genes selected from ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in five genes selected from ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in the ompR, nlpD, enC, enE, yebE, and yihF genes.

In one embodiment, the host cell is a genetically-modified Shigella, Yersinia, Salmonella or Escherichia sp. bacterium containing a mutation in at least one gene which inhibits the extracellular secretory pathway.

Various E. coli strains may be mutated to contain a mutation in one or more genes selected from ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. Wild-type E. coli strains may be any E. coli strains that are found in natural populations. Examples include the E. coli strain BW25113, HB101, UI5374, BLR, TOP10, W3110 (ATCC Accession No. 27325) and the MG1655 (ATCC Accession No. 47076), 294 (ATCC Accession No. 31,446), E. coli B (ATCC Accession No. 11303), X1776 (ATCC Accession No. 31,537), E. coli W (ATCC Accession No. 9637), DHI (ATCC Accession No. 33,849) and KO11 (ATCC Accession No. 55,124).

The E. coli mutant strain can be obtained by any method. In one embodiment, a gene or DNA on the E. coli chromosomal DNA is deleted. For example, a gene can be deleted using homologous recombination in a strain expressing the lambda red recombinease system. In E. coli, homologous recombination usually requires a helper such as the lambda red system developed by Datsenko and Wanner. Proc Natl Acad Sci U S A. 2000 Jun. 6; 97(12):6640-5. Homologous recombination involves the use of DNA fragments located at both outer sides of the gene that is intended to be deleted. An example of a DNA that can be used for homologous recombination includes, but is not limited to, a linear DNA comprising, at both ends of a selectable marker gene, DNA that is homologous to chromosomal DNA into which the introduction of deletion, substitution or addition of nucleotide(s) is desired.

DNA that exists at both ends of the linear DNA is oriented on the linear DNA in the same direction as the chromosomal DNA. The length of the homologous region is suitably about 10 bp to 100 bp, about 20 bp to 50 bp, or about 30 bp to 40 bp. The homologous region will typically be 80% or more, suitably 90% or more, more suitably 100% homology. Homology of the nucleotide sequences can be determined using programs such as BLAST or FASTA. The DNA fragments can be prepared by PCR based on the published sequences of the target gene(s), e.g., ompR, envZ, nlpD, enC, enE, yebE, yihF, yebG, mzaA, ftsK, tnaA, ompC, and ompF. Genomic DNA from the desired host strain can be used as a template for the PCR.

After the DNA for homologous recombination is introduced into a host cell by a conventional method, such as electroporation, transformants are selected using the selectable marker, e.g., antibiotic resistance, as an indicator. The transformants are cultured in a medium that does not contain the antibiotic for several hours to 1 day, and then the cultures are plated on a medium that contains the antibiotic. By determining the nucleotide sequence of a region of the chromosomal DNA in which the gene or DNA to be deleted was present, the deletion of the target gene or DNA on chromosomal DNA can be confirmed.

Any selectable marker gene can be used, provided that such genes impart resistance to an agent to which E. coli shows sensitivity. For example, kanamycin-resistant genes, chloramphenicol-resistant genes, gentamicin-resistant genes, spectinomycin-resistant genes, tetracycline-resistant genes, or ampicillin-resistant genes can be used as the selectable marker genes.

E. coli mutant strains can also be obtained using phase transduction of DNA from a donor strain to a recipient strain. In this case the donor strain mutation has typically been previously characterized and confers at least one selectable phenotype.

Expression Vectors for Secretion of Recombinant Proteins

The secreted recombinant proteins invention can be produced through the application of recombinant DNA technology. Recombinant constructs encoding a protein of interest typically include an expression control sequence operably-linked to the coding sequences of the protein of interest. A “recombinant protein of interest” refers to a protein, the production of which may be deemed desirable for any reason. Such proteins may include enzymes, antibodies, etc., or portions thereof. The protein may be of interest for commercial and/or therapeutic purposes. A nucleotide sequence “encodes” or “codes for” a protein if the nucleotide sequence can be translated into the amino acid sequence of the protein. The nucleotide sequence may or may not contain an actual translation start codon or termination codon.

For expression of the recombinant protein of interest, the nucleic acid containing all or a portion of the nucleotide sequence encoding the protein of interest is inserted into an appropriate cloning vector, or an expression vector (i.e., a vector that contains the necessary elements for the transcription and translation of the inserted nucleotide sequence coding sequence) by recombinant DNA techniques well known in the art and as detailed below. Methods for producing diverse populations of vectors have been described by Lerner et al., U.S. Pat. No. 6,291,160; 6,680,192. Vectors can also encode secretory protein or signal peptide, e.g., YebE, SEC, TAT, peptate lyase, etc., which are useful to direct the secretion of the peptide of interest to the periplasm or extracellular medium.

In general, expression vectors useful in recombinant DNA techniques are often in the form of plasmids. In the present specification, “plasmid” and “vector” can be used interchangeably as the plasmid is the most commonly used form of vector. However, the technology is intended to include such other forms of expression vectors that are not technically plasmids, which serve equivalent functions.

The recombinant expression vectors include a nucleic acid encoding a protein of interest in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operatively-linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence. The term “regulatory sequence” is intended to include promoters, enhancers and
other expression control elements. Such regulatory sequences are described, e.g., in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only under certain conditions, i.e. inducible promoters. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, etc. The expression vectors of the invention can be introduced into host cells thereby produce polypeptides or peptides, including fusion polypeptides, encoded by nucleic acids as described herein. One such example is the expression of heterologous proteins through chromosomal insertion.

Expression of polypeptides in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors serve four purposes: (i) to direct secretion of the polypeptide from the cell; (ii) to increase expression of recombinant polypeptide; (iii) to increase the solubility of the recombinant polypeptide; and (iv) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. In some embodiments, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

In some embodiments, the expression vectors can encode a secretory sequence or signal peptide, e.g., YebF, SEC, TAL, etc. as described above, which are useful to direct the secretion of the peptide of interest. In one embodiment, the secretory sequence is YebF. For example, the recombinant protein of interest may be constructed as a C-terminal fusion to YebF. In one embodiment, YebF has the sequence according to SEQ ID NO: 1 below:

```
MKRAGFLGLLYVSACAVRNNESKEVTHFEKQGLDAAGIAAS
VREDYQQRVARNRQDIQYQ6APVMNSQDIQKEDDNSHPSVPLT
VRKGRADHNYQYVSQKQMAAGAEYQRR
```

In one embodiment, YebF is encoded by the sequence according to SEQ ID NO: 2 below:

```
ATGAAAAAAAGGGGCGCTTTTGAGCTCTTGTTGTTCTGCTCTCCT
GCAGCTATTGGCTGGAATAAGACCTAGACCTGCTCAC
TTCCACACGCAATGCTGCTCGAGATGCTGCAGGACCGAGAC
GTTAAAATCGTAATTACAAAATCTGGCGCTTTGGCGACAGTG
```

In some embodiments, signal peptides may be used to export proteins to the periplasm between the inner and outer membranes. By placing a signal sequence in front of the coding sequence of the desired protein, the expressed protein can be directed to a particular export pathway (U.S. Pat. No. 5,047,334, U.S. Pat. No. 4,953,495.). Known export pathways in E. coli include the SecB-dependent (SEC), the twin-arginine translocation (TAT), and the signal recognition particle (SRP) pathway. Translocation in the SEC or TAT pathway is via a post-translational mechanism, whereas the SRP pathway translocation is co-translational. Proteins translocated by the SEC pathway are unfolded prior to export and then refolded in the periplasm. In the TAT pathway, the proteins are translocated in a folded state. Examples of other signal sequences that could be used to secrete proteins in E. coli include, but are not limited to, Pectate lyase B (PleB) from Erwinia carotovora; Outer membrane protein A (OmpA); Heat-stable enterotoxin 2 (StII); Endoynylase (Endo) from Bacillus sp.; Alkaline phosphatase (PhoA); Outer-membrane pore protein F (OmpF); Outer-membrane pore protein E (PhoE); MalT-binding protein (MalE); Outer-membrane protein C (OmpC); Murein lipoprotein (Lpp); Lambda receptor protein (LamB); Protease VII (OmpT); and Heat-labile enterotoxin subunit B (LTB).

One strategy to maximize recombinant polypeptide expression in E. coli is to express the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide. See, e.g., Gettesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the expression host, e.g., E. coli (see, e.g., Wada, et al., 1992. Nucl. Acids Res. 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

Expression and Secretion of Recombinant Proteins

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

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms “transformation” and “transfection” are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation, biolistics or viral-based transfection can be used for other cellular hosts.
Other methods used to transform mammalian cells include the use of polybrene, protoplast fusion, liposomes, electroporation, and microinjection (see generally, Sambrook et al., Molecular Cloning, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al., and other laboratory manuals. Host cells carrying the expression vector are identified through the use of the selectable marker, and the presence of the gene of interest is confirmed by hybridization, PCR, antibodies, or other techniques.

A mutant host cell that includes an expression vector, such as a prokaryotic host cell in culture, can be used to produce (i.e., express) the recombinant protein of interest. In one embodiment, the method comprises culturing the mutant host cell of invention (into which a recombinant expression vector encoding the protein of interest has been introduced) in a suitable medium such that the protein of interest is produced. In another embodiment, the method further comprises the step of isolating the protein of interest from the medium or the host cell. Once expressed, collections of the protein of interest are purified from culture media and host cells. The protein of interest can be purified according to standard procedures of the art, including HPLC purification, column chromatography, gel electrophoresis and the like. Usually, the protein of interest is expressed with signal sequences and are thus released to the culture media.

The host cells are grown in growth medium until such time as is desired to harvest the secreted protein. The time required depends upon a number of factors relating to the bacterial expression system being used and to the protein produced. The rate of growth of a particular bacterial strain or species; the rate at which the secreted target protein accumulates in the periplasm or extracellular medium; the stability of the secreted protein; and the time at which bacterial lysis begins to occur (which will contaminate the medium) are examples of the types of considerations that will affect when the secreted protein is harvested from the periplasm or extracellular medium.

In the case of intracellular production, the cells are harvested and the protein, polypeptide or peptide is released from the periplasm into the extracellular medium by inducing outer membrane leakage or rupturing the cells using mechanical forces, ultrasound, enzymes, chemicals and/or high pressure. Following secretion into the medium (for example, via YebF'), the protein, polypeptide or peptide may be extracted from the medium. Depending upon the level of purity required, which will again depend upon the application for which the secreted recombinant protein, polypeptide or peptide will be used, the secreted protein may be further purified, for example by chromatography (e.g., affinity chromatography), precipitation, ultrafiltration, electrophoresis, or other suitable techniques.

Purification of recombinant polypeptides is well known in the art and include ammonium sulfate precipitation, affinity chromatography purification technique, column chromatography, ion exchange purification technique, gel electrophoresis and the like (see generally, Scopes, Protein Purification (Springer-Verlag, N.Y., 1982)).

Uses

In one aspect, the bacteria described herein may be useful for manufacturing a variety of proteins. In some embodiments, the bacteria are engineered to produce proteins needed for bioenergy production, therapeutic biologics, and research tools. The present technology provides significant advantages over current techniques. Because the proteins are exported, there is a significantly lower level of contamination, endotoxin, host cell proteins and nucleic acids, making purification easier and thus lowering production costs and durations. Importantly, the invention enables the production of proteins which might otherwise not be expressed due to toxicity and folding errors. The technology may be used for rapid production of proteins at a commercial scale, adapted to high throughput protein production, or readily employed in automated systems.

In one embodiment, the mutant host strains and expression systems are used in the manufacture of cellulolytic biofuels. Cellulosic biofuels are produced using secreted enzyme complexes including cellulases and xylanases. The cellulolytic substrates cannot be imported into the cell. Therefore, the enzyme must be secreted. Providing a microorganism that could supply secreted enzyme complexes would greatly enhance biofuel production.

EXAMPLES

The present invention is further illustrated by the following examples, which should not be construed as limiting in any way.

Example 1—Identification of Mutants Affecting YebF-Mediated Secretion

We identified six *E. coli* genes whereby the deletion of each gene results in improved YebF-mediated secretion: *ompR*, *envZ*, *nlpD*, *entC*, *yebE* and *yihF*. Mutants in each of these genes were identified and tested as described in this Example.

Strains. *E. coli* K-12 BW25113 is the parental strain in the Keio collection of knockouts from which all strain construction was performed. The initial host strain is the Keio *dsbA* knockout with the kanamycin resistance cassette removed. All subsequent deletions (i.e., *entC*, *envZ*, *nlpD*, *ompR*, *yebF*, and *yihF*) and deletion combinations were transduced into this strain. Removal of the kanamycin resistance cassette was performed between each transduction utilizing the FLP recombinase described by Datsenko and Wanner (Proc Natl Acad Sci U S A. 2000 Jun 6; 97(12):6640-5). In addition, each of these knockout strains was picked from the Keio collection to create the phage lysate for transduction.

Plasmids. Three plasmids were used in these Examples and are all contained in the pTRC99a vector backbone. The YebF' sequence was modified to include a 6xHis tag (SEQ ID NO: 6) and a FIAsh tag (CCPGCC-SEQ ID NO: 7) on the protein carboxy terminus. All plasmid maps are shown in the attached sequence listing.

A brief summary of the workflow for the experiment was as follows:

1. Generated lysate of knockout deletion;
2. Transduced deletion into recipient strain;
3. Removed antibiotic resistance marker;
4. Transformed strain with expression construct (e.g., pTRC99a-YebF-FIAsh-His, pTRC99a-(Cm)-YebF-FIAsh-His, or pTRC99a-YebF-Cel5B);
5. Induced expression with 0.1 mM IPTG;
6. Assayed protein secretion by FIAsh fluorescence or western blot of His tag. The FIAsh tag reacts with the FIAsh-EDT reagent (Invitrogen) to produce a fluorescent product. The actual fluorescence assay generated during the screening solicited the use of a construct using an ampicillin drug marker and the subsequent verification of the single and multiple deletion containing strains utilized a chloramphenicol resistance marker.
The western blot utilized a separate plasmid containing the YebF fused with a cellulase gene (i.e. Cel51H)
Table 1 and Fig. 1 shows the result of FlAsH fluorescence for each deletion on YebF-mediated secretion. The strains identified show consistently higher secretion of both tagged YebF as well as YebF-cellulase fusions.

<table>
<thead>
<tr>
<th>Screening Score</th>
<th>Screening Loops</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4 ++ envZ/ompR</td>
<td>2-component osmolality regulator</td>
<td></td>
</tr>
<tr>
<td>12.2 ++ nlpD</td>
<td>Novel lipoprotein, function unknown</td>
<td></td>
</tr>
<tr>
<td>8.9 + muzA</td>
<td>Modulator of EnvZ/OmpR operon</td>
<td></td>
</tr>
<tr>
<td>10.8 ++ fltK</td>
<td>DNA translocase at septal ring sorting daughter chromosome</td>
<td></td>
</tr>
<tr>
<td>6.2 + tusA</td>
<td>Tryptophanase</td>
<td></td>
</tr>
<tr>
<td>27.4 + entC/E</td>
<td>Isochorismate synth 1 &amp; comp of heterosuberin synth cpx</td>
<td></td>
</tr>
<tr>
<td>6.6 0 yihF</td>
<td>Conserved protein, DUP545 family</td>
<td></td>
</tr>
<tr>
<td>N/A NA yebE</td>
<td>Inner membrane protein</td>
<td></td>
</tr>
</tbody>
</table>

Example 2—Comparison of Secretion in Single- and Multiple-Mutant E. coli Strains

96 deep-well plates were inoculated with all transformed secretion strains. A single colony from transformed plate was picked into 1.5 ml LB/Cm35. Plates were incubated at 30°C while shaking in humidified shaker for 18-24 hours. The overnight cultures were subcultured at a 1:40 ratio into 1.5 mL media [LB/Cm35 (negative control) or LB/Cm35+0.1 mM IPTG]. Plates incubated overnight at 30°C while shaking in humidified shaker for ~17-20 hrs. 200 µL of induced culture was assayed for secreted YebF protein by the addition of 10 µL of FlAsH/DTT/BAL cocktail (21 µM FlAsH-EDT, 21 mM DTT, and 5.25 mM 2,3-dimercaptopropanol) for a final concentration of 1 µM FlAsH-EDT, 1 mM DTT, and 250 µM 2,3-dimercaptopropanol. Plate incubated in a spectrophotometer for 20 minutes while measuring the optical density at 600 nm and fluorescence (Ex 508 nm/Em 528 nm) every minute. The data shown in Fig. 2 and Fig. 3 represent the fluorescence measurements after 20 minutes.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All nucleotide sequences provided herein are presented in 5' to 3' direction.

The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including," containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification, improvement and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications, improvements and variations are considered to be within the scope of this invention. The materials, methods, and examples provided here are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

All publications, patent applications, patents, and other references mentioned herein are expressly incorporated by reference in their entirety, to the same extent as if each were incorporated by reference individually. In case of conflict, the present specification, including definitions, will control.
Arg Asp Tyr Gln Gln Aen Arg Val Ala Arg Trp Ala Asp Asp Gln Lys
50
55
Ile Val Gly Gln Ala Asp Pro Val Ala Trp Val Ser Leu Gln Asp Ile
60
65
70
75
80
Gln Gly Lys Asp Asp Lys Trp Ser Val Pro Leu Thr Val Arg Gly Lys
85
90
95
Ser Ala Asp Ile His Tyr Gln Val Ser Val Asp Cys Leu Ala Gly Met
100
105
110
Ala Glu Tyr Gln Arg Arg
115

&lt;210&gt; SEQ ID NO 2
&lt;211&gt; LENGTH: 357
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: Artificial Sequence
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 2
atgaaaaaaa gaggggtgt tttagggctg tttgtagttt ctgtgcgtgca atcagtttttc 60
gctgcaatca atgaacaccag caagtcgctc acttttttttaa aagtgtgaaga tcttgatgct 120
ggcggagtt cgcggagcgt aacacggtgat tatcaacaaaa atcgggtgacg gcgtttggca 180
gagcgtcaca aatagtgctg ttaagggggac cccggtgcgtg agggttcgctt gcaggtataa 240
cagggggttt cttggtgttt ctggtgtgtg cttctgcag caggttggtgaa tctggctgaa 300
cattacagct tccccgggttt cttgacaagc ggtgagagac cccggtttttttt 357

&lt;210&gt; SEQ ID NO 3
&lt;211&gt; LENGTH: 6120
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: Artificial Sequence
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 3
gttgacagc ttctatcctg ctcgaaggcg gcgaagcggt gcgttaaatgt 60
ggacagtgc ggacagcgt gcagcgtgaa acactagcgc taattcgtgt ccagcctggc 120
ggctcgttgt gtttgcctgtg ctcgcgcgtgatatccaccc ccagcgtggtgacacagcagcgc 180
tgacgcagc gcgtcgcgct ggtgagctgc gattgttatgc gcgtcgccga 240
taccacggta acagctgggtg aggagtggc ggcggcagcgc gggttgcgtgagcgcacagc 300
ggacagccc gcagcgcgct ctcggtgcgt gcattttttt gcagctgaaactgc 360
ccggtgatgc cgctggaagt ggcgcgtgaccc gcagtctttt gctggttgcgcgcgg 420
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 480
atcttcgct cttggtgtgtc gcagtcgctg gttgagctgc cttggaggtggt gcgttgcggtg 540
atcttcgct cttggtgtgtc gcagtcgctg gttgagctgc cttggaggtggt gcgttgcggtg 600
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 660
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 720
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 780
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 840
tcagtgccca gcgtgagttc gcacggtgtg gcagcgcctt gcaagcgtggtgagcgc 900
cagcgatcgc aacgacagtg tachggtaaa cagctggtgcc tagaccaaat gcacaccaat 960
atcgatotca ggcgatcgcgg ttccttgcaagag acaatgtcctg cgaagctctg gcgaaacacc 1020
aagtgtcata ctgagacacta aaggtgcctg tagcaggtgag cggcttctac actaccaacc 1080
cgctgtgcaac cggctgttgc ggcggtatcctggcgtatga ccagcaaatgctcttattc 1140
tggtgtagccc gaaccttgagt cagatctccg gatgaccccg attaaggtct 1200
accgagacatt gcctgatatcct tctcattcctg acgctttgctat gcacgcttcaac actacgctctg 1260
ataagcgcaat cgcgcaaccttg aatcgtgcgggt ttgctctggct tctcaccggaa tggggtctgggt 1320
ctatgctgcaat tccaggtgctatcagctgctcta atacgcaacat ccgctgcttg tctccgcttta 1380
tgacatcacc tccacatcctc aacgcaacactt gggcagcattg aacaaagttc aacggcgaac 1440
gcagatggtttt ccggttgctgctt ggcgaaacag ccgtggtgcttt aacggtctgctac ggtgacgact 1500
cgcagcctgc aacgcaacagcc cggcagcgcag ctcagtggtta cagctggtttcg ccgctgctgcttcgct 1560
atcaggtgctg ccgcggtgcctg ccgctgctgcttcgct 1620
gatcggctgg gtcggtatgcc gatggtctgatt gacgacgctgcg tgaactgtgt ctcagcgtgctggc 1680
acgcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag 1740
cgcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag 1800
cgcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag 1860
gctccagccag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 1920
cgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 1980
atcaggtgctg ccgcggtgcctg ccgctgctgcttcgct 2040
gatcggctgg gtcggtatgcc gatggtctgatt gacgacgctgcg tgaactgtgt ctcagcgtgctggc 2100
cgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2160
acgcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag cggcagcgcag 2220
cgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2280
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2340
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2400
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2460
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2520
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2580
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2640
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2700
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2760
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2820
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2880
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 2940
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3000
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3060
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3120
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3180
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3240
acgcagcgcag ctcagccgctg tctccgctgcttcgctgctt ctcagcgtgctggctc 3300
...continued

taccacacga cgacgctgac accacagatgc ctacagcaat ggcaaccaacg ttgcgcaac 3360
tattaactgg cgacacttcct actctagctt cccgaccaaca ataatagac tggatgagg 3420
cggataaatg tgcagaccaacc tttctgctct ggcctgtcct tcggctgtct gtttatgtcgc 3480
ataatatcgg agccgctgag ctgctggcttc gcggtactac tgacacaacct gggccagatg 3540
gtaacgcct cccgtatgcg gctatactca cgcagggagcg tcaaccacact atggatgaaac 3600
gaaaatacag ctacgctgag atagggacct cctgaatgtaa gcatgtgtaa ctgccccggc 3660
aacatctact tatatatcctt tataactcctt ttatattatc aaaaacgatct 3720
agtggaagat ccccttttgc aactctatga cccaaaatcctt ttaacgctgag tttctgtccc 3780
actgacgctcc agacacttgata gaaaaagctt ccagatcttt ttcagatcctt 3840
gcgtaactcg ctgctggctca aacaaaaaacc caacgcttac agcgctttgtt cttcgtgcggc 3960
atcaagagct accacatctct tttctcgcagtc ttaactgtgctt tcacgacgctc cagaatacacc 3960
atacgctgctct ctatgtctag cgatagctag goccacactt caaagaacctg atgacacgccg 4020
tctacattct cgtctgctgca atctcgtaac cggtgctgctc gcggcagctgg gggcagcgtgat 4080
gttccttgctg ttgctggctc agagactatc gtaaagatgg aggagcagctgg ggctggtggtgag 4140
cggcgggtttgc tctgcaacacg cccagctcga cagcgaagctc atcaccgagat cctgaatatc 4200
tacagcctgtct gccatgctgga agcagccgcc caacgccggcgcgc agagctcttt ttcagatcctt 4260
cggtaagctg caggtccgca aacagcggagc goccagaggtt cttcctgcaggg agaagagcct 4320
agtatctttac tagctgctgc ggctttcgcc acaatcagct tgacagctgac ttcgctgttcg 4380
getctgcctag cggcgggcagc ttccggaaaa acagccggccg cggcgccttc ttaactgtgcctt 4440
ttcgtgctgctg ttgctggctc attaagatgc gtaaagatgg aggagcagctgg ggctggtggtgag 4500
ataacgctgtat caccgcttctt gccatctgtg ctaaactgtc ttcagctgtcg aacgacgcagc 4560
gacgagctg agtcgctctgg cagactgcttc cagggagagtg agcgctcttg cttcgtggcctc 4620
atctgtctgg catttcacaa caccatagtt gccatcttctg cttggtctcgt ctgogcctgcag 4680
cagacagctg gtcgcttcag cggcgtgctg gcctgctcag cggcgccttc ttaactgtgcctt 4740
acacccacca acacccgctgc acagccctgtc cggctgtgctc gcgggcttcag cattttctctct 4800
cagacacggc tctgcgcttc cggcgtgctg gcctgctcag cggcgccttc ttaactgtgcctt 4860
gacacagcctgc ccctctggtctg gcggcgccttc ttaactgtgcctt 4920
ctgccccggc aacctctgtcgc ggtatgctgg ctagacgcag cggacgagctg gggcagcgtgat 4980
catttcaggg ttgctgattg cgaacacgtg acgttataacga tttcaacgatg atgtggcgtt 5040
ctgcctctgc ccgctgctgtt ggactcggct gacccccggc caacagcttctt ttcagctgttc 5100
ctgctttaag ccacgtacca ctcctctcttc gtaactgctg cttggtctcgt ctgogcctgcag 5160
cacacccacca acacccgctgc acagccctgtc cggctgtgctc gcgggcttcag cattttctctct 5220
gcccctgctg cggcgatctgc cggcgatctgc cggcgatctgc cggcgatctgc 5280
gtgggctgctg ttgtacagctgc aagcccggct gtcggtcgtc gttcgtcgtc gtggcctgctg 5340
cctgcagcag cggctgctgtt gctgccgttct cttcctgtgctt gcgggcttcag cattttctctct 5400
gtcctgtgctg cggctgctgtt gctgccgttct cttcctgtgctt gcgggcttcag cattttctctct 5460
cgccacacca ccctcaccgctg gagctgctgctg cggcgccttc ttaactgtgcctt 5520
gtcctgtgctg cggctgctgtt gctgccgttct cttcctgtgctt gcgggcttcag cattttctctct 5580
cgcgtgctgctg ttgctgctgctg ccataaatat cttcctgctg cttcaacatg ggcgatcagt 5640
gaacggaag gcagctggag tgcacggtcc gtttcttcac aaccctagca aatgctgaat
5700
gagggcatcg tttccacagt gatgctgtgtt gccacagact acagtcggct ggcgcacattg
5760
cgcgcatta cgcagctcgcc gctgcgctgtt gttgcgcgata tctgcgtagt gggatacgac
5820
gatgccag gaacgcgtat tattatcccc gcgtcaccac ccattaacaac gattttttcg
5880
cgtcctggggc aaccacgctt gagaagtttg tgcgaactctct ctaagggccg gcgggtgaag
5940
gcgcaatgc tgcgtcgcgcgc ctcaacgtttt aaaaaaaacc gcccctttgc gcgcacatcg
6000
caaacgcgct tttccgccgct gcggcgcgat ttctatagtg aagttgacagc accagttttcc
6060
cgactgtgaa ggcggcagcg gcggcaaccgc aatattagttg atggcggcgc aatggtatcg
6120

<210> SEQ ID NO 4
<211> LENGTH: 5994
<212> TYPE: DNA
<220> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 4

gttgcaacgc ttaattctgta tgcacggcttg caacacgtct tcgggcgctca gcgggcaatc 60
ggaacctgttg gtagggcggta acatcgtgca taatccgtgt cgcggcaagc 120
gcagctccct tcccagataa gtttcttcg cgcacatcat acggctacct gcaaatatcc 180
tgaaatgcgc ttggcagcata taatcctcgc gcgcgatataa ttgagctgaa tggagaagga 240
taacaatttc acacaagggc cagagctcaag aattctgggt gccgaaacgt 300
gaggggcttc ttttggcttg tcgggcggct ctatgcgggtg atcggcttcg ttggccgat 360
ggaacacgtt ataacactaa aatcgcgtatg ggcgggctca gatgacaca 420
cggcgagccag aacaccctat tataaatccaa aaccgattag cacgtggggc gcggcacaag 480
aaaggctcgg tcggcgcatt cccgctgtct gccggcagtt ggggctgtttt cgcaggataa 540
agtataaatag ctgacagcgtg gtaacgctgtg gggccggtat cgggtttcatt gacggcaccg 600
tcgggtctgc tgcgcaagcg ggaattgcgc aatatacagc ggcgggctct ataatttggtt 660
cgggcttctgc ctggagcagct acaagaagcg gcaatcctgcag cggagatgggt 720
catggactt gccgcttttt gcggcattag aagaggtttt cgcgcgtgaa cagattaaat 780
cagagcagcg aagcggtgtg atacaacaga atttgcttg ggccgagttg gcgggctctc 840
ccccgacgcc catgcgagca tcggagctta gcgcagctgg cggcagtgtg agtggtgggtt 900
cgccgcagtc gccaggaggg aactgcaggc catcacttttt aaccagggc tcagtttgg 960
gactgggcc ttctctcatt ctggtgttgtc tcggggacgc ctctctcagag tgcgacaatt 1020
cgcggctcgc ccgagctggag cttgcgaaag cgcgcggcgc gcgggcttcg gcgggacgcc 1080
cgcgcataaa tgcgcagcgc tcagccatta aagagggccc tccgagcggg tgggccccct 1140
gencttctac aacctttttt tggttttttt tccaatcata ttagctgcctga 1200
tgagcaataa aaccctaatg aatgcttcag aaccgagatt caatagttatc 1260
aacacccctg tcggtctcctc attctctttt ttgggctact tgcgcttcgctttttttgc 1320
aacccgaac gcgtgggttaa gtagatgatt tgcgaatctc gcggcgattcg aagttggttt 1380
aactcgaact gcagccgaaag tacgcttcgct gcggcggctgc gcggcggtttcg 1440
tttcaggtg atgcgcttccca aactctgttgct tctgtgccc gcggcttcgct 1500
cgcggcataa gcaactcggct cgcgcgata aatatttcga gatgctatgcgttct 1560
caccagtcac agaasaagct cttacggagt gctagcaagt aggagaattta tcgagtcgctg 1620
cctataaac acgcagcagcc aactacttct cagaacacagt gagagaccga 1680
agaggctaac ccgcttcttg cacacaactag ggattcagct atcctgctt ggctggtggt 1740
aaccagagcgt ggaattgcac atacccaaag cgaagctgta caccagctag ccatcacaac 1800	
tgcgaacacc gttgcgtaag aagtttcaca ttcaccaata atgaaaatag acaaaactcg 1860
gctctatatt ttgctccttc gatttcttca gggatccagtg aagctttattat ggaagaaaaa 1920
atcactctagc ataccacgcg tgcataatcc caatgcactta gtaaagaca tttgagggca 1980
ttcacacacg tcgctcatag taticataac cagagctttcc aagcttgattac tgcgcctttt 2040
ttaagacgcg taagaagaaaa taagcccaag ttttatcogg ctttttctca cattcttcgc 2100
cgcttgagtc atgtccatcc gaaatccgtg agttcagttga gctggtgata 2160
tgggtactg ctacaccttg ttacacgctt tcctatgagc aaactgaaac gtttctcatcg 2220
cctcgagttt atcaccacgc cggattcctgg cagtttttcac actatatttc gcaagatgtg 2280
ggttgattc gcgtaaacc acgcctattc cctaaggggt ttatttgaaat tagtttttctc 2340
gtctgccaca atcctctggtt gagtttcacc aagttttgatt taaacgtgccc caaatataggc 2400
aactcttctc cccctgctttt caccattgagc aataattatt cgaagcggga caagttgtcg 2460
atgctccgt cgttccaagt taccatgacc ggcctctggatt tgcgtccagt ggcggatagt 2520
cctatgctat taacacagta cttgagttgc tggccaggggg ggggttattt ttttaaaggg 2580
agttctggtc gcctttaaacc gccgcttcttg aacgtttgat aagctgataa aaggggtattg 2640
atgcagagga tccgagagaa aatctcagcc ggtgttcgggt tccggcgccag gtcgtaaat 2700
agcagcttact gttcattttg tttgttattc acggtacctg cggagcagct tggggctcgt 2760
gtctcttca aatcgcctgaggccgctttgc tccgctcttc cccgctggag ttaataattga 2820
cgatagtact atttactctag ctctcctagag cctagaattaa gctccttacta agcctggtat 2880
agcagatagc tgttctccaa cggagtccag cgcctctctag cgagcagat cggaaacacta 2940
agtgctgcg gctgcctttt gggggttgggt atgactgccc gcctctgcttg cggggtatcgt 3000	
tggggctcgtt cggcgcccttc ttctacgagt tcrgacgtta aagaagacgt ctaagtcrgcg 3060
gcgacagtag tcatgcctgac cgccacccgg aagagctgac cgcacagcgg tcggagctgt 3120
ttgtatctcg aatagcagat gaggcgcttc argcgcctga tcctgctgca tagtctgcgt 3180
gtacacacgg tgggttccag tctctgtgggg gggcaacttc aacgacacoct aaaaattactaa 3240
cgctgcctta aacctcctcta gtctcctcggc aacacttttaa aactacggtat gaggcggata 3300
aagttgcagc accctctctcg cgctgccttc ttcggcgttg ctctgtttatt gctgtataaat 3360
tctgcagcgg tggagctgggg tctctgcgta tcagtcacgcc aatgagccca aatagctggcg 3420
cctcctcctg tctgcgcttat cacccagagc gggacgcctc aacatctgat gaaacgtaata 3480
gcagagatgc ggtgagagat gctctacgtg taaagcctga ttcacagcga gacacagtttt 3540
acctctatat atcttgattat gattttatac attttataag atctaggtgta 3600
agatcttcttc tgcagcatcc atcgccccaa tacctcatacg tgaatctttgc ttcctacgtg 3660
cgtgacgcc ccggatcctta atcagaggtac cttcctgagta ttcttttttttt gctgcctggaa 3720				
tctggtgct gccaatcaca aacaccctgg taccagcggt ggttggttgtt cgagtcaagc 3780
agcagctaac gttttcttcc aaggtcactg gtctcgcagc agcagcagata ccaattactcg 3840
tccctctagg tgaagcttagc ttagccacac atctcaagaa ctctgtagca cgcogctacat 3900
acgctgcctt gctacatcctgttacaggg ctggcctcag tggcggataag tggctgttta 3960
cgggtgtgga ctcagaaca tagttacgga ataaggccga cgggtcgggc tgaaaggggg 4020
gtcgctgcoc acgccccagc tgggagcgcg gcaacccac ccgagtaaag agaagccggc 4080
gtgaactag aacagagccc acgcttcgac aaggagaaaa ggcgcgcagc tatccctgta 4140
gcgggagggg gaggagagac gacagagaa ggcagagacg ggcgctggttctgcc 4200
tttctagctg cggcgctcgt tcggcagcg ctcgactgttt ctcgctgtctc 4260
caggggggag gacgctatgg aaaaaagcca gcaagcgcgc ctggaggggt ccggcttcgct 4320
tttctgggcc cttttcttcc atctgttctt ctgctgttctt ctggatatgcc 4380
gttgacgct cttgtgaagg gcgtgatacg ctggccccgag cggagagccac gacgccccgg 4440
gagtcaagcg cggaggagcc ggagagccgct gatttgccgt tttttctcct cctggctcctgt 4500
gcgcatttccc acgacgccat tcgggcgcct tcgtcctctat ctcgctgtgc gctcgctagtt 4560
taaggcagta tacattacgcg tagctgactg atgggtgaca gcggacagcggc cccgaccc 4620
gccacacgc gcggagcggc ctcgtcaagg cccgactgcag ccgtccagaca 4680
gagtctggacc gccggcggc cgtgcatgtt ccagagctgt ttacagcctt caccagcggc 4740
cgagggcggcg ccggcgggag cggagacgcc gcgcctgggctgcctgtgct 4800
tagatatact gcagagtcct ccgtgagatag cctgcagcttg gcgagtatgg 4860
gggttggagg atgggtgaca gcgttacgctt caggtatggc cgggtcagctt 4920	taacagcggc tttttctggtgctgcgtgcgag cggacagccgc cttctgcgagc aagagggggc 4980
aaagttaggg cgggttgcgg ggggtgatgt tacattcaca cccgctgcgg gacacagctgt 5040
gcgggcaacc agcttgcggt tattgcccct gcgggtccca atgggtcctg gcacgccgagc 5100
tggccagattc tcggtcctgg ctatcatctgc gcgcagatcgc ctggtgctgcag ctgtgtgtgtg 5160	tcgatagatg tgactagagg gcgttgagcc gctttacggc cgggggtccaca ttctctcggc 5220
cacgaggggc gcggagagcc gcgggttgtc atcatcagag caggtcgcagcttgccggtg 5280
gacggtgtct gcacgttagtc ttcgctgttct ttcggtgatc gacaccccata cccctggct 5340
aacaggtatttt catctcttcgcc aagaggggct aggttcgagc ggtggagagcg tgtgcggtgca 5400
ttgggtgaccc agcgcagagc gctgttacgc gcccagatcttt cttccggtgctgcgcgccg 5460
cgtctgcctgg gcggctgctgct aacagctgtt gcagactcag atgggatcgc gcagagagccgc 5520
gggtgccgat gccgagccct gtctggctt gtcgtctctct cacaatcacc gtaaatgctg 5580
atcgtctccag ctcgagcggc ggtggtggat gcgtctagcgc gcgtgctgctg gcggagagccgc 5640
attcaagagt cgggtttgag cgtggctgtc gcgttgggtgt tagctccgtgc cggactctgtgc 5700
agagccacgc atatctgtat cccgctgctca accaccacca aacaggaattt gcggcctgtg 5760
gggcagagcc gcgggttcggc cttctgcgac tcctctgcgg ggggagggtag gaaaggggcaagc 5820
caggtgcaggg ccggtctcctg ggtggagagga aaaaagccgct acggcgcgcagc tacaagagggc 5880
gcctctcccag cgcgcttgcc gcgtttctccgtatggagctg cagcagcggt ctcccgagcgcgctt 5940
gggaggggag gcggagagcc gccagacttta tggagttccg gcggagaattt gcttg 5994

<210> SEQ ID NO 5
<211> LENGTH: 4581
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 5
gttgaaca gc ttatcatca ctgcacggtg ccaaaatgct tctggcgctca gcgcacgctc 60
ggaagcgtgtg gtaaicggct gcaacgtgta aatcaactca taatccgatg cgtcgaaggg 120
gcactctcag tctggaataa ttttttttgc cgcacactat aacgcttgctc gaaataatcc 180
tgaatagcgg tggtaaacaat taatatccgg gctggataaa tgtgtggaat tgtgacgaga 240
taacatattt cccacggagaa cagacccagt aatgtcgact ctggaagaac atccaaaaaa 300
gaggggctgt ttactgggcttg ttttggtgtt cctgctgccc atcagttttc ggcctccaata 360
atggacacag cagatcggatc actttaccaca agtggctaga aagttgatagt gcccgaatag 420
cggcgagcg aacatcgttg tatacacaata aatccggtgc gcgttgggca gatgacaaaa 480
aaatgtcgg tggcagccgat cccgtggcct ggtcatgttt cgcagacaatt cagggtaacc 540
atgatataatt gctcagttcct aaaccgttgcgtc ggttaaaatt cattacaccac 600
tacgctggga ctcggcagcgc ggaatcgcggg aatcatcggc ggcttttgcg caactgtggcc 660
cgggtgtcgtg cagaggcggc ctatcachtac atactccacta aacccgagtg cgcctgccag 720
catcgaacagt tgtgtgttttttgt cggtagtac ggaagattttt cagctgtaat cagatttaa 780
cagaagcgcc aacgtgttgat ataaacaaca atttgagtgc gcggcggttg acgcggtttc 840
ccatctgacc ctcagagatgctc gccgcagtcag gcgtgggtcc cccagcgatt tggcggttt 900
tcctccatgc gcggaggtgag gactgcagg cattacaataa aagcgagggc taactcgcgaa 960
gactgggcttc ctcaggttattc cttggttttgtt gttgactagc tctcagctgag taggacaaat 1020
cgcccgggag gctatgggaaagtccagc gacggtcggc agacaggtccgg gacagcgccg 1080
cgcctaaata ctcgagccgta ctaaatattag cgaaggtcaca tccgacgaga tggccttttt 1140
gctttttcatt aacaccatttt ttgatttttt ttaaatatttt ttcactaatg ttcgctgta 1200
tgacacataacc cccctgtataa atgtctcaat aatattggaa aagagaagttt atgcagttat 1260
aacatccctgcgtgcccttgtt atatcatcgttt tggcgtttgctg tttttttttgctg 1320
acccacgaga gctggtggpaat gtaaaagattg ctagaagatca tgggcgggtc cagatgggtt 1380
acatggaact gatcttcgaac agcggtaaag tccgtggagaattgtcgc cacgacaggtt 1440
ttccaatagttg cgtacatttt aacgcttgctg tattgggtgc gcgtatattcg cgggtgtcag 1500
cgggcaaga gctcaggtttc ggcgcatctaca atatctttca gatgcctggtg cgttgctact 1560
cacagtccgag aagacgtacc ttcgtcgggat gcacagtggc aagagaattg tgcagtgtgcg 1620
ccataacccattggttgagaga cagcgcccaacctctctt gcacgcagttgc ggagggccgg 1680
agggagagag agttttgttttgc ccaacagctgg gcagctgtagt ctgctcctgtcgtggtgg 1740
aacgcaggtc gatgaagcgc atacaaaaac gcagcgcgtg ccaacagctgg cttacagccaa 1800
tagcacaacag tccgctgaaatttaactg ggaactactt taccctgctactt cccggcaca 1860
aatataagca ctcggctgaggt gcgtataagat ctcggctgcac atcgtcgcctctctg 1920
cgcgctgcttg gttttattctg gataaatctg ggcggctgtag gcgggtggctct cgcggatata 1980
ttgacgactctt gcggcgacag ggtgatccct tgcggctgcaat ggttatctag aagacggggg 2040
gtccagcaca ttaggtgaaag caagagagact ggagcgtagt gcattctggc tcaactgata 2100
agcttcatgcgc cggtaatctc ctagattactc catatatgc ggatatgactc ttagaatctcct 2160
atatttatattc cttttaatgctt attttttttg atatctctgc aatatcactc ttaaatcttc 2220
cattcagcga gtttttcctgt gccgctggctc gacgccccgt gaaagagatc aagagatctt 2280
cttggagtgc cttttttttgtg gcgctaatct gcggtttgctc aaccaaaaac cccgctcgtc 2340
cagcggagtgt tatttgccg gatacagacg tacaacatct ttttccgaaag ttaacttggtc 2400
tcagcagacg gcaagataaa aatactggcc ttctagtgta gcctgttaa ggcacaccact 2460
tcaagacttc ttagcagccg cctctacatcc tcgctctcgt taatctgtta ccaatggctg 2520
cctcgacagtg ctaaaggctt tgtctaccgc gcggcgcaga tgcgaccccg ttcctccgaat 2580
agcgcagcgc gtcgggctga acggggtgcttg gtcgaccaaca gcggattg tcgaggaaga 2640
cctacaccgca actgctcagtc tccaacgctg acgatgatgaa aagcgcacac cttccggaag 2700
ggagagaggg gggacagcagtt cccgtaaacg gccgacgcgg gcacggcaggc cgcacggag 2760
agctcagacttg gggaaacgccc cgggtctcttt atagctccgt cgggttctcg cacctgctgac 2820
ttcgagctcag atttttgctg tggcctgcag ggggagggcag ccattggaaca aagcgcacca 2880
agccgccctct cattctgtcctgtcctcttt gcggcccttc tggcctcata gacccgctg 2940
ctggcattctcttcttcag ggaacccggcat aacacccggt gcgcgccttc gcgcgccttg 3000
gccgcacgcc gcacgacacg cagcagcagtt gcctgtagagc gcggcctgaa gcgcgtcta 3060
tgggccctcaa ttcaccctccct gtagtggct ccctcataac ccgctttcgag tgcctct 3120
tcatcactgct catccacgctg gcatagttac gcacgataat ccgtcagtttc gacccgtgta 3180
cggcgctagc tgtgctcggcg gacccggccgc aacacccggt gcgcgccttc gcgcgccttg 3240
tctccctccg gcattgcggtag cagagaacatc ccgcgtcttc gcgcgccttg 3300
aggttttcac cgcctcattcc gcggaaacccgg gcggcagccag atcaccgctgc gcgcggactc 3360
agcgcggcat gcattacgct gcaaccatac gaaggtcagca aaacctttcg aggttaggtaa 3420
tgcacccgc gcgcgctggacct gcaccctcgg tgggcttttag tgggaacccagt aaccttggtaa 3480
aggtgctcag ctagttcctcg tgcagcttctc gacgggcttt gcggcctgaa gcgcgtcttc 3540
agccgcctttc tgtgcatcgg gccgggagaga ggtcgggacac gcgcgccgag gcgctgatttc 3600
atacccaacc gcgcgccttc gcgggagaga ggtcgggacac gcgcgccgag gcgctgatttc 3660
acccgcacgc tgtgcgctccgc gcgcgtctcc gcaccctttcg gcgcgcctgaa gcgcgtcttc 3720
gacccgctcc gcgcgcttcc gcgcgtcttg atggtgaaac gcgcgccttc gcgcgcctgaa gcgcgtcttc 3780
aaagcgcggc tcgcacaattct ctctgcctgaa gcgcgctggct gcgcgttatc taactatcgcg 3840
cggtgtgcac gcggatgtccc gtcgggctttc gcgcgctgaa gcgcgtcttc gcgcgcctgaa gcgcgtcttc 3900
cttggctgtct gcgcgctggtg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 3960
cgcgtgctct gcgcgctgcgg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4020
ccattaagctg cgcgcgtctcc gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4080
aatccacataa aggccgatgca gcgcgcttgaa gcgcgcttgaa gcgcgccttc gcgcgccttc gcgcgccttc 4140
ccacacacgc aacaccgtgaa tgggggcattc gttccccctcg gcgcgatcgc gcgcgcctgcg 4200
cgcgtgctct gcgcgctgcgg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4260
atgcgtgtag cgggctggacct gcgcgctggct gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4320
acccgcacgc gccatggttctc ccgggacgcttg acggcgcgtctgc gcgcgccttc gcgcgccttc gcgcgccttc 4380
ctcgcgcgctac gcgcgctgcgg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4440
acacacgccgc gcgcgctgcgg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4500
cgcgctgctactccgcgctgcgcgcttc gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4560
ggcggtgc gcgggcttc gcgcgctgcg gcgcgccttc gcgcgccttc gcgcgcctgaa gcgcgtcttc 4620
Cys Cys Pro Gly Cys Cys

Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys
Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
Pro Met Met Ser Thr Phe Lys Val Leu Cys Gly Ala Val Leu Ser
Arg Val Asp Ala Gly Gln Glu Glu Leu Gly Arg Arg Ile His Tyr Ser
Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr
Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser
Asp Asn Thr Ala Ala Asn Leu Leu Thr Thr Ile Gly Gly Pro Lys
Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asp Gly Arg
Asp Thr Thr Met Pro Thr Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser
Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp

SEQ ID NO: 9
LENGTH: 319
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

SEQUENCE:
Met Ala Glu Leu Asn Tyr Ile Pro Asn Arg Val Ala Gln Gln Leu Ala
Gly Lys Gln Ser Leu Leu Ile Gly Val Ala Thr Ser Ser Leu Ala Leu
His Ala Pro Ser Gln Ile Val Ala Ala Ile Lys Ser Arg Ala Asp Gln
Leu Gly Ala Ser Val Val Ser Met Val Glu Arg Ser Gly Val Glu
Ala Cys Lys Ala Ala Val His Asn Leu Leu Ala Gln Arg Val Ser Gly
Leu Ile Ile Asn Tyr Pro Leu Asp Gln Asp Ala Ile Ala Val Glu
Ala Ala Cys Thr Asn Val Pro Ala Leu Phe Leu Asp Val Ser Asp Gln
Thr Pro Ile Asn Ser Ile Ile Phe Ser His Glu Asp Gly Thr Arg Leu
Gly Val Glu His Leu Val Ala Leu Gly His Gln Gln Ile Ala Leu
Ala Gly Pro Leu Ser Ser Val Ser Ala Arg Leu Arg Leu Ala Gly Trp
His Lys Tyr Leu Thr Arg Asn Gln Ile Gln Pro Ile Ala Glu Arg Glu
Gly Asp Trp Ser Ala Met Ser Gly Phe Gln Gln Thr Met Gln Met Leu
Asn Glu Gly Ile Val Pro Thr Ala Met Leu Val Ala Asn Asp Gln Met
Ala Leu Gly Ala Met Arg Ala Ile Thr Glu Ser Gly Leu Arg Val Gly
Ala Asp Ile Ser Val Gly Tyr Asp Thr Glu Asp Ser Ser Cys
Tyr Ile Pro Pro Ser Thr Thr Ile Lys Gln Asp Phe Arg Leu Leu Gly
Gln Thr Ser Val Asp Arg Leu Gln Leu Ser Gly Gin Gin Ala Val
Lys Gly Asn Gln Leu Leu Pro Val Ser Leu Val Lys Arg Lys Thr Thr
Leu Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg Ala Leu Ala Asp Ser
What is claimed is:

1. A method for producing a recombinant protein comprising: (a) culturing an E. coli bacterium under conditions in which the bacterium secretes a heterologous protein into a culture medium, wherein the E. coli bacterium comprises: (i) a mutant E. coli bacterium comprising mutations so that at least the NlpD gene product and at least one of the EnvZ, OmpR and YihF gene products is not expressed or is rendered non-functional; and (ii) a recombinant vector comprising a first DNA sequence encoding YebF operatively linked to a second DNA sequence encoding a heterologous protein, and (b) isolating the secreted protein from the culture medium.

2. The method of claim 1 further comprising the step of purifying the secreted protein.

3. A method for producing a heterologous protein comprising: (a) transforming a host cell with a recombinant vector, wherein the host cell is a mutant E. coli bacterium comprising mutations so that at least the NlpD gene product and at least one of the EnvZ, OmpR and YihF gene products is not expressed or is rendered non-functional, and wherein the recombinant vector comprises a first DNA sequence encoding YebF operatively linked to a second DNA sequence encoding a heterologous protein; (b) culturing the host cell under conditions in which the bacterium secretes the heterologous protein into the culture medium; and (c) isolating the secreted protein from the culture medium.

4. The method of claim 3 further comprising the step of purifying the secreted protein.