



US009410157B2

(12) **United States Patent**
Withers, III et al.

(10) **Patent No.:** **US 9,410,157 B2**
(45) **Date of Patent:** **Aug. 9, 2016**

(54) **SYSTEMS AND METHODS FOR THE
SECRETION OF RECOMBINANT PROTEINS
IN GRAM NEGATIVE BACTERIA**

(75) Inventors: **Sydnor T. Withers, III**, Madison, WI
(US); **Miguel A. Dominguez**, Madison,
WI (US); **Matthew P. DeLisa**, Ithaca,
NY (US); **Charles H. Haitjema**, Ithaca,
NY (US)

(73) 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 81 days.

(21) Appl. No.: **13/192,058**

(22) Filed: **Jul. 27, 2011**

(65) **Prior Publication Data**

US 2012/0225453 A1 Sep. 6, 2012

Related U.S. Application Data

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

(51) **Int. Cl.**
C12N 15/70 (2006.01)
C12P 21/02 (2006.01)

(52) **U.S. Cl.**
CPC **C12N 15/70** (2013.01); **C12P 21/02**
(2013.01)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,951,361 B2 * 5/2011 Turner et al. 424/93.1

OTHER PUBLICATIONS

Prehna et al., *Structure* 20, 1154-1166, 2012.*

Ward et al. *Nature* 341:544-546, 1989.*

Zhang, et al., "Extracellular accumulation of recombinant proteins
fused to the carrier protein YebF in *Escherichia coli*," *Nat.*
Biotechnol., Jan. 2006, No. 24, vol. 1, pp. 100-104.

* cited by examiner

Primary Examiner — Nancy T Vogel

(74) *Attorney, Agent, or Firm* — Casimir Jones S.C.

(57) **ABSTRACT**

Disclosed herein are systems and methods for producing
recombinant proteins utilizing mutant *E. coli* strains contain-
ing expression vectors carrying nucleic acids encoding the
proteins, and secretory signal sequences to direct the secre-
tion of the proteins to the culture medium. Host cells trans-
formed with the expression vectors are also provided.

19 Claims, 12 Drawing Sheets

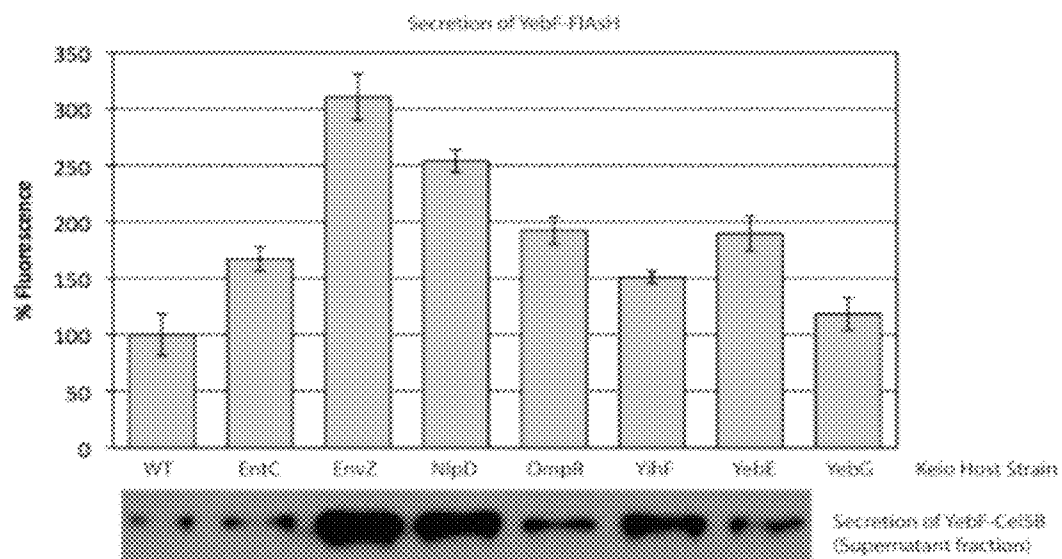
FIG. 1

FIG. 2

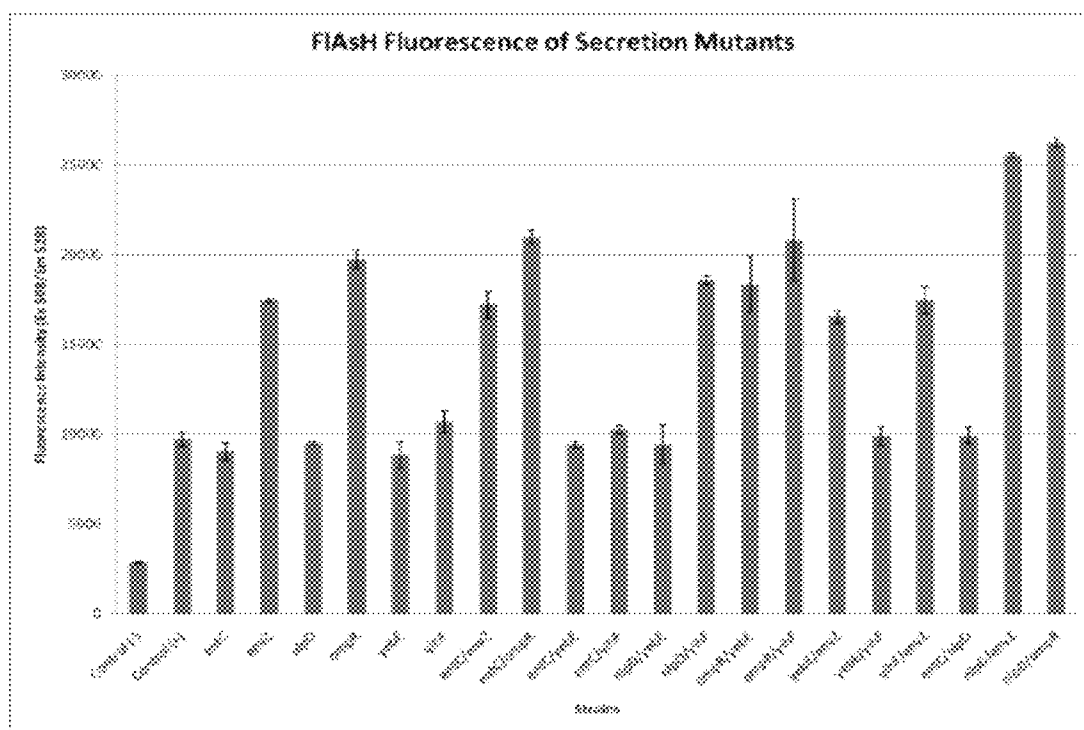


FIG. 3

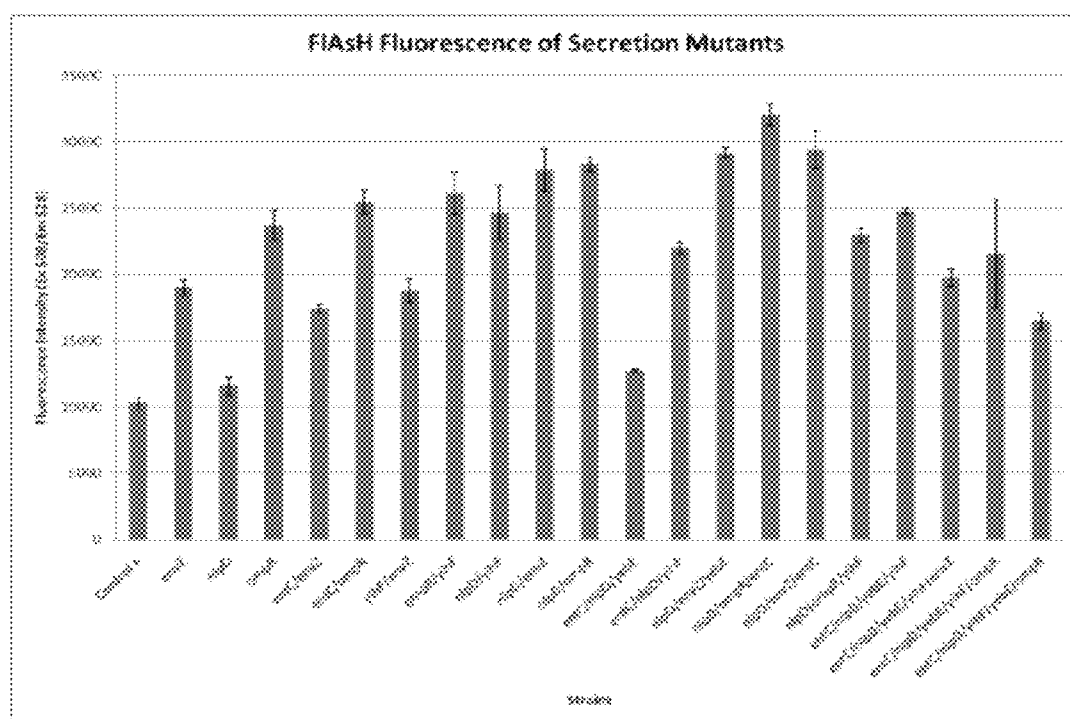


FIG. 4

PLASMID MAPS

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

FEATURES

	Location/Qualifiers
source	1..8120
	/organism="Cloning vector pTrc99a"
	/mol_type="genomic DNA"
	/db_xref="taxon:40992"
	/lab_host="Escherichia coli"
	/note="derived from pBR322-2"
misc_feature	1..17
	/note="derived from cloning vector pBR322"
promoter	18..363
	/note="trc promoter from pBR322-2"
	/citation=[1]
misc_feature	284..270
	/note="NcoI/EcoRI linker"
misc_feature	2273..2697
	/note="3S RNA, T1, T2, rnsB"
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..2261
	/note="6XHis" (SEQ ID NO: 6)
gene	291..644
	/note="YebF"
gene	651..2243
	/note="Cel5B"

ORIGIN

1	gtttgswagc	ttatcataga	cgcacggtg	aaccaatgct	tctgggagca	ggagacuatc
81	ggaagctgtg	gtatggctgt	gcaggttgtt	aatccctgca	taattcgtgt	cgctcaaggg
161	gcactccggt	tctggataat	gtttttttgg	cgcacatcat	aacggtttctg	gcacatattc
181	tgaattgagc	tggtgacaa	taatacatcg	gctcgtataa	tgtgtggaat	tgtgagcgga
241	caacaatttc	acacaggaaa	cagacaaatg	aattcgagct	cGACAAAAAC	ATGAAAAAA
301	CAGCGGCTTT	TTTACGGCTC	TGTTCCTTT	CTGCTCTGTC	ATCAGTTTTC	GCTGCTAATA
361	ATGAAACCAE	CAATCGCTC	ACTTTCCTAA	AGTGTGAAGA	TCTGCTATCT	GCTGCAATTC
421	CCGCGAGCCT	AAACCTGAT	TATCAACAAA	ATCCCTGCTC	GCCTTGCTCA	GATGCTCAAA
481	AAATTCTCTG	TCAGGCTGAT	CCCCTGCTCT	GGTCTCATTT	GCAGGACATT	CAGGCTAAAG
541	ATGATAAATG	GTCAATACCG	CTAACCTGCT	GTGCTAAAGG	TGCTGATATT	CATTACCAAG
601	TCAGCTGCGA	CTGCAAGCG	GGAATGCTG	AAATATCAGC	GCCTTCTAGA	GATGCTGCTC
661	CATTGAGCTT	GCAAGCAAC	AAGATCTCTG	CGAATGCTCA	GCCTGCTGAG	TTCAGCTGTA
721	TCAGCTCTTT	TTGCAACCAAT	ACCGCTGCTG	GTGCTGAGAA	GTACTATAAC	GCTCAACTTC
781	TTTCTCTGTT	GAAATCGGAT	TGGAACGCTA	AGCTGCTCTG	GCCTGCTGAT	GCTGCTGCTG
841	ATGAGGCTCG	TTACTGCAAC	GACCTGCTCA	ATAGGATCTG	CTGACTCAAA	GCTGCTGCTG
901	CAGCGATCTC	AAACGACATG	TACGCTGCTA	TCGACTGCTA	TAGCTATAAT	GCACACCAAT
961	ATCAGTCTCA	GCTCATGCTC	TTCTTCTGAG	AGATGCTCTG	CAATATCTCT	GCTGACCAAC

FIG. 4, con't.

1081 ACGTGATCTA TGAATCTAC AATCAGCCTT TGCAGTGCAG CTGGTCTAAC ACTATCAAAC
1091 CGTATGCGCA AGCGGTGATT GCGGCGATCC GTGCGATTGA CCCAGACAAT CTGATTATCG
1141 TGGGTACGCC GACCTGGAGC CAGGATGTTC ACGTGGCGGC GAATGACCCG ATTACGGGTT
1201 ACCAGAACAT TGGGTATACC CTGCATTTCT ATGCGGGTAC GCACGGTCAA TACCTGCGTG
1241 ATAAGGCACA GACCGCACTG AATCGTGGCA TTGCTCTGTT TGTACCGAA TGGGGCTCGG
1321 TTAATGCCAA TGGTGATGCC GCGGTTCCTA ATAGGUAAC CAATGCTTGG CTGACCTTTA
1391 TGAAAACCAA TCACATCTTC AACGCGAACT GGGCACTGAA TGACAAAGTT GAGGCGGCAA
1441 GCGCATTTGGT CCGGGGTGCC AGCGCAAACT GCGGCTGGGT TAACAGCCAA TTGACCGCCT
1501 CCGGCGCTCT GGCACAAAGC ATCATCAGCG GCTGGGCGAG CTACAAATCC AGCTCCAGCA
1541 GCAGCGCGGT TTCCAGGTCAG ACGCAAGTCA GCAGTTGGTC CTAAGCCCGG GTGCTGTCTA
1621 GCTCTAGCAG CACGCGCTTC AGCGTGGTTA GCTCCGCTGT CAGCGGCGAA CAGTCTAACT
1691 GGTATGGTAC GTTGATATCA CTGTGACGCA CGACCAAGAA CGGTTGGGGT TGGGAAAACA
1741 ACGCGTCTGT CATTTGCTGT GCAACGTGCA GCGGTACAGC GGCACCGTGG GGTATCTGTC
1801 GCGGTATGAC CAGCAGCCAA GGTTCCTTCA GCGTGGCAG CAGCAGCAGC TCTCTGCTCA
1841 GTTCCAGCTG TAGCAGTACG AGCAGCTTTC TTCACTCTAG CAGTGGGCTT TGTGGGTGG
1921 CGACAGCAGC GGCAGCAGC AGCGGCGCTT GCAGCTACAC CGTTACCAAT CAGTGGAGCA
1991 ACGGTTTTAC CGCATCTATC CGTATTGCGA ACAATGGGAC CAGCCCGATC AACGGTTGGA
2041 ATCTCAGCTG GAGCTACTCT GACGCTAGCC GTGTTACCAA TTCTTGGAAC GCGAATGTGT
2101 CTGGCAATAA CCAATACACC GCATCTAACC TGGGTTGGAA TGGCAGCATT CAACCGGCTC
2141 AAGCTGTGGA GTTGGTTTT CAGGCGACCA AGAATAACAG CAGTGGGCTT ATCCGAGCTT
2221 TGACCGGCAA CGTGTGCAAC AACCATCATC ACCATCAGCA CTAAaagctt ggtgtgtttt
2291 ggggtgtgag gaagattttt agcctgtatc agatttaatt gaagcgcaga aggggtctga
2341 taaaacagaa ttgtcctggc ggcagtagag cgggtggctc acctgacccc atggcgaact
2401 cagaagtga agcgcgtagc gccgatggta gtgtggggtc tcccatgctg agagtggga
2441 actgcacagg ctcaaatcaa agcaagagct cagtcgaagg actgggcttt tggttttatc
2521 tgttgtttgt cggtgaaagc tctcctgagt aggaatcaat ccgcgggagc ggatttgaac
2591 gttgcgaagc aagcgcccg aggggtggcg gacggagcgc cgcctaaac tgcacggcat
2641 caaatgaagc agaaagccat cctgacggat ggcctttttg cgtttctaca aactctttt
2701 gtttattttt ctaaatatat tcaaatatgt atcgcctcat gagcaataa cctgtataa
2741 tgtttcaata ctattgaaas aggaagagta tgaattctca acatttcgtg tgcgcctta
2821 ttcccttttt tgcggcattt tgccttctgt tttttgtca cccagaaagc ctgggtgaag
2891 taaaagatgc tgaagatcag ttgggtgcac gagtgggtta cctcgaatg gatctcaaca
2941 ggggtgaagc ccttgagagt ttgcgcacag aagaacgttt tccaatgatg aggaacttta
3001 aagttctgt atgtggcgc gtattatccc gtgttgagc cgggcgaagc caactcggtc
3041 gccgcataca ctatttcag aatgacttgy ttgaagtact accagtca gaaaagcatc
3121 ttacggatgg catgacagta agagaattat gcagtgtctc ctataccatg agtgataaca
3191 ctgaggccaa cttaacttctg acaacgatcg gaggacgga ggaactaac gtttttttgc
3241 acaaatggg ggaatcatgt actgccttg atcgttggga accggagctg aatgaagcca
3301 taccaaagc cgaagctgac accacgatgc ctacagaaat ggcaacaagc ttgcgcgaac
3341 ttttaacgcy cgaactactt actctagctt ccgggcaaca atcaatagac ttgatggagg
3421 cggataaagt tgcaggacca ctctctgctt cggccttcc ggctggctgg ttatttgtg
3491 ataatctctg agcgggtgag vgtgggtctc ggggtatcat tgcagaaatg gggcagatg
3541 gtaagccctc ccgtatcgt gttatctaca cgaaggagg tcaggcaact atggatgaac
3601 gaaatagaca gatcgcgtg ataggtgctt cactgattaa gcaattggtta ctgtcagacc
3641 aagttacttc atatatactt tagattgatt taaaacttca tttttaattt aaaaagatct
3721 aggtgaagat cttttttgat aatctcatga ccaaaatccc ttaacgtgag ttctgttcc
3791 acgagagctc agaccccgta gaaagatca aaggtatctc ttgagatcct tttttctgc
3841 gggtaatctg ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt tgtttgcgg
3901 atcaagagct acaactctt tttccgag taacttggct cagcagagcg cagataccaa
3961 ctactgtctt totagtgtag ccgtagttag gccaccactt caagaactct gtagcaaccg
4021 ctacataact cgtctgtgt atactgttac cagtgggtgc tgcagtgcc gatgaagctg
4091 gtcttaccgg gttggaactc agacgatagt taccggataa ggcgcagcgg tgggctga
4141 cgggggggtt gtgcacacag cccagcttgg agcgaaagac ctacaccgaa ctgagatacc
4201 taccagcgtg gttatgagaa agcgcacgc ttcgggaag gagaaaggcg gacgggtatc
4261 cgttaagcgg cagggtcggg acaggaagga gcacgaagga gattccagg ggaacgcct
4321 ggtatctttt tagtctgtc ggggtttgac acctctgact tgagcgtoga tttttgtgat
4391 gtcgctcagg gggcggagc ctatggaaaa acgcagcaca cggggcctt ttaaggctcc
4441 tggccttttt ctggcctttt gctcacatgt tctttctctg gttatccctt gattctgtg
4501 ataacggtat taacgccttt gagttagctg atacggctcg ccgcagcaga acgcagcagc
4561 ggaagcagtc agtgagcag gaagcggag agcgcctgat ggggtatttt ctacttacc
4621 atctgtcggg tatttcacac cgcataatgt gaactctcag tacaatctgc tctgatgacg

FIG. 4, con't.

4681 cataghtaaq ccagttataca ctccgctatc gctacgtgac tgggtcatgg ctgagccccg
4741 ccaccccgca acacccgctg acgcgccttg acgggcttgt ctgtccnngy cctccgctta
4801 cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcc cgtcatcacc
4861 gaaaacgcgc aggcagcaga tcaattcgcg cgcgaaggcg aagcggcctg ccttcaagtt
4921 gacacccatcg aatgggtgaa aaccttttgc ggtatggcat gatagcgcc ggaagagagt
4981 caatccaggg tggtagaagt gaaacvagt acgttatag atgtcgcaga gtatgccggt
5041 gtctctctct agaccggttc ccgcgtggtg aaccaggcca gcacagttc tgcgaaacag
5101 cgggaaaaaa tggaaagcgcc gatggcgagg ctgaattaca ttcccaaccg cgtggcacaa
5161 caactggcgg gnaaacagtc gttgctgatt ggcgttgcca ctccagctt ggcctgccc
5221 ggcgcgctgc aaattgtgcg ggcgattaaa tctcgcgccg atcaactgg tgcagcggtg
5281 gtggtgtcga tggtagaagc aagcggcgtc gaagcctgta aagcggcggt gcacaaacct
5341 ctgcgcgaac ggcgtcgttg gctgctcatt aactatccgc tggatgaccc ggtatgccatt
5401 gctgttggaag ctgcctgcac taatgttccg gcgttatctc ttgatgtctc tgaccagaca
5461 cccatcaaca gtatcttttt ctcccatgaa gacggtagcc gactggcggt ggaacaactg
5521 gtgcgaattg gtaaccagca aatcgcgctg ttacggggcc cattaagttc tgtctcggcg
5581 cgtctgcgtc tggctgggtg gcatcaatat ctccctcgca ctcaaatcca gcgatagcg
5641 gaacggggaag ggcgctggag tgcctatgac ggttttcaac caaccatgca aatgctgaat
5701 gagggcatcg tcccactgc gatgctggtt gcacacgac agatggcgct gggcgcaatg
5761 cgcgcattta ccgagtrcgg gctgcgcgtt ggtgcggata tctcggtagt gggatcacgac
5821 gataccgaag acagctcatg ttatatcccg ccgtcaacca ccatacaaca ggaatttctc
5881 ctgctggggc aaacacgggt ggaacgcttg ctgcaactct ctcagggcca ggcgggtgag
5941 ggcacatcgc tgttgcccgct ctccactggtg aaagaaaaa ccaaccggc gcccacacg
6001 caaacccgct ctccccgcgc gttgycgat tcatcaatgc agctggcagc acaggtttcc
6061 cgactggaaa ggcggcgctg agcgcacacg aattaatgtg agttagcgcy aatcgaactg

FIG. 4, con't.

pTUC93a(Cm)-YebX-FlaA₂-ShiA (SEQ ID NO: 6) (SEQ ID NO: 4)

FEATURES	Location/Qualifiers
promoter	193..266 /label=trc_promoter /ApEinfo_fwdcolor="##04040" /ApEinfo_revcolor="##04040"
misc_feature	235..257 /label=Mil_pUC_rev_primer /ApEinfo_fwdcolor="##0ff80" /ApEinfo_revcolor="##0ff80"
misc_feature	complement(764..781) /label=pBAD_rev_primer /ApEinfo_fwdcolor="##0ff80" /ApEinfo_revcolor="##0ff80"
misc_feature	complement(764..781) /label=pTtrcHis_rev_primer /ApEinfo_fwdcolor="##0ff80" /ApEinfo_revcolor="##0ff80"
terminator	814..971 /label=rrnB_terminator /ApEinfo_fwdcolor="##f8080" /ApEinfo_revcolor="##f8080"
terminator	937..980 /label=rrnB_T1_terminator /ApEinfo_fwdcolor="##f8080" /ApEinfo_revcolor="##f8080"
terminator	1113..1139 /label=rrnB_T2_terminator /ApEinfo_fwdcolor="##f8080" /ApEinfo_revcolor="##f8080"
promoter	1181..1209 /label=AmpR_promoter /ApEinfo_fwdcolor="##04040" /ApEinfo_revcolor="##04040"
CDS	1251..1816 /gene="Ampicillin" /note="ORF frame 3" /translation="MSIQHFEVALIPFFAASPCLPVFAHFEETLVKVKDAEQQLGARVGY IRLEKNSGHILESRPPEHFFWMTTFKVLICGAVLSRYDAGQEQLRRIRYSQNDLVE YFPTVERHLTDGMTVRLCSAAITMSDNTKANELETTIDGPKELTAFLHSMGDSHVTSL DSWPELNEAIPIQDERDTNPTAMATTLAKLLTCELLTLASRQQLIQWNEADKVAQFL LRGALPAGWFIADKSGAGERSGRIIAALGFDCKPSRIVVIYTTGSGATNCKRNRQIA EIGASLIKHW" (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="##f8040" /ApEinfo_revcolor="##f8040"
rep_origin	3679..4298 /label=pBR322_origin /ApEinfo_fwdcolor="##f8000" /ApEinfo_revcolor="##f8000"
misc_feature	4685..4717 /label=pGEX_3_primer /ApEinfo_fwdcolor="##0ff80"

FIG. 4, con't.

```
/ApEinfo_revcolor="#80ff80"
misc_feature 4864..5955
  /label=lsc1
  /ApEinfo_fwdcolor="#80ff80"
  /ApEinfo_revcolor="#80ff80"
CDS
4956..5955
  /translation="MRRLNYIPNRVAQQLACKQELLIGVATSSSLALHAPESQIVASIKK
KADQLGASVYVSMVERSEVEACKAAVHNLLAGRVSELIINTFLDQQAIAVEAACTNV
PALFLDVEDQTFINSIITSHEGTRLGVENLVALCHQQAIALLAGPLSGVSARLRLAGW
HKYLTRNQIQPIAREGDSKMSGFQQTMQMLNDSIVPTANLVANDQNALGAPRAITE
SCLHVQANHGVVQYDQTEDESCYIIPPSTFIHQGWHLLOQTUVENMLQLSQCCQVKKQKQ
LLPVGLVERKTTIAPNTQFASPRALADLHAGLANQVERLEKQ*" (SEQ ID NO: 9)
  /label=ORF frame 1
  /ApEinfo_fwdcolor="#c0c0c0"
  /ApEinfo_revcolor="#c0c0c0"
CDS
3231..3524
  /gene="Ampicillin"
  /note="ORF frame 3"
  /translation="MSIQHFRVALLIPFPAACLPVFAHPETLVKVKDAEDQLGARVGY
IHLINSEKILDELPPFERHPDMSTFHYLLCNVLSKVDAGQGLGRADNYTQKHLVE
YSEVYDSHLLTGNTVHELDAAIYMSNTANILITTIQGIKELTAPLDKMSDPTKL
DRNEFELNEAIPINDERDTTMTPTAMATTIRRLTGGELLTLASRQQLIDWMEASKVGSL
LRKALFAGRFIARRKRRGRRGGRIIALGVDSKFKRVYVIYTTSGQATMDENRNGIA
EISAEILKHM*" (SEQ ID NO: 8)
  /label=Ampicillin(8)
  /ApEinfo_label="Ampicillin"
  /ApEinfo_fwdcolor="#c0c0c0"
  /ApEinfo_revcolor="#c0c0c0"
gene
3231..3524
  /gene="Ampicillin"
  /label=Ampicillin(3)
  /ApEinfo_label="Ampicillin"
  /ApEinfo_fwdcolor="#ff8040"
  /ApEinfo_revcolor="#ff8040"
misc_feature
1817..3230
  /label=CmR
  /ApEinfo_fwdcolor="#008040"
  /ApEinfo_revcolor="#008040"
misc_feature
681..698
  /note="6XHis" (SEQ ID NO: 6)
misc_binding
645..680
  /note="FlAsH"
gene
291..644
  /note="TubP"
ORIGIN
1 gtttgacagc ttatcataga ctgcacggtg caccaatgct tctggcgtga ggcagccatc
61 ggaagctgtg gtatggctgt gcaggctgtg aatcaactga taattcgtgt cgtccaggc
121 gcaatccgtg ctggtgataa gtttctgtcg ccgscatcat aacggttctg gcaaatctc
181 tgaatagagc tgttgacaat taatcatcgg gctcgtataa tgtgtggaa tgtgagcgg
241 taacaatttc acacagggaa cagccatcgg AATTCGAGCT CGAGAAAAC ATGAAAAAA
301 GAGGGGCGTT TTTAGGSCGT TTGTTGGTTT CTGGCTGCGG ATCAATTTTC GCTGCCAAT
361 ATGAAACCA CAACTCGGTC ACTTTCCCA AATGTGAAGA TCTGATGCT GCGGGAATT
421 CCCTGAGCTT AAAACGTGAT TATCAACAAA ATCCCTGCTG GCTTTGAGCA GATGATCAA
481 AAATTGTGCG TCAGGCGGAT CCGCTGGTTT GCGTCAGTTT CGAGGACATT CAGGGTAAAG
541 ATGATAAATG GTCAATACCG CTACCGTTC GTGGTAAAG TCGGATATT CATACCAAG
601 TCAGCGTGA CTECAAGCGG GAAATGCGCG AATATTAGCG GCTTTTCTG AACTGCTGCC
661 CAGGCTGCTG CATGGAACCG CATCATCAC ATCACACTA Aactagagtc gacctgcagg
721 catgcagctt tggctgtttt ggcggatgag agsagatttt cagcctgata cagattaat
781 cagaacgag aagcgtgtctg atcaaacaga attgctctgg cggcagtagc ggggtggtcc
841 caactgaccc catgcagac toagaagtga aacgcgtgag cgcagatggt agtgtgggt
```

FIG. 4, con't.

901 cccccatgc gagagtaggg aactgcacgg catcaaatca aacgaaggcc ccagtcgaaa
961 gactggggcct ttctgttttat ctgtttgtttg tgggtgaaacg ctctcctgag taggacaaat
1021 ccgcgcgggag cggattttgaa cgttgcgaag caacgggcccg gagggtggcg ggcagggacg
1081 ccgcataaaa ctgcacggca tcaaatlaag cagaaggcca tcttgacgga tggccttttt
1141 gcgtttctac aaactctttt tgttttatitt tctaaataca ttcaaatatg tatccgctca
1201 tgagacaata accctgataa atgcttcaat aatattgaaa aaggaaaggt atgagtattc
1261 aacatttcog tgtcgccctt attccctttt ttggggcatt ttgccttctt gtttttgctc
1321 acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gtcgggtgca cgggtgggtt
1381 acatcgact ggatctcaac agcggtaaga tccctgagag ttttcgcgcc gaagaaagtt
1441 ttccaatgat gagcactttt aaegtctctg tatgtggcgc ggtattatcc cgtgttgacg
1501 ccgggcaaga gcaactcgtt cgcgcatac actattctca gaatgacttg gttgagtact
1561 caccagtcac agaaaagcat cttacggatg gcatgacagt sagagaatta tgcagtgcgt
1621 ccataaccat gagtataac actgcggcca acttacttct gacaaagatc ggaggaacga
1681 aggagtaac cgtttttttg cacacatgg gggatcatgt aactgcctt gatcgttggg
1741 aacggagct gaatgaagc atacaaacg acgagcgtga caccagatg cctcagcaaa
1801 tggcaaaaac gttgcgttaag aggttccaac ttccaccata atgaataaag atcaataccg
1861 ggcgtatttt ttgagttatc gagattttcc ggagctaagg aagctaaat ggagaaaaaa
1921 atcactggat ataccaccgt tgatatatcc caatggcatc gtaaaagaca ttttgagga
1981 tttcagtcag ttgtcaaatg taactataac cagaacgttc agtggatata tacggccttt
2041 ttacagaccy taagaaaaa taagcaacag ttttatccgg ctttattoa cattctgnc
2101 cgcctgatga atgctcatcc ggaattccgt atggcaatga aagacgggtg gctgggtata
2161 tgggtagtg ttcaacctg ttacaccgtt ttccatgagc aaactgaac gtttfoatcg
2221 ctctggagtg aataccacya cgatttccgg cagtttctac acatatattc gcaagatgtg
2281 gcgtgttacg gtgaasact ggctatttcc cctaaagggt ttattgagaa tatgttttcc
2341 gtctcagcca atccctgggt gagtttccac agttttgatt taacgtggc caatgtgac
2401 aacttcttcg ccccgctttt caccatgggc aaatattata ccgaaggcca caaggtgctg
2461 atgcccgtgg cgattcaggt toatcatgac gtctgtgatg gcttccatgt cggcagatg
2521 cttaatgaat tacaacagta ctgcgatgag tggcagggcg gggcgttaatt tcttaaggc
2581 agttattggt gcccttaaac gcttgggtgt acgcctgaat aagtgaat aagcggatga
2641 atggcagaaa ttggaagca aatgcaccc ggttgtcgtt tcaagggcgg gtcgttaaat
2701 agccgcttat gtctattgct ggtttaccgg ttatttgact accggaagca gctgaacgt
2761 gtgcttctca aatgcctgag gccagtttgc tcaggctctc ccggtggagg taataattga
2821 cgatatgata etttattctg cctccagag cctgataaaa accggttagcg cttcgttaat
2881 acagatgtag gtgttcacac gggtagccag cagcatcctg cgatgcagat cgggaacata
2941 atggtgcagg gcgttgtttt vggcgtgggt acggtggcag gccccgtggc cgggggactg
3001 ttggggcgtg ccggcaacct tccacagat tgcattgata agaaagacag cataagtgcg
3061 gcgaagatag tcatgcccgc gccccaccgg aaggagutac cggacagcgg tgcggaatgt
3121 tgtaactcag aataagaaat gaggcgcgtc atggcgttga ctctcagta tagtatcgtg
3181 gtatcaccgg ttggttcacac tctctgtttg gggcaacttc agcagcaacg aaactattaa
3241 ctggcgcaact acttactcta gcttccggc acaaatlaat agactggatg gaggcgata
3301 aagttagcag accactcttg cgttcggccc ttccggctgg ctggtttatt gctgataaat
3361 ctggcgccgg tgagcgtggg tctcgcggtc tcatlgvagc actggggcca gatgtaaac
3421 cctcccgat cgtagttatc tacacgacgg ggagtcaggc aactatggat gaacgaata
3481 gacagatcgc tgagataggt gcttactga ttaagcattg gtaactgtca gaccaagttt
3541 actcatatat actttagatt gatttassac ttcattttta atttaaaagg acttaagtg
3601 agatcctttt tgataatctc atgacaaaaa tcccttaacg tgagttttcg ttccactgag
3661 cgtcagaccc cgtagaaaag atcaaaaggt ctctttgaga tccctttttt ctgcgcgtaa
3721 tctgctgctt gcaaacaaaa aaaccaccgc taccagcgtt ggtttgtttg cggatcaag
3781 agctaccaac tctttttccg aaggttaactg gcttcagcag agncagata cccatactg
3841 tcttctcagt gtagcgttag ttaggccacc acttaaaaga ctctgtaga cgcctaat
3901 acctcgctct gataatctg ttaccagtgg ctgctgcccag tggcgataag tegtgtctta
3961 ccgggttggc ctcaagacya tagttaccgg ataagggcca gcggtcgggc tgaacggggg
4021 gttcgtgac acagcccagc ttggagcgaa cgaactacac cgaactgaga tacctacagc
4081 gtgagctatg agaaagcgc acgcttcccg aagggaagaa ggcggacagg tatccgtaa
4141 gcggcagggt cggaaagcga gagcgcagca gggagcttcc aggggaaac gctgggtac
4201 tttatagttc tgtcgggttt cgcacactct gacttgagcg tggatttttg tgatgctcgt
4261 cagggggggcg ggcctatgg aaaaacgcgc gaaacggggc ctttttaagg ttcctgacct
4321 tttgctggcc ttttgcacac atgttctttc ctgggttata ccttgattct gggataaac
4381 gtattaccgc ctttgagtg gctgataccg ctgcgcgag ccgaacgacc gagcgcagc
4441 agtcagtgag caggaagcgc gaagagcgcc tgatgcgtta tctctcctt acccatctg

FIG. 4, con't.

```
4501 ggggtatttc aaacgcata tggtycaatc tcagtaaat ctgtctgat gccgcatagt
4561 taagccagta taactcagg tatcgctag tgactgggtc atggctgogc ccgscaccc
4621 gccaacaccc gctgacggc cctgacgggc ttgtctgctc ccggcatccg cttacagaca
4681 agctgtgacc gtctccggga gctgcatgtg tcagaggttt tcacgtcat cacgaaacg
4741 ccgagggcag cagatcaatt cgcgcgcgaa ggcgaagcgg catgcatctc cgttgacacv
4801 atcgaaatggt gcaaacctt tcgcggtatg gcctgatagc gcccggaaga gagtcaattc
4861 aggggtggtga atgtgaacc agtaacgtta tacgatgtcg cagagtatgc cgggtgtctct
4921 taccagaccg ttcccggt ggtgaaccag gccagccaag ttctctgcaa aacgcgggaa
4981 aaagtggaa ggcgatggc ggagctgaat tacattccca accggtggc acaacaactg
5041 gggggcaaac agtcgttgt gatggggtt gccaccccca gtctggccct gcaagcgcg
5101 tcgcaaatg tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtggtggtg
5161 tcgatggtag aacgaagcgg cgtcgaagcc tgttaagcgg cggtgcaaac tcttctcgg
5221 caacgcgtca gtgggtgat cattaactat ccgctggatg accaggaagc cattgctgtg
5281 gaagetgctt gcaataatgt tccggcgtta ttctctgatg tctctgacca gacaccatc
5341 aacagtatta tttctccca tgaagacggg ccgcgactgg gcgtggagca tctggtcgca
5401 ttgggtcacc agcaaatcgc gctgttagcg ggccattaa gttctgtctc ggcggtctg
5461 cgtctggctg gctggcctaa atctctcaet cgcactcaaa ttccgccgat agcggaacgg
5521 gaagcgact ggagtgcct gtcgggttt caacaaacca tgcaaatgct gaatgagggc
5581 atcgttccca ctgcgatgct ggttgcaaac gatcagatgg cgtggggcgc aatgcgcgc
5641 attaccgagt ccgggtgctg cgttggtgag gatctctcgg tagtgggata cgaagatccv
5701 gaagacagct catgttatat ccgcgcgtca accaccatca aacaggattt tgcctgctg
5761 gggcaaacca ggtgggccc cttgctgcaa ctctctcagg gccaggcggg gaagggcaat
5821 cagctggttg ccgtctcaat ggtgaaaaga aaacccaccc tggcgcccaa tacgcaacc
5881 gcctctcccc gcgcgttggc cgaattcatta atgcagctgg caagacagg tcccgactg
5941 gaaagcgggc agtyagcca acycaattaa tgtgagtag ccgaatcga tctg
```

FIG. 4, con't.

pTFC99a-YehF-FlaA8-6His (SEQ ID NO: 6) (SEQ ID NO: 5)

FEATURES

	Location/Qualifiers
source	1..4551 /organism="Cloning vector pTFC99A" /mol_type="genomic DNA" /db_xref="taxon:40992" /lab_host="Escherichia coli" /note="derived from pKK233-2"
misc_feature	1..17 /note="derived from cloning vector pBR322"
promoter	18..263 /note="trc promoter from pKK233-2" /citation={1}
misc_feature	264..266 /note="NcoI/BamHI linker"
misc_feature	734..1158 /note="5S rRNA, T1, T2, rrsB"
misc_feature	1159..3337 /note="derived from cloning vector pBR322"
misc_feature	3338..3345 /note="BglII linker"
misc_feature	3346..3348 /note="BglII linker"
misc_feature	3349..3350 /note="EcoRI linker"
misc_feature	3351..3381 /note="derived from plasmid BP4"
misc_feature	3382..4568 /note="lacI-q region"
misc_feature	4569..4575 /note="EcoRI linker"
misc_feature	4576..4581 /note="BglII linker"
misc_feature	681..698 /note="6XHis" (SEQ ID NO: 6)
misc_binding	645..680 /note="FlaA8"
gene	291..644 /note="YehF"

ORIGIN

```
1  gtttgacagc ttatcakaga ctgcacggtg caccaatgct tctggggtca ggcagccatc
61  ggaagctgtg gtatggctgt gcaggctgta aatcactgca taattcgtgt cgtcaaagga
121  gcactccggt tctgggtatg gtttcttgag ccgaatcatc aacgggttctg gcaaatcttc
181  tgaattgagc tghtyacaat taatcatcag gctcgtatca tctgtggaat tctgagcgga
241  taacaaattc acacaggaaa cagccatgag aattcgagct cgaataaaac atgaaaaaaa
301  gagggccttt tttagggctg ttggttggtt ctgctgctgc atcaattttt scctgccaata
361  atgaaacccg cagctcgctc actttcccaa agtctgaaag tctgcatgct cccgggattg
421  ccccgagcgt aaaaagctga tatcaacaaa atcccgctgc cccttgagca gatcatcaaa
481  aaattcttgc tcagcccgat cccgctgctt ccctcagctt cagcagacat cagcgttaac
541  atgataaatt ctacatagcc ctacacgctg ctgctgaaag tcccgatatt cattaccagg
601  tcagcctgca ctgcaaaagc ggaatgctgc aatattcagc ccgtttttctg aactgctgctc
661  cggcctgctc catggaaccc catcatcaac atcaccacta atctagagtc gactgaggg
721  catgcaagct tgggtgtttt ggaggatgag agaatgtttt cagcctgata cagattaat
781  cagaaagcag aagcgttctg ataatacaga atttgcctg cygcagttag cgggtggtcc
841  cacttgaccc catgcgaac taagaagtga aacgcgttag cygcagttag agtgtgggtt
901  ctcccatgca gagagtgagg aactgacagg catcaaatca aacgaaggcc tcactgaaa
961  gactgggctt tctgttttat ctgttggttg tgggtgaaag ctctcctgag taggacaaat
1021  ccgcctggag cggattttaa cgttgctgag caacgggccc ggggttggag ggcaggagcc
```

FIG. 4, con't.

1081 ccggccatase ctgcccaggca tcaaatthaag cagaaggcca tccatgaaggga tggcccttttt
1141 gggtttctctc aaactcttttt tgttttttttt tccaaatata ttcaaatatg tatccgctca
1201 tgagacaata accctgataa atgcttcaat aatattgaaa aagggaagagt atgagtattc
1261 aacattttccg tgcggccctt acttcccttttt ttggggcatt ttgcttctct gttttttgctc
1321 acccagaaac gctgggtgaaa gttaaagatg ctgaagatca gttgggtgaa cgaagtgggtt
1381 acatogaaat ggatctcaac agcggtaaga tccctgagag ttttggcccc gaagaaacgtt
1441 ttccaatgat gaggacttttt aaagtctctg taagtggcgc ggtatttatcc cgtgttgacg
1501 cggggcaaga gcaactcggg ccgcgcatac actattctca gaatgacttg gttgagtact
1561 ccccagucac agaaaaagcat cttaacggatg gcatagcagt aagagaatta tgcagtgcctg
1621 ccataaccat gagtataaac actcgggcca acttaattct gacaaagatc ggaggacoga
1681 eggagctaac cgtttttttt cacaaactgg ggatcatgt aactcgcctt tccgcttgctg
1741 aacgggagct gaatgaagcc ataccaaaag ccgagcgtga cccacagatg cctacagcaa
1801 tggcaaccaa gttgcgcaca ctattcaactg gcgaactact tactctagct tccgggcacg
1861 aattaataga ctggatggag ggggataaag ttgcaggacc acttctgcgc tccgcccctc
1921 cggctggctg gtttatttgt gataaatctg gagccggtga gcgtgggtct cgcggtatca
1981 ttgcagcaat ggggcccgat ggtaaagcct ccgctatcgt agttatctac aagacgggga
2041 gtccaggcaa tatggatgaa cgaatatagc agatcgtgga gatagggtgc tcaactgata
2101 agcatlgtta actgtccgac caagtttact catataact ttgattgaa ttaaaacttc
2161 atttttaatt taaaaggatc taggtgaaga tcccttttga taatctcagc accaaaatcc
2221 cttaacgtga gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt
2281 cttgagatcc tttttttctg ccgctaatct gctgcttga aacaaaaaaa ccccgctac
2341 cagcgggtgt ttgtttgcgc gatcaagagc taccaaactc ttttcgaag gtaactggct
2401 tccagcagcg gcagatccca aatactgtcc ttctagtga gcctagtta ggcacacact
2461 tcaagaactc tctagccgcg cctacatacc tgcctctgct aatctgtta ccagtgcctg
2521 ctgcagctgg cgataagctg tgtcttaccg ggttggactc aagacgatac ttaccggata
2581 eggcgcagcg gtcgggctga accgggggtt cgtgcacaca gccagcttg gagcgaacgc
2641 cctacaccca actgagctac ctacacgcgt agctatgaga aagcgcacag cttcccgaa
2701 ggagaaacga ggcacagtat ccggttaagc gcaggtcgg aacgggaag cgcacaggg
2761 agcttccagc tggtatcttt ttgtctctcag atagtctgt cyggtttcgc caactctgac
2821 ttgagcgtcg atttttgtga tgcctgtcag gggggcggag cctatggaaa aaagccagca
2881 aegcggcctt tttaacggtt ctggcctttt gctggccttt tgcctacatg ttctttctctg
2941 cgttatcccc tgattctgtg gataaacgta ctaccgctt tgagtgaact gataccgctc
3001 gccgcagcgc aacgacagc ccgcagcagt cagtgaagca ggaagcggaa ggcgcctga
3061 ttgcgttattt tctccttaag catctgtgcg gtatttcaaa cgcctatag cgcactctca
3121 gtacaactct ctctgatgac gcatagttaa gccagtatac actccgctat cgtactgtga
3181 ctgggtcctg gctgcgcgcc gacacccgvc aacacccgct gcgcgcctt ggcgggcttg
3241 tctgctcccg gcatccgctt accagacaagc tgcgacgctc tccgggagct gcctgtgcca
3301 gaggttttca ccgtcactca cgaacgcgcg gaggcgcag ctcaattcgc gcgcgaaggc
3361 gaagcggcat gcatctacgt tgacacvata gaatggtgca aaactttcgc cgttatggca
3421 tgatagcgc ccgaagagag tcaattcagg gtggtgaatg tgaaccagc aagttatca
3481 gatgtgcgc agtatgcgc tgtctcttat cagacgctt cccgcgtggt gaaccaggcc
3541 agccacgctt ctgcgaaaaa gcgggaaaaa gtggaagcgc cgtatggcga gctgaattac
3601 attcccaacc gctgggcacc acaactggcg ggcaaacagt cgttgcgtat tggcgttgc
3661 aactccagtc tggccttgca ccgcgcgtcg caaattgctg cgcgattaa atctgcgcgc
3721 gatcaactgg gtgcacagct tctcgcgcga atggtagaac gaagcggcgt cgaagcctgt
3781 aagcggcgc tgcacactct tctcgcgcga ccgctcagtg gctgatacat taactatccg
3841 ctggatgacc aggatgcct tgccttgaaa gctgcctgca ctaatgttcc gggttattt
3901 cttgatgtct ctgacccgac aacctacaa agtattattt tctccatga agcggtagc
3961 agactggcgc tggagcactc ggtcgcattg ggtcaccagc aaatcgcgt gttagcggc
4021 caattaagtt ctgtctcgcc gcgtctgcgt ctggctggct ggcataaaa tctcactcgc
4081 aatcaaatte agccgatagc ggaacgggaa ggccgctgga gtgcctatgc cgggtttcaa
4141 caaaccatgc aatatgtgaa tgaaggcata gttcccaact cgtatgctgt tgcacaagat
4201 cagatggcgc tgggcgcact gncgcvatt acccgctcgc ggtcgcgcgt tggctgggat
4261 atctcgttag cgtacacaga cgtacacaga gacagctcat gttatatcc ccgtcaacc
4321 acaatcaaac aggattttct cctgctgggg caaacccagc tggaccgctt gctgcaactc
4381 tctcagggcc aggcgggtga ggcacactcgt ctggtgcgc tctcactggt gaagagaaa
4441 acaacccctg cgcacactac gcaaacccgc tctcccccgc cgttggcaga tcaattaatg
4501 cagctggcac gacaggtttc ccgactggaa agcgggcagt ggcgcacac caattaatgt
4561 gagttagcgc gaattgatct g

1

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

CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims priority to U.S. Provisional Application Ser. No. 61/369,188, filed Jul. 30, 2010, the entire disclosure of which is 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.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Sep. 12, 2011, is named 32261227.txt and is 30,272 bytes in size.

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 secreted protein production. There are no well-understood secretory pathways in *E. coli* to transport

2

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, entC, entE, yebE, yihF, yebG, mzcA, ftsK, tnaA, 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, EntC, EntE, YebE, YihF, YebG, MzcA, FtsK, TnaA, 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, the NlpD and YihF 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 the 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 het-

erologous protein is selected from the group consisting of: a cellulase, a protease, a lipase, a cutinase, an amylase, a galactosidase, a pullulanase, a glucose isomerase, a protein disulphide isomerase, a cyclodextrin gluconotransferase, a phytase, a glucose oxidase, a glucosyl 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, MzrA, 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, YihF, YebG, MzrA, 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, MzrA, 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 Keio host strain indicates what gene has been deleted from strain BW25113 ΔdsbA. WT indicates the wild-type background (*E. coli* BW25113 ΔdsbA). The upper graph shows the relative fluorescence from FAsH-tagged YebF. Below that are the results of Western blots of secreted YebF-6×His-cellulase ("6×His" disclosed as SEQ ID NO: 6) fusion proteins.

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

FIG. 3 is a graph showing the relative fluorescence from FAsH-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, YebE, yihF, yebG, mzrA, ftsK, tnaA, ompC, and ompF or homologs thereof.

In practicing the present invention, many conventional techniques in molecular biology, protein biochemistry, cell biology, microbiology and recombinant DNA are used. These techniques are well-known and are explained in, e.g., *Current Protocols in Molecular Biology*, Vols. I-III, Ausubel, Ed. (1997); Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989); *DNA Cloning: A Practical Approach*, Vols. I and II, Glover, Ed. (1985); *Oligonucleotide Synthesis*, Gait, Ed. (1984); *Nucleic Acid Hybridization*, Hames & Higgins, Eds. (1985); *Transcription and Translation*, Hames & Higgins, Eds. (1984); Perbal, *A Practical Guide to Molecular Cloning*; the series, *Meth. Enzymol.*, (Academic Press, Inc., 1984); and *Meth. Enzymol.*, Vols. 154 and 155, Wu & Grossman, and Wu, Eds., respectively.

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 to the nucleic acid encoding the heterologous protein interest, to provide for a single amino acid chain when transcription and translation occur. 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., about 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 RNA 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 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 secretory pathway. In some embodiments, the mutation is in one or more genes selected from ompR, envZ, nlpD, entC, entE, yebE, yihF, yebG, mzcA, ftsK, tnaA, ompC, and ompF. In one embodiment, the geneti-

cally modified bacterium contains a single mutation in the ompR, envZ, nlpD, entC, entE, yebE, yihF, yebG, mzhA, ftsK, tnaA, 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, entC, entE, yebE, yihF, yebG, mzhA, 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, entC, entE, yebE, yihF, yebG, mzhA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in four genes selected from ompR, envZ, nlpD, entC, entE, yebE, yihF, yebG, mzhA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in five genes selected from ompR, envZ, nlpD, entC, entE, yebE, yihF, yebG, mzhA, ftsK, tnaA, ompC, and ompF. In one embodiment, the genetically modified bacterium contains mutations in the ompR, nlpD, entC, entE, yebE, and yihF genes.

In one embodiment, the host cell is a genetically-modified *Shigella*, *Yersinia*, *Salmonella* and *Escherichia* sp. bacteria 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, entC, entE, yebE, yihF, yebG, mzhA, 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, HMS174, 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), DH1 (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 recombinase 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 USA*. 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 include, 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 95% 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 upon the published sequences of the target gene(s), e.g., ompR, envZ, nlpD, entC, entE, yebE,

yihF, yebG, mzhA, 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 phage 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 to 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 polypeptide 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., YebF, SEC, TAT, pectate 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, "operably-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 to 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, TAT, 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:

(SEQ ID NO: 1)
 MKKRGAFLGLLLVSACASVFAANNETSKSVTFPKCEGLDAAGIAAS
 VKRDYQQNRVARWADDQKIVGQADPVAWVSLQDIQKDDKWSVPLT
 VRGKSADIHYQVSDCKAGMAEYQRR

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

(SEQ ID NO: 2)
 ATGAAAAAAGAGGGGCGTTTTTAGGGCTGTTGTTGGTTTCTGCCT
 GCGCATCAGTTTTTCGCTGCCAATAATGAAACCAGCAAGTCGGTCAC
 TTTCCCAAGTGTAAGATCTGGATGCTGCCGAATTCGCCGAGC
 GTAAAAACGTGATTATCAACAAAATCGCGTGGCGCGTTGGGCAGATG
 ATCAAAAAATTGTCGGTCAGGCCGATCCCGTGGCTTGGTCAGTTT
 GCAGGACATT CAGGGTAAAGATGATAAATGGTCAGTACCGCTAAC
 GTGCGTGGTAAAAGTGCCGATATTCATTACCAAGTCAGCGTGACT
 GCAAAGCGGAATGGCGGAATATCAGCGGCGTTAA

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. Nos. 5,047,334, 4,963,495.). 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 (PelB) from *Erwinia carotovora*; Outer-membrane protein A (OmpA); Heat-stable enterotoxin 2 (StII); Endoxylanase (Endo) from *Bacillus* sp.; Alkaline phosphatase (PhoA); Outer-membrane pore protein F (OmpF); Outer-membrane pore protein E (PhoE); Maltose-binding protein (MalE); Outer-membrane protein C (OmpC); Murein lipoprotein (Lpp); Lamba 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., Gottesman, *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 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 cost 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 cellulosic biofuels. Cellulosic biofuels are produced using secreted enzyme complexes including cellulases and xylanases. The cellulosic 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-Mediating 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, yebE, 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 USA. 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 6×His 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);

13

(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 screen 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. Cel5B).

Table 1 and FIG. 1 shows the result of FIAsh 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 1

1° Screening Score	2° Screening	Locus	Description
9.4	+++	envZ/ompR	2-component osmolarity regulator
12.2	++	nlpD	Novel lipoprotein, function unknown
8.9	+	mzrA	Modulator of EnvZ/OmpR operon
10.8	++	ftsK	DNA translocase at septal ring sorting daughter chromosome
6.2	+	tnaA	Tryptophanase
27.6	+	entC/E	Isochorismate synth I & comp of enterobactin synth cmplx
6.6	0	yihF	Conserved protein, DUF945 family
N/A	N/A	yebE	Inner membrane protein

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]. Plated 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 FIAsh/DTT/BAL cocktail (21 µM FIAsh-EDT, 21 mM DTT, and 5.25 mM 2,3-dimercaptopropanol) for a final concentration of 1 µM FIAsh-EDT, 1 mM DTT, and 250

14

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 the 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 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.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 9

<210> SEQ ID NO 1

<211> LENGTH: 118

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 1

Met Lys Lys Arg Gly Ala Phe Leu Gly Leu Leu Val Ser Ala Cys
 1 5 10 15

Ala Ser Val Phe Ala Ala Asn Asn Glu Thr Ser Lys Ser Val Thr Phe

-continued

20	25	30	
Pro Lys Cys Glu Gly Leu Asp Ala	Ala Gly Ile Ala Ala	Ser Val Lys	
35	40	45	
Arg Asp Tyr Gln Gln Asn Arg Val	Ala Arg Trp Ala Asp	Asp Gln Lys	
50	55	60	
Ile Val Gly Gln Ala Asp Pro Val	Ala Trp Val Ser Leu Gln Asp	Ile	
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 Lys Ala Gly	Met	
100	105	110	
Ala Glu Tyr Gln Arg Arg			
115			
<210> SEQ ID NO 2			
<211> LENGTH: 357			
<212> TYPE: DNA			
<213> ORGANISM: Artificial Sequence			
<220> FEATURE:			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide			
<400> SEQUENCE: 2			
atgaaaaaaa gagggcgctt tttagggctg ttgttggttt ctgcctgcgc atcagttttc	60		
gctgccaata atgaaccag caagtcggtc actttcccaa agtgtgaaga tctggatgct	120		
gccggaattg ccgcgagcgt aaaacgtgat tatcaacaaa atcgcgtggc gcgttgggca	180		
gatgatcaaa aaattgtcgg tcaggccgat cccgtggctt gggtcagttt gcaggacatt	240		
cagggtaaag atgataaatg gtcagtaccg ctaaccgtgc gtggtaaaag tgccgatatt	300		
cattaccagg tcagcgtgga ctgcaaagcg ggaatggcgg aatatcagcg gcgttaa	357		
<210> SEQ ID NO 3			
<211> LENGTH: 6120			
<212> TYPE: DNA			
<213> ORGANISM: Artificial Sequence			
<220> FEATURE:			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide			
<400> SEQUENCE: 3			
gtttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgta ggcagccatc	60		
ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc	120		
gcactcccg tctggataat gttttttgcg ccgacatcat aacggttctg gcaaattatc	180		
tgaaatgagc tgttgacaat taatcatccg gtcgtataa tgtgtggaat tgtgagcgga	240		
taacaatttc acacaggaag cagaccatgg aattcgagct cgagaaaaac atgaaaaaaa	300		
gagggggcgt tttagggctg ttgttggttt ctgcctgcgc atcagttttc gctgccaata	360		
atgaaccag caagtcggtc actttcccaa agtgtgaaga tctggatgct gccggaattg	420		
ccgcgagcgt aaaacgtgat tatcaacaaa atcgcgtggc gcgttgggca gatgatcaaa	480		
aaattgtcgg tcaggccgat cccgtggctt gggtcagttt gcaggacatt cagggtaaag	540		
atgataaatg gtcagtaccg ctaaccgtgc gtggtaaaag tgccgatatt cattaccagg	600		
tcagcgtgga ctgcaaagcg ggaatggcgg aatatcagcg gcgttctaga gatgtcgccc	660		
cattgagcgt gcaaggcaac aagatcctgg cgaatggtca gccggcgagc ttcagcggtg	720		
tgagcctgtt ttggagcaat accgagtggg gtggcgagaa gtactataac gcgcaagtgt	780		

-continued

tttcttggtt gaaatcggtt tggaacgcca agctgggtccg cgcagcgatg ggtgttgagg	840
atgaaggcgg ttacctgacc gaccgcgcga ataaggatcg cgtgactcaa gtggtggacg	900
cagcgatcgc aaacgacatg tacgtgatca tcgactggca tagccataat gcacaccaat	960
atcagttctca ggccatcgcc ttctttcagg agatggctcg caagtatggt gcgaacaacc	1020
acgtgatcta tgaaatctac aatgagcctt tgcaggtgag ctggtctaac actatcaaac	1080
cgtatgcgca agcggtgatt gcggcgatcc gtgcgattga cccagacaat ctgattatcg	1140
tgggtacgcc gacctggagc caggatgtcg acgtcgcggc gaatgaccgg attacgggtt	1200
accagaacat tgcgtatacc ctgcatttct atgcgggtac gcacgggtcaa tacctgcgtg	1260
ataaggcaca gaccgcactg aatcgtggca ttgctctgtt tgtcaccgaa tgggggtcgg	1320
ttaatgcgaa tgggtgatgc gccgttgcta atagcgaaac caatgcttgg gtgagcttta	1380
tgaaaaccaa tcacatctcc aacgcgaact gggcactgaa tgacaaagtt gagggcgcaa	1440
gcgcattggt cccgggtgcc agcgcgaacg gcggctgggt taacagccaa ttgaccgcgt	1500
ccggcgctct ggccaaaagc atcatcagcg gctggccgag ctacaatacc agctccagca	1560
gcagcgcggt ttccagccag acgcaagtca gcagctcgtc ccaagcccg gtcgtgtcta	1620
gctctagcag caccgctcg agcgtgggta gctccgctgt cagcggccaa cagtgttaact	1680
ggtatggtac gttgtatcca ctgtgcagca cgaccacgaa cgggtggggg tgggaaaaca	1740
acgcgtcgtg cattgcgcgt gcaacgtgca gcggtcagcc ggcaccgtgg ggtatcgtcg	1800
gcggtagcac cagcagccaa gcgtccctca gcgtccgcag cagcagcagc tctctggtca	1860
gctccagccg tagcagcagc agcagctctg ttcagtctag cagcgcgcct tcgtcggtgg	1920
cgagcagcag cggcagcagc agcggccagt gcagctacac cgttaccaat cagtggagca	1980
acgggtttac cgcattccat cgtattcgca acaatggcac cagcccgatc aacgggtgga	2040
atctgagctg gagctactct gacggtagcc gtgttaccac ttcttggaac gcgaatgtgt	2100
ctggcaataa cccatacacc gcattctaacc tgggttgga tggcagcatt caaccgggtc	2160
aagctgtgga gtttggtttt caggggacca agaataacag cgctgcggct atcccagccc	2220
tgagcggcaa cgtgtgcaac aaccatcacc accatcacca ctaaaagctt ggctgttttg	2280
gcggatgaga gaagattttc agcctgatac agattaaatc agaacgcaga agcggctctga	2340
taaaacagaa tttgcctggc ggcagtagcg cgggtgtccc acctgacccc atgccgaact	2400
cagaagtgaac acgccgtagc gccgatggta gtgtggggtc tccccatcg agagtaggga	2460
actgccagcg atcaaataaa acgaaaggct cagtcgaaag actgggcctt tcgtttttatc	2520
tgtgtgttgt cgggtgaacg tctcctgagt aggacaaatc cgccgggagc ggatttgaac	2580
gttgcaagc aacggcccg aggggtggcg gcaggacgcc cgccataaac tgccaggcat	2640
caaattaagc agaaggccat cctgacggat ggcctttttg cgtttctaca aactcttttt	2700
gtttattttt ctaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa	2760
tgcttcaata atattgaaaa aggaagagta tgagtattca acatttccgt gtccgccctta	2820
ttcccttttt tgccgcatth tgccttccgt tttttgctca cccagaaaag ctggtgaaag	2880
taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca	2940
gcggtaagat ccttgagagt tttcgcccc aagaacgttt tccaatgatg agcactttta	3000
aagttctgct atgtggcgcg gtattatccc gtgttgacgc cgggcaagag caactcggtc	3060
gcgcataca ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcatc	3120

-continued

ttacggatgg	catgacagta	agagaattat	gcagtgcctgc	cataacccatg	agtgataaca	3180
ctgcggccaa	cttacttctg	acaacgatcg	gaggaccgaa	ggagctaacc	gcttttttgc	3240
acaacatggg	ggatcatgta	actcgccttg	atcgttggga	accggagctg	aatgaagcca	3300
taccaaacga	cgagcgtgac	accacgatgc	ctacagcaat	ggcaacaacg	ttgcgcaaac	3360
tattaactgg	cgaactactt	actctagctt	cccggcaaca	attaatagac	tggatggagg	3420
cggataaagt	tgcaggacca	cttctgcgct	cggcccttcc	ggctggctgg	tttattgctg	3480
ataaatctgg	agccgggtgag	cgtgggtctc	gcggtatcat	tgcagcactg	gggccagatg	3540
gtaagccctc	ccgtatcgta	gttatctaca	cgacggggag	tcaggcaact	atggatgaac	3600
gaaatagaca	gatcgctgag	atagggtcct	cactgattaa	gcattggtaa	ctgtcagacc	3660
aagtttactc	atatatactt	tagattgatt	taaaacttca	tttttaattt	aaaaggatct	3720
aggtgaagat	cctttttgat	aatctcatga	ccaaaatccc	ttaacgtgag	ttttcgttcc	3780
actgagcgtc	agaccccgta	gaaaagatca	aaggatcttc	ttgagatcct	ttttttctgc	3840
gcgtaatctg	ctgcttgcaa	acaaaaaac	caccgctacc	agcggtggtt	tgtttgccgg	3900
atcaagagct	accaactctt	tttccgaagg	taactggctt	cagcagagcg	cagataccaa	3960
atactgtcct	tctagtgtag	ccgtagttag	gccaccactt	caagaactct	gtagcaccgc	4020
ctacatacct	cgctctgcta	atcctgttac	cagtggctgc	tgccagtggc	gataagtcgt	4080
gtcttaccgg	gttggaactca	agacgatagt	taccggataa	ggcgagcg	tcgggctgaa	4140
cgggggggtc	gtgcacacag	cccagcttgg	agcgaacgac	ctacaccgaa	ctgagatacc	4200
tacagcgtga	gctatgagaa	agcgcacgc	ttcccgaagg	gagaaaggcg	gacaggtatc	4260
cggtaagcgg	cagggctcga	acaggagagc	gcacgaggga	gcttccaggg	ggaaacgcct	4320
ggatatcttta	tagtctctgc	gggtttcgcc	acctctgact	tgagcgtcga	ttttgtgat	4380
gctcgtcagg	ggggcggagc	ctatggaaaa	acgccagcaa	cgcggccttt	ttacggttcc	4440
tggccttttg	ctggcctttt	gtccacatgt	tctttcctgc	gttatccctt	gattctgtgg	4500
ataaccgtat	taccgccttt	gagtgagctg	ataccgctcg	ccgcagccga	acgaccgagc	4560
gcagcgagtc	agtgagcgag	gaagcggaag	agcgcctgat	gcggtatttt	ctccttacgc	4620
atctgtgcgg	tatttcacac	cgcataatgt	gcactctcag	tacaatctgc	tctgatgccg	4680
catagttaag	ccagtataca	ctccgctatc	gctacgtgac	tgggtcatgg	ctgcgccccg	4740
acaccgcga	acaccgcgtg	acgcgccttg	acgggcttgt	ctgctcccg	catccgctta	4800
cagacaagct	gtgaccgtct	cggggagctg	catgtgtcag	aggtttttac	cgtcatcacc	4860
gaaacgcgcg	aggcagcaga	tcaattcgcg	cgcgaaggcg	aagcggcatg	catttacgtt	4920
gacaccatcg	aatgggtgcaa	aacctttcgc	ggtatggcat	gatagcgccc	ggaagagagt	4980
caattcaggg	tgggtaatgt	gaaaccagta	acgttatatg	atgtcgcaga	gtatgccggg	5040
gtctcttate	agaccgtttc	cgcgctgggt	aaccaggcca	gccacgtttc	tgcgaaaacg	5100
cgggaaaaag	tggaaagcgc	gatggcgag	ctgaattaca	ttcccaaccg	cgtggcacia	5160
caactggcgg	gcaaacagtc	gttctgtgatt	ggcgttgcca	cctccagtct	ggccctgcac	5220
gcgcgctcgc	aaattgtcgc	ggcgattaaa	tctcgcgcgg	atcaactggg	tgccagcgtg	5280
gtggtgtcga	tggtagaacg	aagcggcgtc	gaagcctgta	aagcggcggg	gcacaatctt	5340
ctcgcgcaac	gcgtcagtgg	gctgatcatt	aactatccgc	tggatgacca	ggatgccatt	5400
gctgtggaag	ctgcctgcac	taatgttccg	gcgttatctc	ttgatgtctc	tgaccagaca	5460
cccatcaaca	gtattatctt	ctcccatgaa	gacggtacgc	gactgggcgt	ggagcatctg	5520

-continued

```

gtgcattg gtcaccagca aatcgcgctg ttagcgggco cattaagtcc tgtctcgcg 5580
cgtctcgctc tggtcggtg gcataaatat ctcaactcga atcaaatcca gccgatagcg 5640
gaacgggaag gcgactggag tgccatgtcc ggttttcaac aaaccatgca aatgctgaat 5700
gagggcatcg ttcccaactgc gatgctgggt gccaacgac agatggcgct gggcgcaatg 5760
cgcgccatta ccgagtcgag gctgcgcggt ggtgcggata tctcggtagt gggatacgac 5820
gataccgaag acagctcatg ttatatcccg ccgtcaacca ccatcaaaca ggattttcgc 5880
ctgctggggc aaaccagcgt ggaccgcttg ctgcaactct ctcaggggcca ggcgggtgaag 5940
ggcaatcagc tgttgcccgct ctcaactggtg aaaagaaaaa ccacctggc gcccaatacg 6000
caaacgcct ctccccgcgc gttggccgat tcattaatgc agctggcagc acagggttcc 6060
cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg agttagcgcg aattgatctg 6120

```

<210> SEQ ID NO 4

<211> LENGTH: 5994

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 4

```

gtttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgcca ggcagccatc 60
ggaagctgtg gtatggctgt gcaggctgta aatcaactgca taattcgtgt cgctcaaggc 120
gcactcccg tctggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc 180
tgaaatgagc tgttgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcgga 240
taacaatttc acacaggaaa cagaccatgg aattcgagct cgagaaaaac atgaaaaaaa 300
gaggggcggt tttagggctg ttgttggttt ctgcctgcgc atcagttttc gctgccaata 360
atgaaaccag caagtcggtc actttcccaa agtgtgaaga tctggatgct gccggaattg 420
ccgcgagcgt aaaacgtgat tatcaacaaa atcgcgtagc gcgttgggca gatgatcaaa 480
aaattgtcgg tcaggccgat cccgtggcct gggtcagttt gcaggacatt cagggtaaag 540
atgataaatg gtcagtaccg ctaaccgtgc gtggtaaaaag tgccgatatt cattaccagg 600
tcagcgtgga ctgcaaagcg ggaatggcgg aatatcagcg gcgttttctg aactgctgcc 660
cgggctgctg catggaaccg catcatcacc atcaccacta atctagagtc gacctgcagg 720
catgcaagct tggctgtttt ggcggatgag agaagatttt cagcctgata cagattaaat 780
cagaacgcag aagcggctctg ataaaacaga atttgcttgg cggcagtagc gcggtggtcc 840
cacctgaccc catgccgaac tcagaagtga aacgccgtag cgcgatggt agtgtggggt 900
ctccccatgc gagagtaggg aactgccagg catcaataa aacgaaaggc tcagtcgaaa 960
gactgggctc ttctgtttat ctgttgtttg tcggtgaacg ctctcctgag taggacaaat 1020
ccgccgggag cggatttgaa cgttgcaag caacggcccg gaggtggcg ggcaggacgc 1080
ccgccataaa ctgccaggca tcaaattaag cagaaggcca tctgacgga tggccttttt 1140
gcgtttctac aaactctttt tgttattttt tctaaataca ttcaaatatg tatccgctca 1200
tgagacaata accctgataa atgcttcaat aatattgaaa aggaagagt atgagtattc 1260
aacatttccg tgcgccectt attccctttt ttgcggcatt ttgccttct gtttttgctc 1320
accagaaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt 1380
acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt 1440

```


-continued

ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	ggtattatcc	cgtgttgacg	1500
ccgggcaaga	gcaactcggg	cgccgcatac	actattctca	gaatgacttg	gttgagtact	1560
caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	aagagaatta	tgcatgtctg	1620
ccataacccat	gagtgataac	actgcgccca	acttacttct	gacaacgacg	ggaggaccga	1680
aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	aactcgctt	gatcgttggg	1740
aaccggagct	gaatgaagcc	ataccaaaacg	acgagcgtga	caccacgatg	cctacagcaa	1800
tggcaacaac	gttgcgtaag	aggttccaac	tttcaccata	atgaaataag	atcactaccg	1860
ggcgtatttt	ttgagttatc	gagattttca	ggagctaagg	aagctaaaat	ggagaaaaaa	1920
atcactggat	ataccaccgt	tgatatatcc	caatggcatc	gtaaagaaca	ttttgaggca	1980
tttcagtcag	ttgctcaatg	tacctataac	cagaccgttc	agctggatat	tacggccttt	2040
ttaaagaccg	taaagaaaaa	taagcacaag	ttttatccgg	cctttattca	cattcttgcc	2100
cgctgatga	atgctcatcc	ggaattccgt	atggcaatga	aagacgggtga	gctggtgata	2160
tgggatagtg	ttcacccttg	ttacaccgtt	ttccatgagc	aaactgaaac	gttttcatcg	2220
ctctggagtg	aataccacga	cgattttccg	cagtttctac	acatatattc	gcaagatgtg	2280
gcgtgttacg	gtgaaaacct	ggcctatttc	cctaaagggt	ttattgagaa	tatgtttttc	2340
gtctcagcca	atccctgggt	gagtttcacc	agttttgatt	taaacgtggc	caatatggac	2400
aaattcttcg	cccccgtttt	caccatgggc	aaatattata	cgcaaggcga	caagggtgctg	2460
atgcccgctg	cgattcaggt	tcatcatgcc	gtctgtgatg	gcttccatgt	cggcagaaatg	2520
cttaatgaat	tacaacagta	ctgcgatgag	tggcagggcg	gggcgtaatt	tttttaaggc	2580
agttattggt	gcccttaaac	gcctgggtgt	acgcctgaat	aagtgataat	aagcggatga	2640
atggcagaaa	ttcgaaaaca	aattcgaccc	ggtcgctcgt	tcagggcagg	gtcgttaaat	2700
agccgcttat	gtctattgct	ggtttaccgg	tttattgact	accggaagca	gtgtgaccgt	2760
gtgcttctca	aatgcctgag	gccagtttgc	tcaggctctc	cccgtggagg	taataattga	2820
cgatatgatc	atttattctg	cctcccagag	cctgataaaa	acggttagcg	cttcgttaat	2880
acagatgtag	gtgttcacac	gggtagccag	cagcatcctg	cgatgcagat	ccggaacata	2940
atgggtgcagg	gcgcttggtt	cggcgtgggt	atgggtggcag	gccccgtggc	cgggggactg	3000
ttgggcgctg	ccggcacctg	tcctacgagt	tgcataataa	agaagacagt	cataagtgcg	3060
gcgacgatat	tcctgccccg	cgccccaccg	aaggagctac	cgacacagcg	tgccggaactg	3120
tgtaactcag	aataagaaat	gaggccgctc	atggcggtga	ctctcagtea	tagtatcgtg	3180
gtatcacccg	ttggttccac	tctctgttgc	gggcaacttc	agcagcacgc	aaactattaa	3240
ctggcgaaat	acttactcta	gcttcccggc	aacaattaat	agactggatg	gaggcggata	3300
aaagttgcagg	accacttctg	cgtcgggcc	ttccggctgg	ctggtttatt	gctgataaat	3360
ctggagccgg	tgagcgtggg	tctcgcggtg	tcattgcagc	actggggcca	gatggtaagc	3420
cctcccgat	cgtagttatc	tacacgacgg	ggagtcaggc	aactatggat	gaacgaaata	3480
gacagatcgc	tgagataggt	gcctcactga	ttaagcattg	gtaactgtca	gaccaagttt	3540
actcatatat	actttagatt	gatttaaaac	ttcattttta	atttaaaagg	atctaggtga	3600
agatcccttt	tgataatctc	atgacaaaaa	tcctttaacg	tgagttttcg	ttccactgag	3660
cgtcagaccc	cgtagaaaag	atcaaaggat	cttcttgaga	tccttttttt	ctgcgcgtaa	3720
tctgctgctt	gcaaacaaaa	aaaccaccgc	taccagcggg	ggtttggttg	ccgatcaag	3780

-continued

agctaccaac tctttttccg aaggttaactg gcttcagcag agcgcagata ccaaatactg	3840
tccttctagt gtagccgtag ttaggccacc acttcaagaa ctctgtagca ccgcctacat	3900
acctcgctct gctaactctg ttaccagtgg ctgctgccag tggcgataag tcgtgtctta	3960
ccgggttgga ctcaagacga tagttaccgg ataaggcgca gcggtcgggc tgaacggggg	4020
gttcgtgcac acagcccagc ttggagcgaa cgacctacac cgaactgaga tacctacagc	4080
gtgagctatg agaaagcgcc acgcttcccg aaggagagaa ggcgacagg tatccggtaa	4140
gcggcagggg cggaacagga gagcgacga gggagcttcc agggggaaac gcctggatc	4200
tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt	4260
cagggggggc gagcctatgg aaaaacgcca gcaacgggc ctttttacgg ttcttgccct	4320
tttgctggcc ttttgctcac atgttctttc ctgcgttacc cctgattct gtggataacc	4380
gtattaccgc ctttgagtga gctgataccg ctgcgcgcag ccgaacgacc gagcgacgcg	4440
agtcagttag cgaggaagcg gaagagcgcc tgatgcggtt ttttctcctt acgcatctgt	4500
gcggtatttc acaccgcata tggtgcactc tcagtacaat ctgctctgat gccgcatagt	4560
taagccagta tacactccgc tatcgctacg tgactgggtc atggctgcgc cccgacccc	4620
gccaacaccc gctgacgcgc cctgacgggc ttgtctgctc ccggcatccg cttacagaca	4680
agctgtgacc gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg	4740
cgcgaggcag cagatcaatt cgcgcgcgaa ggcgaaagcg catgcattta cgttgacacc	4800
atcgaatggt gcaaaacctt tcgcggtatg gcatgatagc gcccggaaga gagtcaattc	4860
agggtggtga atgtgaaacc agtaacgtta tacgatgtcg cagagtatgc cgggtgtctct	4920
tatcagaccg tttcccgcgt ggtgaaccag gccagccacg tttctgcgaa aacgcgggaa	4980
aaagtggaag cggcgatggc ggagctgaat tacattccca accgcgtggc acaacaactg	5040
gcgggcgaac agtcgttgct gattggcggt gccacctcca gtctggccct gcacgcgcgc	5100
tcgcaaattg tcgcggcgat taaatctcgc gccgatcaac tgggtgccag cgtggtggtg	5160
tcgatggtag aacgaagcgg cgtcgaagcc tgtaaagcgg cgggtgcaca tcttctcgcg	5220
caacgcgtca gtgggctgat cattaactat ccgctggatg accaggatgc cattgctgtg	5280
gaagctgctt gcactaatgt tccggcggtt tttcttgatg tctctgacca gaccccatc	5340
aacagtatta ttttctccca tgaagacggt acgcgactgg gcgtggagca tctggctgca	5400
ttgggtcacc agcaaatcgc gctgttagcg ggcccattaa gttctgtctc ggcgctctg	5460
cgtctggctg gctggcataa atatctcact cgcaatcaaa ttcagccgat agcggaacgg	5520
gaaggcgact ggagtgccat gtccgggttt caacaaacca tgcaaatgct gaatgagggc	5580
atcgttccca ctgcgatgct ggttgccaac gatcagatgg cgctgggcgc aatgcgcgc	5640
attaccgagt ccgggctcgc cgttggtcgc gatatctcgg tagtgggata cgacgatacc	5700
gaagacagct catgttatat ccgcgcgtca accaccatca aacaggattt tcgcctgctg	5760
gggcaaacca gcgtggaccg cttgctgcaa ctctctcagg gccaggcggg gaagggcaat	5820
cagctgttgc ccgtctcact ggtgaaaaga aaaaccaccc tggcgcccaa tacgcaaac	5880
gcctctcccc gcgcgttggc cgattcatta atgcagctgg caccagaggt ttcccactg	5940
gaaagcgggc agtgagcgca acgcaattaa tgtgagttag cgcaattga tctg	5994

<210> SEQ ID NO 5

<211> LENGTH: 4581

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 5

```

gtttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgtea ggcagccatc      60
ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc      120
gcactcccggt tctggataat gttttttgcg ccgacatcat aacggttctg gcaaatattc      180
tgaaatgagc tgttgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcgga      240
taacaatttc acacagggaaa cagaccatgg aattcgagct cgagaaaaac atgaaaaaaa      300
gagggggcgtt tttagggctg ttgttggttt ctgcctgcgc atcagttttc gctgccaaata      360
atgaaaccag caagtcggtc actttcccaa agtgtgaaga tctggatgct gccggaattg      420
ccgcgagcgt aaaacgtgat tatcaacaaa atcgcgtagc gcgttgggca gatgatcaaa      480
aaattgtcgg tcaggccgat cccgtggcct gggtcagttt gcaggacatt cagggtaaag      540
atgataaatg gtcagtaccg ctaaccgtgc gtggtaaaag tgccgatatt cattaccagg      600
tcagcgtgga ctgcaaacg ggaatggcgg aatatcagcg gcgttttctg aactgctgcc      660
cgggctgctg catggaaccg catcatcacc atcaccacta atctagagtc gacctgcagg      720
catgcaagct tggctgtttt ggcgatgag agaagatttt cagcctgata cagattaaat      780
cagaacgcag aagcggctctg ataaaacaga atttgctctg cggcagtagc gcggtgggtcc      840
cacctgaccc catgccgaac tcagaagtga aacgcgtag cgccgatggt agtgtggggt      900
ctccccatgc gagagtaggg aactgccagg catcaaataa aacgaaaggc tcagtcgaaa      960
gactgggcct ttcgttttat ctgttggttg tcggtgaacg ctctcctgag taggacaaat     1020
ccgccgggag cggatttgaa cgttgcaag caacggcccg gaggtggcg ggcaggacgc     1080
ccgccataaa ctgccaggca tcaaattaag cagaaggcca tcctgacgga tggccttttt     1140
gcgtttctac aaactctttt tgttattttt tctaaataca ttcaaatatg tatccgctca     1200
tgagacaata accctgataa atgcttcaat aatattgaaa aagggaagagt atgagtattc     1260
aacatttccg tgcgcctctt attccctttt ttgcggcatt ttgccttcct gtttttgcct     1320
acccagaaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtggggt     1380
acatcgaaat ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt     1440
ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtgttgacg     1500
ccgggcaaga gcaactcggg cgcgcgcatc actattctca gaatgacttg gttgagtact     1560
caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgtctg     1620
ccataacccat gagtgataac actgcggcca acttacttct gacaacgacg ggaggaccga     1680
aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg     1740
aaccggagct gaatgaagcc ataccaaaac acgagcgtga caccacgatg cctacagcaa     1800
tggaacaacac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac     1860
aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc     1920
cggctggctg gtttatgtct gataaatctg gagccggtga gcgtgggtct cgcggtatca     1980
ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga     2040
gtcaggcaac tatggatgaa cgaaatagac agatcgctga gatagggtgc tcaactgatta     2100
agcattggta actgtcagac caagtttact catatatact ttagattgat ttaaaaacttc     2160
atttttaatt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaaatcc     2220

```

-continued

cttaacgtga gttttcggtc cactgagcgt cagaccccggt agaaaagatc aaaggatctt	2280
cttgagatcc tttttttctg cgcgtaactct gctgcttgca aacaaaaaaa ccaccgctac	2340
cagcgggtggt ttgtttgccc gatcaagagc taccaactct ttttccgaag gtaactggct	2400
tcagcagagc gcagatacca aatactgtcc ttctagtgtg gccgtagtta ggccaccact	2460
tcaagaactc tgtagcaccc cctacatacc tcgctctgct aatcctgtta ccagtggctg	2520
ctgccagtgg cgataagtcg tgtcttaccg ggttggaactc aagacgatag ttaccggata	2580
agggcgagcg gtcgggctga acgggggggt cgtgcacaca gccagcttg gagcgaacga	2640
cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag	2700
ggagaaggc ggacaggtat ccggaagcg gcagggtcgg aacaggagag cgcacgaggg	2760
agcttccagg gggaaacgcc tggatatctt atagtctgtt cgggtttcgc cactctgac	2820
ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca	2880
acgcggcctt tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcctg	2940
cgttatcccc tgattctgtg gataaccgta ttaccgcctt tgagtgaagt gataccgctc	3000
gccgcagccg aacgaccgag cgcagcgagt cagtgaagca ggaagcggaa gagcgctga	3060
tgcggtatct tctccttacg catctgtcgg gtatttcaca ccgcatatgg tgcactctca	3120
gtacaactcg ctctgatgcc gcatagttaa gccagtatac actccgctat cgctacgtga	3180
ctgggtcatg gctgcgcccc gacaccgccc aacaccgctt gacgcgccct gacgggcttg	3240
tctgctcccc gcataccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca	3300
gagggttttca ccgtcatcac cgaacacgagc gaggcagcag atcaattcgc gcgcgaaggc	3360
gaagcggcat gcatttacgt tgacaccatc gaatggtgca aaacctttcg cggtatggca	3420
tgatagcgcc cggaagagag tcaattcagg gtgggtgaatg tgaaaccagt aacgttatac	3480
gatgtcgcag agtatgcggg tgtctcttat cagaccgttt cccgcgtggt gaaccaggcc	3540
agccacgttt ctgcgaaaac gcgggaaaaa gtggaagcgg cgatggcgga gctgaattac	3600
attcccaacc cgtgtggcaca acaactggcg ggcaaacagt cgttgctgat tggcggtgcc	3660
acctccagtc tggccctgca cgcgcgctcg caaattgtcg cggcgattaa atctcgcgcc	3720
gatcaactgg gtgccagcgt ggtggtgtcg atggtagaac gaagcggcgt cgaagcctgt	3780
aaagcggcgg tgcacaatct tctcgcgcaa cgcgtcagtg ggctgatcat taactatccg	3840
ctggatgacc aggatgccat tgctgtggaa gctgcctgca ctaatgttcc ggcgttatct	3900
cttgatgtct ctgaccagac acccatcaac agtattatct tctcccatga agacggtagc	3960
cgactgggcg tggagcatct ggtcgcattg ggtcaccagc aaatcgcgct gttagcgggc	4020
ccattaagtt ctgtctcgcc gcgtctcgct ctggctggct ggcatataa tctcactcgc	4080
aatcaaatc agccgatagc ggaacgggaa gccgactgga gtgccatgtc cggttttcaa	4140
caaaccatgc aaatgctgaa tgagggcatc gttcccaact cgatgctggt tgccaacgat	4200
cagatggcgc tgggcgcaat gcgcgcattt accgagtcgg ggctgcgcgt tgggtcggat	4260
atctcggtag tgggatacga cgataccgaa gacagctcat gttatatccc gccgtcaacc	4320
accatcaaac aggtttttcg cctgctgggg caaacacgagc tggaccgctt gctgcaactc	4380
tctcagggcc agggcgtgaa gggcaatcag ctgttgcccg tctcactggg gaaaagaaaa	4440
accacctgg cgcccaatc gcaaacgcc tctccccgcg cgttgggcga ttcattaatg	4500
cagctggcac gacaggtttc ccgactggaa agcgggcagt gagcgcaacg caattaatgt	4560

-continued

gagtttagcgc gaattgatct g

4581

<210> SEQ ID NO 6
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 6xHis tag

<400> SEQUENCE: 6

His His His His His His
 1 5

<210> SEQ ID NO 7
 <211> LENGTH: 6
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 peptide

<400> SEQUENCE: 7

Cys Cys Pro Gly Cys Cys
 1 5

<210> SEQ ID NO 8
 <211> LENGTH: 286
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 polypeptide

<400> SEQUENCE: 8

Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
 1 5 10 15

Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys
 20 25 30

Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
 35 40 45

Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
 50 55 60

Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser
 65 70 75 80

Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
 85 90 95

Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr
 100 105 110

Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser
 115 120 125

Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys
 130 135 140

Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
 145 150 155 160

Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
 165 170 175

Asp Thr Thr Met Pro Thr Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
 180 185 190

Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
 195 200 205

-continued

Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
 210 215 220

Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser
 225 230 235 240

Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
 245 250 255

Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
 260 265 270

Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
 275 280 285

<210> SEQ ID NO 9
 <211> LENGTH: 319
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 polypeptide

<400> SEQUENCE: 9

Met Ala Glu Leu Asn Tyr Ile Pro Asn Arg Val Ala Gln Gln Leu Ala
 1 5 10 15

Gly Lys Gln Ser Leu Leu Ile Gly Val Ala Thr Ser Ser Leu Ala Leu
 20 25 30

His Ala Pro Ser Gln Ile Val Ala Ala Ile Lys Ser Arg Ala Asp Gln
 35 40 45

Leu Gly Ala Ser Val Val Val Ser Met Val Glu Arg Ser Gly Val Glu
 50 55 60

Ala Cys Lys Ala Ala Val His Asn Leu Leu Ala Gln Arg Val Ser Gly
 65 70 75 80

Leu Ile Ile Asn Tyr Pro Leu Asp Asp Gln Asp Ala Ile Ala Val Glu
 85 90 95

Ala Ala Cys Thr Asn Val Pro Ala Leu Phe Leu Asp Val Ser Asp Gln
 100 105 110

Thr Pro Ile Asn Ser Ile Ile Phe Ser His Glu Asp Gly Thr Arg Leu
 115 120 125

Gly Val Glu His Leu Val Ala Leu Gly His Gln Gln Ile Ala Leu Leu
 130 135 140

Ala Gly Pro Leu Ser Ser Val Ser Ala Arg Leu Arg Leu Ala Gly Trp
 145 150 155 160

His Lys Tyr Leu Thr Arg Asn Gln Ile Gln Pro Ile Ala Glu Arg Glu
 165 170 175

Gly Asp Trp Ser Ala Met Ser Gly Phe Gln Gln Thr Met Gln Met Leu
 180 185 190

Asn Glu Gly Ile Val Pro Thr Ala Met Leu Val Ala Asn Asp Gln Met
 195 200 205

Ala Leu Gly Ala Met Arg Ala Ile Thr Glu Ser Gly Leu Arg Val Gly
 210 215 220

Ala Asp Ile Ser Val Val Gly Tyr Asp Asp Thr Glu Asp Ser Ser Cys
 225 230 235 240

Tyr Ile Pro Pro Ser Thr Thr Ile Lys Gln Asp Phe Arg Leu Leu Gly
 245 250 255

Gln Thr Ser Val Asp Arg Leu Leu Gln Leu Ser Gln Gly Gln Ala Val
 260 265 270

Lys Gly Asn Gln Leu Leu Pro Val Ser Leu Val Lys Arg Lys Thr Thr

-continued

275	280	285
Leu Ala Pro Asn Thr Gln Thr Ala Ser Pro Arg Ala Leu Ala Asp Ser		
290	295	300
Leu Met Gln Leu Ala Arg Gln Val Ser Arg Leu Glu Ser Gly Gln		
305	310	315

What is claimed is:

1. A recombinant bacterium transformed with a recombinant vector comprising a first DNA sequence encoding a YebF linked to a second DNA sequence encoding a heterologous protein, wherein the mutant bacterium comprises mutations so that at least the NlpD gene product and at least one of the EnvZ, OmpR and YihF gene products are not expressed or are rendered non-functional.

2. The recombinant bacterium of claim 1 wherein the bacterium is a gram negative bacterium.

3. The recombinant bacterium of claim 2, wherein the bacterium is selected from the group consisting of *Escherichia*, *Salmonella*, *Yersinia*, and *Shigella*.

4. The recombinant bacterium of claim 1, wherein both the NlpD and EnvZ gene products are not expressed or are rendered non-functional.

5. The recombinant bacterium of claim 1, wherein both the NlpD and OmpR gene products are not expressed or are rendered non-functional.

6. The recombinant bacterium of claim 1, wherein the NlpD and YihF gene products are not expressed or are rendered non-functional.

7. An expression system for secreting a recombinant protein into a culture medium, the system comprising: (a) a mutant *E. coli* bacterium, wherein the NlpD gene product and at least one of the EnvZ, OmpR and YihF gene products are not expressed or are rendered non-functional; and (b) a recombinant vector comprising a first DNA sequence encoding YebF linked to a second DNA sequence encoding a heterologous protein.

8. The system of claim 7, wherein both the NlpD and EnvZ gene products are not expressed or are rendered non-functional.

9. The system of claim 7, wherein both the NlpD and OmpR gene products are not expressed or are rendered non-functional.

10. The system of claim 7, wherein the NlpD and YihF gene products are not expressed or are rendered non-functional.

11. The system of claim 7, wherein the 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.

12. The system of claim 7, wherein the at least one gene product is not expressed or is rendered non-functional by way of alteration of a promoter control sequence.

13. The system of claim 7, wherein said recombinant vector further comprises an inducible promoter sequence element.

14. The system of claim 7, wherein said recombinant vector further comprises a repressor element.

15. The system of claim 7, wherein the heterologous protein that is secreted is biologically active.

16. The system of claim 7, wherein the heterologous protein is selected from the group consisting of: a cellulase, a protease, a lipase, a cutinase, an amylase, a galactosidase, a pullulanase, a glucose isomerase, a protein disulfide isomerase, a cyclodextrin gluconotransferase, a phytase, a glucose oxidase, a glucosyl transferase, laccase, bilirubin oxidase, a xylanase, an antigenic microbial or protozoan protein, a bacterial protein toxin, a viral protein, and a pharmaceutical.

17. The system of claim 7, wherein 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.

18. The system of claim 7, wherein the expression of both DNA sequences is under the control of an inducible promoter.

19. The system of claim 18, wherein the inducible promoter is a lac promoter.

* * * * *