US 20110217740A1

(19) United States

(12) Patent Application Publication

Adams et al.

(10) Pub. No.: US 2011/0217740 A1

(43) Pub. Date: Sep. 8, 2011

(54) **METHODS, MICROORGANISMS, AND COMPOSITIONS FOR PLANT BIOMASS PROCESSING**

(75) Inventors: **Michael W.W. Adams**, Athens, GA (US); **Janet Westpheling**, Bogart, GA (US); **Scott Hamilton-Brehm**, Oak Ridge, TN (US); **Irina Kataeva**, Athens, GA (US); **Sung-Jae Yang**, Athens, GA (US); **Farris Poole**, Bogart, GA (US)

(73) Assignee: **University of Georgia Research**, Athens, GA (US)

(21) Appl. No.: **13/061,278**

(22) PCT Filed: **Aug. 26, 2009**

(86) PCT No.: **PCT/US09/55049**

§ 371 (c)(1), (2), (4) Date: **May 19, 2011**

Related U.S. Application Data

(60) Provisional application No. 61/190,181, filed on Aug. 26, 2008.

Publication Classification

(51) **Int. Cl.**

C12P 1/04	(2006.01)
C12P 7/04	(2006.01)
C12P 7/06	(2006.01)
C12P 7/16	(2006.01)
C12P 7/64	(2006.01)
C12P 3/00	(2006.01)
C12P 7/46	(2006.01)
C12P 7/48	(2006.01)
C12P 7/50	(2006.01)
C12P 7/44	(2006.01)
C12P 7/40	(2006.01)
C12P 19/12	(2006.01)
C12P 19/02	(2006.01)
C12N 15/63	(2006.01)
C12N 15/81	(2006.01)
C12N 1/00	(2006.01)
C12N 1/21	(2006.01)
C12N 9/14	(2006.01)

(52) **U.S. Cl.** **435/100**; 435/170; 435/157; 435/161; 435/160; 435/134; 435/168; 435/145; 435/144; 435/143; 435/142; 435/136; 435/105; 435/471; 435/243; 435/252.3; 435/195

(57) **ABSTRACT**

Disclosed herein are methods of degrading plant biomass, and microorganisms and polypeptides used in such methods, hi certain embodiments, the methods include growing *Anaerocellum thermophilum* on a substrate that comprises plant biomass under conditions effective for the *A. thermophilum* to convert at least a portion of the plant biomass to a water soluble product or a water insoluble product, hi some cases, the method can further include one or more steps to further process the water soluble product or a water insoluble product to produce, for example, a biofuel or commodity chemical. In another aspect, microorganisms that include at least one *A. thermophilum* plant biomass utilization polynucleotide are disclosed. Also disclosed are methods of transferring one or more *A. thermophilum* plant biomass utilization polynucleotides to a recipient microorganism. *A. thermophilum* plant biomass utilization polynucleotides and polypeptides encoded by such polynucleotides are also disclosed. Also disclosed are methods of degrading plant biomass by providing an isolated *A. thermophilum* polypeptide capable of degrading unprocessed plant biomass, and contacting the *A. thermophilum* polypeptide with plant biomass under conditions effective for the *A. thermophilum* polypeptide to at least partially degrade the plant biomass.

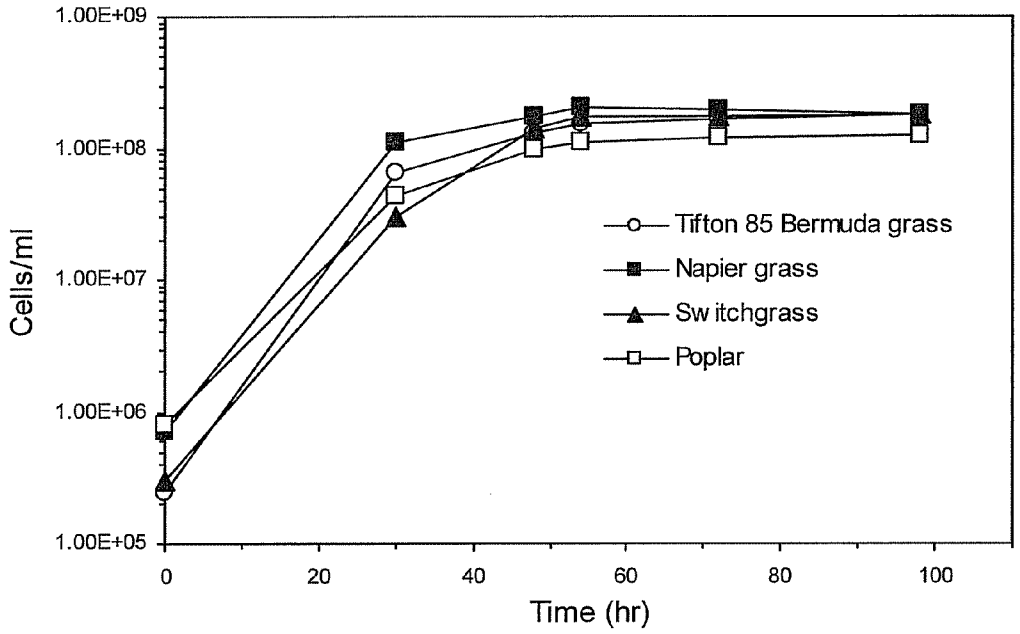

1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
Cells/ml
0
20
40
60
80
100
Time (hr)
Tifton 85 Bermuda grass
Napier grass
Switchgrass
Poplar

Figure 1

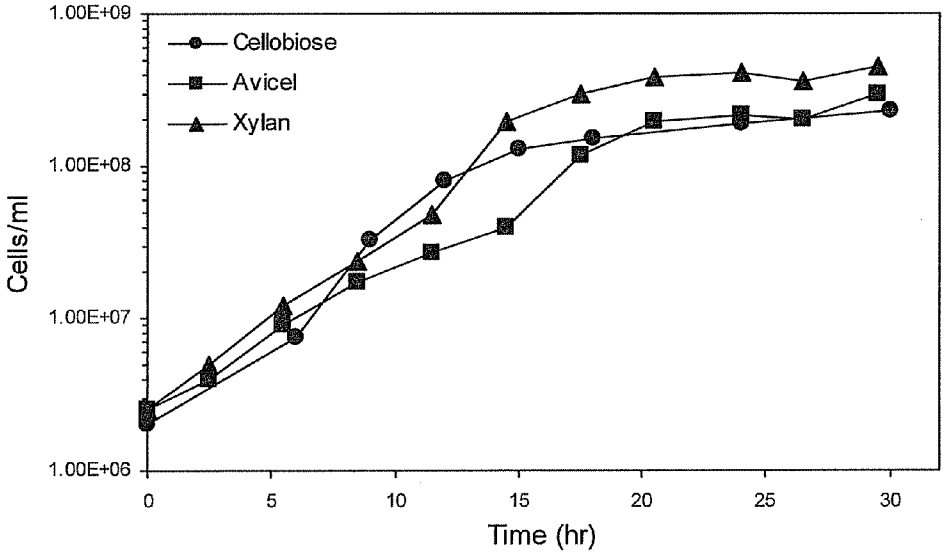

1.00E+09
1.00E+08
1.00E+07
1.00E+06
Cells/ml
Cellobiose
Avicel
Xylan
0
5
10
15
20
25
30
Time (hr)

Figure 2

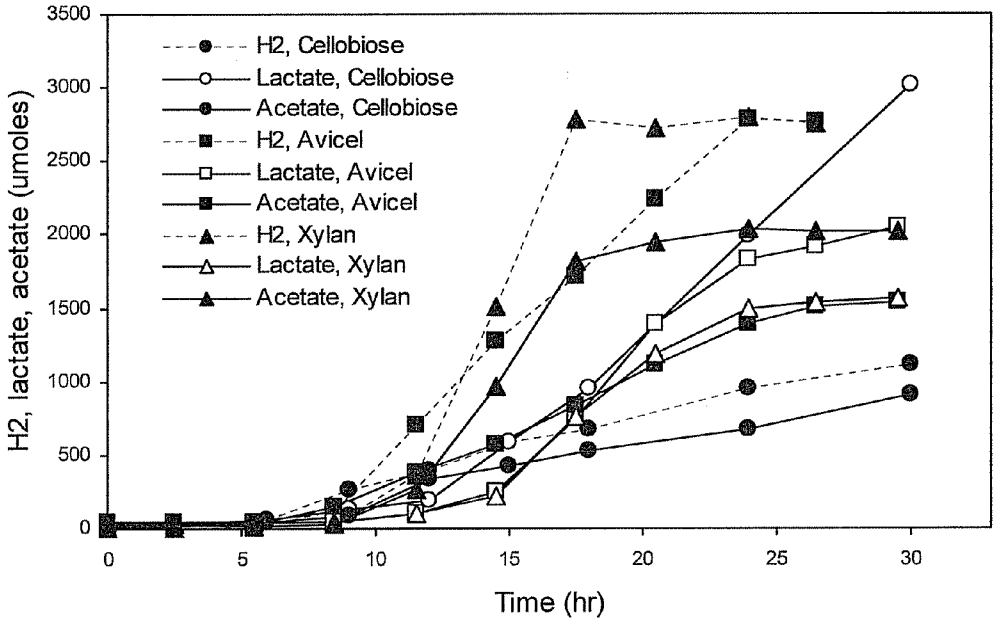

H2, Cellobiose
Lactate, Cellobiose
Acetate, Cellobiose
H2, Avicel
Lactate, Avicel
Acetate, Avicel
H2, Xylan
Lactate, Xylan
Acetate, Xylan
H2, lactate, acetate (umoles)
Time (hr)
0
500
1000
1500
2000
2500
3000
3500
0
5
10
15
20
25
30

Figure 3

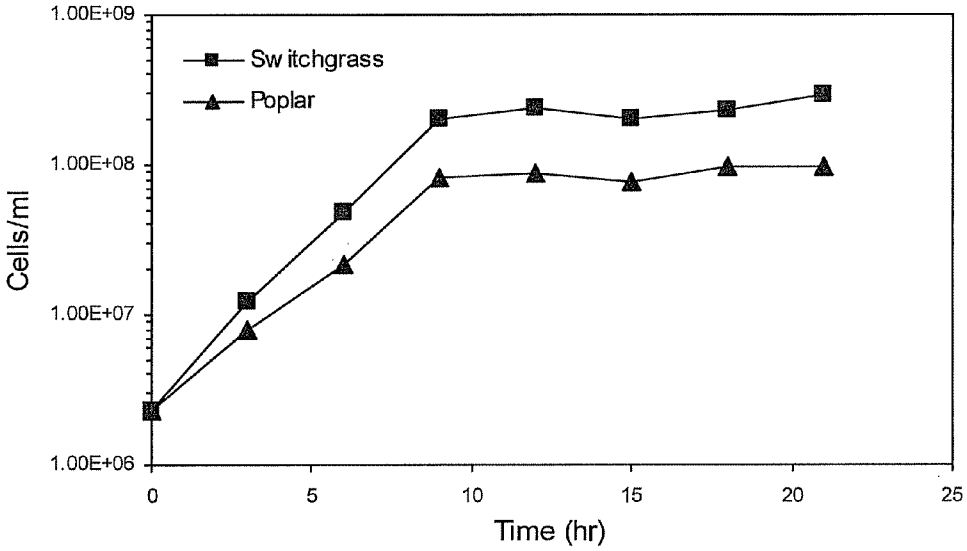

1.00E+09
1.00E+08
1.00E+07
1.00E+06
Cells/ml
Switchgrass
Poplar
0
5
10
15
20
25
Time (hr)

Figure 4

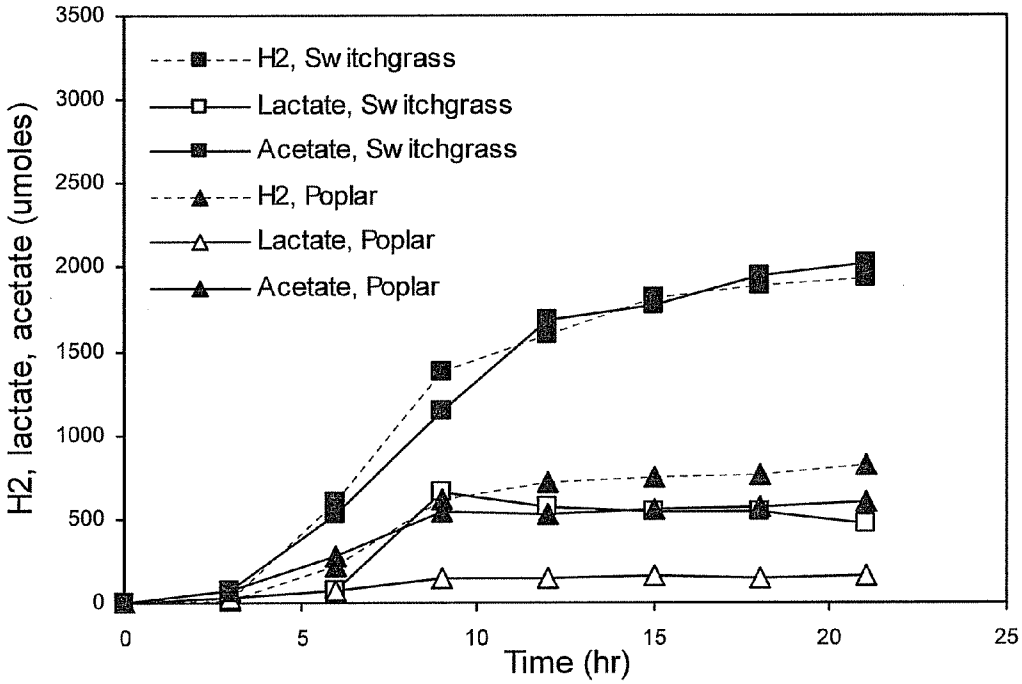

H2, Switchgrass
Lactate, Switchgrass
Acetate, Switchgrass
H2, Poplar
Lactate, Poplar
Acetate, Poplar
H2, lactate, acetate (umoles)
Time (hr)
0
500
1000
1500
2000
2500
3000
3500
0
5
10
15
20
25

Figure 5

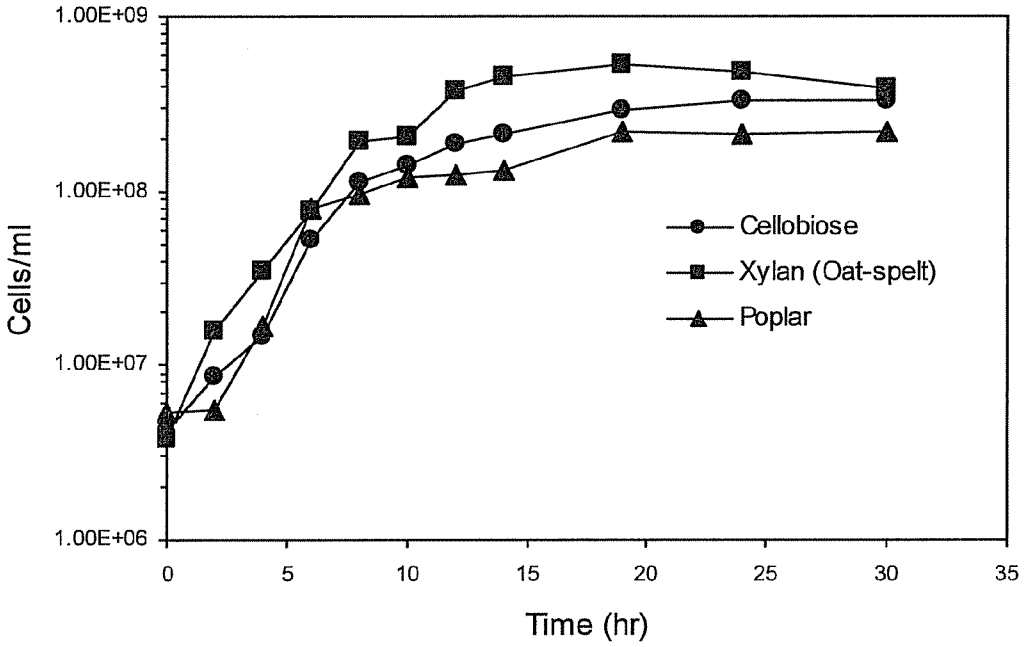

1.00E+09
1.00E+08
1.00E+07
1.00E+06
Cells/ml
0
5
10
15
20
25
30
35
Time (hr)
Cellobiose
Xylan (Oat-spelt)
Poplar

Figure 6

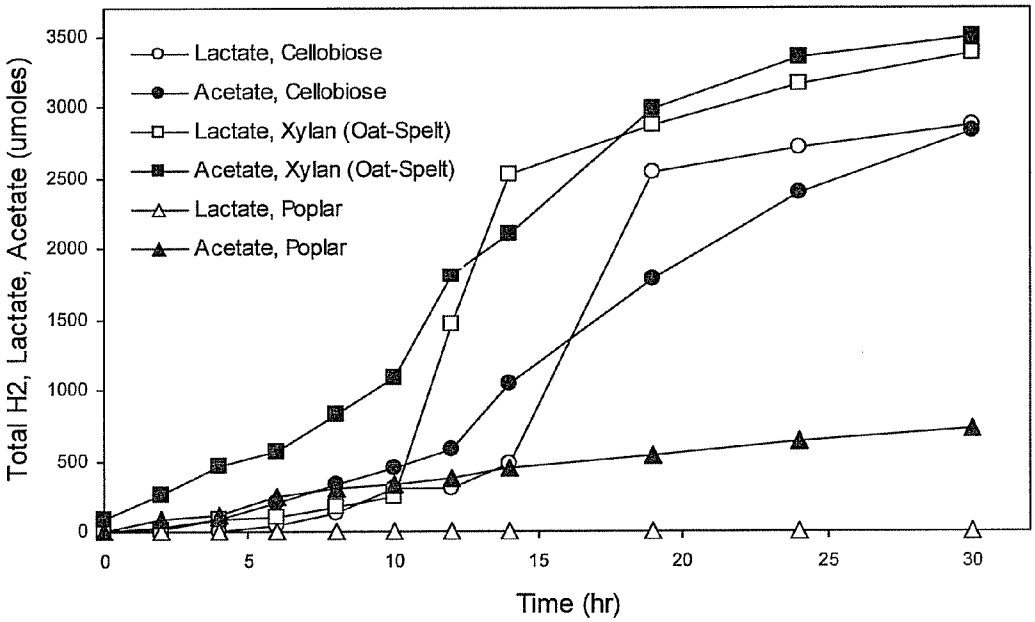

Lactate, Cellobiose
Acetate, Cellobiose
Lactate, Xylan (Oat-Spelt)
Acetate, Xylan (Oat-Spelt)
Lactate, Poplar
Acetate, Poplar
Total H2, Lactate, Acetate (umoles)
3500
3000
2500
2000
1500
1000
500
0
0
5
10
15
20
25
30
Time (hr)

Figure 7

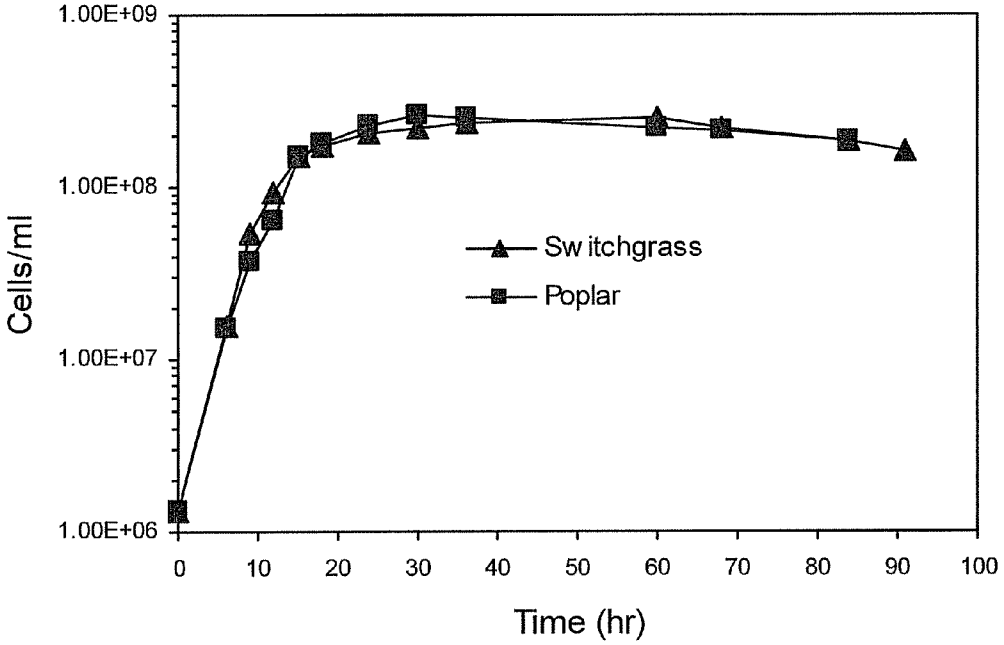

1.00E+09
1.00E+08
1.00E+07
1.00E+06
Cells/ml
0
10
20
30
40
50
60
70
80
90
100
Time (hr)
Switchgrass
Poplar

Figure 8

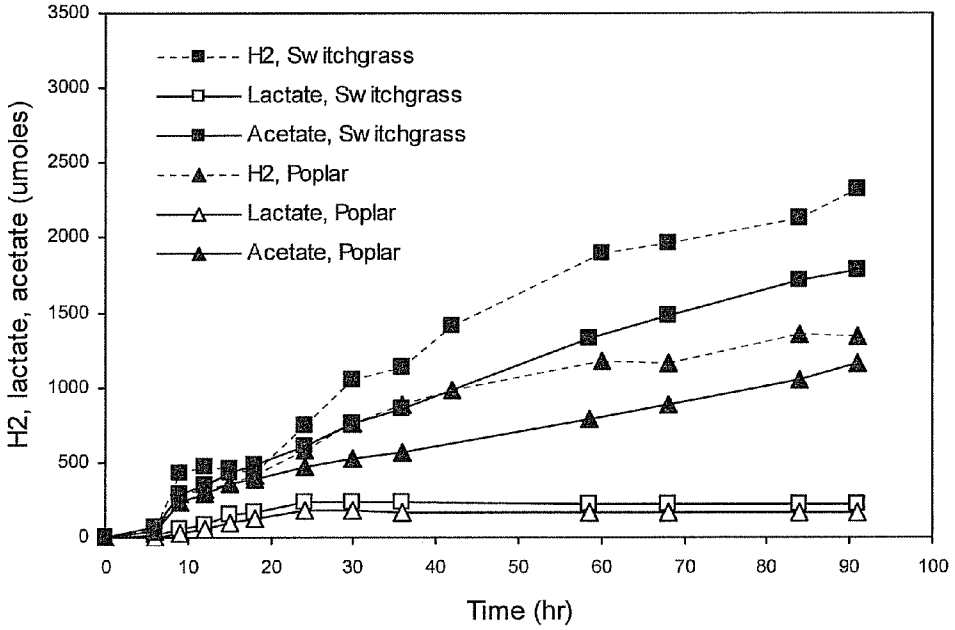

H2, Switchgrass
Lactate, Switchgrass
Acetate, Switchgrass
H2, Poplar
Lactate, Poplar
Acetate, Poplar
H2, lactate, acetate (umoles)
Time (hr)
0
500
1000
1500
2000
2500
3000
3500
0
10
20
30
40
50
60
70
80
90
100

Figure 9

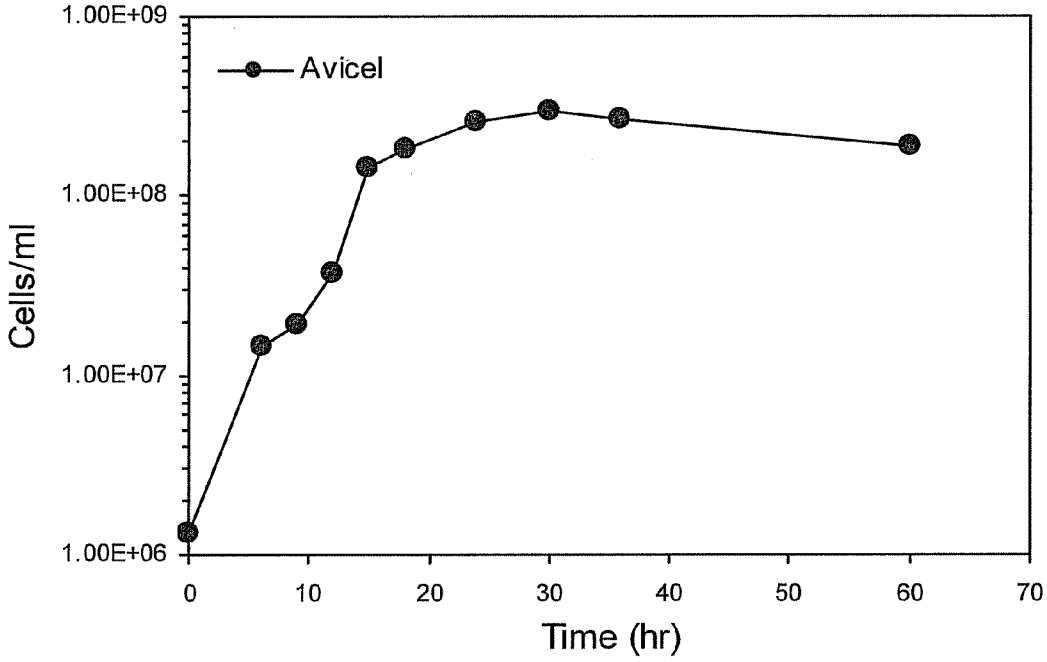

1.00E+09
1.00E+08
1.00E+07
1.00E+06
Cells/ml
Avicel
0
10
20
30
40
50
60
70
Time (hr)

Figure 10

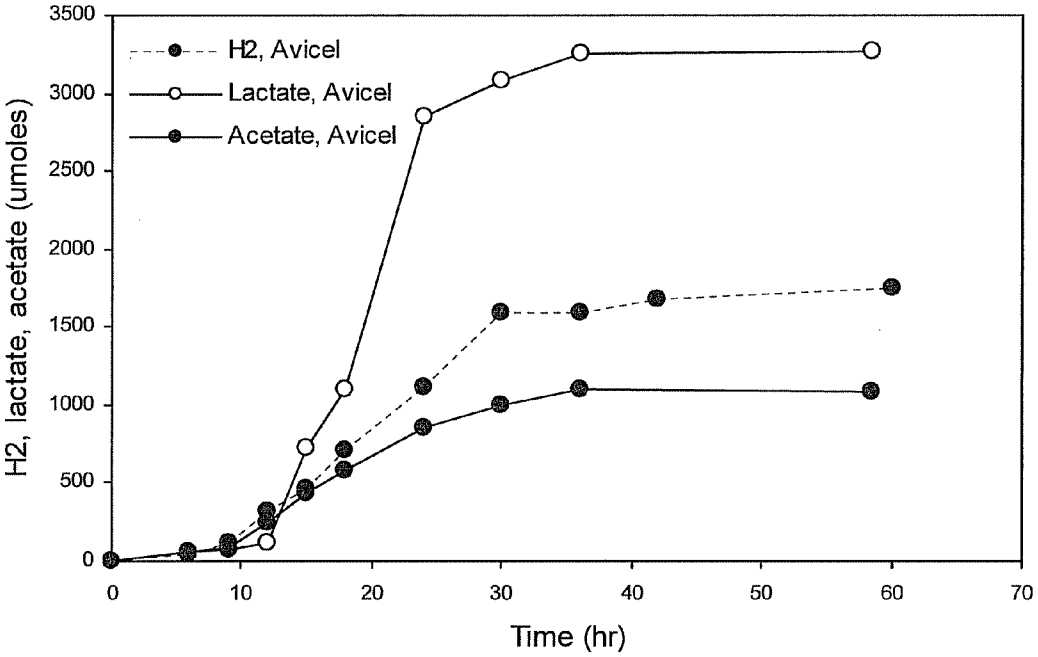

H2, Avicel
Lactate, Avicel
Acetate, Avicel
H2, lactate, acetate (umoles)
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
Time (hr)

Figure 11

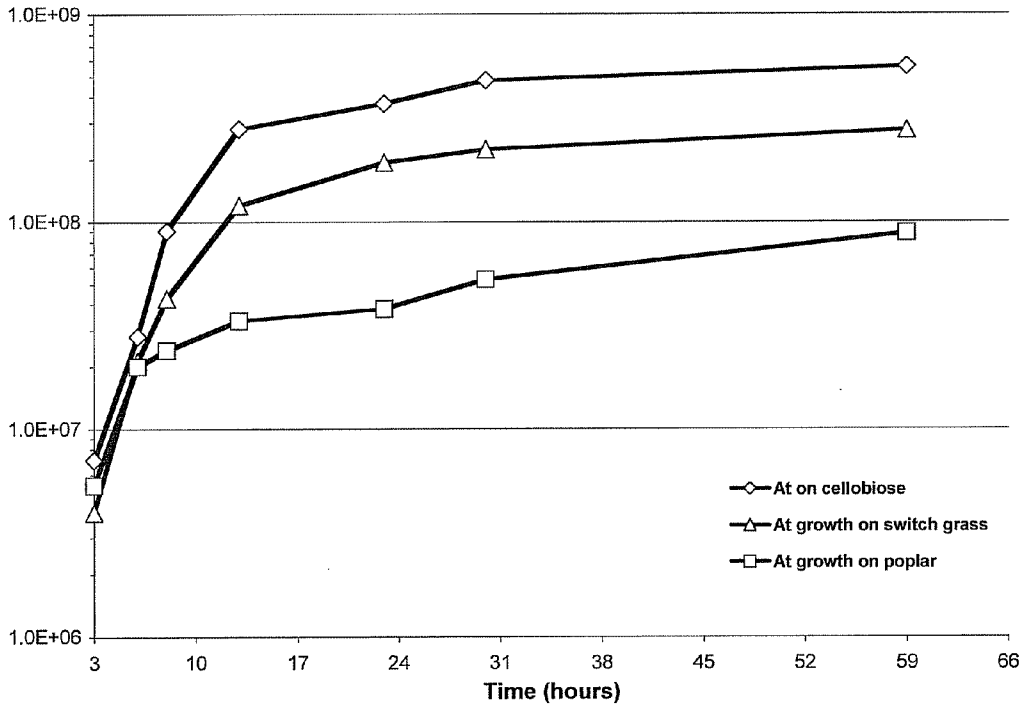

1.0E+09
1.0E+08
1.0E+07
1.0E+06
3
10
17
24
31
38
45
52
59
66
Time (hours)
At on cellobiose
At growth on switch grass
At growth on poplar

Figure 12

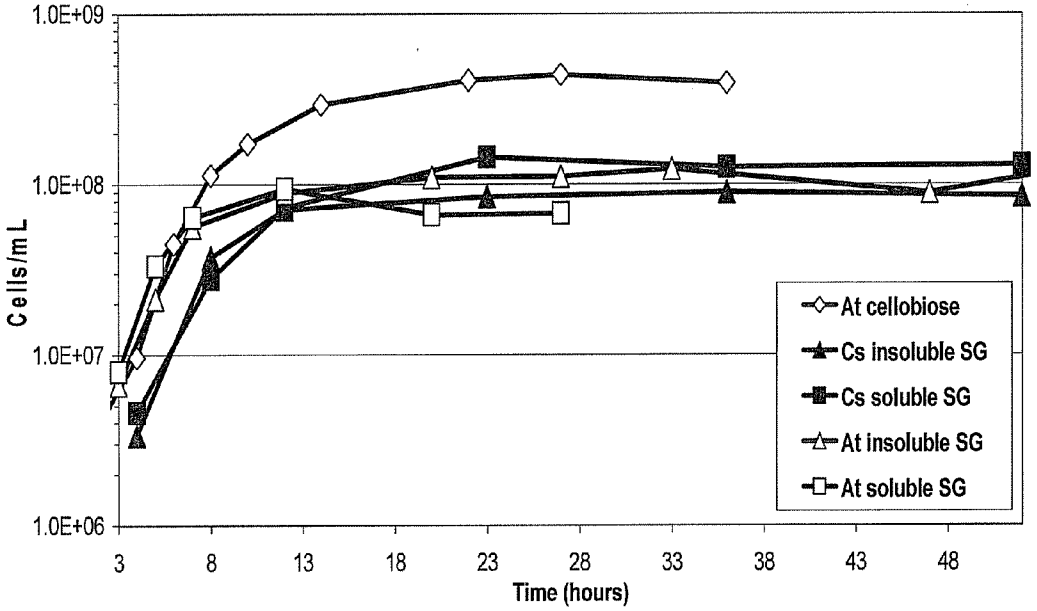

1.0E+09
1.0E+08
1.0E+07
1.0E+06
Cells/mL
3
8
13
18
23
28
33
38
43
48
Time (hours)
At cellobiose
Cs insoluble SG
Cs soluble SG
At insoluble SG
At soluble SG

Figure 13

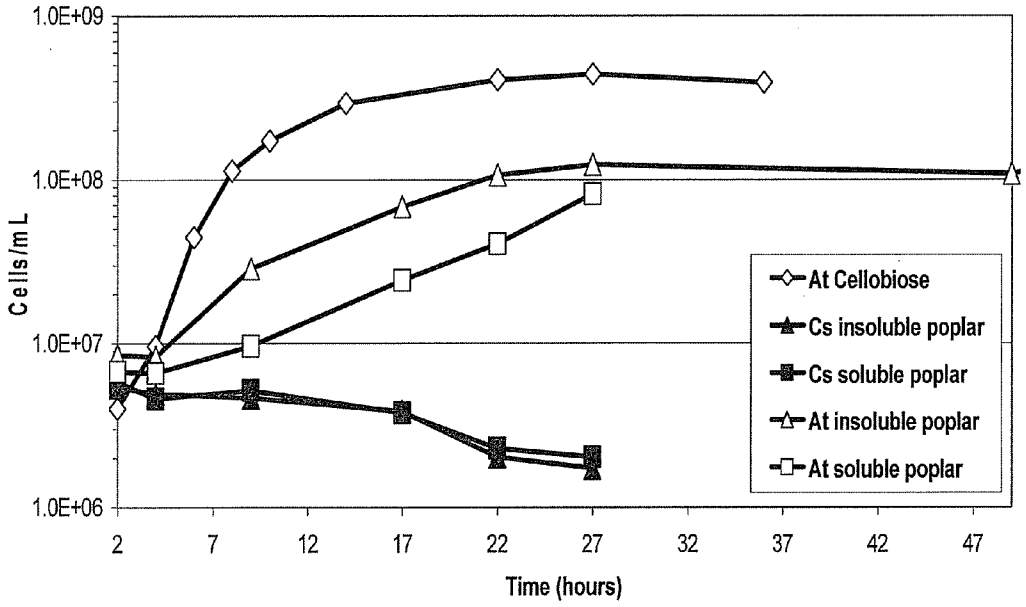

1.0E+09
1.0E+08
1.0E+07
1.0E+06
Cells/mL
2
7
12
17
22
27
32
37
42
47
Time (hours)
At Cellobiose
Cs insoluble poplar
Cs soluble poplar
At insoluble poplar
At soluble poplar

Figure 14

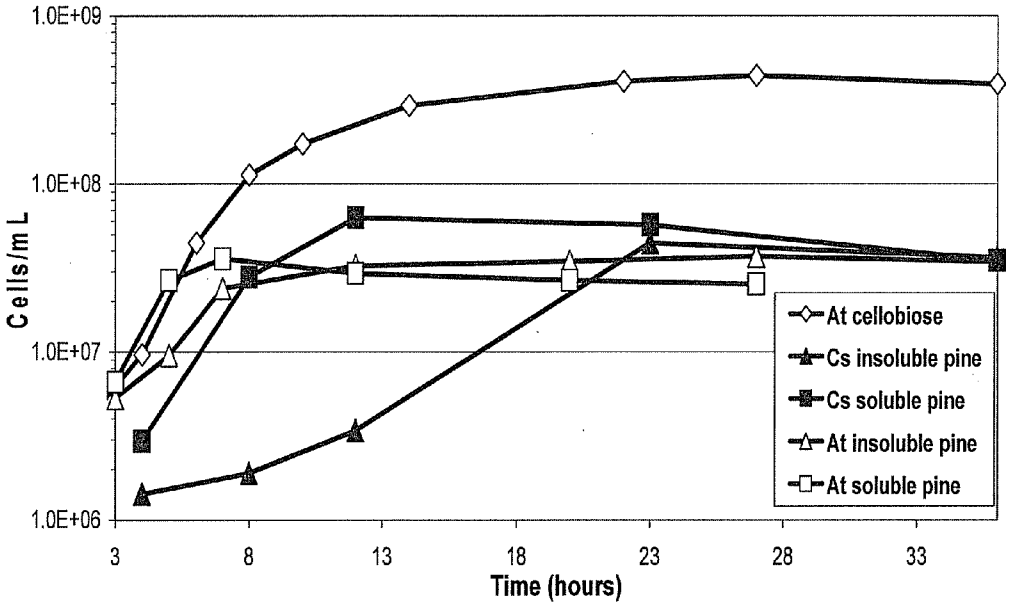

1.0E+09
1.0E+08
1.0E+07
1.0E+06
Cells/m L
3
8
13
18
23
28
33
Time (hours)
At cellobiose
Cs insoluble pine
Cs soluble pine
At insoluble pine
At soluble pine

Figure 15

```
   1 - GGTTCGTTTAACTATGGTGAAGCTTTACAAAAAGCTATCATGTTTTACGAATTTCAAATG - 60
   1 - G  S  F  N  Y  G  E  A  L  Q  K  A  I  M  F  Y  E  F  Q  M   - 20

  61 - TCTGGTAAACTTCCGAATTGGGTACGCAACAACTGGCGTGGCGACTCAGCATTAAAGGAT - 120
  21 - S  G  K  L  P  N  W  V  R  N  N  W  R  G  D  S  A  L  K  D   - 40

 121 - GGTCAAGACAATGGGCTTGATTTGACAGGTGGTTGGTTTGACGCAGGTGATCACGTCAAG - 180
  41 - G  Q  D  N  G  L  D  L  T  G  G  W  F  D  A  G  D  H  V  K   - 60

 181 - TTTAACCTTCCAATGTCATACACTGGTACAATGTTGTCATGGGCAGTGTATGAGTACAAA - 240
  61 - F  N  L  P  M  S  Y  T  G  T  M  L  S  W  A  V  Y  E  Y  K   - 80

 241 - GATGCATTTGTCAAGAGTGGTCAATTGGAACATATCTTAAATCAAATCGAATGGGTTAAT - 300
  81 - D  A  F  V  K  S  G  Q  L  E  H  I  L  N  Q  I  E  W  V  N   - 100

 301 - GACTATTTTGTAAAATGCCATCCAAGCAAATATGTATACTATTACCAGGTTGGGGATGGA - 360
 101 - D  Y  F  V  K  C  H  P  S  K  Y  V  Y  Y  Y  Q  V  G  D  G   - 120

 361 - AGTAAAGATCATGCATGGTGGGGACCTGCTGAGGTTATGCAAATGGAGAGACCTTCATTT - 420
 121 - S  K  D  H  A  W  W  G  P  A  E  V  M  Q  M  E  R  P  S  F   - 140

 421 - AAGGTCACCCAAAGCAGTCCTGGATCTACAGTAGTAACAGAGACAGCAGCTTCCTTAGCA - 480
 141 - K  V  T  Q  S  S  P  G  S  T  V  V  T  E  T  A  A  S  L  A   - 160

 481 - GCAGCTTCAATTGTTTTGAAAGACAGAAATCCCACTAAAGCAGCAACATATCTGCAACAT - 540
 161 - A  A  S  I  V  L  K  D  R  N  P  T  K  A  A  T  Y  L  Q  H   - 180

 541 - GCAAAAGAATTATATGAGTTTGCAGAAGTAACAAAAAGCGATGCAGGTTACACTGCTGCA - 600
 181 - A  K  E  L  Y  E  F  A  E  V  T  K  S  D  A  G  Y  T  A  A   - 200

 601 - AATGGATATTACAATTCATGGAGCGGTTTCTATGATGAGCTTTCTTGGGCAGCAGTTTGG - 660
 201 - N  G  Y  Y  N  S  W  S  G  F  Y  D  E  L  S  W  A  A  V  W   - 220

 661 - TTGTATTTGGCAACAAATGATTCAACATATCTCACAAAAGCTGAGTCATATGTCCAAAAT - 720
 221 - L  Y  L  A  T  N  D  S  T  Y  L  T  K  A  E  S  Y  V  Q  N   - 240

 721 - TGGCCAAAAATTTCTGGCAGTAACACAATTGACTACAAGTGGGCTCATTGCTGGGATGAT - 780
 241 - W  P  K  I  S  G  S  N  T  I  D  Y  K  W  A  H  C  W  D  D   - 260

 781 - GTTCACAATGGAGCGGCATTATTGTTAGCAAAAATTACCGGTAAGGATATTTATAAGCAA - 840
 261 - V  H  N  G  A  A  L  L  L  A  K  I  T  G  K  D  I  Y  K  Q   - 280

 841 - ATTATTGAAAGTCACTTAGATTACTGGATCACAGGATACAATGGCGAAAGGATTAAGTAT - 900
 281 - I  I  E  S  H  L  D  Y  W  I  T  G  Y  N  G  E  R  I  K  Y   - 300

 901 - ACACCAAAAGGATTAGCATGGCTTGATCAATGGGGTTCGTTGAGATATGCAACAACTACA - 960
 301 - T  P  K  G  L  A  W  L  D  Q  W  G  S  L  R  Y  A  T  T  T   - 320

 961 - GCATTTTTGGCATTTGTTTATAGCGATTGGGTTGGCTGTCCAAGCACAAAAAAAGAAATA - 1020
 321 - A  F  L  A  F  V  Y  S  D  W  V  G  C  P  S  T  K  K  E  I   - 340

1021 - TATAGAAAATTTGGAGAAAGCCAGATTGATTATGCGTTAGGCTCAGCTGGAAGAAGCTTT - 1080
 341 - Y  R  K  F  G  E  S  Q  I  D  Y  A  L  G  S  A  G  R  S  F   - 360

1081 - GTTGTTGGATTTGGTACAAATCCACCAAAGAGACCGCATCACAGAACTGCTCATAGCTCA - 1140
 361 - V  V  G  F  G  T  N  P  P  K  R  P  H  H  R  T  A  H  S  S   - 380

1141 - TGGGCAGACAGTCAGAGTATACCTTCATATCACAGACATACATTATATGGAGCGCTTGTT - 1200
 381 - W  A  D  S  Q  S  I  P  S  Y  H  R  H  T  L  Y  G  A  L  V   - 400

1201 - GGTGGTCCAGGCTCTGATGATAGCTACACAGATGATATAAGCAACTATGTGAACAATGAG - 1260
 401 - G  G  P  G  S  D  D  S  Y  T  D  D  I  S  N  Y  V  N  N  E   - 420

1261 - GTTGCATGTGATTATAATGCAGGGTTTGTGGGTGCATTAGCAAAGATGTATCAATTGTAC - 1320
 421 - V  A  C  D  Y  N  A  G  F  V  G  A  L  A  K  M  Y  Q  L  Y   - 440

1321 - GGTGGGAATCCAATACCAGATTTCAAAGCTATTGAAACTCCAACAAACGACGAATTCTTT - 1380
 441 - G  G  N  P  I  P  D  F  K  A  I  E  T  P  T  N  D  E  F  F   - 460

1381 - GTTGAAGCTGGTATAAATGCATCCGGAACTAACTTTATTGAAATTAAAGCGATAGTTAAT - 1440
 461 - V  E  A  G  I  N  A  S  G  T  N  F  I  E  I  K  A  I  V  N   - 480
```

```
1441 - AACCAAAGTGGTTGGCCTGCCAAAGCAACAGATAAGCTTAAATTTAGATATTTTGTTGAC - 1500
 481 - N  Q  S  G  W  P  A  K  A  T  D  K  L  K  F  R  Y  F  V  D  - 500

1501 - CTGAGTGAATTAATTAAAGCAGGATATTCACCAAATCAATTAACCTTGAGCACCAATTAT - 1560
 501 - L  S  E  L  I  K  A  G  Y  S  P  N  Q  L  T  L  S  T  N  Y  - 520

1561 - AATCAAGGTGCAAAAGTAAGTGGACCTTATGTATGGGATGCAAGCAAAAATATATACTAC - 1620
 521 - N  Q  G  A  K  V  S  G  P  Y  V  W  D  A  S  K  N  I  Y  Y  - 540

1621 - ATTTTAGTAGACTTTACTGGCACATTGATTTATCCAGGTGGTCAAGACAAATATAAGAAA - 1680
 541 - I  L  V  D  F  T  G  T  L  I  Y  P  G  G  Q  D  K  Y  K  K  - 560
```

Figure 16

TACGTA
ATGAGTAGGAAGCTTTCTCTCCTCTTAGTATTTCTAATTTTTGGAAGCATGCTGGGAGCA
AACAACATTGTAAAAGCT (ATG) GGTTCGTTTAACTATGGTGAAGCTTTACAAAAAGCTATCATG
TTTTACGAATTTCAAATGTCTGGTAAACTTCCGAATTGGGTACGCAACAACTGGCGTGGC
GACTCAGCATTAAAGGATGGTCAAGACAATGGGCTTGATTTGACAGGTGGTTGGTTTGAC
GCAGGTGATCACGTCAAGTTTAACCTTCCAATGTCATACACTGGTACAATGTTGTCATGG
GCAGTGTATGAGTACAAAGATGCATTTGTCAAGAGTGGTCAATTGGAACATATCTTAAAT
CAAATCGAATGGGTTAATGACTATTTTGTAAAATGCCATCCAAGCAAATATGTATACTAT
TACCAGGTTGGGGATGGAAGTAAAGATCATGCATGGTGGGGACCTGCTGAGGTTATGCAA
ATGGAGAGACCTTCATTTAAGGTCACCCAAAGCAGTCCTGGATCTACAGTAGTAACAGAG
ACAGCAGCTTCCTTAGCAGCAGCTTCAATTGTTTTGAAAGACAGAAATCCCACTAAAGCA
GCAACATATCTGCAACATGCAAAAGAATTATATGAGTTTGCAGAAGTAACAAAAAGCGAT
GCAGGTTACACTGCTGCAAATGGATATTACAATTCATGGAGCGGTTTCTATGATGAGCTT
TCTTGGGCAGCAGTTTGGTTGTATTTGGCAACAAATGATTCAACATATCTCACAAAAGCT
GAGTCATATGTCCAAAATTGGCCAAAAATTTCTGGCAGTAACACAATTGACTACAAGTGG
GCTCATTGCTGGGATGATGTTCACAATGGAGCGGCATTATTGTTAGCAAAAATTACCGGT
AAGGATATTTATAAGCAAATTATTGAAAGTCACTTAGATTACTGGATCACAGGATACAAT
GGCGAAAGGATTAAGTATACACCAAAAGGATTAGCATGGCTTGATCAATGGGGTTCGTTG
AGATATGCAACAACTACAGCATTTTTGGCATTTGTTTATAGCGATTGGGTTGGCTGTCCA
AGCACAAAAAAAGAAATATATAGAAAATTTGGAGAAAGCCAGATTGATTATGCGTTAGGC
TCAGCTGGAAGAAGCTTTGTTGTTGGATTTGGTACAAATCCACCAAAGAGACCGCATCAC
AGAACTGCTCATAGCTCATGGGCAGACAGTCAGAGTATACCTTCATATCACAGACATACA
TTATATGGAGCGCTTGTTGGTGGTCCAGGCTCTGATGATAGCTACACAGATGATATAAGC
AACTATGTGAACAATGAGGTTGCATGTGATTATAATGCAGGGTTTGTGGGTGCATTAGCA
AAGATGTATCAATTGTACGGTGGGAATCCAATACCAGATTTCAAAGCTATTGAAACTCCA
ACAAACGACGAATTCTTTGTTGAAGCTGGTATAAATGCATCCGGAACTAACTTTATTGAA
ATTAAAGCGATAGTTAATAACCAAAGTGGTTGGCCTGCCAAAGCAACAGATAAGCTTAAA
TTTAGATATTTTGTTGACCTGAGTGAATTAATTAAAGCAGGATATTCACCAAATCAATTA
ACCTTGAGCACCAATTATAATCAAGGTGCAAAAGTAAGTGGACCTTATGTATGGGATGCA
AGCAAAAATATATACTAC ATTTTAGTAGACTTTACTGGCACATTGATTTATCCAGGT**TAA**
GGATCC

Figure 17

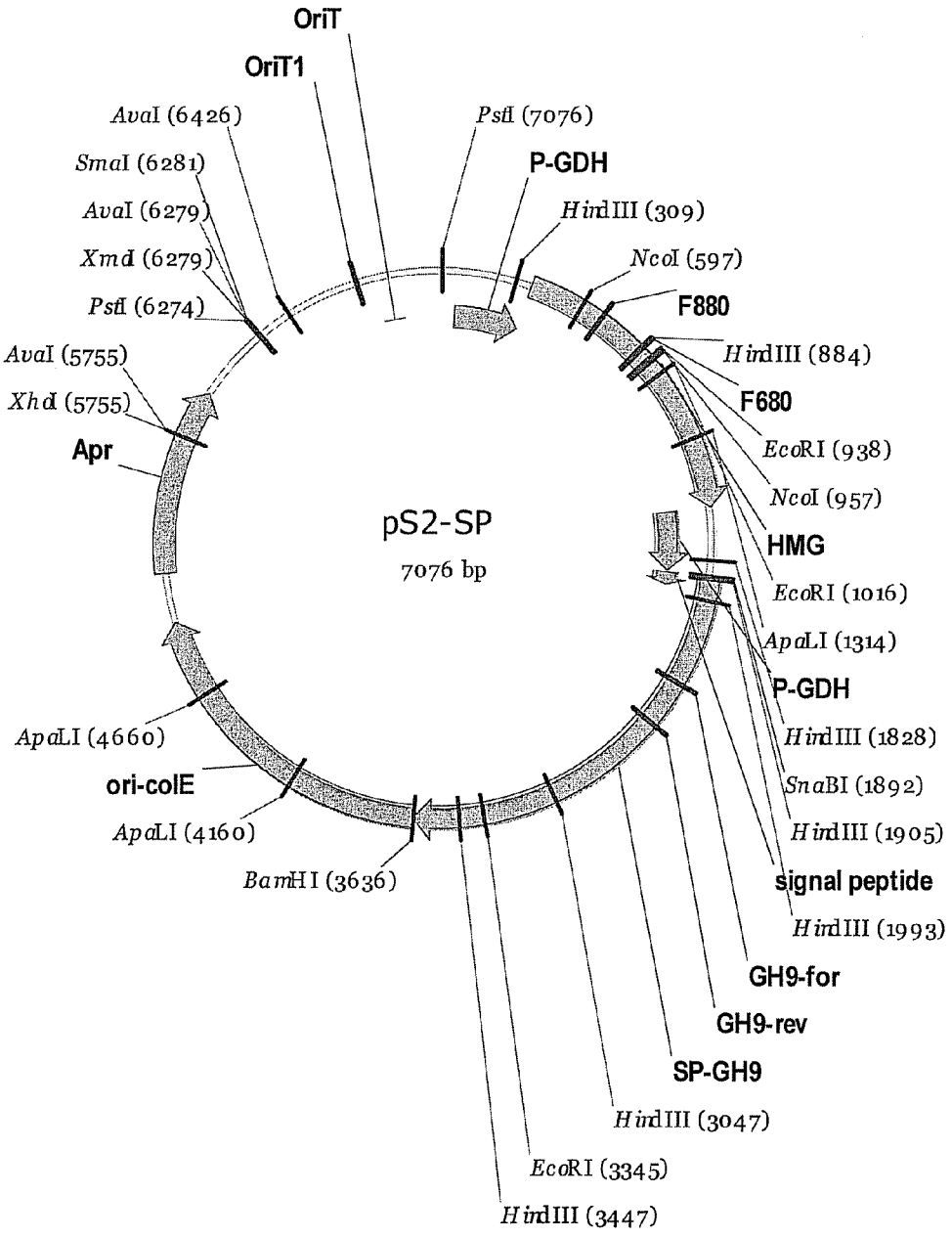

OriT
OriT1
AvaI (6426)
SmaI (6281)
AvaI (6279)
XmaI (6279)
PstI (6274)
AvaI (5755)
XhoI (5755)
Apr
PstI (7076)
P-GDH
HindIII (309)
NcoI (597)
F880
HindIII (884)
F680
EcoRI (938)
NcoI (957)
HMG
EcoRI (1016)
ApaLI (1314)
P-GDH
HindIII (1828)
SnaBI (1892)
HindIII (1905)
signal peptide
HindIII (1993)
GH9-for
GH9-rev
SP-GH9
HindIII (3047)
EcoRI (3345)
HindIII (3447)
BamHI (3636)
ApaLI (4160)
ori-colE
ApaLI (4660)
pS2-SP
7076 bp

Figure 18

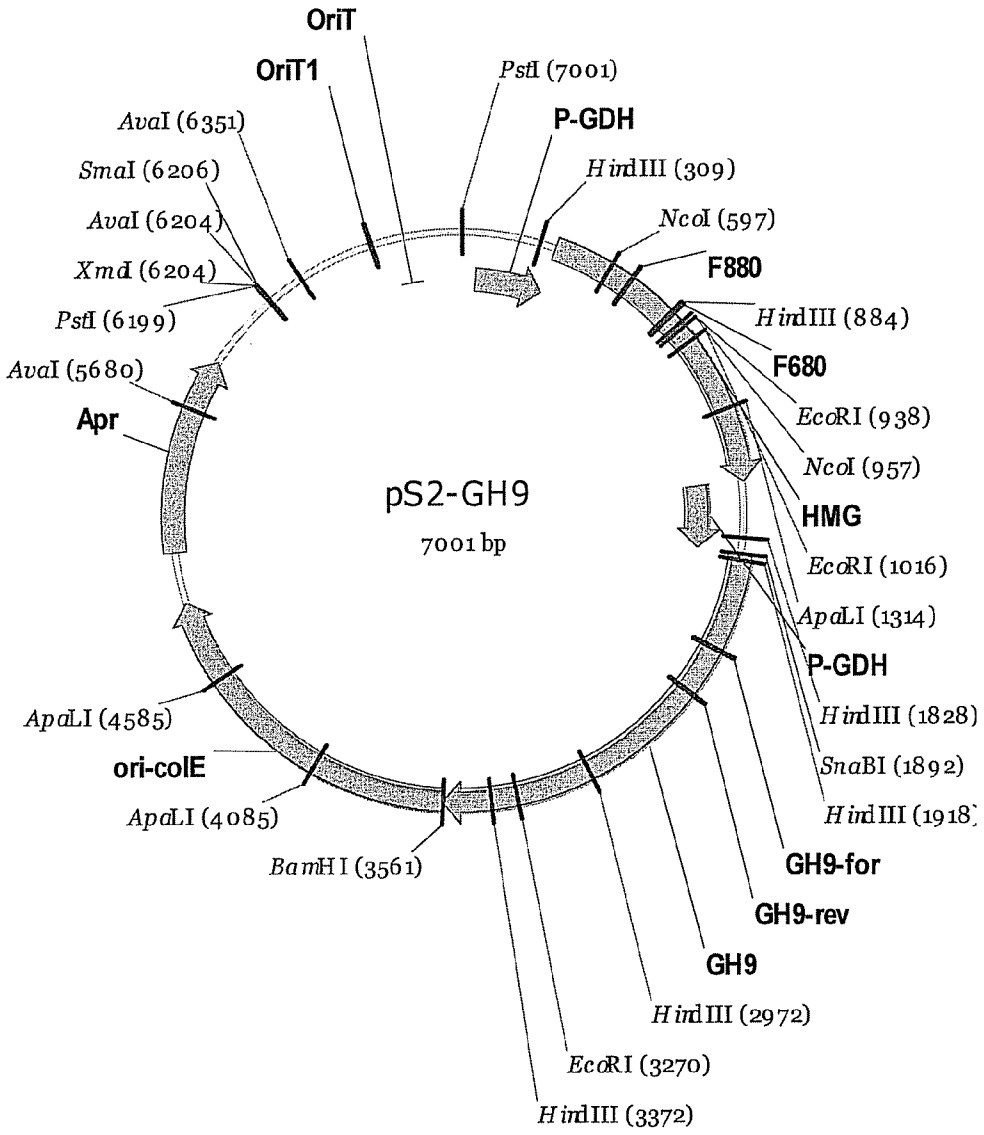

OriT
OriT1
PstI (7001)
AvaI (6351)
P-GDH
SmaI (6206)
HindIII (309)
AvaI (6204)
NcoI (597)
XmaI (6204)
F880
PstI (6199)
HindIII (884)
AvaI (5680)
F680
Apr
EcoRI (938)
NcoI (957)
pS2-GH9
7001 bp
HMG
EcoRI (1016)
ApaLI (1314)
P-GDH
ApaLI (4585)
HindIII (1828)
ori-colE
SnaBI (1892)
ApaLI (4085)
HindIII (1918)
BamHI (3561)
GH9-for
GH9-rev
GH9
HindIII (2972)
EcoRI (3270)
HindIII (3372)

Figure 19

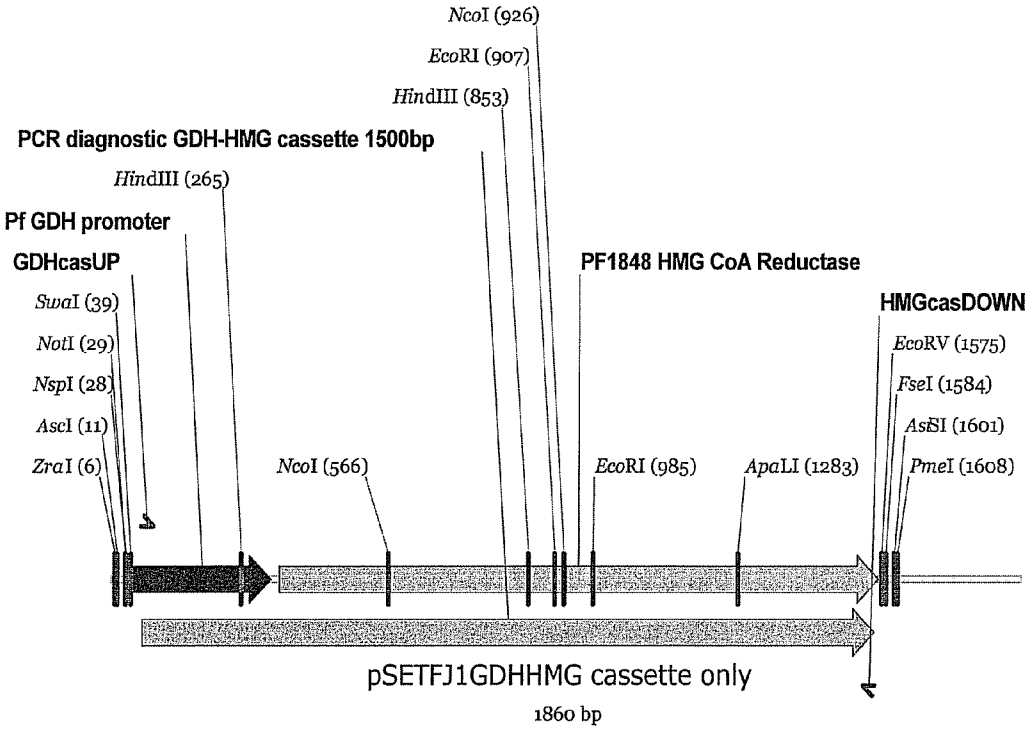
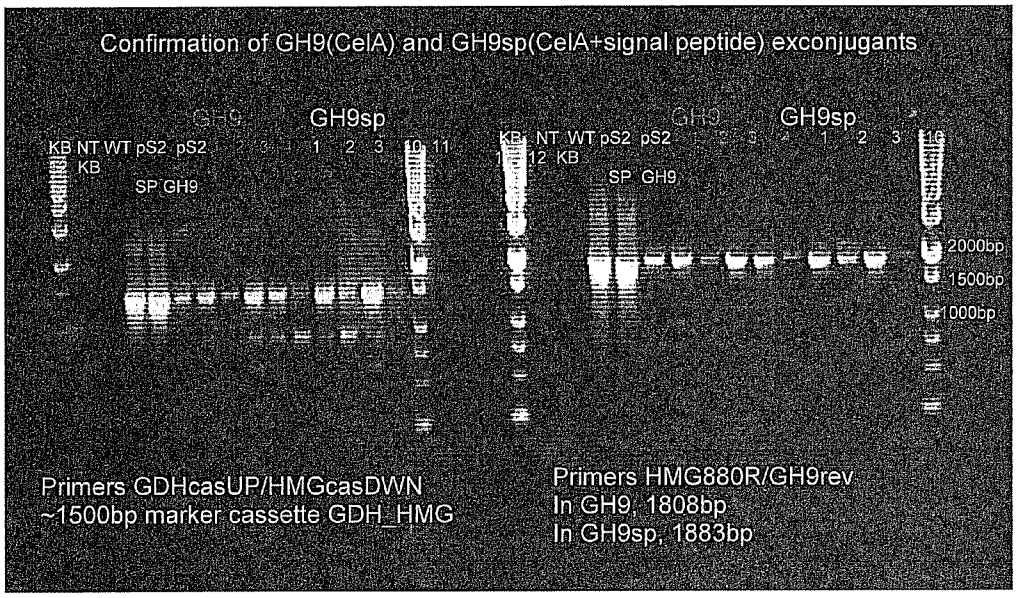
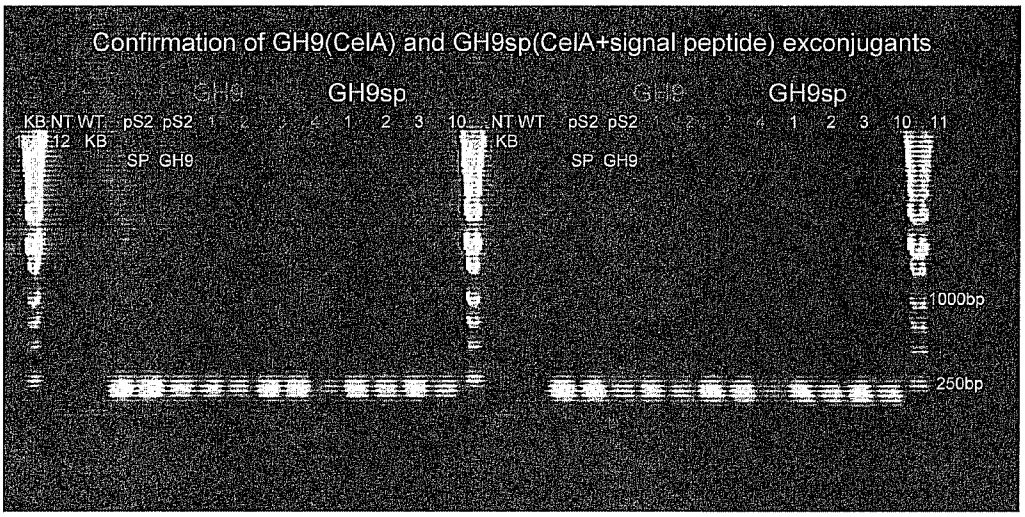

NcoI (926)
EcoRI (907)
HindIII (853)
PCR diagnostic GDH-HMG cassette 1500bp
HindIII (265)
Pf GDH promoter
GDHcasUP
PF1848 HMG CoA Reductase
HMGcasDOWN
SwaI (39)
NotI (29)
NspI (28)
AscI (11)
ZraI (6)
NcoI (566)
EcoRI (985)
ApaLI (1283)
EcoRV (1575)
FseI (1584)
AsSI (1601)
PmeI (1608)
pSETFJ1GDHHMG cassette only
1860 bp

Figure 20



- NT, WT negative controls empty
- pS2-SP plasmid, pS2-GH9 plasmid positive controls all correct
- GDH_HMG marker cassette is present in all 10 exconjugants
- HMG_celA fragment present in all 10 exconjugants

Figure 21



- Primers GH9forward, GH9reverse
- 219bp fragment, CelA
- Negative controls empty, positive controls correct
- Transgene present in ALL exconjugants

Figure 22

Athe Gene Number:
Athe_0010

Product Function:
hypothetical protein

Nucleotide Sequence:
atgaaagttaagaaaaagaagataggtagtgttattatgatttttgcaatattgt
tgagcgttattgttccagttttaacggctacatatcaaagtaaaagttatttaga
aactacataa

Amino Acid Sequence:
MKVKKKKIGSVIMIFAILLSVIVPVLTATYQSKSYLETT*

Figure 23-01

Athe Gene Number:
Athe_0011

Product Function:
hypothetical protein

Nucleotide Sequence:
atggtttttaaaagtgataaaaacttaatttacttggttacaggaaaagtgaaga
aaaatgaccttggcatgggaatttacatttgcgaataa

Amino Acid Sequence:
MVFKSDKNLIYLVTGKVKKNDLGMGIYICE*

Figure 23-02

Athe Gene Number:
Athe_0012

Product Function:
YD repeat protein

Nucleotide Sequence:
gtgagatttccaaagaaattgagagtattgttaacatttcttattcttttaatct
tcacccttaatacaaatctatatgtcttttcacaaagtagtaatctcaagaacca
agagtcaaattacaaagatttaataggacactgggcagaggatgaatttagatgg
ttgatagagaagggaataattagtggagtgaaggcaaatggagtaatgttagcaa
agccagatgagaagttaacaagagcggaagctgttgtgattatactaagaactat
atttgagaaggagcttcttgaggaggaaataaagaaattaaaagaggatagtttt
aaagatatttcaaatcattggagcagagattacataaatgtagctgcgagatatg
gtttggttaaaggttattcagacaagacatttagaccaaatcagagaattacacg
tgaagagtttgttttaatggtgataagggtgagcaaatatagggagcaagcagca
agtgagggcaagcaaaacgataagaaggtaaaatttaaagatgtgagtgttaaca
attttggatataacgaaataatttttgcagttgagaaaggcttgatcaaaggcta
tagcgatggtacatttagaccaagaagttatataagccgagcagaagcagcagta
atagttgcaagggcattgaaagcggatttattcattacctatagagcaaagaagc
ctcaaataactaaagaaaatgcgagcgaatttttgaacttgttgttgataaaga
ggaagtaggtataatagaaaaggcgaagctcacactaaagagcaaagtaaagggt
gtcaagtttacagcagagtggaaggcaagtggtggtaaacttgaggtagctaaag
ataatcagagtgcgatttggagcccgacagatgcagaagagggtaagaattatat
tgtgactgttgaagtaaccatcctgacagataatggggaaaagataaaaatacaa
aaggttgcaaggatacgagtgaacgagggcgtccaactgacaaagaatgaaaaca
tagaagaaccggtgaataaaatagatccattatttacccaaagagaatcaaacaa
tttagttcaagaaaattctagttatagcccattagaagaaaaactttcagcgcca
atattgaaaatagcaaaaaaaggtgagtatgtagagcttacttggaatgaggtaa
aaggtgcaaggagttacattgtaaaaagagggaaggtaagcggggagtatgaggt
attggcaagcggagttattactacggcgtatattgatggtcctatggatggaaaa
acaacgtattactatgtggttgcagcagtgggagcaaatggtacgaccagtatga
attcaaatgaggtagtgtacaaagcactgccacaatcaccaaagttatttgggac
atacaatgggacaagtgcacgtcttacatggacgaaagcgaatggagcagagagg
tatactctttacaggagcacagtcagtggaggaccatactatgcaattgcagaaa
acttacttaccaatgtgtttgaagatactaatttgacatcaggtacagtgtatta
ctatgtagtaaaagccattaacgaattaggagaaagtgaatactcaaatgaggta
gcgatgggggaaggagttacagttgcagctaaatttaatcctgatgacgacgagg
atggagatgggttgacaaatgtagatgagttaaaatatgcaacaagtttgaagaa
gaatgatagtgatggagatgggttgagtgacggttatgaagtaaagaaaggaaca
aatccattagtaccggatacagataatgacgggatatatgatggagcagaggttg
tgatgggaacagacccattgacaaagaatcctttgacaagtgccgagaaatatgc
ggtatcggaagatggtaaagtttttgtaagagcactttctgatgcaaatatttta
atagctcccttacaggtaaagaggtctgacaatgtgtttattaactcattgaaag
gaatagtaggtaaagcgatagagataactgcagggggatttgatataaagaaggc
agagatagtagtgaactatgatgaagcagaacttaatggtgtagaagagaataat

Figure 23-03

ttaatgctgtattatgtcaattatgacaagaagatattagagccgttagaagatg
ttgtagtagatacagtgtataatcgagtttcgggtaagacagagcattttagcac
atttttacttggtgataagaatatgccagttgatctttccaaagtagatattgtg
tttgtacttgacaactcaggtagtatgtcgtctaatgacccaaattattacagaa
tagaagcaacaaagaaatttatacaaaatatagatgaacttaacaatagagttgg
gttagtagatttcgatagctctgtaagtgttagatcaaatttgacatctgacaag
agcaaattgttacaagccctaaatgcgatgagatggacaggtggttctacgaata
taggaggaggattgaaagctgcgctgggattattcgaccaagagcaatctaaaaa
gataatagtacttttatcagacggatatcacaacacaggaattcatccaaacgat
gtactaccagaattaataaaacaagaaatagtagtaaataccatcgcattgggta
aagattgtgatagggagttattacatgatatagctgataagacaaaaggaggcta
tttttacgttgataatacaggaggactttctcaagaagatgtagacaagcaaata
gagctaatatacgagaaactaactaaatggataacccttcaaaaggaagcagaaa
agaatctcaaacctcaagaagtgttaagtatagagtacaatgatgtaggtttaga
caatgaagagtttcataaatggataaccacagctatgactaatttacttacggga
aattatatggaagagtttgatgatattagtatagaggggaatggtcctgagttta
aatttgtgaggtattacaattcttttgctaatcaacaaaagacgataataggtaa
aggttttagaaccaattttgacagcaagcttacaaaggtagcgggagtcggaata
gtccaagcaggggtattgaatgtaagagaaggaccgaatgtaaataccaaaaaga
tagggtatttgactaaggggacaaaggtgcagattgaagaggatggcaacaagaa
cggcagtggttggcacagaattgtgtataagggtaaatctgcatatatttgcgca
gcatatgtcaaagagctaaacaatggaattgaagtaacatatccaagtggtagta
tgatagtatttgtggatgataatggagatgggatttacttgtcagacagtaacaa
agctgacaaggtagttttgtcaggtaatgaatatatattgatacaaagagatttg
acgagatatgtttttgacaagagtggaaaacttatcaaaattggtgatagaaatg
aaaattacattactattgagtattctggtgaaaagatatctgcggcaagagatgt
ttttggcagaaaattagaatttatctttcaaggggataacctggtatgtataaga
gagaacatcaagggtaaaataggtaggaaagtagaatttgtatatgatgacaagg
atagactaataaaagtaattggggttgatggcgcagagacaagatatgaatatga
tgagaaagacagactcaagaggataattgatgcgaacgggcaccaggtggtaagg
aatgaatatgatattttaggtaggatagtgaggcaatacgacggggaagatataa
tcagatttttcatatacgatgatgaagacagagtaagatattatatagacgaaaa
tggtaatgagagtatggtagtattcaatgaagagctaaaaccaataaaagagaga
aatgcattgggtgggggttgtgactacaagtatgagattaatgacggttcaaagt
ggatagacgtgacaactcctgatttggataaggatgtagttgtgagcggtcttac
acgggaaaaatatcaagagctgaaagagaaaggtagcatgacaaagcaggtcaca
attcagataataaagacttctcctagtttagaaacagcaaagactacacaaatat
acgatggacgtggcaatattatccaagtgatagatgcttatggtaattcaataaa
aatgaaatatgataacaataataatttaattgagcaaacagacaggattggtgct
actaccaaatatatttatgatgccgagggtataaatttaattgagaaaattgatc
cacttggaaacaaggaaagatatgaatattattctataaattcgggcataaaact
caatggtttacttgcaaagtatatagataaaaatgggaatgagacgagatactat
tatgaagatgaatataacaatttaacacgagtggttgacgcagaaggttatgaga
caaaatatgaatatgaccagttaggcagaaaaataaaagaaataaatgaaagagg
atatgtaacaaggtttgagtatgatccagaaggcagaattacaaaagaaatagac
gcgtttggtaaaacaaaagtatatgtatatgacaaagtaggtaatttaatagaag

Figure 23-03

aaatcgatcgccttggcaacaaaacaagatatgtatacgatgataaaaacaggct
tataaaagaaattgatgcgatgggtggagagtatcaatacttttatgatcctgta
gggaacaagataaaggagatagaccctgaaggaagagtaaccaaatatacttatg
atgaactgaacagacttgttgagatagaagatgcagaaggcaataaaaccaaatt
caagtatgatttagcaggcaggaaaatatcagaggtaaatgccttaggaaaagag
acaagatatgaatatgatttattaggaagacttacgaaagtgattgatcctttag
gtaaaacccgaagttaccaatacaatgctgaagggtataaaatatcagagacgaa
taaaaatggtgccactacttcatatgtatatgatttagcaggaaggttaataaca
gcatattatccagatggaactagaaggagctataactacgataataatggcaatg
taatcagtatcatcaatccaaaaggatatgtgacaaaatactactacgacaaatt
aaacagggtaatcaaggtagaagacagcaacggcaaagcagtaacatatgagtat
gacggttgtggaaatgtaatttgctttaaagacaaaaaaggaagagaaacaaggt
acgaatatgatgctcttgacagagtaaaaagagtaatagcaccaaacggagcaca
aacagagtatgagtacgatgcagagggcagagttgtcaaggtcacagatgcaaaa
ggaagaagcgaggaatatatgtatgatgagcttggtagagttgttgtatataagg
acaagttaggaaatgtaatcaaatacgcatatgacaaggtgggtaatagaacaca
gcttattgatagaagaggaaatgcaacaaagtatgagtatgataaattgaatagg
gtagtaaaggtgattgacgcgtatggaaatgaaagtaggctggagtatgatgctg
ttgggaataatatagcaaagacagacagaagaggaaacacaacaaagtatgaata
tgatgcaaataacagactagtgacaataattgatccatatggaaataagattaga
tttgagtatgatggagaaggcaatgtaatatgcagaatagatgcaagagggaaca
gaatgtattatagttatgatggattgaacaggttgaggactgtacaagacaacga
tggcaggaaaactgtttatgaatatgatgagaatgggaatatagtaaagataata
cgaccagacggcaagtatgtaacctacaggtatgacagtttggacaggttggtaa
gagttacacaagagaatggagctgtgacagagtataggtatgacgaagaggataa
tttaattgaggtcaaggatgggaatggcaatataacaaggtatgaatacaatgag
atagacagacctgtcaaagtaatagatgcaatcggcaatgaagaaagatatagct
atgacttagtagggaatatagtatatgcgatagacaagaatggagttaggataga
gtacagctatgaccaacttgacagggtagtgcatgttaaagctggcggagtagag
gtgagatatagttatgatgaggaaggcaacagagtccagatgtcagacaaaacag
ggataaacaaatatgaatatgataagctaaataggttgataaggaaaatataccc
agatggcaagagtatagaatatgaatacgaccaagaaggtaatgtaataaaggta
aaagacccgagcggatatgtgacgcaatacaagtatgataagatgaatcggcttg
aagaggtgataacctcagatggtagcacaagatattcatacgatgagaatggaaa
tgtgaaatcaatagagtatccgaataagctgaagtttgagtacgaatatgatagc
aggaacttacttaaaggattagtggttacagcgagagatgttgtaaggaatgaga
tagataagtattatactcctagcagtgtgatagaacaaggtggtggcacaagcac
atatatttacaagtatgaatatggatatgatgacaatggcaatatgatatacaaa
caggagcctataggtaggacagagtacaagtatgatgaggtaggtagagtagctg
aagtcaaagatagatttggacgattaactcgatatgaatatgacaatgtaggcaa
tagaataaaggagatagtagaggtcccgacaggtataagcagagatagtcttaat
tctgaagggctgaatattagatacgattatggagaagtatacagcgttgagaaga
cttacacatatgataaattaaatagactattgagcattgaatctcgagacagaga
agggaatattgtaggtgttaacagctacggctatgacaacaatgggaatttgata
agagcagaggaaagatggcgcacaagagtatacaagggtcagggaagcaaagtag
ttgcaaaagagatagaaaagcaagatttacagactgtatatcaagatgcatatac

Figure 23-03

tgcagagaatgttcaggtacagacaagcgtatataactcagcatatagtttggac
aatagcacagtagtgatttcagtatatgatagtgtgtacaatagtgcatacaatt
tacaaaattcaagttctgagcagagtattactgtaaaagacgtagtttacgagac
ggtatatgaagatgtttatgagatagaagagaagattaaagtaagtgaatttaga
tatgatgagcttggcaggatggtatgggcaaaggttgacaataatgtggtagaat
ttgaatacgacggagatgggctcagaacaaagaagataacagcaaatgatgtaaa
aacatactactggtctggcagcaatcttatttatgaaagcgatgcaacgggcaaa
ggattcagcagcatatggggcttgagtatgataggaaggacggacggtagtaata
cagagtattttatgaaagatgggcatggagatgtacttatcacgtttgacaaaag
tgggcagaggaagaatatatatgaatatgatttgtatggtaatgtaataaaagaa
atagagacagggcaagagaacccgataagatatgctggatattactatgacaagg
aattagaatggtattatttaaagacaaggtattacgatagtaggatagggaggtt
tgtgaaagaggatggcataaagggagatataacagacccagaaagtttgaaccta
tacacttactgtacaaacaatcctgtgaatttatatgatccggatggagagtttg
cgatagtgccactgttagtaggattagcagtgcagactgtaagtggagtattgtt
agattatataatagataggaaaaatttcaatctttggaagagtattggaacaaat
ttagtagtgtcaatggttggtatagtaacaggaggaatagcaagttcaataaagc
ttggtacaaaaataactaaagttgcaccgaaagtagcaaaactaacagaaagatt
agtaagatggggaagtgaaaaagctgttaaaaaacctaactctaaggttgcaaaa
tttataaataaatttacagaagatgttgcaaaacctttattgataaattaggca
ctccaaagacatgggttatggctgggttaaaagcactgacccaaacaggtgcgga
tgcgacgatagatatgatatctggtgaaaaagttaccgcaggatcaatttctctc
gatatacttttgagaactgcatttggtgcgtttggctcagcaacaggaaacgtga
aactatttggaggtggattgaggagaatatttaaattacctccctctaagaaagc
ggaaagttgggatgctgctgctaaaagagtagttgcaaaatatttgatagaagat
ttagaggcgggattatcaaattttggcaaaaagtgttcacttgatgcattaaagg
gtattaacttaggaaggatattgggacgagcaggaagaaagtttacaaaatatat
agtaaataaattggaaatgcaaaagtattcaataacaaatgtttcagggatccaa
ggggaaagtggtggaggagtattagacgagtttattgttcaaaataataaaaaaa
caggcgaggatgtaaaatggacaagaggaagtggagtaagtagtccagcgacaat
tggtatctggggcactgcaggtgaacatgaggttggacaaataaaaggagtagca
gcacaaggattgaattttagctggagcatacaatcaaagatgggagctgatttaa
acttaggggcggtaagaggaggcgagccgaggacgccgacagtagatttaaaatt
cggctatatggcacctatgagaaggatgatagtgccgagggtaggcagggtatta
gcatattga

Amino Acid Sequence:
VRFPKKLRVLLTFLILLIFTLNTNLYVFSQSSNLKNQESNYKDLIGHWAEDEFRW
LIEKGIISGVKANGVMLAKPDEKLTRAEAVVIILRTIFEKELLEEEIKKLKEDSF
KDISNHWSRDYINVAARYGLVKGYSDKTFRPNQRITREEFVLMVIRVSKYREQAA
SEGKQNDKKVKFKDVSVNNFGYNEIIFAVEKGLIKGYSDGTFRPRSYISRAEAAV
IVARALKADLFITYRAKKPQITKENASEFFELVVDKEEVGIIEKAKLTLKSKVKG
VKFTAEWKASGGKLEVAKDNQSAIWSPTDAEEGKNYIVTVEVTILTDNGEKIKIQ
KVARIRVNEGVQLTKNENIEEPVNKIDPLFTQRESNNLVQENSSYSPLEEKLSAP
ILKIAKKGEYVELTWNEVKGARSYIVKRGKVSGEYEVLASGVITTAYIDGPMDGK
TTYYYVVAAVGANGTTSMNSNEVVYKALPQSPKLFGTYNGTSARLTWTKANGAER

Figure 23-03

YTLYRSTVSGGPYYAIAENLLTNVFEDTNLTSGTVYYYVVKAINELGESEYSNEV
AMGEGVTVAAKFNPDDDEDGDGLTNVDELKYATSLKKNDSDGDGLSDGYEVKKGT
NPLVPDTDNDGIYDGAEVVMGTDPLTKNPLTSAEKYAVSEDGKVFVRALSDANIL
IAPLQVKRSDNVFINSLKGIVGKAIEITAGGFDIKKAEIVVNYDEAELNGVEENN
LMLYYVNYDKKILEPLEDVVVDTVYNRVSGKTEHFSTFLLGDKNMPVDLSKVDIV
FVLDNSGSMSSNDPNYYRIEATKKFIQNIDELNNRVGLVDFDSSVSVRSNLTSDK
SKLLQALNAMRWTGGSTNIGGGLKAALGLFDQEQSKKIIVLLSDGYHNTGIHPND
VLPELIKQEIVVNTIALGKDCDRELLHDIADKTKGGYFYVDNTGGLSQEDVDKQI
ELIYEKLTKWITLQKEAEKNLKPQEVLSIEYNDVGLDNEEFHKWITTAMTNLLTG
NYMEEFDDISIEGNGPEFKFVRYYNSFANQQKTIIGKGFRTNFDSKLTKVAGVGI
VQAGVLNVREGPNVNTKKIGYLTKGTKVQIEEDGNKNGSGWHRIVYKGKSAYICA
AYVKELNNGIEVTYPSGSMIVFVDDNGDGIYLSDSNKADKVVLSGNEYILIQRDL
TRYVFDKSGKLIKIGDRNENYITIEYSGEKISAARDVFGRKLEFIFQGDNLVCIR
ENIKGKIGRKVEFVYDDKDRLIKVIGVDGAETRYEYDEKDRLKRIIDANGHQVVR
NEYDILGRIVRQYDGEDIIRFFIYDDEDRVRYYIDENGNESMVVFNEELKPIKER
NALGGGCDYKYEINDGSKWIDVTTPDLDKDVVVSGLTREKYQELKEKGSMTKQVT
IQIIKTSPSLETAKTTQIYDGRGNIIQVIDAYGNSIKMKYDNNNNLIEQTDRIGA
TTKYIYDAEGINLIEKIDPLGNKERYEYYSINSGIKLNGLLAKYIDKNGNETRYY
YEDEYNNLTRVVDAEGYETKYEYDQLGRKIKEINERGYVTRFEYDPEGRITKEID
AFGKTKVYVYDKVGNLIEEIDRLGNKTRYVYDDKNRLIKEIDAMGGEYQYFYDPV
GNKIKEIDPEGRVTKYTYDELNRLVEIEDAEGNKTKFKYDLAGRKISEVNALGKE
TRYEYDLLGRLTKVIDPLGKTRSYQYNAEGYKISETNKNGATTSYVYDLAGRLIT
AYYPDGTRRSYNYDNNGNVISIINPKGYVTKYYYDKLNRVIKVEDSNGKAVTYEY
DGCGNVICFKDKKGRETRYEYDALDRVKRVIAPNGAQTEYEYDAEGRVVKVTDAK
GRSEEYMYDELGRVVVYKDKLGNVIKYAYDKVGNRTQLIDRRGNATKYEYDKLNR
VVKVIDAYGNESRLEYDAVGNNIAKTDRRGNTTKYEYDANNRLVTIIDPYGNKIR
FEYDGEGNVICRIDARGNRMYYSYDGLNRLRTVQDNDGRKTVYEYDENGNIVKII
RPDGKYVTYRYDSLDRLVRVTQENGAVTEYRYDEEDNLIEVKDGNGNITRYEYNE
IDRPVKVIDAIGNEERYSYDLVGNIVYAIDKNGVRIEYSYDQLDRVVHVKAGGVE
VRYSYDEEGNRVQMSDKTGINKYEYDKLNRLIRKIYPDGKSIEYEYDQEGNVIKV
KDPSGYVTQYKYDKMNRLEEVITSDGSTRYSYDENGNVKSIEYPNKLKFEYEYDS
RNLLKGLVVTARDVVRNEIDKYYTPSSVIEQGGGTSTYIYKYEYGYDDNGNMIYK
QEPIGRTEYKYDEVGRVAEVKDRFGRLTRYEYDNVGNRIKEIVEVPTGISRDSLN
SEGLNIRYDYGEVYSVEKTYTYDKLNRLLSIESRDREGNIVGVNSYGYDNNGNLI
RAEERWRTRVYKGQGSKVVAKEIEKQDLQTVYQDAYTAENVQVQTSVYNSAYSLD
NSTVVISVYDSVYNSAYNLQNSSSEQSITVKDVVYETVYEDVYEIEEKIKVSEFR
YDELGRMVWAKVDNNVVEFEYDGDGLRTKKITANDVKTYYWSGSNLIYESDATGK
GFSSIWGLSMIGRTDGSNTEYFMKDGHGDVLITFDKSGQRKNIYEYDLYGNVIKE
IETGQENPIRYAGYYYDKELEWYYLKTRYYDSRIGRFVKEDGIKGDITDPESLNL
YTYCTNNPVNLYDPDGEFAIVPLLVGLAVQTVSGVLLDYIIDRKNFNLWKSIGTN
LVVSMVGIVTGGIASSIKLGTKITKVAPKVAKLTERLVRWGSEKAVKKPNSKVAK
FINKFTEDVAKPFIDKLGTPKTWVMAGLKALTQTGADATIDMISGEKVTAGSISL
DILLRTAFGAFGSATGNVKLFGGGLRRIFKLPPSKKAESWDAAAKRVVAKYLIED
LEAGLSNFGKKCSLDALKGINLGRILGRAGRKFTKYIVNKLEMQKYSITNVSGIQ
GESGGGVLDEFIVQNNKKTGEDVKWTRGSGVSSPATIGIWGTAGEHEVGQIKGVA
AQGLNFSWSIQSKMGADLNLGAVRGGEPRTPTVDLKFGYMAPMRRMIVPRVGRVL

Figure 23-03

AY*

Figure 23-03

Athe Gene Number:
Athe_0013

Product Function:
hypothetical protein

Nucleotide Sequence:
atggagagtagagagagggtagtgaaagcgctgaataagctaataaaggcattgg
cagtaatttggttagtgttaattgttctgaggcttatatggtattttgcagacat
aacttatatatttgagagattatcgatagatacgacagatgttctatcattgttc
ttggaacaaacaataggttctttttggcttatagtgtggctatattgattgtag
ctttatttactgtagtgttattttttttattaagttcaataattacggtgttgat
tgctggtttcacaacaacaatttcagaaaagattgaagcaatttttgtcaagaag
gaatttggtgcgatgcaggtaaatgaagaagtaattgaagctagtttagatatag
atgaagagatttatattgataggagtaagaggaaaataagcgtatttggcagata
tgaaaaggccaaaaaaacggattcacttttaggagcggagcacaggattgatgga
gagtttttttattttaagtgtattaaagtgaaaggagaaaaattatatttaccgt
actcattaattgaaaatagatatgtaaacggatatgttgtaggtaagaatatatc
atactactgtgatggcaatgaatcagaaagaatgttattggtagacgcaattaaa
gaagaaataatagaaaatgaaaaagagcatgttgaagaaaaaatttgtggtgaaa
ttctagcagaaaaatgtgctaaagctgctgcaaagttaattaagagtataacaga
agaaagttttgatactgaaccaagatttaaggatatagtactaacgttattagtt
acaggggtatttttaataacaaattttgtgtcttggttgatatctaagatagcaa
gtagtgagataaatttgcaaggaatattatggattataaatataatgattttagt
gttaatattagtactttacgtagttaagagattttatgaaaaagaaaaaccgctt
aagtttgatatagagatagataaggtacatgtaatattttgtgaaagttctgaag
gaataatagaaacagtaataaaagaaaaagaagcatttgtaaaagatgaatacat
agtgttagtaaaaaaaggagatgaagaaggcgagagatttttggggaaattaaaa
gaagaaaaagaaaaaggattgtatgaagtaagattaatatatggtaccaaaaaag
agaacaaaataaagtattgggtaaaattaaggagttga

Amino Acid Sequence:
MESRERVVKALNKLIKALAVIWLVLIVLRLIWYFADITYIFERLSIDTTDVLSLF
LEQTIGSFLAYSVAILIVALFTVVLFFLLSSIITVLIAGFTTTISEKIEAIFVKK
EFGAMQVNEEVIEASLDIDEEIYIDRSKRKISVFGRYEKAKKTDSLLGAEHRIDG
EFFIFKCIKVKGEKLYLPYSLIENRYVNGYVVGKNISYYCDGNESERMLLVDAIK
EEIIENEKEHVEEKICGEILAEKCAKAAAKLIKSITEESFDTEPRFKDIVLTLLV
TGVFLITNFVSWLISKIASSEINLQGILWIINIMILVLILVLYVVKRFYEKEKPL
KFDIEIDKVHVIFCESSEGIIETVIKEKEAFVKDEYIVLVKKGDEEGERFLGKLK
EEKEKGLYEVRLIYGTKKENKIKYWVKLRS*

Figure 23-04

Athe Gene Number:
Athe_0014

Product Function:
hypothetical protein

Nucleotide Sequence:
atgagcaagatattagtgaaaggtgtaaaaaatccaatatataacagattgcgta
ttaagtcatatttggtaagtttcaaagttttgatagtagttgttactataattat
tgctttttgcttattaactatgcaactgattgggttaatagaattggattggat
gtagatgaaaaaacaaggtcgtttataaatacattggcttttagtttatcgctag
atatcttatccccatgctctatcgatgacactattttgggataccctattacaga
agctattttcacaccaataatatacatggttacttggtttaagggtgatagcaaa
gaaaaaatagaatatttttcaagaaataaatcagtaatagaactgctgttgagca
agcttcaatatgttttaatggacaactctccagaggaaaattttatgaattgggt
gaaatcgaaagatatagatttaaaaaaatatgaaaaggaagatattttaacgttt
tattgggcattccactttttagatgtcaatgtaatagtagtgatattcagtactg
tagtgttgataggatgtattggagcaccagtaagtgtattagtatttacattaat
attcagcatatctacgttaatagaggatataagagatttggtctttaagagaaga
gaagcaattttaaggggagaggaaggaattgaagatgaagatgaaaggaagatat
atgagaaaataaaagaagagataaaaatagcaatttataaggaagaagaagcaag
aaatcaaagttag

Amino Acid Sequence:
MSKILVKGVKNPIYNRLRIKSYLVSFKVLIVVVTIIIAFLLINYATDWVNRIGLD
VDEKTRSFINTLAFSLSLDILSPCSIDDTILGYPITEAIFTPIIYMVTWFKGDSK
EKIEYFSRNKSVIELLLSKLQYVLMDNSPEENFMNWVKSKDIDLKKYEKEDILTF
YWAFHFLDVNVIVVIFSTVVLIGCIGAPVSVLVFTLIFSISTLIEDIRDLVFKRR
EAILRGEEGIEDEDERKIYEKIKEEIKIAIYKEEEARNQS*

Figure 23-05

Athe Gene Number:
Athe_0015

Product Function:
hypothetical protein

Nucleotide Sequence:
atgtatgtatcggcatttatattgttaatatattctttgtatccatttttacatg
aacatgttattaatatacttaaaaagggaataaaagaagagtcaaggctaataag
tatatggcttgtaatggttataatatcttatcttcttaaaaggtatgacattata
actaggtctgatgaaggggatttatga

Amino Acid Sequence:
MYVSAFILLIYSLYPFLHEHVINILKKGIKEESRLISIWLVMVIISYLLKRYDII
TRSDEGDL*

Figure 23-06

Athe Gene Number:
Athe_0016

Product Function:
hypothetical protein

Nucleotide Sequence:
atggagagtagagagagggtagtgaaagcgctgaataagctaataaaggcattgg
cagtaatttggttagtgttgattgttctgaggcttatatggtattttgcagatat
agcgtatgtgagcgaaagaataacaggggattcatcagatggtattatagtattt
tttaagcagatatttttgtacattctttttttgggtttagtaggagtaatattac
cgtttttcatttttctgggggagggtatattccatatttagcattgtatgaact
gttgacaatttttaaaataatagttgtagtaactgtagcagttgtggtaaggaag
ttttgggttgtaaataaatattttgagacaacagcgtttaagagagatttttag
tgaatagtagcaaaaaagaaatagtactaatattaaaatgtgacgaatcatattg
taaatataaaacagaagaagagttttatatttttaagagaattgtgactgaagga
ggaggagagatatatttaccttactctgtaatagaagatagatatgtagatggat
atgttataggcaaaaatatatggtatcatgaagaagattcaggtagaaaaatatt
gatagaagagttattagaagaaaaaatagtagaaaataaagaaaataaaaaggaa
caagctgaagagagaattaatgaagaaatattagcagaaaaagtggcaagagctg
ctataaaattgcttaaattaagaaggcaagagaagacaatggtcagttgtagtaa
tgtttcacccatttattatgtaacagctttaggtaattcaagaggtttttttgcg
aaaataattccatcgttaattgtaggctcgttaattgcggggatatttgtaacaa
tagaatttacatattggttaatacaggatataacgaatagtagtcgtataaccgt
acacaagatcatatggattatgagtatgctactgatggtaatgctattagcaatt
tttgtggtcaaaggtttttataaaagagaaaaaaaagatgaagaaaaacttaaca
tagaaatagtaaaaataccagtgatactatacagggaaaataagaaagctgaggt
aaaagaaggaggtatatttgtaaaagacgacgaaaatatagtgttagtgaaaaag
ggagacaaagaagttgaagagcttttgagtgaattaaaaataaaaaaagaaagag
tagaaggattgtatgaggtgagattaataagtaagtattcttctgatgaagaagt
aaaaagcacgctatggtgcttggtaagaataggtggtgaagcttaa

Amino Acid Sequence:
MESRERVVKALNKLIKALAVIWLVLIVLRLIWYFADIAYVSERITGDSSDGIIVF
FKQIFLYILFLGLVGVILPFFIFSGGGYIPYLALYELLTIFKIIVVVTVAVVVRK
FWVVNKYFETTAFKRDFLVNSSKKEIVLILKCDESYCKYKTEEEFYIFKRIVTEG
GGEIYLPYSVIEDRYVDGYVIGKNIWYHEEDSGRKILIEELLEEKIVENKENKKE
QAEERINEEILAEKVARAAIKLLKLRRQEKTMVSCSNVSPIYYVTALGNSRGFFA
KIIPSLIVGSLIAGIFVTIEFTYWLIQDITNSSRITVHKIIWIMSMLLMVMLLAI
FVVKGFYKREKKDEEKLNIEIVKIPVILYRENKKAEVKEGGIFVKDDENIVLVKK
GDKEVEELLSELKIKKERVEGLYEVRLISKYSSDEEVKSTLWCLVRIGGEA*

Figure 23-07

Athe Gene Number:
Athe_0017

Product Function:
hypothetical protein

Nucleotide Sequence:
atgaataggatagtagtaagagacatagaagacaaaaaaggaaaaagcactatta
caaagattagtacattgcaaaaagctgtagtgttagttgttgctgtaatttttac
tttcttattattcaagtacgcaaaagaggtagttaataggaatgagctaaatgtg
gatgataggataaggcctattacaaaaacattagtttttggattatcaatagaac
ctgggtttataggtaatgaacttttttgactttcctactgtgggatgttatagttt
gattgcatttatatggcggcacgtaacggaagatgacaaaattactctttcagaa
aatttaatagaattgacctcaaaagaaaaacaggcaatagaaatgatgttgaaaa
aacttaactatattttattggaaaatccaaaagaagcttttatgagatggattaa
atcagagaatattaaaatgaacagaaaagagtataaatatgatattttagaagat
ttttgggcatttcatttttagatatgaatgtaatatttatcggagttagtatta
taagttttatttttggactacatatgcattttgcaatatgttatttttaattat
agtaactatggtagagtttgtaataaacttgtgggtagtaattttgaagaaagaa
gaaggagtagaagatgaagaagaaagagagatatatgagatgataaaagaagaag
taagaaaaataatttatgatggaaaagaagcaagaaatttaagacaggaaagcaa
ttaa

Amino Acid Sequence:
MNRIVVRDIEDKKGKSTITKISTLQKAVVLVVAVIFTFLLFKYAKEVVNRNELNV
DDRIRPITKTLVFGLSIEPGFIGNELFDFPTVGCYSLIAFIWRHVTEDDKITLSE
NLIELTSKEKQAIEMMLKKLNYILLENPKEAFMRWIKSENIKMNRKEYKYDILED
FWAFHFLDMNVIFIGVSIISFIFWTTYAFCNMLFLIIVTMVEFVINLWVVILKKE
EGVEDEEEREIYEMIKEEVRKIIYDGKEARNLRQESN*

Figure 23-08

Athe Gene Number:
Athe_0052

Product Function:
ABC-2 type transporter

Nucleotide Sequence:
ttgaatagatttttggcaaactttctgaaatacaaagatcttttatatgagcttg
ttttgagagatataaaaattaaatatagacggtctattttaggtatgttttggag
cttgctaaatcctcttttgatgatgattgttttaacgattgtattttctcacctt
ttcagatttgacataaaaaattatcctatttacttattgaccggacaaattatgt
ttgctttttttctgaatcaacctcaagtgcaatgagggcaattatagataacgc
ttcactcattaaaaaggtatatataccaaaatatatatttcctgtttccaaagtt
ttgtcatctttttttaatctaattttctcacttttggcaattgttttagttacaa
taggaatggtagtattgggcaataatgttaatttaacgtggacatttttgttatt
tcctattcctttaatattcatattaattttttcaataggcataggcttaattctg
tcttgctatgcagtatttttcagagatttgattcatctatactcggttggattaa
ctgcttggatgtatttgacacccatattttatcctgtgagtataattccacaaaa
atatctgatattgattaagctaaatccgatgtattattttatagagtactttaga
aaagttacattttatggtacactacctacactaaaggaaactctaatttgcattt
tggtagacataatatttttagtaataggacttttagtttttatcgaaaacaaaa
taaatttatcttgtatgtgtaa

Amino Acid Sequence:
LNRFLANFLKYKDLLYELVLRDIKIKYRRSILGMFWSLLNPLLMMIVLTIVFSHL
FRFDIKNYPIYLLTGQIMFAFFSESTSSAMRAIIDNASLIKKVYIPKYIFPVSKV
LSSFFNLIFSLLAIVLVTIGMVVLGNNVNLTWTFLLFPIPLIFILIFSIGIGLIL
SCYAVFFRDLIHLYSVGLTAWMYLTPIFYPVSIIPQKYLILIKLNPMYYFIEYFR
KVTFYGTLPTLKETLICILVDIIFLVIGLLVFYRKQNKFILYV*

Figure 23-09

Athe Gene Number:
Athe_0053

Product Function:
ABC transporter related

Nucleotide Sequence:
atggacaaaaacatagctgtaaagattgaaaatgtttcaatgatgtttaatatgg
catctgaaaagatttatagtattaaagagtactttataaaacttgtgtcaggtaa
gctatactttagagaattttgggccttgaaggatataagttttaaaatcaaaaaa
ggtgaaatttttggtataataggcttaaacggggcaggcaaaagtaccctgctca
aaataattgcaggagttttaaagccgacaatgggtagagtctatgtaaacggtac
tatggcaccattgattgaacttggagcaggttttgactttgaacttactgcaaga
gagaatatattcttaaacggcgctattttgggctattcaagaaagttcatgaaag
aaaaatttgacgagatagtagagtttgcagaactgagagatttttagatgttcc
tctaaaaaacttttcgtctggtatgcaggcaagacttggttttgctattgctaca
attgttgaccctgacattttaattgttgatgagattctggcagtaggagactttc
attttcaggaaaaatgcgaaagaagaattaatagtatgcttgaaaaagggacaac
tattgtgatggtgtcccattcaatagatcaaattgagagaatgtgccaaagagtt
ttgtggcttgaaaaaggcagaatgaaaatgataggcgatgcaaaagaggtttgtg
aggcttacaggaactcttga

Amino Acid Sequence:
MDKNIAVKIENVSMMFNMASEKIYSIKEYFIKLVSGKLYFREFWALKDISFKIKK
GEIFGIIGLNGAGKSTLLKIIAGVLKPTMGRVYVNGTMAPLIELGAGFDFELTAR
ENIFLNGAILGYSRKFMKEKFDEIVEFAELRDFLDVPLKNFSSGMQARLGFAIAT
IVDPDILIVDEILAVGDFHFQEKCERRINSMLEKGTTIVMVSHSIDQIERMCQRV
LWLEKGRMKMIGDAKEVCEAYRNS*

Figure 23-10

Athe Gene Number:
Athe_0054

Product Function:
Nucleotidyl transferase

Nucleotide Sequence:
atgaaaggaatagttttggcaggaggaacaggttcacgtttatatccgttaacta
aggttacaaacaagcatttgttgcctgttggaaaatatccaatgatttattatcc
aatttttaaattaaaacaagcaggtatcaaagagataatgataataactggtaaa
gaacacatgggagctgttgtcaatttattaggaagtggtcgtgagtttggattag
aatttacttataggattcaagatgaagctggtggaattgctcaagctctgggact
gtgtagctttttgctgggaatgataagtgtgttgtgatactgggcgacaatgtt
tttgaggatgatataactgagtatgtaaggaatttcgaaaagcaggagcggggcg
ctaggattcttcttaaggaagttccggacccgcatagatttggagttgctgaact
gaaagatgggaaaatagtctcgattgaagaaaagcctaagaatccgaaaagtaat
tttattgtcacaggaatctatatgtatgatagccaagtttttgatattatcaaga
cacttaagccttcacagaggggtgaattggaaattacggatgtaaataacgagta
cataagaagaggagagttgtattttgactttttaaagggctggtggactgatgca
gggacttttgagtcactaaaaagagcaaatgagcttgctgaaaatatggtgttag
attttaatggaatataa

Amino Acid Sequence:
MKGIVLAGGTGSRLYPLTKVTNKHLLPVGKYPMIYYPIFKLKQAGIKEIMIITGK
EHMGAVVNLLGSGREFGLEFTYRIQDEAGGIAQALGLCSFFAGNDKCVVILGDNV
FEDDITEYVRNFEKQERGARILLKEVPDPHRFGVAELKDGKIVSIEEKPKNPKSN
FIVTGIYMYDSQVFDIIKTLKPSQRGELEITDVNNEYIRRGELYFDFLKGWWTDA
GTFESLKRANELAENMVLDFNGI*

Figure 23-11

Athe Gene Number:
Athe_0055

Product Function:
dTDP-4-dehydrorhamnose 35-epimerase related

Nucleotide Sequence:
atggagcttattgaaggtgttaaggtaaaaagactcgagaaatttgctgatgata
gaggtttttcatggagattttgagagacgaagatggctttttagaaaaatttgg
acaggcttcaatgtctttgacttatccaggggtaataaaagcttttcactatcat
aagttgcaagacgatgtttggtttttcccaaaaggcaatgcacaggttgttcttt
acgatttgcggccagattcgcctacatacaaaaaaacaaatgtattctacatggg
agaacacaaccctatagtacttttaatacccagaatggttgctcatggctatagg
gtattggggaatgagccggcaataattgtgtattttacaacagagcattataata
gagaaaatccagatgaatacagaattccttgggatgacaaggaaattaattttga
ttggacaacaaaatttagataa

Amino Acid Sequence:
MELIEGVKVKRLEKFADDRGFFMEILRDEDGFLEKFGQASMSLTYPGVIKAFHYH
KLQDDVWFFPKGNAQVVLYDLRPDSPTYKKTNVFYMGEHNPIVLLIPRMVAHGYR
VLGNEPAIIVYFTTEHYNRENPDEYRIPWDDKEINFDWTTKFR*

Figure 23-12

Athe Gene Number:
Athe_0056

Product Function:
dTDP-glucose 4,6-dehydratase

Nucleotide Sequence:
atggagacaattcttgtagcaggtggggcaggttttataggaagcaattttgtta
agtacatgattagcaaagaagaatacaaaataatcaattatgatgcattaaccta
tgcaggaaatcttgagaatttgaaagaggtagaaaaccacccttattatacattt
attaaaggagatattgttgatagatctaaagttgaagaggtttttaaaaattatc
aaattgactatgtaataaactttgccgcagagtcgcatgtggacagaagtataaa
ggaccctgatatattcgttaaaacaaatgttttgggaacacaagttttattagat
gtgtcgaggaaattcgggataaaaaagtttattcaaatttcaacagatgaagtat
atggttccttagggcctgaaggatattttacagaagaaagtccgcttgcaccaaa
cagtccttattctgccagcaaagcaggggctgatatgcttgtgagagcatatttt
aagacatatggtctgcctgtgaacataacaaggtgttcaaacaattttggtccac
atcaacacccagaaaagtttataccgaccgtaattttgaatgcgctgcaaaacaa
gccgataccaatttatggtgacgggcaaaatataagagactggctatatgtagaa
gaccactgcagagcaattgagcttgtgctcaaaaaaggtagaataggtgaagtat
acaatattggcgggaataatgagtggaggaatatagatatagccaaattgatttt
aaaactactcgggaaaccagagaatctaatacaatttgtggctgacaggccagga
catgataggagatatgcaattgactcgagtaagattcaaaaggaattggggtgga
aggttgaatataagtttgatgaagctattagaaagactatagaatggtataaaaa
tgaatttttcaaaggagaatga

Amino Acid Sequence:
METILVAGGAGFIGSNFVKYMISKEEYKIINYDALTYAGNLENLKEVENHPYYTF
IKGDIVDRSKVEEVFKNYQIDYVINFAAESHVDRSIKDPDIFVKTNVLGTQVLLD
VSRKFGIKKFIQISTDEVYGSLGPEGYFTEESPLAPNSPYSASKAGADMLVRAYF
KTYGLPVNITRCSNNFGPHQHPEKFIPTVILNALQNKPIPIYGDGQNIRDWLYVE
DHCRAIELVLKKGRIGEVYNIGGNNEWRNIDIAKLILKLLGKPENLIQFVADRPG
HDRRYAIDSSKIQKELGWKVEYKFDEAIRKTIEWYKNEFFKGE*

Figure 23-13

Athe Gene Number:
Athe_0057

Product Function:
methyltransferase FkbM family

Nucleotide Sequence:
atgagaagtataaaccaattaaatattggtctaacagagacaaagtacggcaaat
tctattatttaaagagtgatgcaataataggcaaatctttagaaatatacggaga
atgggcggtaccagaaatagagttgttaacaagttttatagaagaaggggatata
gtaattgatgtaggcgcatacataggaactcatactattccatttgctcaaaagt
tgaatggaaaaggttttgtaattgcattcgaaccccaagagattatattcaatat
actagcaaagaatataaggacaaataacgcagataatgtgcaaatatttaacaaa
gctgttttagataaaaatacaataacctacatagaaacttttaactacaatgaaa
ctaataattttggcagtgctaaaataataactgatgatattgaaactcaaggaat
tatcaaaaaaattgaagcaatcactattgatagtttagaattgaataattgtaaa
ttaataaaaattgacgtagaaggacaagaagaatttgtattgagaggctcagaaa
agacaataaaaagttttatgccaattatttattttgaaagtaatgagcttgagaa
aacatggaatagtatttgcttagtaaagaaatggggatacgatactttcttgttt
agatttccagcatttaatccatttaacattaaaggtacacaagataatatttttg
gttatgcatgtgaaacaggtattttggggattcataaatcaaaactgagtactta
tggaaatgatactttgaaatggcgaaaattaaattatctatttgaaattaccacg
cttgacgatttagctttttattaatagaaacttctcgtggacatgatcccatta
tacctgatattaaatataatgaaatcagaaacataaagtttcttcaaaattatat
aattgaggtggtaaaagaaattgaaaaaaaagataaactattacaaacaattaaa
gaagatttaaaaataaaagaacaacacttacaaagtatttattgttcggaaggtt
ggaaactgttaaccaagtgttataaaataagagataagatattcccaccgaatag
taagagaagagaattagcaaagtttattttatttattatgaaaaaactgaattta
aaatttgttcaggaaataataagatttgcaaggatgtatggtattaaagctcttt
ataatcgaagtcgaattgctttgcactctatgaaaagttattcgagtaaccaaga
aaattattttgttcctgaacaatttgaaagagattatcttttaagtgaaatcagt
gttgatataattgtacctatttataatgcttatgaggatttaaagaggtgtgtag
aaagtattcttaagcacactgatttgaaaaaacatagattggtattaattaatga
ttgttcgactgatgagagaatatattctttcttaaaaaatcttgagaacgaaaga
accgaagaaaatttacttgttattcataataaggagaatttgggttttgtaaaaa
cagttaataaaggaattagtttgagtaataaagatgttattattcttaactctga
tacaattgtaactgcgagatgggtagaaaagcttatcagagcagcgtattcaagg
tctaatgttgcaacagttacacctttttcaaataatgccacaatttgttcattac
caataatgttaaaagataattcgcttcctttggattgggatattgattattttgc
taaagttgtagataggatttcacttttaaaatatcctgaaattccaacagcagtt
ggattttgtatgtatattaaaagagaagttattgaaaaaataggaatgtttaacg
aagagaaatttggcaaaggttatggagaagaaaatgatttttgtatgcgagcgct
aaacgaagggtatgtaaatattttatgtgataatttatttatttatcacaaagga
agtcaaagctttacggaagaagttaaaaggaaaagagaaatggagagcttaaagg
ttatcaatcaattacatccattttactctgagatggttaaggcctttatagaaaa
aaatcctttaaaatattatcattctacacttgaagaaatcatgtccctttataac

Figure 23-14

ttattaaaaagtgaggagaaaaagtga

Amino Acid Sequence:
MRSINQLNIGLTETKYGKFYYLKSDAIIGKSLEIYGEWAVPEIELLTSFIEEGDI
VIDVGAYIGTHTIPFAQKLNGKGFVIAFEPQEIIFNILAKNIRTNNADNVQIFNK
AVLDKNTITYIETFNYNETNNFGSAKIITDDIETQGIIKKIEAITIDSLELNNCK
LIKIDVEGQEEFVLRGSEKTIKSFMPIIYFESNELEKTWNSICLVKKWGYDTFLF
RFPAFNPFNIKGTQDNIFGYACETGILGIHKSKLSTYGNDTLKWRKLNYLFEITT
LDDLAFLLIETSRGHDPIIPDIKYNEIRNIKFLQNYIIEVVKEIEKKDKLLQTIK
EDLKIKEQHLQSIYCSEGWKLLTKCYKIRDKIFPPNSKRRELAKFILFIMKKLNL
KFVQEIIRFARMYGIKALYNRSRIALHSMKSYSSNQENYFVPEQFERDYLLSEIS
VDIIVPIYNAYEDLKRCVESILKHTDLKKHRLVLINDCSTDERIYSFLKNLENER
TEENLLVIHNKENLGFVKTVNKGISLSNKDVIILNSDTIVTARWVEKLIRAAYSR
SNVATVTPFSNNATICSLPIMLKDNSLPLDWDIDYFAKVVDRISLLKYPEIPTAV
GFCMYIKREVIEKIGMFNEEKFGKGYGEENDFCMRALNEGYVNILCDNLFIYHKG
SQSFTEEVKRKREMESLKVINQLHPFYSEMVKAFIEKNPLKYYHSTLEEIMSLYN
LLKSEEKK*

Figure 23-14

Athe Gene Number:
Athe_0058

Product Function:
glycosyl transferase group 1

Nucleotide Sequence:
atgagcacaattgtttatttagtccattcaatacccaaatatgaaaagagtggaa
ccccaatagctgcttggagagtggcaaagggtgtaaaagaaaaatataatcaaaa
tgtagcatttataattccttcacctgatggagaagagggaaaagaaaaggttgac
gatattttggtttacaaagtaaaaagaatagattggcatgaaaattttttcatg
attttgatatagatagagaaacgtatattagaaaaataaagaatattttaaaaga
agttaattgtaatattttgcatatttataatttagtatttagctcttatcaggta
atgaaacttaggaaagaaggaatcagaattgttcgtacgataacacatactgagg
atatttgttttaatgttgatccttttgttaaagttggtgataaaattgagatatg
tagtgggcctgatccgatagcaaaatgcgcttgtcattataaacagatgtatggt
ggtaataatttgatggagtttataattaaaaagatatctaagcattttacgagcg
tagaattattatattctaatttttgtgatataataacctttactaatgaagaatt
tgctaaatattttaccaactacgtaaatattccaagagatgtaatccgaataata
ccacatggagtagaaaataaattagaaaaatatatattgccgaatatgccaaaaa
acgaaggatttagatttttatatcttggaggagataattttagaaaaggatttgt
tatattagataatgcgttaaattctttaaatggggagttattcaacaagattaag
gaaataactatagtgggtaaaacaacaaaggaatttagagaaaggtttaataatg
ataaatatatgtttaagggggtattgccagaagaagaattatataacgaaataag
taatgctgatcttgtgatattgcccacatttttcgaaacatataacatatcttta
agagaagctatcaagttaggaaaacctgttataactactaaaacttttggttcga
atattgtagtggatggatataatgggtttagatttgatattggtgatagcttaca
attaaaaaacataatagaattaatattaagcaatccccaaatactggtagatatg
agcaaaaattgtttaaatactcatataactgatattgaagaagaaataaagttat
ttatgaaagtttacaatgaattaaaagattag

Amino Acid Sequence:
MSTIVYLVHSIPKYEKSGTPIAAWRVAKGVKEKYNQNVAFIIPSPDGEEGKEKVD
DILVYKVKRIDWHENFFHDFDIDRETYIRKIKNILKEVNCNILHIYNLVFSSYQV
MKLRKEGIRIVRTITHTEDICFNVDPFVKVGDKIEICSGPDPIAKCACHYKQMYG
GNNLMEFIIKKISKHFTSVELLYSNFCDIITFTNEEFAKYFTNYVNIPRDVIRII
PHGVENKLEKYILPNMPKNEGFRFLYLGGDNFRKGFVILDNALNSLNGELFNKIK
EITIVGKTTKEFRERFNNDKYMFKGVLPEEELYNEISNADLVILPTFFETYNISL
REAIKLGKPVITTKTFGSNIVVDGYNGFRFDIGDSLQLKNIIELILSNPQILVDM
SKNCLNTHITDIEEEIKLFMKVYNELKD*

Figure 23-15

Athe Gene Number:
Athe_0059

Product Function:
glycosyl transferase family 2

Nucleotide Sequence:
atgagaaatttaacagttataatacctacatataatcaaaaagatttattagaaa
gagcaataaattcgttaatatctaagtgcaatgacgaaataaatatatatgtatt
agtcaacaatacggaaagtaattacatcgtaagaaataaaaattatgcaaatcta
catgttgaatatttgaacacaaactgtggtttttgtaaagcagtcaattatggat
tacgattaataaaaaaatcaagatttatttttctttaaacgatgatactgaagt
aataaatcagatagatatagataatattattcatgaattgattgaaaaaggcaac
atttttctatatcgttgaaaatgctgaagggtaattatccaaatcttttagacg
acgcaggtgatatgtacactatcttaggttggcagtttaaaagaggcaatggtct
tccaaaagaactttatgatagaccgtgtgaaattatttccgcatgtggtggtgct
gcaatctataacaaaaaaattcttgatgagataggttacttcgatgaagattttt
ttgcatatcttgaggatgtagatttaggtttgagagcgctcatgaggggatataa
aaatttatattatccttacataagcgtattgcatgttggaagtgcgacaacagga
ggaaaatataacgatattactatcagacttacagcaagaaactcgatatatgtta
tatacaaaaatcttcccttaccccttttaataattaattttccctttattttatt
gggatacttaatcaaatttatattctttgctaaaaaaggaaaaggaaaaatttac
ataagtggagttcttgaaggactaaaaaatttgcccaaatttaaagaaaaaagaa
gagaaaatatgagaaaaaagaaaatttctaacataaagcttgaatggattcttat
taaagctacatttgaatattttcaccaatatataaaaagagcattttacacttta
agaggtgcaaagaaatga

Amino Acid Sequence:
MRNLTVIIPTYNQKDLLERAINSLISKCNDEINIYVLVNNTESNYIVRNKNYANL
HVEYLNTNCGFCKAVNYGLRLIKKSRFIFLLNDDTEVINQIDIDNIIHELIEKGN
IFSISLKMLKGNYPNLLDDAGDMYTILGWQFKRGNGLPKELYDRPCEIISACGGA
AIYNKKILDEIGYFDEDFFAYLEDVDLGLRALMRGYKNLYYPYISVLHVGSATTG
GKYNDITIRLTARNSIYVIYKNLPLPLLIINFPFILLGYLIKFIFFAKKGKGKIY
ISGVLEGLKNLPKFKEKRRENMRKKKISNIKLEWILIKATFEYFHQYIKRAFYTL
RGAKK*

Figure 23-16

Athe Gene Number:
Athe_0060

Product Function:
dTDP-4-dehydrorhamnose reductase

Nucleotide Sequence:
atgatactaatcacaggttccaaaggccaacttggaagcgaatttataaagcagt
ttgaaaataaatataatgtgaaaggtattgacattgaacaagttgatattactga
tcttgacagtacagttagttatatttcagccacaaagcccaacatcatcatccac
tgtgcggcttacaccaatgtagatgggtgtgagagtgacaaagacactgctttta
aagtaaatgccattgggacacgaaatgttgcaatggctgcagaaaaagttggtgc
aaagcttgtatatatatctaccgattatgtatttgacggtgaaaaagagaagcca
tataatgaatttgatagaccaaatccgataagcatatatggcctttctaagcttg
cgggagaggaatttgtaaaaactttttgtagcagatattttatagtgaggattgc
ctggctgtatggtgaaaatggtaacaactttgtaaaaacaattgtaaaacttgct
aaagaaaaaggcgagattgatgttgttaatgatcagagaggcaatcccacattta
ccaaagatgttgttcaagctgtagaagtaataatgaactcagagaaatatggaac
ataccatgtaacaaatgaaggcataacttcttggtacgattttgcatataagatt
gtaagcacgtttgggatagactgcaaagtcaatccaactacaagtgataaattta
ttcgcccggcaaaacgtcctaagaactcagcacttgacaagatgatgttaagact
tgaatttggatataaaatgagacattgggaagaggcttttgaagagtttgctatg
ctgatgaaaggaagaatatag

Amino Acid Sequence:
MILITGSKGQLGSEFIKQFENKYNVKGIDIEQVDITDLDSTVSYISATKPNIIIH
CAAYTNVDGCESDKDTAFKVNAIGTRNVAMAAEKVGAKLVYISTDYVFDGEKEKP
YNEFDRPNPISIYGLSKLAGEEFVKTFCSRYFIVRIAWLYGENGNNFVKTIVKLA
KEKGEIDVVNDQRGNPTFTKDVVQAVEVIMNSEKYGTYHVTNEGITSWYDFAYKI
VSTFGIDCKVNPTTSDKFIRPAKRPKNSALDKMMLRLEFGYKMRHWEEAFEEFAM
LMKGRI*

Figure 23-17

Athe Gene Number:
Athe_0061

Product Function:
glycosyl transferase family 2

Nucleotide Sequence:
atggatttgtctataattatagttaattacaataccaggaatttgctcagaaaaa
ctcttgagtcaatatataagaatcctacctatagagaatttgaaataattgtggt
agacaatgcatcaagtgatggcagccaagagatggtaaaaaaagagtttccaaat
gttattttgattaaaaataaacaaaatttaggatttgccaaagccaacaatatag
ggataagaattgcaaagggaaagtacatattgcttttaaattcagatacagaagt
tttaaggggtacacttgatagttgtatagatttttggagaaagatgaagctaaa
gaaattggaattcttggatgtaaagtggtgcttccagatggcaagcttgatttag
catgtagaagagggtttcctactcctaagaattcgtttttaagatatttgggct
tgcgaagctgtttcctaaaagtcgatttttgcaggttacaatcttacatatctt
gatgaaaatcaatcttatgaagttgattcagttgttggggcgtttatgttaataa
ggcgcgaagttatagataaaataggtttacttgatgaggattatttcatgtttgg
cgaggacatagatttttgttttagagccaaacaaaatggatttaaagtttactac
tatgctgatgcaaaaatcattcatcacaaaagaggttctggcaggaatttgaagg
ttttgtcagctttttatgactcaatgtggatattttataaaaaacattactacaa
cagatacccaaaaaccttagccttattaatatttataacaattaagttgattaaa
tcattaaagcttgcccatgcgaaagtacggaactttcctattagaaagggaaaaa
aaccatgcatagcctaa

Amino Acid Sequence:
MDLSIIIVNYNTRNLLRKTLESIYKNPTYREFEIIVVDNASSDGSQEMVKKEFPN
VILIKNKQNLGFAKANNIGIRIAKGKYILLLNSDTEVLRGTLDSCIDFLEKDEAK
EIGILGCKVVLPDGKLDLACRRGFPTPKNSFFKIFGLAKLFPKSRFFAGYNLTYL
DENQSYEVDSVVGAFMLIRREVIDKIGLLDEDYFMFGEDIDFCFRAKQNGFKVYY
YADAKIIHHKRGSGRNLKVLSAFYDSMWIFYKKHYYNRYPKTLALLIFITIKLIK
SLKLAHAKVRNFPIRKGKKPCIA*

Figure 23-18

Athe Gene Number:
Athe_0077

Product Function:
S-layer domain protein

Nucleotide Sequence:
ttgaattcaaagaaagcttttagactgctatcatgggtagtgataatatcttttg
tattgggtttataaacccagctgtttttgcaaaaggttttaaagacacatcaaa
tcactgggcaaaagatgtaattgagagatgggcaaatacatacaatgtggcaaat
ggttattctgatggcacttttaaaccctcaaatgctattacaagggctgaatttg
ctcagcttgtgagcagagtaattggtgaggctttagtcaggtctgaaattaactt
taaggatgtaaaagaaaatgattggttttactcagccgtcaaaaatcttgctgat
tatataagtgggtatccagatggtacatttaaaccaaagaacagtataacaagag
aagaagcagcatgtattttggcaaaggtatttggtattgacaaatcacagagcaa
tgttctttcgaagttttctgattataaacaagtgtcagagtgggcaaaagagtat
ttagcagctatggttgaaaatggatatattaatggctataaagataaaacattaa
gacctaaaaattacattacaaaagctgaggcactgactattttagataacattgt
tgggcttttgttatcaaaaaagggtatgtatgtaggaagagaagtcaagggaaat
ttaattataagtgcaccttatgttagtgtgtcaggatttaatgtaagtaaaaatt
gcttgataacagaaggagtaaaagatggaaatgttacaattactaatacgcaggt
agacggtaatatcattgttcgtggcggaggagaacacagtgtcgttttgaaaaat
gtcaaagcttctgcagttatagttgtgaacaaacaagcagctacaaaagtaaata
ttagtggaaattcaaagatagataaggttgtaattgaaaagccagctaatatttt
agttgagaagactgcagaagtgaacacagtggatatcagagctgacagaacaatt
tttaaggcagaaggcaaaatagagaggatagcagtaaatgcaaaagatgtaaaag
taaatgataaggaggtccagcaaggcgctatactaacaattactaataccaatga
aaatgagaaaaagactcaatctacaaataccaatacttcgcaggagtcaacacaa
aatgaaactacgtctcaaacatctaataatacttcaagtgtgcaaacaaatactg
gaacatcaactggaaatatatcttcaggcacttcaactggtggctatacaggtgg
aggaagtgtatatagtggcggttcttcgggttctggaacaagtgtccaatatgga
cctgtaagtagagtagaggttgacaaagaaaccatccctattggtcagaacgtcc
aagtaacagttacagtgaaggatgcagcggggaatttattgccaaacaaggttgt
aaggattgaaggacaggctgcttatacaaattccttaggtagtgcaacatttact
ttatcagttcaggatagcaaagaaatagttataaatgtagatggcaatgactatt
atggtcttttgtacgcaattaaaccaactgaaggagttttgacatttaagttaaa
gacggcaagtggtaactttttaagtggatttaatgtgaaactagtaaatcaagct
aagcaattttctgagttaaagtctgccacagctgaacaagtgtcattcattgtac
cagcagttgatggatacaaagctttaatatggtcatatgatagtcaaaatgggct
tatatataccatcttagataatctgtcagttggaaacggtattgtgagacttgtt
gcagacacaacattaccaaattatgtacaagctacactagattttagcataaata
accaacccctcaataactatgagttttcaattatcaatagctcaaaatctaatac
tttttcaattgatgatgtaaagctaaatatttcttcaaatcagcttaaaattaca
gcagatgcgggcagttattcaatcaaggtagcaaaagatactcaggatggtaagc
tatattttatgagagattttaatctccaacagtcaggacaagctttttcatttaa
tttcagctcttttaaaagagttatatttaatttttctggtattggcactttacaa

Figure 23-19

aagagtgtatctgtaaaattaaatggtagctggtttgaactctccggggaaaatg
atatatatctgcaaagaggagtttataacctggaagaagtttttataaagacata
cgatgagaacagtaaagaggttgattattcatataaggtgccagcaggcagtagt
ccaattgtagttgatgcgacaggaagtaacgaagagtgctctgtgaatgtagacc
tttcaatagataatgtctcaagcagtgtatatgctactaatatagatggaaatga
atttgcttctataaaggcaggttcacttatagattttgtagtggttttaaagacg
aaaagtgggctgaacttagctttgttaaagccatctgaaaagatgtcagatacag
caccattgtttaaagcaagtgatattgtggcatacatatcaagtggaaatgtaca
gaataatattgaacttctttgtacatcggtagaagataataatatttcgcatgta
ctcggatatttgccaaggaatcttcaagatggtactgttagcgtagaatttacgg
cagaaataggtttgctttatggaagctctattacaagcagtatttcgctttctat
agacaatcaaaatggtgacagtagaggttatgtatatccacaaaatacaactaca
agtgctgtatactttgtatgtgataaaattgacagtggagcaagtgactttttag
cacttagtcttccggcaagtgaaacttgtgacgtcaatcttattgccaacaaaat
atattttgaactcatgacagaaagaccgcttacttttgaatacggattgaatgga
atgtttggactgagctttaatctcaccagtgaaaacctctacatgatatggcctg
tttatcttttcagcaaagacgagggtcttacaagaagacaagctgttgaacaaaa
gtccttgcaaatagtatctacagttgtagaccaaacttacaatgactatgataaa
gttttaggtttgcatgactggcttgttttgcacacccagtacgatttggaagggt
atcttaacaacaacataccttatgagtcacatacagcttatggtgcactaattaa
tggtattgcagtttgcaacggttatgcaacagctatgcttgcgcttttagaggat
gcaggtatagactcaaaagagatctatggtatggcaggtgtaggtaattcaaaag
aagcccatgcatggaacatggtaagcttagaaaataattggtatcatttagatgc
aacatgggatgacccagactggggcaattatgttgaacatgccttttcaatgtt
ccagatagtaaaattgagttaacgcatgattgggaaagaaacctttatcctgctg
ctacagctatagattacagctatggtaactattacaatgtagaaaccattccaca
ggatgtttattacaacgaagacaatattgtaacaatagttgtcaaaaactataaa
ggtgaactacaagccaataaacttattacaataacgaaatttaaccaagcaggag
ttgaagatattgtatttactggctatacttctgctgatggtacaataacattcaa
tttaaagcctactgagatgatatcttaccagatttatataacctcttttatggaa
gacaaaggatacattaaagttgttgaaaaactgaagccggtgacattaagtttaa
atattgacaatcaaccaattgatgaattttatatctcagatggtttgaacttctt
gaaagtcacgggtggtcagcgacaggttggacttagcaggtggcgcgagaatgat
ctaattttctatggtttaggttttgtaataaagaaaagtttaacatttgacagct
atgacccgatccaacttacagtttgtaacgatgcatacgattcttatgtactttt
aagtgtgtataacagtgcttatgcagttgtcggtgcagacgtttatgtgattgac
aaagataccttcaaggagctgtttgtaggcagcacagactcagaaggaaatcttc
cgcttgttcttacaaacggagaatacattttaaaagttgctaattataattcaga
aacaaattctcaagatttttactattcgattttacaggtaacacaagatagtgta
acttcaattgatttgtctcagtttaatcagattcaatatataccaggatttaccg
aaggcggcgttggaatgggagacaagctgttacttggtttcgactttgactatga
tggaaaatttgttgtgtcatcgatggggcactcgagaaaggcttattttgcaccg
ggtgattatggttgggtatggattttggcccataacccgtatgatatagaagata
atgttgctacatcattagtgtataaacctttgcaaaaactaattattccagaagg
tacgcttcagcaatatactgttgatttgagagttgacagaagcaaaatagtcttg
aactgtgaacgtcgaccagaaggaagctgggattttcttccaacaacgcttgata

Figure 23-19

tggtatatgaaaaggatgacattataattgatatttatattccaaccaatagtca
gagctacgttgtgggtggggatactgtacctcagcagcttatagaaaataatgga
gatatgcatattatgtggatgggtattgctggcatttatccaattagaacgtata
ctatttcaaatagtactccatatgaattgtttacaggacatcttggttttattga
caatactaatagttggcgtataattgtacttgatgtagtaggtgatctaaattca
acatttagaatagtattaccactatcaccgtatgacattgatgttattatcgaca
aagaaattacaatttacccacttactgcaggatcaggtggtttgcgaaaagagtc
atatattcaaaaggctaagatttacaatcaaataatcaacacacagcttattaaa
aaagagaataaaaaataa

Amino Acid Sequence:
LNSKKAFRLLSWVVIISFVLGFINPAVFAKGFKDTSNHWAKDVIERWANTYNVAN
GYSDGTFKPSNAITRAEFAQLVSRVIGEALVRSEINFKDVKENDWFYSAVKNLAD
YISGYPDGTFKPKNSITREEAACILAKVFGIDKSQSNVLSKFSDYKQVSEWAKEY
LAAMVENGYINGYKDKTLRPKNYITKAEALTILDNIVGLLLSKKGMYVGREVKGN
LIISAPYVSVSGFNVSKNCLITEGVKDGNVTITNTQVDGNIIVRGGGEHSVVLKN
VKASAVIVVNKQAATKVNISGNSKIDKVVIEKPANILVEKTAEVNTVDIRADRTI
FKAEGKIERIAVNAKDVKVNDKEVQQGAILTITNTNENEKKTQSTNTNTSQESTQ
NETTSQTSNNTSSVQTNTGTSTGNISSGTSTGGYTGGGSVYSGGSSGSGTSVQYG
PVSRVEVDKETIPIGQNVQVTVTVKDAAGNLLPNKVVRIEGQAAYTNSLGSATFT
LSVQDSKEIVINVDGNDYYGLLYAIKPTEGVLTFKLKTASGNFLSGFNVKLVNQA
KQFSELKSATAEQVSFIVPAVDGYKALIWSYDSQNGLIYTILDNLSVGNGIVRLV
ADTTLPNYVQATLDFSINNQPLNNYEFSIINSSKSNTFSIDDVKLNISSNQLKIT
ADAGSYSIKVAKDTQDGKLYFMRDFNLQQSGQAFSFNFSSFKRVIFNFSGIGTLQ
KSVSVKLNGSWFELSGENDIYLQRGVYNLEEVFIKTYDENSKEVDYSYKVPAGSS
PIVVDATGSNEECSVNVDLSIDNVSSSVYATNIDGNEFASIKAGSLIDFVVVLKT
KSGLNLALLKPSEKMSDTAPLFKASDIVAYISSGNVQNNIELLCTSVEDNNISHV
LGYLPRNLQDGTVSVEFTAEIGLLYGSSITSSISLSIDNQNGDSRGYVYPQNTTT
SAVYFVCDKIDSGASDFLALSLPASETCDVNLIANKIYFELMTERPLTFEYGLNG
MFGLSFNLTSENLYMIWPVYLFSKDEGLTRRQAVEQKSLQIVSTVVDQTYNDYDK
VLGLHDWLVLHTQYDLEGYLNNNIPYESHTAYGALINGIAVCNGYATAMLALLED
AGIDSKEIYGMAGVGNSKEAHAWNMVSLENNWYHLDATWDDPDWGNYVEHAFFNV
PDSKIELTHDWERNLYPAATAIDYSYGNYYNVETIPQDVYYNEDNIVTIVVKNYK
GELQANKLITITKFNQAGVEDIVFTGYTSADGTITFNLKPTEMISYQIYITSFME
DKGYIKVVEKLKPVTLSLNIDNQPIDEFYISDGLNFLKVTGGQRQVGLSRWREND
LIFYGLGFVIKKSLTFDSYDPIQLTVCNDAYDSYVLLSVYNSAYAVVGADVYVID
KDTFKELFVGSTDSEGNLPLVLTNGEYILKVANYNSETNSQDFYYSILQVTQDSV
TSIDLSQFNQIQYIPGFTEGGVGMGDKLLLGFDFDYDGKFVVSSMGHSRKAYFAP
GDYGWVWILAHNPYDIEDNVATSLVYKPLQKLIIPEGTLQQYTVDLRVDRSKIVL
NCERRPEGSWDFLPTTLDMVYEKDDIIIDIYIPTNSQSYVVGGDTVPQQLIENNG
DMHIMWMGIAGIYPIRTYTISNSTPYELFTGHLGFIDNTNSWRIIVLDVVGDLNS
TFRIVLPLSPYDIDVIIDKEITIYPLTAGSGGLRKESYIQKAKIYNQIINTQLIK
KENKK*

Figure 23-19

Athe Gene Number:
Athe_0088

Product Function:
hypothetical protein

Nucleotide Sequence:
gtgacagacttatatataaccattttatttgtctcgataattatagcatttggga
ttgaaattttacaaagtaaaacctttcataaaagttgtttaacaggagacaaaaa
tttatttctgctcatagggatgactttttaggaattcatttgtttataacattt
ccgggcatgatttcgataaagatatggatgatgctaaatcttattgtattgtgga
tcttgtcatattacagagaaactataaaaaaagaaatattattcaagaatatttc
taaacttttaaaatctcaaagaagaaaaatatttttataagtagcatattgtgt
tacataattgtctgtcttcttttaaattattacatagtttacaagttcatatatt
catctgcagtttcaaaattaacaataacattcactatttcctttcttatactaac
agaagctaaaaatatagcttgttttattacaattttaaaattttattgcatatct
ttggtaacgatttggggagaaaatttagtgaaatatgagggatggcttattggag
caacaagagattattacatattaaaagaaaaattttcggggaaattaataacagt
aaaaaaagatatagttaatcaaatagtgataatgggaaaagtttttgaaagataa

Amino Acid Sequence:
VTDLYITILFVSIIIAFGIEILQSKTFHKSCLTGDKNLFLLIGMTFLGIHLFITF
PGMISIKIWMMLNLIVLWILSYYRETIKKEILFKNISKLLKSQRRKIFFISSILC
YIIVCLLLNYYIVYKFIYSSAVSKLTITFTISFLILTEAKNIACFITILKFYCIS
LVTIWGENLVKYEGWLIGATRDYYILKEKFSGKLITVKKDIVNQIVIMGKVFER*

Figure 23-20

Athe Gene Number:
Athe_0089

Product Function:
Endo-1,4-beta-xylanase

Nucleotide Sequence:
atgaggtttaaaaagtttttaaaagttttgatagctgttttaatgtgctttatgc
ttggaaatccgttttatgcgcaagctgcaataaccctcacatccaatgctagtgg
aacttatgatggctattactatgaattgtggaaggattctggaaatacaacaatg
acagttgacacaggaggaagatttagctgtcaatggagtaacattaacaatgcac
ttttcagaacaggtaaaaaattcagtacagcatggaaccagctcggaacagtgaa
aatcacctactctgctacctacaatccaaatggtaattcatatctatgtatctat
ggatggtcaagaaatccactggttgaattttacattgtcgaaagctggggcacat
ggcgtccgcccggggcaacatcattgggtactgtaacgattgatggaggaacata
cgatatttacaagacaactcgtgtaaatcaaccatctatagaagggacaactaca
tttgatcagtattggagtgttagaacatcaaaaagaacaagtggcactgttactg
tgactgatcattttaaagcatgggctgcaaaaggcctgaacttgggtacaattga
ccagattacactctgtgtagaaggttaccagagcagtggttcagctaatataaca
caaaatacattttctataacaagtgcctcttcaggtggaacaacacccactacca
caaaggtagagtgtgaaaatatgtcactgagtgggccgtatgcatcaaaaattac
aagtccattttacggcatggctctttatgcaaacggagataaagcaacaacaaat
ataaacttttcagcaagccgtaactatacttttaaattacgaggatgtggaaata
acaataatttagcatcagttgatttactaatagatgggaagaaagtaggttcgtt
ctattatcgggggacatatccttgggaagctcctatagagaatgtgtatgtgagt
gcaggttcgcacaaagtggaaattgtagtttctgctgataatggtacatgggatg
tttatgcagattatttgttaatacaatga

Amino Acid Sequence:
MRFKKFLKVLIAVLMCFMLGNPFYAQAAITLTSNASGTYDGYYYELWKDSGNTTM
TVDTGGRFSCQWSNINNALFRTGKKFSTAWNQLGTVKITYSATYNPNGNSYLCIY
GWSRNPLVEFYIVESWGTWRPPGATSLGTVTIDGGTYDIYKTTRVNQPSIEGTTT
FDQYWSVRTSKRTSGTVTVTDHFKAWAAKGLNLGTIDQITLCVEGYQSSGSANIT
QNTFSITSASSGGTTPTTTKVECENMSLSGPYASKITSPFYGMALYANGDKATTN
INFSASRNYTFKLRGCGNNNNLASVDLLIDGKKVGSFYYRGTYPWEAPIENVYVS
AGSHKVEIVVSADNGTWDVYADYLLIQ*

Figure 23-21

Athe Gene Number:
Athe_0090

Product Function:
hypothetical protein

Nucleotide Sequence:
atgtttggcggaagtacaatatccaaatatgttgaaggttatacaacattaaaga
atatctatggaacagaaatcaattcttcgaccattcaatcttgtgataggaatat
tggtcaggctacaatttatgtaagcccttatgtaactgcatttacacatacttta
actcagcgaagaggtattttaatgcatgagatgggacacgctatgggcttaagac
atcctaattattcggactcaagaaatagctattcagctgatagttatggaagtat
tatggattacagttataccgaagaatacccaactatacatgatatatgtgacatt
gaaataatgtatggttttaactgttacaattaa

Amino Acid Sequence:
MFGGSTISKYVEGYTTLKNIYGTEINSSTIQSCDRNIGQATIYVSPYVTAFTHTL
TQRRGILMHEMGHAMGLRHPNYSDSRNSYSADSYGSIMDYSYTEEYPTIHDICDI
EIMYGFNCYN*

Figure 23-22

Athe Gene Number:
Athe_0153

Product Function:
hypothetical protein

Nucleotide Sequence:
atgtttaaagatttcttaaaagactcatttttcttcacaaaaagatactatggat
acatatttggtctattggttatatcatttattcttggtactttggcagttgtggt
tttactaatacttggtattttattggctatgttaatgggtgtggatatgtctaca
ttaggtttattaaatgataaaaatgtttttccatttttaagtagcaaatttattt
tctatatgattttaatgattctactttttataatgtgggtattgttagtattagc
ttttatccaatatccacttgtaaagacctttattgaaataacacgagaaaaggat
atatacaaaaaacctttgaaatttctttgccggaataaaagaaaagaatctga
tgatgggcttgaaaatagttggactgggatttttattaacacttatagttgggtc
aatccttttataggtattataggtatagcttttagtaccaatactgtcgataat
ttagcacgctttgctttgattataattggtattgttggtatagtattaggtgttt
acctgttattaaggttaatgtttgctaatgcagcgttggtggataaagatatagg
agttatagaaagtattaaagagagtttaaagcttacaaaggggaaaatgggattt
gtaatagcagcttttgtttatagtattttggtatccatagttttccaaataccag
tttatattgttgatacattatttaaaacagaaacatctcaagagagtatattatt
tattattttaagtttgataggatttgtttttaccttatagcagtgccttacttt
attgtgttacagtatctaccgtataataccttgaaaagccatgttgaaaatgcaa
accatatagatagccagttttacggaacttataataatattattgggtaa

Amino Acid Sequence:
MFKDFLKDSFFFTKRYYGYIFGLLVISFILGTLAVVVLLILGILLAMLMGVDMST
LGLLNDKNVFPFLSSKFIFYMILMILLFIMWVLLVLAFIQYPLVKTFIEITREKD
IYKKPFEIFFAGIKEKNLMMGLKIVGLGFLLTLIVGSILFIGIIGIAFSTNTVDN
LARFALIIIGIVGIVLGVYLLLRLMFANAALVDKDIGVIESIKESLKLTKGKMGF
VIAAFVYSILVSIVFQIPVYIVDTLFKTETSQESILFIILSLIGFVFYLIAVPYF
IVLQYLPYNTLKSHVENANHIDSQFYGTYNNIIG*

Figure 23-23

Athe Gene Number:
Athe_0154

Product Function:
glycosyl transferase family 2

Nucleotide Sequence:
atgagtgcaaaagaatttccaaaaaacgaactaataagcgttataattccctgtt
ataacgaggcacaaaatattggacaaacactcaaagaaatttatgattatttaga
tgagtttgtacctaactatgaagtaattgttgtagttgaaaaaagcacagacaat
actttggaagtaataaactcaagaaaaaatcaaaaaactatcgtattagagaata
caaaaaaatatggcaagggatatagtttaaaaagaggtatatattttgcaaaagg
acaatacatacttacctgcgatgcggaccttcctgtggacataaaaaaatatttt
cttcctatgttagaactattaaagagagatgaaaaggttgctgcagtttttgcca
ctgccttagctataaaaacatgcaggaaagaaagaggctttgtaagaagtattgt
ttcactgatatttttttgtttttagacagttgttattacagtttcctgtaagtgat
acacaattaggttttaaattgtttagagcagatgtagcaaaaaagttctgtcaaa
aagttaatgagaatgggtttttgtttgacttgatattaacagatttgatgttaaa
tgaaggatatcaaattgaagaagtaaatgtaaaagttgtagaaaggaaaatcaag
tcttctgtttctgtcgctgaaataataaaaactacttataagttttgcaaatata
tattgtttacgcgttcaaaattacttagaaaatgtacagataaagttatgataaa
tgataaaagtaaaagcataaaagctacaagtatttga

Amino Acid Sequence:
MSAKEFPKNELISVIIPCYNEAQNIGQTLKEIYDYLDEFVPNYEVIVVVEKSTDN
TLEVINSRKNQKTIVLENTKKYGKGYSLKRGIYFAKGQYILTCDADLPVDIKKYF
LPMLELLKRDEKVAAVFATALAIKTCRKERGFVRSIVSLIFFVFRQLLLQFPVSD
TQLGFKLFRADVAKKFCQKVNENGFLFDLILTDLMLNEGYQIEEVNVKVVERKIK
SSVSVAEIIKTTYKFCKYILFTRSKLLRKCTDKVMINDKSKSIKATSI*

Figure 23-24

Athe Gene Number:
Athe_0155

Product Function:
hypothetical protein

Nucleotide Sequence:
gtgaacgaacgaaataaagaaattgcatatattattcttgacttattcgttgctg
ccgttccatttttgtatattacatacctttacagattaggagttacagaattaga
aagtttaaaagaagttgtatatgctaattcccttttaagtacactttttatatt
tctaacatagttatgtttctcaaaatagaatataatattagcttctacgcaactg
caaaagtaataagtgtaatttcattaattttaggaattatattgcaagtcttata
tccatcttatggatgggcaataagtatacttttattaatggacagttttgtagtt
tttctcagaggaactatgttgagagagaaaaattacatgaaagcggctttaatta
attttctgttgcacacaagtagatttgggatattaatactgtttcctttagatgg
tcttttgacaagatggattgtatctaatatcggtggatatttgctggtattggca
atgtatttatctttgataaagtggatttttgaaagagaagataaaaatagcatta
tgagtttaaacatagatgaaattagaaaaagagcttattatttactttttgtaat
gataattttgtaattcaaaatattgatctattagcactaaagggaatgaatttt
tataacttgcttgttttgtcaaggccgtggggcttagttgcttatgttgttacag
tatcaattggaaatttagtcttgattaataaaaaaacatatccctattggattta
tttaactgtagcatgggctggaattttcattggatatatttctttgggtaaacac
ataaattactatcttttcaaaattccatattcgtttgatttaagtatattaccaa
ttttattttaccttttaatggctacgatattatttacaaattacttaaattatgc
aaaaaagaagcagttaataattgaacctttgtgcttattcctattaattgaagtc
ttaaaaaataaaatctctctaatgagagaattgaaatttgaggtattaattttaa
ttttaacatttttgataattgttcacttgtttttaattttttatattcaactta
taacaaatgtatgtag

Amino Acid Sequence:
VNERNKEIAYIILDLFVAAVPFLYITYLYRLGVTELESLKEVVYANSLLSTLFYI
SNIVMFLKIEYNISFYATAKVISVISLILGIILQVLYPSYGWAISILLLMDSFVV
FLRGTMLREKNYMKAALINFLLHTSRFGILILFPLDGLLTRWIVSNIGGYLLVLA
MYLSLIKWIFEREDKNSIMSLNIDEIRKRAYYLLFVMIIFVIQNIDLLALKGMNF
YNLLVLSRPWGLVAYVVTVSIGNLVLINKKTYPYWIYLTVAWAGIFIGYISLGKH
INYYLFKIPYSFDLSILPILFYLLMATILFTNYLNYAKKKQLIIEPLCLFLLIEV
LKNKISLMRELKFEVLILILTFLIIVHLFLIFLYSTYNKCM*

Figure 23-25

Athe Gene Number:
Athe_0156

Product Function:
glycosyl transferase group 1

Nucleotide Sequence:
atgaagactattttagtgttaacgcaaagggatatataccataaaaaagctggtg
gtgcagaaagatatttgtttaatgtgctaaaagggttgagtgatcattacaaaat
cacttgtctgtgtcaaaatgatggtactcagaaagattatgagatatatgataat
ataacgtttattcgaattaaaactaatttaattagcctaatttttaaagcaatgt
tctattataaaagaaataaggaaaacattgacttaataattgaccatacaaatac
acatcagtttttttactttttatacgttccgaaaaataagagactattaatagtt
caccagcttgcacttgagatttgggaatattattttccaaagtacattggaaaag
cgctaaaattacttgaaaaacttctttggcgtctatcatccggaatggcagtaac
agttagccggtcaactaaggaggatttacagaggtttggttttaaagaagaattt
atatggattgtaaaaaattcattaaagcataaatatacttcacttccacatatgg
agaaagaagattttcttgttagtgtcgggagattagttccatataaaagatttga
agatgcaatttatttagcaaaaaaagttaacaaaaaaatatttattataggagaa
ggtcaggaaaattataaaagaaaattaagaaattatgcaaaaaaaattaacgcag
atgtgatttttactggatatatttctgaagagaaaaaacaagacatagtagagaa
agcctatatgcacatttttccatctataagagaaggttggggacttgtaataagc
gaagcagctaacttagggacaccttccttagtatatcctgttcctggctgcctgg
atgcagtaaattacggaaaagctggttttgtaacaaagagaataggcaaagaaca
tttgttggaaagatttctaactattaatagggaagaatatgaaaggatgagaatt
tgtgcatttgaatttacttcacagctaaactataataaacaatgtgaagagtttc
aacaggttatacaaagtataatagaaaatacatag

Amino Acid Sequence:
MKTILVLTQRDIYHKKAGGAERYLFNVLKGLSDHYKITCLCQNDGTQKDYEIYDN
ITFIRIKTNLISLIFKAMFYYKRNKENIDLIIDHTNTHQFFTFLYVPKNKRLLIV
HQLALEIWEYYFPKYIGKALKLLEKLLWRLSSGMAVTVSRSTKEDLQRFGFKEEF
IWIVKNSLKHKYTSLPHMEKEDFLVSVGRLVPYKRFEDAIYLAKKVNKKIFIIGE
GQENYKRKLRNYAKKINADVIFTGYISEEKKQDIVEKAYMHIFPSIREGWGLVIS
EAANLGTPSLVYPVPGCLDAVNYGKAGFVTKRIGKEHLLERFLTINREEYERMRI
CAFEFTSQLNYNKQCEEFQQVIQSIIENT*

Figure 23-26

Athe Gene Number:
Athe_0157

Product Function:
glycosyl transferase family 2

Nucleotide Sequence:
atgaagctgggaattttgataacaacatataatgatgactatattattttaagat
gcttagattctatctataaccaacttgatgagatagattttccaatttatgttgt
atgtgttgatgatggttcagatttgcctcttacatatcctcattttgatattcta
agaactgaacatagggggaagaagctatgcgagaattgagggactgaaaaaaattt
tagctgaaaattgcacacatttcttatttttagatagcgatatggtccttccacc
aggctttctcaaaaagttgaagacagtagttgaaaattacgatagtgatgctttc
attatccctgaagtggcttttagtagttataacaatttttggacaaaagttaagg
tctttgaaagaaatttatatagagtaagttattgcaaagaaagtggaaatattga
agctgccagattatggaaaacaaatgcatttccgggttttgttgaaggattagaa
gcatttgaagagattcagccaacaatattgggtgttaaaaaaggtttaaagattt
taaaaatccaagagatttttattcttcatgatgaaaagaaggtaactttcaaaga
cttaataagaaaaaagaatacttacttttgttgtatgattggttctgaaaaatgt
tcaaagtgggagataatgaaaaggtattatttctttcgtccccatctttaccata
aggaaaatttaaagaaatacataagacatcccattttagctattggagttgtgtt
tatgtaccttgtattgacattgaattttatatggacgagtatatcaataaattta
attagaaaaggattgatggaaaaatga

Amino Acid Sequence:
MKLGILITTYNDDYIILRCLDSIYNQLDEIDFPIYVVCVDDGSDLPLTYPHFDIL
RTEHRGRSYARIEGLKKILAENCTHFLFLDSDMVLPPGFLKKLKTVVENYDSDAF
IIPEVAFSSYNNFWTKVKVFERNLYRVSYCKESGNIEAARLWKTNAFPGFVEGLE
AFEEIQPTILGVKKGLKILKIQEIFILHDEKKVTFKDLIRKKNTYFCCMIGSEKC
SKWEIMKRYYFFRPHLYHKENLKKYIRHPILAIGVVFMYLVLTLNFIWTSISINL
IRKGLMEK*

Figure 23-27

Athe Gene Number:
Athe_0158

Product Function:
hypothetical protein

Nucleotide Sequence:
atgagaagaagggggagaaatgttctatataaggtaagacgaagggaaaattatg
caaaaatagttataatagctgtattaatagggtgccttattttattaggatggca
tacggttgtaaagtcaagtaattggtttaaatcaattattatcgagaacatggac
tacaactatggaaaagattcaaatagtatccagccttatgcacctaaaaaaacac
atcctcgaggagaataa

Amino Acid Sequence:
MRRRGRNVLYKVRRRENYAKIVIIAVLIGCLILLGWHTVVKSSNWFKSIIIENMD
YNYGKDSNSIQPYAPKKTHPRGE*

Figure 23-28

Athe Gene Number:
Athe_0159

Product Function:
hypothetical protein

Nucleotide Sequence:
atgaaacaagaagagatattgcaaagacaaaggaagatagcaaagactatttcgt
tgattgctgtagtagttataatcctcatagatacaaaaaatatttggttgaataa
tggctttattattttcagtgatcttgactttagtggaattaatgataaaaggtat
attgagagaatatggggtgtttttaatccgcatttttcttcaaccaattttttta
atttatctcgtttgtttttatatttccttttatgccctaaacttttttggtctc
ttctttttacaatcaagcaatgcttaagttgataattttgtcttgtttactaata
tcaggaataggtatgtttttattatgtgagtatctattgataaaaaccttcagag
gttttcctgattatgtgcattattttggattaataataccggctgtgtattatgc
attaaatccatgggttatattcagaattcagcatatatttcttttgagcgggtac
agtgtatttccgcttataataaaagaatttttgaggttgtatgaaataaagattt
ttgactttcaaaaagatattgaagaatatcaacaaagaagaagagttccgaaaga
gcatattattatagatgaaaaaataaatcttaaggacttcttcaaaatagttttg
tactccatgattggttctgcagctattcactatttttctttttatgtcattttcc
ttaccatattatttataggcattatgccaagaatatggaaattaaataggtattc
taaagttgtacttttgttttatctaaaaaagtactttttaatgatttttttgatg
atagttggaaatatgtattggtttttgacatattctatagcactattatttgtta
atattgaacctcaaaatgtaaatgtggttgatacaattcaattattttcacgaaa
tagctctgtacagaatgtactctatttaatttcctattggcttccatttttttaac
ttagaaatgtttttagacaaaatgttttgggttgcaggaggagtcttcctgctta
taattgcatacataatattttatcgctttgggtggcatttttatgtaagactttt
caccctagtaacaagtctggtaatcttactctctttaggaacaaacgttgacccc
ttagctgatgtatatgttaaagtagtaacaaacgttccgataataggtcatctct
ttagagaccctaacaaaattgtaggtgtactggcttgtttttagctataatcct
gggttttggagtagacagaatagtatttttcctgttaaatgcaggctttggcaga
aaatttgcagtagcttttgttttagtcatgttgctttgtttctatttctatcaaa
gaccagttaaaattttctacatcgataattattacaaaggtgtgccaattccaaa
agaatataaagaaataagtaaatattatacaaaagatggtaaaatattatggata
cctactatggacaacatggtactttcaaatggcatttcggcttatagatggaata
taaacaaggaaatgcctggttcaatgaaagcggctggggattttcatgtatatgc
atcttctaaaaatacaatatttcaacatgaaaataatgtagggacagtttcttat
ttttattcttatttacagcatctattggataaaggtggaactgacaaaataggca
ggttattaaaacttacaggtttcaatcaagtggcttatcacagagatgtccaaaa
tcaagaagagagacaaaagtttaatctttttgttctggaaagacagagaggaatt
gttctgagaaagaaattaggatttattcacctttacaatgtaaagaagaattctg
ataacgctaaggtggacatttacaataccaagggattaaaccattttgttcagct
ctttgactttgaaaatatactgggatcaaaaggatataatataatctggtctcag
ggtaaaagggagaaatttgatcttaacaaggtaaatttagtggtaggcgataata
aatttgatatttttcaacaaaatatacctgagaaatatataatttcattatttga
caagataaatacaggaaatccatatgtaggttgggctaagtcaatgtgcaaggag

Figure 23-29

tcagattggttttggattcacaaagtaaacggtttaaatcgtctgccatgggatt
atgattatggtaaaggatttatattcacatatacaccctataaagtggaacttcc
tccatataagttaaataagaatgttggggaggttatactcaactccaatgatatt
gtaaaaacagatttttttgaaattgacaaagaggagacaaattttaaacttgaga
ttacaaaagatgtaatgggcaatagcatcttagcagcggaagttgctcccaaaaa
atttattggtaatgttttctggaaaattgcaaggtcaaaacttatgtcgataaaa
cctgggtttgtaaatcttcgggcagtggtttccggtgtaaatgctggaaaagttc
attttaaactcaaatttttcaatcaaaacatgaatgaattgggagtggtctatgg
taatgttccgggagaaatagctgaatttaataaagcacttataactgctaatgct
gttgtaccgcctgcaacaaagtatatgagggtggaactttggctacttgaagatc
aaaaaacaccagtatatttttggattcatgacattgagttaagatatttttcaag
catttcatctaatgaactgagaataaaaattccttcatatattaaaggagaatac
catctttacgtgagagctctggtaaatgagaatggtgggaaggtcaagatcaatg
ataccaaaataaacttaaaaggtgatgagaaccagtttaaatggttgtatgttgg
taaatataccagagatactattgttataaaaccagaagaagggtttgtactattg
aaccttctttgtgctttacctgattggttatataaaaaaattcaaaacgaagatt
tgaagagtaaacaagcaatagtattgtttagcaatgaatttgcattaaaagaagg
ttttaaagttaaaaaaatggatgattacattattcatccaaattttgttgatagt
cttttagacgtaataagtgatgggatttatactaaaaaaatagatatattaagaa
gtgatcactataaactatattttataggaaacgttacaagtcaagttacaataac
attgaagaataaagacgataaagttgtaagatttggatatttgaaatttttaaa
ctaattaataataactttaatggtataagattttatgtttctaacaagccttact
cgtactttttaaatattgaaaaagaagaaaatccaaaatggcgttcacaattgta
ttatattgaccttggtttttgaaaaaagggcaatatactttagagataaagttt
gaaaatggaataaaatctttaacagaacctaatgatattcatgttttacgtccag
aagaggttaaagttaatttaaaagttgacgacactctttatatattaatttcaaa
cattcttcaagaaaatgtgacgactcaaaagaaagaaaataaaatttttagcagc
aagtattcagacatatattctgataaagtaaaatactggctcatctatggaacaa
aacaggtagagtcaaaaaagaatgagttatattacgtgaactttgatattgaaac
gaaaggtttggcagaggttagtgctaagattttgtttatggataagaataaagta
ctttacaatagccaatatattgacatatttaacaataaaggacaaagtatatttt
cagctaaaaaaaatggctttatccaaatagtatttttgtgcgtagtaacggtaa
agaagatggtgtgtttgaattaaagcaaattagattactagagataagtaaaatg
aacaaattttcatcaatagtaattttaccaagtagtatacaaaaaggtagtaaat
tgaatatactgaatgaaacctataatcctctatgggaaaacaatggtgaaaaagc
gaaacaagtaaacattttcttaaatggtttttacactcaaagcaacaaatatgta
tttggtaaaaaaatatggatagcttatatatttggtttactaatatcaattttat
atattacagtaaatttgtggttgctattaagaagaaaaaaataa

Amino Acid Sequence:
MKQEEILQRQRKIAKTISLIAVVVIILIDTKNIWLNNGFIIFSDLDFSGINDKRY
IERIWGVFNPHFSSTNFFNLSRLFFIFPFYALNFLVSSFYNQAMLKLIILSCLLI
SGIGMFLLCEYLLIKTFRGFPDYVHYFGLIIPAVYYALNPWVIFRIQHIFLLSGY
SVFPLIIKEFLRLYEIKIFDFQKDIEEYQQRRRVPKEHIIIDEKINLKDFFKIVL
YSMIGSAAIHYFFFYVIFLTILFIGIMPRIWKLNRYSKVVLLFYLKKYFLMIFLM
IVGNMYWFLTYSIALLFVNIEPQNVNVVDTIQLFSRNSSVQNVLYLISYWLPFFN

Figure 23-29

LEMFLDKMFWVAGGVFLLIIAYIIFYRFGWHFYVRLFTLVTSLVILLSLGTNVDP
LADVYVKVVTNVPIIGHLFRDPNKIVGVLACFLAIILGFGVDRIVFFLLNAGFGR
KFAVAFVLVMLLCFYFYQRPVKIFYIDNYYKGVPIPKEYKEISKYYTKDGKILWI
PTMDNMVLSNGISAYRWNINKEMPGSMKAAGDFHVYASSKNTIFQHENNVGTVSY
FYSYLQHLLDKGGTDKIGRLLKLTGFNQVAYHRDVQNQEERQKFNLFVLERQRGI
VLRKKLGFIHLYNVKKNSDNAKVDIYNTKGLNHFVQLFDFENILGSKGYNIIWSQ
GKREKFDLNKVNLVVGDNKFDIFQQNIPEKYIISLFDKINTGNPYVGWAKSMCKE
SDWFWIHKVNGLNRLPWDYDYGKGFIFTYTPYKVELPPYKLNKNVGEVILNSNDI
VKTDFFEIDKEETNFKLEITKDVMGNSILAAEVAPKKFIGNVFWKIARSKLMSIK
PGFVNLRAVVSGVNAGKVHFKLKFFNQNMNELGVVYGNVPGEIAEFNKALITANA
VVPPATKYMRVELWLLEDQKTPVYFWIHDIELRYFSSISSNELRIKIPSYIKGEY
HLYVRALVNENGGKVKINDTKINLKGDENQFKWLYVGKYTRDTIVIKPEEGFVLL
NLLCALPDWLYKKIQNEDLKSKQAIVLFSNEFALKEGFKVKKMDDYIIHPNFVDS
LLDVISDGIYTKKIDILRSDHYKLYFIGNVTSQVTITLKNKDDKVVRFGYLKFFK
LINNNFNGIRFYVSNKPYSYFLNIEKEENPKWRSQLYYIDLGFLKKGQYTLEIKF
ENGIKSLTEPNDIHVLRPEEVKVNLKVDDTLYILISNILQENVTTQKKENKIFSS
KYSDIYSDKVKYWLIYGTKQVESKKNELYYVNFDIETKGLAEVSAKILFMDKNKV
LYNSQYIDIFNNKGQSIFSAKKNGFIQIVFFVRSNGKEDGVFELKQIRLLEISKM
NKFSSIVILPSSIQKGSKLNILNETYNPLWENNGEKAKQVNIFLNGFYTQSNKYV
FGKKIWIAYIFGLLISILYITVNLWLLLRRKK*

Figure 23-29

Athe Gene Number:
Athe_0160

Product Function:
hypothetical protein

Nucleotide Sequence:
atgccaggtaaaaaaagatctgttgtgggtgtgttgatattctcaataatcactt
tgggtatctattacttatactggatttatgtaacctcaaaagaaacccaacggta
cttagaaaagaatactacaagcccaggtcttgaacttttactatgtataattaca
tgtggtctttattggttttattggatttttaaatatagcaaaattgctgttgaat
gccaacaaaaagctggtttgcctacagaagataatgctgttataaaccttattct
tagcatcattggacttggaattataagttctatgattctccaatcaagtttaaac
aaagtttgggaatttgaaagttctaaaaatagtccagttatttcaacaactaata
cttctgaaaacaataagtaa

Amino Acid Sequence:
MPGKKRSVVGVLIFSIITLGIYYLYWIYVTSKETQRYLEKNTTSPGLELLLCIIT
CGLYWFYWIFKYSKIAVECQQKAGLPTEDNAVINLILSIIGLGIISSMILQSSLN
KVWEFESSKNSPVISTTNTSENNK*

Figure 23-30

Athe Gene Number:
Athe_0450

Product Function:
response regulator receiver protein

Nucleotide Sequence:
atgatagttaaaaatgtgttgattgcagatgaaaatgaatacataagaaaagcaa
taattgaaaagtttgaaaattcttttgattcagtacatcactttgtatttttga
ggctgctgatggtgaagaagctttgaacaaaattggtgagaacagtattcatgtt
gcaataattgatacaaacttgcctaaaattagtggatttgaagttttgagaacta
tcaaaaaaagttctgttaaagcttacattccagtaattttgctcagtagcaatat
tcatcgtgcaacaagagcaaaggcatacgagcttggagctattggagttattcca
aagccttttcaactttagaagtttacaatatggtaaggtcacttttgtacactc
aagatgaatatttacatattaatgaagttgtacatcttatttctttttgaatga
aatggcaaatgaacatatcattgtagttcccaaaattatagatgaatttttatca
gaatattttgtttcatatcacttgttggctttaaaagatactacaaaagaggtga
tttataacagaggttttgatgaatatgaaattaaaaaggttattggctatttaaa
tgatgagaacagaataaattcaagctgttattcgattataaaacttaacaataag
gaagaaacttattactttatattcaaaattctgaatgataataaaatattaattc
ttctaaagaaagttttagaactttggagtgaattaaatgatagataa

Amino Acid Sequence:
MIVKNVLIADENEYIRKAIIEKFENSFDSVHHFVFFEAADGEEALNKIGENSIHV
AIIDTNLPKISGFEVLRTIKKSSVKAYIPVILLSSNIHRATRAKAYELGAIGVIP
KPFSTLEVYNMVRSLLYTQDEYLHINEVVHLISFLNEMANEHIIVVPKIIDEFLS
EYFVSYHLLALKDTTKEVIYNRGFDEYEIKKVIGYLNDENRINSSCYSIIKLNNK
EETYYFIFKILNDNKILILLKKVLELWSELNDR*

Figure 23-31

Athe Gene Number:
Athe_0451

Product Function:
hypothetical protein

Nucleotide Sequence:
atgatagataatataaattttgttatcttgtgtattattattgttattcttttag
ttataacttcgggagttgtttttttatacaaagttataaatagtatccgtcaaga
aaaaataaagcagatttttaaaaattacagtaaaaatgtctatgatgtgataaca
ggaaagaaaaatgttattgattcttcaaacatttacattctgagtgatgttataa
atatatattattcatgggtacatggagaagaaaaaaaacggctatatatagcttt
gaagaaattgaattttttgacattgctatcaagatggttcaaaaagggaacaaa
gttcgaaaacttaggtttgcaaaagttattagtatagtgggcgaagaagatgagc
taaaaacgcttttaaaaataagcattaaagagccgtatctgatagatacaactgc
tgaagccattttcaaaaacatagatgagatacagaatttaagtgtttttaaacct
tatttaaagactatttttctaaacattgataaatatccagattcggttagaaaaa
gaatagaattttttacagtttatggtggcgaaaaaataaaagatataattttata
cgtgataaaagaaaacccgtcggataaagttctgatatcatgtttgaatatattt
tcagagattgcggcattagaagatttagaacacattgattttcttataaatcatc
catcaccagaagtcagatcagctttctgcagagtgattgaaaaaataggatgtag
aaactgcaaagaaaaacttgaaactctgataaaaaacgaaaatataaattttgtg
aagttacgagccttaagggcgttgagcaatgtttcatctaagggttctttaaaat
atttacttgcctcccttgaagatgactggttctacatgagagactttgcaagaaa
gatgctatctgagtttggaccagttattttaaatgatcttctaaaattttactat
acaacaaatgataaatttgcaagggacaaaataagagaggtttttatagtccag
tgaattttgattatattatcaagagtgctttgaattatagaactgaacaggagaa
agatttagctttagaaataattaaaatacttaaatcgtctaacccccactgtttt
tatcaaagaatgaaagatgaaggatttgaatatcttataaccgagaaagaggtag
tgaataacgaatgcgacatataa

Amino Acid Sequence:
MIDNINFVILCIIIVILLVITSGVVFLYKVINSIRQEKIKQIFKNYSKNVYDVIT
GKKNVIDSSNIYILSDVINIYYSWVHGEEKKRLYIALKKLNFFDIAIKMVQKGNK
VRKLRFAKVISIVGEEDELKTLLKISIKEPYLIDTTAEAIFKNIDEIQNLSVFKP
YLKTIFLNIDKYPDSVRKRIEFFTVYGGEKIKDIILYVIKENPSDKVLISCLNIF
SEIAALEDLEHIDFLINHPSPEVRSAFCRVIEKIGCRNCKEKLETLIKNENINFV
KLRALRALSNVSSKGSLKYLLASLEDDWFYMRDFARKMLSEFGPVILNDLLKFYY
TTNDKFARDKIREVFYSPVNFDYIIKSALNYRTEQEKDLALEIIKILKSSNPHCF
YQRMKDEGFEYLITEKEVVNNECDI*

Figure 23-32

Athe Gene Number:
Athe_0452

Product Function:
glycosyl transferase family 2

Nucleotide Sequence:
atgcgacatataataaactggttttcatggtttgtatcttattatgttttagttt
tgaatactgtttatgccattttaattttaatatctctttttggtattgttagtta
ctggagaaataagataaaaggaaggattgtggaggttgtctcatcagactttgca
ctccctgtatcactgctggtgccggcatacaatgaagaagaaacaatagcaaaat
ctgtaaaatctttttgcagattgaatatccagaatatgaagtagtagtaattaa
tgatgggtcaaaagatgggacattggatgtattaaaaaacgaatttgacctttac
attgtagataggaaatttagaaagattttgtcaacaaaagagataaaggtcatat
actattccaaaaagtattcaaatttaatcgtagtggataaagaaaatgggggaaa
agctgatgctctcaatgcagggataaacgtatgtacatacccatatgtttgttca
cttgatgccgattcaattttagaaagagattcaatagccaaggttatgcaaccat
tttttgataacccttatgaagttgtagccacaacgggtattgttaggattgtaaa
tggaaccgaattggattcctttggcaatataaaaaaattaaagctaccaagttca
agtcttgcaagatttcagataatagaatacttgagagctttttgggtgcgagaa
aaggtctttctatgataggaagtcttgttattgcatctggagcatttgcagcatt
taataaaaatgctattataaaggttggagggttttctgatagaactgttggcgag
gatatggagattgttgtaaagctcagaaaaaactcttataaagaaggcgcactgg
gaagagttgaatttgtaccagacccgattgtatggactcaatgtccagagacttt
aaaagacctttcaaaacaaagaagaagatggcaaagaggtctttgccaagttatt
ttcatgcacaaagatattctatttaatcctaagtatggcatattaggtctttttg
ctatgccatatcagcttatgtttgaactattggggccgtttgtggagatgttggg
ttatatttttatacctatatcgtattttgctcacataatcaatttagaagtggcc
ttatttttctttgcagttgagataatgtacggaatagttatttcgattttagcag
ttcttcttggagaattttctgatagaaaatatgaagggtggagagagttcggtat
acttgtattgtttgcgatattagaaaattttggctacagacagatgacaatgtta
ttcagaattgttggtacatttgaagctatacttagaaagaaaggttgggcaaagc
ctgagagaaagaagttataa

Amino Acid Sequence:
MRHIINWFSWFVSYYVLVLNTVYAILILISLFGIVSYWRNKIKGRIVEVVSSDFA
LPVSLLVPAYNEEETIAKSVKSFLQIEYPEYEVVVINDGSKDGTLDVLKNEFDLY
IVDRKFRKILSTKEIKVIYYSKKYSNLIVVDKENGGKADALNAGINVCTYPYVCS
LDADSILERDSIAKVMQPFFDNPYEVVATTGIVRIVNGTELDSFGNIKKLKLPSS
SLARFQIIEYLRAFLGARKGLSMIGSLVIASGAFAAFNKNAIIKVGGFSDRTVGE
DMEIVVKLRKNSYKEGALGRVEFVPDPIVWTQCPETLKDLSKQRRRWQRGLCQVI
FMHKDILFNPKYGILGLFAMPYQLMFELLGPFVEMLGYIFIPISYFAHIINLEVA
LFFFAVEIMYGIVISILAVLLGEFSDRKYEGWREFGILVLFAILENFGYRQMTML
FRIVGTFEAILRKKGWAKPERKKL*

Figure 23-33

Athe Gene Number:
Athe_0607

Product Function:
Ig domain protein

Nucleotide Sequence:
atgttctatcggcaggacccagatatcgaaaagaatatagtgagtgttcaaagat
tagatattacaacagaggtagggcaaaagccacagcttccgccaactgtcatggt
ggtttatgatgatgatacatcctctgaggcaaatgttacatgggaagctattgat
gagagtaaatactcagcgattggtgaatttactgtgactggtactgtagaaggca
caactctaaaagcatatgccaaagtcaaggttatcaacagcaaaaaccttgtaaa
aaattacagttttgaagaaggactaaagtattggataagcgaaggaaatacaaat
gcagtcaaaactgaaagtggtgggcattctgggacagcgcagcttactcactgga
gtgacaaaccttacaaggtggtcacatatcaaacaattgagaatataccaaatgg
tatatacattttcagagcctttgcaaatggaggaggcggtcagaatgctaactac
ctgtttgtaaaagactatggtggagaagaattgagaataaatatgccaactacat
gggtcgctgagtggcacagaatggtgattgcaaacattaaggttacgaccggacg
tgttacaatagggctttattcagattcagagaatgcaggaggtacatggtgcaac
cttgacgatgttgaatttttcaaggttgctgacagtgaatttgaaattttaaatg
caaatctgcaaaaggataaaggaattgttgcgagtgttacagtgaaacaatcaga
aggtatccagcatgaggggaaagaggccattgtatttgagctacttaaagggaca
actcctgtttcaattgttgctattgaaaaagatataatagatgctgaagatttca
aggcatactttaatgtaaatgattatcttagccctgattacagggtaaaagtatt
tgtatttgacaaatttgacacaagcctttctgttccaaatagtcttgctgagcct
gttgagcttcgatga

Amino Acid Sequence:
MFYRQDPDIEKNIVSVQRLDITTEVGQKPQLPPTVMVVYDDDTSSEANVTWEAID
ESKYSAIGEFTVTGTVEGTTLKAYAKVKVINSKNLVKNYSFEEGLKYWISEGNTN
AVKTESGGHSGTAQLTHWSDKPYKVVTYQTIENIPNGIYIFRAFANGGGGQNANY
LFVKDYGGEELRINMPTTWVAEWHRMVIANIKVTTGRVTIGLYSDSENAGGTWCN
LDDVEFFKVADSEFEILNANLQKDKGIVASVTVKQSEGIQHEGKEAIVFELLKGT
TPVSIVAIEKDIIDAEDFKAYFNVNDYLSPDYRVKVFVFDKFDTSLSVPNSLAEP
VELR*

Figure 23-34

Athe Gene Number:
Athe_0608

Product Function:
S-layer domain protein

Nucleotide Sequence:
atgaaaaaaagaacaaggatagcttcaagtctattgcttgtttttgtgtttttgt
tctctttattgtctgtgggatattctgcccaggaagttattttaaattccatacc
tgataaaaatccaggtgaagacgttgtaatttctgggcaaacgatgtttgatgag
attgttatcaaagtgttgagaccaaattctaccatactttatataaatacagtga
aggggaaaaacttcagtgacaaatttactttaccagcagacttgcctgagggtac
ttacacagttgttactggaaaaggttctattgtcgatataaaaactttcaatgta
gtaaggaaacaaccatctttacctaattcgtttcctgagattttaatccctaatt
tcaaagaagctgaaaaacaggcaggtgtacagccatctcaaaaggaagatattaa
acaaactgataaagctgctttgcctgaaattacaattgataatagtgggataatt
aggcccaaaatatatcagctgtcagagggaagattagacgttagattagatgagt
ctttagtacaaaaagcactgcagatgcaggtgtcaaaaaacaaaattttaacgtt
ggatttaaaaaacagcaagcctttagtacagtacaatattgtgcttccatcgaaa
gtatttacagcttcttttgatgtgcaggaagttttagttgatacagatgaattag
atttgaaacttccagctgacctattaaaggaaaagaacatagacaagaacaagca
aatagagatattgataaaaaaagtagaaaattctaatataccttcagagattcgt
gcggcagttgggaaaagaccaatatatgaaattgggttttatcaagatttgaaaa
aggtgaatgttgaaagacctacaaattctataaaaatatctattgcttatactcc
tggaatagaagagctgagcggtattgaaaacttggttttgtttcatataaaagag
gacgggaaaacagagatttatcaaatagtaaatatgtaaggacatcgaaatgtg
tgattgcaaatgtgaatggttttagcaagtttgcaatagggtatgttaccaaaaa
atttgacgacttaaaaaattattcgtgggctgaaaaagctgtgtcttccctggcc
gcaagaaatataataagtggtgttgacaaaagtagttttagacctcaagagtata
taaagcgtggagagtttgtcaagtggctggtaaatgctcttggccttgatgctca
gtattcttcgaactttgaggatgtcaaaaaagatagcagctactggcgtgaagtt
gcaattgcaaaagcacttggcattgcaaaaggttatgcgaataagtttaaaccag
aagattatataacaaggcaagatatgatggttcttgttaaaagagcattggaggt
tgcaaacaaaccaattgtaaaaacaaaaagtagccttgcaactaaattttctgat
tcatctgatatatctttgtatgctcaagatagtatttctattttggttgcaaatg
atttaataaaaggaaatagcaaaaatcaaatactgcctcgaaagtttgcaactcg
tgcagaagctgctcagcttttgttcaggatatttttcaaatgggaatag

Amino Acid Sequence:
MKKRTRIASSLLLVFVFLFSLLSVGYSAQEVILNSIPDKNPGEDVVISGQTMFDE
IVIKVLRPNSTILYINTVKGKNFSDKFTLPADLPEGTYTVVTGKGSIVDIKTFNV
VRKQPSLPNSFPEILIPNFKEAEKQAGVQPSQKEDIKQTDKAALPEITIDNSGII
RPKIYQLSEGRLDVRLDESLVQKALQMQVSKNKILTLDLKNSKPLVQYNIVLPSK
VFTASFDVQEVLVDTDELDLKLPADLLKEKNIDKNKQIEILIKKVENSNIPSEIR
AAVGKRPIYEIGFYQDLKKVNVERPTNSIKISIAYTPGIEELSGIENLVLFHIKE
DGKTEILSNSKYVRTSKCVIANVNGFSKFAIGYVTKKFDDLKNYSWAEKAVSSLA

Figure 23-35

ARNIISGVDKSSFRPQEYIKRGEFVKWLVNALGLDAQYSSNFEDVKKDSSYWREV
AIAKALGIAKGYANKFKPEDYITRQDMMVLVKRALEVANKPIVKTKSSLATKFSD
SSDISLYAQDSISILVANDLIKGNSKNQILPRKFATRAEAAQLLFRIFFKWE*

Figure 23-35

Athe Gene Number:
Athe_1853

Product Function:
Cellulose 1,4-beta-cellobiosidase

Nucleotide Sequence:
atgttaactaataaaaagcttagggagtatgtatattttgtatcgattatttttt
tcatgactattatattgttaaatggacagataaatgtgagcaaaagtaatttagc
aatggcggcaacaggaagtcaaattgagaagttaaatagaggattaattgcaata
aaagtaaccaacggagtatttttgagttggagaatgtttggttcagatttagcga
atatagggtttaatatttacagaaatggagtgaagataactaatacgccgattca
gaacagcacaaactatgttgatactgctgggacagctgcttcgaagtattatgta
aaagcagtgataaatggtgtagaggtagagcaatctgaagaagtaagtgtgctta
gtagcaattatattgaaataagattaaataaaccagcgaattctccactcggtgc
ttcatattcgccaaatgatgcaagtgttggcgatttagatggtgatggggaatat
gagattgttctgaaatgggatccaagtgattcaaaggataactcacaatctggat
acacaagcaatgtgtatttagacgcttacaaattaaatggcaagtttttatggag
aattgatttgggtaggaatataagagcaggagcacattatacacaatttatagta
tatgatttagatggtgacggaaaagctgaagttgcatgtaaaacagctgatggaa
caatagatggacaagggaatgtgataggtgatccaaatgctgattggaggaattc
ctctggttatattttatcagggcctgaatatttgactatatttgaaggtgcgaca
ggacgagcaataaagacggttaattatattccgccacgagggaatgtttcatcat
ggggtgattcttatggaaacagggttgacaggtttttagcagcagtagcttattt
ggatgggaatagacctagcttaattatgtgccgaggatattacacaaaaacatat
atagttgcttggaattggcgaaatggtgagttaacaaagttatggcaatttgaca
caggagagattagagatggatatagagatgattacgaaggacaaggaaatcacaa
tttgagtgtggctgatgttgacaatgatggtaaagatgagataatatatggtgct
atggtagtagatgataatggagcaccattatattcaactaaattaggtcatggtg
atgcaatgcatgtgacagatattgatccagacaggccaggattagaagtttggca
gtgtcatgaaggaagtacaggagcgagtttaagggatgcacgaactggacagata
ttagtgagagttttaacatctggagatagtggacgtgctttgacggcagatatta
acccgcgatatagaggattagaaatgtgggcggcaggtggaataagtgttagaga
ttgtagaggtaatgtaatcagtaatgcgacaccaccaattaattttgcaatatgg
tgggatggagatttaggtagagaattgttggataatgtatatatttataaatggg
attataacaataataggagtaatactatattcacagcaagtggatgttcatcaaa
caatggcacaaaagcaacaccttgcttgagtgcagatatacttggggattggcga
gaagaggtcatatttaggactagtgataacaatgcaatcaggatatatacaacca
ccacattgacagattataagatacctacgcttatgcataacaggcaatatagggt
gtctatagcatggcagaacgttgcatataatcaaccacctcacgtaagtttttat
ttagggtatgagactaatgtgaataacatatatcaatattttgaaggttatgggc
aacaaccaattgttacaccgtcgctaaccccgacaaaaacaccaacgcctacatc
aactccattgccaacatcaactgcaacatctacgccaactccaacagcaacagca
acaccaacaccgacaccaacagcaacaccaacaccgacgccgagcagcacacctg
tagcaggtggacagataaaggtattgtatgctaacaaggagacaaatagcacaac
aaacacgataaggccatggttgaaggtagtgaacactggaagtagcagcatagat

Figure 23-36

ttgagcagggtaacgataaggtactggtacacggtagatggggacaaggcacaga
gtgcgatatcagactgggcacagataggagcaagcaatgtgacattcaagtttgt
gaagctgagcagtagcgtaagtggagcggactattatttagagataggatttaag
agtggagctgggcagttgcaggctggtaaagacacaggggagatacagataaggt
tcaacaagagtgattggagcaattacaatcaagggaatgactggtcatggatgca
gagcatgacgaattatggagagaatgtgaaggtaacagcgtatgtagatggggtg
ctggtatgggggcaagaaccaaaataa

Amino Acid Sequence:
MLTNKKLREYVYFVSIIFFMTIILLNGQINVSKSNLAMAATGSQIEKLNRGLIAI
KVTNGVFLSWRMFGSDLANIGFNIYRNGVKITNTPIQNSTNYVDTAGTAASKYYV
KAVINGVEVEQSEEVSVLSSNYIEIRLNKPANSPLGASYSPNDASVGDLDGDGEY
EIVLKWDPSDSKDNSQSGYTSNVYLDAYKLNGKFLWRIDLGRNIRAGAHYTQFIV
YDLDGDGKAEVACKTADGTIDGQGNVIGDPNADWRNSSGYILSGPEYLTIFEGAT
GRAIKTVNYIPPRGNVSSWGDSYGNRVDRFLAAVAYLDGNRPSLIMCRGYYTKTY
IVAWNWRNGELTKLWQFDTGEIRDGYRDDYEGQGNHNLSVADVDNDGKDEIIYGA
MVVDDNGAPLYSTKLGHGDAMHVTDIDPDRPGLEVWQCHEGSTGASLRDARTGQI
LVRVLTSGDSGRALTADINPRYRGLEMWAAGGISVRDCRGNVISNATPPINFAIW
WDGDLGRELLDNVYIYKWDYNNNRSNTIFTASGCSSNNGTKATPCLSADILGDWR
EEVIFRTSDNNAIRIYTTTTLTDYKIPTLMHNRQYRVSIAWQNVAYNQPPHVSFY
LGYETNVNNIYQYFEGYGQQPIVTPSLTPTKTPTPTSTPLPTSTATSTPTPTATA
TPTPTPTATPTPTPSSTPVAGGQIKVLYANKETNSTTNTIRPWLKVVNTGSSSID
LSRVTIRYWYTVDGDKAQSAISDWAQIGASNVTFKFVKLSSSVSGADYYLEIGFK
SGAGQLQAGKDTGEIQIRFNKSDWSNYNQGNDWSWMQSMTNYGENVKVTAYVDGV
LVWGQEPK*

Figure 23-36

Athe Gene Number:
Athe_1854

Product Function:
Pectate lyase

Nucleotide Sequence:
atgataaaatctaagaataaaaaagaggaggtttgggtgatgagtaacaggaaga
ttttagccattgtagtcagtttgataatggttgtttcattgtttacagggattgg
gttgcgtaatgaagttgcaaaggcagcgacacttttaacagatgattttgaagat
ggcaacagagatggatggtcgacatcgaacggtagttggagtgtagtagtggatg
ggagcaaggttttaaagcaggctagcacaggttctgaggcgagagcatatactgg
ttcatctgattggagtgattatacagttgaagcgaaagttaaagtattaaatgtg
aaggattcgagttcaggtgcgggagtgatagtgagatataaaaactcaggtaact
tttatgcgttggtgctaaggggttcaaagatagaaatagggaagaaattaaacag
taactggagtacattggcgttcaagtcatttacgttggatcaggatacctggtat
aatgtgaaattagaagtaaatgggagcaagttagttggatatgttaatgggagtc
aagtattaagtgcaagtgatttatcgattacgacaggaaaagcaggtttaatagc
tgacaggtgtgttgctgaatttgatgatgttgttgtaaattcaagtgtgagcggt
acagcacctactccgacaccaacaccgacttcatcagtgacaccaacaccgacat
cgactccaacgccaaccaaaacacctactccaacttccacaccagtaccaacaca
gaccccagcagtaacaccgacgccgaccccaaatacgggtggtgttttagttatt
acagatacaataattgtaaaatccggtcaaacatatgatggtaaaggaataaaaa
taatagctcaaggaatgggtgacggaagtcaatctgaaaatcaaaagcccatatt
taaacttgaaaaaggggcaaatttgaaaaatgtaataattggagcgccaggttgt
gacgggatacattgttatggtgataatgtggttgaaaatgttgtatgggaagatg
ttggagaggatgcgttgactgtaaaaagtgagggggtagtggaagttattggtgg
ttcagcaaaagaagctgctgacaaggtgttccaacttaatgcaccgtgtacattc
aaagtaaaaaacttcacagctacaaatataggaaagcttgtaagacaaaatggta
atactactttcaaagtagttatttatcttgaagatgtaacattaaacaatgtaaa
aagctgtgttgcaaaatctgatagtccagtatcagaactgtggtatcataacttg
aatgtaaacaattgtaaaacattatttgaatttccgtcccaatcacagatacatc
aatactaa

Amino Acid Sequence:
MIKSKNKKEEVWVMSNRKILAIVVSLIMVVSLFTGIGLRNEVAKAATLLTDDFED
GNRDGWSTSNGSWSVVVDGSKVLKQASTGSEARAYTGSSDWSDYTVEAKVKVLNV
KDSSSGAGVIVRYKNSGNFYALVLRGSKIEIGKKLNSNWSTLAFKSFTLDQDTWY
NVKLEVNGSKLVGYVNGSQVLSASDLSITTGKAGLIADRCVAEFDDVVVNSSVSG
TAPTPTPTPTSSVTPTPTSTPTPTKTPTPTSTPVPTQTPAVTPTPTPNTGGVLVI
TDTIIVKSGQTYDGKGIKIIAQGMGDGSQSENQKPIFKLEKGANLKNVIIGAPGC
DGIHCYGDNVVENVVWEDVGEDALTVKSEGVVEVIGGSAKEAADKVFQLNAPCTF
KVKNFTATNIGKLVRQNGNTTFKVVIYLEDVTLNNVKSCVAKSDSPVSELWYHNL
NVNNCKTLFEFPSQSQIHQY*

Figure 23-37

Athe Gene Number:
Athe_1855

Product Function:
Pectate disaccharide-lyase

Nucleotide Sequence:
atgagtaacaggaagattttagccattgtagtcagtttgataatggttgtttcat
tgtttacagggattgggttgcgtaatgaagttgcaaaggcagcgacacttttaac
agatgattttgaagatggcaacagagatggatggtcgacatcgaacggtagttgg
agtgtagtagtggatgggagcaaggttttaaagcaggctagcacaggttctgagg
cgagagcatatactggttcatctgattggagtgattatacagttgaagcgaaagt
taaagtattaaatgtgaaggattcgagttcaggtgcgggagtgatagtgagatat
aaaaactcaggtaacttttatgcgttggtgctaaggggttcaaagatagaaatag
ggaagaaattaaacagtaactggagtacattggcgttcaagtcatttacgttgga
tcaggatacctggtataatgtgaaattagaagtaaatgggagcaagttagttgga
tatgttaatgggagtcaagtattaagtgcaagtgatttatcgattacgacaggaa
aagcaggtttaatagctgacaggtgtgttgctgaatttgatgatgttgttgtaaa
ttcaagtgtgagcggtacagcacctactccgacaccaacaccgacttcatcagtg
acaccaacaccgacatcgactccaacgccaaccaaaacacctactccaacttcca
caccagtaccaacacagaccccagcagtaacaccgacgccgaccccaacgccgac
gacagttccaacacctgccccgacacctgtacctggggtgaatgctatttatgtg
gcaccaagtgggagctcagataatcctggtaccattgatcgacctactacattag
aaaaagcaatcacgatagtacaacctgggcagataatctacatgagaggtgggac
gtataagtattctgcgcagatcacaattgaaagaaataatagtggtacaagcaat
gcaagaaaatgtatttatgcatttccaaatgaaaggccaatattggacttttcat
ctcaaacatatgggagtgtggactcaaatccaagaggattacagattaatgggaa
ctattggcacataaaaggattagaagtcatgggagctgcggacaacggaatcttc
gtaggaggcagctacaatataattgaacaatgtgaaattcatcataacagagatt
caggtttgcagataagcaggtatataagttctgcaaccagagatgagtggccaag
ttataacttgatattgaattgtacatcacatgacaatatggatccagataacggt
gaagatgcagatggttttgcatgcaaactaacagcaggaccaggaaatgtattcc
gaggttgtgtagcgtactacaatgttgatgatggttgggatttatacacaaaaag
tgagacaggagctattggtgaagtattaattgaggattgtgtggcatatggtcac
gggcaaacatcaaccgggagtgccacatctagcagtgatggaaatggctttaagc
taggaggcagtaatataaaggtcaatcatacagtgagaagatgtatagcatttaa
taacaacaaacatggatttacttataatagtaatccgggtagcataacagtggaa
aattgtacgggctataataacggtttaaaggtaagtggaaggaacttttattttg
aagaaggtacacacgtgttgaagaattgtttatcctacaaagagagtgcatcgag
tgatttggtaagtggaacgataattaattgtgttttgtggagtaataggcaagca
ataaagctaaatggtcaactggtaaccgataatgacttttacagcttaacaccaa
ccataacaaggaatagtgatgggggtttaaacttaggagactttttaaagccaaa
gcctggtagtggtttagaaggaataggagcaaggtaa

Amino Acid Sequence:
MSNRKILAIVVSLIMVVSLFTGIGLRNEVAKAATLLTDDFEDGNRDGWSTSNGSW

Figure 23-38

SVVVDGSKVLKQASTGSEARAYTGSSDWSDYTVEAKVKVLNVKDSSSGAGVIVRY
KNSGNFYALVLRGSKIEIGKKLNSNWSTLAFKSFTLDQDTWYNVKLEVNGSKLVG
YVNGSQVLSASDLSITTGKAGLIADRCVAEFDDVVVNSSVSGTAPTPTPTPTSSV
TPTPTSTPTPTKTPTPTSTPVPTQTPAVTPTPTPTPTTVPTPAPTPVPGVNAIYV
APSGSSDNPGTIDRPTTLEKAITIVQPGQIIYMRGGTYKYSAQITIERNNSGTSN
ARKCIYAFPNERPILDFSSQTYGSVDSNPRGLQINGNYWHIKGLEVMGAADNGIF
VGGSYNIIEQCEIHHNRDSGLQISRYISSATRDEWPSYNLILNCTSHDNMDPDNG
EDADGFACKLTAGPGNVFRGCVAYYNVDDGWDLYTKSETGAIGEVLIEDCVAYGH
GQTSTGSATSSSDGNGFKLGGSNIKVNHTVRRCIAFNNNKHGFTYNSNPGSITVE
NCTGYNNGLKVSGRNFYFEEGTHVLKNCLSYKESASSDLVSGTIINCVLWSNRQA
IKLNGQLVTDNDFYSLTPTITRNSDGGLNLGDFLKPKPGSGLEGIGAR*

Figure 23-38

Athe Gene Number:
Athe_1856

Product Function:
transcriptional regulator, AraC family

Nucleotide Sequence:
atgggaagaaccatatgtattaaaaaaagtgcaaatatgttcttaatatttttag
ttatggtttttctttttcattttctaatgtcagttttcttattatatatgactaa
agaacttattttgagagacataattaatgataagtcacttcaattagataaactt
cgcgaaaaaatagacaatgaactaaagaaaatgaatgagatagttacacgtattc
ttactgatgatgaattaaaatgggtagggaatctaagggggttaagagaaaattc
tttagatttatgggagtactttgaatattataaacactttaaagacataggcatg
attaatggagagttaaaacctatcattgtactttttttacgcgagggagaaatag
tatattttgcttccgaagaatttggctctttttttacatttggatttgaaaattt
ttgtaattacttttcacctgacagaatgaactgccgcgtttggctcaataaaata
ttcgataaaaacaacagtaaatgcaaaaataagttaatatcacaagaatattcaa
ttagtggccaaagaatattggcattgcatgaaacatattactttccatttgataa
cattggaaatcaattggcaattcttctggtaattattgatgttgctaaattttgt
aagctactaaaggaaaacaaattaaacagccaagattcaataattttaatttatg
ataggtgtgagaaaaaaatattaacttcaaataaagcagatattaataatacaat
aaataacatattagaaaaacttagcaaaaataagaaatatatttctaatctgtat
aatgtaataacaatacgaaataaaaagtatatatttttacggttggcttcaaatg
tatatgattgggattatgtttatctagtaccatataacagcataaaaagtgaaat
atacatttcaagaactgtttatacattgtatctaacggagtttacacttttcata
attattataagtttatatgctatttatatcaaattatataaagtaaggacatata
aaataactaagggcttaaaaaatgtaaacgaggagaatgagaataaaatacttac
ttttaggtgcataaacaagttaaaggagaaagaccgtaggttaatacacactaaa
attacaagttacaggattcttaataaaataaccagtaaacaagaattattaaacg
atatgataatagaaaagcttatttatggttggtcttatagcaaaggggcaataga
agaaaaaatccaaagtataggcttaaaaattggaggtaaaaaatttttagttgcc
ataatcaaaatgctctctcttaccaaacaagttagaacaagagatataataaaaa
atgagttagaaagtattagggttgactcaaaaattaacttggtattttattatgt
ttaccaacttgcagatagtgatcttgctttaattttagcatttgatgaagatgaa
gatgaaaaagtaagtcaaaatgtaaatctatggctgagattgataggtgatagat
tagtaattaaaatatgtcacaaatatttaattgcagttggtcgtatagttaataa
cattgacgaatgtagaatttcatttttagatgcaaaagaaattattgaagctaat
aaattaatcacaacagagaacgttgagaatggcatactgtggtattataatgttg
ttaagaaagataataatatatggtatccaatagaaatagaagagagattaatgtt
gctagttaatcttgggaaaatatctgagattgaagaaattctacacctattatat
tataaaaactttgaagaaaaagatatatcactgagtttaaagtatgtgctaatta
gtgagttaattggaacaattattaaaattgcaaatatgacaaaagtaaatattga
caatctatttgacattaagaactttatttttattgaagaaaatgactttgacagg
ttatttgaagatataaagcaagtatttattaatataacagaacaaatcaagagaa
acaggaaggcttctaaggaaaaattaatagaggagatattagaatttataaacaa
aaacttatttaatccaaacatgagtatatcccttgtcgctgaaagatttaattta

Figure 23-39

tctgaatcttatttttctaatatttttaaaaatgcagtaggtataaagttcagtg
attatgtagaaaagttaagaatagaagaagcatataaattgataaagcaaaaaaa
gtggaatttagacgatattagtaaaatggtaggatatactaacattaaaaccttt
agaagagcttttaaaagggtaaaaggttgtttacccagtgaaattttaaacataa
atgaacatgatatatag

Amino Acid Sequence:
MGRTICIKKSANMFLIFLVMVFLFHFLMSVFLLYMTKELILRDIINDKSLQLDKL
REKIDNELKKMNEIVTRILTDDELKWVGNLRGLRENSLDLWEYFEYYKHFKDIGM
INGELKPIIVLFLREGEIVYFASEEFGSFFTFGFENFCNYFSPDRMNCRVWLNKI
FDKNNSKCKNKLISQEYSISGQRILALHETYYFPFDNIGNQLAILLVIIDVAKFC
KLLKENKLNSQDSIILIYDRCEKKILTSNKADINNTINNILEKLSKNKKYISNLY
NVITIRNKKYIFLRLASNVYDWDYVYLVPYNSIKSEIYISRTVYTLYLTEFTLFI
IIISLYAIYIKLYKVRTYKITKGLKNVNEENENKILTFRCINKLKEKDRRLIHTK
ITSYRILNKITSKQELLNDMIIEKLIYGWSYSKGAIEEKIQSIGLKIGGKKFLVA
IIKMLSLTKQVRTRDIIKNELESIRVDSKINLVFYYVYQLADSDLALILAFDEDE
DEKVSQNVNLWLRLIGDRLVIKICHKYLIAVGRIVNNIDECRISFLDAKEIIEAN
KLITTENVENGILWYYNVVKKDNNIWYPIEIEERLMLLVNLGKISEIEEILHLLY
YKNFEEKDISLSLKYVLISELIGTIIKIANMTKVNIDNLFDIKNFIFIEENDFDR
LFEDIKQVFINITEQIKRNRKASKEKLIEEILEFINKNLFNPNMSISLVAERFNL
SESYFSNIFKNAVGIKFSDYVEKLRIEEAYKLIKQKKWNLDDISKMVGYTNIKTF
RRAFKRVKGCLPSEILNINEHDI*

Figure 23-39

Athe Gene Number:
Athe_1989

Product Function:
Superfamily I DNA and RNA helicase and helicase subunits-like protein

Nucleotide Sequence:
gtgggtgatgtaaattctgtcaattccgctttaaaaaacttcaatattgtagttg
cttctgccaacaacaaagctgtggaaaatgtcaccaaagaaataccagttctaaa
ttcagttgaccaatcctgtcttgaaaaatatgaactccattatttcaaagatgga
gccgaactggtatacgactacagaggggaaaacaatgaaagcagtgagaaaactg
aatctttagtatcagttgaaaatgggcagcaatgctgggcattgatttcagcagt
tttgggtaaaaaagaaaacagagaaaagtttttcagtgctctggaaaagtatatt
aatgaattattttcttctataccgccagtcaagtgggaacagtgtaaaagaagtt
tcaatcatgtcttcaagaagtttagacatgtccagaaactttacaggactatgga
actttttctgaaactcgaagaaaacaggtttaaatttacccgtatgatagctgct
gcggggaagatttgcctaaaaaaatttttgccagatattttcccgatgacatga
aacttccttcctcagagttctggtcccaaaatgagcacgaaattcacaaatcatc
tccatggatgagcaggtatctcaatgatatacgcaatgaagtttttgcaagagct
cttcagttacatcaggcagctattgctgcaaacaaagagaagttcggggagaatt
tagaaaaattcattaaatatatgagaaaaaacgaatcactgcctgcagaaaaggc
aaaagaactgtggcatactttctttatgattgttccagttgtctcaacaacattt
gcgtcgctgagcaatatgtttaaagatgttgacgacgaaatcattgactggctga
ttgttgatgaggctgggcaggcactaccccagcattttatcggcgcactcctcag
aagtaaaagagcaataatagtgggagaccctctacaaattccacctgttgttaaa
ataccaccatttgttattaatgatgtgttcaaagcatatgggattttcaaatgga
gacaggaagacagtaattccagtacaccacgtataactgaaactgattccgtaca
aattgtagccgatagagccaacaaattcggtgcaaagattggcgatatgtgggtt
gggtgtcctctgagagtacataagcgttgtacggaaccaatgtttacagtagcga
atagaattgcttataaaaaccttatgatttttgatgtgaataaacctgaaggatt
gcaaactgtttttaaagatagcttctgggtagatgtaaagggtaagtgtgtttac
aagcattatgtaagagaacagggaaaagtagtgaaagctatcgtacaagaatttt
tgcagcgaaccttgacaactcaaaatgaggttaaattaacagaagagctttttat
catttcgccttttaaagcagtcaaaagtagtatatcttctatactgaaaagaatg
cctctgtataacacaaatttaaagaaagaagaatgggagaatatagttaatgaaa
tagttggcacaattcacagttttcagggcaagcaagccaacaatgtaatcatatg
ccttggtgccgatgaaagtaatgaaggagctgtaagatgggcatcttcagaaccg
aatattttaaatgtggcgctgaccagagctaaatatagagtcatagtaataggtg
ataaagacctgtggggaaagcataaatattttgatacactcttagaagaactagg
tgaaaaagtgattgagtataccacagaaaaagatctggtcaacaagatttttgct
tag

Amino Acid Sequence:
VGDVNSVNSALKNFNIVVASANNKAVENVTKEIPVLNSVDQSCLEKYELHYFKDG
AELVYDYRGENNESSEKTESLVSVENGQQCWALISAVLGKKENREKFFSALEKYI

Figure 23-40

NELFSSIPPVKWEQCKRSFNHVFKKFRHVQKLYRTMELFLKLEENRFKFTRMIAA
AGKICLKKFFARYFPDDMKLPSSEFWSQNEHEIHKSSPWMSRYLNDIRNEVFARA
LQLHQAAIAANKEKFGENLEKFIKYMRKNESLPAEKAKELWHTFFMIVPVVSTTF
ASLSNMFKDVDDEIIDWLIVDEAGQALPQHFIGALLRSKRAIIVGDPLQIPPVVK
IPPFVINDVFKAYGIFKWRQEDSNSSTPRITETDSVQIVADRANKFGAKIGDMWV
GCPLRVHKRCTEPMFTVANRIAYKNLMIFDVNKPEGLQTVFKDSFWVDVKGKCVY
KHYVREQGKVVKAIVQEFLQRTLTTQNEVKLTEELFIISPFKAVKSSISSILKRM
PLYNTNLKKEEWENIVNEIVGTIHSFQGKQANNVIICLGADESNEGAVRWASSEP
NILNVALTRAKYRVIVIGDKDLWGKHKYFDTLLEELGEKVIEYTTEKDLVNKIFA
*

Figure 23-40

Athe Gene Number:
Athe_1990

Product Function:
hypothetical protein

Nucleotide Sequence:
ttgaaagatacttttgaacagttaattaaaaatcagaaaaatattagagaaatct
atatgaatttactccagaaattttgttacgatatttatccagtcaggaacaaaga
atttacaaattcattcggtgttacatggcttaattccaaaaaaggcaaacaacac
gaacaaagaagaaaagtcgagatagaaaacgaaatagaggaagaaattgaaacgc
cactgcaagacctttttgccagtttttatctaaaagacattgaagaagcacttaa
acgccttgagaatgataatctaaaccctgtcttgtttcgcttctttactgagtaa

Amino Acid Sequence:
LKDTFEQLIKNQKNIREIYMNLLQKFCYDIYPVRNKEFTNSFGVTWLNSKKGKQH
EQRRKVEIENEIEEEIETPLQDLFASFYLKDIEEALKRLENDNLNPVLFRFFTE*

Figure 23-41

Athe Gene Number:
Athe_1991

Product Function:
hypothetical protein

Nucleotide Sequence:
atgataaatacaattgcaaaagatgagatgaaaagaatactgctggcataccaga
aacttgaattgttgcagcctgctcctattcccaatggagtttttataaacgactt
tactgaaatctggagtcctaactatttcacagattgcagagtctatcttggaatt
tatgattaccaggaagtaaaagaggcttttgaaaaaatttataaccagcaatatg
accatcccgaaaaagaagatgaaaaacaaaacagaaaacaaaaactttactggtt
ctatttcgatattgatgagaacgggaaatataaagatgacagtttaaaaatttct
tcaactttgtgggctcttaaacaactggaactggaaaatacaaacaaaagaaggc
tgtcagaaactgcggagaaaaaagatgaagaaaagcaaactgaagattatgatct
tggtaaaaatttacagagtttgctcaaaacttga

Amino Acid Sequence:
MINTIAKDEMKRILLAYQKLELLQPAPIPNGVFINDFTEIWSPNYFTDCRVYLGI
YDYQEVKEAFEKIYNQQYDHPEKEDEKQNRKQKLYWFYFDIDENGKYKDDSLKIS
STLWALKQLELENTNKRRLSETAEKKDEEKQTEDYDLGKNLQSLLKT*

Figure 23-42

Athe Gene Number:
Athe_1992

Product Function:
hypothetical protein

Nucleotide Sequence:
atggctgactataagaaagcatatacattacgaatagatgaaactttgcttgata
agattagagtaatcgccgagaaaaataaacgctcaattaatgctcaaattgaata
tttgattcaacagtgtgtcgaagaatttgaagctgagcacggtgagataaaaatt
gaagaagaaaatactgaatga

Amino Acid Sequence:
MADYKKAYTLRIDETLLDKIRVIAEKNKRSINAQIEYLIQQCVEEFEAEHGEIKI
EEENTE*

Figure 23-43

Athe Gene Number:
Athe_1993

Product Function:
transcriptional regulator, RpiR family

Nucleotide Sequence:
atggatatatttaccttaataagaacaaaatataatactttatctcaaactcaaa
agaaaattgcagattggattttagaacatggaaatgaaattattttaatgtcaat
aggcgaaatcgcagcaaagtgttcaacaagtgaagctactgtaatgagattttta
agaaaacttggttttgattcatatcagctctttaaggtaaaagtggctcaagata
ttgttgatgtgactccccgagccgtttatgaggatgtaagtagtgaagatagtgt
taatgagataaaacagaaagtaattcaatctacaattgatgctatcagagatata
gataagttaattgaagatagcacaatagaaaaagctatagagataatggctaatg
caaaaagaatctttttctttggagtgggagcatcaggagcaatagcaaaggacgc
atttcacaagtttttaagattaggcataaataccatatattgtagtgattctcat
attatgagcataatgtgtagccatatgacagaaaaagatgccgttttggcgattt
cacactcaggagaaagtcgagaaataattgatgcaatagaacttgcaaaagaaaa
caaagcaaaagtaatttcgtttacaagttacccgaattctaccttggcaaaactt
tcagatgtggtttttctaagtgctaccaaagagactaaattcagatctgacgcga
tggtttcaagaattgtgcaatgtgtgattatagacattctctatgtttctctagt
tttaaagttaggaagtgaagcaattgtaaatgtgaataaatcaagattagcagtc
gcaaaaaagaaaaaataa

Amino Acid Sequence:
MDIFTLIRTKYNTLSQTQKKIADWILEHGNEIILMSIGEIAAKCSTSEATVMRFL
RKLGFDSYQLFKVKVAQDIVDVTPRAVYEDVSSEDSVNEIKQKVIQSTIDAIRDI
DKLIEDSTIEKAIEIMANAKRIFFFGVGASGAIAKDAFHKFLRLGINTIYCSDSH
IMSIMCSHMTEKDAVLAISHSGESREIIDAIELAKENKAKVISFTSYPNSTLAKL
SDVVFLSATKETKFRSDAMVSRIVQCVIIDILYVSLVLKLGSEAIVNVNKSRLAV
AKKKK*

Figure 23-44

Athe Gene Number:
Athe_1994

Product Function:
D-isomer specific 2-hydroxyacid dehydrogenase
NAD-binding

Nucleotide Sequence:
atgagtaaaaaaatgaaaataatggtaattggagacgcaatgataccgggtaaag
attttgaatcagcagctaaaaaatatttatctgattatgtggaagaaataattac
aggagattgggaaaataattgggacaatttacaaagaagaagattggaagtagaa
aagaaagggcccgagattgaggaagtagttcctttaataaaggaaaaagggcaag
atgtttcaatgttgtttggtttatttgttcccatttccaaagaaacattcaatta
cttgccaaaggtaaagattattggggtttcgcgagcaggcttagaaaatgtaaac
gtaaaagaagcaacccaacgaggagttttagtgttcaatgtccagggaagaaatg
cagaagctgtttctgactttgcaataggtttgcttttggcagaatgtagaaacat
tgcgagagcccactatgcaataaagaatggccagtggcggaaagaattttctaat
tctgattggattccggaactaaaaggcaaaacagttggtattattggttttggat
atattggtagactggtagcaaaaaaactctctggatttgaagttagaagacttgt
gtacgatccttatgtaagtgaagaggaaattagagaatgcggatgtataccagta
gacaaagagactttgttcaaagaaagtgattttattactctccatgcacgcctca
cagaagagaataaaaatttggttggcaaatatgagatttcattgatgaaaccaac
agcatacattattaacactgcacgggcaggtctaattgataaagaagcattaata
gaggctctaaagacaaagagaatagcaggagcagcactggatgtgttctgggaag
aacctattccttcggacagtgagttgttagaattggacaatgttactcttacaag
tcatttagcaggaacaaccaaagaagcacttacaagatcacctgagcttttaatg
gaggatgtcaagaagtttattgaagggcagaaagcaagatttattgtgaatccag
aggttttggaaaaccaagagttcaagaaatggctggagggtgtgaagaaatga

Amino Acid Sequence:
MSKKMKIMVIGDAMIPGKDFESAAKKYLSDYVEEIITGDWENNWDNLQRRRLEVE
KKGPEIEEVVPLIKEKGQDVSMLFGLFVPISKETFNYLPKVKIIGVSRAGLENVN
VKEATQRGVLVFNVQGRNAEAVSDFAIGLLLAECRNIARAHYAIKNGQWRKEFSN
SDWIPELKGKTVGIIGFGYIGRLVAKKLSGFEVRRLVYDPYVSEEEIRECGCIPV
DKETLFKESDFITLHARLTEENKNLVGKYEISLMKPTAYIINTARAGLIDKEALI
EALKTKRIAGAALDVFWEEPIPSDSELLELDNVTLTSHLAGTTKEALTRSPELLM
EDVKKFIEGQKARFIVNPEVLENQEFKKWLEGVKK*

Figure 23-45

Athe Gene Number:
Athe_2076

Product Function:
glycoside hydrolase family 2 sugar binding

Nucleotide Sequence:
atgaacatcaactttaaaataaaaccattctggttttggaatgggaaaatggaaa
atgatgaaatagcagatcaaatagctcagatgcatgaaaaaggaattggtggatt
cttcattcatccacggcaaggtctcgaaataccctatctttctcacgagtggttt
gaaaaggtttctgttgcaattgaatgtgcaaaaaaatataacatggaagtgtggc
tttatgatgaatatccttatccaagtggaatttcagctggtgaagtagttgttca
gcatcctgaatatcaagcttttatattagattataaagtgtttgaggcaaaagat
aatgaagaaatttgcattgaaattcccatgtgtgaagtgttattagcaagagctt
atagaataaggaacaatattatagtatggaatgaatacatagatttgattgatta
tattggtgtaatctacaaagaacacatttaccaagaaagtgggcttacgttttac
aataggaaaagatactttgttggagatggggcaaaagcacttaaatggaaagtac
caaaaggaaagtggaagatatttttgttttatcagtatccattaaaaaattttaa
atatttcgggacatttattgatcctctaaataaagatgctgtaagactgtttatt
caaaccactcatgaaaaatataaaaaatacttgggtcatgagtttggaaaaacaa
ttaaaggaatttttacagatgaaacagctccagttgctggcaaacttccttggtc
aaaattgcttcctaagctttttgagcaaacatatggggaaaatcttattgaaaaa
ttacctcaaattatttgcacagacatatttgatacagccggttcaaaaattaggt
atcagttttggaaacttgttgtggacacatttattgagagctacgacaaacaaat
tcttgaatggtgtcatcagaacaatcttttgtatgtgggcgaaaagccaattttg
agaagctcacaattagcttttatggacattcccggaatagatgcagggcaccaaa
aggctggtgacattccacaggttgtatctgaaaactacagagcaaatccaaagat
agcatcatctgccgctcacttttataaaaaggaaagggttttgtgtgaatgcttt
cacagtattggctggagcatgacaatgcaagatatgaaatggatatttgactggc
taatattgcagggaatagatatgtttgtcccccatgccttttattatagtgcaga
tggactcaaaaaacacgatgccccaccttcagcctttttccaaatgccttggtgg
aaacatcaaaaaatattgtctgagtatgtagaaaatgtaactaaaatgcttaaaa
attgtaaaagaaaagttgatgtacttatagtagatccaattacaagccagtggac
ctgttttaacgacaaagaagtaaaagagaagatttcgatggatttctgtagaatt
cagcaaattcttttagaagaaaatgtagactattatgtgattgaccagtcattag
taggaagtttggaatgcaggaatcagaagatttattatgacaatgaaaaatttga
attattgatcattccacctgtgaccaatttagaaaaagaggcttacatgaagata
aaagatctaatattgaaaggatgtaaagttgtatttattggctgtttgccattcc
aaaccatcgaagattttgatgttgccaaagatattagcaattttcttggagtaaa
ttctatggacatcgcaaaagcatataaaacaggttctaaattgaacaatacagtt
tttcttaacagttgtatcttcattggtaatatagaagatttagtagcaaaaattg
acaagatttgtaaaaagcctgtaagtatatcatatgaatcttctaatgaccgtgg
tattctctgtgcttactttgaggacgcagaacacgactttttatttatgattaac
ccaactaatgaaaagaagatctgcaaagtatacttgcggtacaggcctgatgaaa
taaacaaaatttattcggttccattgacatcgcaagaacttgataaggaaataaa
tttcgaaaattctttagataaaaagcaaataactttttccatgaattttgaacct

Figure 23-46

tttcaatcgtatttaattaaattagaaaaaagttttgttaagaaaaacgaccata
aaaatgatgttgaaaagagagtttttgagtacaaaatccctcttgcaactgtttg
ggaattttcaattgagagtttaaatcctttaaggctcggcagatggaatttaaag
ttgattttcaacaacgaaaatgaacagtattcaatatcttcaaaaattccagtta
caccaaaaccaataattgatcagatagaagaggcaaaaattccaataccactaaa
aacaaaaagtttctttggatgtccaaaagaaataacattgccaagctttgaggca
atatacactacttcattttttatagatgcagcaagccaaaaattctggcttgtaa
ttgaagatgaaggtataaaaggtgaatgggttgtgcttttgaacaatcatactat
tttaccaagagattttgtacttaaaagatttattctcatactaatttggcatat
gatattagcaacttaattaaactgggggaaaatcagctttgtgtatgtgtaaaga
ttagcaggtcttttgatggacttcttacacccatttatatcttcagcacggcagg
tgtattcaaagttgatgatagctggcatatagacaaacttccaactcaaggctgc
tttggtaaagaccttgaaaatggcattccttttatgctggttttataaagtacg
aaaaagaagttcaaatgccatcttttgaaaatggttttgtggaattctttattga
agataacataaatcagtgtgtaagtctttatataaatgatgaatttataggtaca
aggtgctggcagccttatagatggaaggtagattctgatttactttcttcaaaga
aggtaaagctcacacttgaagtatcaacttcgagcctgcagctgttcgaaggtga
agttattgaaccaataacacataaaattaagacaatataa

Amino Acid Sequence:
MNINFKIKPFWFWNGKMENDEIADQIAQMHEKGIGGFFIHPRQGLEIPYLSHEWF
EKVSVAIECAKKYNMEVWLYDEYPYPSGISAGEVVVQHPEYQAFILDYKVFEAKD
NEEICIEIPMCEVLLARAYRIRNNIIVWNEYIDLIDYIGVIYKEHIYQESGLTFY
NRKRYFVGDGAKALKWKVPKGKWKIFLFYQYPLKNFKYFGTFIDPLNKDAVRLFI
QTTHEKYKKYLGHEFGKTIKGIFTDETAPVAGKLPWSKLLPKLFEQTYGENLIEK
LPQIICTDIFDTAGSKIRYQFWKLVVDTFIESYDKQILEWCHQNNLLYVGEKPIL
RSSQLAFMDIPGIDAGHQKAGDIPQVVSENYRANPKIASSAAHFYKKERVLCECF
HSIGWSMTMQDMKWIFDWLILQGIDMFVPHAFYYSADGLKKHDAPPSAFFQMPWW
KHQKILSEYVENVTKMLKNCKRKVDVLIVDPITSQWTCFNDKEVKEKISMDFCRI
QQILLEENVDYYVIDQSLVGSLECRNQKIYYDNEKFELLIIPPVTNLEKEAYMKI
KDLILKGCKVVFIGCLPFQTIEDFDVAKDISNFLGVNSMDIAKAYKTGSKLNNTV
FLNSCIFIGNIEDLVAKIDKICKKPVSISYESSNDRGILCAYFEDAEHDFLFMIN
PTNEKKICKVYLRYRPDEINKIYSVPLTSQELDKEINFENSLDKKQITFSMNFEP
FQSYLIKLEKSFVKKNDHKNDVEKRVFEYKIPLATVWEFSIESLNPLRLGRWNLK
LIFNNENEQYSISSKIPVTPKPIIDQIEEAKIPIPLKTKSFFGCPKEITLPSFEA
IYTTSFFIDAASQKFWLVIEDEGIKGEWVVLLNNHTILPRDFVLKRFYSHTNLAY
DISNLIKLGENQLCVCVKISRSFDGLLTPIYIFSTAGVFKVDDSWHIDKLPTQGC
FGKDLENGIPFYAGFIKYEKEVQMPSFENGFVEFFIEDNINQCVSLYINDEFIGT
RCWQPYRWKVDSDLLSSKKVKLTLEVSTSSLQLFEGEVIEPITHKIKTI*

Figure 23-46

Athe Gene Number:
Athe_2077

Product Function:
lipolytic protein G-D-S-L family

Nucleotide Sequence:
atgatgcgaattgtagaaaaaattagacaaaaaaatgaaaattatcactccggtc
caatacctacaattgccttttgggtgacagtataacccacggttgttttgaggt
cattgaaggaacaaaaagaacattagaagttgtgtgcgattttgaagctgtctat
cataatcaatttaaaaaaatcttatcaatgatgtttccatttgctcaaattaata
ttgtcaatgcaggcataagtggtgatacagcacagggcggcttacaaagacttga
aagagatgttttaagatttaacccagaccttgtggtagtttgttatggattaaat
gatagtaacaaaggaaaagagtatttgaatgaatatttagatggtttagcaggga
tatttatagaacttaaaaaacacgatattgaagtaattttttaactcccaatat
gaaaaatacttatatttcaccagctataaagagcttgcctttgattgagatggca
aaattaaatatggaaagtcaaatcaatggaacgttagatctttacatggattcag
ctaaggagctttgcaaaaaggaaaaggttgtagtttgtgattgttatgaaaaatg
gaaaaagctttatcacagtggcgttgacacaacaaatcttctttcaaattttata
aatcaccccaatcgcccgatgcataagctctttgcatggtcattatttgagacaa
ttatgttttaa

Amino Acid Sequence:
MMRIVEKIRQKNENYHSGPIPTIAFLGDSITHGCFEVIEGTKRTLEVVCDFEAVY
HNQFKKILSMMFPFAQINIVNAGISGDTAQGGLQRLERDVLRFNPDLVVVCYGLN
DSNKGKEYLNEYLDGLAGIFIELKKHDIEVIFLTPNMKNTYISPAIKSLPLIEMA
KLNMESQINGTLDLYMDSAKELCKKEKVVVCDCYEKWKKLYHSGVDTTNLLSNFI
NHPNRPMHKLFAWSLFETIMF*

Figure 23-47

Athe Gene Number:
Athe_2078

Product Function:
oxidoreductase domain protein

Nucleotide Sequence:
atgaaaatttgcatagtaggcagtagtggacactatgtatatgctttaagaggaa
taaaagaagaccctcatgcccaaattgtgggaatctctcctggatgtgaaggaga
gaatattgaaaggttacattctcaagtaaatgaaatgggattcacacctgtggtt
tatagcaatcctataaggatgtttgaagatctcaaacctgacattgctgtgatta
atacatttttttataaaaattctgagcttgcaattgaggctatgaaaagaggaat
ccacgtatatatggaaaagcctgttgcactatcaatagaaaaacttgaagaacta
aagagtgtgtggaggcaaacaaaagtaaaactctcatcaatgctgggattgcgct
atacaccccatttttggactgcttataaacttataaatgaaaacaagataggtag
aataagactgatacatgcccaaaaatcttataaacttggaactcgacctgacttt
tataaacatagaagaacatatggcggaacaattccctgggttggcattcatgcta
ttgattggatttattggctaagtggcaagaaatttaaatcggtctttgcaggaca
ttcaaaactttataataatgatcatggtgagcttgaatctactgctttttgtagt
tttgtaatggaagatgagatttttgcaacggtgaacattgactatctgcgtcctg
ctactgcccctactcatgatgatgatagaattagaattgtgggaacaagaggaat
ttttgaagttttaaatggaaaagttttcttgctaaatgataccactaaagagatc
tcagaagtctctttagaaaaaccacctattgtgtttttagatttcttaaatgagg
taagaggtacagataagtgcttagttagtagcgaggatagcttttatgtaacctt
tgcttcgcttttagcaaggcagtctgctgatgaggataaggtaattgaattttaa

Amino Acid Sequence:
MKICIVGSSGHYVYALRGIKEDPHAQIVGISPGCEGENIERLHSQVNEMGFTPVV
YSNPIRMFEDLKPDIAVINTFFYKNSELAIEAMKRGIHVYMEKPVALSIEKLEEL
KSVWRQTKVKLSSMLGLRYTPHFWTAYKLINENKIGRIRLIHAQKSYKLGTRPDF
YKHRRTYGGTIPWVGIHAIDWIYWLSGKKFKSVFAGHSKLYNNDHGELESTAFCS
FVMEDEIFATVNIDYLRPATAPTHDDDRIRIVGTRGIFEVLNGKVFLLNDTTKEI
SEVSLEKPPIVFLDFLNEVRGTDKCLVSSEDSFYVTFASLLARQSADEDKVIEF*

Figure 23-48

Athe Gene Number:
Athe_2079

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
gtgagaagaacaaagacaatgaaagcattattgtttcatctgctaacaattctat
ttggttacataatgctgtatccacttctttggatgtttttcagttccttcaaaga
aaatagtgaaatttttctaaatgcgcacgaattgctacccaaaaagtggcttttc
aaaaattatattgatggatggagaggttttgcgggatacccattttcagtttttt
ttaaaaattcatttattgttactattattggtactgttggtgctgtcatttcctc
agctattgtagcttatggttttgcaagatgtaagttcaaaggcaaaggattttgg
tttggttgtatgattataactatgcttttaccatatcaggttgttatgattccgc
aatacatcatgtttcaaaagatgggatgggtaaatacgttcaagcctttacttgt
tccagctttttgggtcagccatttttatctttttaatgattcaatttataaaga
ggaattccaaacgaattagatgaagcggctaaaattgatggttgtagtaaatatt
caatttttacaagaattattttgccattaatttctccggctttaattacatcagc
aatcttttcatttttgtggcgttgggatgactttttagggcctcttttgtacctc
agcaaacctgagttgtatacagtatctttagggttgagaatgttctctgacccga
cagcagtttctaactggggagctgcctttgctatggcaacattatcgctcgttcc
ctcatttataatatttatattcttccaacgctatctagtggaaggaattgttact
acaggtttaaagggttaa

Amino Acid Sequence:
VRRTKTMKALLFHLLTILFGYIMLYPLLWMFFSSFKENSEIFLNAHELLPKKWLF
KNYIDGWRGFAGYPFSVFFKNSFIVTIIGTVGAVISSAIVAYGFARCKFKGKGFW
FGCMIITMLLPYQVVMIPQYIMFQKMGWVNTFKPLLVPAFLGQPFFIFLMIQFIR
GIPNELDEAAKIDGCSKYSIFTRIILPLISPALITSAIFSFLWRWDDFLGPLLYL
SKPELYTVSLGLRMFSDPTAVSNWGAAFAMATLSLVPSFIIFIFFQRYLVEGIVT
TGLKG*

Figure 23-49

Athe Gene Number:
Athe_2080

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
atgtcgaatattaatcgaaaaaagtctttcagagaacttttgatgagttcagaaa
atgttgctggttatgtttttatatctccctggcttattggctttttgtgttcac
tttgattcctattgcagcaacctttttatttgtcttttactcaatatgatttatta
tcatctcctaaatttgtaggattacaaaattatgtacaaatgtttaaagaagatc
cgttattttggaaatcaatgtcagtaactttttctatgtgtttgtaactgtgcc
attaaagctggcttttgcattgcttcttgccctttggctttcttacaaaagcaga
ctaacaccattttacagggctgtatactatgttccttctatgatgggtggcagtg
tggctgtggcagtgctttggcaaagactttttacaagtgatggtgttataaattc
aatattgaaactatttggaattcaaagtgagacttcatggataggaaatccaaga
actgctatatggacgttgatattacttgcggtttggcaatttggttcgccgatgt
tgatatttttagcaggtttaaagcaaataccagaaagctattatgaagcagctat
tattgatggagcaaatagctggcaaaagtttgttaaaataaccttgccgatgctc
acaccaataatatttttcaacttgattatgcagatgataggaagctttatgactt
ttactcaaggattcattattacaaatggcggccctgtgaacagcacactctttta
cgctatttacctctacagaagagcattccaattttatgacatgggctacagctgt
gctatgtcgtgggtaatgcttattatcattggaatactcacagcttttatattca
aatcatctacattttgggtatattatgagtccaaggaaggtgaataa

Amino Acid Sequence:
MSNINRKKSFRELLMSSENVAGYVFISPWLIGFFVFTLIPIAATFYLSFTQYDLL
SSPKFVGLQNYVQMFKEDPLFWKSMSVTFFYVFVTVPLKLAFALLLALWLSYKSR
LTPFYRAVYYVPSMMGGSVAVAVLWQRLFTSDGVINSILKLFGIQSETSWIGNPR
TAIWTLILLAVWQFGSPMLIFLAGLKQIPESYYEAAIIDGANSWQKFVKITLPML
TPIIFFNLIMQMIGSFMTFTQGFIITNGGPVNSTLFYAIYLYRRAFQFYDMGYSC
AMSWVMLIIIGILTAFIFKSSTFWVYYESKEGE*

Figure 23-50

Athe Gene Number:
Athe_2081

Product Function:
extracellular solute-binding protein family 1

Nucleotide Sequence:
atgaaaagaatattgtgcattggaactattataatatttttattttcaatcctta
tatttccttttactaaaccttctgaaaaggcttttggttcttctcagtcaatcaa
actcagagttgcgtggtggggaagccaaacgcgtcatgacagaactcaaaaggtg
cttgaactataccgagcaaaggttaatcgaaaagtaagttttgtaacagaatttg
gaagctggtcgggatattgggataaacttacaactcaagcagctgcaaagaattt
gcctgacattattcaaatggactatatgtatttagctcaatatgttcaaaaagga
ttgttggcagatttaacaccttacacaaaaaatggaatattgaatctgaaagatg
taagtgaagcaagtataaaaagtggatcagttggtggaaggatttatgccataag
tttagggactaacgctttagcaataatttatgaccctgctgtggctcaaaaagca
ggtgttaagatacctgaagatggcaactggacctggaatgactacaaagagatta
tcaaaaaagtttttcaaaaaaccaaaatcagagcagacttggcattaacagctga
cccaaaattcttacttgaatactatgtaaggcagcaaggcaaaagcttatataaa
cacgatggaacgggtttgggatttactcaagaaaaatttgttattgatgtattta
atatcaatttagagttattgaaaggcggatatacagcaaaacctgatgaagtttc
agcaacctccacaattgaagacagcttatttgtgaaaggcaagacatggatagga
tggacatggagcaacatgtttgttgcaacagcaaatgctgcaaaacgcccactgg
cattagctcttccacccaagggggtataagaccagggttgtatttaaaaccttc
tcagttcttctctattgctgcaacctccaaatataaaactgaagcagctaaagtt
ataaacttctttacaaatagcatagaggcgaataatattttgcttgctgaaagag
gtgtgcctatttcagcaaaagtaagagaaggtatcaagaatgcagtcgaacctgc
agtgagacagacatttgattatatagccctcgctgaaaagaattgttcacctatt
gatccaccagatccaccaggcggaactgaagtagggcagacgtttaaagatcttt
acgaccaagtcttgtatggtcagataaaacctgaaacagcggcaaagatgtttat
gcaaaaagcaaatcaaatacttgccaaaaacaagaaataa

Amino Acid Sequence:
MKRILCIGTIIIFLFSILIFPFTKPSEKAFGSSQSIKLRVAWWGSQTRHDRTQKV
LELYRAKVNRKVSFVTEFGSWSGYWDKLTTQAAAKNLPDIIQMDYMYLAQYVQKG
LLADLTPYTKNGILNLKDVSEASIKSGSVGGRIYAISLGTNALAIIYDPAVAQKA
GVKIPEDGNWTWNDYKEIIKKVFQKTKIRADLALTADPKFLLEYYVRQQGKSLYK
HDGTGLGFTQEKFVIDVFNINLELLKGGYTAKPDEVSATSTIEDSLFVKGKTWIG
WTWSNMFVATANAAKRPLALALPPKGGIRPGLYLKPSQFFSIAATSKYKTEAAKV
INFFTNSIEANNILLAERGVPISAKVREGIKNAVEPAVRQTFDYIALAEKNCSPI
DPPDPPGGTEVGQTFKDLYDQVLYGQIKPETAAKMFMQKANQILAKNKK*

Figure 23-51

Athe Gene Number:
Athe_2082

Product Function:
oxidoreductase domain protein

Nucleotide Sequence:
ttgaatattgcaaagattggtcttattggaattagtggttttgggagtatacact
tgcggtcaatagaacagcttcaagggaagatggttgacttgagagcaattgttgc
aacaagctacgaaaaaaataaagaagtgattgatagattggcttctcgaggtgtt
gagtattatcaggattatagattaatgcttgaaaatcataaagacttagattttg
ttgccatctcaacgcccattcatttacatgctccaatggcaattgatgcaatgga
aagaggttttaatgtcctgcttgaaaagccgcctgctgtgacaattcaggatatt
gacgctataattgagacgaagagaaaaacaaaaagggtttgtagcgtgaactttc
aaaacacctctggtaaagcatttagaaaacttcttgagtatataagagaaggcag
gcttggcaggataaaatcaattattggtgttggacgttggaaaagggatgaaagc
tattatcaaagaaatgcgtgggctgggaagctaattgttgatggtaactatgtct
tggatggaacaataaacaatcctcttgcacatcttttgaacaatgaattgattat
tgctgagacctcagaagaaaatggaggtgtacctcagaaagttactgcagaactt
tatcacggtcataaaattgaaggtgaagacacagcgtgtgtgagaatcatcacaa
aaacaggaatcgaggtctatttttattcaacgttatgtaacagggaagaagagtc
accctatatcataatagaagctgaaaaagcaagagcttattggacatttgcaaat
aagtttaagatagaatattttgatggcaatacagaagagtttgatggtggacgcg
aagatttgtttgtgaatatgtacattaatatggtggagcatctgtttgaaggcaa
acaactttactgtcctctggaagtgactcgcaattttgttttggcatccaacggg
gctttcgaatcgtcagggtgtatttatgatattccagacgaatatttagagatta
gcaatgaaaatggtaagatatatacctatatcaaaaatataaaagaaatcataga
tgaggcagctgagaatagaaaactattttctgaaataggagtaccgtgggcgaaa
caaacagaagagtttgatttaattgattactgttgttttagtatgtttaaaaggt
aa

Amino Acid Sequence:
LNIAKIGLIGISGFGSIHLRSIEQLQGKMVDLRAIVATSYEKNKEVIDRLASRGV
EYYQDYRLMLENHKDLDFVAISTPIHLHAPMAIDAMERGFNVLLEKPPAVTIQDI
DAIIETKRKTKRVCSVNFQNTSGKAFRKLLEYIREGRLGRIKSIIGVGRWKRDES
YYQRNAWAGKLIVDGNYVLDGTINNPLAHLLNNELIIAETSEENGGVPQKVTAEL
YHGHKIEGEDTACVRIITKTGIEVYFYSTLCNREEESPYIIIEAEKARAYWTFAN
KFKIEYFDGNTEEFDGGREDLFVNMYINMVEHLFEGKQLYCPLEVTRNFVLASNG
AFESSGCIYDIPDEYLEISNENGKIYTYIKNIKEIIDEAAENRKLFSEIGVPWAK
QTEEFDLIDYCCFSMFKR*

Figure 23-52

Athe Gene Number:
Athe_2083

Product Function:
Alcohol dehydrogenase zinc-binding domain protein

Nucleotide Sequence:
atgaaagcttatgcaatggtattagaagaatttaacaagccgttaaaagcaaaag
agtttgaactaatcaagccttctgatggtgaacttcttgttaaaattgaagcggc
gggtgtttgtggatctgatgtgcatatgttcagaggtaatgacccgcgtacaaaa
cttcccatgattttagggcatgaaggtgttggacgtgtgtatgctatttcaggtc
agtggcgtgatataaatggcgagaaaattcaagagggagatttgataatttggga
caggggtgttgtgtgtggtaggtgctacttttgtgctgtcaaaaaagaaagctat
ctgtgtccgcacagatggacatatgggataagcgttagctgcgcagagcctccgc
atttgagaggctgctattcggagtacatttatcttcacaaagatacgaaagtgat
aaaaataaaagagaatgttgatccagaaattttagtatctgcctcatgttctggt
gcaacgtgtgctcatgcttttgacattgtttcacctgattttggtgacagtgtcc
taattcaagggccaggtcctatagggctttatgcaatcatttttgcaaaacttag
aggagcacgaaatataattgtgattggtggcacaaaagaaagacttaaaatgtgt
gaagaatttggggcaacgcatgtgcttgatagaaattcaactacagcttgccaaa
gacaggaaataataatggatatcacaaatgggcgtggagtcgatttggcaattga
agctgtgggacatccatcagcagtaagtgagggaataaaacttgttcgaaatggt
ggaagctacttatcacttggttttggtgacccaaacggcagcgttacactcgatt
gttactatgatattgtgagaaaaaatttaagatatcaaggggtatgggtcagcga
tacaaaacatttatatatggcagtgaatgttgtgctccagaacagggaacttttc
aaaaagatgattacaaatgtttataagttgactgatgcgacaaaagctcttgagg
atatggaaaacaaaaatacaataaaatctgttctaaagccttga

Amino Acid Sequence:
MKAYAMVLEEFNKPLKAKEFELIKPSDGELLVKIEAAGVCGSDVHMFRGNDPRTK
LPMILGHEGVGRVYAISGQWRDINGEKIQEGDLIIWDRGVVCGRCYFCAVKKESY
LCPHRWTYGISVSCAEPPHLRGCYSEYIYLHKDTKVIKIKENVDPEILVSASCSG
ATCAHAFDIVSPDFGDSVLIQGPGPIGLYAIIFAKLRGARNIIVIGGTKERLKMC
EEFGATHVLDRNSTTACQRQEIIMDITNGRGVDLAIEAVGHPSAVSEGIKLVRNG
GSYLSLGFGDPNGSVTLDCYYDIVRKNLRYQGVWVSDTKHLYMAVNVVLQNRELF
KKMITNVYKLTDATKALEDMENKNTIKSVLKP*

Figure 23-53

Athe Gene Number:
Athe_2084

Product Function:
fumarate reductase/succinate dehydrogenase flavoprotein domain protein

Nucleotide Sequence:
gtgatatacaaagcagatgtgctggtagtaggtagtggcggagcaggtttgagag
ctgcaattgcagcgtgtgaaagagcatatgagagccgaaaaagaataaaagtttt
gctggcaagtaaaggaaaaataggaagttgtggtactacagctcttgcatactct
gatagaatggcattccatgtcacacttcccacaacagagcctaaaggtgaagata
actggaaatatcatgcaaaagatatctatgagattggcggacttgtttctgatta
tgacttggctgagattttagcaaaaaactcagccgatgcttatttttacttagat
agtctcggtgttccgtttgtaaaagaaaatggcgtacctgctcaatttgtgacag
acggttctatatatgcgcgtgcatgttttacaggacctgatactgctgttcaaat
agaaaaggctttgattcgaaagcttggtgagatgaaggatattgaagttttagaa
gatgtgatgataagtgatttgattgttgtgaataacaaagtttgcggagcaattg
cttttaaagggaatcaaaatatcataatacttgcaaaagccattgttttagcaac
aggtggggcaggaagtatttataaaagcaatgtgtttccaccacgcatgacaggt
gatgggtatgcaatggcacttcgtgcaggggctttgcttgtgaatatggagttta
ttcaaataggtctttcatcccccaaaacaaaacttgcgtgttcaggaagtataat
gagatgtgtccccaggtttgtaaatgaaaaaggagaagaatttttattaaattat
cctatcgcgtacaatgatgtatttgaaaaaggtgcaacgtggccaataagctacg
agcacaagacatgtttgatagacattgcagtgttcagagagattgcccgtggggg
taaagtgtttttagactttactcaaaatccaaaaggttttgaatttaaacaccta
agagaagacttaaaacaaagatattactcagaagttaaaaatcttacaaataaaa
gatctactccatacgaaagactttgtgaaataaatcctcagacagttgagtggtt
tttgaaaagaggaattgaccttagaaaccaaatgttagaaattgcgccatcaatt
cagcatttccagggtggtgtaaaaattagagaaaatgcaaatacagcaattagtg
gcttgtatgcctgtggcgagtgtgcaggcggacagcatggagcaaacagaccagg
cggaaatgcacttttggatactcaggtttttggtaagatagcaggggaaagtagc
tttgaatttgcctcaaacacttcgattgatgaagaatctgcaatttctgaagcaa
accgcctgtttgaaagctataaaatctatatagctgaagatggcattgatttgga
gaatgctatatcagaactaaatagggttatggatttgtacgcaagtgttgtgaga
catcaggatggacttcaaaaagcacttatgaaaattgaggagttgaaaacaagaa
aaatcaagcctgttgagtatgaatatcttttagagctaaaaaacatgcttttgtg
tgcagaggcagtggtaaaaagttgcattttaagagatgaaagcagaggaccgcac
ttgatgttcgaaaattacagcgatttgtggccaaagccaagagatgaaagataca
acatatatcacgtatgcaagctaaacaaagagacaaatcaagttgaagtctttcc
tatggaaccagtaaaacccgaaaccttagggggcaaagtaaaatga

Amino Acid Sequence:
VIYKADVLVVGSGGAGLRAAIAACERAYESRKRIKVLLASKGKIGSCGTTALAYS
DRMAFHVTLPTTEPKGEDNWKYHAKDIYEIGGLVSDYDLAEILAKNSADAYFYLD
SLGVPFVKENGVPAQFVTDGSIYARACFTGPDTAVQIEKALIRKLGEMKDIEVLE

Figure 23-54

DVMISDLIVVNNKVCGAIAFKGNQNIIILAKAIVLATGGAGSIYKSNVFPPRMTG
DGYAMALRAGALLVNMEFIQIGLSSPKTKLACSGSIMRCVPRFVNEKGEEFLLNY
PIAYNDVFEKGATWPISYEHKTCLIDIAVFREIARGGKVFLDFTQNPKGFEFKHL
REDLKQRYYSEVKNLTNKRSTPYERLCEINPQTVEWFLKRGIDLRNQMLEIAPSI
QHFQGGVKIRENANTAISGLYACGECAGGQHGANRPGGNALLDTQVFGKIAGESS
FEFASNTSIDEESAISEANRLFESYKIYIAEDGIDLENAISELNRVMDLYASVVR
HQDGLQKALMKIEELKTRKIKPVEYEYLLELKNMLLCAEAVVKSCILRDESRGPH
LMFENYSDLWPKPRDERYNIYHVCKLNKETNQVEVFPMEPVKPETLGGKVK*

Figure 23-54

Athe Gene Number:
Athe_2085

Product Function:
dihydroorotate dehydrogenase family protein

Nucleotide Sequence:
atgccaaacctatcaacgacgtatgcaaagctgaatttaagaacacctgtaattg
ttgcatctgctggcattactggaactgtggagaggcttcaaagatgcgaagaaaa
cggtgctggggcagttgtgacaaaaagtctttttcaaaaggaaatatgcagaatt
gcacccactccacggtttaaaatagtcaagcatgaaaacacgtttacgctttact
catatgaacaggcaagcgaatttaaccctcaagagtatgctgaatttatattcaa
agcaaaacaaaagctaagcattccagttattgcgagtataaactgctacacagat
gatgcatggcttgagtatagcaagcttatggagcaggcaggggctgatgcgatag
agctaaacctttcatgtcctcacggtgtgcatataatgtctggtatggatgtaat
tgaagagatggtcaacacaacaaaacttgtcaaaagcaatgttaagataccagtg
atacccaaaatgactcctcaatctacaaatccgggatctgatgccttaagactcg
acagtgcaggagcagacgggcttgtaatgttcaatagatttacagggcttgacat
tgatattgagaaagaagcacccattttgcacggcggttatgcagggcatggtggt
ccgtgggcaattatgtatggtttgaggtggataagcgctgtatcgccaaaagtga
aatgtagtatcagtgcaagcggcggtgccatgaatggagaagatgttgtcaaata
catattggcaggtgcgtcggctgttcaagtttgcacaactgttattttgaatggc
tatggggttataaaaaagataaacaagtatttagaagagtacatggagagaaaag
gttacaacacaattgatgattttaaaggaaaggtgtgcagtagaattcttgacat
ggactctgttgacagaacgcactgggctgttgcaaggattgacaaagaaaaatgc
acatcttgtggcaagtgcttcacagtttgcatatatgatgcaattgaaaaggatg
atggaaagtttaaagtaaatcaaaactgcgatggctgcggactttgtgcagaact
gtgcccagccaaggcaatcttaatggtaagaagaggtgaagtttaa

Amino Acid Sequence:
MPNLSTTYAKLNLRTPVIVASAGITGTVERLQRCEENGAGAVVTKSLFQKEICRI
APTPRFKIVKHENTFTLYSYEQASEFNPQEYAEFIFKAKQKLSIPVIASINCYTD
DAWLEYSKLMEQAGADAIELNLSCPHGVHIMSGMDVIEEMVNTTKLVKSNVKIPV
IPKMTPQSTNPGSDALRLDSAGADGLVMFNRFTGLDIDIEKEAPILHGGYAGHGG
PWAIMYGLRWISAVSPKVKCSISASGGAMNGEDVVKYILAGASAVQVCTTVILNG
YGVIKKINKYLEEYMERKGYNTIDDFKGKVCSRILDMDSVDRTHWAVARIDKEKC
TSCGKCFTVCIYDAIEKDDGKFKVNQNCDGCGLCAELCPAKAILMVRRGEV*

Figure 23-55

Athe Gene Number:
Athe_2086

Product Function:
Beta-glucuronidase

Nucleotide Sequence:
ttgtacatcattatattccactttcagtacatttttattttatcattaacacctt
ttgaagttgagattaataaatttgcccaaattggctgtgagaatagacttacaat
tgttgtaaacaacatcttggactggagttgtcttccaccagggtttataagggaa
tacaatgacccaatgcatccagaagggtataaaactcaggaatatctttttgact
ttttcaactattcaggtattcacagaccagttttgctctacaccacttccaaaac
atatattgaggatattaagattgaaacccagattgagggtcaaaagggtatagtt
tgctttaaggtggctgtaagtggcgaaaaaaaggatgaatgtcagatagcagtag
ctttgtatgacaaagatggaaagcaaatagcaaaggtcgaagggccagagggtat
gatagaggttggagatgcgatattttgggagccttcaaatccatatctttacaaa
ctaaatgtaactttaatacacgatgaaaaggtggtagatgaatattatcttcctg
tgggaataaggacagttgaggtaaaaggcaaaagacttttcctaaatggtaagcc
agtgtatcttaaaggtttggcaaagcatgaagacagtgatataaggggcaaggga
tacgaccctgtgatagctgtgaaagatttcaacctcctaaaatggataggagcaa
actcattcagaacatcacattatccttacgcagaagagattttaaacttggcaga
cgagtatggtttttggtaattgacgaggcaccagctgttggcatgaatttcttt
aacaaaaacgaaaaagtgtttaccgcggagagagtaaaccaaaagacattagaac
atcacttagaagttataagacaacttattgcaagggataaaaaccatccaagtgt
gattatgtggagtgtggcaaatgaggctgcaacatatgaagatggggcatatgaa
tatttcaaaagagtaatagatgaggtgagaaagcttgacccgacaagaccggtga
cgctggttgaatcctcttttccagatgagaccaaagtgggaagtcttgttgatgt
tatatgtgtaaacaggtactattcatggtattctgatcctggcagactggatttg
atagagttccagcttgaaaaggagctgaaaaggtggtttgagctttatcaaaaac
cagtgataataacagagtatggggcagatacaattgcaggatttcattcaagtcc
tccaatgatgttttctgaggaatatcagtgtgagatgcttgaaagatatcatagg
gtgtttgacaggctggattttgtgataggcgaacacatatggaactttgcagact
ttgcaacaaaacaagaggttcgaaggattatgggcaacaggaaaggaatctttac
aaggcaaagacagccaaaagccgcagctttcttgctcaaaaaaagatggcaaaat
tcagagcacaaaaggctggaggaaaatgtttcagaagataaaacacgtaattaa

Amino Acid Sequence:
LYIIIFHFQYIFILSLTPFEVEINKFAQIGCENRLTIVVNNILDWSCLPPGFIRE
YNDPMHPEGYKTQEYLFDFFNYSGIHRPVLLYTTSKTYIEDIKIETQIEGQKGIV
CFKVAVSGEKKDECQIAVALYDKDGKQIAKVEGPEGMIEVGDAIFWEPSNPYLYK
LNVTLIHDEKVVDEYYLPVGIRTVEVKGKRLFLNGKPVYLKGLAKHEDSDIRGKG
YDPVIAVKDFNLLKWIGANSFRTSHYPYAEEILNLADEYGFLVIDEAPAVGMNFF
NKNEKVFTAERVNQKTLEHHLEVIRQLIARDKNHPSVIMWSVANEAATYEDGAYE
YFKRVIDEVRKLDPTRPVTLVESSFPDETKVGSLVDVICVNRYYSWYSDPGRLDL
IEFQLEKELKRWFELYQKPVIITEYGADTIAGFHSSPPMMFSEEYQCEMLERYHR
VFDRLDFVIGEHIWNFADFATKQEVRRIMGNRKGIFTRQRQPKAAAFLLKKRWQN

Figure 23-56

SEHKRLEENVSEDKTRN*

Figure 23-56

Athe Gene Number:
Athe_2087

Product Function:
Integrase catalytic region

Nucleotide Sequence:
atgcagaacttaacagcacatttaaacatgataagggcgatgaaaatgaaaccta
acttttcagaacttgcaagaatatatgggatggatagaagaacagttaaaaaata
ttatgagggttatgaaggaaaacctaagaatagaaataaaccaagtaaattggac
aaatactatgatgagataaaatcaaagcttgctatcaaaggagttacagtcaagg
gtgtttatgagtatttaaaatcaaaagatgagacaataggaacatattcaaactt
caataagtatgttaagaaaaaaggattaaagccagagaagaaaataaaaggtcac
ccaagatttgagacagatccaggtgagcaagcgcaagttgattggaaagagaata
taaagcttgtctcaagaaatggagaggagtttatcattaatgttcttgattttaa
attaggttattcaaggtattgctgctttgagataaacaggacaaaaactcaagaa
gaattaatagaaactctaataagaatattcaaagatataggcggagtaccgagag
agattttatttgacaatacagcagcagttgttgatataacaggtgagaaaattaa
agtaaattcaagatttaaaagttttgcaaaagactttgggtttgaagtaaaactg
tgcaaaccaagacattcgtacacaaaaggaaaagttgaagcagcaaacaagttta
tagattggatactgccatatcagggtgaatttgaaacagaagaggacttagtaag
gataataaaagagataaacgcaaaggtcaatatgcagccaaatcaaacaactcaa
gttccacctgctcttctgtttcaaaaagaaaaagagtatttacaacccttgccag
acaaaaggttaatagacagttacctaaattcctacaagtcagttaaagtccaaaa
ggactctctgatttactacaagggaagtaaatactctgttccacccgaatacata
ggaaagacagtccaagtaaaggaggtggaaaacaaaatttatatttattataaca
caaacctgttaaggatacatgttattgatgaaaaaaatatcaattatcacgatga
agattacaagcagctaatgctaatgagagttggtcaaagagaagagcttaacaag
atatgtgaggaaaacctaaagaaatttgataatctgttgaaaacctaa

Amino Acid Sequence:
MQNLTAHLNMIRAMKMKPNFSELARIYGMDRRTVKKYYEGYEGKPKNRNKPSKLD
KYYDEIKSKLAIKGVTVKGVYEYLKSKDETIGTYSNFNKYVKKKGLKPEKKIKGH
PRFETDPGEQAQVDWKENIKLVSRNGEEFIINVLDFKLGYSRYCCFEINRTKTQE
ELIETLIRIFKDIGGVPREILFDNTAAVVDITGEKIKVNSRFKSFAKDFGFEVKL
CKPRHSYTKGKVEAANKFIDWILPYQGEFETEEDLVRIIKEINAKVNMQPNQTTQ
VPPALLFQKEKEYLQPLPDKRLIDSYLNSYKSVKVQKDSLIYYKGSKYSVPPEYI
GKTVQVKEVENKIYIYYNTNLLRIHVIDEKNINYHDEDYKQLMLMRVGQREELNK
ICEENLKKFDNLLKT*

Figure 23-57

Athe Gene Number:
Athe_2088

Product Function:
IstB domain protein ATP-binding protein

Nucleotide Sequence:
atgagcaactatgtgaaactacttaacaacttagaagagttaggtcttcacaaca
taaagaataaccttgacaaatacttagatttagtggcaagcggagaaaaaagtat
gacagatgcattatatgaacttagtaatttagagataaaagccaaagaagaaagg
gcgatattaggatgtgtgagggtggcaaatttcccatttataaaaggtatagaag
attttgatttttcatttcagccaagtataaacaagcaacagataatggacttaat
gagtttgagatttttagagggtaatgaaaatatactatttgtcggaacaccaggg
gtagggaaaacgcatctagccacagcaataggtatagagtgtgcaaaacgaaggt
attcaacatattttatacattttcaagagttaatagcccagctaaagaaagcatt
attggagaacagattagagtacagacttaagcatttttcgaaatacaaagtttta
ataatagatgagataggttatttgccaatagacaatgatggagcaaatttatttt
tccagctgatatcgagcagatatgagaagagcagtacaataataacaactaatgt
tgtattctcagaatggggagagatatttggtggagcgacaatagcaaatgcaatt
ttagataggctactgcatcattcttacgtgattttcataaaaggtccttcataca
gattacagtcaaaaacagcatattttagcaatacaaaccagcaaagttaa

Amino Acid Sequence:
MSNYVKLLNNLEELGLHNIKNNLDKYLDLVASGEKSMTDALYELSNLEIKAKEER
AILGCVRVANFPFIKGIEDFDFSFQPSINKQQIMDLMSLRFLEGNENILFVGTPG
VGKTHLATAIGIECAKRRYSTYFIHFQELIAQLKKALLENRLEYRLKHFSKYKVL
IIDEIGYLPIDNDGANLFFQLISSRYEKSSTIITTNVVFSEWGEIFGGATIANAI
LDRLLHHSYVIFIKGPSYRLQSKTAYFSNTNQQS*

Figure 23-58

Athe Gene Number:
Athe_2089

Product Function:
Beta-glucuronidase

Nucleotide Sequence:
atgctttatccacgtgagacaagaacaagagagataaaagacttaagcggtatat
ggaggtttaaagttgacagggacaataaaggctatgaacagaggtggtttgaaaa
acccctttgaagatgctattttgatgcctgtgccgtcaagctacaatgacataacc
caggatgagagcataagagatcacataggtgatgtgtggtatgaaaggacatttt
acatacctgaaggttggaaggacaaaagaatagttttgagagttggaagtgctac
tcatcacgcaagagtatttttaaatggtaaggaaatagcccagaacaaaggtggg
tttttgccttttgtaaatgtcaatatcaaaatgaacaaaaaatcgaaaataaaaa
tgtacaaaaaataa

Amino Acid Sequence:
MLYPRETRTREIKDLSGIWRFKVDRDNKGYEQRWFEKPLEDAILMPVPSSYNDIT
QDESIRDHIGDVWYERTFYIPEGWKDKRIVLRVGSATHHARVFLNGKEIAQNKGG
FLPFVNVNIKMNKKSKIKMYKK*

Figure 23-59

Athe Gene Number:
Athe_2090

Product Function:
hypothetical protein

Nucleotide Sequence:
atgcaagaaataaagctaaaatggcttgtcaaaccgcaggtaggcaatattggtg
caacgttttgtattccatggaaaaaagctcagctctggaatacagaaaatgttat
tgtcagaaacaaggatggaagtattattcctactcaaaactgggcactttcatac
tggggtgatgggtcggtaaaggtgagtagccacgctgctttctttggcaccaatt
tacctgatgagctctttgcatgtattgaagaagataccgtaacaaaattcgaaac
aatggttgaagtttatgaagacaaggagcatattagcgtgacaacaggcaagctt
gagtgcaaaattagcaagaccggagataaaattattgaatatttaaaagtaggac
aaaagattatttgctctaatctcacgcaaattcttattacagaaagaatttctaa
ttttacaggatataagactaaagttgaagaagaatacattcctcaaatattaggt
gcaaaagtagagtctcacggacctttgagatgtgttataagggtgtggggaagac
acttgcggaataccaagatgacatttgatggctctttactcaaacaagggtggtt
gccttttgagttgagattatatttctatgcaaattcagataaaattaagattgtt
catacatttatttacaacggaaatccccatcaagattttataaaaggattgggct
taaaatttgatcttcctttgcactctcctctgtacaatagatttgtaagatttgg
aggggacagtggactttttgtgagtcaccaaaaaatttgatagatgtgtacaga
aaagagagatacaaaataatgtttcaaaatcaggtcaatggacaaccagtagatt
ttgatattgacgaggaccaccattatattagggagattaacaatactccagagtg
ggatgaatttattctttttcaagattcttccgagcactatgttattaaaaaacgt
acatcttgccagtgcagttttgtaaaagcaacagaaggcagaagatctatggggt
ttgcatatattggtgatgagaatggtgggcttggagtagctttaaaggatttttg
gcaaaaatatccttcaggatttgaggcaaaagatttaacttcttctgaggcaaat
ttaatagtttggctttggccaccttatgctgaagttatggatttgagacactatg
attttggtacatacacatcttcttcatatgaagggtttgatgagtatagaagtac
tccctatggtatagctaatacaaatgagctgtggctgtactgttttgattatgct
ccaacaagtgagcagcttttggaatatgcaaaaatgcaaggattttctcctcttc
ttgtatgtagccctgaaaggtataaagagacagagatatttggaggattaaattt
accggataatagcaactccaagaaagaaaagattgaaagaatactgacagctatt
gttgatttttatctaaaggaagttgagagaagaaagtggtacggtttttgggact
atggtgattttatgcatagctatgatatggtaagacatatgtggaaatatgatat
tggtggatatgcttggcagaacacagagcttgtgccaaacatttggctttggctt
atgtttttaagaacaggacgttttgatatatttaaaatggctgaggcaatgacaa
gacatacatcagaagtagatgtgtatcatcttggggaatacaagggactgggttc
aaggcacaatgttgttcattggggatgtggttgcaaagaagtgagaattagcatg
gcataccttcacaggtattactattatcttactgccgatgaaagaatagggcagc
ttatggatgatgtaaaagatattgacaaacaaattatgcacatggaccctatgcg
tgcgtattttacgaatgatcctgaaaatagagtacacataagggttggaccggat
gttatgacctttgcgccaattggtttgtaagatgggaaaggcacaaggaagaaa
tttacaaacaaaagataataaagatactggactttttgaagaaaaatccagcagc
atttatttcgggtggggtatttgactatgatcctgaaaaaacccagttaaaacct

Figure 23-60

gttgaatacactggtggttcaaattttgtattttgttttggaaatacacttgttt
ggcttgagatagccaagaattttgaagataaagaatttgaagaattgtgtgctca
gcagggcctttttatactgaatttaaagaaaacaaaaatgagattttaaaaagc
tggggtgttccaagctttggatttaagttgaacatgctcaatattgggatggctg
cctttgctgcaatgaagaaaaatattcctgaattgaaaaaagaaatatggcagat
gtttttagaccataacaaaaatccatggctaaaatttatttgtgatggaggtata
aatttacaattggcaactgtacctgcgcctgttgctgagattccttttatttcaa
cgaatatagcttctcaatggtcattaaacgcattgcttgcgctagaatttattgg
agatgagctatga

Amino Acid Sequence:
MQEIKLKWLVKPQVGNIGATFCIPWKKAQLWNTENVIVRNKDGSIIPTQNWALSY
WGDGSVKVSSHAAFFGTNLPDELFACIEEDTVTKFETMVEVYEDKEHISVTTGKL
ECKISKTGDKIIEYLKVGQKIICSNLTQILITERISNFTGYKTKVEEEYIPQILG
AKVESHGPLRCVIRVWGRHLRNTKMTFDGSLLKQGWLPFELRLYFYANSDKIKIV
HTFIYNGNPHQDFIKGLGLKFDLPLHSPLYNRFVRFGGDSGLFCESPKNLIDVYR
KERYKIMFQNQVNGQPVDFDIDEDHHYIREINNTPEWDEFILFQDSSEHYVIKKR
TSCQCSFVKATEGRRSMGFAYIGDENGGLGVALKDFWQKYPSGFEAKDLTSSEAN
LIVWLWPPYAEVMDLRHYDFGTYTSSSYEGFDEYRSTPYGIANTNELWLYCFDYA
PTSEQLLEYAKMQGFSPLLVCSPERYKETEIFGGLNLPDNSNSKKEKIERILTAI
VDFYLKEVERRKWYGFWDYGDFMHSYDMVRHMWKYDIGGYAWQNTELVPNIWLWL
MFLRTGRFDIFKMAEAMTRHTSEVDVYHLGEYKGLGSRHNVVHWGCGCKEVRISM
AYLHRYYYYLTADERIGQLMDDVKDIDKQIMHMDPMRAYFTNDPENRVHIRVGPD
VMTFCANWFVRWERHKEEIYKQKIIKILDFLKKNPAAFISGGVFDYDPEKTQLKP
VEYTGGSNFVFCFGNTLVWLEIAKNFEDKEFEELCAQQGLFYTEFKENKNEILKS
WGVPSFGFKLNMLNIGMAAFAAMKKNIPELKKEIWQMFLDHNKNPWLKFICDGGI
NLQLATVPAPVAEIPFISTNIASQWSLNALLALEFIGDEL*

Figure 23-60

Athe Gene Number:
Athe_2091

Product Function:
extracellular solute-binding protein family 1

Nucleotide Sequence:
atgagtaaagttaagaagattttgaggttatttgtagttagtatctgtattgcag
gtttgttgttaacatcaataggagtattgggtgcatcaaaatcaaagtattcgtt
taaactcacaatcatggctcagtattttgggacagagcctgcgccatcaaatagt
ccagtaattttaaaagcagaacagtacctaaaaactgacctggagtttacatggg
tacctgccgatggttacaatgacaaattaaacattatgttagcaagtggaaatct
tccaatggtggtttatgtaccaggaaaaactgcgtctataattggtgcatgtaaa
gcaggagcattctgggaacttggaccgtatataaagcaatataagaatttaaaag
caatcccggatatagttctgtggaactcttccattgatggtaaaatatatggtat
tccacgttcaaggaccttgggaagaaatggtattgtttatagaaaagactgggct
aaaaatgttggtatcactaagcttgaaactattgatgatttgtacaacatgttaa
agaagtttacttacaatgatccagataaaaatggaaaaaatgacacatatggaat
gattgtttgcaactacaatggaccattctatattactctcacatggtttggtggt
ccgaatggatggggattgaacaaaaacggacagttggtgccaactttcttgacaa
atgcatatatagaaaatttaaaattctggcgaaagatgtatcaagaaaagctgtt
taaccatgacttcccgagtgttccaggtgcaagatgggaggattattactcacag
ggtaaaggtggagtaaagattgacgttatagactctgcaaacagaatttacaatg
gcctcaagaacaacggtatcattccaaaagatgcaaaggatacagatataatgga
cgttgttgtgtctgtgaaaggtaaatatggtctaagaaatatgccgacttctggt
tatgcggggtatctgatggtttcaaagacaagtgtaaaagatatgaatacattca
agaaagtaatgtcgatattagataagtttggcgacagaacaatgcaggatctgtt
tggttatggactaccgaacagacattacaaacttgtagatggtaaaatagatcct
attccaaacttgccagcagatttgtcaagagaaataagcggacttaaccaggttc
tacatttttatcctgcaaacggtggaacaccaagatatatgactccacttttgca
gctgcaggcagatatgcaagctttgaatgaaaagctcaatattcttgtgccaaat
ccggcagaagcattgataggtatgtcagatacttatatcaaacgcggtgtcacac
ttgataatatgatagaggatgcgcgcgttaaatatattactggtcagctcaatga
ccagggctttaagaaagttcttgataactggagaaggcagggaggagaccagatt
ataaaagaagtaaatgcactgtatcgtaaatataaaaagaatattccttacaaag
aagacatatataagattctcaatccttag

Amino Acid Sequence:
MSKVKKILRLFVVSICIAGLLLTSIGVLGASKSKYSFKLTIMAQYFGTEPAPSNS
PVILKAEQYLKTDLEFTWVPADGYNDKLNIMLASGNLPMVVYVPGKTASIIGACK
AGAFWELGPYIKQYKNLKAIPDIVLWNSSIDGKIYGIPRSRTLGRNGIVYRKDWA
KNVGITKLETIDDLYNMLKKFTYNDPDKNGKNDTYGMIVCNYNGPFYITLTWFGG
PNGWGLNKNGQLVPTFLTNAYIENLKFWRKMYQEKLFNHDFPSVPGARWEDYYSQ
GKGGVKIDVIDSANRIYNGLKNNGIIPKDAKDTDIMDVVVSVKGKYGLRNMPTSG
YAGYLMVSKTSVKDMNTFKKVMSILDKFGDRTMQDLFGYGLPNRHYKLVDGKIDP
IPNLPADLSREISGLNQVLHFYPANGGTPRYMTPLLQLQADMQALNEKLNILVPN

Figure 23-61

PAEALIGMSDTYIKRGVTLDNMIEDARVKYITGQLNDQGFKKVLDNWRRQGGDQI
IKEVNALYRKYKKNIPYKEDIYKILNP*

Figure 23-61

Athe Gene Number:
Athe_2092

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
atgagacagaataagacagtagcaagtactatttttgatgtttttaatcatatat
tccttggaatatgggcgataataactgttttacctttcctatacgttttggctgc
atcttttgctccagattcagaaataaaaacgaggacattctttataattcctcat
aatcctactcttctcacctataaatttatttttgcttcaaactattttcttcgta
gtatgctcaacagtgtgattatcacggtaggaggtaccttggtaaatctgttttt
cacatttacaatggcatatgctctttcaaaaaagcactttataggcagaagtatt
gtcttaaatggcataatatttacaatgctttttggtggaggaatgattcctacat
atttgctggtaaaaagcctgggtctcttaaattcttattgggcactatggttgcc
aggtgcaattagtccctttaacttctttgttgtcaaaaactttttccaggaaatg
ccacaggatttagaggatgcagcaagaattgatggatgtacagaagctcaagtgt
tgtggaagataatacttcctctttcaaagcctattattgcgacatttgctttgtt
ctatggagtggggcattggaattcatggtttggagcgcttttgtatataaacgat
gctgagaaatggcctgtgcagctaattttaaggcaaatagtaatgttgtcaacca
cgcttgcttcggacctaacacaatttgatcctaatttccaacctccacaggagtc
gctgaagatggcaataattgtcgtagcaactttaccaataatgcttttgtatcca
tggcttcagaagtacttcataaaaggtatgtttattggttcattgaaggagtga

Amino Acid Sequence:
MRQNKTVASTIFDVFNHIFLGIWAIITVLPFLYVLAASFAPDSEIKTRTFFIIPH
NPTLLTYKFIFASNYFLRSMLNSVIITVGGTLVNLFFTFTMAYALSKKHFIGRSI
VLNGIIFTMLFGGGMIPTYLLVKSLGLLNSYWALWLPGAISPFNFFVVKNFFQEM
PQDLEDAARIDGCTEAQVLWKIILPLSKPIIATFALFYGVGHWNSWFGALLYIND
AEKWPVQLILRQIVMLSTTLASDLTQFDPNFQPPQESLKMAIIVVATLPIMLLYP
WLQKYFIKGMFIGSLKE*

Figure 23-62

Athe Gene Number:
Athe_2093

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
atgaaaaacaatattatcaagggaccaaaaaagaagggagagaatgatttaatga
cagctaccacttctatatggaaaagactgaaaaaagacaaatggctgtatatctt
agctctaccgggtattttgtacttcatcatttttaggtacattccaatgtttggt
atagtagttgcctttcaagattttaatccgtttttagggttctggaaaagtccgt
gggtaggttttgagcattttaaaacacttttcaccgaccctgattttccgatgct
gtttagaaatacactgttaatttcattttacaatatacttttctatttccctgtg
ccaattattttagcgctactgatcaacgaggtgcgaaatcaggtttataagagaa
ttgtccagacgtgtgtttatgttcctcactttgtttcaatggtggtcatagcaag
tattacatacgtgcttctctcaagtgaggcaggggttataaataatattctgtat
agcttaacaggtaaaaaaattgaatttttgacagatccaagatggttcagacctc
ttataataattcagagcatatggaaagaagcaggatggggaacaataatcttctt
ggcagcactatcaaatgttgacccaacattatatgaggctgctattgtggatggg
gcaacgaggtggcagcagacgtggcatatcacaataccctcaattatgagtacgg
taatcatacttttcattttgcgattgggacatcttcttgatactggctttgaaca
gatatttttgatgaaaaatcctattaacaggtcggtagcagaagtgtttgacaca
tatgtttatcaagtaggggttacccaaggagcgtacagttacagtacagcggttg
gtctttttaagtctgttgttggactgattttgattcaggtttctaactatctgtc
caagaaatttactgaaacttcattgttctaa

Amino Acid Sequence:
MKNNIIKGPKKKGENDLMTATTSIWKRLKKDKWLYILALPGILYFIIFRYIPMFG
IVVAFQDFNPFLGFWKSPWVGFEHFKTLFTDPDFPMLFRNTLLISFYNILFYFPV
PIILALLINEVRNQVYKRIVQTCVYVPHFVSMVVIASITYVLLSSEAGVINNILY
SLTGKKIEFLTDPRWFRPLIIIQSIWKEAGWGTIIFLAALSNVDPTLYEAAIVDG
ATRWQQTWHITIPSIMSTVIILFILRLGHLLDTGFEQIFLMKNPINRSVAEVFDT
YVYQVGVTQGAYSYSTAVGLFKSVVGLILIQVSNYLSKKFTETSLF*

Figure 23-63

Athe Gene Number:
Athe_2094

Product Function:
glycoside hydrolase family 2 sugar binding

Nucleotide Sequence:
atgaatgagagtataaaattgaattctcttttgatgaaatcaacaaccctccag
ctactgccagtattattcattggtggatttttctgatgagatgaacgagaacag
aataaatgctgagcttgattatatttcaaatcttggctttaagcaagtattaatt
gcagtaggacacaatgtttcgcctaaatatttgacacatggttggtttgaaatgg
taaaatttgcagttctccaagctaaaaaaagaggagttaaagtgtggattgccga
tgaagggacatatccaagtggctttgctggcgaaacttttaataagaagtatcct
cacaaaaggatgaaggctattattgttgagaaggagtttattattgagggtaatt
tatgtgaagttgaacctcactctggtacaattgggattttggccaaagacatgaa
ccagaataaatactttgcttttgaaaagcttgaatttagtagcggatttttatac
ttgccctatcattcgacttggcaaataaaagtaatatcttcagcttacaggacat
ctccaacaagatacgttcaccatccaacaggtgcaaaagatactacattttcact
ttgtgattatcttgactatgaagctgtcaatctattcataagtgaggtatatgaa
aaatataaagcttatatgggaaatgaatttggaaagacaataattggatttttcg
ctgacgaacctgattattctatttctggactaccatatacggataatatatttga
tatattttacaatgaacaaggatacgacgttaaaaagtacataccgtatttcttt
aaagagcaattagatgaaaaaataaaaagagtaaaagcagattactgggatgtat
ggagcaatatttttacaaatactttctttaagcagatctacaaatggtgtgaggc
aaatggcctcaaatttgtagtacatctaaatcatgaagatatgatagaacacctt
accaaatctgaaggacagttcttttcgcatatgaagtatgttcatattccagcaa
ttgatgtaatttggagacaaatctggtatgacaaaaaagcaatattccctaaata
cgcttcttctgtttctcatattaaaaatattgctcagacctttcagagagtttt
gcagtatatggacaaggtatatctgttgagcaaatcaaatgggtagttgattacc
agtttgcaatggacataaatctatttttgacctcaatcttcaagtatctttatga
ccatccgcaaaattatttctttccagaggtaattaagtatattaataccatttca
tatcttctctatgtaagcacccttgtacaaaggttctggtttactttcctacac
cggatctgtgggcaggtgaaaatatgtctgcttcaaaagcaatggaaattggcaa
tgcacttttagagaaccagattgattttgatttttttgaccattctcttttagaa
tatctggaaattaaaaaccatagaatatacgctaacaatagaaaagaatacgaca
ttgttattcttccgcctataaagtatttgccacaagatctgttcagatttttaaa
gcttttctcaagcaaaggagggaagattattttcttcgagaactctcctttgttt
gtttataacaaaacctttacatcgtttttccactttgtagatagagaaataggtg
tggttgttgaaagtatcgagcagctttcaaaaatggttgaaaaagatgtcactgt
tgtagacagcaaagatgttagagttcttcataaaagaatagaaggcaataatctg
atttttctcttcaatgtttcaggtacttcatttttgggtaagataatattaaaat
tttctaagaaaaatgtatatatatgggatcatatacagaataaatttttaatggt
ttcaaatatcaaaagtaataaaaaaaacatacaattagaactctatatacatcca
tatcagactttggttttaatagcaagtgatgagtatgtagatggaattcaaaaaa
caacactgcttggaagcttaccgagaacagtcttggaattaaacgataactggga
aattcattttgataaagattttgttttgttttcagatttaaaagattggcaaagc

Figure 23-64

ttgggctttggtgactattctggcagtgtagtttatagaaaaatattttcgtttt
ctcatgatgactttattaaaaataaacatcttttcctcaactgccccaatgtaaa
gtactctgcaaaggtttggttaaataaaagatatcttggtgtaagagctttttcg
ccttttatgtgggatataacagaggcattgaaaattggtgagaatgaacttgtga
ttgaagttcaaaacacccctgcagcagctctacttggaacacaagaaaaattgga
aaaattaagaaaagaggcagagaagaacttttatctttctatttctctaaaattt
gacctggaaatggtccaatcaggattgttgcctccagttgctattgtttctttag
aatga

Amino Acid Sequence:
MNESIKLNSLFDEINNPPATASIIHWWIFSDEMNENRINAELDYISNLGFKQVLI
AVGHNVSPKYLTHGWFEMVKFAVLQAKKRGVKVWIADEGTYPSGFAGETFNKKYP
HKRMKAIIVEKEFIIEGNLCEVEPHSGTIGILAKDMNQNKYFAFEKLEFSSGFLY
LPYHSTWQIKVISSAYRTSPTRYVHHPTGAKDTTFSLCDYLDYEAVNLFISEVYE
KYKAYMGNEFGKTIIGFFADEPDYSISGLPYTDNIFDIFYNEQGYDVKKYIPYFF
KEQLDEKIKRVKADYWDVWSNIFTNTFFKQIYKWCEANGLKFVVHLNHEDMIEHL
TKSEGQFFSHMKYVHIPAIDVIWRQIWYDKKAIFPKYASSVSHIKNIAQTFSESF
AVYGQGISVEQIKWVVDYQFAMDINLFLTSIFKYLYDHPQNYFFPEVIKYINTIS
YLLYVSTPCTKVLVYFPTPDLWAGENMSASKAMEIGNALLENQIDFDFFDHSLLE
YLEIKNHRIYANNRKEYDIVILPPIKYLPQDLFRFLKLFSSKGGKIIFFENSPLF
VYNKTFTSFFHFVDREIGVVVESIEQLSKMVEKDVTVVDSKDVRVLHKRIEGNNL
IFLFNVSGTSFLGKIILKFSKKNVYIWDHIQNKFLMVSNIKSNKKNIQLELYIHP
YQTLVLIASDEYVDGIQKTTLLGSLPRTVLELNDNWEIHFDKDFVLFSDLKDWQS
LGFGDYSGSVVYRKIFSFSHDDFIKNKHLFLNCPNVKYSAKVWLNKRYLGVRAFS
PFMWDITEALKIGENELVIEVQNTPAAALLGTQEKLEKLRKEAEKNFYLSISLKF
DLEMVQSGLLPPVAIVSLE*

Figure 23-64

Athe Gene Number:
Athe_2371

Product Function:
glycoside hydrolase family 43

Nucleotide Sequence:
atgctaaaaagacaggatatttatatccgtgacccttttattatgcctgtaaaag
aaaaaagggtgtactatatgtttggtacaactgataaaaattgctggggaaatga
aaaagcaacaggctttgattattatgtaacctatgaccttgaaaattttgatggg
ccatatccagcattcaggcctcctcagaatttttgggcagacagaaatttctggg
cgccagaggtgcacttttataacggcaaatattatatgtttgcaacattttgttc
accagaaggaaaaaggggaactgcgattcttgtttcagatactgtagaaggacca
tatgttgaacatagtcttggacctgttacaccaaaggattggatgtgccttgatg
gaacactttacatagatgaagattcaaatccctggcttgtattttgccgtgaatg
ggttgaggtttatgatggtcagattttggcagcaagactttcaaatgatttgaaa
agagttattgatgagcccgttttactctttacagcctccagcgcaccttggacag
caattataagagttaaagatggcagagaatgttttgtgacagatggtccgttttt
attcagaacccaaaacggtactcttttgatgctatggtcaagttttgatagagaa
agaagatattgtgttggagttgcaaaatcactgtcaggcaatattttaggtccat
ggcagcacaaaacagagccaatttttctaatgacggtggtcatggcatgctttt
tagaacctttgaaggtgaccttatgctttcaattcatactccaaattcaaggggc
aatgagaggcctttgtttataaaaattgatgataaaaatttagaagacgagtttt
ga

Amino Acid Sequence:
MLKRQDIYIRDPFIMPVKEKRVYYMFGTTDKNCWGNEKATGFDYYVTYDLENFDG
PYPAFRPPQNFWADRNFWAPEVHFYNGKYYMFATFCSPEGKRGTAILVSDTVEGP
YVEHSLGPVTPKDWMCLDGTLYIDEDSNPWLVFCREWVEVYDGQILAARLSNDLK
RVIDEPVLLFTASSAPWTAIIRVKDGRECFVTDGPFLFRTQNGTLLMLWSSFDRE
RRYCVGVAKSLSGNILGPWQHKTEPIFSNDGGHGMLFRTFEGDLMLSIHTPNSRG
NERPLFIKIDDKNLEDEF*

Figure 23-65

Athe Gene Number:
Athe_2372

Product Function:
histidine kinase internal region

Nucleotide Sequence:
atgataagaaaaaatggagtttataaaagcaagattttcaaaaagaattttgtgc
agatagttattgttcccattgtgataataactattcttgggcttttctcatgcat
cattatagaacaatacgttaaaaatgaaataaacaaaaatttagagacaatgcta
atacaaagcaaaaacaatgtcgagcttatgctcggtgagatagactatctttata
tggtatttgggataaacaaagatgtgacccttcagataaagaggattttgaactc
aatgtatttttctttagaagatatctggcagattaacatggtcaaaaatgtttta
aattcaatctcatattcaaagccgttcatacattccatctatgtttatttcgaaa
atcctgaagggaattttatagttaccccagatggaatgactaattttcagtattt
ttatgacaaatggtggtttgatcagtataaagaaaataaagcattaatgtgggta
gagagaagaaaaattcaaccttacaattttactggagaatcaattgatgttttga
ccatctacaaaaggataaaatctgcatattctgatgtgaatgagggtgttattgt
tcttaatctgtattacgaccaggtaaaaaagctcttaagccttaaaagttcgctc
cctcagcatgcaatgtacatattagatcaaaatggaaatgttttggtatcaaatg
aatcagataactctaatacctcaagtatggccctcctaaaaaaagaaacagacaa
ctatctcacaaaaagattagagtcaaaaaaatacaacttaacctttgtttcagta
attcccaaaaattatctttacagcatccctatcagacttttcaaggtgacactgg
tgctacttttaatttttatagttatcgcttttgctgcctcatactacattgccaa
agtaaattacaggaatattaaaaagattatagatacaataaattcagcaacagaa
ggaaaaccaccaaaagaaattaaaattacttcaaatgatgaatatgggtatatca
tgtacaatgtaatcaagaactttattgaaaaacattatctgacaacacgccttca
agctttggagcttttagcattgcaggctcagattaaccctcattttcttttcaat
actttagaacacatatatcttaaaactttagcacttacaggcaccccaaacgaga
ttacaaaaatgatagaaaacctttcggctatactcaaatattctctgagcaatcc
aaaaattactatcttcctaagggaagaaattaaagctacacaggcatatattgag
cttgtaaaagcaagatataaagataagtttgatgtgttttgggactatagtgaag
atgtgcttgagataaaagtgatgaagcttttattccagccgctcatagaaaattc
aatctatcatgggataaaaccttgcgaaaagagatgtggaataaaaatcaggata
agaaaattaaaagataccagtgattggctttgtatatgggtaattgacaatggaa
ttgggatgagcaaagaaaagttagaggaggtacaaggcaggctttcacaggattt
tgacttttcagatcatattgggcttttaaacaccaatgaaaggttaaagctcaac
tatgggggtaactttaaactcaaggtttggagcaagctgggtttggggacaattg
taaaaataattcttcctgtgaattttgaggaccgaaaggagaatgaaatagatgc
taaaaagacaggatatttatatccgtga

Amino Acid Sequence:
MIRKNGVYKSKIFKKNFVQIVIVPIVIITILGLFSCIIIEQYVKNEINKNLETML
IQSKNNVELMLGEIDYLYMVFGINKDVTLQIKRILNSMYFSLEDIWQINMVKNVL
NSISYSKPFIHSIYVYFENPEGNFIVTPDGMTNFQYFYDKWWFDQYKENKALMWV
ERRKIQPYNFTGESIDVLTIYKRIKSAYSDVNEGVIVLNLYYDQVKKLLSLKSSL

Figure 23-66

PQHAMYILDQNGNVLVSNESDNSNTSSMALLKKETDNYLTKRLESKKYNLTFVSV
IPKNYLYSIPIRLFKVTLVLLLIFIVIAFAASYYIAKVNYRNIKKIIDTINSATE
GKPPKEIKITSNDEYGYIMYNVIKNFIEKHYLTTRLQALELLALQAQINPHFLFN
TLEHIYLKTLALTGTPNEITKMIENLSAILKYSLSNPKITIFLREEIKATQAYIE
LVKARYKDKFDVFWDYSEDVLEIKVMKLLFQPLIENSIYHGIKPCEKRCGIKIRI
RKLKDTSDWLCIWVIDNGIGMSKEKLEEVQGRLSQDFDFSDHIGLLNTNERLKLN
YGGNFKLKVWSKLGLGTIVKIILPVNFEDRKENEIDAKKTGYLYP*

Figure 23-66

Athe Gene Number:
Athe_2373

Product Function:
two component transcriptional regulator, AraC family

Nucleotide Sequence:
atgtcttacaagatgattatcgtggaggatgaaagtgagataagacaggggcttt
ttacctgctttccatggaataaacttggttttgaagttgttggcctgtttgaaaa
tggaaagcaggctttagattatattttgcaaaataaggttgatgttgtttttgc
gatattaaaatgcctgttatgagtggaattgaccttgcaagggttttgtttttaa
ataacatccctgtaaaagttgtttttataagtggttatcaggactttgaatttgc
tcaaaaagcaatgaaatacggggtaaggtactacataacaaaaccagcaacgtat
gatgaagtaatagagatatttgggctaataaaaaaagagcttgacctacagtcaa
aaaatccgcaaactcacagtgaagaggtaaaaaaacaagatttaatttctattga
caatcctgaaaatgtgattgaaactgtgaaagagtatataaaaagaaactataaa
gaggccacgcttgaggatgctgcaaagcttgtctatatgaatccttattatctta
gccggcttttcaaatgcaaaactggtaaaaacttttctgattatctgatggaagt
caggatgagaaaagctctggaacttatgaaaattcctatatataaacttcatgag
ataggtgaaatggttggttacaagaaccctaagaactttacaagagcttttaaaa
aatatttcggcaagtctccttcagagtttgttcaggagagataa

Amino Acid Sequence:
MSYKMIIVEDESEIRQGLFTCFPWNKLGFEVVGLFENGKQALDYILQNKVDVVFC
DIKMPVMSGIDLARVLFLNNIPVKVVFISGYQDFEFAQKAMKYGVRYYITKPATY
DEVIEIFGLIKKELDLQSKNPQTHSEEVKKQDLISIDNPENVIETVKEYIKRNYK
EATLEDAAKLVYMNPYYLSRLFKCKTGKNFSDYLMEVRMRKALELMKIPIYKLHE
IGEMVGYKNPKNFTRAFKKYFGKSPSEFVQER*

Figure 23-67

Athe Gene Number:
Athe_2374

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
atgtactcccaaaacaaaatagtaaaaaatactatttctctttttataacttatg
catttttgattctatttggactgtttatgatttatccacttttgtgggttgtatc
tgcggcatttaaatcaaatgatgagatattcaaatcgctgtctttatttcctcaa
aaaattgttacagactcatttataaaaggatggcaggggacagggcaatatacat
ttggtagatttttgcaaataaccttatacttgttattccagttgtatttttac
aataatatcctctacacttgtggcatatgggtttgcaaggtttaattttccactg
aaaagattatttttgtgatactaatttcaacattaatgctgccagactctgtaa
accttattccaagatatattcttttcaatgcatttggatgggtagatagttacaa
accatttataattccatcaatgtttgcatcaactccatttttgtgtttatgatg
atacagtttatgagagggcttccaagagaaatagaagaagctgctataattgatg
gctgcaactctttccagatacttttgcgtataacgggtcctctttgtaaaacagc
aatgatttcaatgggaattttccagtttatctggacttggaatgactttttagga
ccccttatatatatcaacagcgttgaaaaatacacaattgcccttggacttagaa
tgtgtgttgacagtgcagctgcaattgcttggaaccagatcatggctatgactgt
tatagcaatgctaccgtgtattatcatcttctttgctgctcaaaaatacttcgtt
gaaggaattgcaacaagcggaataaaaggttaa

Amino Acid Sequence:
MYSQNKIVKNTISLFITYAFLILFGLFMIYPLLWVVSAAFKSNDEIFKSLSLFPQ
KIVTDSFIKGWQGTGQYTFGRFFANNLILVIPVVFFTIISSTLVAYGFARFNFPL
KRLFFVILISTLMLPDSVNLIPRYILFNAFGWVDSYKPFIIPSMFASTPFFVFMM
IQFMRGLPREIEEAAIIDGCNSFQILLRITGPLCKTAMISMGIFQFIWTWNDFLG
PLIYINSVEKYTIALGLRMCVDSAAAIAWNQIMAMTVIAMLPCIIIFFAAQKYFV
EGIATSGIKG*

Figure 23-68

Athe Gene Number:
Athe_2375

Product Function:
binding-protein-dependent transport systems inner membrane component

Nucleotide Sequence:
atggtagttcgctatagaaaaaaagactttgttggtttttatatatactacctt
ggcttattggatttttgatttttaggctatatccatttattatgtctttctacta
ctctttcagtgattatacaatgttgaagccgcctcgctatgttggattatataac
ttcatttatatgttcacaaaggatgaactgtttccaaaagcacttttaaacacta
taaagtatgttataataactgttccgcttaaaatctcgtttgcactttttgttgc
aataatattgaacatgaaactgaaaggaataaatcttttcagaacagtatattat
cttccttctatcttcggtgggtctgttgcaatctcgattttgtggagatttttgt
ttatgaaagaaggtatagtaaacaagttcttaagtctttttagaatagaaggtat
aaactggcttggagacccaagaatagctatgttttcagtaagccttcttgcggtg
tggcagtttgggtcatctatggtactatttttagcaagacttaaagagataccat
cagaactttacgaagcagcactggttgatggagcatcaagactaaaaatgtttac
aaagataactcttcctatgatttcacctataatgtttttcaaccttgtgatgcag
accataaatgctttccaagaatttactggaccgtacatcatcacaggcggtggac
ctgtcaattccacctatcttttgagtatgctcatatatgacaatgcatttaagta
ttttagaatgggttatgcggcagcactttcctgggttcagtttgtgataatatta
atctttactgcatttatatttaggtcttctacttattggacatattacgagtatg
atgaagggaggttctaa

Amino Acid Sequence:
MVVRYRKKDFVGFLYILPWLIGFLIFRLYPFIMSFYYSFSDYTMLKPPRYVGLYN
FIYMFTKDELFPKALLNTIKYVIITVPLKISFALFVAIILNMKLKGINLFRTVYY
LPSIFGGSVAISILWRFLFMKEGIVNKFLSLFRIEGINWLGDPRIAMFSVSLLAV
WQFGSSMVLFLARLKEIPSELYEAALVDGASRLKMFTKITLPMISPIMFFNLVMQ
TINAFQEFTGPYIITGGGPVNSTYLLSMLIYDNAFKYFRMGYAAALSWVQFVIIL
IFTAFIFRSSTYWTYYEYDEGRF*

Figure 23-69

Athe Gene Number:
Athe_2376

Product Function:
extracellular solute-binding protein family 1

Nucleotide Sequence:
gtgataaaaagtaaaaggttaattgcaactcttgtgttagtagtttttactatgt
cgatcttctttgctttttcgaccgctgggtctgagaaagctaaggcagcatcgaa
aaaggttacactcaggtttatgtggtggggcggagaggcaagacacaaagccact
ttggcagcaattcaggcgtatatgaagaaataccctaatgtaagaattaatgcag
agtatggcggtattgaaggttacatgcagaagctcattacccagcttgtgggaag
aaccgctccagatataatccagattgacgttacatggattggtgagctgagcagc
cagggagatttctttgcagaccttaaaactttcaaagaggtcaacttaaagccat
ttgaagagaagtttttaaaagactggtgctattcaaacggaaaacttattggact
tccaacaggtgttaatgcttcggtacttcaatataataaagagtttttaagaag
tttaatatcgacgaaaatacagtttggacgtgggataacttactatcaatagctg
aaaaagtacacaaaaaagataaaaatagttatttgcttaattttgatcaaatcct
ctgttactatgttttgacatcgtatattggtcaaaaaacaggaaaggattggatt
ttagatgattatacattaggatttaataggaatcagttgatagaagcttttactt
atttgaaaaaattatttgatgtaggagctattcaacctttgcagaaagtgcgcc
atttcaaggtaagccagaacaaaatccaaaatggctaaaaggagaattagggatt
ttatggaattggacttcaacttatgctgcaaataaagctatgattccaagtttgg
cgatgacattaccacccaggggcaacaacttgaaaaactatgcagtaactgtcag
accgtcgcaattgttatctgtaaataaactttcgaagaatgctaaagaagctgca
aaatttattaactggtttttaaacgataaacaagctgctctaatactcactgatg
taagaggagttcctgcaagttcaagtgccagagatgcattgttaaaggcaaataa
attagatccagaaatattgagggttacaaacgaagcggtaaagtatgcagcaaaa
ccacagaatgcactatcacagaatcaagaaatagcaaatatagcatatgatatca
ttcagcagcttgcatacaaacagctaacaccgacacaggctgcagataaattgat
agcattatataaacaaaaactttctgagcttaaaagaatgcagtctcgatag

Amino Acid Sequence:
VIKSKRLIATLVLVVFTMSIFFAFSTAGSEKAKAASKKVTLRFMWWGGEARHKAT
LAAIQAYMKKYPNVRINAEYGGIEGYMQKLITQLVGRTAPDIIQIDVTWIGELSS
QGDFFADLKTFKEVNLKPFEEKFLKDWCYSNGKLIGLPTGVNASVLQYNKEFFKK
FNIDENTVWTWDNLLSIAEKVHKKDKNSYLLNFDQILCYYVLTSYIGQKTGKDWI
LDDYTLGFNRNQLIEAFTYLKKLFDVGAIQPFAESAPFQGKPEQNPKWLKGELGI
LWNWTSTYAANKAMIPSLAMTLPPRGNNLKNYAVTVRPSQLLSVNKLSKNAKEAA
KFINWFLNDKQAALILTDVRGVPASSSARDALLKANKLDPEILRVTNEAVKYAAK
PQNALSQNQEIANIAYDIIQQLAYKQLTPTQAADKLIALYKQKLSELKRMQSR*

Figure 23-70

Athe Gene Number:
Athe_0423

Product Function:
glycosyltransferase 28-like protein

Nucleotide Sequence:
atgaaaaaagtattcttacaatttggttctgggttaggtcccatgtctcgatccc
ttccaatcgcagaagggttacaaagagaaggatatatcgtaaagtattttggatt
tgaaaatgcaaaaccatatatgaataaaataggaattgaagagttatcagaaaat
tttaatatcaaggatattaaaaaaggagtgcaaactcctaattggtattgtgcag
agcaattttgggaaataataggatatggtaatatggaatgggtagaaaaaaaagt
tgaagaattaatagaatatttaaaagatttttctcctgattttataatatcagac
cttggtatattaagttgtattgctgcaagaataatggacatacctttgatagcta
taactcaaagttgttatcatcctaacattgcttttggaagaataagatggtggga
agaagaacaaaatttaaagtttacattaactgagaaattaaataattattttaag
aaaaaaggtgtttcacaattaaattcttttgaagaaatttttactggtagtttaa
ccataattcccagttttcctgaatttgatccaataaataatccttcagaatttaa
cacatattatgttggtcccatattatgggatccattagacatggctaaagaagag
tatataaaattgtttaacagagataaaaataagcctacaatttttgctatacag
caagattttacgacaatgtgggtgaaagtggaattattatttttaaaacattact
ttcagcattaaaaaaatttgatgctaacattatttttctacagggagtgattcg
gacaggaaaatagcaaaagagattttaaactcttacggaattgatgaagagaaat
ttagcattattgattgggttccaatgggaattgcttatggaaactctgatgttgt
tatccatcatggaggccatggaagttgtttaggtcaatttttgtatgaggtacct
tcattaattatacctactcatactgaacgagagtataatgcaagaatttgcacca
atatgggagtttctaaatttataaaaagagaagacattgaaaaagcagatatatt
agctgaaattgttgagattttaactaactcaagttttaaagaaagattacacttt
tggcatactaaactaaatgaatataattttacaggtgtaaataaagttttagaat
tgattcaaaaattataa

Amino Acid Sequence:
MKKVFLQFGSGLGPMSRSLPIAEGLQREGYIVKYFGFENAKPYMNKIGIEELSEN
FNIKDIKKGVQTPNWYCAEQFWEIIGYGNMEWVEKKVEELIEYLKDFSPDFIISD
LGILSCIAARIMDIPLIAITQSCYHPNIAFGRIRWWEEEQNLKFTLTEKLNNYFK
KKGVSQLNSFEEIFTGSLTIIPSFPEFDPINNPSEFNTYYVGPILWDPLDMAKEE
YIKLFNRDKNKPTIFCYTARFYDNVGESGIIIFKTLLSALKKFDANIIFSTGSDS
DRKIAKEILNSYGIDEEKFSIIDWVPMGIAYGNSDVVIHHGGHGSCLGQFLYEVP
SLIIPTHTEREYNARICTNMGVSKFIKREDIEKADILAEIVEILTNSSFKERLHF
WHTKLNEYNFTGVNKVLELIQKL*

Figure 23-71

Athe Gene Number:
Athe_0603

Product Function:
xylose isomerase

Nucleotide Sequence:
atgaagtacttcaaagacattccagaagtaaaatatgaaggaccacagtcggaca
acccatttgctttcaagtactacaatcctgacgaaatcattgacggcaagccttt
gaaagaccaccttcgttttgctattgcttactggcacacattctgtgcaacagga
agcgatccgtttggacaacctacaattgttcgtccttgggataagttttcaaacc
gaatggacaacgcaaaagcaagggttgaggcagcatttgaatttttgaactgtt
agatgtaccatttttctgcttccatgacagagatattgcacctgaaggggaaaat
ttaaaagagtcaaataagaatttggatgagattgtttctttaataaaagagtatt
tgaaaaccagcaagacaaaagtattatggggaacagcaaacctattttcacatcc
gcgatatgttcatggtgctgcaacatcctgcaatgccgatgtttttgcatatgca
gcagcgcaagtgaaaaaggcgttagaggttacaaaagagcttggcggcgaaaact
atgtgttctggggcggaagggaaggttatgagacacttctaaatacagatatggg
attggaacttgataaccttgcaagatttttgcatatggcggttgagtatgcaaag
gaaataggttttgacggacagtttttaatagaaccaaaaccaaaagagccaacta
agcatcagtacgattttgattcggctcatgtttatggatttttgaaaaagtatga
tcttgacaaatacttcaagctcaacatagaggtaaaccatgcaaccttagcagga
catgatttccaccatgagttgagatttgcgcgaataaacaacatgcttggttcaa
ttgacgctaacatgggcgatttgcttttgggctgggatacagatcagttcccaac
agatgtaagacttactacacttgctatgtatgaggttattaaagctggtggtttt
gacaaaggtggacttaactttgacgcaaaggtaagaagaggttcttttgagcttg
aagacttggtcattggtcacattgctggcatggatgcttttgctaaaggcttcaa
gattgcgtataagcttgttaaagatggcgtatttgataaatttatagatgagaga
tacaagagctacaaagaaggaatcggtgctaagattgtaagcggtgaagcaaact
tcaagatgttagaggaatatgctctgtctcttgacaagatagaaaataaatctgg
caagcaagagcttcttgagatgattttgaacaaatatatgttcagcgaataa

Amino Acid Sequence:
MKYFKDIPEVKYEGPQSDNPFAFKYYNPDEIIDGKPLKDHLRFAIAYWHTFCATG
SDPFGQPTIVRPWDKFSNRMDNAKARVEAAFEFFELLDVPFFCFHDRDIAPEGEN
LKESNKNLDEIVSLIKEYLKTSKTKVLWGTANLFSHPRYVHGAATSCNADVFAYA
AAQVKKALEVTKELGGENYVFWGGREGYETLLNTDMGLELDNLARFLHMAVEYAK
EIGFDGQFLIEPKPKEPTKHQYDFDSAHVYGFLKKYDLDKYFKLNIEVNHATLAG
HDFHHELRFARINNMLGSIDANMGDLLLGWDTDQFPTDVRLTTLAMYEVIKAGGF
DKGGLNFDAKVRRGSFELEDLVIGHIAGMDAFAKGFKIAYKLVKDGVFDKFIDER
YKSYKEGIGAKIVSGEANFKMLEEYALSLDKIENKSGKQELLEMILNKYMFSE*

Figure 23-72

Athe Gene Number:
Athe_0610

Product Function:
glycoside hydrolase starch-binding

Nucleotide Sequence:
atgagaatgctattaaaaaggtctcttgctctgctggtaagtattgtccttgtat
tttcgttgttcttaagtgtttttccacagcaagcaagagcacaagataccataaa
aattgtaggtaactggcaggatgctggtaactggaattttgacagctctaatatt
gttttgagtgaaacctcaactccgggactttattatggtgagtatacttttaaaa
ctggtggttcttatgagttcaaagcagttataaatggtagtatatggtgtacagg
cgctccaaaagtagcagataataatactaacatacctcttaatgttacagatggt
cagacagttaagttttggttcttcaaaaattctcaacttgtaattgacagcactc
attttccaaatgggcctgatagtttagtagggcaaaattcttttaaatttgtagg
agtaaataatgaatggaaccctaatgatggtaagtatcaatttataagagtccca
gatgcaacatatacctatttatatgataatagttacaatgatttttcgataccat
atggatttaagattataataggtggttttgggaacctatcatgggcatggaatgg
tgcaaaagaaggttcagttgttaaatttaaagagggtggagataatattgacctg
aaagaatttaaagaatctaatgataatcttctaaaaacaaaatttttccttgata
tcctaaacggctggctctttacagaaaaagatttgacaaatatccaaccagtaga
ttttgcaaataacggctctgtaattggtggaacttcagttcatcttacatggaca
ccatatacaccaaatgcaaatccactttcagccaagctttactacaaaataattg
atgtgagcaatgatagcgagttggttagtctcaccgattacagcacaatccacaa
aatagatatcccaagagagtggataggaaagacaattaaaattattgcaaatgcc
aagataggtgaaataacaggacctaatgttgagtttacattaaatattgttgact
tgccacaaaacttaatagttgaagcagctaactatataacatacaactctataga
tgagacatttttaaatttggctcaaggtgatacagttcaaaacataacttctaac
ttcgcagttgttacatcatatgtttacaatgtggtgtacgagggaaatagttatc
cactgacatttaatattgactgggagagcagtaattcatctgtattaacaataaa
tggtagtacagttgtggttacaaggccaacacaaggtgaccttcaggttgagttg
agagcaagagcaagatttggtgctatttcagcagatggacaaaaaatatttactc
ttacagtaaagaaattcctattaggagtagatggtggaatacctgttacatttaa
cgtaacagttccagattatactcctgataatgacaatatttacattgctggagat
tttaagactgataagcttccaaaatgggaccctgtaggtatcaaattgataaagg
tcggagataaaaaatatagtataacaatgtatcttccaccgaatgtaacaattga
atataaatatacccgcggaagttggtcaaaagtagaaaaggatgcatttggtaat
gaaatatcaaatagagttctgcagataaataatcaggctgtgacaaaaaatgatg
tagttgaagcatttgcggaccttggagctgtcaagcaaggtcttccaacggtgaa
tttggtcatcaatgttcctcaacaaactgtaattgatacaggtgatggaggaagg
gtaataaaagcagaaggtggtgctatttcagttccaaatactaagaacactgttt
tgatagatattcggagtgcattaaatgctattatattaaagtaa

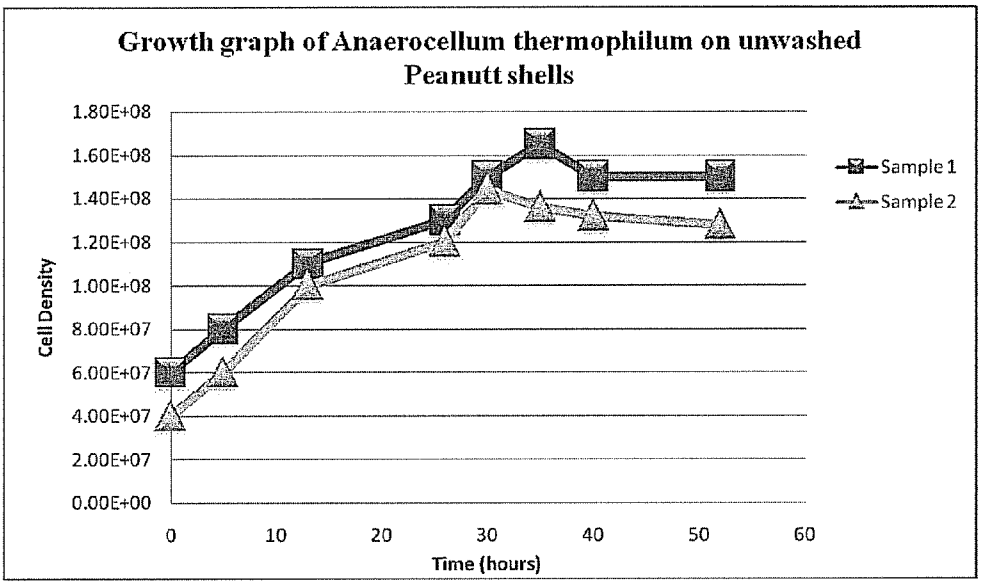
Amino Acid Sequence:
MRMLLKRSLALLVSIVLVFSLFLSVFPQQARAQDTIKIVGNWQDAGNWNFDSSNI
VLSETSTPGLYYGEYTFKTGGSYEFKAVINGSIWCTGAPKVADNNTNIPLNVTDG

Figure 23-73

QTVKFWFFKNSQLVIDSTHFPNGPDSLVGQNSFKFVGVNNEWNPNDGKYQFIRVP
DATYTYLYDNSYNDFSIPYGFKIIIGGFGNLSWAWNGAKEGSVVKFKEGGDNIDL
KEFKESNDNLLKTKFFLDILNGWLFTEKDLTNIQPVDFANNGSVIGGTSVHLTWT
PYTPNANPLSAKLYYKIIDVSNDSELVSLTDYSTIHKIDIPREWIGKTIKIIANA
KIGEITGPNVEFTLNIVDLPQNLIVEAANYITYNSIDETFLNLAQGDTVQNITSN
FAVVTSYVYNVVYEGNSYPLTFNIDWESSNSSVLTINGSTVVVTRPTQGDLQVEL
RARARFGAISADGQKIFTLTVKKFLLGVDGGIPVTFNVTVPDYTPDNDNIYIAGD
FKTDKLPKWDPVGIKLIKVGDKKYSITMYLPPNVTIEYKYTRGSWSKVEKDAFGN
EISNRVLQINNQAVTKNDVVEAFADLGAVKQGLPTVNLVINVPQQTVIDTGDGGR
VIKAEGGAISVPNTKNTVLIDIRSALNAIILK*

Figure 23-73

Unwashed shells



Washed shells

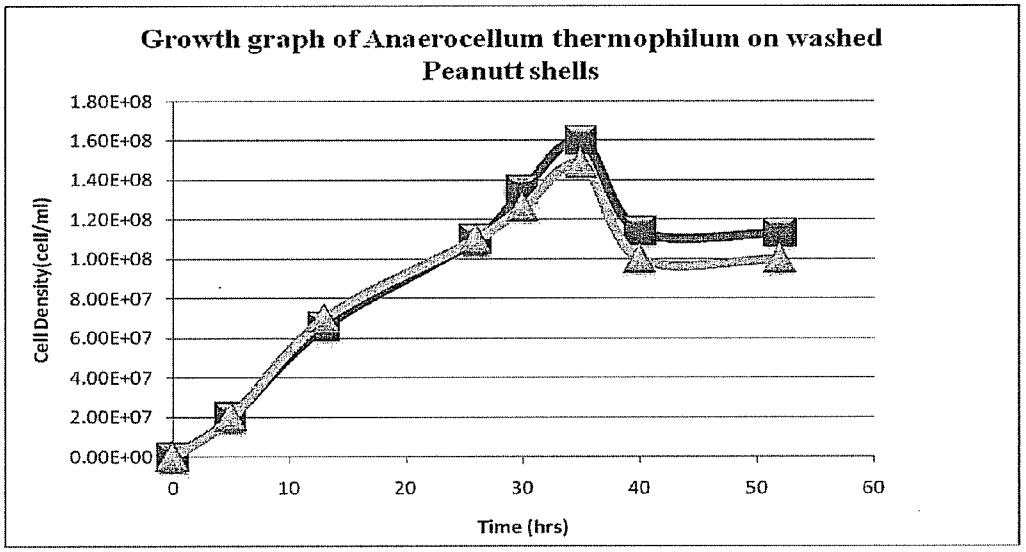
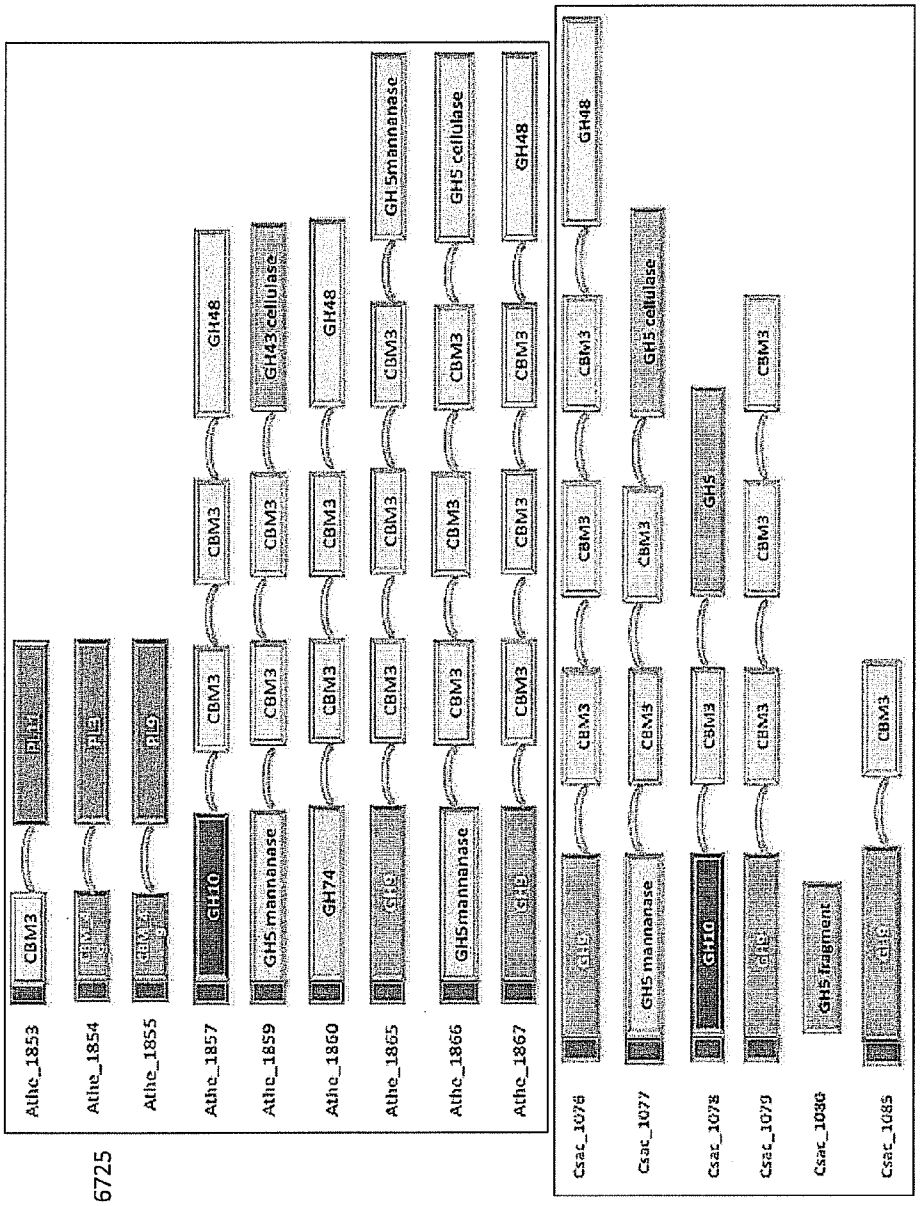


Figure 24



A. thermophilum DSM 6725

C. saccharolyticus DSM 8903

Figure 25

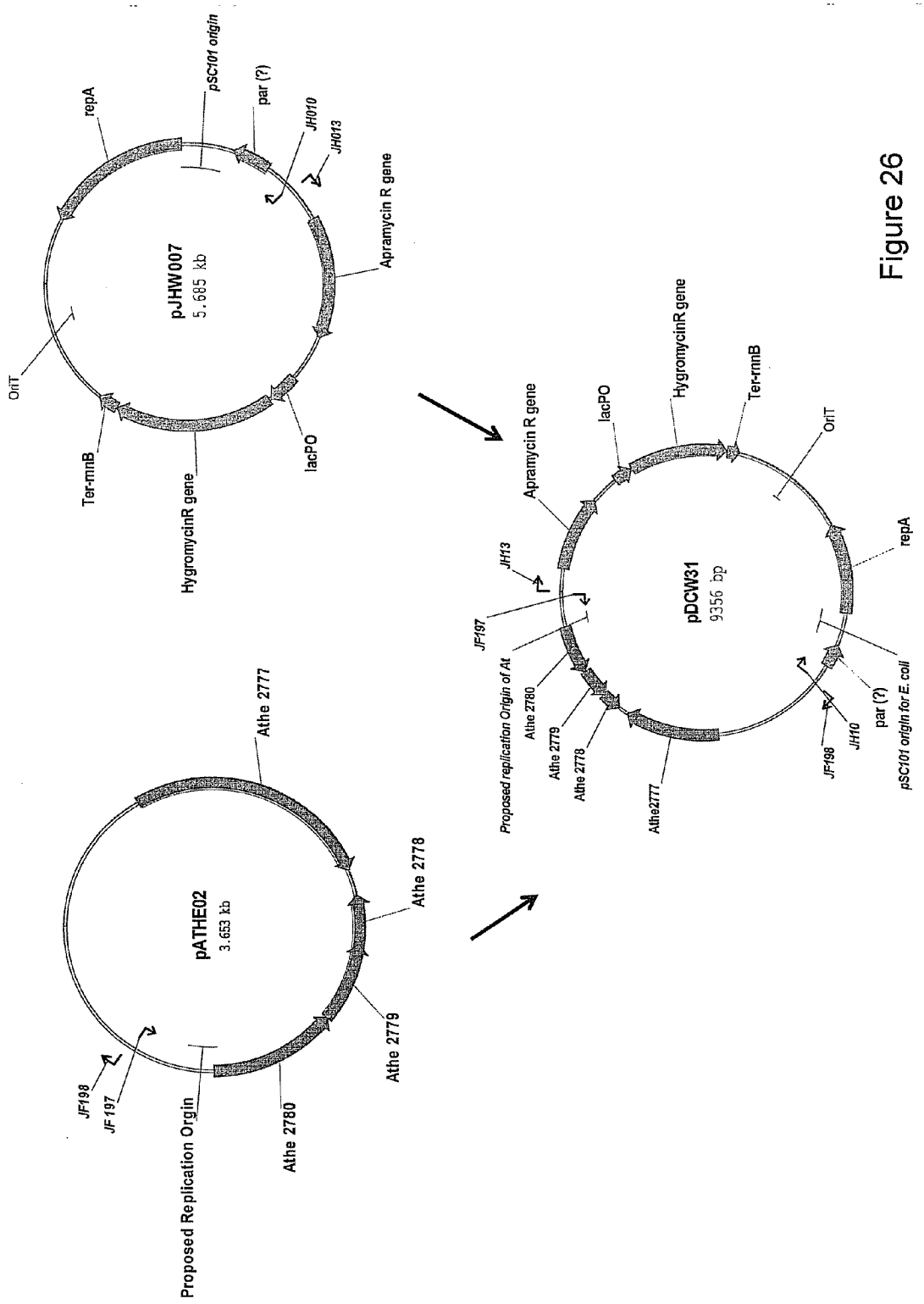

pJHW007
5.685 kb
repA
pSC101 origin
par (?)
JH010
JH013
Apramycin R gene
lacPO
HygromycinR gene
Ter-rrnB
OriT
pATHE02
3.653 kb
Athe 2777
Athe 2778
Athe 2779
Athe 2780
JF198
JF197
Proposed Replication Orgin
pDCW31
9356 bp
Apramycin R gene
lacPO
HygromycinR gene
Ter-rrnB
OriT
repA
JH13
JF197
Proposed replication Origin of At
Athe 2780
Athe 2779
Athe 2778
Athe2777
JF198
JH10
par (?)
pSC101 origin for E. coli

Figure 26

Common domains in DSM 6725 and DSM 8903.

Domain	DSM 6725	DSM 8903	Total	Diff
CBM22	5	11	16	6
CBM3	16	10	26	6
GT2	11	8	19	3
GH43	3	6	9	3
GT4	10	12	22	2
GH48	3	1	4	2
GH10	5	6	11	1
CBM48	4	3	7	1
GH3	1	2	3	1
GT51	1	2	3	1
CBM20	2	1	3	1
CBM6	2	1	3	1
GT1	2	1	3	1
GH13	7	7	14	0
GH5	6	6	12	0
CE4	4	4	8	0
CE9_inc	4	4	8	0
GH2	4	4	8	0
GH23	2	2	4	0
GH28	2	2	4	0
GH39	2	2	4	0
GH4	2	2	4	0
GH65	2	2	4	0
GH9	2	2	4	0
GH94	2	2	4	0
GT28	2	2	4	0
GT35	2	2	4	0
CBM28	1	1	2	0
CBM41	1	1	2	0
CE1	1	1	2	0
CE7	1	1	2	0
GH1	1	1	2	0
GH105	1	1	2	0
GH15	1	1	2	0
GH29	1	1	2	0
GH30	1	1	2	0
GH31	1	1	2	0
GH36	1	1	2	0
GH42	1	1	2	0
GH51	1	1	2	0
GH67	1	1	2	0
GT26	1	1	2	0
GT39	1	1	2	0
GT5	1	1	2	0

Figure 27

Domain	DSM 6725
CBM_X	2
CBM36	1
CE12	1
CE9	1
GH11	1
GH44	1
GH74	1
PL11	1
PL3	1
PL9	1

Figure 28

DSM 6725	Domains				
Gene name					
Athe_0459	CBM_X	GH94			
Athe_0460	CBM_X	GH94			
Athe_1857	GH10	CBM3	CBM3	GH48	
Athe_0182	GH43	CBM22	GH43	CBM6	
Athe_1859	GH5	CBM3	CBM3	GH44	
Athe_1866	GH5	CBM3	CBM3	CBM3	GH5
Athe_1865	GH9	CBM3	CBM3	CBM3	GH5

DSM 8903	Domains				
Gene name					
Csac_1090	GH94				
Csac_1091	GH94				
Csac_1078	GH10	CBM3	GH5		
Csac_2411	GH43	CBM22	GH43	GH43	CBM6
Csac_1077	GH5	CBM3	CBM3		
Csaz_2528	GH5				
Csac_1079	GH9	CBM3	CBM3	CBM3	

Figure 29

Domains enriched in each genome

Domain	DSM 6725	DSM 8903	Diff
CBM22	5	11	6
CBM3	16	10	6
GT2	11	8	3
GH43	3	6	3
GT4	10	12	2
GH48	3	1	2

Figure 30

	Protein ID	SS	TMDs	Culture time (hr) 21	24	30	42	45	48	CAZy family	Description
	Athe_0165							2		GH13	alpha amylase
1	Athe_0459			45	31	33	26	64	46	CBM_X-GH94	Cellobiose phosphorylase
1	Athe_0460			25	21	16	16	76	25	CBM_X-GH94	Cellobiose phosphorylase
	Athe_0468			3		2	4		2	GH36	alpha-Galactosidase
	Athe_0594	Yes	1					11		GH5-CBM28-SLH-SLH-SLH	Endoglucanase
	Athe_0609	Yes	1		7	4		2		CBM41-CBM48-GH13-CBM20	Pullulanase, type I
2	Athe_1853		1				30	30		PL11-CBM3	Rhamnogalacturonan lyase
1	Athe_1854	Yes	1	30	49	51	42	61		CBM_4_9-PL3	Pectate lyase
1	Athe_1855	Yes	1	78	117	77	104	64		CBM_4_9-PL9	Pectate lyase
2	Athe_1857		1	11	25	20	113	118	5	GH10-CBM3-CBM3-GH48	Xylanase/cellobiohydrolase
2	Athe_1859	Yes	1	10	14	10	130	99	4	GH5-CBM3-CBM3-GH44	Mannanase/endoglucanase(xyloglucanase)
2	Athe_1860	Yes	1	10	10	4	285	110	5	GH74-CBM3-CBM3-GH48	Xyloglucanase/cellobiohydrolase
3	Athe_1865	Yes	1	35	52	25	155	149	4	GH9-CBM3-CBM3-CBM3-GH5	Endoglucanase/mannanase
3	Athe_1866	Yes	1	9	7	7	84	88	4	GH5-CBM3-CBM3-CBM3-GH5	Mannanase/cellulase
2	Athe_1867	Yes	1	210	218	114	462	302	11	GH9-CBM3-CBM3-CBM3-GH48	Endogluconase/cellobiohydrolase
	Athe_2028							8	7	GH4	alpha-galactosidase
	Athe_2354			6	7	8	5	17	5	GH3	beta-glucosidase

1	high all the time
2	high at late stationary phase
3	steady increase but decrease at the end

SS: signal peptide, indicating that protein is secreted
TMDs: transmembrane domains, indicating membrane-related or -bound

Figure 31

Gene name	AAs	SP	Domain structure	Proteomics	
				ExtP	Memb
Athe_0012	3027	Y	**SLH-SLH-SLH-Fn3**-VWA-(19x **RHS**_repeat)	C	CXP
Athe_0077	1710	Y	**SLH-SLH-SLH**-Transglut_core	C	CP
Athe_0174	951	Y	**CBM6**-X-**SBP1**		CXSP
Athe_0181	595	N	**SBP1-CBM22**[1]	C	CXSP
Athe_0594	755	Y	GH5-**CBM28-SLH-SLH-SLH**	C	
Athe_0608	547	Y	X-**SLH-SLH-SLH**		
Athe_0438	1157	Y	**SLH-SLH**-X	C	C
Athe_1839	575	Y	**SLH-SLH**-X		P
Athe_1943	277	Y	**SLH-SLH**		
Athe_2295	1074	Y	**SLH-SLH**-X		ASP
Athe_2341	484	Y	X-**SLH-SLH**		
Athe_1573	1055	Y	**SLH-SLH-** SpoVT_AbrB	C	CP
Athe_2303	1018	Y	**SLH**-X	CXS	CXSP
Athe_2342	1010	Y	**SLH**-X		C
Athe_1944	1201	Y	X-**Fn3**-X	C	
Athe_1945	1265	Y	X-**Fn3**-X	C	
Athe_0190	582	N	X-**LysM**-G5-peptidase M3		
Athe_0508	203	Y	**LysM**-SH3_3		
Athe-0560	507	Y	**LysM-LysM**		
Athe_1391	109	Y	**LysM**		
Athe_2402	511	N	X-**LysM**		

SLH: surface layer homology domain, binds to cell wall
CBM: binds to carbohydrates (6, cellulose, xylan: 22, xylan: 28, cellulose)
SBP1: solute-binding protein, binds to membrane
Fn3: fibronectin type 3 homology domain, found in many CAZy enzymes
LysM: binds to peptidoglycan
RHS: has a YD motif involved in binding to carbohydrates
X: PFAM unrecognizable domain/s

C: cellulose
X: xylan
S: switchgrass
P: poplar

Figure 32

Gene Name	Length	MW (kDa)	SS	Annotated Function	LC-MS/MS
Athe_023	220	24.9	-	Alkyl hydroperoxide reductase	CSP
Athe_0287	110	12.7	+	Hypothetical Protein	CXSP
Athe_0416	122	13.8	-	Conserved hypothetical protein	P
Athe_1869	307	34.1	+	NHL repeat containing protein	XS
Athe_2076	1039	121.4	-	Putative uncharacterized protein	S
Athe_2209	132	15.0	-	Thioesterase superfamily protein	S
Athe_2653	784	90.5	+	Hypothetical Protein	P
Athe_2688	552	64.8	-	Hypothetical Protein	S
Athe_2719	784	90.6	+	Hypothetical Protein	P

C: cellulose; X: xylan; S: switchgrass, P: poplar

Figure 33

METHODS, MICROORGANISMS, AND COMPOSITIONS FOR PLANT BIOMASS PROCESSING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Ser. No. 61/190,181, filed Aug. 26, 2008.

STATEMENT OF GOVERNMENT RIGHTS

[0002] This invention was made with government support under a grant from the Department of Energy, Grant No. DE-PS02-06ER64304. The U.S. Government has certain rights in this invention.

BACKGROUND

[0003] Biofuel can be broadly defined as solid, liquid, or gas fuel derived from recently dead biological material. The derivation of biofuel from recently dead biological material distinguishes it from fossil fuels, which are derived from long dead biological material. Biofuel can be theoretically produced from any biological carbon source, but a common source of biofuel is photosynthetic plants. Many different plants and plant-derived materials may be used for biofuel manufacture.

[0004] One strategy for producing biofuel involves growing crops high in either sugar (e.g., sugar cane, sugar beet, and sweet sorghum) or starch (e.g., corn/maize), and then using yeast fermentation to produce ethyl alcohol (ethanol). One challenge associated with this strategy is that competition between food markets and energy markets for the crops can increase food costs.

[0005] Thus, a second strategy involves converting biological material such as, for example, wood and its byproducts into biofuels such as, for example, woodgas, methanol, or ethanol fuel. It is also possible to make cellulosic biofuel—e.g., cellulosic ethanol—from non-edible plant parts. Cellulosic biofuel production can use non-food crops or inedible waste products. Thus, producing cellulosic biofuel need not divert food crops away from the animal or human food chain. Moreover, in some cases, biofuel can be produced from material that would otherwise present a disposal problem.

[0006] Producing biofuel from cellulose can be economically challenging, however. It often involves multiple processing steps to break down the cellulose and convert the biological material into material that is, or can be readily converted to, biofuel. Each processing step can make the overall process more costly and, therefore, decrease the economic feasibility of producing biofuel from cellulosic biological material. Thus, there is a need to develop methods that reduce the number of processing steps needed to convert cellulosic biological material to biofuel and other commercially desirable materials.

[0007] *Anaerocellum thermophilum* was first described in 1990. *A. thermophilum* DSM 6725 is a strict anaerobic microorganism with a temperature optimum at 72-75° C. It is freely available from a public culture collection at DSM-Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH, Mascheroder Weg 1b, D-3300 Braunschweig, Germany, under the accession number DSM 6725.

SUMMARY OF THE INVENTION

[0008] The present invention relates to methods, microorganisms, and compositions useful for processing plant biomass. The application of this technology has the potential to render production of biofuels more economically feasible and to allow any microorganism to utilize recalcitrant biomass. The use of cellulosic materials as sources of bioenergy is currently limited by typically requiring pretreatment of the cellulosic material. Such pretreatments can be expensive. Thus, methods that reduce dependence of existing pretreatments of cellulosic materials may have a dramatic impact on the economics of the use of recalcitrant biomass for biofuels production.

[0009] In one aspect, the methods described herein involve processing plant biomass. Generally, the methods include growing *Anaerocellum thermophilum* on a substrate that comprises plant biomass under conditions effective for the A. thermophilum to convert at least a portion of the plant biomass to a product that may be water soluble or water insoluble. In some cases, methods described herein can yield both soluble and insoluble products that are more readily converted to biofuel, a polymer, or commodity chemicals than unprocessed plant biomass. In other cases, the methods themselves can include converting the plant biomass to biofuel, a polymer, and/or a commodity chemical.

[0010] In another aspect, methods described herein include transferring one or more polynucleotides that include at least one *A. thermophilum* coding region to a recipient microorganism. In some embodiments, the method involves direct or indirect cloning of an *A. thermophilum* polynucleotide, then introducing the A. thermophilum polynucleotide into a recipient microorganism. In other embodiments, *A. thermophilum* is co-cultivated with a recipient microorganism, wherein the *A. thermophilum* comprises a conjugative polynucleotide, and wherein the co-cultivation is under conditions suitable for conjugative transfer of at least a portion of the conjugative polynucleotide from the *A. thermophilum* to the recipient microorganism; and identifying a recipient microorganism exconjugant.

[0011] In another aspect, the present invention provides a genetically-modified microorganism comprising one or more *A. thermophilum* plant biomass utilization (PBU) coding regions. In some cases, the PBU coding region comprises a polysaccharide hydrolases and related enzymes (PHR) coding rgion.

[0012] In another aspect, the methods described herein involve using a microorganism for processing plant biomass. Generally, the methods include growing microorganisms comprising one or more *A. thermophilum* plant biomass utilization (PBU) coding regions on a substrate that comprises unprocessed or spent plant biomass under conditions effective for the microorganism to convert at least a portion of the plant biomass to a soluble product.

[0013] In another aspect, the present invention provides an isolated polypeptide, and compositions comprising the isolated polypeptide, in which the isolated polypeptide includes an amino acid sequence that is at least 80% identical to the amino acid sequence of a PBU polypeptide. In some embodiments, the PBU polypeptide comprises a PHR polypeptide.

[0014] In another aspect, the invention provides a method of making an isolated *A. thermophilum* polypeptide. Gener-

ally, the method includes growing a microorganism comprising at least one coding region encoding an *A. thermophilum* polypeptide under conditions effective for the microorganism to produce the *A. thermophilum* polypeptide, and isolating the *A. thermophilum* polypeptide.

[0015] In yet another aspect, the present invention provides a method of processing plant biomass using an isolated *A. thermophilum* polypeptide. Generally, the method includes providing an isolated *A. thermophilum* polypeptide; and contacting the *A. thermophilum* polypeptide with plant biomass under conditions effective for the *A. thermophilum* polypeptide to at least partially degrade the plant biomass.

[0016] The above summary of the present invention is not intended to describe each disclosed embodiment or every implementation of the present invention. The description that follows more particularly exemplifies illustrative embodiments. However, embodiments other than those expressly described are possible and may be made, used, and/or practiced under circumstances and/or conditions that are the same or different from the circumstances and/or conditions described in connection with the illustrative embodiments. In several places throughout the application, guidance is provided through lists of examples, which examples can be used in various combinations. In each instance, the recited list serves only as a representative group and should not be interpreted as an exclusive list.

BRIEF DESCRIPTION OF THE FIGURES

[0017] FIG. **1**. Growth of *A. thermophilum* on unprocessed wood and grass biomass.

[0018] FIG. **2**. Growth of *A. thermophilum* on defined substrates: cellobiose, crystalline cellulose (Avicel), and xylan (oat spelt).

[0019] FIG. **3**. End products of growth of *A. thermophilum* on defined substrates: cellobiose, crystalline cellulose (Avicel) and xylan (oat spelt).

[0020] FIG. **4**. Growth of *A. thermophilum* on unprocessed switchgrass and poplar.

[0021] FIG. **5**. End products of growth of *A. thermophilum* on unprocessed switchgrass or poplar.

[0022] FIG. **6**. Growth of *A. thermophilum* in flushed cultures on defined and undefined substrates (poplar, xylan and cellobiose).

[0023] FIG. **7**. End products of growth of *A. thermophilum* in flushed cultures on defined and undefined substrates (poplar, xylan and cellobiose).

[0024] FIG. **8**. Growth of *A. thermophilum* on 'spent' poplar and switchgrass.

[0025] FIG. **9**. End products of growth of *A. thermophilum* on 'spent' poplar and switchgrass.

[0026] FIG. **10**. Growth of *A. thermophilum* on 'spent' crystalline cellulose (Avicel).

[0027] FIG. **11**. End products of growth of *A. thermophilum* on 'spent' crystalline cellulose (Avicel).

[0028] FIG. **12**. Growth of *A. thermophilum* on a defined medium (on cellobiose) and on untreated switchgrass and poplar in the absence of yeast extract.

[0029] FIG. **13**. Growth of *A. thermophilum* and *C. saccharolyticus* on soluble and insoluble heat-treated (98° C./2 min) extracts of switchgrass.

[0030] FIG. **14**. Growth of *A. thermophilum* and *C. saccharolyticus* on soluble and insoluble heat-treated extracts of poplar.

[0031] FIG. **15**. Growth of *A. thermophilum* and *C. saccharolyticus* on soluble and insoluble heat-treated extracts of pine.

[0032] FIG. **16**. CelA fragment encoding GH9-CBM (GH9 is catalytic domain, CBM is carbohydrate-binding domain).

[0033] FIG. **17**. Signal sequence of *P. furiosus* amylase coding region.

[0034] FIG. **18**. Plasmid pS2-SP used to generate the recombinant *P. furiosus* strain containing *A. thermophilum* CelA.

[0035] FIG. **19**. Plasmid pS2-GH9 used to generate the recombinant *P. furiosus* strain containing *A thermophilum* CelA.

[0036] FIG. **20**. PCR using primers GDHcasUP-HMGcas-DOWN will amplify a 1500 bp fragment diagnostic of PF GDH-HMG cassette.

[0037] FIG. **21**. Confirmation of GH9(CelA) and GH9sp (CelA+signal peptide) exconjugants.

[0038] FIG. **22**. Confirmation of GH9(CelA) and GH9sp (CelA+signal peptide) exconjugants.

[0039] FIG. **23**. Nucleotide and amino acid sequences of selected A. thermophilum plant biomass utilization (PBU) coding regions.

[0040] FIG. **23-01**: Nucleotide sequence (SEQ ID NO:18) and amino acid sequence (SEQ ID NO:19) of Athe_0010.

[0041] FIG. **23-02**: Nucleotide sequence (SEQ ID NO:20) and amino acid sequence (SEQ ID NO:21) of Athe_0011.

[0042] FIG. **23-03**: Nucleotide sequence (SEQ ID NO:22) and amino acid sequence (SEQ ID NO:23) of Athe_0012.

[0043] FIG. **23-04**: Nucleotide sequence (SEQ ID NO:24) and amino acid sequence (SEQ ID NO:25) of Athe_0013.

[0044] FIG. **23-05**: Nucleotide sequence (SEQ ID NO:26) and amino acid sequence (SEQ ID NO:27) of Athe_0014.

[0045] FIG. **23-06**: Nucleotide sequence (SEQ ID NO:28) and amino acid sequence (SEQ ID NO:29) of Athe_0015.

[0046] FIG. **23-07**: Nucleotide sequence (SEQ ID NO:30) and amino acid sequence (SEQ ID NO:31) of Athe_0016.

[0047] FIG. **23-08**: Nucleotide sequence (SEQ ID NO:32) and amino acid sequence (SEQ ID NO:33) of Athe_0017.

[0048] FIG. **23-09**: Nucleotide sequence (SEQ ID NO:34) and amino acid sequence (SEQ ID NO:35) of Athe_0052.

[0049] FIG. **23-10**: Nucleotide sequence (SEQ ID NO:36) and amino acid sequence (SEQ ID NO:37) of Athe_0053.

[0050] FIG. **23-11**: Nucleotide sequence (SEQ ID NO:38) and amino acid sequence (SEQ ID NO:39) of Athe_0054.

[0051] FIG. **23-12**: Nucleotide sequence (SEQ ID NO:40) and amino acid sequence (SEQ ID NO:41) of Athe_0055.

[0052] FIG. **23-13**: Nucleotide sequence (SEQ ID NO:42) and amino acid sequence (SEQ ID NO:43) of Athe_0056.

[0053] FIG. **23-14**: Nucleotide sequence (SEQ ID NO:44) and amino acid sequence (SEQ ID NO:45) of Athe_0057.

[0054] FIG. **23-15**: Nucleotide sequence (SEQ ID NO:46) and amino acid sequence (SEQ ID NO:47) of Athe_0058.

[0055] FIG. **23-16**: Nucleotide sequence (SEQ ID NO:48) and amino acid sequence (SEQ ID NO:49) of Athe_0059.

[0056] FIG. **23-17**: Nucleotide sequence (SEQ ID NO:50) and amino acid sequence (SEQ ID NO:51) of Athe_0060.

[0057] FIG. **23-18**: Nucleotide sequence (SEQ ID NO:52) and amino acid sequence (SEQ ID NO:53) of Athe_0061.

[0058] FIG. **23-19**: Nucleotide sequence (SEQ ID NO:54) and amino acid sequence (SEQ ID NO:55) of Athe_0077.

[0059] FIG. **23-20**: Nucleotide sequence (SEQ ID NO:56) and amino acid sequence (SEQ ID NO:57) of Athe_0088.

[0060] FIG. **23-21**: Nucleotide sequence (SEQ ID NO:58) and amino acid sequence (SEQ ID NO:59) of Athe_0089.
[0061] FIG. **23-22**: Nucleotide sequence (SEQ ID NO:60) and amino acid sequence (SEQ ID NO:61) of Athe_0090.
[0062] FIG. **23-23**: Nucleotide sequence (SEQ ID NO:62) and amino acid sequence (SEQ ID NO:63) of Athe_0153.
[0063] FIG. **23-24**: Nucleotide sequence (SEQ ID NO:64) and amino acid sequence (SEQ ID NO:65) of Athe_0154.
[0064] FIG. **23-25**: Nucleotide sequence (SEQ ID NO:66) and amino acid sequence (SEQ ID NO:67) of Athe_0155.
[0065] FIG. **23-26**: Nucleotide sequence (SEQ ID NO:68) and amino acid sequence (SEQ ID NO:69) of Athe_0156.
[0066] FIG. **23-27**: Nucleotide sequence (SEQ ID NO:70) and amino acid sequence (SEQ ID NO:71) of Athe_0157.
[0067] FIG. **23-28**: Nucleotide sequence (SEQ ID NO:72) and amino acid sequence (SEQ ID NO:73) of Athe_0158.
[0068] FIG. **23-29**: Nucleotide sequence (SEQ ID NO:74) and amino acid sequence (SEQ ID NO:75) of Athe_0159.
[0069] FIG. **23-30**: Nucleotide sequence (SEQ ID NO:76) and amino acid sequence (SEQ ID NO:77) of Athe_0160.
[0070] FIG. **23-31**: Nucleotide sequence (SEQ ID NO:78) and amino acid sequence (SEQ ID NO:79) of Athe_0450.
[0071] FIG. **23-32**: Nucleotide sequence (SEQ ID NO:80) and amino acid sequence (SEQ ID NO:81) of Athe_0451.
[0072] FIG. **23-33**: Nucleotide sequence (SEQ ID NO:82) and amino acid sequence (SEQ ID NO:83) of Athe_0452.
[0073] FIG. **23-34**: Nucleotide sequence (SEQ ID NO:84) and amino acid sequence (SEQ ID NO:85) of Athe_0607.
[0074] FIG. **23-35**: Nucleotide sequence (SEQ ID NO:86) and amino acid sequence (SEQ ID NO:87) of Athe_0608.
[0075] FIG. **23-36**: Nucleotide sequence (SEQ ID NO:88) and amino acid sequence (SEQ ID NO:89) of Athe_1853.
[0076] FIG. **23-37**: Nucleotide sequence (SEQ ID NO:90) and amino acid sequence (SEQ ID NO:91) of Athe_1854.
[0077] FIG. **23-38**: Nucleotide sequence (SEQ ID NO:92) and amino acid sequence (SEQ ID NO:93) of Athe_1855.
[0078] FIG. **23-39**: Nucleotide sequence (SEQ ID NO:94) and amino acid sequence (SEQ ID NO:95) of Athe_1856.
[0079] FIG. **23-40**: Nucleotide sequence (SEQ ID NO:96) and amino acid sequence (SEQ ID NO:97) of Athe_1989.
[0080] FIG. **23-41**: Nucleotide sequence (SEQ ID NO:98) and amino acid sequence (SEQ ID NO:99) of Athe_1990.
[0081] FIG. **23-42**: Nucleotide sequence (SEQ ID NO:100) and amino acid sequence (SEQ ID NO:101) of Athe_1991.
[0082] FIG. **23-43**: Nucleotide sequence (SEQ ID NO:102) and amino acid sequence (SEQ ID NO:103) of Athe_1992.
[0083] FIG. **23-44**: Nucleotide sequence (SEQ ID NO:104) and amino acid sequence (SEQ ID NO:105) of Athe_1993.
[0084] FIG. **23-45**: Nucleotide sequence (SEQ ID NO:106) and amino acid sequence (SEQ ID NO:107) of Athe_1994.
[0085] FIG. **23-46**: Nucleotide sequence (SEQ ID NO:108) and amino acid sequence (SEQ ID NO:109) of Athe_2076.
[0086] FIG. **23-47**: Nucleotide sequence (SEQ ID NO:110) and amino acid sequence (SEQ ID NO:111) of Athe_2077.
[0087] FIG. **23-48**: Nucleotide sequence (SEQ ID NO:112) and amino acid sequence (SEQ ID NO:113) of Athe_2078.
[0088] FIG. **23-49**: Nucleotide sequence (SEQ ID NO:114) and amino acid sequence (SEQ ID NO:115) of Athe_2079.
[0089] FIG. **23-50**: Nucleotide sequence (SEQ ID NO:116) and amino acid sequence (SEQ ID NO:117) of Athe_2080.
[0090] FIG. **23-51**: Nucleotide sequence (SEQ ID NO:118) and amino acid sequence (SEQ ID NO:119) of Athe_2081.
[0091] FIG. **23-52**: Nucleotide sequence (SEQ ID NO:120) and amino acid sequence (SEQ ID NO:121) of Athe_2082.
[0092] FIG. **23-53**: Nucleotide sequence (SEQ ID NO:122) and amino acid sequence (SEQ ID NO:123) of Athe_2083.
[0093] FIG. **23-54**: Nucleotide sequence (SEQ ID NO:124) and amino acid sequence (SEQ ID NO:125) of Athe_2084.
[0094] FIG. **23-55**: Nucleotide sequence (SEQ ID NO:126) and amino acid sequence (SEQ ID NO:127) of Athe_2085.
[0095] FIG. **23-56**: Nucleotide sequence (SEQ ID NO:128) and amino acid sequence (SEQ ID NO:129) of Athe_2086.
[0096] FIG. **23-57**: Nucleotide sequence (SEQ ID NO:130) and amino acid sequence (SEQ ID NO:131) of Athe_2087.
[0097] FIG. **23-58**: Nucleotide sequence (SEQ ID NO:132) and amino acid sequence (SEQ ID NO:133) of Athe_2088.
[0098] FIG. **23-59**: Nucleotide sequence (SEQ ID NO:134) and amino acid sequence (SEQ ID NO:135) of Athe_2089.
[0099] FIG. **23-60**: Nucleotide sequence (SEQ ID NO:136) and amino acid sequence (SEQ ID NO:137) of Athe_2090.
[0100] FIG. **23-61**: Nucleotide sequence (SEQ ID NO:138) and amino acid sequence (SEQ ID NO:139) of Athe_2091.
[0101] FIG. **23-62**: Nucleotide sequence (SEQ ID NO:140) and amino acid sequence (SEQ ID NO:141) of Athe_2092.
[0102] FIG. **23-63**: Nucleotide sequence (SEQ ID NO:142) and amino acid sequence (SEQ ID NO:143) of Athe_2093.
[0103] FIG. **23-64**: Nucleotide sequence (SEQ ID NO:144) and amino acid sequence (SEQ ID NO:145) of Athe_2094.
[0104] FIG. **23-65**: Nucleotide sequence (SEQ ID NO:146) and amino acid sequence (SEQ ID NO:147) of Athe_2371.
[0105] FIG. **23-66**: Nucleotide sequence (SEQ ID NO:148) and amino acid sequence (SEQ ID NO:149) of Athe_2372.
[0106] FIG. **23-67**: Nucleotide sequence (SEQ ID NO:150) and amino acid sequence (SEQ ID NO:151) of Athe_2373.
[0107] FIG. **23-68**: Nucleotide sequence (SEQ ID NO:152) and amino acid sequence (SEQ ID NO:153) of Athe_2374.
[0108] FIG. **23-69**: Nucleotide sequence (SEQ ID NO:154) and amino acid sequence (SEQ ID NO:155) of Athe_2375.
[0109] FIG. **23-70**: Nucleotide sequence (SEQ ID NO:156) and amino acid sequence (SEQ ID NO:157) of Athe_2376.
[0110] FIG. **23-71**: Nucleotide sequence (SEQ ID NO:158) and amino acid sequence (SEQ ID NO:159) of Athe_0423.
[0111] FIG. **23-72**: Nucleotide sequence (SEQ ID NO:160) and amino acid sequence (SEQ ID NO:161) of Athe_0603.
[0112] FIG. **23-73**: Nucleotide sequence (SEQ ID NO:162) and amino acid sequence (SEQ ID NO:163) of Athe_0610.
[0113] FIG. **24**. Growth of *A. thermophilum* on washed and unwashed peanut shells.
[0114] FIG. **25**. Gene clusters encoding multi-domain carbohydrate active enzymes from *A. thermophilum* and C. saccharolyticus.
[0115] FIG. **26**. Construction of Shuttle Vector pDCW 31.
[0116] FIG. **27**. Peptide domains common to *A. thermophilum* DSM6725 and *C. saccharolyticus* DSM8903.
[0117] FIG. **28**. Peptide domains unique to *A. thermophilum* DSM 6725.
[0118] FIG. **29**. Peptide domain re-arrangements in *A. thermophilum* compared to *C. saccharolyticus*.
[0119] FIG. **30**. Peptide domains enriched in *A. thermophilum* DSM6725 and *C. saccharolyticus* DSM8903.
[0120] FIG. **31**. Differential expression of extracellular proteins during growth of *A. thermophilum* DSM 6725 on crystalline cellulose.
[0121] FIG. **32**. Non-catalytic extracellular (ExtP) or membrane-associated (Memb) proteins in *A. thermophilum* DSM 6750.

[0122] FIG. **33**. Exemplary proteins produced by *A. thermophilum* during growth on cellulose, xylan, poplar and/or switchgrass that are not encoded in the *C. saccharolyticus* genome.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0123] The present invention relates to methods, microorganisms, and compositions useful for processing plant biomass. The invention relates, in certain aspects, to a group of coding regions, the expression of which can enable a microorganism to convert plant biomass such as, for example, poplar wood chips, to soluble products that can be used by the same or by another microorganism to produce an economically desirable product such as, for example, a biofuel (e.g., an alcohol and/or hydrogen gas (H_2)), polymer, or commodity chemical.

[0124] The application of this technology has the potential to render production of biofuels more economically feasible and to allow a broader range of microorganisms to utilize recalcitrant biomass. The use of cellulosic materials as sources of bioenergy is currently limited by typically requiring preprocessing of the cellulosic material. Such preprocessing methods can be expensive. Thus, methods that reduce dependence on preprocessing of cellulosic materials may have a dramatic impact on the economics of the use of recalcitrant biomass for biofuels production.

[0125] One challenge in converting biomass into liquid (e.g., ethanol, biodiesel) and gaseous (e.g., H_2) fuels is the recalcitrance and heterogeneity of the biological material. Consequently, effective and efficient conversion of the biological material cannot be achieved by a single naturally-occurring microorganism, a mixture of naturally-occurring microorganisms, or a mixture of enzymes. In certain aspects, the present invention involves exploiting a specific group of coding regions, the so-called plant biomass utilization (PBU) gene set of *Anaerocellum thermophilum*. Expression of one or more of these coding regions can enable processed, unprocessed, and/or spent samples of plant biomass to be utilized directly for biomass conversion. These coding regions can be expressed by various microorganisms by the appropriate genetic manipulations. The microorganisms may be thermophilic microorganisms such as, for example, *A. thermophilum* or may be mesophilic microorganisms. Moreover, the products of biomass conversion are not limited to biofuels, but extend to any polymer or commodity chemical derived from plant cell biomass.

[0126] In the description that follows, the following terms shall have the meanings set forth below.

[0127] "Biofuel" refers to a combustible material that can be produced through chemical, enzymatic, or microbiotic fermentation or processing of plant biomass (e.g., processed biomass, unprocessed biomass, spent biomass, etc.) and that can be used, alone or in combination with other materials, for the generation of energy.

[0128] "Commodity chemical" refers to any product (e.g., oxalic acid, succinic acid, lactic acid, pyruvic acid, salts thereof, amino acids, etc.) from the fermentation of plant biomass (e.g., processed biomass, unprocessed biomass, spent biomass, etc.) that can be the starting material for the production of other chemicals and/or materials.

[0129] "Extremophilic" refers to a microorganism that can thrive in, and may require, specific conditions that are unfavorable to other microorganisms.

[0130] "Exconjugant" refers to a cell that, after conjugation, has received DNA from a conjugation partner cell.

[0131] "Mesophilic" refers to a microorganism that has a temperature optimum for growth of from 20-37° C.

[0132] "Processed plant biomass" refers to plant biomass that has been subjected to chemical, physical, microbial, or enzymatic processing under conditions such that at least some of the complex organic polymers originally present in the plant biomass are degraded to smaller chemical subunits.

[0133] "Spent biomass" refers to water insoluble material that remains after a microbial culture is permitted to grow on plant biomass to late stationary phase. As one example, spent biomass can refer to water insoluble material remaining after a culture of *A. thermophilum* is permitted to grow to approximately 10^8 cells/mL on plant biomass.

[0134] "Thermophilic" refers to a microorganism that has a temperature optimum for growth of from 50° C.-100° C. "Extremely thermophilic" refers to a microorganism that has a temperature optimum for growth of from 70° C.-100° C.

[0135] "Untreated plant biomass" refers to plant biomass that contains complex organic polymer such as, for example, lignin or a complex polysaccharide or heteropolysaccharide (e.g., cellulose, a hemicellulose such as xylan, pectin, etc.) that has not been subjected to chemical, physical, microbial, or enzymatic processing to degrade the biomass—i.e., degrade the complex organic polymer to smaller chemical subunits.

[0136] The term "and/or" means one or all of the listed elements or a combination of any two or more of the listed elements.

[0137] The terms "comprises" and variations thereof do not have a limiting meaning where these terms appear in the description and claims.

[0138] Unless otherwise specified, "a," "an," "the," "one or more," and "at least one" are used interchangeably and mean one or more than one.

[0139] Also herein, the recitations of numerical ranges by endpoints include all numbers subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, 5, etc.). Unless otherwise indicated, all numbers expressing quantities of components, molecular weights, and so forth used in the specification and claims may be modified in each instance by the term "about." Accordingly, unless otherwise indicated to the contrary, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties sought to be obtained by the present invention. At the very least, and not as an attempt to limit the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques.

[0140] Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. All numerical values, however, inherently contain a range necessarily resulting from the standard deviation found in their respective testing measurements.

[0141] For any method disclosed herein that includes discrete steps, the steps may be conducted in any feasible order. And, as appropriate, any combination of two or more steps may be conducted simultaneously.

[0142] It has been found that *A. thermophilum* can grow efficiently on various types of untreated biomass (e.g., poplar

woodchips, various types of grasses, and on the insoluble extracts of such biomass) (FIGS. **1-7**). As used herein "efficient" growth refers to growth in which cells may be cultivated to a specified density within a specified time. For example, *A. thermophilum* can grow to a density of at least 5×10^7 cells/milliliter (mL) such as, for example, a density of 10^8 cells/mL. Methods for determining cell density of a culture are routine and known to those skilled in the art. Efficient growth of *A. thermophilum* on a substrate can be determined by measuring the cell density of the culture at a time no greater than 60 hours after the culture medium is inoculated. For example, efficient growth of *A. thermophilum* can be determined by measuring the cell density of the culture no greater than 30 hours, no greater than 24 hours, no greater than 16 hours, no greater than 12 hours, or no greater than 8 hours after inoculation of the culture.

[0143] *A. thermophilum* can grow efficiently on crystalline cellulose and, in contrast to original reports (Svetlichnyi, V. A., T. P. Svetlichnaya, N. A. Chernykh, and G. A. Zavarzin. 1990. *Anaerocellum thermophilum* gen. nov., sp. nov., an extremely thermophilic cellulolytic eubacterium isolated from hot-springs in the valley of Geysers. Microbiology 59:598-604), can grow efficiently on xylan (oat spelt) (e.g., FIGS. **2** and **6**). The main products when grown on untreated biomass substrates were lactate, acetate, and hydrogen gas (FIGS. **3** and **6**). Moreover, the primary product is influenced at least somewhat by the biomass substrate. For example, FIG. **3** shows that when *A. thermophilum* is grown on a substrate of cellobiose, lactate is favored as a product over acetate and H_2. In contrast, FIG. **9** shows that when *A. thermophilum* is grown on a substrate of switchgrass, acetate and H_2 are favored products over lactate.

[0144] *A. thermophilum* also can grow efficiently on spent biomass—insoluble material that remains after a culture has grown to late stationary phase (e.g., greater than 10^8 cells/mL) on untreated biomass (FIGS. **8** and **10**). *A. thermophilum* also grew efficiently on cellobiose, untreated switchgrass, and untreated poplar (FIG. **12**). *A. thermophilum* also grew on switchgrass and poplar that had been heated at 98° C. for two minutes. As shown in FIG. **13** and FIG. **14**, *A. thermophilum* grew efficiently (greater than 10^8 cells/ml) on both the soluble and insoluble materials obtained after heat treating the biomass. The microorganism also grew efficiently on the insoluble material obtained from pine wood after a similar heat treatment (FIG. **15**). *A. thermophilum* also grew efficiently on peanut shells regardless of whether the peanut shells were first washed for 18 hours at 75° C. (FIG. **24**).

[0145] Thus, in one aspect, the present invention provides methods of processing biomass—particularly but not exclusively water insoluble untreated plant biomass and/or water insoluble spent biomass. Generally, the methods include growing *A. thermophilum* on a substrate that includes plant biomass under conditions effective for the *A. thermophilum* to convert at least a portion of the plant biomass to a less complex water soluble product such as, for example, organic compounds (e.g., organic acids and/or simple carbohydrates such as, for example, monosaccharides and disaccharides) that are readily metabolizable by *A. thermophilum* and/or another microorganism. In some embodiments, the method can further include converting at least a portion of the water soluble product to a biofuel, a polymer, or a commodity chemical. In other cases, the water soluble product may itself be a biofuel, a polymer, and/or a commodity chemical. In other cases, the product of processing the biomass may be a water insoluble product that may itself be a biofuel. In particular embodiments, the methods include growing *A. thermophilum* on a substrate that includes plant biomass under conditions effective for the *A. thermophilum* to degrade cellulose present in the plant biomass.

[0146] The plant biomass can be any plant biomass that is degradable by *A. thermophilum*—i.e., any plant biomass in which *A. thermophilum* is capable of breaking down a complex organic polymer (e.g., lignin or a complex polysaccharide or heteropolysaccharide) component of the biomass to smaller, constituent subunits. In some embodiments, the plant biomass can include plant biomass not utilizable by *Caldicellulosiruptor saccharolyticus* such as, for example, *C. saccharolyticus* (DSM 8903). As used herein, plant biomass that is not utilizable by *C. saccharolyticus* refers to biomass on which *C. saccharolyticus* does not grow efficiently (e.g., soluble and/or insoluble heat-treated poplar, FIG. **14**).

[0147] The plant biomass can include lignocellulosic material. Lignocellulosic material may be found, for example, in the stems, leaves, hulls, husks, and/or cobs of plants or leaves, branches, and wood of trees. Lignocellulosic material can also be, for example, herbaceous material, agricultural residues, forestry residues, municipal solid wastes, waste paper, and pulp and paper mill residues. In some cases, lignocellulosic material may be in the form of plant cell wall material containing lignin, cellulose, and hemicellulose in a mixed matrix. In some aspects the lignocellulosic material may include grass such as switchgrass, Bermudagrass, napiergrass; paper and/or pulp processing waste; corn waste such as corn stover and/or corn fiber; hardwood such as poplar and/or birch; softwood such as Douglas fir, pine (e.g., *Pinus taeda*) and/or spruce; cereal straw such as wheat straw and/or rice straw; municipal solid waste; industrial organic waste; sugarcane and/or bagasse; sugarbeets and/or pulp; sweet potatoes; food processing wastes; or any mixtures thereof.

[0148] Thus, in some embodiments, the plant biomass can include woody plant biomass such as, for example, treated and/or untreated wood, woodchips, sawdust, etc. The woody plant biomass may be, or be derived from, any species of woody plant. In some embodiments, the woody plant biomass may be derived from poplar (i.e., *Populus* spp.) or pine (i.e., *Pinus* spp.), but the methods may be practiced using woody plant biomass derived from other species of woody plants.

[0149] In other embodiments, the plant biomass may be, or be derived from, treated or untreated sources such as, for example, grasses, peanut shells (washed or unwashed), crystalline cellulose, cellobiose, or xylan.

[0150] In some embodiments, the plant biomass may include spent biomass. Thus, the methods offer the possibility of extracting compounds and/or energy from plant biomass that is commonly left unexploited.

[0151] In some embodiments, the plant biomass can include a combination of plant biomass from various sources (e.g., hardwood, softwood, grass, straw, pulp, etc.). Thus, a combination of plant biomass can include, for example, poplar and pine woodchips. Alternatively, in some embodiments, a combination of plant biomass can include, for example, plant biomass that excludes, for example, softwood sawdust (e.g., pine sawdust). As one example, such a combination of plant biomass can include grass (e.g., switchgrass, Bermudagrass, and/or napiergrass), straw (e.g., wheat straw and/or rice straw), and/or corn stover.

[0152] Also, the plant biomass can include a combination of treated, untreated, and spent biomass, with the nature (i.e.,

treated, untreated, or spent) of biomass from each source being independent of the nature of biomass from other sources in the combination.

[0153] The methods of processing biomass can include growing *A. thermophilum* on a substrate that includes plant biomass under conditions effective for the *A. thermophilum* to convert at least a portion of the plant biomass to a less complex—e.g., water soluble—product. Such conditions include conditions under which *A. thermophilum* may be grown in culture. Because *A. thermophilum* is a thermophilic microbe, in some embodiments, the conditions include a temperature of at least 70° C. such as, for example, at least 75° C., at least 80° C., at least 85° C., or at least 90° C. However, the methods described herein may be practiced at lower temperatures including, for example, a temperature of at least 37° C. or at least 30° C. Also, the growing conditions may be anaerobic. As used herein, "anaerobic" conditions refer to conditions in which the partial pressure of O_2 in the gas phase is less than 10 ppm, such as, for example, 1 ppm.

[0154] In another aspect, the invention provides a method of pretreating plant biomass. Generally, the method includes growing *Anaerocellum thermophilum* on a substrate that comprises plant biomass under conditions effective for the *A. thermophilum* to degrade cellulose of the plant biomass, thereby preparing the plant biomass for further processing by another biomass processing method. Pretreating plant biomass using *A. thermophilum* can reduce the need for chemical and/or heat pretreatments in order to make most efficient use of the plant biomass. Thus, in this aspect, the method can reduce, for example, the time, cost, and environmental impact of processing plant biomass and can increase, for example, the efficiency at which the plant biomass is processed.

[0155] In some aspects, described in more detail below, the invention can involve one or more coding regions that can encode polypeptides involved in the degradation of plant biomass and/or the synthesis of certain metabolic products (e.g., biofuels, commodity chemicals, and/or intermediates for the production of either biofuels or commodity chemicals). As used herein, "coding region" refers to a nucleotide sequence that encodes a polypeptide and, when placed under the control of appropriate regulatory sequences expresses the encoded polypeptide. The boundaries of a coding region are generally determined by a translation start codon at its 5' end and a translation stop codon at its 3' end. A "regulatory sequence" is a nucleotide sequence that regulates expression of a coding sequence to which it is operably linked. Regulatory sequences include, for example, promoters, enhancers, transcription initiation sites, translation start sites, translation stop sites, and transcription terminators. The term "operably linked" refers to a juxtaposition of components such that they are in a relationship permitting them to function in their intended manner. A regulatory sequence is "operably linked" to a coding region when it is joined in such a way that expression of the coding region is achieved under conditions compatible with the regulatory sequence.

[0156] In some embodiments, the coding region can include a nucleotide sequence having at least 80% identity to a reference nucleotide sequence such as, for example, an *A. thermophilum* PBU coding region, an *A. thermophilum* PHR coding region, or any other identified coding region (each of which is described herein below). Nucleotide sequences of *A. thermophilum* coding regions such as, for example, PBU coding regions and PHR coding regions, are accessible via GenBank Accession No. CP001395 (version 1, created Feb. 5, 2009). In certain embodiments, a coding region can have at least 85% identity to the nucleotide sequence of a reference coding region such as for example, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identity to the nucleotide sequence of a reference coding region. Such nucleotide sequences may include one or more modifications relative to the nucleotide sequence of the reference coding region. As used herein, two nucleotide sequences may be compared and the nucleotide identity is resulting from that comparison may be referred to as "identities." Two nucleotide sequences may be compared using the Blastn program of the BLAST 2 search algorithm, as described by Tatusova, et al. (FEMS Microbiol Lett, 174, 247250 (1999)), and available through the World Wide Web, for instance at the internet site maintained by the National Center for Biotechnology Information, National Institutes of Health. Preferably, the default values for all BLAST 2 search parameters are used, including reward for match=1, penalty for mismatch=−2, open gap penalty=5, extension gap penalty=2, gap x dropoff=50, expect=10, wordsize=11, and optionally, filter on.

[0157] In other aspects, the invention can involve the expression of an *A. thermophilum* polypeptide or a biologically active analog, subunit, or derivative thereof. An *A. thermophilum* polypeptide or a biologically active analog, subunit, or derivative thereof encoded by a PBU coding region may be referred to as a PBU polypeptide. Similarly, an *A. thermophilum* polypeptide or a biologically active analog, subunit, or derivative thereof encoded by a PHR coding region may be referred to as a PHR polypeptide.

[0158] In some embodiments, the *A. thermophilum* polypeptide may be isolated. As used herein, an "isolated" polypeptide is one that is separated from its natural environment to any degree. An isolated polypeptide may be, for example, at least 60% free, at least 75% free, at least 90% free, at least 91% free, at least 92% free, at least 93% free, at least 94% free, at least 95% free, at least 96%, at least 97% free, at least 98% free, or at least 99% free from other components with which it is naturally associated. Polypeptides that are produced outside the microorganism in which they naturally occur, e.g., through chemical or recombinant means, are considered to be isolated and purified by definition, since they were never present in a natural environment.

[0159] A "biologically active" analog, subunit, or derivative of an *A. thermophilum* polypeptide is a polypeptide that exhibits the ability to degrade water insoluble plant biomass material. A biologically active "analog" of an *A. thermophilum* polypeptide includes, for example, an *A. thermophilum* polypeptide that has been modified by the addition, substitution, or deletion of one or more contiguous or noncontiguous amino acids, or that has been chemically or enzymatically modified, e.g., by attachment of a reporter group, by an N-terminal, C-terminal or other functional group modification or derivatization, or by cyclization, as long as the analog retains biological activity. An analog can thus include additional amino acids at one or both of the termini of a polypeptide.

[0160] Substitutes for an amino acid in an *A. thermophilum* polypeptide are preferably conservative substitutions, which are selected from other members of the class to which the amino acid belongs. For example, it is well-known in the art of protein biochemistry that an amino acid belonging to a grouping of amino acids having a particular size or characteristic (such as charge, hydrophobicity and hydrophilicity)

can generally be substituted for another amino acid without substantially altering the structure of a polypeptide. For the purposes of this invention, conservative amino acid substitutions are defined to result from exchange of amino acids residues from within one of the following classes of residues: Class I: Ala, Gly, Ser, Thr, and Pro (representing small aliphatic side chains and hydroxyl group side chains); Class H: Cys, Ser, Thr and Tyr (representing side chains including an —OH or —SH group); Class III: Glu, Asp, Asn and Gln (carboxyl group containing side chains): Class IV: His, Arg and Lys (representing basic side chains); Class V: Ile, Val, Leu, Phe and Met (representing hydrophobic side chains); and Class VI: Phe, Trp, Tyr and His (representing aromatic side chains). The classes also include related amino acids such as 3Hyp and 4Hyp in Class I; homocysteine in Class II; 2-aminoadipic acid, 2-aminopimelic acid, γ-carboxyglutamic acid, β-carboxyaspartic acid, and the corresponding amino acid amides in Class III; ornithine, homoarginine, N-methyl lysine, dimethyl lysine, trimethyl lysine, 2,3-diaminopropionic acid, 2,4-diaminobutyric acid, homoarginine, sarcosine and hydroxylysine in Class IV; substituted phenylalanines, norleucine, norvaline, 2-aminooctanoic acid, 2-aminoheptanoic acid, statine and β-valine in Class V; and naphthylalanines, substituted phenylalanines, tetrahydroisoquinoline-3-carboxylic acid, and halogenated tyrosines in Class VI.

[0161] The amino acid sequences of exemplary *A. thermophilum* polypeptides are accessible via GenBank Accession No. CP001395 (version 1, created Feb. 5, 2009). Certain biologically active analogs, subunits, or derivatives of a reference *A. thermophilum* polypeptide can include those analogs, subunits, or derivatives that have at least 80% identity to the reference *A. thermophilum* polypeptide. In some embodiments, the biologically active analog, subunit, or derivative can have at least 85% identity to a reference *A. thermophilum* polypeptide such as, for example, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to a reference *A. thermophilum* polypeptide. Such analogs, subunits, or derivatives can contain one or more amino acid deletions, insertions, and/or substitutions relative to the reference A. thermophilum polypeptide, and may further include chemical and/or enzymatic modifications and/or derivatizations, as described above.

[0162] The degree of identity between two amino acid sequences can be determined using commercially available algorithms. Preferably, two amino acid sequences are compared using the BLASTP program of the BLAST 2 search algorithm, as described by Tatusova, et al., (*FEMS Microbiol Lett* 1999, 174:247-250), and available through the World Wide Web, for instance at the internet site maintained by the National Center for Biotechnology Information, National Institutes of Health. Preferably, the default values for all BLAST 2 search parameters are used, including matrix=BLOSUM62; open gap penalty=11, extension gap penalty=1, gap x_dropoff=50, expect=10, wordsize=3, and optionally, filter on.

[0163] Thus, modification of a nucleotide sequence encoding an *A. thermophilum* polypeptide may provide the synthesis of a polypeptide that is substantially similar to the *A. thermophilum* polypeptide. The term "substantially similar" to the *A. thermophilum* polypeptide refers to a non-naturally occurring form of the *A. thermophilum* polypeptide. Such a polypeptide may differ in some engineered way from the *A. thermophilum* polypeptide isolated from a native source—e.g., the variant may differ in specific activity, thermostability, pH optimum, or the like. The variant sequence may be constructed on the basis of the nucleotide sequence presented as the polypeptide encoding region of any one of the nucleotide sequences depicted in FIG. **23**, a subsequence thereof, and/or by introduction of nucleotide substitutions which do not give rise to another amino acid sequence of the *A. thermophilum* polypeptide encoded by the nucleotide sequence, but which correspond to the codon usage of the recipient microorganism, or by introduction of nucleotide substitutions which may give rise to a different amino acid sequence. For a general description of nucleotide substitution, see, e.g., Ford et al., 1991, Protein Expression and Purification 2: 95-107.

[0164] In some embodiments, a *A. thermophilum* polynucleotide can include the nucleotide sequence of one or more PHR coding regions such as, for example, Athe_0423 (or2161) (SEQ ID NO:158), Athe_0603 (or1720) (SEQ ID NO:160), or Athe_0610 (or1727) (SEQ ID NO:162). As used herein, the Athe_#### coding region designations refer to the locus tag associated with the identified coding region, as provided in GenBank Accession No. CP001393, version 1 for the *A. thermophilum* chromosome, CP001394, version 1 for pATHE01, and CP001395 for pATHE02 (SEQ ID NO:1). The or#### designations refer to the coding region identifiers used in the draft *A. thermophilum* sequence. Table 1 correlates both designations. Consequently, the *A. thermophilum* polynucleotide can encode a PHR polypeptide—including, as defined herein, a biologically active analog, subunit, or derivative—such as, for example, a PHR polypeptide that includes the amino acid sequence of one or more of: Athe_0423 (or2161) (SEQ ID NO:159), Athe_0603 (or1720) (SEQ ID NO:161), or Athe_0610 (or1727) (SEQ ID NO:163).

[0165] As described in more detail below, many of the coding regions, including PHR coding regions, that confer the ability of *A. thermophilum* to grow efficiently on plant biomass that cannot be utilized by *C. saccharolyticus* are present as gene clusters (106 clusters, defined as two or more adjacent coding regions, most of which are likely to be present as operons). Consequently, in certain embodiments, an *A. thermophilum* polynucleotide can include one or more coding regions from one or more of gene clusters such as, for example, SYb004 (e.g., one or more of Athe_0052-Athe_0061 (or1895-or1905), SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44, SEQ ID NO:46, SEQ ID NO:48, SEQ ID NO:50, and SEQ ID NO:52), SYb007 (e.g., one or more of Athe_0088-Athe_0090 (or2788-or2790), SEQ ID NO:56, SEQ ID NO:58, and SEQ ID NO:60), SYb012 (e.g., one or more of Athe_0153-Athe_0160 (or1387-or1394), SEQ ID NO:62, SEQ ID NO:64, SEQ ID NO:66, SEQ ID NO:68, SEQ ID NO:70, SEQ ID NO:72, SEQ ID NO:74, and SEQ ID NO:76), SYb032 (e.g., one or more of Athe_0450-Athe_0452 (or2132-or2130), SEQ ID NO:78, SEQ ID NO:80, and SEQ ID NO:82), SYb059 (e.g., one or more of Athe_1853-Athe_1856 (or2888-or2885, and or2910), SEQ ID NO:88, SEQ ID NO:90, SEQ ID NO:92, and SEQ ID NO:94), SYb063 (e.g., one or more of Athe1989-Athe_1994 (or1187-or1182), SEQ ID NO:96, SEQ ID NO:98, SEQ ID NO:100, SEQ ID NO:102, SEQ ID NO:104, and SEQ ID NO:106), SYb067 (e.g., one or more of Athe_2076-Athe_2094 (or1093-or1071), SEQ ID NO:108, SEQ ID NO:110, SEQ ID NO:112, SEQ ID NO:114, SEQ ID NO:116, SEQ ID

NO:118, SEQ ID NO:120, SEQ ID NO:122, SEQ ID NO:124, SEQ ID NO:126, SEQ ID NO:128, SEQ ID NO:130, SEQ ID NO:132, SEQ ID NO:134, SEQ ID NO:136, SEQ ID NO:138, SEQ ID NO:140, SEQ ID NO:142, and SEQ ID NO:144), and SYb082 (e.g., one or more of Athe_2371-Athe_2376 (or1921-or1926), SEQ ID NO:146, SEQ ID NO:148, SEQ ID NO:150, SEQ ID NO:152, SEQ ID NO:154, and SEQ ID NO:156). Thus, the *A. thermophilum* polynucleotide can encode a PHR polypeptide-including, as defined herein, a biologically active analog, subunit, or derivative-such as, for example, a PHR polypeptide that includes the amino acid sequence of one or more of: SEQ ID NO:35, SEQ ID NO:37, SEQ ID NO:39, SEQ ID NO:41, SEQ ID NO:43, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, SEQ ID NO:57, SEQ ID NO:59, SEQ ID NO:61, SEQ ID NO:63, SEQ ID NO:65, SEQ ID NO:67, SEQ ID NO:69, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:75, SEQ ID NO:77, SEQ ID NO:79, SEQ ID NO:81, SEQ ID NO:83, SEQ ID NO:89, SEQ ID NO:91, SEQ ID NO:93, SEQ ID NO:95, SEQ ID NO:97, SEQ ID NO:99, SEQ ID NO:101, SEQ ID NO:103, SEQ ID NO:105, SEQ ID NO:107, SEQ ID NO:109, SEQ ID NO:111, SEQ ID NO:113, SEQ ID NO:115, SEQ ID NO:117, SEQ ID NO:119, SEQ ID NO:121, SEQ ID NO:123, SEQ ID NO:125, SEQ ID NO:127, SEQ ID NO:129, SEQ ID NO:131, SEQ ID NO:133, SEQ ID NO:135, SEQ ID NO:137, SEQ ID NO:139, SEQ ID NO:141, SEQ ID NO:143, and SEQ ID NO:145, SEQ ID NO:147, SEQ ID NO:149, SEQ ID NO:151, SEQ ID NO:153, SEQ ID NO:155, and SEQ ID NO:157.

[0166] In some embodiments, an *A. thermophilum* polynucleotide can include the nucleotide sequence of one or more of the remaining PBU coding regions such as, for example, Athe_0077 (or2776), SEQ ID NO:54). Consequently, the *A. thermophilum* polynucleotide can encode a PBU polypeptide-including, as defined herein, a biologically active analog, subunit, or derivative-such as, for example, a PBU polypeptide that includes the amino acid sequence of SEQ ID NO:55.

[0167] Here again, many of the remaining PBU coding regions are present as gene clusters. Consequently, in certain embodiments, an *A. thermophilum* polynucleotide can include one or more coding regions from one or more of gene clusters such as, for example, SYb001 (e.g., one or more of Athe_0010-Athe_0017 (or1851-or1859), SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:32) and SYb037 (e.g., one or more of Athe_0607-Athe_0608 (ori1724-or1724), SEQ ID NO:84 and SEQ ID NO:86). Thus, an *A. thermophilum* polynucleotide can encode a PBU polypeptide—including, as defined herein, a biologically active analog, subunit, or derivative—such as, for example, a PBU polypeptide that includes the amino acid sequence of one or more of SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:85, and SEQ ID NO:87.

[0168] Some methods described herein exploit the PBU coding regions of *A. thermophilum* to convert plant biomass into water soluble or water insoluble product. A water soluble product may have value in itself, or as a starting material from which some other material may be prepared in one or more subsequent processes. For example, in some embodiments, the water soluble product can include an alcohol such as, for example, ethanol, n-butanol, 1,4-butanediol, sec-butanol, and/or methanol. In other embodiments, the water soluble product can include, for example, hydrogen gas (H_2). In still other embodiments, the water soluble product can include one or more small organic (e.g., C1-C8) acids such as, for example, succinic acid, lactic acid, citric acid, oxaloacetic acid, malic acid, adipic acid, fumaric acid, pyruvic acid, or a salt thereof). In still other embodiments, the water soluble product can include simple saccharides such as, for example, monosaccharides and/or disaccharides. Small organic acids and/or simple saccharides can serve as metabolic intermediates for the production of other organic compounds such as, for example, alcohols, fatty acids, and polymers. Ethanol, methanol, a butanol, and/or hydrogen gas may be used as biofuels. Ethanol, methanol, a butanol, or an organic acid or a salt thereof may be used as a commodity chemical. In still other embodiments, the water soluble product can include a water soluble polymer material such as, for example, a soluble lipid such as, for example, a fatty acid or a polyisoprenoid. In other embodiments, the product may be water insoluble, such as, for example, the production of a biodiesel (alkyl fatty acid esters), which may be used as a biofuel.

[0169] In some embodiments, the product, whether water soluble or water insoluble, may be released by the *A. thermophilum* into the culture medium, from which the product may be isolated, purified, or otherwise recovered using a method or process appropriate for the product. In this context, "isolated" refers to increasing the proportion (e.g., concentration, w/v%, etc.) of the product to any degree regardless of the way in which the product is isolated. Thus, in some cases, a product may be isolated by, for example, removing at least a portion of the product from the culture medium. In other cases, a product may be isolated by, for example, removing one or more components (e.g., cells, spent biomass, medium components, etc.) of the culture medium, leaving behind an increased proportion of the product compared to the sum of non-product constituents of the culture medium. In other embodiments, the product, whether water soluble or water insoluble, may be sequestered within the *A. thermophilum*. In such cases, the methods described herein can further include solubilizing the *A. thermophilum* before the product may be recovered. As used herein, the term "solubilizing" refers to dissolving cellular materials (e.g., polypeptides, nucleic acids, carbohydrates) into the aqueous phase of a buffer in which the microbe was disrupted, and the formation of aggregates of insoluble cellular materials. Methods for solubilizing cells are routine and known to those skilled in the art.

[0170] The chromosomal genome of *A. thermophilum* is 2.97 Mb in size and is predicted to contain 2,824 genes, of which 2,654 are predicted to be protein coding regions. The *A. thermophilum* genome further includes two native plasmids: pATHE01 (approximately 8.3 Kb in size and containing eight coding regions) and pATHE02 (approximately 3.7 Kb in size and containing four coding regions, SEQ ID NO:1). A preliminary bioinfoiniatics analysis of the *A. thermophilum* DSM 6725 coding regions revealed that the closest homologs for 2,284 coding regions in the *A. thermophilum* genome are found in the genome of *Caldicellulosiruptor saccharolyticus* (DSM 8903). *C. saccharolyticus* was discovered in 1994 and, like *A. thermophilum*, is a strict anaerobe that grows optimally near 75° C. Its genome sequence was reported in 2007 and contains 2,679 coding regions (2.97 Mb). *C. saccharolyticus* and *A. thermophilum* appear to be close relatives and

may be members of the same bacterial genus. Indeed, it has been proposed that *A. thermophilum* DSM 6725 be reclassified as *Caldicellulosiruptor bescii*. Thus, as used herein, the term *A. thermophulim* DSM 6725 refers to the bacterial strain deposited Aug. 12, 2009 with the American Type Culture Collection (ATCC), Manassas, Va., regardless of whether the microorganism is classified as *A. thermophilum* or *C. bescii*. The deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. The deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

[0171] Despite the apparent relatedness of *A. thermophulim* DSM 6725 and *C. saccharolyticus*, only one of the species, *A. thermophilum*, is able to grow efficiently on certain forms of plant biomass. The coding regions that confer this property to *A. thermophilum* DSM 6725 are termed PBU for plant biomass utilization. Certain *A. thermophilum* DSM 6725 coding regions that are not specific to *A. thermophilum* may, in conjunction with one or more PBU coding regions, also be involved in plant biomass utilization. Many of the PBU coding regions are present in *A. thermophilum* DSM 6725 as gene clusters.

[0172] Biomass utilization in *C. saccharolyticus* has been partially characterized and *C. saccharolyticus* may grow on a variety of polysaccharides, including crystalline cellulose and xylan. However, growth on untreated biomass has not been reported. *C. saccharolyticus* can grow on soluble and insoluble heat-treated switchgrass (i.e., after heat treatment; FIG. **13**). However, in contrast to *A. thermophilum*, *C. saccharolyticus* cannot utilize either the soluble or insoluble material derived from poplar (FIG. **14**), and it grows much less efficiently than *A. thermophilum* on insoluble material derived from heat-treated pine (FIG. **15**). *A. thermophilum* has also been shown to grow efficiently on both washed and unwashed peanut shells (FIG. **24**).

[0173] The ability of *A. thermophilum* to grow efficiently on untreated and treated biomass that cannot be utilized by *C. saccharolyticus* is a consequence, at least in part, of coding regions present in *A. thermophilum* that lack homologs in *C. saccharolyticus*.

[0174] Table 1 lists a total of 550 such coding regions. Many of these coding regions are present as gene clusters (106 clusters, defined as adjacent coding regions, most of which are likely to be present as operons). The 106 gene clusters are labeled SYa001-SYa106 and contain 436 coding regions. The remaining 114 coding regions that lack close homologs in *C. saccharolyticus* that are not part of gene clusters SYa001-SYa106 are labeled FPa001-FPa114. More than 30 of the clusters contain five or more coding regions, with one cluster containing 19 coding regions (SYa067; Table 2). The 550 coding regions also include nine coding regions encoding transposases. These are similar to those found in both Gram negative bacteria and other Gram positive bacteria, suggesting that at least some of the gene clusters were acquired by *A. thermophilum* through lateral gene transfer. Of the 550 coding regions found in *A. thermophilum* DSM 6725 that are not found in *C. saccharolyticus,* 332 of them are annotated as conserved/hypothetical/unknown function proteins, leaving 218 coding regions with a proposed function. These include 21 DNA binding proteins (11 putative transcriptional regulators/10 containing helix-turn-helix motifs) indicating that many of these coding regions may respond to and regulate carbon source utilization for growth on substrates such as plant biomass.

TABLE 1

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
FPb001	Athe_0002	or1843
FPb002	Athe_0007	or1848
SYb001	Athe_0010	or1851
SYb001	Athe_0011	or1852
SYb001	Athe_0012	or1853, or1854
SYb001	Athe_0013	or1855
SYb001	Athe_0014	or1856
SYb001	Athe_0015	or1857
SYb001	Athe_0016	or1858
SYb001	Athe_0017	or1859
FPb003	Athe_0020	or1862
SYb002	Athe_0022	or1865
SYb002	Athe_0023	or1866
SYb002	Athe_0024	or1867
SYb002	Athe_0025	or1868
SYb003	Athe_0028	or1870
SYb003	Athe_0029	or1871
FPb004	Athe_0035	or1877
SYb004	Athe_0052	or1895
SYb004	Athe_0053	or1896
SYb004	Athe_0054	or1897
SYb004	Athe_0055	or1898
SYb004	Athe_0056	or1899
SYb004	Athe_0057	or1900
SYb004	Athe_0058	or1901
SYb004	Athe_0059	or1902, or1903
SYb004	Athe_0060	or1904, or1903
SYb004	Athe_0061	or1905
SYb005	Athe_0066	or1910
SYb005	Athe_0067	or1911, or1912
SYb005	Athe_0068	
SYb005	Athe_0069	or1914
SYb005	Athe_0070	
SYb006	Athe_0072	or2770
SYb006	Athe_0073	or2771
SYb006	Athe_0074	or2772
FPb005	Athe_0077	or2776
SYb007	Athe_0088	or2788
SYb007	Athe_0089	or2789
SYb007	Athe_0090	or2790
FPb006	Athe_0092	
SYb008	Athe_0109	or2529
SYb008	Athe_0110	or2530
SYb008	Athe_0111	or2531
SYb009	Athe_0130	or2555
SYb009	Athe_0131	or1363
SYb010	Athe_0135	or1368
SYb010	Athe_0136	or1369
SYb011	Athe_0139	or1372
SYb011	Athe_0140	
FPb007	Athe_0142	or1376, or1374, or1375
SYb012	Athe_0153	or1387
SYb012	Athe_0154	or1388
SYb012	Athe_0155	or1389
SYb012	Athe_0156	or1390
SYb012	Athe_0157	or1391
SYb012	Athe_0158	or1392
SYb012	Athe_0159	or1393
SYb012	Athe_0160	or1394
FPb008	Athe_0188	or1208, or1423
FPb009	Athe_0201	or1436
SYb013	Athe_0204	or1440
SYb013	Athe_0205	or1441
FPb010	Athe_0224	or1460
FPb011	Athe_0229	or1465
SYb014	Athe_0235	or1471

TABLE 1-continued

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
SYb014	Athe_0236	or1472
SYb014	Athe_0237	or1473
FPb012	Athe_0241	
SYb015	Athe_0247	or1482
SYb015	Athe_0248	or1483, or1484
SYb016	Athe_0252	or2645, or2646
SYb016	Athe_0253	or2647
SYb016	Athe_0254	or2648
SYb017	Athe_0258	or2652
SYb017	Athe_0259	
SYb018	Athe_0261	or2655
SYb018	Athe_0262	or2656
SYb019	Athe_0266	or2661
SYb019	Athe_0267	or2662
SYb019	Athe_0268	or2663
SYb019	Athe_0269	or2664
SYb020	Athe_0271	or2665
SYb020	Athe_0272	or2666
SYb020	Athe_0273	or2667
SYb021	Athe_0279	or2673
SYb021	Athe_0280	or2674
SYb021	Athe_0281	or2675
SYb022	Athe_0285	or2680
SYb022	Athe_0286	or2681
SYb022	Athe_0287	or2682
SYb023	Athe_0310	or2367
SYb023	Athe_0311	or2368
FPb013	Athe_0328	or2385
SYb024	Athe_0330	or2387
SYb024	Athe_0331	
SYb025	Athe_0336	or2394
SYb025	Athe_0337	or2395
SYb025	Athe_0338	or2396
SYb026	Athe_0347	
SYb026	Athe_0348	or2920
SYb026	Athe_0349	or2919
SYb026	Athe_0350	or2918
SYb026	Athe_0351	or2917
SYb026	Athe_0352	
SYb026	Athe_0353	or2916
SYb026	Athe_0354	or2915
SYb026	Athe_0355	or2914
SYb026	Athe_0356	
SYb026	Athe_0357	or0501
FPb014	Athe_0366	or0510
SYb027	Athe_0375	or0520
SYb027	Athe_0376	or0521
SYb027	Athe_0377	or0522
SYb027	Athe_0378	or0523
SYb027	Athe_0379	or0524
SYb028	Athe_0384	or0529
SYb028	Athe_0385	or0530
SYb029	Athe_0406	or2843
SYb029	Athe_0407	or2842
SYb029	Athe_0408	or2841
SYb029	Athe_0409	or2840
SYb029	Athe_0410	or2839
SYb029	Athe_0411	or2838
SYb029	Athe_0412	or2837, or2836
SYb029	Athe_0413	or2835, or2836
SYb030	Athe_0416	or2168
SYb030	Athe_0417	or2167
SYb031	Athe_0419	or2165
SYb031	Athe_0420	or2164
SYb031	Athe_0421	or2163
FPb015	Athe_0423	or2161
SYb032	Athe_0450	or2132
SYb032	Athe_0451	or2131
SYb032	Athe_0452	or2130
FPb016	Athe_0456	or2126

TABLE 1-continued

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
FPb017	Athe_0464	or2118
SYb033	Athe_0481	or2097, or2098, or2099, or2599
SYb033	Athe_0482	or2600
SYb033	Athe_0483	or2601
SYb034	Athe_0485	or2604
SYb034	Athe_0486	or2605
SYb034	Athe_0487	or2606
SYb034	Athe_0488	or2607, or2608
FPb018	Athe_0490	or2611
SYb035	Athe_0492	or2614
SYb035	Athe_0493	or2615
SYb036	Athe_0496	or2618
SYb036	Athe_0497	or2619
SYb036	Athe_0498	or2620
FPb019	Athe_0506	or2629
FPb020	Athe_0549	or1663
FPb021	Athe_0590	
FPb022	Athe_0603	or1720
SYb037	Athe_0607	or1724
SYb037	Athe_0608	or1725
FPb023	Athe_0610	or1727
SYb038	Athe_0644	or2728, or2729
SYb038	Athe_0645	or1835, or2729
SYb039	Athe_0673	or1805
SYb039	Athe_0674	or1804
SYb039	Athe_0675	or1803
SYb039	Athe_0676	or1802
SYb039	Athe_0677	or1801
SYb039	Athe_0678	or1800
FPb024	Athe_0681	or1796
SYb040	Athe_0718	or1754
SYb040	Athe_0719	or1753
SYb040	Athe_0720	or1752
SYb040	Athe_0721	or1751
SYb040	Athe_0722	or1750
SYb040	Athe_0723	or1749
SYb040	Athe_0724	or1748
SYb040	Athe_0725	or1747
SYb040	Athe_0726	or1746
FPb025	Athe_0729	or1742
FPb026	Athe_0732	or1739
SYb041	Athe_0737	or1734
SYb041	Athe_0738	or1733
SYb042	Athe_0744	or1362
SYb042	Athe_0745	or1361
SYb042	Athe_0746	or1360
FPb027	Athe_0759	
FPb028	Athe_0768	or1338
FPb029	Athe_0864	or1239
FPb030	Athe_0868	
FPb031	Athe_0871	or1230
FPb032	Athe_0888	or1212
SYb043	Athe_0892	
SYb043	Athe_0893	or1207
SYb043	Athe_0894	
FPb033	Athe_0896	or1204
SYb044	Athe_0899	or1202
SYb044	Athe_0900	or1201
SYb044	Athe_0901	or1200
SYb045	Athe_0903	or1197
SYb045	Athe_0904	or1196
FPb034	Athe_0906	or1195
FPb035	Athe_0908	or1193
SYb046	Athe_0911	or0498
SYb046	Athe_0912	or0497
SYb046	Athe_0913	or0496
FPb036	Athe_0916	or0492, or0493
FPb037	Athe_0923	or0485
FPb038	Athe_0945	or0463

TABLE 1-continued

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
SYb047	Athe_0947	or0460
SYb047	Athe_0948	or0459
SYb047	Athe_0949	or0458
SYb047	Athe_0950	or0457
FPb039	Athe_0956	or0450, or0451
FPb040	Athe_0965	or0440
SYb048	Athe_1024	or0379
SYb048	Athe_1025	or0378
SYb048	Athe_1026	or0377
SYb048	Athe_1027	
SYb049	Athe_1106	or0296
SYb049	Athe_1107	or0295
SYb049	Athe_1108	or0294
SYb049	Athe_1109	or0293
SYb049	Athe_1110	or0292
SYb049	Athe_1111	or0291
SYb049	Athe_1112	or0290
FPb041	Athe_1122	or0279
FPb042	Athe_1130	or0271
FPb043	Athe_1146	or0255
FPb044	Athe_1165	or0236
FPb045	Athe_1174	or0227
SYb050	Athe_1178	
SYb050	Athe_1179	or0222
FPb046	Athe_1203	or0197
FPb047	Athe_1256	or0142
FPb048	Athe_1317	or0080
FPb049	Athe_1329	or0068
SYb051	Athe_1351	or0046
SYb051	Athe_1352	or0045
SYb052	Athe_1364	or0033
SYb052	Athe_1365	or0032
SYb052	Athe_1366	or0029
SYb052	Athe_1367	or0030
SYb052	Athe_1368	or0031
SYb052	Athe_1369	or0028
SYb052	Athe_1370	or0027
FPb050	Athe_1383	or0014
FPb051	Athe_1392	or0005
SYb053	Athe_1394	or0004
SYb053	Athe_1395	or0003
SYb053	Athe_1396	or0002
SYb053	Athe_1397	or0001
FPb052	Athe_1408	or0853
FPb053	Athe_1431	
FPb054	Athe_1468	or0792
FPb055	Athe_1519	or0739
FPb056	Athe_1572	or0685
SYb054	Athe_1581	or0675
SYb054	Athe_1582	or0674
SYb055	Athe_1590	or0666
SYb055	Athe_1591	or0665
SYb055	Athe_1592	or0664
SYb056	Athe_1597	or0658
SYb056	Athe_1598	or0657
SYb056	Athe_1599	or0656
SYb056	Athe_1600	or0655
SYb056	Athe_1601	or0654
SYb056	Athe_1602	or0653
SYb056	Athe_1603	or0652
SYb056	Athe_1604	or0651
SYb056	Athe_1605	or0650
SYb056	Athe_1606	or0649
SYb056	Athe_1607	or0648
FPb057	Athe_1621	or0634
FPb058	Athe_1633	or0622
SYb057	Athe_1658	or0596
SYb057	Athe_1659	or0595
SYb057	Athe_1660	or0594

TABLE 1-continued

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
SYb057	Athe_1661	or0593, or0592
SYb057	Athe_1662	or0591
SYb057	Athe_1663	or0590
SYb057	Athe_1664	or0589
SYb057	Athe_1665	or0588
SYb058	Athe_1683	
SYb058	Athe_1684	or0570
FPb059	Athe_1768	or1570
FPb060	Athe_1771	or1567
FPb061	Athe_1776	or1562
FPb062	Athe_1817	or1519
FPb063	Athe_1845	or1490
SYb059	Athe_1853	or2887, or2888
SYb059	Athe_1854	or2886
SYb059	Athe_1855	or2885
SYb059	Athe_1856	or2910
FPb064	Athe_1858	or2856
FPb065	Athe_1869	or2230
FPb066	Athe_1907	or2192
FPb067	Athe_1931	or2508
SYb060	Athe_1933	or2506
SYb060	Athe_1934	or2505
SYb060	Athe_1935	or2504
SYb060	Athe_1936	or2503
SYb060	Athe_1937	or2502
FPb068	Athe_1957	or2482
SYb061	Athe_1962	or2477
SYb061	Athe_1963	or2476, or2475
SYb061	Athe_1964	or2474, or2475
SYb061	Athe_1965	or2473
SYb061	Athe_1966	or2472
SYb061	Athe_1967	or2471
SYb061	Athe_1968	or2470
SYb061	Athe_1969	or2469
SYb061	Athe_1970	or2468
FPb069	Athe_1977	or2899
SYb062	Athe_1985	or1191
SYb062	Athe_1986	or1190
SYb063	Athe_1989	or1187
SYb063	Athe_1990	or1186
SYb063	Athe_1991	or1185
SYb063	Athe_1992	or1184
SYb063	Athe_1993	or1183
SYb063	Athe_1994	or1182
SYb064	Athe_1996	or1180
SYb064	Athe_1997	or1179
SYb064	Athe_1998	or1178
SYb064	Athe_1999	or1177
SYb064	Athe_2000	or1176
FPb070	Athe_2005	or1171
FPb071	Athe_2013	or1159
SYb065	Athe_2022	or1149
SYb065	Athe_2023	or1148
FPb072	Athe_2025	or1146
SYb066	Athe_2029	or1142
SYb066	Athe_2030	or1141
SYb066	Athe_2031	or1140
FPb073	Athe_2033	or1138
FPb074	Athe_2063	or1107
SYb067	Athe_2076	or1093
SYb067	Athe_2077	or1092
SYb067	Athe_2078	or1091
SYb067	Athe_2079	or1090, or1088, or1089
SYb067	Athe_2080	or1087
SYb067	Athe_2081	or1086
SYb067	Athe_2082	or1085
SYb067	Athe_2083	or1084, or1083
SYb067	Athe_2084	or1082, or1083
SYb067	Athe_2085	or1081
SYb067	Athe_2086	or1080

TABLE 1-continued

PBU Coding Regions		
Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
SYb067	Athe_2087	or1079
SYb067	Athe_2088	or1078
SYb067	Athe_2089	or1077
SYb067	Athe_2090	or1076
SYb067	Athe_2091	or1075
SYb067	Athe_2092	or1074
SYb067	Athe_2093	or1073
SYb067	Athe_2094	or1071, or1072
FPb075	Athe_2103	
FPb076	Athe_2145	or1018
FPb077	Athe_2153	or1010
SYb068	Athe_2187	or0975
SYb068	Athe_2188	or0974
FPb078	Athe_2194	or0968
FPb079	Athe_2196	or0966
SYb069	Athe_2200	or0962
SYb069	Athe_2201	or0961
FPb080	Athe_2203	or0959
FPb081	Athe_2209	or0953
FPb082	Athe_2212	or0950
SYb070	Athe_2216	or0946
SYb070	Athe_2217	or0944
SYb071	Athe_2223	or0937
SYb071	Athe_2224	or0936
SYb072	Athe_2230	or0930
SYb072	Athe_2231	or0929, or0930
SYb072	Athe_2232	or0928
SYb072	Athe_2233	or0927
SYb072	Athe_2234	or0926
SYb072	Athe_2235	or0925
SYb072	Athe_2236	or0923, or0924
SYb072	Athe_2237	or0922
SYb072	Athe_2238	or0921
SYb072	Athe_2239	or0920
SYb073	Athe_2247	or0912
SYb073	Athe_2248	or0911
SYb073	Athe_2249	or0910
SYb073	Athe_2250	or0909
SYb074	Athe_2257	or0901
SYb074	Athe_2258	or0900
SYb074	Athe_2259	or0899
SYb075	Athe_2261	
SYb075	Athe_2262	or0896
SYb075	Athe_2263	or0895
FPb083	Athe_2275	or0883
FPb084	Athe_2290	or0866
SYb076	Athe_2292	or0863, or0864, or2908
SYb076	Athe_2293	or2096
SYb077	Athe_2300	or2088
SYb077	Athe_2301	or2087
SYb078	Athe_2312	or2075
SYb078	Athe_2313	or2074
SYb078	Athe_2314	or2073
SYb078	Athe_2315	or2072
FPb085	Athe_2320	or2067
FPb086	Athe_2325	or2060, or2061
SYb079	Athe_2328	or2057
SYb079	Athe_2329	or2056
SYb080	Athe_2331	or2054
SYb080	Athe_2332	or2053
FPb087	Athe_2344	or2041
SYb081	Athe_2349	or2036
SYb081	Athe_2350	or2035
FPb088	Athe_2353	or2032
SYb082	Athe_2371	or1921
SYb082	Athe_2372	or1922
SYb082	Athe_2373	or1923
SYb082	Athe_2374	or1924
SYb082	Athe_2375	or1925
SYb082	Athe_2376	or1926

TABLE 1-continued

PBU Coding Regions		
Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
FPb089	Athe_2379	or1930
FPb090	Athe_2382	or1933
FPb091	Athe_2404	or1956
SYb083	Athe_2407	or1959
SYb083	Athe_2408	or1960
SYb083	Athe_2409	or1961
SYb083	Athe_2410	or1962
SYb084	Athe_2412	or1964
SYb084	Athe_2413	or1965
SYb084	Athe_2414	or1966
SYb084	Athe_2415	or1967
SYb085	Athe_2417	or1969
SYb085	Athe_2418	or1970
SYb085	Athe_2419	or1971
SYb085	Athe_2420	or1972
SYb085	Athe_2421	or1973
SYb085	Athe_2422	or1974
SYb085	Athe_2423	or1975
SYb085	Athe_2424	or1976
SYb085	Athe_2425	or1977
SYb085	Athe_2426	or1978
SYb085	Athe_2427	or1979
SYb085	Athe_2428	or1980
SYb085	Athe_2429	or1981
SYb086	Athe_2431	or1983
SYb086	Athe_2432	or1984
SYb086	Athe_2433	or1985
SYb086	Athe_2434	or1986
SYb087	Athe_2436	or1988
SYb087	Athe_2437	or1989
SYb087	Athe_2438	or1990
SYb087	Athe_2439	or1991
SYb087	Athe_2440	or1992, or1993
SYb088	Athe_2442	or1996
SYb088	Athe_2443	or1997
SYb088	Athe_2444	or1998
SYb088	Athe_2445	or1999
SYb088	Athe_2446	or2000
FPb092	Athe_2462	or2016
SYb089	Athe_2468	or2913
SYb089	Athe_2469	or2912
SYb090	Athe_2471	
SYb090	Athe_2472	or2834
SYb090	Athe_2473	or2833
SYb091	Athe_2475	or2831
SYb091	Athe_2476	or2830
SYb091	Athe_2477	or2829
SYb091	Athe_2478	or2828
SYb091	Athe_2479	or2827
SYb091	Athe_2480	or2826
FPb093	Athe_2484	or2822
SYb092	Athe_2486	or2820
SYb092	Athe_2487	or2818, or2819
SYb092	Athe_2488	or2817
SYb092	Athe_2489	or2816
SYb092	Athe_2490	or2815
SYb092	Athe_2491	or2814
SYb092	Athe_2492	or2813
SYb093	Athe_2494	or2811
SYb093	Athe_2495	or2810
SYb093	Athe_2496	or2809
SYb093	Athe_2497	or2808
SYb093	Athe_2498	or2807
SYb093	Athe_2499	or2806
SYb093	Athe_2500	or2805
SYb094	Athe_2504	or2801
SYb094	Athe_2505	or2800
SYb094	Athe_2506	or2799
SYb094	Athe_2507	or2798
SYb094	Athe_2508	or2797

TABLE 1-continued

PBU Coding Regions

Cluster/Single Number	GenBank CP001393.1 locus tag	Draft sequence locus tag
SYb094	Athe_2509	or2796
SYb094	Athe_2510	or2795
SYb095	Athe_2512	
SYb095	Athe_2513	
SYb095	Athe_2514	or2464
SYb095	Athe_2515	or2463
SYb095	Athe_2516	or2462
FPb094	Athe_2518	or2460
FPb095	Athe_2525	or2453
FPb096	Athe_2527	or2451
SYb096	Athe_2530	or2448
SYb096	Athe_2531	or2447
SYb096	Athe_2532	or2446
SYb096	Athe_2533	or2445
SYb097	Athe_2536	or2442
SYb097	Athe_2537	or2441
SYb097	Athe_2538	or2440
SYb097	Athe_2539	or2439
SYb097	Athe_2540	or2438
FPb097	Athe_2545	or2432, or2433
SYb098	Athe_2547	or2430
SYb098	Athe_2548	or2429
FPb098	Athe_2556	or2421
SYb099	Athe_2586	or2248
SYb099	Athe_2587	or2249
SYb099	Athe_2588	or2250
FPb099	Athe_2604	or2267
FPb100	Athe_2613	or2276
FPb101	Athe_2622	or2286
SYb100	Athe_2628	or2292
SYb100	Athe_2629	or2293
SYb101	Athe_2634	or2557
SYb101	Athe_2635	or2558
FPb102	Athe_2637	or2560
FPb103	Athe_2647	or2572
SYb102	Athe_2653	or2579, or2580
SYb102	Athe_2654	or2581, or2582
FPb104	Athe_2665	or2591
FPb105	Athe_2667	or2593
FPb106	Athe_2672	or2598
FPb107	Athe_2678	or2346
SYb103	Athe_2686	or2336
SYb103	Athe_2687	or2335
SYb103	Athe_2688	or2334
SYb103	Athe_2689	or2333
SYb103	Athe_2690	or2332
SYb104	Athe_2692	or2329
SYb104	Athe_2693	or2328
SYb104	Athe_2694	or2327
SYb104	Athe_2695	or2326
SYb104	Athe_2696	or2325
SYb104	Athe_2697	or2324
FPb108	Athe_2706	or2315
FPb109	Athe_2709	or2311
SYb105	Athe_2711	or2309
SYb105	Athe_2712	or2308
SYb105	Athe_2713	or2307
FPb110	Athe_2716	or2304
SYb106	Athe_2718	or2299
SYb106	Athe_2719	or2298, or2877
SYb106	Athe_2720	or2876
SYb106	Athe_2721	
FPb111	Athe_2728	or2767
FPb112	Athe_2743	or2752
FPb113	Athe_2764	or2730
FPb114	Athe_2768	or1841

TABLE 2

Exemplary PBU Gene Clusters

Cluster/Single Number	GenBank CP001393.1 locus tag
SYb001	Athe_0010
SYb001	Athe_0011
SYb001	Athe_0012
SYb001	Athe_0013
SYb001	Athe_0014
SYb001	Athe_0015
SYb001	Athe_0016
SYb001	Athe_0017
SYb004	Athe_0052
SYb004	Athe_0053
SYb004	Athe_0054
SYb004	Athe_0055
SYb004	Athe_0056
SYb004	Athe_0057
SYb004	Athe_0058
SYb004	Athe_0059
SYb004	Athe_0060
SYb004	Athe_0061
SYb012	Athe_0153
SYb012	Athe_0154
SYb012	Athe_0155
SYb012	Athe_0156
SYb012	Athe_0157
SYb012	Athe_0158
SYb012	Athe_0159
SYb012	Athe_0160
SYb026	Athe_0347
SYb026	Athe_0348
SYb026	Athe_0349
SYb026	Athe_0350
SYb026	Athe_0351
SYb026	Athe_0352
SYb026	Athe_0353
SYb026	Athe_0354
SYb026	Athe_0355
SYb026	Athe_0356
SYb026	Athe_0357
SYb029	Athe_0406
SYb029	Athe_0407
SYb029	Athe_0408
SYb029	Athe_0409
SYb029	Athe_0410
SYb029	Athe_0411
SYb029	Athe_0412
SYb029	Athe_0413
SYb040	Athe_0718
SYb040	Athe_0719
SYb040	Athe_0720
SYb040	Athe_0721
SYb040	Athe_0722
SYb040	Athe_0723
SYb040	Athe_0724
SYb040	Athe_0725
SYb040	Athe_0726
SYb056	Athe_1597
SYb056	Athe_1598
SYb056	Athe_1599
SYb056	Athe_1600
SYb056	Athe_1601
SYb056	Athe_1602
SYb056	Athe_1603
SYb056	Athe_1604
SYb056	Athe_1605
SYb056	Athe_1606
SYb056	Athe_1607
SYb057	Athe_1658
SYb057	Athe_1659
SYb057	Athe_1660
SYb057	Athe_1661
SYb057	Athe_1662
SYb057	Athe_1663

TABLE 2-continued

Exemplary PBU Gene Clusters	
Cluster/Single Number	GenBank CP001393.1 locus tag
SYb057	Athe_1664
SYb057	Athe_1665
SYb061	Athe_1962
SYb061	Athe_1963
SYb061	Athe_1964
SYb061	Athe_1965
SYb061	Athe_1966
SYb061	Athe_1967
SYb061	Athe_1968
SYb061	Athe_1969
SYb061	Athe_1970
SYb067	Athe_2076
SYb067	Athe_2077
SYb067	Athe_2078
SYb067	Athe_2079
SYb067	Athe_2080
SYb067	Athe_2081
SYb067	Athe_2082
SYb067	Athe_2083
SYb067	Athe_2084
SYb067	Athe_2085
SYb067	Athe_2086
SYb067	Athe_2087
SYb067	Athe_2088
SYb067	Athe_2089
SYb067	Athe_2090
SYb067	Athe_2091
SYb067	Athe_2092
SYb067	Athe_2093
SYb067	Athe_2094
SYb072	Athe_2230
SYb072	Athe_2231
SYb072	Athe_2232
SYb072	Athe_2233
SYb072	Athe_2234
SYb072	Athe_2235
SYb072	Athe_2236
SYb072	Athe_2237
SYb072	Athe_2238
SYb072	Athe_2239
SYb085	Athe_2417
SYb085	Athe_2418
SYb085	Athe_2419
SYb085	Athe_2420
SYb085	Athe_2421
SYb085	Athe_2422
SYb085	Athe_2423
SYb085	Athe_2424
SYb085	Athe_2425
SYb085	Athe_2426
SYb085	Athe_2427
SYb085	Athe_2428
SYb085	Athe_2429

[0175] Of the 218 functionally-annotated coding regions (rather than having an unknown function) found in *A. thermophilum* that are not found in *C. saccharolyticus,* 20 of them encode polysaccharide hydrolases and related (PIM) enzymes (Table 3). Several of the coding regions that encode PHR enzymes are part of eight so-called PHR gene clusters (Table 4). These include clusters of six (SYb082), 19 (SYb067), six (SbYb063) eight (SYb012) and 10 (SYb004) coding regions (see Table 4). The PHR clusters contain almost 60 coding regions (including the 20 PHR coding regions).

TABLE 3

PHR Coding Regions	
Cluster/Single Number	GenBank CP001393.1 locus tag
SYb004	Athe_0058
SYb004	Athe_0059
SYb004	Athe_0061
SYb007	Athe_0089
SYb012	Athe_0154
SYb012	Athe_0156
SYb012	Athe_0157
FPb015	Athe_0423
SYb032	Athe_0452
FPb022	Athe_0603
FPb023	Athe_0610
SYb059	Athe_1853
SYb059	Athe_1854
SYb059	Athe_1855
SYb063	Athe_1993
SYb067	Athe_2076
SYb067	Athe_2086
SYb067	Athe_2089
SYb067	Athe_2094
SYb082	Athe_2371

TABLE 4

PHR Gene Clusters	
Cluster/Single Number	GenBank CP001393.1 locus tag
SYb004	Athe_0052
SYb004	Athe_0053
SYb004	Athe_0054
SYb004	Athe_0055
SYb004	Athe_0056
SYb004	Athe_0057
SYb004	Athe_0058
SYb004	Athe_0059
SYb004	Athe_0060
SYb004	Athe_0061
SYb007	Athe_0088
SYb007	Athe_0089
SYb007	Athe_0090
SYb012	Athe_0153
SYb012	Athe_0154
SYb012	Athe_0155
SYb012	Athe_0156
SYb012	Athe_0157
SYb012	Athe_0158
SYb012	Athe_0159
SYb012	Athe_0160
SYb032	Athe_0450
SYb032	Athe_0451
SYb032	Athe_0452
SYb059	Athe_1853
SYb059	Athe_1854
SYb059	Athe_1855
SYb059	Athe_1856
SYb063	Athe_1989
SYb063	Athe_1990
SYb063	Athe_1991
SYb063	Athe_1992
SYb063	Athe_1993
SYb063	Athe_1994
SYb067	Athe_2076
SYb067	Athe_2077
SYb067	Athe_2078

TABLE 4-continued

PHR Gene Clusters	
Cluster/Single Number	GenBank CP001393.1 locus tag
SYb067	Athe_2079
SYb067	Athe_2080
SYb067	Athe_2081
SYb067	Athe_2082
SYb067	Athe_2083
SYb067	Athe_2084
SYb067	Athe_2085
SYb067	Athe_2086
SYb067	Athe_2087
SYb067	Athe_2088
SYb067	Athe_2089
SYb067	Athe_2090
SYb067	Athe_2091
SYb067	Athe_2092
SYb067	Athe_2093
SYb067	Athe_2094
SYb082	Athe_2371
SYb082	Athe_2372
SYb082	Athe_2373
SYb082	Athe_2374
SYb082	Athe_2375
SYb082	Athe_2376

[0176] The PHR coding regions and particularly the PHR clusters together with other coding regions in the 550 gene set found in *A. thermophilum* that are not found in *C. saccharolyticus* form what are referred to herein as the plant biomass utilization, or PBU, coding regions. The PBU coding regions are directly and indirectly involved in enabling *A. thermophilum* to efficiently utilize untreated, treated, and spent plant biomass. Thus, the ability to confer to other microorganisms the ability to utilize untreated and/or spent biomass can be achieved by directly transferring certain PBU polynucleotides to microorganisms known to utilize, for example, cellulose and xylan. Since *A. thermophilum* grows at moderate temperatures (75° C. optimum, but remain viable at, for example 90° C.), the microorganisms receiving an *A. thermophilum* PBU polynucleotide can include thermophilic microorganisms, including extreme thermophiles, as well as microorganisms that grow at more moderate temperatures (mesophiles).

[0177] Coding regions that enable *A. thermophilum* to efficiently breakdown plant biomass encode various types of proteins, including what are referred to herein as carbohydrate-active enzymes (CAZy) as well as proteins that may not be catalytic but allow the microorganism to attach to the insoluble biomass prior to and during degradation. FIG. **27** lists CAZy-related domains—found in enzymes such as glycoside hydrolases, glycosyl transferases, and carbohydrate esterases—that are present in the genomes of *A. thermophilum* and *C. saccharolyticus*. Such domains can be highly conserved between functionally related proteins and between species. Thus, the structure and function of many CAZy-related domains are well characterized. FIG. **28** lists CAZy-related domains that are uniquely present in *A. thermophilum*. In addition, *A. thermophilum* has some unique combinations of these domains that are not present in *C. saccharolyticus* (FIG. **25** and FIG. **29**). Some of these and other CAZy-related coding regions are expressed at different times throughout the growth phase when *A. thermophilum* is grown on crystalline cellulose, as shown by proteomic identification of the proteins released by the microorganism into the growth medium (FIG. **31**). Numerous non-catalytic extracellular and membrane-associated proteins were also identified in the *A. thermophilum* genome that could potentially mediate its attachment to biomass (FIG. **32**). Using the same proteomics analyses, several of these have been measured in either the extracellular fraction or the membrane fraction of A. thermophilum when grown on cellulose, xylan, switchgrass, and/or poplar (FIG. **32**). FIG. **33** lists some other proteins, measured by proteomic analysis, that are not encoded in the genome of *C. saccharolyticus* but are produced by A. thermophilum when the microorganism is grown on cellulose, xylan, switchgrass, and/or poplar.

[0178] An *A. thermophilum* PBU polynucleotide can include one or more of the PBU coding regions identified in Table 1. In some embodiments, the *A. thermophilum* PBU polynucleotide can include one or more coding regions of a PBU gene cluster as identified in Table 2. In certain embodiments, the *A. thermophilum* PBU polynucleotide may be an *A. thermophilum* PHR polynucleotide—i.e., include one or more of the *A. thermophilum* PHR coding regions identified in Table 3. In some embodiments, the *A. thermophilum* PHR polynucleotide can include one or more coding regions of a PHR gene cluster as identified in Table 4. The complete nucleotide sequence—and the predicted amino sequence encoded by the nucleotide sequence—of every remaining *A. thermophilum* PBU coding region is accessible via GenBank Accession No. CP001395 (version 1, created Feb. 5, 2009).

[0179] An *A. thermophilum* polynucleotide can include one or more *A. thermophilum* coding regions that encode products that are involved in plant biomass utilization, but may not necessarily be specific to *A. thermophilum* compared to *C. saccharolyticus*. Such coding regions can include, for example, Athe1867 (SEQ ID NO:6). Consequently, the *A. thermophilus* polynucleotide can encode a polypeptide having the amino acid sequence of, for example, SEQ ID NO:7.

[0180] Thus, in another aspect, the present invention provides methods of transferring one or more polynucleotides of *A. thermophilum* to a recipient microorganism. In some cases, such methods can include the cloning and direct transfer of one or more polynucleotides from *A. thermophilum* to the recipient microorganism. Such methods are routine and known to those skilled in the art. (See, e.g., Sambrook et al, (1989) *Molecular Cloning: A Laboratory Manual*., Cold Spring Harbor Laboratory Press or Ausubel, R. M., ed. (1994). *Current Protocols in Molecular Biology*).

[0181] When direct cloning methods are used to transfer one or more polynucleotides from *A. thermophilum* to a recipient microorganism, the recipient microorganism may be any microorganism suitable for cloning transfer of polynucleotides. Suitable recipient microorganisms include, for example, members of the family Enterobacteriaceae such as, for example, members of the genus *Escherichia* or *Salmonella*. In certain embodiments, a suitable recipient microorganism may include *E. coli*. In other embodiments, the recipient microorganism can include a eukaryote such as, for example, a yeast such as, for example, *Saccharomyces cerevisiae*.

[0182] In other cases, such methods can include the cloning and transfer of one or more polynucleotides from *A. thermophilum* to an intermediate, or "vector," microbe, followed by transfer of the one or more *A. thermophilum* polynucleotides from the vector microbe to the recipient microorganism. The

cloning of the one or more *A. thermophilum* polynucleotides into the vector microbe may be accomplished using routine methods referred to in the immediately preceding paragraph. Alternatively, the cloning of one or more *A. thermophilum* polynucleotides into the vector microbe may be accomplished using a shuttle vector that permits the movement of nucleotide sequences cloned into the shuttle vector to be shuttled between *A. thermophilum* and another microorganism. One such shuttle vector is pDCW 31, the construction of which is described in Example 5 and is shown in FIG. **26**. The pCDW 31 shuttle vector contains elements from the naturally-occurring *A. thermophilum* plasmid pAthe02 (SEQ ID NO:1) and the pSC101-based plasmid pJHW007. While components of the pJHW007 plasmid were used to construct pCDW 31, analogous components of any pSC101-based plasmid can be used to construct a similar shuttle vector.

[0183] The subsequent transfer of the one or more *A. thermophilum* polynucleotides to a recipient microorganism may be accomplished by any method appropriate for transferring a polynucleotide to the particular recipient microorganism. In some cases, an appropriate method may include routine cloning methods already described. In other cases, an appropriate method may include methods described in U.S. Provisional Patent Application Ser. No. 61/000,338, filed, Oct. 25, 2007, entitled "METHODS FOR GENETIC MANIPULATION OF EXTREMOPHILES," which describes the transfer of polynucleotides by conjugation. Conjugation is a polynucleotide transfer process in which a donor microbe (e.g., a vector microbe) makes contact with and transfers a polynucleotide to a recipient (Frost et al., Microbiol. Rev., 1994, 58:162-210); Willets and Skurray, In: *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, Neidhardt et al. (eds.), 1987, American Society for Microbiology, Washington, D.C., 1110-1133). Generally, such methods include co-cultivating a vector microbe and a recipient microorganism, wherein the vector microbe includes a conjugative polynucleotide, and wherein the co-cultivation is under conditions suitable for conjugative transfer of at least a portion of the conjugative polynucleotide from the vector microbe to the recipient microorganism, and identifying a recipient microorganism exconjugant. Conjugation from a vector microbe to a recipient microorganism can result in the transfer of a plasmid or in the transfer of part of the vector microbe's chromosome. Preferably, the methods described herein result in transfer of a plasmid from vector microbe to the recipient microorganism.

[0184] In particular, conjugative methods may be appropriate if the recipient microorganism is, for example, an extremophile or a mesophile. Examples of extremophiles include, but are not limited to, thermophiles and extreme thermophiles (microorganisms that grow in environments at temperatures of between 50° C. and 100° C., and between 70° C. and 100° C., respectively), hyperthermophiles (microorganisms that grow in environments at temperatures above 80° C.), acidophiles (microorganisms that grow in environments at low pH, such as less than pH 3), and halophiles (microorganisms that grow in environments of at least 1 M NaCl). The extremophile may be an obligate anaerobe. The extremophile may be a member of the kingdom Archaea such as, for instance, a member of phylum Crenarchaeota, Euryarchaeota, Korarchaeota, or Nanoarchaeota, preferably Crenarchaeota or Euryarchaeota, more preferably, Euryarchaeota. Examples of such microorganisms include, but are not limited to, *Pyrococcus* spp., such as *P. furiosus, Sulfolobus* spp, such as *S. solfataricus*, and *Thermococcus* spp., such as *T kodakaraensis*. The extremophile may be a member of the family Thermotogaceae, such as, for example, *Thermotoga* spp. such as, for example, *T. maritima*, or a member of the family Aquificaceae, such as, for example, *Aquifex* spp such as, for example, *A. aeolicus*. Examples of thermophiles that are not extreme thermophiles include, for example, *A. thermophilum, Caldicellulosiruptor saccharolyticus*, and *Clostridium thermocellum*. Examples of mesophiles include, for example, members of the family Enterobacteriaceae such as, for example, members of the genus *Escherichia* or *Salmonella*. In certain embodiments, a suitable mesophile may include *E. coli*.

[0185] The vector microbe may be a member of the family Enterobacteriaceae and may be, but is not limited to, *E. coli* and *Salmonella* spp. The member of the family Enterobacteriaceae is one that is able to transfer polynucleotides by conjugation with the recipient microorganism. Alternatively, the vector microbe may be a member of the family Bacillaceae such as, for example, *Bacillus* spp.

[0186] In some embodiments, the polynucleotide to be transferred to the recipient microorganism (e.g., the cloning vector or conjugative polynucleotide) can include an *A. thermophilum* PBU coding region as defined above. The transfer of a polynucleotide that includes an *A. thermophilum* PBU coding region can permit the recipient microorganism (e.g., the cloning recipient or the exconjugant) to express an *A. thermophilum* polypeptide—as defined above—encoded by the *A. thermophilum* PBU coding region. Exemplary PBU polypeptides are encoded by *A. thermophilum* PBU coding regions identified in Table 1. The amino acid sequences of PBU polypeptides encoded by the exemplary PBU coding regions are accessible via GenBank Accession No. CP001395 (version 1, created Feb. 5, 2009).

[0187] In some embodiments, the polynucleotide to be transferred to the recipient microorganism (e.g., the cloning vector or conjugative polynucleotide) can include a PHR coding region as defined above—i.e., a member of a subset of PBU coding regions. The transfer of a polynucleotide that includes an *A. thermophilum* PHR coding region can permit the recipient microorganism (e.g., the cloning recipient or the exconjugant) to express an *A. thermophilum* polypeptide—as defined above—encoded by the *A. thermophilum* PHR coding region. Exemplary PHR coding regions are identified in Table 3. The amino acid sequences of PHR polypeptides encoded by the exemplary PHR coding regions are accessible via GenBank Accession No. CP001395 (version 1, created Feb. 5, 2009).

[0188] The recombinantly expressed *A. thermophilum* polypeptide (e.g., a PBU polypeptide or a PHR polypeptide) may be isolated from the recipient cell—whether a cloning recipient or an exconjugant—using methods well-known in the art. Consequently, in another aspect, the present invention provides an isolated polypeptide encoded by an *A. thermophilum* PBU polynucleotide or a PHR polynucleotide.

[0189] In another aspect, the present invention provides a genetically-modified microorganism that includes one or more *Anaerocellum thermophilum* plant biomass utilization (PBU) polynucleotides. The genetically-modified microorganism may be derived from one of the recipient microorganisms described above with respect to methods of transferring at least a portion of an *A. thermophilum* polynucleotide to a recipient microorganism. Also, the genetically-modified microorganism may include one or more PBU coding

regions, PHR coding regions, or one or more coding regions from a gene cluster identified above.

[0190] In some embodiments, the genetically-modified microorganism may be modified in a way to promote the production and/or accumulation of a particular metabolic product. As noted above, such genetic modifications can include the introduction of one or more heterologous coding regions that promote the production of one or more desired products or intermediates. In other cases, such genetic modifications can include disrupting the activity of one or more endogenous coding regions in a way that inhibits the production of non-desired metabolic products and/or redirects the metabolism of intermediates toward the production of desired metabolic products.

[0191] For example, metabolic pathways that supply or are supplied by the citric acid cycle are well known to those skilled in the art. Thus, disrupting—either by reducing or eliminating the activity of products encoded by certain coding regions—a metabolic pathway that is, at least in part, supplied by the citric acid cycle can shunt metabolism away from the disrupted pathway (and its product) in favor of accumulating other intermediates of the citric acid cycle and/or pathways supplied by those alternative intermediates. Examples of modifications that disrupt a metabolic pathway include, for example, "knock out" mutations that significantly reduce or eliminate biological activity of the mutated coding region (and/or the polypeptide encoded by the mutated coding region). Methods for introducing knock out mutations in many cellular models are routine and known to those skilled in the art. In other words, one may direct metabolism toward pathways that produce desired products by reducing or eliminating metabolism via pathways that compete with the desired pathway for metabolic resources.

[0192] For example, modifications that disrupt one or more metabolic enzymes involved in a pathway supplied by the citric acid cycle can promote the accumulation of, for example, succinate that would otherwise be metabolized—either directly by the disrupted pathway or indirectly to form the citric acid cycle intermediate that would be directly metabolized by the disrupted pathway. Disrupting activity in other well known metabolic pathways can promote production of, for example, ethanol, acetate, lactate, hydrogen gas, etc. Exemplary targets for such knock out mutations in *A. thermophilum* include, for example, Athe_1918 (SEQ ID NO:8), Athe_2388 (SEQ ID NO:10), Athe_1493 (SEQ ID NO:12), Athe_1494 (SEQ ID NO:14), Athe_1223 (SEQ ID NO:16), but those skilled in the art can readily determine additional targets in *A. thermophilum* by identifying coding regions in *A. thermophilum* that correspond to known components of known and conserved metabolic pathways other microorganisms.

[0193] Such modifications may be provided alone or in combination with one or more additional modifications such as, for example, introduction of a heterologous coding region that promotes the conversion of an intermediate (e.g., an intermediate accumulated due to a knock out modification) to a desired product (e.g., a metabolic product not produced—or produced inefficiently—by the wild type of the genetically-modified microorganism. In some cases, the production of one or more butanols may be promoted in *A. thermophilum* by a combination of disrupting one or more *A. thermophilum* metabolic pathways and introducing one or more heterologous coding regions that promote the production of butanol from. In one exemplary embodiment, a knock out modification in one or more of SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, or SEQ ID NO:16 may be combined with introducing one or more coding regions of *Clostridium acetobutylicum* that are known to confer the ability to produce 1-butanol in *E. coli* such as, for example, the coding region for *C. acetobutylicum* thiolase (Atsumi et al., *Metab. Eng.* 2008, 10:305-311.

[0194] In yet another aspect, the present invention provides a method of processing plant biomass. In this aspect, the method includes growing genetically-modified microorganisms comprising one or more *A. thermophilum* PBU polynucleotides on a substrate that comprises plant biomass under conditions effective for the microorganism to convert at least a portion of the plant biomass to a water soluble product.

[0195] Generally, the plant biomass, the cultivation conditions, the microorganisms, and PBU polynucleotides may be those described above in connection with various embodiments of other aspects of the present invention. In some embodiments, the genetically-modified microorganism may be *A. thermophilum*. In other embodiments, the genetically-modified microorganism may be a microorganism other than *A. thermophilum*.

[0196] Another utility of *A. thermophilum* and/or the genetically-modified microorganisms described above may be for the production of one or more *A. thermophilum* polypeptides that possesses acellular plant biomass degrading activity—i.e., is able to degrade plant biomass when isolated from *A. thermophilum*. Thus, in another aspect, the present invention provides a method of making an isolated *A. thermophilum* polypeptide. Generally, the method includes growing a microorganism comprising at least one polynucleotide encoding an *Anaerocellum thermophilum* polypeptide possessing plant biomass degrading activity under conditions effective for the microorganism to produce the *A. thermophilum* polypeptide, and isolating the *A. thermophilum* polypeptide.

[0197] In some embodiments, the microorganism may be *A. thermophilum*. In other embodiments, the microorganism may be genetically engineered to include one or more *A. thermophilum* PBU polynucleotides, PHR polynucleotides, or one or more coding regions from a gene cluster identified above. Methods for isolating polypeptides produced by microorganisms in culture are well known to those skilled in the art. Polypeptides and fragments thereof useful in the present invention may be produced using recombinant DNA techniques, such as an expression vector present in a cell. Such methods are routine and known in the art. The polypeptides and fragments thereof may also be synthesized in vitro, e.g., by solid phase peptide synthetic methods. The solid phase peptide synthetic methods are routine and known in the art. A polypeptide produced using recombinant techniques or by solid phase peptide synthetic methods may be further purified by routine methods, such as fractionation on immunoaffmity or ion-exchange columns, ethanol precipitation, reverse phase HPLC, chromatography on silica or on an anion-exchange resin such as DEAE, chromatofocusing, SDS-PAGE, ammonium sulfate precipitation, gel filtration using, for example, Sephadex G-75, or ligand affinity.

[0198] In some cases, the isolated polypeptide may be used to directly for biomass conversion. Thus, in yet another aspect, the present invention provides a method of processing plant biomass. Generally, the method includes providing an isolated *A. thermophilum* polypeptide possessing plant biomass degrading activity, and contacting the *A. thermophilum*

polypeptide with plant biomass under conditions effective for the *A. thermophilum* polypeptide to at least partially degrade the plant biomass.

[0199] In certain circumstances, it may be desirable to have the *A. thermophilum* utilization of plant biomass result in the production of an product that *A. thermophilum* is not naturally capable of producing. In such cases, the water soluble product produced by methods described herein may be recovered and subsequently processed to produce a desired end product. In other cases, the desired end product may be a product of a metabolic process native to another microorganism that is made possible by expression of one or more coding regions from that microorganism. Transfer of a polynucleotide that includes one or more such coding regions to *A. thermophilum* may permit the *A. thermophilum* to perform one or more additional metabolic steps to convert the water soluble product to the desired product.

[0200] Thus, in yet another aspect, the present invention provides methods of transferring one or more polynucleotides that include heterologous coding regions—e.g., carbohydrate metabolism coding regions or butanol synthesis coding regions—to *A. thermophilum*. Metabolic pathways in *E. coli* for producing, for example, various biofuels are known and coding regions of the *E. coli* genome that promote the production of the various biofuels are similarly known. (See, e.g., Connor et al., Curr. Opin. Biotech. 2009, 20:307-315 and Atsumi et al., Metab. Eng. 2008, 10:305-311).

[0201] One or more heterologous coding regions may be introduced into *A. thermophilum* using any suitable method including, for example, routine cloning and direct transfer of polynucleotides containing the heterologous coding region, cloning and transfer of one or more polynucleotides to *A. thermophilum* via an intermediate, or "vector," microbe, or the transfer of polynucleotides by conjugation, as described above. In addition, a polynucleotide that includes one or more heterologous coding regions may be introduced into *A. thermophilum* by, for example, electroporation as described in Example 6, below.

[0202] Generally, the plant biomass, the processing conditions (e.g., temperature), and the *A. thermophilum* polypeptide may be those described above in connection with various embodiments of other aspects of the present invention.

[0203] The present invention is illustrated by the following examples. It is to be understood that the particular examples, materials, amounts, and procedures are to be interpreted broadly in accordance with the scope and spirit of the invention as set forth herein.

EXAMPLES

Example 1

[0204] *Anaerocellum thermophilum* strain DSM 6725 (Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH (DSMZ), Braunschweig, Germany) was grown in 0.5% modified 516 medium (DSMZ). The medium was modified by adding vitamins and trace minerals solutions and the method to reduce the medium. The modified medium contained, per liter: 0.5 g yeast extract, 0.33 g NH_4C1, 0.33 g

[0205] KH_2PO_4, 0.33 g KCl, 0.33 g $MgCl_2$×6 H_2O, 0.33 g $CaCl_2$×2 H_2O, 0.5 mg resazurin, 5 mL vitamin solution, and 1 mL trace minerals solution. The vitamin solution contained: 4 mg/L biotin , 4 mg/L folic acid, 20 mg/L pyridoxine-HCl, 10 mg/L thiamine-HCl, 10 mg/L riboflavin, 10 mg/L nicotinic acid, 10 mg/L calcium panthothenate, 0.2 mg/L vitamin B_{12}, 10 mg/L p-aminobenzoic acid, and 10 mg/L lipoic acid. The trace minerals solution contained: 2 g/L $FeCl_3$, 0.05 g/L $ZnCl_2$, 0.05 g/L $MnCl_2$×4H_2O, 0.05 g/L H_3BO_3, 0.05 g/L $CoCl_2$×6H_2O, 0.03 g/L $CuCl_2$×2H_2O, 0.05 g/L $NiCl_2$×6H_2O, 0.5 g/L Na_4EDTA (tetrasodium salt), 0.05 g/L $(NH_4)2MoO_4$, and 0.05 g/L $AlK(SO_4)_2$.12H_2O. Both vitamin and trace minerals solutions were filtered through 0.22 pm membrane and stored at 4° C. The reducing system was composed of 0.5 g cysteine, 0.5 g N_2S, and 1 g $NaHCO_3$. The final pH was 7.2. The medium was filtered through 0.22 μM membrane and prepared anaerobically under 80% N_2 +20% CO_2 (N_2/CO_2) gas atmosphere. Soluble growth substrates were added into the medium prior to filtration. Insoluble growth substrates were weighed and added into sterilized culture bottles individually.

[0206] The growth substrates and their sources were: D-(+)-cellobiose (cat. C7252) and oat spelts xylan (cat. X0627) were from Sigma Chemical Company, St. Louis, Mo., and Avicel PH-101 (cat. 11365) was from Fluka, Switzerland), Poplar and switchgrass (sieved, −20/+80 mesh fraction) were provided by Dr. Brian Davison of Oak Ridge National Laboratory (Oak Ridge, Tenn.), Tifton 85 bermuda grass and napier grass (sieved, −20/+80 mesh fraction) were provided by Dr. Joy Peterson (Department of Microbiology, University of Georgia, Athens, Ga.), and the pine wood was provided by Dr. Alan Darvill (Department of Biochemistry and Complex Carbohydrate Research Center, University of Georgia, Athens, Ga.).

[0207] *A. thermophilum* was grown at 75° C. with shaking at 150 rpm unless specified otherwise. To test the ability of *A. thermophilum* to grow on untreated plant biomass, *A. thermophilum* was grown in 50 mL 0.5% modified 516 medium in sealed 100-mL serum bottles without shaking. For the kinetic analyses, A. thermophilum was grown in either 0.5 L or 0.25 L cultures in 1 L or 0.5 L sealed bottles, respectively. "Flushed" cultures were grown in the same conditions, but the cultures were purged with N_2/CO_2. For growth on "spent" insoluble substrates (from poplar, switchgrass and Avicel), the insoluble material that was left over after cells had grown on that substrate was collected in late stationary phase (when cell growth had stopped). The residual insoluble substrate was separated from the cells by filtering through glass filters with a pore size 40-60 μm. The material was washed with distilled water and dried at 50° C. overnight. This was then used as the growth substrate for new cultures.

[0208] During growth of *A. thermophilum* on different complex and defined substrates, samples were removed from the cultures at various time intervals (FIGS. **1-4**). Some or all of the following parameters were measured: pH, cell density, cell protein, hydrogen, acetate, lactate, ethanol, and in some cases, reducing sugars. The cell count was determined using a phase-contrast microscope with 40× magnification. Cell protein was determined by the Bradford method. For cell protein assay in cultures growing on insoluble substrate, the cells were separated from the substrate by a low speed centrifugation. To measure protein, the cell pellet resuspended in 50 mM Tris-HCl (pH 7.0) buffer with lysozyme (0.2 mg/ml) was incubated at 10° C. for 6 hours and then subjected to three freeze-thaw cycles. Acetate and lactate were measured in the growth medium after removing cells (and the insoluble substrate if present) by HPLC (Waters 2690 Separations Module, Waters Corp., Milford, Mass.) equipped with a Aminex HPX-87H column (300 mm 7.8 mm, Bio-Rad Corp., Hercules, Calif.) at 40° C. with 5 mM H_2SO_4 as the mobile phase at a

flow rate of 0.6 ml min^{-1} with a refractive index detector (Waters 2410, Waters Corp., Milford, Mass.). Ethanol was measured enzymatically using the Ethanol Kit (Megazyme International Ireland Ltd., Wicklow, Ireland). Hydrogen producing during cell growth was determined by gas chromatography (Shimadzu GC-8A, Shimadzu Scientific Instruments, Inc., Columbia, Md.) equipped with a thermal conductivity detector and a molecular sieve column (Alltech 5A 80/100, Grace Davison Discovery Sciences, Waukegan, Ill.) with argon as the carrier gas. Reducing sugars were determined with dinitrosalicylic acid (DNS) reagent as previously described (Miller, G. L., 1959, Anal. Chem., 31:426-428).

[0209] The data shown in FIGS. **12-15** used the defined medium that we developed for *A. thermophilum* (DSMZ 6725). The same medium was also used to grow *Caldicellulosiruptor saccharolyticus* (DSMZ 8903). Both microorganisms were grown in 50 mL culture volumes in a medium containing: 0.33 g/L $MgCl_2$, 0.33 g/L KCl, 0.25 g/L NH_4Cl, 0.14 g/L $CaCl_2$, trace minerals (Na_4EDTA, $FeCl_3$, $ZnCl_2$, $MnCl_2$, H_3B0_3, $CoCl_2$, $CuCl_2$, $NiCl_2$, $(NH_4)_2MoO_4$, AlK (SO_4)), vitamin mix (0.02 mg/L biotin, 0.02 mg/L folic acid, 0.1 mg/L pyridoxine-HCl, 0.05 mg/L thiamine, 0.05 mg/L riboflavin, 0.05 mg/L nicotinic acid, 0.05 mg/L D-Ca-pantothenate, 0.001 mg/L vitamin B_{12}, 0.05 mg/L p-aminobenzoic acid, 0.05 mg/L lipoic acid), 20 amino acids (0.076 g/L alanine, 0.124 g/L arginine, 0.1 g/L asparagine, 0.048 g/L aspartic acid, 0.2 g/L glutamic acid, 0.048 g/L glutamine, 0.2 g/L glycine, 0.1 g/L histidine, 0.1 g/L isoleucine, 0.1 g/L leucine, 0.1 g/L lysine, 0.076 g/L methionine, 0.076 g/L phenylalanine, 0.125 g/L proline, 0.076 g/L serine, 0.1 g/L threonine, 0.076 g/L tryptophan, 0.012 g/L tyrosine, 0.052 g/L valine, 0.5 g/L cysteine), 0.25 mg/mL resazurin, 1 mM KH_2PO_4, 0.5 g/L Na_2S, and 1.0 g/L $NaHCO_3$. The heat-treated biomass samples were prepared by taking switchgrass, poplar or pine (100 mg) and extracting them for 2 minutes with 2 mL sterile water at 98° C. The soluble material was removed and used as a growth substrate for one culture and the insoluble solid was used as the growth substrate for a separate culture. Cultures were grown in triplicate at 75° C. without stirring or shaking. The cell density was measured as described above.

Example 2

[0210] CelA (Athe_1867, or2232, SEQ ID NO:6) encodes a cellulase coding region in *A. thermophilum* with an activity not present in the hyperthermophile *P. furiosus* , a microorganism that grows optimally at 100° C. The CelA coding region contains two cellulase enzymatic domains intermixed with carbohydrate binding domains. Two forms of the CelA coding region from *A. thermophilum* are generated and introduced into *P. furiosus* by mating as described in U.S. Provisional Patent Application Ser. No. 61/000,338, entitled "METHODS FOR GENETIC MANIPULATION OF EXTREMOPHILES," filed Oct. 25, 2007. The first form consists of part of the native CelA nucleotide sequence itself (a single cellulase enzymatic domain and a single carbohydrate binding domain adjacent to it). This truncated form of CelA is cloned by PCR amplification from *A. thermophilum* into *E. coli* in a vector for mating into *P furiosus*. The second form of CelA consists of these domains proceeded by a signal sequence for protein localization. The signal sequence is from the *P. furiosus* alpha amylase coding region.

[0211] The DNA sequence of the CelA coding region and signal sequence are shown in FIGS. **16** and **17** respectively. Plasmid maps of these constructions are shown in FIGS. **18** and **19**.

[0212] These plasmids are mated into *P. furiosus* and exconjugants are selected on simvastatin using methods described as follows:

Media Components

[0213] 1000× (1 mL/L) Trace Minerals Solution: 1.00 mL/L HCl (concentrated), 0.50 g/L
Na_4EDTA (tetrasodium), 2.00 g/L $FeCl_3$, 0.05 g/L H_3BO_3, 0.05 g/L $ZnCl_2$, 0.03 g/L

$CuCl_2.2H_2O$, 0.05 g/L $MnCl_2.4H_2O$, 0.05 g/L $(NH_4)_2MoO_4$, 0.05 g/L

$AlK(SO_4).2H_2O$, 0.05 g/L $CoCl_2.6H_2O$, and 0.05 g/L $NiCl_2.6H_2O$.

[0214] 5× Base Salts: 140.00 g/L NaCl, 17.50 g/L, $MgSO_4.7H_2O$, 13.50 g/L $MgCl_2.6H_2O$, 1.65 g/L KCl, 1.25 g/L NH_4Cl, 0.70 g/L $CaCl_2.2H_2O$.
Liquid complex cellobiose (CC) media (pH 6.8): 200 mL/L 5× Base salts, 1 mL/L 1000× Trace minerals, 100 μL/L 100 mM $Na_2WO_4*2H_2O$, 50 μL/L Resazurin (5 mg/mL), 5 mL/L 10% w/v Yeast Extract, 50 mL/L 10% w/v Casein hydrolysate, 35 mL/L 10% w/v Cellobiose, 0.5 g/L Cysteine, 0.5g Na_2S, 1 g/L $NaHCO_3$, 1 mL/L 1M K_2HPO_4 buffer.
Solid complex cellobiose (CC) media: 1× media +1% phytagel solution (Sigma Chemical Company, St. Louis, Mo.).
CC plates containing 5-fluoroorotic acid (5-FOA): to ensure complete 5-FOA solvation, 1M NaOH is dripped into the solution until a murky consistency is reached at around pH 10, cysteine is then used to lower the pH to 7, where the solution turns transparent.
Simvastatin plates: solid complex cellobiose plates with the indicated amount of simvastatin added.
A. thermophilum is sensitive to 8 millimolar (mM) 5-FOA, 30 mM hygromycin, 8 micromolar (μM) simvastatin, and 50 μM apramycin.

Growth Conditions.

[0215] *P. furiosus* strain (DSM 3638) (DSMZ, Braunschweig, Germany) is grown in liquid complex cellobiose (CC) media and on solid CC plates containing 1% phytagel. 50 mL liquid cultures are incubated in serum bottles and phytagel-containing plates of solid media are cultivated in anaerobic jars. Both types of media are grown at 90° C. under an argon atmosphere introduced through a vacuum manifold. Single crossover mutants containing an up-regulated HMG CoA reductase coding region are selected for on CC plates containing 8 μM Simvastatin (Sigma Chemical Company, St. Louis, Mo.). PyrF deletion mutants are selected for on CC plates containing 0.25% 5-FOA (Zymo Research Corp., Orange, Calif.). *P. furiosus* cells are plated on solid media by adding 50 μL of cell suspension to a pool of 800 μL 1× base salts. The plates are then spun by hand to spread the cells by centrifugal force. *E. coli* strains XL10 (Stratagene, LaJolla, Calif.) and ET12576 (Beirman et al., Gene 1992, 116L43-49) are grown in both liquid LB media and on solid LB plates at 37° C.

Growth Measurements.

[0216] Cell counts are estimated by direct observation 2 μL of cell sample using a Petroff-Hauser counting chamber

under 40× magnification. Viable cell count is determined by plating 1/100 and 1/1000 dilutions of cell culture and recording the number of colony forming units.

Conjugation Procedure.

[0217] *P. furiosus* strain (DSM 3638) (DSMZ, Braunschweig, Germany) is used as the recipient strain in the conjugation experiments. 100 mL of a 1% v/v inoculum *P. furiosus* are incubated for nine hours to a cell density of approximately 10^8 cells/mL. The cells are then pelleted at 5100 rpm for 15 minutes and washed twice with 1× base salts before resuspending in a final volume of 3 mL 1× base salts. *E. coli* strain ET12576, carrying the helper plasmid PUZ8002 and the conjugation plasmid, was used as the donor. An *E. coli* culture of 50 mL LB media containing 50 μg/mL kanamycin (selection for PUZ8002) and 50 μg/mL apramycin (selection for conjugation plasmid) is incubated overnight until a cell density of approximately 10^9 cells/mL is reached. The *E. coli* is then pelleted at 2500 rpm for 10 minutes and washed twice with LB. 1 mL of the *P. furiosus* cell suspension is used to resuspend the *E. coli* control pellet, carrying only the PUZ8002 plasmid. The remaining 2 mL of *P. furiosus* are combined with the pellet of *E. coli* cells containing both the PUZ8002 plasmid and the conjugation plasmid. Once the *E. coli* cells have been resuspended with *P. furiosus* cells, the mixture is allowed to shake at 37° C. at 200 rpm for one hour. The cells are then plated on CC media containing Simvastatin as previously described and incubated aerobically at 37° C. for two hours to allow conjugation to occur. After the two hour incubation, the plates are transferred to anaerobic jars. Additional reductants, in the form of solid Na_2S and cysteine crystals, are added directly to the anaerobic jar as it is filled with the plates. Once the jars have been degassed and filled with an argon atmosphere, they are transferred to 90° C. incubators and allowed to grow for 40 hours.

Mutant Selection.

[0218] After incubating for 40 hours, the anaerobic jars are placed in water baths to cool to room temperature before opening. Colonies growing on plates with selection are restreaked on fresh selective plates and incubated for another 40 hours to test for stability of transformation. In concert with the restreaks, mutants are inoculated into 5 mL of liquid CC cultures with no selection to create cell stocks. Genomic DNA is isolated from the cell stocks for further analysis by PCR after examination of the restreaked selective plates to identify potential transformants demonstrating stability with new growth. To select for double crossover mutants, exconjugants demonstrating resistance to the first selection (8 μM Simvastatin) are passaged through non-selective liquid CC media and plated on media containing the second selective reagent (0.25% 5-FOA). Colonies growing on the second selection are restreaked and inoculated into liquid cultures as previously described.

DNA isolation. *Pyrococcus Furiosus* Genomic DNA Mini Prep Protocol

[0219] 1-2 mL of *P. furiosus* cell culture is pelleted at 5000 rpm for 10 minutes and resuspend in 200 μL of buffer A (25% w/v sucrose, 50 mM Tris-HCl pH 7.8, 40 mM EDTA) w/RNase A by vortexing. 250 μL of 6M guanidinium pH 8.5 is added to the pellet, mixed by gentle inversion, and allowed to sit for 5 minutes. The pellet is washed twice with 200 μL phenol/chloroform. The aqueous layers are combined and washed with 200 μL chloroform/isoamylalcohol (24:1). 20 μL of 3M sodium acetate is added and mixed by gentle inversion. 0.6 volumes of isopropanol is added and allowed to sit at −80° C. for 15 minutes after mixing by inversion. The sample is centrifuged at 14,000 rpm for 30 minutes. The supernatant is carefully removed and the pellet washed with 70% ethanol. The pellet is centrifuged at 5000 rpm for 2 minutes. The supernatant is removed and the pellet is allowed to air dry. The pellet is resuspened in 50 μL dH_2O or an appropriate buffer.

Example 3

[0220] The presence of the celA coding region in the *P. furiosus* chromosome was confirmed by PCR. Primers for PCR were designed to amplify the GDH-CelA cassette with and without a signal sequence upstream of the CelA coding region (FIG. **20**). The expected products were obtained from the *P. furiosus* exconjugants but not wild type *P. furiosus* strain (FIGS. **21** and **22**). These results indicate that the GDH-CelA construction is integrated into the *P. furiosus* chromosome. As these plasmids do not replicate in *P. furiosus* , it is expected that the cassette integrated at either the GDH or HMG locus. The plasmid also contains a GDH-HMG cassette for simvastatin selection and as both these coding regions are from *P. furiosus* they provide an area of homology for crossing over.

[0221] In addition, quantitative PCR assays (qPCR) were performed on the *P. furiosus* exconjugants to detect the presence of *A. thermophilum* CelA specific transcript. These assays detect relative transcript levels as compared to an internal standard. In this case the constitutively expressed POR transcript was used as an internal control. CelA transcript was clearly detected in the exconjugants but not in the wild type strain. Since there is no "wild type" level of CelA transcript to compare it to there is no "x-fold" level of increase in this case. The detection of the CelA transcript confirms the presence of the coding region in *P. furiosus* and indicates that it is in fact expressed at the level of transcription.

Example 4

[0222] *A. thermophilum* was grown as described in Example 1, except that the growth substrate was peanut shells (0.5%, w/v) that were used either with or without prior washing at 75° C. for 18 hours. Results are shown in FIG. **24**.

Example 5

[0223] Construction of pDCW 31, *Anaerocellum-E. coli* Shuttle Vector

[0224] The native *A. thermophilum* plasmid pAthe02 (SEQ ID No:1) has been sequenced (GenBank Accession No. CP001395, version 1, created Feb. 5, 2009) and is described in Kataeva et al. (2009), *J. Bact.,* 191(11):3760-3761. The entire 3.653 kb pAthe02 plasmid was amplified by PCR using the primers JF 197 and JF198:

```
JF197 5'-CAGCGTTAGCAAAGTGTTGT-3'    (SEQ ID NO: 2)

JF198 5'-AGCTAACGGACAGCTCAACGT-3'   (SEQ ID NO: 3)
```

[0225] A 5.601 kb fragment from the pJHW007 plasmid was amplified by PCR using the primer set JH010 and JH013:

```
                                                (SEQ ID NO: 4)
JH10 5'-AGAGAGATGCATACCAGCCTAACTTCGATCATTGGA-3'
              Nsi I

                                                (SEQ ID NO: 5)
JH13 5'-AGAGAGGGTACCAGGATCTCAAGAAGATCCTTTGAT-3'
              Kpn I
```

[0226] All PCR amplifications were performed using the High Fidelity Pfu DNA polymerase (Stratagene, La Jolla, Calif.) as described in the manufacturer's direction. The two amplified DNA fragments were treated with FAST-LINK DNA ligase (Epicentre Biotechnologies, Madison, Wis.) to construct pDCW 31 (9.356 kb) by blunt-end Ligation. The pDCW 31 plasmid includes the pSC101 origin of replication and the apramycin resistance coding regions that function in *E. coli*, and a replication origin and hygromycin resistance cassette that function in Anaerocellum. It also contains an oriT. Construction of pDCW 31 is shown in FIG. **26**.

Example 6

[0227] *Anaerocellum thermophilum* (At) Electroporation Protocol

[0228] 0.1 mL of an *Anaerocellum thermophilum* culture (approximately 2 10^8 cells per mL) is inoculated into a bottle with 50 mLs of defined At medium+uracil. Growth medium components are prepared as separate sterile stock solutions. Stock solutions are as follows: 50× salts prepared in a final volume of 1 L, 16.5 g of $MgCl_2.6H_2O$, 16.5 g of KCl, 12.5 g of NH_4Cl, 7.0 g of $CaCl_2$.2H2O; 1000× trace minerals prepared in a final volume of 1 L, 1.0 ml of HCl (25%: 7.7M), 0.5 g of Na_4EDTA tetrasodium, 2.0 g $FeCl_3.4H_2O$, 0.05 g of $ZnCl_2$, 0.05 g of $MnCl_2.4H_2O$, 0.05 g of H_3BO_3, 0.05 g of $CoCl_2.6H_2O$, 0.03 g of $CuCl_2.2H_2O$, 0.05 g of $NiCl_2.6H_2O$, 0.05 g of $(NH_4)_2Mo0_4$, 0.05 g of $AlK(SO_4).2H_2O$; 500× vitamin solution prepared in a final volume of 1 L, 0.010 g of biotin, 0.010 g of folic acid, 0.50 g of pyridoxine-HCl, 0.025 g of thiamine-HCl, 0.025 g of riboflavin (cocarboxylase), 0.025 g of nicotinic acid, 0.025 g of D-Ca-pantothenate, 0.50 g of vitamin B12, 0.025 g of p-aminobenzoic acid, 0.025 g of lipoic acid (6,8-thioctic acid); 25 amino acid solution in a final volume of 1 L, 1.9 g of L-alanine, 3.1 g of L-arginine, 2.5 g of L-asparagine, 1.2 g of L-aspartic acid, 5.0 g of L-glutamic acid, 1.2 g of L-glutamine, 5.0 g of glycine, 2.5 g of L-histidine, 2.5 g of L-isoleucine, 2.5 g of L-leucine, 2.5 g of L-lysine, 1.9 g of L-methionine, 1.9 g of L-phenylalanine, 3.1 g of L-proline, 1.9 g of L-serine, 2.5 g of L-threonine, 1.9 g of L-tryptophan, 0.3 g of L-tyrosine, 1.3 g of L-valine; 5 mg/ml resazurin sodium salt; 10% (w/v) D-(+)-cellobiose consisting of 100 g in a final volume of 1 L; 1 M KH_2PO_4, adjusted to pH 6.8 with 10 M NaOH; 0.142 M $MgSO_4.7H_2O$; 0.544 M $CaCl_2.2H_2O$; 10% (w/v) yeast extract (Difco, BD Diagnostic Systems, Sparks, Md.) consisting of 100 g in a final volume of 1 L; 10% (w/v) casein hydrolysate (enzymatic; USB Corp., Cleveland, Ohio) consisting of 100 g in a final volume of 1 L.

[0229] Each liter of defined liquid medium is composed of 20 ml of 50× salts, 2 ml of 500× vitamin mix, 1 ml of 1000× trace minerals, 40 ml of 25× amino acid solution, 50 μl of 5 mg/ml resazurin, 50 ml of 10% cellobiose, and 2.4 ml of 1 M KH_2PO_4. When complex medium is desired, 5 ml of 10% yeast extract and 50 ml of 10% casein hydrolysate is added. The medium is brought to 1 L with distilled water. To reduce the oxygen in the medium, 3 g of L-cysteine HCL, 1 g of Na_2S, and 2 g of $NaHCO_3$ is added and adjusted to pH 6.4 with 1 N NaOH at room temperature. The medium is filtered through a 0.2 μm filter, distributed into smaller bottles, and the headspace flushed with at least three times with argon. To make 1 L of solid medium, the medium is prepared the same as above except the final volume is adjusted to 500 ml, and 2.5 ml of 0.142 M $MgSO_4.7H_2O$ and 1 ml of 0.544 M $CaCl_2$. $2H_2O$ are added to aid in polymerization. The headspace of the bottle is flushed with argon and placed at 95° C. Another bottle of 500 ml of distilled water with 10 g of phytagel is autoclaved and immediately combined with the first bottle. The medium is poured into polystyrene Petri dishes and inoculated immediately after solidification. The plates are put in modified paint tanks which are flushed with four to five times with argon before incubating.

[0230] The culture is incubated at 75° C. for 16 hours. Following the incubation, the culture is centrifuged at 3500 g for 15 minutes at 23° C. The supernatant is discarded and the pelleted cells are resuspended cells in 25 mL of room temperature 10% glycerol. The cells are washed twice by repeating the centrifugation and resuspension in 10% glycerol. After the final wash, the cell pellet is resuspended in 1 mL of 10% glycerol.

[0231] 50 μL of cells are transferred to room temperature tubes for each electroporation. 30 ng of either replicating or non-replicating plasmid DNA in a total volume of 5 μL is added to each tube and mixed with the cell suspension. The cell/plasmid mixture is transferred to a 1 mm gap electroporation cuvette (to get 18 kV/cm). The cells are electroporated using an electroporator (Bio-Rad Gene Pulser, Bio-Rad Laboratories, Hercules, Calif.)) set to 1.80 V, 400 Ω resistance, 125 F capacitance, and 25 F capacitance at bottom.

[0232] The electroporated cells are transferred to 10 mL of complex medium with uracil and cytosine (described above) and incubated at 75° C. overnight. Following the overnight incubation, the cells are centrifuged at 3500 g for 15 minutes. The cell pellet is washed once by resuspension in 5 mL of 1× At salts (see above) and then recentrifuged. The washed cells are resuspended in 300 μL of 1× At salts.

[0233] The cells are plated by adding 100 μL of the cell suspension to a 4 mL tube containing 0.3% agar, then overlaying the cell/agar suspension onto either defmed medium with uracil (one plate) or defmed medium with uracil and 20 μg/mL hygromycin (two plates). The plates are placed in a jar and degassed by flushing the headspace with argon three to five times, then incubated at 75° C. for 60 hours. After 60 hours incubation, growth on plates with and without hygromycin is observed.

[0234] The efficiency of transformation is 1000 transformants per μg of replicating plasmid DNA and 100 transformants per μg of non-replicating plasmid DNA based on an average of at least three independent transformation experiments. The replicating plasmid is stably maintained after approximately 100 generations without selection.

[0235] The complete disclosure of all patents, patent applications, and publications, and electronically available material (including, for instance, nucleotide sequence submissions in, e.g., GenBank and RefSeq, and amino acid sequence submissions in, e.g., SwissProt, PIR, PRF, PDB, and translations from annotated coding regions in GenBank and RefSeq) cited herein are incorporated by reference. In the event that any inconsistency exists between the disclosure of the present application and the disclosure(s) of any document incorpo-

rated herein by reference, the disclosure of the present application shall govern. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.

[0236] All headings are for the convenience of the reader and should not be used to limit the meaning of the text that follows the heading, unless so specified.

SEQUENCE LISTING

```
<160> NUMBER OF SEQ ID NOS: 163

<210> SEQ ID NO 1
<211> LENGTH: 3653
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 1

aaaaaaatta gagcaggatt tgagcaagag gggcgctgca accccctttg aaaccgcttt     60

cataacccga aaagctcttc acccaccgca tcaatattca tacccgcttg cataattatt    120

catccctgat tgtaccacaa aacccagaaa agtcctggga aaatcatgaa gattcttcca    180

ttttccaatt tccaggaagc gaattgaaat tcccagggat tttttgcctt tttccaggga    240

atgaactgga atttccagcg gtttttccag ggacttaacc acacgaaagg tactttccgc    300

catcaaaggc ttgcaaaccc ataggaaact actggcaggc aaaaccttcc acccagtagc    360

aaatctcttc taccttgttt tgcgagcgtt agcagggagc tgtccgttag ctttgtgtta    420

gcatccttct gttagccttt ggcacgttag tagacttttg ttagttctct cttcttgttt    480

ttatcttcat tgtgattgtt ttccagaaga ctgcttgaca ttttgttatt gtagtgttat    540

gatctatttc agaagcgaaa agtaaaccgc ggtagcgggt tactttttgc gtaatatttg    600

gaattgtcag tttggcacgg gaaatggaat cgggcaggta tagggaaagg caatagccgt    660

ggacggcgtt ccgattccgc tttgtttgtg cgttagcaca tgagtcttgc tgttatcacg    720

ctctgctctg tggtagtaca ccgcaagggg tacaccattt ggagttttgc cgcatctgag    780

agtttagcag tttgcaaaac aaagttcgca gttagaaaag ttaggcaagg ggcgatgacg    840

tgcacgtctg acacgcttga cgtgcaccga gccgagagta gagcgtgatg acatagaaaa    900

agcccagtag caggttgtgc gtaccgtgac atgctactgt ggaagacaag taaatagtca    960

aaaaacggtg cgcttactgt tcgtgggctc atgggaatac cgtgagcatt ctggacaggt   1020

tcaacagcct gaaagggcaa aaccccccttt gaaagggttt aaatacacac gtccttttttg  1080

tcgtttttg tcggtttaaa tatttatgca ggggatgata gggaatggga aagccgtcca    1140

taatcaagca agtactcaat gaattcgaaa agcagataag gttcggggaa agtaagcatg   1200

aagcaaagag agaagagcgg gagagatgtg aagttactgg agagacgtgg aatcctgctc   1260

gtgtggaagg catttttagt ttctcaacct acagggagta cgttaaggag gcgttagaat   1320

ttgcgaattg ggctcgtact gaaaaggggt gtaaggattt agaacaagca cgggcctatg   1380

tgtctgaata tttgcagtcg catatagaca aagggtatag cgcgtggact gttaagaaag   1440

aagcagcagc cctggcgaaa ctgtatcatt gtcgtacaac tgactttaaa gtagagcttc   1500

ccgcaagaca cagggaagag attgagagaa gcaggggata caaagaccac gatagggagt   1560

ttagcaaaga gaggaataga gacattatca tcttttcaaa agctactggg ctgagaagaa   1620

gggaattgga aagagtgagt tctcgggata tctttcgtgg gcctgacgga agattatatg   1680

tgcacgtgag caacggcaag ggcggtagag aaagggatgt tcatgttttg cagaaatacg   1740
```

-continued

```
agagagaggt tgagaggata gtcagagagc gggaaggaag agacaggctg tttgacaggg   1800
tccccataag gatggacgta cacagctata ggagggaata tgcaagagag cgttacagag   1860
aagttgagcg tgagataagt cgtgagagaa agcttttcga cagagttgag gatctcgttc   1920
gtagtaggct tacaaggctc tatcctgaca ggtttagaga aattggcgaa agacaactta   1980
ctcgtgaact cacaagagct gatgggcttt atcatcgcag tgatggtagg gagtttgacc   2040
gcctggcatt gtgggaagtt tcaaacgact tgggacataa tcgaattgac gttgttgcaa   2100
gacactatct ggattaagcg aataaaggct caagaaagtg gataaaaaaa cagggggtat   2160
attgatatat ccccccctgtt ttttgtgcgt ctacaggacc ttatttgcgt ttcaaggcta   2220
ttcttgtggg gtagtgcagt aaaaaaggtt gcctttgtat cttgcctgta gaatcttgcc   2280
gttttcaaaa acaaattcaa cgttgcactt atcagcaagt ttcttgatta ttccctgctc   2340
aaattcgtcg aacctgtacc cttccattgc tcttgctgtg aaataaccca gaaggtacat   2400
tgcttcttca cgtgtaaact caattgttag ctttttcact ctcttttacc ccctttacga   2460
gactgagata ttctttgtac aatctgtcaa tacgttctgt atcgtcaggt tgtacgtttt   2520
tgagtttttc gtagagatta tgaaacaaca ggttctctgc tgttccaata gggtatttct   2580
ttctttccgc gtggactagc cttttcagtt ggtttagctt tacgttccgg atcttgctca   2640
gtaggtcttc gacgatattt tcaatctgtt ttgcttcgtc ttcttcgatg tagaagcata   2700
gttgtttgat ttcaagtagc ctctgaataa gttcctgctt tgttaacatt ccttgtcacc   2760
cttccccttt tcagtctttt ttattttgtc gtattctgat tcgaactttc tcagttggtc   2820
agctctgccg taacgtatgc gtcgaggatt tagcatgaat cgtccacggc ctattttaat   2880
caacaaatct ttctcgatta gctttttgat tgcaagatgg tattgctttt ttgacatgtc   2940
gatgtttcct ttgatgtcgt attcgtatga aagtgcaaac gtgtagttgt taaacgagat   3000
tttgtgttgt acaatgtatg agagaattcg cattgctgat tttgcaaggt ctttatcttc   3060
cattatgagt ttgatagaaa tgaatggaag tttgacaaag tctgtgtctt caaaatcatg   3120
tccgatcagt atgaaagttt gtacttcgcc tgtttcaggg tcgaccattt ttttgtaccc   3180
tatcttcttc atttgaatcc cccttttttt accttttttcc aattctaaag ccattataat   3240
acacttaaag taacttgtca agtaacttca ggtgtaaaaa attacacttc aggtgtaata   3300
aattacactt tccattcagt attttcaagg ctttgtgggt aactttattc ttatctatgt   3360
atatatcgcc tgcgttagca ggcttgaaaa tttccagtta ggataagcag gaacaacggt   3420
cgctgacgct gaacactgac gaaatagctg acgccccaaa gtccacaaca gtgccaaacc   3480
gataacaaaa acatgctaac gcaaacatag actaacgcac gactgacgtc gtgatgtgtg   3540
tgtgggccta cctacacaca aaaagaacta acaacagctg actaacgtct gaagagctct   3600
aacaacactt tgctaacgct gagctaacgg acagctcaac gttaacaccc gct          3653
```

<210> SEQ ID NO 2
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
Synthetic primer"

<400> SEQUENCE: 2

-continued

```
cagcgttagc aaagtgttgt                                                20


<210> SEQ ID NO 3
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 3

agctaacgga cagctcaacg t                                              21


<210> SEQ ID NO 4
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 4

agagagatgc ataccagcct aacttcgatc attgga                              36


<210> SEQ ID NO 5
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<221> NAME/KEY: source
<223> OTHER INFORMATION: /note="Description of Artificial Sequence:
      Synthetic primer"

<400> SEQUENCE: 5

agagagggta ccaggatctc aagaagatcc tttgat                              36


<210> SEQ ID NO 6
<211> LENGTH: 5280
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 6

atgaagcgtt acagaagaat tattgccatg gttgtaacct tcatatttat tttaggagtg     60

gtatatggag ttaaaccatg gcaagaggtt agggctggtt cgtttaacta tggggaagct    120

ttacaaaaag ctatcatgtt ttacgaattt caaatgtctg gtaaacttcc gaattgggta    180

cgcaacaact ggcgtggcga ctcagcatta aaggatggtc aagacaatgg gcttgatttg    240

acaggtggtt ggtttgacgc aggtgatcac gtcaagttta accttccaat gtcatacact    300

ggtacaatgt tgtcatgggc agtgtatgag tacaaagatg catttgtcaa gagtggtcaa    360

ttggaacata tcttaaatca aatcgaatgg gttaatgact attttgtaaa atgtcatcca    420

agcaaatatg tatactatta ccaggttggg gatggaagta aagatcatgc atggtgggga    480

cctgctgagg ttatgcaaat ggagagacct tcatttaagg tcacccaaag cagtcctgga    540

tctacagtag tagcagagac agcagcttcc ttagcagcag cttcaattgt tttgaaagac    600

agaaatccca ctaaagcagc aacatatctg caacatgcaa aagaattata tgagtttgca    660

gaagtaacaa aaagcgatgc aggttacact gctgcaaatg gatattacaa ttcatggagc    720
```

-continued

```
ggtttctatg atgagctttc ttgggcagca gtttggttgt atttggcaac aaatgattca    780
acatatctca caaaagctga gtcatatgtc caaaattggc ccaaaatttc tggcagtaac    840
acaattgact acaagtgggc tcattgctgg gatgatgttc acaatggagc ggcattattg    900
ttagcaaaaa ttaccggtaa ggatatttat aaacaaatta ttgaaagtca cttagattac    960
tggactacag gatacaatgg cgaaaggatt aagtatacac caaaaggatt agcatggctt   1020
gatcaatggg gttcgttgag atatgcaaca actacagcat ttttggcatt tgtttatagc   1080
gattgggttg gctgtccaag cacaaaaaaa gaaatatata gaaaatttgg agaaagccag   1140
attgattatg cgttaggctc agctggaaga agctttgttg ttggatttgg tacaaatcca   1200
ccaaagagac cgcatcacag aactgctcat agctcatggg cagacagtca gagtatacct   1260
tcatatcaca gacatacatt atatggagcg cttgttggtg gtccaggctc tgatgatagc   1320
tacacagatg atataagtaa ctatgtgaac aatgaggttg catgtgatta taatgcaggg   1380
tttgtgggtg cattagcaaa gatgtatcaa ttgtacggtg ggaatccaat accagatttc   1440
aaagctattg aaactccaac aaacgacgaa ttctttgttg aagctggtat aaatgcatcc   1500
ggaactaact ttattgaaat taaagcgata gttaataacc aaagtggttg gcctgccaga   1560
gcaacagata agcttaaatt tagatatttt gttgacctga gtgaattaat taaagcagga   1620
tattcaccaa atcaattaac cttgagcacc aattataatc aaggtgcaaa agtaagtgga   1680
ccttatgtat gggatgcaag caaaaatata tactacattt tagtagactt tactggcaca   1740
ttgatttatc caggtggtca agacaaatat aagaaagaag tccaattcag aattgcagca   1800
ccacagaatg tacagtggga taattctaac gactattctt tccaggatat aaagggagtt   1860
tcaagtggtt cagttgttaa aactaaatat attccacttt atgatggaga tgtgaaagta   1920
tggggtgaag aaccaggaac ttctggagca acaccgacac caacagcaac agcaacacca   1980
acaccaacgc cgacagtaac accaacaccg actccaacac caacatcaac tgctacacca   2040
acaccgacac caacaccgac agtaacacca accccgactc cgacaccgac tgctacacca   2100
acagcaacgc caacaccaac atcgacgccg agcagcacac ctgtagcagg tggacagata   2160
aaggtattgt atgctaacaa ggagacaaat agcacaacta atacgataag gccatggttg   2220
aaggtagtga acactggaag cagcagcata gatttgagca gggtaacgat aaggtactgg   2280
tacacggtag atggggacaa ggcacagagt gcgatatcag actgggcaca gataggagca   2340
agcaatgtga cattcaagtt tgtgaagctg agcagtagcg taagtggagc ggactattat   2400
ttagagatag gatttaagag tggagctggg cagttgcagg ctggcaaaga cacaggggag   2460
atacagataa ggtttaacaa gagtgattgg agcaattaca atcaggggaa tgactggtca   2520
tggatgcaga gcatgacgaa ttatggagag aatgtgaagg taacagcgta tatagatggt   2580
gtattggtat ggggacagga gccgagtgga gcgacaccaa caccgacagc gacaccagca   2640
ccgacagtga caccgacacc tacaccaaca ccaacgtcaa caccaactgc tacaccaaca   2700
gcaacgccaa caccaacacc gacgccgagc agcacacctg tagcaggcgg gcagataaag   2760
gtattgtatg ctaacaagga gacaaatagc acaacaaaca cgataaggcc atggttgaag   2820
gtagtgaaca ctggaagcag cagcatagat ttgagcaggg taacgataag gtactggtac   2880
acggtagatg gggacaaggc acagagtgcg atatcagact gggcacagat aggagcaagc   2940
aatgtgacat tcaagtttgt gaagctgagc agtagcgtaa gtggagcgga ctattattta   3000
```

-continued

```
gagataggat ttaagagtgg agctgggcag ttgcaggctg gtaaagacac aggggagata   3060
cagataaggt ttaacaagag tgactggagc aattacaatc aggggaatga ctggtcatgg   3120
atgcagagca tgacgaatta tggagagaat gtgaaggtaa cagcgtatat agatggtgta   3180
ttggtatggg gacaggagcc gagtggagcg acaccaacac cgacagcgac accagcaccg   3240
acagtgacac cgacacctac accagcacca actccaaccc cgacaccaac accaactgct   3300
acaccaacac caacgccaac accaacccca accgcgacac caacagtaac agcaacacca   3360
acaccgacgc cgagcagcac accgagtgtg cttggcgaat atgggcagag gtttatgtgg   3420
ttatggaaca agatacatga tcctgcgaac gggtatttta accaggatgg gataccatat   3480
cattcggtag agacattgat atgcgaagca cctgattatg gtcatttgac cacgagtgag   3540
gcattttcgt actatgtatg gttagaggca gtgtatggta agttaacggg tgactggagc   3600
aaatttaaga cagcatggga cacattagag aagtatatga taccatcagc ggaagatcag   3660
ccgatgaggt catatgatcc taacaagcca gcgacatacg caggggagtg ggagacaccg   3720
gacaagtatc catcgccgtt ggagtttaat gtacctgttg gcaaagaccc gttgcataat   3780
gaacttgtga gcacatatgg tagcacatta atgtatggta tgcactggtt gatggacgta   3840
gacaactggt atggatatgg caagagaggg gacggagtaa gtcgggcatc atttatcaac   3900
acgttccaga gagggcctga ggagtctgta tgggagacgg tgccgcatcc gagctgggag   3960
gaattcaagt ggggcggacc gaatggattt ttagatttgt ttattaagga tcagaactat   4020
tcgaagcagt ggagatatac ggatgcacca gatgctgatg cgagagctat tcaggctact   4080
tattgggcga aagtatgggc gaaggagcaa ggtaagttta atgagataag cagctatgta   4140
gcgaaggcag cgaagatggg agactattta aggtatgcga tgtttgacaa gtatttcaag   4200
ccattaggat gtcaggataa gaatgcggct ggaggaacgg ggtatgacag tgcacattat   4260
ctgctatcat ggtattatgc atggggtgga gcattggatg gagcatggtc atggaagata   4320
gggagcagcc atgtgcactt tggatatcag aatccgatgg cggcatgggc attagcgaat   4380
gatagtgata tgaagccgaa gtcgccgaat ggagcgagtg actgggcaaa gagtttgaag   4440
aggcagatag aattttacag gtggttacag tcagcggagg gagcgatagc aggaggcgcg   4500
acaaattcat ggaatggcag atatgagaag tatccagcag ggacagcaac attttatgga   4560
atggcatatg aaccgaatcc ggtatatcat gatcctggga gcaacacatg gtttggattc   4620
caggcatggt cgatgcagag ggtagcggag tattactatg tgacaggaga taaggacgca   4680
ggagcactgc ttgagaagtg ggtaagctgg gttaagagtg tagtgaagtt gaatagtgat   4740
ggtacgtttg cgataccgtc gacgcttgat tggagcggac aacctgatac atggaacggg   4800
gcgtatacag ggaatagcaa cttacatgtt aaggtagtgg actatggtac tgacttagga   4860
ataacagcgt cattggcgaa tgcgttgttg tactatagtg cagggacgaa gaagtatggg   4920
gtatttgatg agggagcgaa gaatttagcg aaggaattgc tggacaggat gtggaagttg   4980
tacagggatg agaagggatt gtcagcgcca gagaagagag cggactacaa gaggttcttt   5040
gagcaagagg tatatatacc ggcaggatgg atagggaaga tgccgaatgg agatgtaata   5100
aagagtggag ttaagtttat agacataagg agcaagtata aacaagatcc tgattggccg   5160
aagttagagg cggcatacaa gtcagggcag gcacctgagt tcagatatca caggttctgg   5220
gcacagtgcg acatagcaat agctaatgca acatatgaaa tactgtttgg caatcaataa   5280
```

-continued

<210> SEQ ID NO 7
<211> LENGTH: 1759
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 7

```
Met Lys Arg Tyr Arg Arg Ile Ile Ala Met Val Val Thr Phe Ile Phe
1               5                   10                  15

Ile Leu Gly Val Val Tyr Gly Val Lys Pro Trp Gln Glu Val Arg Ala
            20                  25                  30

Gly Ser Phe Asn Tyr Gly Glu Ala Leu Gln Lys Ala Ile Met Phe Tyr
        35                  40                  45

Glu Phe Gln Met Ser Gly Lys Leu Pro Asn Trp Val Arg Asn Asn Trp
    50                  55                  60

Arg Gly Asp Ser Ala Leu Lys Asp Gly Gln Asp Asn Gly Leu Asp Leu
65                  70                  75                  80

Thr Gly Gly Trp Phe Asp Ala Gly Asp His Val Lys Phe Asn Leu Pro
                85                  90                  95

Met Ser Tyr Thr Gly Thr Met Leu Ser Trp Ala Val Tyr Glu Tyr Lys
            100                 105                 110

Asp Ala Phe Val Lys Ser Gly Gln Leu Glu His Ile Leu Asn Gln Ile
        115                 120                 125

Glu Trp Val Asn Asp Tyr Phe Val Lys Cys His Pro Ser Lys Tyr Val
    130                 135                 140

Tyr Tyr Tyr Gln Val Gly Asp Gly Ser Lys Asp His Ala Trp Trp Gly
145                 150                 155                 160

Pro Ala Glu Val Met Gln Met Glu Arg Pro Ser Phe Lys Val Thr Gln
                165                 170                 175

Ser Ser Pro Gly Ser Thr Val Val Ala Glu Thr Ala Ala Ser Leu Ala
            180                 185                 190

Ala Ala Ser Ile Val Leu Lys Asp Arg Asn Pro Thr Lys Ala Ala Thr
        195                 200                 205

Tyr Leu Gln His Ala Lys Glu Leu Tyr Glu Phe Ala Glu Val Thr Lys
    210                 215                 220

Ser Asp Ala Gly Tyr Thr Ala Ala Asn Gly Tyr Tyr Asn Ser Trp Ser
225                 230                 235                 240

Gly Phe Tyr Asp Glu Leu Ser Trp Ala Ala Val Trp Leu Tyr Leu Ala
                245                 250                 255

Thr Asn Asp Ser Thr Tyr Leu Thr Lys Ala Glu Ser Tyr Val Gln Asn
            260                 265                 270

Trp Pro Lys Ile Ser Gly Ser Asn Thr Ile Asp Tyr Lys Trp Ala His
        275                 280                 285

Cys Trp Asp Asp Val His Asn Gly Ala Ala Leu Leu Leu Ala Lys Ile
    290                 295                 300

Thr Gly Lys Asp Ile Tyr Lys Gln Ile Ile Glu Ser His Leu Asp Tyr
305                 310                 315                 320

Trp Thr Thr Gly Tyr Asn Gly Glu Arg Ile Lys Tyr Thr Pro Lys Gly
                325                 330                 335

Leu Ala Trp Leu Asp Gln Trp Gly Ser Leu Arg Tyr Ala Thr Thr Thr
            340                 345                 350

Ala Phe Leu Ala Phe Val Tyr Ser Asp Trp Val Gly Cys Pro Ser Thr
```

-continued

```
        355                 360                 365

Lys Lys Glu Ile Tyr Arg Lys Phe Gly Glu Ser Gln Ile Asp Tyr Ala
    370                 375                 380

Leu Gly Ser Ala Gly Arg Ser Phe Val Val Gly Phe Gly Thr Asn Pro
385                 390                 395                 400

Pro Lys Arg Pro His His Arg Thr Ala His Ser Ser Trp Ala Asp Ser
                405                 410                 415

Gln Ser Ile Pro Ser Tyr His Arg His Thr Leu Tyr Gly Ala Leu Val
            420                 425                 430

Gly Gly Pro Gly Ser Asp Asp Ser Tyr Thr Asp Asp Ile Ser Asn Tyr
        435                 440                 445

Val Asn Asn Glu Val Ala Cys Asp Tyr Asn Ala Gly Phe Val Gly Ala
    450                 455                 460

Leu Ala Lys Met Tyr Gln Leu Tyr Gly Gly Asn Pro Ile Pro Asp Phe
465                 470                 475                 480

Lys Ala Ile Glu Thr Pro Thr Asn Asp Glu Phe Phe Val Glu Ala Gly
                485                 490                 495

Ile Asn Ala Ser Gly Thr Asn Phe Ile Glu Ile Lys Ala Ile Val Asn
            500                 505                 510

Asn Gln Ser Gly Trp Pro Ala Arg Ala Thr Asp Lys Leu Lys Phe Arg
        515                 520                 525

Tyr Phe Val Asp Leu Ser Glu Leu Ile Lys Ala Gly Tyr Ser Pro Asn
    530                 535                 540

Gln Leu Thr Leu Ser Thr Asn Tyr Asn Gln Gly Ala Lys Val Ser Gly
545                 550                 555                 560

Pro Tyr Val Trp Asp Ala Ser Lys Asn Ile Tyr Tyr Ile Leu Val Asp
                565                 570                 575

Phe Thr Gly Thr Leu Ile Tyr Pro Gly Gly Gln Asp Lys Tyr Lys Lys
            580                 585                 590

Glu Val Gln Phe Arg Ile Ala Ala Pro Gln Asn Val Gln Trp Asp Asn
        595                 600                 605

Ser Asn Asp Tyr Ser Phe Gln Asp Ile Lys Gly Val Ser Ser Gly Ser
    610                 615                 620

Val Val Lys Thr Lys Tyr Ile Pro Leu Tyr Asp Gly Asp Val Lys Val
625                 630                 635                 640

Trp Gly Glu Glu Pro Gly Thr Ser Gly Ala Thr Pro Thr Pro Thr Ala
                645                 650                 655

Thr Ala Thr Pro Thr Pro Thr Pro Thr Val Thr Pro Thr Pro Thr Pro
            660                 665                 670

Thr Pro Thr Ser Thr Ala Thr Pro Thr Pro Thr Pro Thr Pro Thr Val
        675                 680                 685

Thr Pro Thr Pro Thr Pro Thr Pro Thr Ala Thr Pro Thr Ala Thr Pro
    690                 695                 700

Thr Pro Thr Ser Thr Pro Ser Ser Thr Pro Val Ala Gly Gly Gln Ile
705                 710                 715                 720

Lys Val Leu Tyr Ala Asn Lys Glu Thr Asn Ser Thr Thr Asn Thr Ile
                725                 730                 735

Arg Pro Trp Leu Lys Val Val Asn Thr Gly Ser Ser Ser Ile Asp Leu
            740                 745                 750

Ser Arg Val Thr Ile Arg Tyr Trp Tyr Thr Val Asp Gly Asp Lys Ala
        755                 760                 765
```

-continued

```
Gln Ser Ala Ile Ser Asp Trp Ala Gln Ile Gly Ala Ser Asn Val Thr
    770                 775                 780

Phe Lys Phe Val Lys Leu Ser Ser Ser Val Ser Gly Ala Asp Tyr Tyr
785                 790                 795                 800

Leu Glu Ile Gly Phe Lys Ser Gly Ala Gly Gln Leu Gln Ala Gly Lys
                805                 810                 815

Asp Thr Gly Glu Ile Gln Ile Arg Phe Asn Lys Ser Asp Trp Ser Asn
            820                 825                 830

Tyr Asn Gln Gly Asn Asp Trp Ser Trp Met Gln Ser Met Thr Asn Tyr
        835                 840                 845

Gly Glu Asn Val Lys Val Thr Ala Tyr Ile Asp Gly Val Leu Val Trp
    850                 855                 860

Gly Gln Glu Pro Ser Gly Ala Thr Pro Thr Pro Thr Ala Thr Pro Ala
865                 870                 875                 880

Pro Thr Val Thr Pro Thr Pro Thr Pro Thr Pro Thr Ser Thr Pro Thr
                885                 890                 895

Ala Thr Pro Thr Ala Thr Pro Thr Pro Thr Pro Thr Pro Ser Ser Thr
            900                 905                 910

Pro Val Ala Gly Gly Gln Ile Lys Val Leu Tyr Ala Asn Lys Glu Thr
        915                 920                 925

Asn Ser Thr Thr Asn Thr Ile Arg Pro Trp Leu Lys Val Val Asn Thr
    930                 935                 940

Gly Ser Ser Ser Ile Asp Leu Ser Arg Val Thr Ile Arg Tyr Trp Tyr
945                 950                 955                 960

Thr Val Asp Gly Asp Lys Ala Gln Ser Ala Ile Ser Asp Trp Ala Gln
                965                 970                 975

Ile Gly Ala Ser Asn Val Thr Phe Lys Phe Val Lys Leu Ser Ser Ser
            980                 985                 990

Val Ser Gly Ala Asp Tyr Tyr Leu  Glu Ile Gly Phe Lys  Ser Gly Ala
        995                 1000                 1005

Gly Gln  Leu Gln Ala Gly Lys  Asp Thr Gly Glu Ile  Gln Ile Arg
    1010                 1015                 1020

Phe Asn  Lys Ser Asp Trp Ser  Asn Tyr Asn Gln Gly  Asn Asp Trp
    1025                 1030                 1035

Ser Trp  Met Gln Ser Met Thr  Asn Tyr Gly Glu Asn  Val Lys Val
    1040                 1045                 1050

Thr Ala  Tyr Ile Asp Gly Val  Leu Val Trp Gly Gln  Glu Pro Ser
    1055                 1060                 1065

Gly Ala  Thr Pro Thr Pro Thr  Ala Thr Pro Ala Pro  Thr Val Thr
    1070                 1075                 1080

Pro Thr  Pro Thr Pro Ala Pro  Thr Pro Thr Pro Thr  Pro Thr Pro
    1085                 1090                 1095

Thr Ala  Thr Pro Thr Pro Thr  Pro Thr Pro Thr Pro  Thr Ala Thr
    1100                 1105                 1110

Pro Thr  Val Thr Ala Thr Pro  Thr Pro Thr Pro Ser  Ser Thr Pro
    1115                 1120                 1125

Ser Val  Leu Gly Glu Tyr Gly  Gln Arg Phe Met Trp  Leu Trp Asn
    1130                 1135                 1140

Lys Ile  His Asp Pro Ala Asn  Gly Tyr Phe Asn Gln  Asp Gly Ile
    1145                 1150                 1155
```

-continued

```
Pro Tyr  His Ser Val Glu Thr  Leu Ile Cys Glu Ala  Pro Asp Tyr
    1160                 1165                 1170

Gly His  Leu Thr Thr Ser Glu  Ala Phe Ser Tyr Tyr  Val Trp Leu
    1175                 1180                 1185

Glu Ala  Val Tyr Gly Lys Leu  Thr Gly Asp Trp Ser  Lys Phe Lys
    1190                 1195                 1200

Thr Ala  Trp Asp Thr Leu Glu  Lys Tyr Met Ile Pro  Ser Ala Glu
    1205                 1210                 1215

Asp Gln  Pro Met Arg Ser Tyr  Asp Pro Asn Lys Pro  Ala Thr Tyr
    1220                 1225                 1230

Ala Gly  Glu Trp Glu Thr Pro  Asp Lys Tyr Pro Ser  Pro Leu Glu
    1235                 1240                 1245

Phe Asn  Val Pro Val Gly Lys  Asp Pro Leu His Asn  Glu Leu Val
    1250                 1255                 1260

Ser Thr  Tyr Gly Ser Thr Leu  Met Tyr Gly Met His  Trp Leu Met
    1265                 1270                 1275

Asp Val  Asp Asn Trp Tyr Gly  Tyr Gly Lys Arg Gly  Asp Gly Val
    1280                 1285                 1290

Ser Arg  Ala Ser Phe Ile Asn  Thr Phe Gln Arg Gly  Pro Glu Glu
    1295                 1300                 1305

Ser Val  Trp Glu Thr Val Pro  His Pro Ser Trp Glu  Glu Phe Lys
    1310                 1315                 1320

Trp Gly  Gly Pro Asn Gly Phe  Leu Asp Leu Phe Ile  Lys Asp Gln
    1325                 1330                 1335

Asn Tyr  Ser Lys Gln Trp Arg  Tyr Thr Asp Ala Pro  Asp Ala Asp
    1340                 1345                 1350

Ala Arg  Ala Ile Gln Ala Thr  Tyr Trp Ala Lys Val  Trp Ala Lys
    1355                 1360                 1365

Glu Gln  Gly Lys Phe Asn Glu  Ile Ser Ser Tyr Val  Ala Lys Ala
    1370                 1375                 1380

Ala Lys  Met Gly Asp Tyr Leu  Arg Tyr Ala Met Phe  Asp Lys Tyr
    1385                 1390                 1395

Phe Lys  Pro Leu Gly Cys Gln  Asp Lys Asn Ala Ala  Gly Gly Thr
    1400                 1405                 1410

Gly Tyr  Asp Ser Ala His Tyr  Leu Leu Ser Trp Tyr  Tyr Ala Trp
    1415                 1420                 1425

Gly Gly  Ala Leu Asp Gly Ala  Trp Ser Trp Lys Ile  Gly Ser Ser
    1430                 1435                 1440

His Val  His Phe Gly Tyr Gln  Asn Pro Met Ala Ala  Trp Ala Leu
    1445                 1450                 1455

Ala Asn  Asp Ser Asp Met Lys  Pro Lys Ser Pro Asn  Gly Ala Ser
    1460                 1465                 1470

Asp Trp  Ala Lys Ser Leu Lys  Arg Gln Ile Glu Phe  Tyr Arg Trp
    1475                 1480                 1485

Leu Gln  Ser Ala Glu Gly Ala  Ile Ala Gly Gly Ala  Thr Asn Ser
    1490                 1495                 1500

Trp Asn  Gly Arg Tyr Glu Lys  Tyr Pro Ala Gly Thr  Ala Thr Phe
    1505                 1510                 1515

Tyr Gly  Met Ala Tyr Glu Pro  Asn Pro Val Tyr His  Asp Pro Gly
    1520                 1525                 1530

Ser Asn  Thr Trp Phe Gly Phe  Gln Ala Trp Ser Met  Gln Arg Val
```

-continued

```
    1535                1540                1545

Ala Glu  Tyr Tyr Tyr Val Thr  Gly Asp Lys Asp Ala  Gly Ala Leu
    1550                1555                1560

Leu Glu  Lys Trp Val Ser Trp  Val Lys Ser Val Val  Lys Leu Asn
    1565                1570                1575

Ser Asp  Gly Thr Phe Ala Ile  Pro Ser Thr Leu Asp  Trp Ser Gly
    1580                1585                1590

Gln Pro  Asp Thr Trp Asn Gly  Ala Tyr Thr Gly Asn  Ser Asn Leu
    1595                1600                1605

His Val  Lys Val Val Asp Tyr  Gly Thr Asp Leu Gly  Ile Thr Ala
    1610                1615                1620

Ser Leu  Ala Asn Ala Leu Leu  Tyr Tyr Ser Ala Gly  Thr Lys Lys
    1625                1630                1635

Tyr Gly  Val Phe Asp Glu Gly  Ala Lys Asn Leu Ala  Lys Glu Leu
    1640                1645                1650

Leu Asp  Arg Met Trp Lys Leu  Tyr Arg Asp Glu Lys  Gly Leu Ser
    1655                1660                1665

Ala Pro  Glu Lys Arg Ala Asp  Tyr Lys Arg Phe Phe  Glu Gln Glu
    1670                1675                1680

Val Tyr  Ile Pro Ala Gly Trp  Ile Gly Lys Met Pro  Asn Gly Asp
    1685                1690                1695

Val Ile  Lys Ser Gly Val Lys  Phe Ile Asp Ile Arg  Ser Lys Tyr
    1700                1705                1710

Lys Gln  Asp Pro Asp Trp Pro  Lys Leu Glu Ala Ala  Tyr Lys Ser
    1715                1720                1725

Gly Gln  Ala Pro Glu Phe Arg  Tyr His Arg Phe Trp  Ala Gln Cys
    1730                1735                1740

Asp Ile  Ala Ile Ala Asn Ala  Thr Tyr Glu Ile Leu  Phe Gly Asn
    1745                1750                1755

Gln


<210> SEQ ID NO 8
<211> LENGTH: 945
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 8

atgagaaaac cgggtaaaat tgtaattatt ggaactggct ttgtaggctc atcaactgct      60

tttgctcttg tagatgccgg gcttgcaaca gaacttgttt taattgatgt aaaccgtgca     120

aaagccgaag gtgaagccat ggatttaaat cacggaatat cctttgtaaa acccgtcaag     180

atatgggcag gtgattatga agattgcaaa gatgctgata taataataat cactgctggt     240

gccaaccaaa agcctggtga aacaaggctt gatttgactt ataaaaatgc acaaattaca     300

aagtcgataa ttgaaaatat tatcaaatat acgcatgatg caatactttt aatggtaaca     360

aaccctgttg atgttctcac gtatgtaatg tataaagttt caggcctgcc aaaaaatcag     420

gttataggtt ctggaacagt cttagactca tcacgattta gatacctttt ggcacaacac     480

tgccaggttg atgtgagaaa tgttcatgca tatatattgg gcgaacatgg agacagtgaa     540

attgctgcct ggtctcttac aaatataggc ggcgtgaatt ttatgcagga gtgtctatta     600

tgcgggaaaa attgctcacc tgaagtaaaa gagcaaatat tcaacaaagt aaaaaatgct     660
```

-continued

```
gcatacgaaa taattgaaag aaaaggagca acatattacg ccattgcatt ggctgttaga    720

agaattgttg aagctataat cagagatgaa aattctatac tgcctgtatc atcaatagtt    780

gatgacgtat atggtgtaaa agacgttgca atttcccttc ctgcaattat caacaaaagc    840

ggagttgtaa aagtatttga tataccactg acagatgagg aaaaagaaaa gcttaaaaac    900

tctgctcagg taataaaaag tgtgatagag tctttaaaac tataa                   945


<210> SEQ ID NO 9
<211> LENGTH: 314
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 9

Met Arg Lys Pro Gly Lys Ile Val Ile Ile Gly Thr Gly Phe Val Gly
1               5                   10                  15

Ser Ser Thr Ala Phe Ala Leu Val Asp Ala Gly Leu Ala Thr Glu Leu
            20                  25                  30

Val Leu Ile Asp Val Asn Arg Ala Lys Ala Glu Gly Glu Ala Met Asp
        35                  40                  45

Leu Asn His Gly Ile Ser Phe Val Lys Pro Val Lys Ile Trp Ala Gly
    50                  55                  60

Asp Tyr Glu Asp Cys Lys Asp Ala Asp Ile Ile Ile Ile Thr Ala Gly
65                  70                  75                  80

Ala Asn Gln Lys Pro Gly Glu Thr Arg Leu Asp Leu Thr Tyr Lys Asn
                85                  90                  95

Ala Gln Ile Thr Lys Ser Ile Ile Glu Asn Ile Ile Lys Tyr Thr His
            100                 105                 110

Asp Ala Ile Leu Leu Met Val Thr Asn Pro Val Asp Val Leu Thr Tyr
        115                 120                 125

Val Met Tyr Lys Val Ser Gly Leu Pro Lys Asn Gln Val Ile Gly Ser
    130                 135                 140

Gly Thr Val Leu Asp Ser Ser Arg Phe Arg Tyr Leu Leu Ala Gln His
145                 150                 155                 160

Cys Gln Val Asp Val Arg Asn Val His Ala Tyr Ile Leu Gly Glu His
                165                 170                 175

Gly Asp Ser Glu Ile Ala Ala Trp Ser Leu Thr Asn Ile Gly Gly Val
            180                 185                 190

Asn Phe Met Gln Glu Cys Leu Leu Cys Gly Lys Asn Cys Ser Pro Glu
        195                 200                 205

Val Lys Glu Gln Ile Phe Asn Lys Val Lys Asn Ala Ala Tyr Glu Ile
    210                 215                 220

Ile Glu Arg Lys Gly Ala Thr Tyr Tyr Ala Ile Ala Leu Ala Val Arg
225                 230                 235                 240

Arg Ile Val Glu Ala Ile Ile Arg Asp Glu Asn Ser Ile Leu Pro Val
                245                 250                 255

Ser Ser Ile Val Asp Asp Val Tyr Gly Val Lys Asp Val Ala Ile Ser
            260                 265                 270

Leu Pro Ala Ile Ile Asn Lys Ser Gly Val Val Lys Val Phe Asp Ile
        275                 280                 285

Pro Leu Thr Asp Glu Glu Lys Glu Lys Leu Lys Asn Ser Ala Gln Val
    290                 295                 300

Ile Lys Ser Val Ile Glu Ser Leu Lys Leu
```

-continued

305 310

<210> SEQ ID NO 10
<211> LENGTH: 972
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 10

atgaagatac ttgtaacaag aagaataatg gaacctgcga ttgagctttt gaaaaaatat 60

ggtgaggttg aagtaaatcc acacgacaga ccaatgacaa gagaagaact tctgtcagca 120

ataaaagaca aggatgcggt tttaacccag cttgttgaca aggttgacaa agagtttttc 180

gaccatgcac caaatgtcaa gattgttgca aactatgcag tggggtacga taacatagat 240

attgaagagg caacaagaag aggtgtttat gtcacaaaca cacctgatgt tcttaccaac 300

gcaacagctg agcttgcatg ggcgctgttg tttgctgcgg caagaagaat agttgaagct 360

gacaagttca tgagaggtgg acattacaaa ggctggggac caatgctgtt tttgggcaaa 420

ggtgtgacag gtaaaacgct tggtgtgatt ggtgcaggta gaattgggca ggcttttgca 480

agaatgtcaa gagggtttaa tatgaagatt ttgtactatg actttgaaag aaaagaaagt 540

tttgaaaaag agcttggtgc ccagtatgtg ccgcttgatg aactattaaa agaagctgat 600

tttatttcaa tacatgtacc tctcacacca cagacaaggc atttaattgg tgaaagagaa 660

ttttctctca tgaaaccatc ggcaatttta attaacacag cacgtggacc aattgtagat 720

gaaaaggctt tagtcaaagc gcttaaagaa aagaagattt atgctgcagg acttgacgtg 780

tacgagagag aacctgagtt tgagccagaa ctggctgagc ttgacaatgt tgtaatgctt 840

cctcatattg gttctgcaac agaagagtcg aggcttgaca tggcaatgct tgcagcaaac 900

aatatagtag atttcattga aggaagggtt ccaagaacac ttgtcaataa agaggttttg 960

aacaagaaat aa 972

<210> SEQ ID NO 11
<211> LENGTH: 323
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 11

Met Lys Ile Leu Val Thr Arg Arg Ile Met Glu Pro Ala Ile Glu Leu
1 5 10 15

Leu Lys Lys Tyr Gly Glu Val Glu Val Asn Pro His Asp Arg Pro Met
20 25 30

Thr Arg Glu Glu Leu Leu Ser Ala Ile Lys Asp Lys Asp Ala Val Leu
35 40 45

Thr Gln Leu Val Asp Lys Val Asp Lys Glu Phe Phe Asp His Ala Pro
50 55 60

Asn Val Lys Ile Val Ala Asn Tyr Ala Val Gly Tyr Asp Asn Ile Asp
65 70 75 80

Ile Glu Glu Ala Thr Arg Arg Gly Val Tyr Val Thr Asn Thr Pro Asp
85 90 95

Val Leu Thr Asn Ala Thr Ala Glu Leu Ala Trp Ala Leu Leu Phe Ala
100 105 110

Ala Ala Arg Arg Ile Val Glu Ala Asp Lys Phe Met Arg Gly Gly His
115 120 125

-continued

```
Tyr Lys Gly Trp Gly Pro Met Leu Phe Leu Gly Lys Gly Val Thr Gly
    130                 135                 140

Lys Thr Leu Gly Val Ile Gly Ala Gly Arg Ile Gly Gln Ala Phe Ala
145                 150                 155                 160

Arg Met Ser Arg Gly Phe Asn Met Lys Ile Leu Tyr Tyr Asp Phe Glu
                165                 170                 175

Arg Lys Glu Ser Phe Glu Lys Glu Leu Gly Ala Gln Tyr Val Pro Leu
            180                 185                 190

Asp Glu Leu Leu Lys Glu Ala Asp Phe Ile Ser Ile His Val Pro Leu
        195                 200                 205

Thr Pro Gln Thr Arg His Leu Ile Gly Glu Arg Glu Phe Ser Leu Met
    210                 215                 220

Lys Pro Ser Ala Ile Leu Ile Asn Thr Ala Arg Gly Pro Ile Val Asp
225                 230                 235                 240

Glu Lys Ala Leu Val Lys Ala Leu Lys Glu Lys Lys Ile Tyr Ala Ala
                245                 250                 255

Gly Leu Asp Val Tyr Glu Arg Glu Pro Glu Phe Glu Pro Glu Leu Ala
            260                 265                 270

Glu Leu Asp Asn Val Val Met Leu Pro His Ile Gly Ser Ala Thr Glu
        275                 280                 285

Glu Ser Arg Leu Asp Met Ala Met Leu Ala Ala Asn Asn Ile Val Asp
    290                 295                 300

Phe Ile Glu Gly Arg Val Pro Arg Thr Leu Val Asn Lys Glu Val Leu
305                 310                 315                 320

Asn Lys Lys


<210> SEQ ID NO 12
<211> LENGTH: 1200
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 12

atgaaggttt tagtattaaa ttcaggaagt tcatctttaa agtatcaatt tattgatacc      60

gatacagagg ttgctctttg taaaggtgtt gttgacagga ttggtttgcc gggggcattt     120

attagacatc aaaagaatgg tcaagagatt gtaaaagaac aggaaataaa tgaccacaat     180

gttgctatta agcttgtgtt agagatgctt acacatgaag aagctggtat tatccattcc     240

atggatgaaa ttgatgcaat tggtcacagg gttgttcatg gtggggaata ttttagtgat     300

gcggtaattg tcaatgaaga ggtaaagaaa gcaataaggg aatgtattga acttgcgcct     360

cttcacaatc ctgctaattt aatggggatt gaagcatgtg aaaaagagat tcctgggaaa     420

cctaatgtgg ctgtatttga tacagcattt catcaaacaa tgccaagata tgcgtatatg     480

tattcgcttc catatgaggt gtatgaaaaa tataaaatta gaaaatatgg attccatgga     540

acatcacaca aatatgttgc aattaaagct gcagagtact taagaagacc tcttgaggag     600

ttaaaactta taacatgtca tcttggaaat ggttcgagtg tatgtgcaat aaagtacgga     660

aaatcagttg atacaagcat gggatttact cctttggctg gtcttgcaat gggaacaaga     720

agcggaacaa ttgaccctgc tgtgatactc tatcttatgg aaaaagaaaa aatggatgta     780

aagcagatga atgattttct gaataaaaag tcgggtgtgc ttggtatatc aggtgtgagc     840

agtgacttta gagatttaga aaaagctgca aatgagggta atgaaagagc acagcttgca     900
```

-continued

```
attgacatgt tctgttacag ggttaaaaag tatattggtg agtacgcagc ggtcttgggt    960

ggagtagatg caataatatt tactgctgga ataggtgaaa ataatgctct tgtgagagat   1020

aaatgtttga ctgatttaga gtatatgggt gtcctgtacg atagagaaag aaacttcaat   1080

gtagaaaaag gcaaggtttt tgaaataaac aaacctgaga gtaaggtaaa ggttttaata   1140

gttcctacaa atgaggaact tatgattgca agagagacaa aaagacttct ttcaaaataa   1200


<210> SEQ ID NO 13
<211> LENGTH: 399
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 13

Met Lys Val Leu Val Leu Asn Ser Gly Ser Ser Ser Leu Lys Tyr Gln
1               5                   10                  15

Phe Ile Asp Thr Asp Thr Glu Val Ala Leu Cys Lys Gly Val Val Asp
            20                  25                  30

Arg Ile Gly Leu Pro Gly Ala Phe Ile Arg His Gln Lys Asn Gly Gln
        35                  40                  45

Glu Ile Val Lys Glu Gln Glu Ile Asn Asp His Asn Val Ala Ile Lys
    50                  55                  60

Leu Val Leu Glu Met Leu Thr His Glu Glu Ala Gly Ile Ile His Ser
65                  70                  75                  80

Met Asp Glu Ile Asp Ala Ile Gly His Arg Val Val His Gly Gly Glu
                85                  90                  95

Tyr Phe Ser Asp Ala Val Ile Val Asn Glu Glu Val Lys Lys Ala Ile
            100                 105                 110

Arg Glu Cys Ile Glu Leu Ala Pro Leu His Asn Pro Ala Asn Leu Met
        115                 120                 125

Gly Ile Glu Ala Cys Glu Lys Glu Ile Pro Gly Lys Pro Asn Val Ala
    130                 135                 140

Val Phe Asp Thr Ala Phe His Gln Thr Met Pro Arg Tyr Ala Tyr Met
145                 150                 155                 160

Tyr Ser Leu Pro Tyr Glu Val Tyr Glu Lys Tyr Lys Ile Arg Lys Tyr
                165                 170                 175

Gly Phe His Gly Thr Ser His Lys Tyr Val Ala Ile Lys Ala Ala Glu
            180                 185                 190

Tyr Leu Arg Arg Pro Leu Glu Glu Leu Lys Leu Ile Thr Cys His Leu
        195                 200                 205

Gly Asn Gly Ser Ser Val Cys Ala Ile Lys Tyr Gly Lys Ser Val Asp
    210                 215                 220

Thr Ser Met Gly Phe Thr Pro Leu Ala Gly Leu Ala Met Gly Thr Arg
225                 230                 235                 240

Ser Gly Thr Ile Asp Pro Ala Val Ile Leu Tyr Leu Met Glu Lys Glu
                245                 250                 255

Lys Met Asp Val Lys Gln Met Asn Asp Phe Leu Asn Lys Lys Ser Gly
            260                 265                 270

Val Leu Gly Ile Ser Gly Val Ser Ser Asp Phe Arg Asp Leu Glu Lys
        275                 280                 285

Ala Ala Asn Glu Gly Asn Glu Arg Ala Gln Leu Ala Ile Asp Met Phe
    290                 295                 300

Cys Tyr Arg Val Lys Lys Tyr Ile Gly Glu Tyr Ala Ala Val Leu Gly
```

-continued

```
305                 310                 315                 320

Gly Val Asp Ala Ile Ile Phe Thr Ala Gly Ile Gly Glu Asn Asn Ala
                325                 330                 335

Leu Val Arg Asp Lys Cys Leu Thr Asp Leu Glu Tyr Met Gly Val Leu
            340                 345                 350

Tyr Asp Arg Glu Arg Asn Phe Asn Val Glu Lys Gly Lys Val Phe Glu
        355                 360                 365

Ile Asn Lys Pro Glu Ser Lys Val Lys Val Leu Ile Val Pro Thr Asn
    370                 375                 380

Glu Glu Leu Met Ile Ala Arg Glu Thr Lys Arg Leu Leu Ser Lys
385                 390                 395


<210> SEQ ID NO 14
<211> LENGTH: 999
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 14

atggcattta ttgatacaat catagaaaaa gcaaaatcag atataaaaac aattgtacta     60

cccgaaagct acgaagaaag aaatttaaag gctgcttcta tagctttgaa agaaaagata    120

gctaagatag ttttgattgg caaagaagat gagataaaaa aagaggcaac aaagtttggt    180

gcagatgtag atgaagctat ttttattgac ccagacaatt ttgatagatt tgatgaattt    240

gtaaatgaat tttatgaact aagaaaaaac aagggtgtaa cattggaaga tgcaaaaaag    300

tttatgaaag acccgatgta ttttggtgtt atgcttgtgt acaaaggttt ggcagatggt    360

atggtgtctg gtgctattca ctcaacagca gatacattaa gaccggctct gcagatatta    420

aaaactgcac ctggggtaaa acttgtttca agcttcttta ttatggttgt accaaactgc    480

gaatatggtg aaaatggagt ttttgtatat gctgatgcag gtttgaatcc aaatccaaca    540

gcagaagagc ttgctgatat agctattaca tctgcaaaga gctttgaagc tttagttggt    600

aaaactccaa aagtagcaat gctttcatat tcaaccaaag gttctgcaaa gtctgagatg    660

gttgacaagg ttgttgaggc aacaaggatt gcaaaagaga aagcaccaga cattttaata    720

gatggcgaac ttcaagcaga cgcagcaata gttccttctg ttgcaaagct gaaagcgcca    780

ggaagtcctg ttgcaggaca agcaaatgtt ctaatcttcc ctgatttgga tgctggcaac    840

attgcataca aacttacaga aaggcttgca aaagcagaag cgtacggacc tattacccag    900

ggaatagcaa aacctgtaaa tgatttgtcc cgaggttgca aagctgaaga cattgtgggg    960

gttattgcta ttactgctgt acaggctatg atgaaataa                           999


<210> SEQ ID NO 15
<211> LENGTH: 332
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 15

Met Ala Phe Ile Asp Thr Ile Ile Glu Lys Ala Lys Ser Asp Ile Lys
1               5                   10                  15

Thr Ile Val Leu Pro Glu Ser Tyr Glu Glu Arg Asn Leu Lys Ala Ala
            20                  25                  30

Ser Ile Ala Leu Lys Glu Lys Ile Ala Lys Ile Val Leu Ile Gly Lys
        35                  40                  45
```

-continued

```
Glu Asp Glu Ile Lys Lys Glu Ala Thr Lys Phe Gly Ala Asp Val Asp
    50                  55                  60

Glu Ala Ile Phe Ile Asp Pro Asp Asn Phe Asp Arg Phe Asp Glu Phe
65                  70                  75                  80

Val Asn Glu Phe Tyr Glu Leu Arg Lys Asn Lys Gly Val Thr Leu Glu
                85                  90                  95

Asp Ala Lys Lys Phe Met Lys Asp Pro Met Tyr Phe Gly Val Met Leu
            100                 105                 110

Val Tyr Lys Gly Leu Ala Asp Gly Met Val Ser Gly Ala Ile His Ser
        115                 120                 125

Thr Ala Asp Thr Leu Arg Pro Ala Leu Gln Ile Leu Lys Thr Ala Pro
    130                 135                 140

Gly Val Lys Leu Val Ser Ser Phe Phe Ile Met Val Val Pro Asn Cys
145                 150                 155                 160

Glu Tyr Gly Glu Asn Gly Val Phe Val Tyr Ala Asp Ala Gly Leu Asn
                165                 170                 175

Pro Asn Pro Thr Ala Glu Glu Leu Ala Asp Ile Ala Ile Thr Ser Ala
            180                 185                 190

Lys Ser Phe Glu Ala Leu Val Gly Lys Thr Pro Lys Val Ala Met Leu
        195                 200                 205

Ser Tyr Ser Thr Lys Gly Ser Ala Lys Ser Glu Met Val Asp Lys Val
    210                 215                 220

Val Glu Ala Thr Arg Ile Ala Lys Glu Lys Ala Pro Asp Ile Leu Ile
225                 230                 235                 240

Asp Gly Glu Leu Gln Ala Asp Ala Ala Ile Val Pro Ser Val Ala Lys
                245                 250                 255

Leu Lys Ala Pro Gly Ser Pro Val Ala Gly Gln Ala Asn Val Leu Ile
            260                 265                 270

Phe Pro Asp Leu Asp Ala Gly Asn Ile Ala Tyr Lys Leu Thr Glu Arg
        275                 280                 285

Leu Ala Lys Ala Glu Ala Tyr Gly Pro Ile Thr Gln Gly Ile Ala Lys
    290                 295                 300

Pro Val Asn Asp Leu Ser Arg Gly Cys Lys Ala Glu Asp Ile Val Gly
305                 310                 315                 320

Val Ile Ala Ile Thr Ala Val Gln Ala Met Met Lys
                325                 330
```

<210> SEQ ID NO 16
<211> LENGTH: 1641
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 16

```
atgaaatttt taactgcaga tgtattaaaa gatatgttaa aagctgcaaa taattattta      60

aaattgcaca tagataagat aaactcatta aacgtctttc cagtaccaga tggtgacaca     120

ggcaccaata tgtctgccac tcttgacagc agcataaaag aaataaatgg aaagactttc     180

gaaaatgtgg acaaacttat gaatgcagtt gcgtttggca gcttaaaagg tgcacgcggt     240

aattctggtg ttattctttc tcagctttta cgcggatttg ccaaagagct aaaaggcaaa     300

gatgttatag atataccaac atttgttgct tgtttaaaat ctgcgtctgc aagtgcttac     360

aaagcagtga tgaagcctac agaaggcact atgctcacag ttgcacgcgg gattgcagag     420
```

-continued

```
gatgttgaaa aagaagtggc agaaggcatt gtgagtgaaa tagaggattt gctggaagtg    480

tgtgtttcaa gcgggaagaa gtggcttgca aagacaccag agatgctttc tattttaaaa    540

gaggcaaatg tagttgacag tggcggtatg ggtcttgtaa taatttttga agggatgtat    600

aaattcttaa aagaaggaat ggtatttgaa gagccatcac agcaggaagt ttatacagtc    660

ctcactttta aacctgaaga tattaagttt acttactgta cagagttttt tattaccggt    720

ttgaaaaaga atattgaaaa agaatttaag gaatatcttg agacaattgg tgattcaatt    780

attgtaatcc aagatggcga cattctcaaa acacacgttc acacaaattc acctggcaag    840

gtaatagaaa aggctttgaa atatggtgag cttataaata taaagattga taatatgaaa    900

tatcagcacc aggagtttat aagtaaaaga gaaaaccatg agacagaact tcaaacacag    960

gctgaagtta ttataaaaga atatggtttt gtagctgtat cacaaggaga aggattcaat   1020

gaaatattaa aaggcttggg tgttgatttt gtaattgaag gcggacagac tatgaatcca   1080

agcgctgagg actttgtaaa tgctataaag aatgtaccag ccaaaaatgt atttattttc   1140

ccgaacaata aaaacgtgat tatgtcagca gagctttctt tacagcttat taatacaaat   1200

aaaaatatag tgattatgaa gacaaccaat attcctgagt gcattactgc aatgataaag   1260

tttgatttga acaagagtat tgaagaaaat ataaagctca tgcagcaagc tataaactca   1320

gtaaaggttg tagaaataac taaggcagtg agaaatacaa aaataaacgg gtttgagatt   1380

gaagaaggcg attttatagg gatttccaaa aaggaaataa ttgcatgtga caaagatatg   1440

ttaaaagtag ctttggcttg tgtcgaaaag attgttgatt ctacaaccca gattttgagt   1500

atttactatg gcaaaggtgt agccttagaa gatatagagg tgcttgttaa aaacatacaa   1560

gaaatatacc cgaaaattga cattgagagc tatgaaagtg gaaatgaaat ttatcaatta   1620

attattgtag ctgagatgtg a                                             1641
```

```
<210> SEQ ID NO 17
<211> LENGTH: 546
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 17

Met Lys Phe Leu Thr Ala Asp Val Leu Lys Asp Met Leu Lys Ala Ala
1               5                   10                  15

Asn Asn Tyr Leu Lys Leu His Ile Asp Lys Ile Asn Ser Leu Asn Val
            20                  25                  30

Phe Pro Val Pro Asp Gly Asp Thr Gly Thr Asn Met Ser Ala Thr Leu
        35                  40                  45

Asp Ser Ser Ile Lys Glu Ile Asn Gly Lys Thr Phe Glu Asn Val Asp
    50                  55                  60

Lys Leu Met Asn Ala Val Ala Phe Gly Ser Leu Lys Gly Ala Arg Gly
65                  70                  75                  80

Asn Ser Gly Val Ile Leu Ser Gln Leu Leu Arg Gly Phe Ala Lys Glu
                85                  90                  95

Leu Lys Gly Lys Asp Val Ile Asp Ile Pro Thr Phe Val Ala Cys Leu
            100                 105                 110

Lys Ser Ala Ser Ala Ser Ala Tyr Lys Ala Val Met Lys Pro Thr Glu
        115                 120                 125

Gly Thr Met Leu Thr Val Ala Arg Gly Ile Ala Glu Asp Val Glu Lys
    130                 135                 140
```

-continued

```
Glu Val Ala Glu Gly Ile Val Ser Glu Ile Glu Asp Leu Leu Glu Val
145                 150                 155                 160

Cys Val Ser Ser Gly Lys Lys Trp Leu Ala Lys Thr Pro Glu Met Leu
                165                 170                 175

Ser Ile Leu Lys Glu Ala Asn Val Val Asp Ser Gly Gly Met Gly Leu
            180                 185                 190

Val Ile Ile Phe Glu Gly Met Tyr Lys Phe Leu Lys Glu Gly Met Val
        195                 200                 205

Phe Glu Glu Pro Ser Gln Gln Glu Val Tyr Thr Val Leu Thr Phe Lys
    210                 215                 220

Pro Glu Asp Ile Lys Phe Thr Tyr Cys Thr Glu Phe Phe Ile Thr Gly
225                 230                 235                 240

Leu Lys Lys Asn Ile Glu Lys Glu Phe Lys Glu Tyr Leu Glu Thr Ile
                245                 250                 255

Gly Asp Ser Ile Ile Val Ile Gln Asp Gly Asp Ile Leu Lys Thr His
            260                 265                 270

Val His Thr Asn Ser Pro Gly Lys Val Ile Glu Lys Ala Leu Lys Tyr
        275                 280                 285

Gly Glu Leu Ile Asn Ile Lys Ile Asp Asn Met Lys Tyr Gln His Gln
    290                 295                 300

Glu Phe Ile Ser Lys Arg Glu Asn His Glu Thr Glu Leu Gln Thr Gln
305                 310                 315                 320

Ala Glu Val Ile Ile Lys Glu Tyr Gly Phe Val Ala Val Ser Gln Gly
                325                 330                 335

Glu Gly Phe Asn Glu Ile Leu Lys Gly Leu Gly Val Asp Phe Val Ile
            340                 345                 350

Glu Gly Gly Gln Thr Met Asn Pro Ser Ala Glu Asp Phe Val Asn Ala
        355                 360                 365

Ile Lys Asn Val Pro Ala Lys Asn Val Phe Ile Phe Pro Asn Asn Lys
    370                 375                 380

Asn Val Ile Met Ser Ala Glu Leu Ser Leu Gln Leu Ile Asn Thr Asn
385                 390                 395                 400

Lys Asn Ile Val Ile Met Lys Thr Thr Asn Ile Pro Glu Cys Ile Thr
                405                 410                 415

Ala Met Ile Lys Phe Asp Leu Asn Lys Ser Ile Glu Glu Asn Ile Lys
            420                 425                 430

Leu Met Gln Gln Ala Ile Asn Ser Val Lys Val Val Glu Ile Thr Lys
        435                 440                 445

Ala Val Arg Asn Thr Lys Ile Asn Gly Phe Glu Ile Glu Glu Gly Asp
    450                 455                 460

Phe Ile Gly Ile Ser Lys Lys Glu Ile Ile Ala Cys Asp Lys Asp Met
465                 470                 475                 480

Leu Lys Val Ala Leu Ala Cys Val Glu Lys Ile Val Asp Ser Thr Thr
                485                 490                 495

Gln Ile Leu Ser Ile Tyr Tyr Gly Lys Gly Val Ala Leu Glu Asp Ile
            500                 505                 510

Glu Val Leu Val Lys Asn Ile Gln Glu Ile Tyr Pro Lys Ile Asp Ile
        515                 520                 525

Glu Ser Tyr Glu Ser Gly Asn Glu Ile Tyr Gln Leu Ile Ile Val Ala
    530                 535                 540
```

-continued

Glu Met
545

<210> SEQ ID NO 18
<211> LENGTH: 120
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 18

atgaaagtta agaaaaagaa gataggtagt gttattatga tttttgcaat attgttgagc 60

gttattgttc cagttttaac ggctacatat caaagtaaaa gttatttaga aactacataa 120

<210> SEQ ID NO 19
<211> LENGTH: 39
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 19

Met Lys Val Lys Lys Lys Lys Ile Gly Ser Val Ile Met Ile Phe Ala
1 5 10 15

Ile Leu Leu Ser Val Ile Val Pro Val Leu Thr Ala Thr Tyr Gln Ser
20 25 30

Lys Ser Tyr Leu Glu Thr Thr
35

<210> SEQ ID NO 20
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 20

atggttttta aaagtgataa aaacttaatt tacttggtta caggaaaagt gaagaaaaat 60

gaccttggca tgggaattta catttgcgaa taa 93

<210> SEQ ID NO 21
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 21

Met Val Phe Lys Ser Asp Lys Asn Leu Ile Tyr Leu Val Thr Gly Lys
1 5 10 15

Val Lys Lys Asn Asp Leu Gly Met Gly Ile Tyr Ile Cys Glu
20 25 30

<210> SEQ ID NO 22
<211> LENGTH: 9084
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 22

gtgagatttc caaagaaatt gagagtattg ttaacatttc ttattctttt aatcttcacc 60

cttaatacaa atctatatgt cttttcacaa agtagtaatc tcaagaacca agagtcaaat 120

tacaaagatt taataggaca ctgggcagag gatgaattta gatggttgat agagaaggga 180

ataattagtg gagtgaaggc aaatggagta atgttagcaa agccagatga gaagttaaca 240

agagcggaag ctgttgtgat tatactaaga actatatttg agaaggagct tcttgaggag 300

-continued

```
gaaataaaga aattaaaaga ggatagtttt aaagatattt caaatcattg gagcagagat    360
tacataaatg tagctgcgag atatggtttg gttaaaggtt attcagacaa gacatttaga    420
ccaaatcaga gaattacacg tgaagagttt gttttaatgg tgataagggt gagcaaatat    480
agggagcaag cagcaagtga gggcaagcaa aacgataaga aggtaaaatt taaagatgtg    540
agtgttaaca attttggata taacgaaata atttttgcag ttgagaaagg cttgatcaaa    600
ggctatagcg atggtacatt tagaccaaga agttatataa gccgagcaga agcagcagta    660
atagttgcaa gggcattgaa agcggattta ttcattacct atagagcaaa gaagcctcaa    720
ataactaaag aaaatgcgag cgaatttttt gaacttgttg ttgataaaga ggaagtaggt    780
ataatagaaa aggcgaagct cacactaaag agcaaagtaa agggtgtcaa gtttacagca    840
gagtggaagg caagtggtgg taaacttgag gtagctaaag ataatcagag tgcgatttgg    900
agcccgacag atgcagaaga gggtaagaat tatattgtga ctgttgaagt aaccatcctg    960
acagataatg gggaaaagat aaaaatacaa aaggttgcaa ggatacgagt gaacgagggc   1020
gtccaactga caaagaatga aaacatagaa gaaccggtga ataaaataga tccattattt   1080
acccaaagag aatcaaacaa tttagttcaa gaaaattcta gttatagccc attagaagaa   1140
aaactttcag cgccaatatt gaaaatagca aaaaaaggtg agtatgtaga gcttacttgg   1200
aatgaggtaa aaggtgcaag gagttacatt gtaaaaagag ggaaggtaag cggggagtat   1260
gaggtattgg caagcggagt tattactacg gcgtatattg atggtcctat ggatggaaaa   1320
acaacgtatt actatgtggt tgcagcagtg ggagcaaatg gtacgaccag tatgaattca   1380
aatgaggtag tgtacaaagc actgccacaa tcaccaaagt tatttgggac atacaatggg   1440
acaagtgcac gtcttacatg gacgaaagcg aatggagcag agaggtatac tctttacagg   1500
agcacagtca gtggaggacc atactatgca attgcagaaa acttacttac caatgtgttt   1560
gaagatacta atttgacatc aggtacagtg tattactatg tagtaaaagc cattaacgaa   1620
ttaggagaaa gtgaatactc aaatgaggta gcgatggggg aaggagttac agttgcagct   1680
aaatttaatc ctgatgacga cgaggatgga gatgggttga caaatgtaga tgagttaaaa   1740
tatgcaacaa gtttgaagaa gaatgatagt gatggagatg ggttgagtga cggttatgaa   1800
gtaaagaaag gaacaaatcc attagtaccg gatacagata atgacgggat atatgatgga   1860
gcagaggttg tgatgggaac agacccattg acaaagaatc ctttgacaag tgccgagaaa   1920
tatgcggtat cggaagatgg taaagttttt gtaagagcac tttctgatgc aaatatttta   1980
atagctccct tacaggtaaa gaggtctgac aatgtgttta ttaactcatt gaaaggaata   2040
gtaggtaaag cgatagagat aactgcaggg ggatttgata taaagaaggc agagatagta   2100
gtgaactatg atgaagcaga acttaatggt gtagaagaga ataatttaat gctgtattat   2160
gtcaattatg acaagaagat attagagccg ttagaagatg ttgtagtaga tacagtgtat   2220
aatcgagttt cgggtaagac agagcatttt agcacatttt tacttggtga taagaatatg   2280
ccagttgatc tttccaaagt agatattgtg tttgtacttg acaactcagg tagtatgtcg   2340
tctaatgacc caaattatta cagaatagaa gcaacaaaga aatttataca aaatatagat   2400
gaacttaaca atagagttgg gttagtagat ttcgatagct ctgtaagtgt tagatcaaat   2460
ttgacatctg acaagagcaa attgttacaa gccctaaatg cgatgagatg gacaggtggt   2520
tctacgaata taggaggagg attgaaagct gcgctgggat tattcgacca agagcaatct   2580
```

-continued

```
aaaaagataa tagtactttt atcagacgga tatcacaaca caggaattca tccaaacgat   2640
gtactaccag aattaataaa acaagaaata gtagtaaata ccatcgcatt gggtaaagat   2700
tgtgataggg agttattaca tgatatagct gataagacaa aaggaggcta tttttacgtt   2760
gataatacag gaggactttc tcaagaagat gtagacaagc aaatagagct aatatacgag   2820
aaactaacta aatggataac ccttcaaaag gaagcagaaa agaatctcaa acctcaagaa   2880
gtgttaagta tagagtacaa tgatgtaggt ttagacaatg aagagtttca taaatggata   2940
accacagcta tgactaattt acttacggga aattatatgg aagagtttga tgatattagt   3000
atagagggga atggtcctga gtttaaattt gtgaggtatt acaattcttt tgctaatcaa   3060
caaaagacga taataggtaa aggttttaga accaattttg acagcaagct tacaaaggta   3120
gcgggagtcg gaatagtcca agcaggggta ttgaatgtaa gagaaggacc gaatgtaaat   3180
accaaaaaga tagggtattt gactaagggg acaaaggtgc agattgaaga ggatggcaac   3240
aagaacggca gtggttggca cagaattgtg tataagggta aatctgcata tatttgcgca   3300
gcatatgtca aagagctaaa caatggaatt gaagtaacat atccaagtgg tagtatgata   3360
gtatttgtgg atgataatgg agatgggatt tacttgtcag acagtaacaa agctgacaag   3420
gtagttttgt caggtaatga atatatattg atacaaagag atttgacgag atatgttttt   3480
gacaagagtg gaaaacttat caaaattggt gatagaaatg aaaattacat tactattgag   3540
tattctggtg aaaagatatc tgcggcaaga gatgtttttg gcagaaaatt agaatttatc   3600
tttcaagggg ataacctggt atgtataaga gagaacatca agggtaaaat aggtaggaaa   3660
gtagaatttg tatatgatga caaggataga ctaataaaag taattggggt tgatggcgca   3720
gagacaagat atgaatatga tgagaaagac agactcaaga ggataattga tgcgaacggg   3780
caccaggtgg taaggaatga atatgatatt ttaggtagga tagtgaggca atacgacggg   3840
gaagatataa tcagattttt catatacgat gatgaagaca gagtaagata ttatatagac   3900
gaaaatggta atgagagtat ggtagtattc aatgaagagc taaaaccaat aaaagagaga   3960
aatgcattgg gtgggggttg tgactacaag tatgagatta atgacggttc aaagtggata   4020
gacgtgacaa ctcctgattt ggataaggat gtagttgtga gcggtcttac acgggaaaaa   4080
tatcaagagc tgaaagagaa aggtagcatg acaaagcagg tcacaattca gataataaag   4140
acttctccta gtttagaaac agcaaagact acacaaatat acgatggacg tggcaatatt   4200
atccaagtga tagatgctta tggtaattca ataaaaatga aatatgataa caataataat   4260
ttaattgagc aaacagacag gattggtgct actaccaaat atatttatga tgccgagggt   4320
ataaatttaa ttgagaaaat tgatccactt ggaaacaagg aaagatatga atattattct   4380
ataaattcgg gcataaaact caatggttta cttgcaaagt atatagataa aaatgggaat   4440
gagacgagat actattatga agatgaatat aacaatttaa cacgagtggt tgacgcagaa   4500
ggttatgaga caaaatatga atatgaccag ttaggcagaa aaataaaaga aataaatgaa   4560
agaggatatg taacaaggtt tgagtatgat ccagaaggca gaattacaaa agaaatagac   4620
gcgtttggta aaacaaaagt atatgtatat gacaaagtag gtaatttaat agaagaaatc   4680
gatcgccttg gcaacaaaac aagatatgta tacgatgata aaaacaggct tataaaagaa   4740
attgatgcga tgggtggaga gtatcaatac ttttatgatc ctgtagggaa caagataaag   4800
gagatagacc ctgaaggaag agtaaccaaa tatacttatg atgaactgaa cagacttgtt   4860
```

-continued

```
gagatagaag atgcagaagg caataaaacc aaattcaagt atgatttagc aggcaggaaa   4920
atatcagagg taaatgcctt aggaaaagag acaagatatg aatatgattt attaggaaga   4980
cttacgaaag tgattgatcc tttaggtaaa acccgaagtt accaatacaa tgctgaaggg   5040
tataaaatat cagagacgaa taaaaatggt gccactactt catatgtata tgatttagca   5100
ggaaggttaa taacagcata ttatccagat ggaactagaa ggagctataa ctacgataat   5160
aatggcaatg taatcagtat catcaatcca aaaggatatg tgacaaaata ctactacgac   5220
aaattaaaca gggtaatcaa ggtagaagac agcaacggca aagcagtaac atatgagtat   5280
gacggttgtg gaaatgtaat ttgctttaaa gacaaaaaag gaagagaaac aaggtacgaa   5340
tatgatgctc ttgacagagt aaaaagagta atagcaccaa acggagcaca aacagagtat   5400
gagtacgatg cagagggcag agttgtcaag gtcacagatg caaaaggaag aagcgaggaa   5460
tatatgtatg atgagcttgg tagagttgtt gtatataagg acaagttagg aaatgtaatc   5520
aaatacgcat atgacaaggt gggtaataga acacagctta ttgatagaag aggaaatgca   5580
acaaagtatg agtatgataa attgaatagg gtagtaaagg tgattgacgc gtatggaaat   5640
gaaagtaggc tggagtatga tgctgttggg aataatatag caaagacaga cagaagagga   5700
aacacaacaa agtatgaata tgatgcaaat aacagactag tgacaataat tgatccatat   5760
ggaaataaga ttagatttga gtatgatgga gaaggcaatg taatatgcag aatagatgca   5820
agagggaaca gaatgtatta tagttatgat ggattgaaca ggttgaggac tgtacaagac   5880
aacgatggca ggaaaactgt ttatgaatat gatgagaatg ggaatatagt aaagataata   5940
cgaccagacg gcaagtatgt aacctacagg tatgacagtt tggacaggtt ggtaagagtt   6000
acacaagaga atggagctgt gacagagtat aggtatgacg aagaggataa tttaattgag   6060
gtcaaggatg ggaatggcaa tataacaagg tatgaataca atgagataga cagacctgtc   6120
aaagtaatag atgcaatcgg caatgaagaa agatatagct atgacttagt agggaatata   6180
gtatatgcga tagacaagaa tggagttagg atagagtaca gctatgacca acttgacagg   6240
gtagtgcatg ttaaagctgg cggagtagag gtgagatata gttatgatga ggaaggcaac   6300
agagtccaga tgtcagacaa aacagggata aacaaatatg aatatgataa gctaaatagg   6360
ttgataagga aaatataccc agatggcaag agtatagaat atgaatacga ccaagaaggt   6420
aatgtaataa aggtaaaaga cccgagcgga tatgtgacgc aatacaagta tgataagatg   6480
aatcggcttg aagaggtgat aacctcagat ggtagcacaa gatattcata cgatgagaat   6540
ggaaatgtga aatcaataga gtatccgaat aagctgaagt ttgagtacga atatgatagc   6600
aggaacttac ttaaaggatt agtggttaca gcgagagatg ttgtaaggaa tgagatagat   6660
aagtattata ctcctagcag tgtgatagaa caaggtggtg gcacaagcac atatatttac   6720
aagtatgaat atggatatga tgacaatggc aatatgatat acaaacagga gcctataggt   6780
aggacagagt acaagtatga tgaggtaggt agagtagctg aagtcaaaga tagatttgga   6840
cgattaactc gatatgaata tgacaatgta ggcaatagaa taaaggagat agtagaggtc   6900
ccgacaggta taagcagaga tagtcttaat tctgaagggc tgaatattag atacgattat   6960
ggagaagtat acagcgttga gaagacttac acatatgata aattaaatag actattgagc   7020
attgaatctc gagacagaga agggaatatt gtaggtgtta acagctacgg ctatgacaac   7080
aatgggaatt tgataagagc agaggaaaga tggcgcacaa gagtatacaa gggtcaggga   7140
```

-continued

```
agcaaagtag ttgcaaaaga gatagaaaag caagatttac agactgtata tcaagatgca   7200
tatactgcag agaatgttca ggtacagaca agcgtatata actcagcata tagtttggac   7260
aatagcacag tagtgatttc agtatatgat agtgtgtaca atagtgcata caatttacaa   7320
aattcaagtt ctgagcagag tattactgta aaagacgtag tttacgagac ggtatatgaa   7380
gatgtttatg agatagaaga gaagattaaa gtaagtgaat ttagatatga tgagcttggc   7440
aggatggtat gggcaaaggt tgacaataat gtggtagaat ttgaatacga cggagatggg   7500
ctcagaacaa agaagataac agcaaatgat gtaaaaacat actactggtc tggcagcaat   7560
cttatttatg aaagcgatgc aacgggcaaa ggattcagca gcatatgggg cttgagtatg   7620
ataggaagga cggacggtag taatacagag tattttatga aagatgggca tggagatgta   7680
cttatcacgt ttgacaaaag tgggcagagg aagaatatat atgaatatga tttgtatggt   7740
aatgtaataa aagaaataga gacagggcaa gagaacccga taagatatgc tggatattac   7800
tatgacaagg aattagaatg gtattattta aagacaaggt attacgatag taggataggg   7860
aggtttgtga aagaggatgg cataaaggga gatataacag acccagaaag tttgaaccta   7920
tacacttact gtacaaacaa tcctgtgaat ttatatgatc cggatggaga gtttgcgata   7980
gtgccactgt tagtaggatt agcagtgcag actgtaagtg gagtattgtt agattatata   8040
atagatagga aaaatttcaa tctttggaag agtattggaa caaatttagt agtgtcaatg   8100
gttggtatag taacaggagg aatagcaagt tcaataaagc ttggtacaaa aataactaaa   8160
gttgcaccga aagtagcaaa actaacagaa agattagtaa gatggggaag tgaaaaagct   8220
gttaaaaaac ctaactctaa ggttgcaaaa tttataaata aatttacaga agatgttgca   8280
aaacctttta ttgataaatt aggcactcca aagacatggg ttatggctgg gttaaaagca   8340
ctgacccaaa caggtgcgga tgcgacgata gatatgatat ctggtgaaaa agttaccgca   8400
ggatcaattt ctctcgatat acttttgaga actgcatttg gtgcgtttgg ctcagcaaca   8460
ggaaacgtga aactatttgg aggtggattg aggagaatat ttaaattacc tccctctaag   8520
aaagcggaaa gttgggatgc tgctgctaaa agagtagttg caaaatattt gatagaagat   8580
ttagaggcgg gattatcaaa ttttggcaaa aagtgttcac ttgatgcatt aaagggtatt   8640
aacttaggaa ggatattggg acgagcagga agaaagttta caaaatatat agtaaataaa   8700
ttggaaatgc aaaagtattc aataacaaat gtttcaggga tccaagggga aagtggtgga   8760
ggagtattag acgagtttat tgttcaaaat aataaaaaaa caggcgagga tgtaaaatgg   8820
acaagaggaa gtggagtaag tagtccagcg acaattggta tctggggcac tgcaggtgaa   8880
catgaggttg gacaaataaa aggagtagca gcacaaggat tgaattttag ctggagcata   8940
caatcaaaga tgggagctga tttaaactta ggggcggtaa gaggaggcga gccgaggacg   9000
ccgacagtag atttaaaatt cggctatatg gcacctatga gaaggatgat agtgccgagg   9060
gtaggcaggg tattagcata ttga                                          9084
```

<210> SEQ ID NO 23
<211> LENGTH: 3027
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 23

```
Val Arg Phe Pro Lys Lys Leu Arg Val Leu Leu Thr Phe Leu Ile Leu
1               5                   10                  15
```

-continued

```
Leu Ile Phe Thr Leu Asn Thr Asn Leu Tyr Val Phe Ser Gln Ser Ser
            20                  25                  30

Asn Leu Lys Asn Gln Glu Ser Asn Tyr Lys Asp Leu Ile Gly His Trp
        35                  40                  45

Ala Glu Asp Glu Phe Arg Trp Leu Ile Glu Lys Gly Ile Ile Ser Gly
    50                  55                  60

Val Lys Ala Asn Gly Val Met Leu Ala Lys Pro Asp Glu Lys Leu Thr
65                  70                  75                  80

Arg Ala Glu Ala Val Val Ile Ile Leu Arg Thr Ile Phe Glu Lys Glu
                85                  90                  95

Leu Leu Glu Glu Glu Ile Lys Lys Leu Lys Glu Asp Ser Phe Lys Asp
            100                 105                 110

Ile Ser Asn His Trp Ser Arg Asp Tyr Ile Asn Val Ala Ala Arg Tyr
        115                 120                 125

Gly Leu Val Lys Gly Tyr Ser Asp Lys Thr Phe Arg Pro Asn Gln Arg
    130                 135                 140

Ile Thr Arg Glu Glu Phe Val Leu Met Val Ile Arg Val Ser Lys Tyr
145                 150                 155                 160

Arg Glu Gln Ala Ala Ser Glu Gly Lys Gln Asn Asp Lys Lys Val Lys
                165                 170                 175

Phe Lys Asp Val Ser Val Asn Asn Phe Gly Tyr Asn Glu Ile Ile Phe
            180                 185                 190

Ala Val Glu Lys Gly Leu Ile Lys Gly Tyr Ser Asp Gly Thr Phe Arg
        195                 200                 205

Pro Arg Ser Tyr Ile Ser Arg Ala Glu Ala Ala Val Ile Val Ala Arg
    210                 215                 220

Ala Leu Lys Ala Asp Leu Phe Ile Thr Tyr Arg Ala Lys Lys Pro Gln
225                 230                 235                 240

Ile Thr Lys Glu Asn Ala Ser Glu Phe Phe Glu Leu Val Val Asp Lys
                245                 250                 255

Glu Glu Val Gly Ile Ile Glu Lys Ala Lys Leu Thr Leu Lys Ser Lys
            260                 265                 270

Val Lys Gly Val Lys Phe Thr Ala Glu Trp Lys Ala Ser Gly Gly Lys
        275                 280                 285

Leu Glu Val Ala Lys Asp Asn Gln Ser Ala Ile Trp Ser Pro Thr Asp
    290                 295                 300

Ala Glu Glu Gly Lys Asn Tyr Ile Val Thr Val Glu Val Thr Ile Leu
305                 310                 315                 320

Thr Asp Asn Gly Glu Lys Ile Lys Ile Gln Lys Val Ala Arg Ile Arg
                325                 330                 335

Val Asn Glu Gly Val Gln Leu Thr Lys Asn Glu Asn Ile Glu Glu Pro
            340                 345                 350

Val Asn Lys Ile Asp Pro Leu Phe Thr Gln Arg Glu Ser Asn Asn Leu
        355                 360                 365

Val Gln Glu Asn Ser Ser Tyr Ser Pro Leu Glu Glu Lys Leu Ser Ala
    370                 375                 380

Pro Ile Leu Lys Ile Ala Lys Lys Gly Glu Tyr Val Glu Leu Thr Trp
385                 390                 395                 400

Asn Glu Val Lys Gly Ala Arg Ser Tyr Ile Val Lys Arg Gly Lys Val
                405                 410                 415
```

-continued

```
Ser Gly Glu Tyr Glu Val Leu Ala Ser Gly Val Ile Thr Thr Ala Tyr
            420                 425                 430

Ile Asp Gly Pro Met Asp Gly Lys Thr Thr Tyr Tyr Tyr Val Val Ala
        435                 440                 445

Ala Val Gly Ala Asn Gly Thr Thr Ser Met Asn Ser Asn Glu Val Val
    450                 455                 460

Tyr Lys Ala Leu Pro Gln Ser Pro Lys Leu Phe Gly Thr Tyr Asn Gly
465                 470                 475                 480

Thr Ser Ala Arg Leu Thr Trp Thr Lys Ala Asn Gly Ala Glu Arg Tyr
                485                 490                 495

Thr Leu Tyr Arg Ser Thr Val Ser Gly Gly Pro Tyr Tyr Ala Ile Ala
            500                 505                 510

Glu Asn Leu Leu Thr Asn Val Phe Glu Asp Thr Asn Leu Thr Ser Gly
        515                 520                 525

Thr Val Tyr Tyr Tyr Val Val Lys Ala Ile Asn Glu Leu Gly Glu Ser
    530                 535                 540

Glu Tyr Ser Asn Glu Val Ala Met Gly Glu Gly Val Thr Val Ala Ala
545                 550                 555                 560

Lys Phe Asn Pro Asp Asp Asp Glu Asp Gly Asp Gly Leu Thr Asn Val
                565                 570                 575

Asp Glu Leu Lys Tyr Ala Thr Ser Leu Lys Lys Asn Asp Ser Asp Gly
            580                 585                 590

Asp Gly Leu Ser Asp Gly Tyr Glu Val Lys Lys Gly Thr Asn Pro Leu
        595                 600                 605

Val Pro Asp Thr Asp Asn Asp Gly Ile Tyr Asp Gly Ala Glu Val Val
    610                 615                 620

Met Gly Thr Asp Pro Leu Thr Lys Asn Pro Leu Thr Ser Ala Glu Lys
625                 630                 635                 640

Tyr Ala Val Ser Glu Asp Gly Lys Val Phe Val Arg Ala Leu Ser Asp
                645                 650                 655

Ala Asn Ile Leu Ile Ala Pro Leu Gln Val Lys Arg Ser Asp Asn Val
            660                 665                 670

Phe Ile Asn Ser Leu Lys Gly Ile Val Gly Lys Ala Ile Glu Ile Thr
        675                 680                 685

Ala Gly Gly Phe Asp Ile Lys Lys Ala Glu Ile Val Val Asn Tyr Asp
    690                 695                 700

Glu Ala Glu Leu Asn Gly Val Glu Glu Asn Asn Leu Met Leu Tyr Tyr
705                 710                 715                 720

Val Asn Tyr Asp Lys Lys Ile Leu Glu Pro Leu Glu Asp Val Val Val
                725                 730                 735

Asp Thr Val Tyr Asn Arg Val Ser Gly Lys Thr Glu His Phe Ser Thr
            740                 745                 750

Phe Leu Leu Gly Asp Lys Asn Met Pro Val Asp Leu Ser Lys Val Asp
        755                 760                 765

Ile Val Phe Val Leu Asp Asn Ser Gly Ser Met Ser Ser Asn Asp Pro
    770                 775                 780

Asn Tyr Tyr Arg Ile Glu Ala Thr Lys Lys Phe Ile Gln Asn Ile Asp
785                 790                 795                 800

Glu Leu Asn Asn Arg Val Gly Leu Val Asp Phe Asp Ser Ser Val Ser
                805                 810                 815

Val Arg Ser Asn Leu Thr Ser Asp Lys Ser Lys Leu Leu Gln Ala Leu
```

-continued

```
            820                 825                 830

Asn Ala Met Arg Trp Thr Gly Gly Ser Thr Asn Ile Gly Gly Gly Leu
        835                 840                 845

Lys Ala Ala Leu Gly Leu Phe Asp Gln Glu Gln Ser Lys Lys Ile Ile
    850                 855                 860

Val Leu Leu Ser Asp Gly Tyr His Asn Thr Gly Ile His Pro Asn Asp
865                 870                 875                 880

Val Leu Pro Glu Leu Ile Lys Gln Glu Ile Val Val Asn Thr Ile Ala
                885                 890                 895

Leu Gly Lys Asp Cys Asp Arg Glu Leu Leu His Asp Ile Ala Asp Lys
            900                 905                 910

Thr Lys Gly Gly Tyr Phe Tyr Val Asp Asn Thr Gly Gly Leu Ser Gln
        915                 920                 925

Glu Asp Val Asp Lys Gln Ile Glu Leu Ile Tyr Glu Lys Leu Thr Lys
    930                 935                 940

Trp Ile Thr Leu Gln Lys Glu Ala Glu Lys Asn Leu Lys Pro Gln Glu
945                 950                 955                 960

Val Leu Ser Ile Glu Tyr Asn Asp Val Gly Leu Asp Asn Glu Glu Phe
                965                 970                 975

His Lys Trp Ile Thr Thr Ala Met Thr Asn Leu Leu Thr Gly Asn Tyr
            980                 985                 990

Met Glu Glu Phe Asp Asp Ile Ser  Ile Glu Gly Asn Gly  Pro Glu Phe
        995                 1000                 1005

Lys Phe  Val Arg Tyr Tyr Asn  Ser Phe Ala Asn Gln  Gln Lys Thr
    1010                 1015                 1020

Ile Ile  Gly Lys Gly Phe Arg  Thr Asn Phe Asp Ser  Lys Leu Thr
    1025                 1030                 1035

Lys Val  Ala Gly Val Gly Ile  Val Gln Ala Gly Val  Leu Asn Val
    1040                 1045                 1050

Arg Glu  Gly Pro Asn Val Asn  Thr Lys Lys Ile Gly  Tyr Leu Thr
    1055                 1060                 1065

Lys Gly  Thr Lys Val Gln Ile  Glu Glu Asp Gly Asn  Lys Asn Gly
    1070                 1075                 1080

Ser Gly  Trp His Arg Ile Val  Tyr Lys Gly Lys Ser  Ala Tyr Ile
    1085                 1090                 1095

Cys Ala  Ala Tyr Val Lys Glu  Leu Asn Asn Gly Ile  Glu Val Thr
    1100                 1105                 1110

Tyr Pro  Ser Gly Ser Met Ile  Val Phe Val Asp Asp  Asn Gly Asp
    1115                 1120                 1125

Gly Ile  Tyr Leu Ser Asp Ser  Asn Lys Ala Asp Lys  Val Val Leu
    1130                 1135                 1140

Ser Gly  Asn Glu Tyr Ile Leu  Ile Gln Arg Asp Leu  Thr Arg Tyr
    1145                 1150                 1155

Val Phe  Asp Lys Ser Gly Lys  Leu Ile Lys Ile Gly  Asp Arg Asn
    1160                 1165                 1170

Glu Asn  Tyr Ile Thr Ile Glu  Tyr Ser Gly Glu Lys  Ile Ser Ala
    1175                 1180                 1185

Ala Arg  Asp Val Phe Gly Arg  Lys Leu Glu Phe Ile  Phe Gln Gly
    1190                 1195                 1200

Asp Asn  Leu Val Cys Ile Arg  Glu Asn Ile Lys Gly  Lys Ile Gly
    1205                 1210                 1215
```

-continued

```
Arg Lys  Val Glu Phe Val Tyr  Asp Asp Lys Asp Arg  Leu Ile Lys
    1220                 1225                 1230

Val Ile  Gly Val Asp Gly Ala  Glu Thr Arg Tyr Glu  Tyr Asp Glu
    1235                 1240                 1245

Lys Asp  Arg Leu Lys Arg Ile  Ile Asp Ala Asn Gly  His Gln Val
    1250                 1255                 1260

Val Arg  Asn Glu Tyr Asp Ile  Leu Gly Arg Ile Val  Arg Gln Tyr
    1265                 1270                 1275

Asp Gly  Glu Asp Ile Ile Arg  Phe Phe Ile Tyr Asp  Asp Glu Asp
    1280                 1285                 1290

Arg Val  Arg Tyr Tyr Ile Asp  Glu Asn Gly Asn Glu  Ser Met Val
    1295                 1300                 1305

Val Phe  Asn Glu Glu Leu Lys  Pro Ile Lys Glu Arg  Asn Ala Leu
    1310                 1315                 1320

Gly Gly  Gly Cys Asp Tyr Lys  Tyr Glu Ile Asn Asp  Gly Ser Lys
    1325                 1330                 1335

Trp Ile  Asp Val Thr Thr Pro  Asp Leu Asp Lys Asp  Val Val Val
    1340                 1345                 1350

Ser Gly  Leu Thr Arg Glu Lys  Tyr Gln Glu Leu Lys  Glu Lys Gly
    1355                 1360                 1365

Ser Met  Thr Lys Gln Val Thr  Ile Gln Ile Ile Lys  Thr Ser Pro
    1370                 1375                 1380

Ser Leu  Glu Thr Ala Lys Thr  Thr Gln Ile Tyr Asp  Gly Arg Gly
    1385                 1390                 1395

Asn Ile  Ile Gln Val Ile Asp  Ala Tyr Gly Asn Ser  Ile Lys Met
    1400                 1405                 1410

Lys Tyr  Asp Asn Asn Asn Asn  Leu Ile Glu Gln Thr  Asp Arg Ile
    1415                 1420                 1425

Gly Ala  Thr Thr Lys Tyr Ile  Tyr Asp Ala Glu Gly  Ile Asn Leu
    1430                 1435                 1440

Ile Glu  Lys Ile Asp Pro Leu  Gly Asn Lys Glu Arg  Tyr Glu Tyr
    1445                 1450                 1455

Tyr Ser  Ile Asn Ser Gly Ile  Lys Leu Asn Gly Leu  Leu Ala Lys
    1460                 1465                 1470

Tyr Ile  Asp Lys Asn Gly Asn  Glu Thr Arg Tyr Tyr  Tyr Glu Asp
    1475                 1480                 1485

Glu Tyr  Asn Asn Leu Thr Arg  Val Val Asp Ala Glu  Gly Tyr Glu
    1490                 1495                 1500

Thr Lys  Tyr Glu Tyr Asp Gln  Leu Gly Arg Lys Ile  Lys Glu Ile
    1505                 1510                 1515

Asn Glu  Arg Gly Tyr Val Thr  Arg Phe Glu Tyr Asp  Pro Glu Gly
    1520                 1525                 1530

Arg Ile  Thr Lys Glu Ile Asp  Ala Phe Gly Lys Thr  Lys Val Tyr
    1535                 1540                 1545

Val Tyr  Asp Lys Val Gly Asn  Leu Ile Glu Glu Ile  Asp Arg Leu
    1550                 1555                 1560

Gly Asn  Lys Thr Arg Tyr Val  Tyr Asp Asp Lys Asn  Arg Leu Ile
    1565                 1570                 1575

Lys Glu  Ile Asp Ala Met Gly  Gly Glu Tyr Gln Tyr  Phe Tyr Asp
    1580                 1585                 1590
```

-continued

```
Pro Val  Gly Asn Lys Ile Lys  Glu Ile Asp Pro Glu  Gly Arg Val
    1595                 1600                 1605

Thr Lys  Tyr Thr Tyr Asp Glu  Leu Asn Arg Leu Val  Glu Ile Glu
    1610                 1615                 1620

Asp Ala  Glu Gly Asn Lys Thr  Lys Phe Lys Tyr Asp  Leu Ala Gly
    1625                 1630                 1635

Arg Lys  Ile Ser Glu Val Asn  Ala Leu Gly Lys Glu  Thr Arg Tyr
    1640                 1645                 1650

Glu Tyr  Asp Leu Leu Gly Arg  Leu Thr Lys Val Ile  Asp Pro Leu
    1655                 1660                 1665

Gly Lys  Thr Arg Ser Tyr Gln  Tyr Asn Ala Glu Gly  Tyr Lys Ile
    1670                 1675                 1680

Ser Glu  Thr Asn Lys Asn Gly  Ala Thr Thr Ser Tyr  Val Tyr Asp
    1685                 1690                 1695

Leu Ala  Gly Arg Leu Ile Thr  Ala Tyr Tyr Pro Asp  Gly Thr Arg
    1700                 1705                 1710

Arg Ser  Tyr Asn Tyr Asp Asn  Asn Gly Asn Val Ile  Ser Ile Ile
    1715                 1720                 1725

Asn Pro  Lys Gly Tyr Val Thr  Lys Tyr Tyr Tyr Asp  Lys Leu Asn
    1730                 1735                 1740

Arg Val  Ile Lys Val Glu Asp  Ser Asn Gly Lys Ala  Val Thr Tyr
    1745                 1750                 1755

Glu Tyr  Asp Gly Cys Gly Asn  Val Ile Cys Phe Lys  Asp Lys Lys
    1760                 1765                 1770

Gly Arg  Glu Thr Arg Tyr Glu  Tyr Asp Ala Leu Asp  Arg Val Lys
    1775                 1780                 1785

Arg Val  Ile Ala Pro Asn Gly  Ala Gln Thr Glu Tyr  Glu Tyr Asp
    1790                 1795                 1800

Ala Glu  Gly Arg Val Val Lys  Val Thr Asp Ala Lys  Gly Arg Ser
    1805                 1810                 1815

Glu Glu  Tyr Met Tyr Asp Glu  Leu Gly Arg Val Val  Val Tyr Lys
    1820                 1825                 1830

Asp Lys  Leu Gly Asn Val Ile  Lys Tyr Ala Tyr Asp  Lys Val Gly
    1835                 1840                 1845

Asn Arg  Thr Gln Leu Ile Asp  Arg Arg Gly Asn Ala  Thr Lys Tyr
    1850                 1855                 1860

Glu Tyr  Asp Lys Leu Asn Arg  Val Val Lys Val Ile  Asp Ala Tyr
    1865                 1870                 1875

Gly Asn  Glu Ser Arg Leu Glu  Tyr Asp Ala Val Gly  Asn Asn Ile
    1880                 1885                 1890

Ala Lys  Thr Asp Arg Arg Gly  Asn Thr Thr Lys Tyr  Glu Tyr Asp
    1895                 1900                 1905

Ala Asn  Asn Arg Leu Val Thr  Ile Ile Asp Pro Tyr  Gly Asn Lys
    1910                 1915                 1920

Ile Arg  Phe Glu Tyr Asp Gly  Glu Gly Asn Val Ile  Cys Arg Ile
    1925                 1930                 1935

Asp Ala  Arg Gly Asn Arg Met  Tyr Tyr Ser Tyr Asp  Gly Leu Asn
    1940                 1945                 1950

Arg Leu  Arg Thr Val Gln Asp  Asn Asp Gly Arg Lys  Thr Val Tyr
    1955                 1960                 1965

Glu Tyr  Asp Glu Asn Gly Asn  Ile Val Lys Ile Ile  Arg Pro Asp
```

-continued

```
    1970                1975                1980

Gly Lys  Tyr Val Thr Tyr Arg  Tyr Asp Ser Leu Asp  Arg Leu Val
    1985                1990                1995

Arg Val  Thr Gln Glu Asn Gly  Ala Val Thr Glu Tyr  Arg Tyr Asp
    2000                2005                2010

Glu Glu  Asp Asn Leu Ile Glu  Val Lys Asp Gly Asn  Gly Asn Ile
    2015                2020                2025

Thr Arg  Tyr Glu Tyr Asn Glu  Ile Asp Arg Pro Val  Lys Val Ile
    2030                2035                2040

Asp Ala  Ile Gly Asn Glu Glu  Arg Tyr Ser Tyr Asp  Leu Val Gly
    2045                2050                2055

Asn Ile  Val Tyr Ala Ile Asp  Lys Asn Gly Val Arg  Ile Glu Tyr
    2060                2065                2070

Ser Tyr  Asp Gln Leu Asp Arg  Val Val His Val Lys  Ala Gly Gly
    2075                2080                2085

Val Glu  Val Arg Tyr Ser Tyr  Asp Glu Glu Gly Asn  Arg Val Gln
    2090                2095                2100

Met Ser  Asp Lys Thr Gly Ile  Asn Lys Tyr Glu Tyr  Asp Lys Leu
    2105                2110                2115

Asn Arg  Leu Ile Arg Lys Ile  Tyr Pro Asp Gly Lys  Ser Ile Glu
    2120                2125                2130

Tyr Glu  Tyr Asp Gln Glu Gly  Asn Val Ile Lys Val  Lys Asp Pro
    2135                2140                2145

Ser Gly  Tyr Val Thr Gln Tyr  Lys Tyr Asp Lys Met  Asn Arg Leu
    2150                2155                2160

Glu Glu  Val Ile Thr Ser Asp  Gly Ser Thr Arg Tyr  Ser Tyr Asp
    2165                2170                2175

Glu Asn  Gly Asn Val Lys Ser  Ile Glu Tyr Pro Asn  Lys Leu Lys
    2180                2185                2190

Phe Glu  Tyr Glu Tyr Asp Ser  Arg Asn Leu Leu Lys  Gly Leu Val
    2195                2200                2205

Val Thr  Ala Arg Asp Val Val  Arg Asn Glu Ile Asp  Lys Tyr Tyr
    2210                2215                2220

Thr Pro  Ser Ser Val Ile Glu  Gln Gly Gly Gly Thr  Ser Thr Tyr
    2225                2230                2235

Ile Tyr  Lys Tyr Glu Tyr Gly  Tyr Asp Asp Asn Gly  Asn Met Ile
    2240                2245                2250

Tyr Lys  Gln Glu Pro Ile Gly  Arg Thr Glu Tyr Lys  Tyr Asp Glu
    2255                2260                2265

Val Gly  Arg Val Ala Glu Val  Lys Asp Arg Phe Gly  Arg Leu Thr
    2270                2275                2280

Arg Tyr  Glu Tyr Asp Asn Val  Gly Asn Arg Ile Lys  Glu Ile Val
    2285                2290                2295

Glu Val  Pro Thr Gly Ile Ser  Arg Asp Ser Leu Asn  Ser Glu Gly
    2300                2305                2310

Leu Asn  Ile Arg Tyr Asp Tyr  Gly Glu Val Tyr Ser  Val Glu Lys
    2315                2320                2325

Thr Tyr  Thr Tyr Asp Lys Leu  Asn Arg Leu Leu Ser  Ile Glu Ser
    2330                2335                2340

Arg Asp  Arg Glu Gly Asn Ile  Val Gly Val Asn Ser  Tyr Gly Tyr
    2345                2350                2355
```

-continued

```
Asp Asn  Asn Gly Asn Leu Ile  Arg Ala Glu Glu Arg  Trp Arg Thr
    2360                 2365                 2370

Arg Val  Tyr Lys Gly Gln Gly  Ser Lys Val Val Ala  Lys Glu Ile
    2375                 2380                 2385

Glu Lys  Gln Asp Leu Gln Thr  Val Tyr Gln Asp Ala  Tyr Thr Ala
    2390                 2395                 2400

Glu Asn  Val Gln Val Gln Thr  Ser Val Tyr Asn Ser  Ala Tyr Ser
    2405                 2410                 2415

Leu Asp  Asn Ser Thr Val Val  Ile Ser Val Tyr Asp  Ser Val Tyr
    2420                 2425                 2430

Asn Ser  Ala Tyr Asn Leu Gln  Asn Ser Ser Ser Glu  Gln Ser Ile
    2435                 2440                 2445

Thr Val  Lys Asp Val Val Tyr  Glu Thr Val Tyr Glu  Asp Val Tyr
    2450                 2455                 2460

Glu Ile  Glu Glu Lys Ile Lys  Val Ser Glu Phe Arg  Tyr Asp Glu
    2465                 2470                 2475

Leu Gly  Arg Met Val Trp Ala  Lys Val Asp Asn Asn  Val Val Glu
    2480                 2485                 2490

Phe Glu  Tyr Asp Gly Asp Gly  Leu Arg Thr Lys Lys  Ile Thr Ala
    2495                 2500                 2505

Asn Asp  Val Lys Thr Tyr Tyr  Trp Ser Gly Ser Asn  Leu Ile Tyr
    2510                 2515                 2520

Glu Ser  Asp Ala Thr Gly Lys  Gly Phe Ser Ser Ile  Trp Gly Leu
    2525                 2530                 2535

Ser Met  Ile Gly Arg Thr Asp  Gly Ser Asn Thr Glu  Tyr Phe Met
    2540                 2545                 2550

Lys Asp  Gly His Gly Asp Val  Leu Ile Thr Phe Asp  Lys Ser Gly
    2555                 2560                 2565

Gln Arg  Lys Asn Ile Tyr Glu  Tyr Asp Leu Tyr Gly  Asn Val Ile
    2570                 2575                 2580

Lys Glu  Ile Glu Thr Gly Gln  Glu Asn Pro Ile Arg  Tyr Ala Gly
    2585                 2590                 2595

Tyr Tyr  Tyr Asp Lys Glu Leu  Glu Trp Tyr Tyr Leu  Lys Thr Arg
    2600                 2605                 2610

Tyr Tyr  Asp Ser Arg Ile Gly  Arg Phe Val Lys Glu  Asp Gly Ile
    2615                 2620                 2625

Lys Gly  Asp Ile Thr Asp Pro  Glu Ser Leu Asn Leu  Tyr Thr Tyr
    2630                 2635                 2640

Cys Thr  Asn Asn Pro Val Asn  Leu Tyr Asp Pro Asp  Gly Glu Phe
    2645                 2650                 2655

Ala Ile  Val Pro Leu Leu Val  Gly Leu Ala Val Gln  Thr Val Ser
    2660                 2665                 2670

Gly Val  Leu Leu Asp Tyr Ile  Ile Asp Arg Lys Asn  Phe Asn Leu
    2675                 2680                 2685

Trp Lys  Ser Ile Gly Thr Asn  Leu Val Val Ser Met  Val Gly Ile
    2690                 2695                 2700

Val Thr  Gly Gly Ile Ala Ser  Ser Ile Lys Leu Gly  Thr Lys Ile
    2705                 2710                 2715

Thr Lys  Val Ala Pro Lys Val  Ala Lys Leu Thr Glu  Arg Leu Val
    2720                 2725                 2730
```

-continued

```
Arg Trp  Gly Ser Glu Lys Ala  Val Lys Lys Pro Asn  Ser Lys Val
    2735                 2740                 2745

Ala Lys  Phe Ile Asn Lys Phe  Thr Glu Asp Val Ala  Lys Pro Phe
    2750                 2755                 2760

Ile Asp  Lys Leu Gly Thr Pro  Lys Thr Trp Val Met  Ala Gly Leu
    2765                 2770                 2775

Lys Ala  Leu Thr Gln Thr Gly  Ala Asp Ala Thr Ile  Asp Met Ile
    2780                 2785                 2790

Ser Gly  Glu Lys Val Thr Ala  Gly Ser Ile Ser Leu  Asp Ile Leu
    2795                 2800                 2805

Leu Arg  Thr Ala Phe Gly Ala  Phe Gly Ser Ala Thr  Gly Asn Val
    2810                 2815                 2820

Lys Leu  Phe Gly Gly Gly Leu  Arg Arg Ile Phe Lys  Leu Pro Pro
    2825                 2830                 2835

Ser Lys  Lys Ala Glu Ser Trp  Asp Ala Ala Ala Lys  Arg Val Val
    2840                 2845                 2850

Ala Lys  Tyr Leu Ile Glu Asp  Leu Glu Ala Gly Leu  Ser Asn Phe
    2855                 2860                 2865

Gly Lys  Lys Cys Ser Leu Asp  Ala Leu Lys Gly Ile  Asn Leu Gly
    2870                 2875                 2880

Arg Ile  Leu Gly Arg Ala Gly  Arg Lys Phe Thr Lys  Tyr Ile Val
    2885                 2890                 2895

Asn Lys  Leu Glu Met Gln Lys  Tyr Ser Ile Thr Asn  Val Ser Gly
    2900                 2905                 2910

Ile Gln  Gly Glu Ser Gly Gly  Gly Val Leu Asp Glu  Phe Ile Val
    2915                 2920                 2925

Gln Asn  Asn Lys Lys Thr Gly  Glu Asp Val Lys Trp  Thr Arg Gly
    2930                 2935                 2940

Ser Gly  Val Ser Ser Pro Ala  Thr Ile Gly Ile Trp  Gly Thr Ala
    2945                 2950                 2955

Gly Glu  His Glu Val Gly Gln  Ile Lys Gly Val Ala  Ala Gln Gly
    2960                 2965                 2970

Leu Asn  Phe Ser Trp Ser Ile  Gln Ser Lys Met Gly  Ala Asp Leu
    2975                 2980                 2985

Asn Leu  Gly Ala Val Arg Gly  Gly Glu Pro Arg Thr  Pro Thr Val
    2990                 2995                 3000

Asp Leu  Lys Phe Gly Tyr Met  Ala Pro Met Arg Arg  Met Ile Val
    3005                 3010                 3015

Pro Arg  Val Gly Arg Val Leu  Ala Tyr
    3020                 3025
```

<210> SEQ ID NO 24
<211> LENGTH: 1248
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 24

```
atggagagta gagagagggt agtgaaagcg ctgaataagc taataaaggc attggcagta      60

atttggttag tgttaattgt tctgaggctt atatggtatt ttgcagacat aacttatata     120

tttgagagat tatcgataga tacgacagat gttctatcat tgttcttgga acaaacaata     180

ggttcttttt tggcttatag tgtggctata ttgattgtag ctttatttac tgtagtgtta     240
```

-continued

```
tttttttat taagttcaat aattacggtg ttgattgctg gtttcacaac aacaatttca    300

gaaaagattg aagcaatttt tgtcaagaag gaatttggtg cgatgcaggt aaatgaagaa    360

gtaattgaag ctagtttaga tatagatgaa gagatttata ttgataggag taagaggaaa    420

ataagcgtat ttggcagata tgaaaaggcc aaaaaaacgg attcactttt aggagcggag    480

cacaggattg atggagagtt ttttattttt aagtgtatta aagtgaaagg agaaaaatta    540

tatttaccgt actcattaat tgaaaataga tatgtaaacg gatatgttgt aggtaagaat    600

atatcatact actgtgatgg caatgaatca gaaagaatgt tattggtaga cgcaattaaa    660

gaagaaataa tagaaaatga aaaagagcat gttgaagaaa aaatttgtgg tgaaattcta    720

gcagaaaaat gtgctaaagc tgctgcaaag ttaattaaga gtataacaga agaaagtttt    780

gatactgaac caagatttaa ggatatagta ctaacgttat tagttacagg ggtattttta    840

ataacaaatt ttgtgtcttg gttgatatct aagatagcaa gtagtgagat aaatttgcaa    900

ggaatattat ggattataaa tataatgatt ttagtgttaa tattagtact ttacgtagtt    960

aagagatttt atgaaaaaga aaaaccgctt aagtttgata tagagataga taaggtacat   1020

gtaatatttt gtgaaagttc tgaaggaata atagaaacag taataaaaga aaaagaagca   1080

tttgtaaaag atgaatacat agtgttagta aaaaaaggag atgaagaagg cgagagattt   1140

ttggggaaat taaaagaaga aaaagaaaaa ggattgtatg aagtaagatt aatatatggt   1200

accaaaaaag agaacaaaat aaagtattgg gtaaaattaa ggagttga                1248
```

<210> SEQ ID NO 25
<211> LENGTH: 415
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 25

```
Met Glu Ser Arg Glu Arg Val Val Lys Ala Leu Asn Lys Leu Ile Lys
1               5                   10                  15

Ala Leu Ala Val Ile Trp Leu Val Leu Ile Val Leu Arg Leu Ile Trp
            20                  25                  30

Tyr Phe Ala Asp Ile Thr Tyr Ile Phe Glu Arg Leu Ser Ile Asp Thr
        35                  40                  45

Thr Asp Val Leu Ser Leu Phe Leu Glu Gln Thr Ile Gly Ser Phe Leu
    50                  55                  60

Ala Tyr Ser Val Ala Ile Leu Ile Val Ala Leu Phe Thr Val Val Leu
65                  70                  75                  80

Phe Phe Leu Leu Ser Ser Ile Ile Thr Val Leu Ile Ala Gly Phe Thr
                85                  90                  95

Thr Thr Ile Ser Glu Lys Ile Glu Ala Ile Phe Val Lys Lys Glu Phe
            100                 105                 110

Gly Ala Met Gln Val Asn Glu Glu Val Ile Glu Ala Ser Leu Asp Ile
        115                 120                 125

Asp Glu Glu Ile Tyr Ile Asp Arg Ser Lys Arg Lys Ile Ser Val Phe
    130                 135                 140

Gly Arg Tyr Glu Lys Ala Lys Lys Thr Asp Ser Leu Leu Gly Ala Glu
145                 150                 155                 160

His Arg Ile Asp Gly Glu Phe Phe Ile Phe Lys Cys Ile Lys Val Lys
                165                 170                 175

Gly Glu Lys Leu Tyr Leu Pro Tyr Ser Leu Ile Glu Asn Arg Tyr Val
```

-continued

```
            180                 185                 190

Asn Gly Tyr Val Val Gly Lys Asn Ile Ser Tyr Tyr Cys Asp Gly Asn
        195                 200                 205

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

Glu Asn Glu Lys Glu His Val Glu Glu Lys Ile Cys Gly Glu Ile Leu
225                 230                 235                 240

Ala Glu Lys Cys Ala Lys Ala Ala Ala Lys Leu Ile Lys Ser Ile Thr
                245                 250                 255

Glu Glu Ser Phe Asp Thr Glu Pro Arg Phe Lys Asp Ile Val Leu Thr
            260                 265                 270

Leu Leu Val Thr Gly Val Phe Leu Ile Thr Asn Phe Val Ser Trp Leu
        275                 280                 285

Ile Ser Lys Ile Ala Ser Ser Glu Ile Asn Leu Gln Gly Ile Leu Trp
    290                 295                 300

Ile Ile Asn Ile Met Ile Leu Val Leu Ile Leu Val Leu Tyr Val Val
305                 310                 315                 320

Lys Arg Phe Tyr Glu Lys Glu Lys Pro Leu Lys Phe Asp Ile Glu Ile
                325                 330                 335

Asp Lys Val His Val Ile Phe Cys Glu Ser Ser Glu Gly Ile Ile Glu
            340                 345                 350

Thr Val Ile Lys Glu Lys Glu Ala Phe Val Lys Asp Glu Tyr Ile Val
        355                 360                 365

Leu Val Lys Lys Gly Asp Glu Glu Gly Glu Arg Phe Leu Gly Lys Leu
    370                 375                 380

Lys Glu Glu Lys Glu Lys Gly Leu Tyr Glu Val Arg Leu Ile Tyr Gly
385                 390                 395                 400

Thr Lys Lys Glu Asn Lys Ile Lys Tyr Trp Val Lys Leu Arg Ser
                405                 410                 415


<210> SEQ ID NO 26
<211> LENGTH: 783
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 26

atgagcaaga tattagtgaa aggtgtaaaa aatccaatat ataacagatt gcgtattaag      60

tcatatttgg taagtttcaa agttttgata gtagttgtta ctataattat tgctttttg     120

cttattaact atgcaactga ttgggttaat agaattggat tggatgtaga tgaaaaaaca     180

aggtcgttta taaatacatt ggcttttagt ttatcgctag atatcttatc cccatgctct     240

atcgatgaca ctattttggg ataccctatt acagaagcta ttttcacacc aataatatac     300

atggttactt ggtttaaggg tgatagcaaa gaaaaaatag aatatttttc aagaaataaa     360

tcagtaatag aactgctgtt gagcaagctt caatatgttt taatggacaa ctctccagag     420

gaaaatttta tgaattgggt gaaatcgaaa gatatagatt taaaaaaata tgaaaaggaa     480

gatattttaa cgttttattg ggcattccac tttttagatg tcaatgtaat agtagtgata     540

ttcagtactg tagtgttgat aggatgtatt ggagcaccag taagtgtatt agtatttaca     600

ttaatattca gcatatctac gttaatagag gatataagag atttggtctt taagagaaga     660

gaagcaattt taaggggaga ggaaggaatt gaagatgaag atgaaaggaa gatatatgag     720
```

-continued

```
aaaataaaag aagagataaa aatagcaatt tataaggaag aagaagcaag aaatcaaagt    780

tag                                                                  783


<210> SEQ ID NO 27
<211> LENGTH: 260
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 27

Met Ser Lys Ile Leu Val Lys Gly Val Lys Asn Pro Ile Tyr Asn Arg
1               5                   10                  15

Leu Arg Ile Lys Ser Tyr Leu Val Ser Phe Lys Val Leu Ile Val Val
            20                  25                  30

Val Thr Ile Ile Ile Ala Phe Leu Leu Ile Asn Tyr Ala Thr Asp Trp
        35                  40                  45

Val Asn Arg Ile Gly Leu Asp Val Asp Glu Lys Thr Arg Ser Phe Ile
    50                  55                  60

Asn Thr Leu Ala Phe Ser Leu Ser Leu Asp Ile Leu Ser Pro Cys Ser
65                  70                  75                  80

Ile Asp Asp Thr Ile Leu Gly Tyr Pro Ile Thr Glu Ala Ile Phe Thr
                85                  90                  95

Pro Ile Ile Tyr Met Val Thr Trp Phe Lys Gly Asp Ser Lys Glu Lys
            100                 105                 110

Ile Glu Tyr Phe Ser Arg Asn Lys Ser Val Ile Glu Leu Leu Leu Ser
        115                 120                 125

Lys Leu Gln Tyr Val Leu Met Asp Asn Ser Pro Glu Glu Asn Phe Met
    130                 135                 140

Asn Trp Val Lys Ser Lys Asp Ile Asp Leu Lys Lys Tyr Glu Lys Glu
145                 150                 155                 160

Asp Ile Leu Thr Phe Tyr Trp Ala Phe His Phe Leu Asp Val Asn Val
                165                 170                 175

Ile Val Val Ile Phe Ser Thr Val Val Leu Ile Gly Cys Ile Gly Ala
            180                 185                 190

Pro Val Ser Val Leu Val Phe Thr Leu Ile Phe Ser Ile Ser Thr Leu
        195                 200                 205

Ile Glu Asp Ile Arg Asp Leu Val Phe Lys Arg Arg Glu Ala Ile Leu
    210                 215                 220

Arg Gly Glu Glu Gly Ile Glu Asp Glu Asp Glu Arg Lys Ile Tyr Glu
225                 230                 235                 240

Lys Ile Lys Glu Glu Ile Lys Ile Ala Ile Tyr Lys Glu Glu Glu Ala
                245                 250                 255

Arg Asn Gln Ser
            260


<210> SEQ ID NO 28
<211> LENGTH: 192
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 28

atgtatgtat cggcatttat attgttaata tattctttgt atccattttt acatgaacat     60

gttattaata tacttaaaaa gggaataaaa gaagagtcaa ggctaataag tatatggctt    120

gtaatggtta taatatctta tcttcttaaa aggtatgaca ttataactag gtctgatgaa    180
```

-continued

```
ggggatttat ga                                                        192


<210> SEQ ID NO 29
<211> LENGTH: 63
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 29

Met Tyr Val Ser Ala Phe Ile Leu Leu Ile Tyr Ser Leu Tyr Pro Phe
1               5                   10                  15

Leu His Glu His Val Ile Asn Ile Leu Lys Lys Gly Ile Lys Glu Glu
            20                  25                  30

Ser Arg Leu Ile Ser Ile Trp Leu Val Met Val Ile Ile Ser Tyr Leu
        35                  40                  45

Leu Lys Arg Tyr Asp Ile Ile Thr Arg Ser Asp Glu Gly Asp Leu
    50                  55                  60


<210> SEQ ID NO 30
<211> LENGTH: 1311
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 30

atggagagta gagagagggt agtgaaagcg ctgaataagc taataaaggc attggcagta     60

atttggttag tgttgattgt tctgaggctt atatggtatt ttgcagatat agcgtatgtg    120

agcgaaagaa taacagggga ttcatcagat ggtattatag tatttttaa gcagatattt    180

ttgtacattc tttttttggg tttagtagga gtaatattac cgtttttcat tttttctggg    240

ggagggtata ttccatattt agcattgtat gaactgttga caatttttaa aataatagtt    300

gtagtaactg tagcagttgt ggtaaggaag ttttgggttg taaataaata ttttgagaca    360

acagcgttta agagagattt tttagtgaat agtagcaaaa aagaaatagt actaatatta    420

aaatgtgacg aatcatattg taaatataaa acagaagaag agttttatat ttttaagaga    480

attgtgactg aaggaggagg agagatatat ttaccttact ctgtaataga agatagatat    540

gtagatggat atgttatagg caaaaatata tggtatcatg aagaagattc aggtagaaaa    600

atattgatag aagagttatt agaagaaaaa atagtagaaa ataaagaaaa taaaaaggaa    660

caagctgaag agagaattaa tgaagaaata ttagcagaaa aagtggcaag agctgctata    720

aaattgctta aattaagaag gcaagagaag acaatggtca gttgtagtaa tgtttcaccc    780

atttattatg taacagcttt aggtaattca agaggttttt ttgcgaaaat aattccatcg    840

ttaattgtag gctcgttaat tgcggggata tttgtaacaa tagaatttac atattggtta    900

atacaggata taacgaatag tagtcgtata accgtacaca agatcatatg gattatgagt    960

atgctactga tggtaatgct attagcaatt tttgtggtca aaggttttta taaaagagaa   1020

aaaaaagatg aagaaaaact taacatagaa atagtaaaaa taccagtgat actatacagg   1080

gaaaataaga aagctgaggt aaaagaagga ggtatatttg taaaagacga cgaaaatata   1140

gtgttagtga aaaagggaga caaagaagtt gaagagcttt tgagtgaatt aaaaataaaa   1200

aaagaaagag tagaaggatt gtatgaggtg agattaataa gtaagtattc ttctgatgaa   1260

gaagtaaaaa gcacgctatg gtgcttggta agaataggtg gtgaagctta a            1311
```

-continued

<210> SEQ ID NO 31
<211> LENGTH: 436
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 31

Met Glu Ser Arg Glu Arg Val Val Lys Ala Leu Asn Lys Leu Ile Lys
1 5 10 15

Ala Leu Ala Val Ile Trp Leu Val Leu Ile Val Leu Arg Leu Ile Trp
20 25 30

Tyr Phe Ala Asp Ile Ala Tyr Val Ser Glu Arg Ile Thr Gly Asp Ser
35 40 45

Ser Asp Gly Ile Ile Val Phe Phe Lys Gln Ile Phe Leu Tyr Ile Leu
50 55 60

Phe Leu Gly Leu Val Gly Val Ile Leu Pro Phe Phe Ile Phe Ser Gly
65 70 75 80

Gly Gly Tyr Ile Pro Tyr Leu Ala Leu Tyr Glu Leu Leu Thr Ile Phe
85 90 95

Lys Ile Ile Val Val Val Thr Val Ala Val Val Val Arg Lys Phe Trp
100 105 110

Val Val Asn Lys Tyr Phe Glu Thr Thr Ala Phe Lys Arg Asp Phe Leu
115 120 125

Val Asn Ser Ser Lys Lys Glu Ile Val Leu Ile Leu Lys Cys Asp Glu
130 135 140

Ser Tyr Cys Lys Tyr Lys Thr Glu Glu Glu Phe Tyr Ile Phe Lys Arg
145 150 155 160

Ile Val Thr Glu Gly Gly Gly Glu Ile Tyr Leu Pro Tyr Ser Val Ile
165 170 175

Glu Asp Arg Tyr Val Asp Gly Tyr Val Ile Gly Lys Asn Ile Trp Tyr
180 185 190

His Glu Glu Asp Ser Gly Arg Lys Ile Leu Ile Glu Glu Leu Leu Glu
195 200 205

Glu Lys Ile Val Glu Asn Lys Glu Asn Lys Lys Glu Gln Ala Glu Glu
210 215 220

Arg Ile Asn Glu Glu Ile Leu Ala Glu Lys Val Ala Arg Ala Ala Ile
225 230 235 240

Lys Leu Leu Lys Leu Arg Arg Gln Glu Lys Thr Met Val Ser Cys Ser
245 250 255

Asn Val Ser Pro Ile Tyr Tyr Val Thr Ala Leu Gly Asn Ser Arg Gly
260 265 270

Phe Phe Ala Lys Ile Ile Pro Ser Leu Ile Val Gly Ser Leu Ile Ala
275 280 285

Gly Ile Phe Val Thr Ile Glu Phe Thr Tyr Trp Leu Ile Gln Asp Ile
290 295 300

Thr Asn Ser Ser Arg Ile Thr Val His Lys Ile Ile Trp Ile Met Ser
305 310 315 320

Met Leu Leu Met Val Met Leu Leu Ala Ile Phe Val Val Lys Gly Phe
325 330 335

Tyr Lys Arg Glu Lys Lys Asp Glu Glu Lys Leu Asn Ile Glu Ile Val
340 345 350

Lys Ile Pro Val Ile Leu Tyr Arg Glu Asn Lys Lys Ala Glu Val Lys
355 360 365

-continued

```
Glu Gly Gly Ile Phe Val Lys Asp Asp Glu Asn Ile Val Leu Val Lys
    370                 375                 380

Lys Gly Asp Lys Glu Val Glu Glu Leu Leu Ser Glu Leu Lys Ile Lys
385                 390                 395                 400

Lys Glu Arg Val Glu Gly Leu Tyr Glu Val Arg Leu Ile Ser Lys Tyr
                405                 410                 415

Ser Ser Asp Glu Glu Val Lys Ser Thr Leu Trp Cys Leu Val Arg Ile
            420                 425                 430

Gly Gly Glu Ala
        435


<210> SEQ ID NO 32
<211> LENGTH: 774
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 32

atgaatagga tagtagtaag agacatagaa gacaaaaaag gaaaaagcac tattacaaag     60

attagtacat tgcaaaaagc tgtagtgtta gttgttgctg taatttttac tttcttatta    120

ttcaagtacg caaaagaggt agttaatagg aatgagctaa atgtggatga taggataagg    180

cctattacaa aaacattagt ttttggatta tcaatagaac ctgggtttat aggtaatgaa    240

ctttttgact ttcctactgt gggatgttat agtttgattg catttatatg gcggcacgta    300

acggaagatg acaaaattac tctttcagaa aatttaatag aattgacctc aaaagaaaaa    360

caggcaatag aaatgatgtt gaaaaaactt aactatattt tattggaaaa tccaaaagaa    420

gcttttatga gatggattaa atcagagaat attaaaatga acagaaaaga gtataaatat    480

gatattttag aagatttttg ggcatttcat tttttagata tgaatgtaat atttatcgga    540

gttagtatta taagttttat tttttggact acatatgcat tttgcaatat gttattttta    600

attatagtaa ctatggtaga gtttgtaata aacttgtggg tagtaatttt gaagaaagaa    660

gaaggagtag aagatgaaga agaaagagag atatatgaga tgataaaaga agaagtaaga    720

aaaataattt atgatggaaa agaagcaaga aatttaagac aggaaagcaa ttaa          774


<210> SEQ ID NO 33
<211> LENGTH: 257
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 33

Met Asn Arg Ile Val Val Arg Asp Ile Glu Asp Lys Lys Gly Lys Ser
1               5                   10                  15

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

Ala Val Ile Phe Thr Phe Leu Leu Phe Lys Tyr Ala Lys Glu Val Val
        35                  40                  45

Asn Arg Asn Glu Leu Asn Val Asp Asp Arg Ile Arg Pro Ile Thr Lys
    50                  55                  60

Thr Leu Val Phe Gly Leu Ser Ile Glu Pro Gly Phe Ile Gly Asn Glu
65                  70                  75                  80

Leu Phe Asp Phe Pro Thr Val Gly Cys Tyr Ser Leu Ile Ala Phe Ile
                85                  90                  95

Trp Arg His Val Thr Glu Asp Asp Lys Ile Thr Leu Ser Glu Asn Leu
```

-continued

```
            100                 105                 110

Ile Glu Leu Thr Ser Lys Glu Lys Gln Ala Ile Glu Met Met Leu Lys
        115                 120                 125

Lys Leu Asn Tyr Ile Leu Leu Glu Asn Pro Lys Glu Ala Phe Met Arg
    130                 135                 140

Trp Ile Lys Ser Glu Asn Ile Lys Met Asn Arg Lys Glu Tyr Lys Tyr
145                 150                 155                 160

Asp Ile Leu Glu Asp Phe Trp Ala Phe His Phe Leu Asp Met Asn Val
                165                 170                 175

Ile Phe Ile Gly Val Ser Ile Ile Ser Phe Ile Phe Trp Thr Thr Tyr
            180                 185                 190

Ala Phe Cys Asn Met Leu Phe Leu Ile Ile Val Thr Met Val Glu Phe
        195                 200                 205

Val Ile Asn Leu Trp Val Val Ile Leu Lys Lys Glu Glu Gly Val Glu
    210                 215                 220

Asp Glu Glu Glu Arg Glu Ile Tyr Glu Met Ile Lys Glu Glu Val Arg
225                 230                 235                 240

Lys Ile Ile Tyr Asp Gly Lys Glu Ala Arg Asn Leu Arg Gln Glu Ser
                245                 250                 255

Asn


<210> SEQ ID NO 34
<211> LENGTH: 792
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 34

ttgaatagat ttttggcaaa ctttctgaaa tacaaagatc ttttatatga gcttgttttg     60

agagatataa aaattaaata tagacggtct attttaggta tgttttggag cttgctaaat    120

cctcttttga tgatgattgt tttaacgatt gtattttctc acctttttcag atttgacata    180

aaaaattatc ctatttactt attgaccgga caaattatgt ttgctttttt ttctgaatca    240

acctcaagtg caatgagggc aattatagat aacgcttcac tcattaaaaa ggtatatata    300

ccaaaatata tatttcctgt ttccaaagtt ttgtcatctt tttttaatct aattttctca    360

cttttggcaa ttgttttagt tacaatagga atggtagtat tgggcaataa tgttaattta    420

acgtggacat ttttgttatt tcctattcct ttaatattca tattaatttt ttcaataggc    480

ataggcttaa ttctgtcttg ctatgcagta tttttcagag atttgattca tctatactcg    540

gttggattaa ctgcttggat gtatttgaca cccatatttt atcctgtgag tataattcca    600

caaaaatatc tgatattgat taagctaaat ccgatgtatt attttataga gtactttaga    660

aaagttacat tttatggtac actacctaca ctaaaggaaa ctctaatttg cattttggta    720

gacataatat ttttagtaat aggactttta gtttttttatc gaaaacaaaa taaatttatc    780

ttgtatgtgt aa                                                        792


<210> SEQ ID NO 35
<211> LENGTH: 263
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 35

Leu Asn Arg Phe Leu Ala Asn Phe Leu Lys Tyr Lys Asp Leu Leu Tyr
```

-continued

```
1               5                   10                  15

Glu Leu Val Leu Arg Asp Ile Lys Ile Lys Tyr Arg Arg Ser Ile Leu
            20                  25                  30

Gly Met Phe Trp Ser Leu Leu Asn Pro Leu Leu Met Met Ile Val Leu
        35                  40                  45

Thr Ile Val Phe Ser His Leu Phe Arg Phe Asp Ile Lys Asn Tyr Pro
    50                  55                  60

Ile Tyr Leu Leu Thr Gly Gln Ile Met Phe Ala Phe Phe Ser Glu Ser
65                  70                  75                  80

Thr Ser Ser Ala Met Arg Ala Ile Ile Asp Asn Ala Ser Leu Ile Lys
                85                  90                  95

Lys Val Tyr Ile Pro Lys Tyr Ile Phe Pro Val Ser Lys Val Leu Ser
            100                 105                 110

Ser Phe Phe Asn Leu Ile Phe Ser Leu Leu Ala Ile Val Leu Val Thr
        115                 120                 125

Ile Gly Met Val Val Leu Gly Asn Asn Val Asn Leu Thr Trp Thr Phe
    130                 135                 140

Leu Leu Phe Pro Ile Pro Leu Ile Phe Ile Leu Ile Phe Ser Ile Gly
145                 150                 155                 160

Ile Gly Leu Ile Leu Ser Cys Tyr Ala Val Phe Phe Arg Asp Leu Ile
                165                 170                 175

His Leu Tyr Ser Val Gly Leu Thr Ala Trp Met Tyr Leu Thr Pro Ile
            180                 185                 190

Phe Tyr Pro Val Ser Ile Ile Pro Gln Lys Tyr Leu Ile Leu Ile Lys
        195                 200                 205

Leu Asn Pro Met Tyr Tyr Phe Ile Glu Tyr Phe Arg Lys Val Thr Phe
    210                 215                 220

Tyr Gly Thr Leu Pro Thr Leu Lys Glu Thr Leu Ile Cys Ile Leu Val
225                 230                 235                 240

Asp Ile Ile Phe Leu Val Ile Gly Leu Leu Val Phe Tyr Arg Lys Gln
                245                 250                 255

Asn Lys Phe Ile Leu Tyr Val
            260


<210> SEQ ID NO 36
<211> LENGTH: 735
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 36

atggacaaaa acatagctgt aaagattgaa aatgtttcaa tgatgtttaa tatggcatct      60

gaaaagattt atagtattaa agagtacttt ataaaacttg tgtcaggtaa gctatacttt     120

agagaatttt gggccttgaa ggatataagt tttaaaatca aaaaaggtga aatttttggt     180

ataataggct taaacggggc aggcaaaagt accctgctca aaataattgc aggagtttta     240

aagccgacaa tgggtagagt ctatgtaaac ggtactatgg caccattgat tgaacttgga     300

gcaggttttg actttgaact tactgcaaga gagaatatat tcttaaacgg cgctattttg     360

ggctattcaa gaaagttcat gaaagaaaaa tttgacgaga tagtagagtt tgcagaactg     420

agagattttt tagatgttcc tctaaaaaac ttttcgtctg gtatgcaggc aagacttggt     480

tttgctattg ctacaattgt tgaccctgac attttaattg ttgatgagat tctggcagta     540
```

-continued

```
ggagactttc attttcagga aaaatgcgaa agaagaatta atagtatgct tgaaaaaggg    600

acaactattg tgatggtgtc ccattcaata gatcaaattg agagaatgtg ccaaagagtt    660

ttgtggcttg aaaaaggcag aatgaaaatg ataggcgatg caaaagaggt ttgtgaggct    720

tacaggaact cttga                                                     735


<210> SEQ ID NO 37
<211> LENGTH: 244
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 37

Met Asp Lys Asn Ile Ala Val Lys Ile Glu Asn Val Ser Met Met Phe
1               5                   10                  15

Asn Met Ala Ser Glu Lys Ile Tyr Ser Ile Lys Glu Tyr Phe Ile Lys
            20                  25                  30

Leu Val Ser Gly Lys Leu Tyr Phe Arg Glu Phe Trp Ala Leu Lys Asp
        35                  40                  45

Ile Ser Phe Lys Ile Lys Lys Gly Glu Ile Phe Gly Ile Ile Gly Leu
    50                  55                  60

Asn Gly Ala Gly Lys Ser Thr Leu Leu Lys Ile Ile Ala Gly Val Leu
65                  70                  75                  80

Lys Pro Thr Met Gly Arg Val Tyr Val Asn Gly Thr Met Ala Pro Leu
                85                  90                  95

Ile Glu Leu Gly Ala Gly Phe Asp Phe Glu Leu Thr Ala Arg Glu Asn
            100                 105                 110

Ile Phe Leu Asn Gly Ala Ile Leu Gly Tyr Ser Arg Lys Phe Met Lys
        115                 120                 125

Glu Lys Phe Asp Glu Ile Val Glu Phe Ala Glu Leu Arg Asp Phe Leu
    130                 135                 140

Asp Val Pro Leu Lys Asn Phe Ser Ser Gly Met Gln Ala Arg Leu Gly
145                 150                 155                 160

Phe Ala Ile Ala Thr Ile Val Asp Pro Asp Ile Leu Ile Val Asp Glu
                165                 170                 175

Ile Leu Ala Val Gly Asp Phe His Phe Gln Glu Lys Cys Glu Arg Arg
            180                 185                 190

Ile Asn Ser Met Leu Glu Lys Gly Thr Thr Ile Val Met Val Ser His
        195                 200                 205

Ser Ile Asp Gln Ile Glu Arg Met Cys Gln Arg Val Leu Trp Leu Glu
    210                 215                 220

Lys Gly Arg Met Lys Met Ile Gly Asp Ala Lys Glu Val Cys Glu Ala
225                 230                 235                 240

Tyr Arg Asn Ser


<210> SEQ ID NO 38
<211> LENGTH: 732
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 38

atgaaaggaa tagttttggc aggaggaaca ggttcacgtt tatatccgtt aactaaggtt     60

acaaacaagc atttgttgcc tgttggaaaa tatccaatga tttattatcc aatttttaaa    120

ttaaaacaag caggtatcaa agagataatg ataataactg gtaaagaaca catgggagct    180
```

-continued

```
gttgtcaatt tattaggaag tggtcgtgag tttggattag aatttactta taggattcaa    240

gatgaagctg gtggaattgc tcaagctctg ggactgtgta gctttttgc tgggaatgat    300

aagtgtgttg tgatactggg cgacaatgtt tttgaggatg atataactga gtatgtaagg    360

aatttcgaaa agcaggagcg gggcgctagg attcttctta aggaagttcc ggacccgcat    420

agatttggag ttgctgaact gaaagatggg aaaatagtct cgattgaaga aaagcctaag    480

aatccgaaaa gtaattttat tgtcacagga atctatatgt atgatagcca agtttttgat    540

attatcaaga cacttaagcc ttcacagagg ggtgaattgg aaattacgga tgtaaataac    600

gagtacataa gaagaggaga gttgtatttt gactttttaa agggctggtg gactgatgca    660

gggacttttg agtcactaaa aagagcaaat gagcttgctg aaaatatggt gttagatttt    720

aatggaatat aa                                                        732
```

```
<210> SEQ ID NO 39
<211> LENGTH: 243
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 39

Met Lys Gly Ile Val Leu Ala Gly Gly Thr Gly Ser Arg Leu Tyr Pro
1               5                   10                  15

Leu Thr Lys Val Thr Asn Lys His Leu Leu Pro Val Gly Lys Tyr Pro
            20                  25                  30

Met Ile Tyr Tyr Pro Ile Phe Lys Leu Lys Gln Ala Gly Ile Lys Glu
        35                  40                  45

Ile Met Ile Ile Thr Gly Lys Glu His Met Gly Ala Val Val Asn Leu
    50                  55                  60

Leu Gly Ser Gly Arg Glu Phe Gly Leu Glu Phe Thr Tyr Arg Ile Gln
65                  70                  75                  80

Asp Glu Ala Gly Gly Ile Ala Gln Ala Leu Gly Leu Cys Ser Phe Phe
                85                  90                  95

Ala Gly Asn Asp Lys Cys Val Val Ile Leu Gly Asp Asn Val Phe Glu
            100                 105                 110

Asp Asp Ile Thr Glu Tyr Val Arg Asn Phe Glu Lys Gln Glu Arg Gly
        115                 120                 125

Ala Arg Ile Leu Leu Lys Glu Val Pro Asp Pro His Arg Phe Gly Val
    130                 135                 140

Ala Glu Leu Lys Asp Gly Lys Ile Val Ser Ile Glu Glu Lys Pro Lys
145                 150                 155                 160

Asn Pro Lys Ser Asn Phe Ile Val Thr Gly Ile Tyr Met Tyr Asp Ser
                165                 170                 175

Gln Val Phe Asp Ile Ile Lys Thr Leu Lys Pro Ser Gln Arg Gly Glu
            180                 185                 190

Leu Glu Ile Thr Asp Val Asn Asn Glu Tyr Ile Arg Arg Gly Glu Leu
        195                 200                 205

Tyr Phe Asp Phe Leu Lys Gly Trp Trp Thr Asp Ala Gly Thr Phe Glu
    210                 215                 220

Ser Leu Lys Arg Ala Asn Glu Leu Ala Glu Asn Met Val Leu Asp Phe
225                 230                 235                 240

Asn Gly Ile
```

-continued

<210> SEQ ID NO 40
<211> LENGTH: 462
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 40

atggagctta ttgaaggtgt taaggtaaaa agactcgaga aatttgctga tgatagaggt 60

tttttcatgg agattttgag agacgaagat ggctttttag aaaaatttgg acaggcttca 120

atgtctttga cttatccagg ggtaataaaa gcttttcact atcataagtt gcaagacgat 180

gtttggtttt tcccaaaagg caatgcacag gttgttcttt acgatttgcg gccagattcg 240

cctacataca aaaaaacaaa tgtattctac atgggagaac acaaccctat agtactttta 300

atacccagaa tggttgctca tggctatagg gtattgggga atgagccggc aataattgtg 360

tattttacaa cagagcatta taatagagaa aatccagatg aatacagaat tccttgggat 420

gacaaggaaa ttaattttga ttggacaaca aaatttagat aa 462

<210> SEQ ID NO 41
<211> LENGTH: 153
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 41

Met Glu Leu Ile Glu Gly Val Lys Val Lys Arg Leu Glu Lys Phe Ala
1 5 10 15

Asp Asp Arg Gly Phe Phe Met Glu Ile Leu Arg Asp Glu Asp Gly Phe
20 25 30

Leu Glu Lys Phe Gly Gln Ala Ser Met Ser Leu Thr Tyr Pro Gly Val
35 40 45

Ile Lys Ala Phe His Tyr His Lys Leu Gln Asp Asp Val Trp Phe Phe
50 55 60

Pro Lys Gly Asn Ala Gln Val Val Leu Tyr Asp Leu Arg Pro Asp Ser
65 70 75 80

Pro Thr Tyr Lys Lys Thr Asn Val Phe Tyr Met Gly Glu His Asn Pro
85 90 95

Ile Val Leu Leu Ile Pro Arg Met Val Ala His Gly Tyr Arg Val Leu
100 105 110

Gly Asn Glu Pro Ala Ile Ile Val Tyr Phe Thr Thr Glu His Tyr Asn
115 120 125

Arg Glu Asn Pro Asp Glu Tyr Arg Ile Pro Trp Asp Asp Lys Glu Ile
130 135 140

Asn Phe Asp Trp Thr Thr Lys Phe Arg
145 150

<210> SEQ ID NO 42
<211> LENGTH: 957
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 42

atggagacaa ttcttgtagc aggtggggca ggttttatag gaagcaattt tgttaagtac 60

atgattagca aagaagaata caaaataatc aattatgatg cattaaccta tgcaggaaat 120

cttgagaatt tgaaagaggt agaaaaccac ccttattata catttattaa aggagatatt 180

-continued

```
gttgatagat ctaaagttga agaggttttt aaaaattatc aaattgacta tgtaataaac    240

tttgccgcag agtcgcatgt ggacagaagt ataaaggacc ctgatatatt cgttaaaaca    300

aatgttttgg gaacacaagt tttattagat gtgtcgagga aattcgggat aaaaaagttt    360

attcaaattt caacagatga agtatatggt tccttagggc ctgaaggata ttttacagaa    420

gaaagtccgc ttgcaccaaa cagtccttat tctgccagca aagcaggggc tgatatgctt    480

gtgagagcat attttaagac atatggtctg cctgtgaaca taacaaggtg ttcaaacaat    540

tttggtccac atcaacaccc agaaaagttt ataccgaccg taattttgaa tgcgctgcaa    600

aacaagccga taccaattta tggtgacggg caaaatataa gagactggct atatgtagaa    660

gaccactgca gagcaattga gcttgtgctc aaaaaaggta gaataggtga agtatacaat    720

attggcggga ataatgagtg gaggaatata gatatagcca aattgatttt aaaactactc    780

gggaaaccag agaatctaat acaatttgtg gctgacaggc caggacatga taggagatat    840

gcaattgact cgagtaagat tcaaaaggaa ttggggtgga aggttgaata taagtttgat    900

gaagctatta gaaagactat agaatggtat aaaaatgaat ttttcaaagg agaatga       957
```

```
<210> SEQ ID NO 43
<211> LENGTH: 318
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 43

Met Glu Thr Ile Leu Val Ala Gly Gly Ala Gly Phe Ile Gly Ser Asn
1               5                   10                  15

Phe Val Lys Tyr Met Ile Ser Lys Glu Glu Tyr Lys Ile Ile Asn Tyr
            20                  25                  30

Asp Ala Leu Thr Tyr Ala Gly Asn Leu Glu Asn Leu Lys Glu Val Glu
        35                  40                  45

Asn His Pro Tyr Tyr Thr Phe Ile Lys Gly Asp Ile Val Asp Arg Ser
    50                  55                  60

Lys Val Glu Glu Val Phe Lys Asn Tyr Gln Ile Asp Tyr Val Ile Asn
65                  70                  75                  80

Phe Ala Ala Glu Ser His Val Asp Arg Ser Ile Lys Asp Pro Asp Ile
                85                  90                  95

Phe Val Lys Thr Asn Val Leu Gly Thr Gln Val Leu Leu Asp Val Ser
            100                 105                 110

Arg Lys Phe Gly Ile Lys Lys Phe Ile Gln Ile Ser Thr Asp Glu Val
        115                 120                 125

Tyr Gly Ser Leu Gly Pro Glu Gly Tyr Phe Thr Glu Glu Ser Pro Leu
    130                 135                 140

Ala Pro Asn Ser Pro Tyr Ser Ala Ser Lys Ala Gly Ala Asp Met Leu
145                 150                 155                 160

Val Arg Ala Tyr Phe Lys Thr Tyr Gly Leu Pro Val Asn Ile Thr Arg
                165                 170                 175

Cys Ser Asn Asn Phe Gly Pro His Gln His Pro Glu Lys Phe Ile Pro
            180                 185                 190

Thr Val Ile Leu Asn Ala Leu Gln Asn Lys Pro Ile Pro Ile Tyr Gly
        195                 200                 205

Asp Gly Gln Asn Ile Arg Asp Trp Leu Tyr Val Glu Asp His Cys Arg
    210                 215                 220
```

-continued

```
Ala Ile Glu Leu Val Leu Lys Lys Gly Arg Ile Gly Glu Val Tyr Asn
225                 230                 235                 240

Ile Gly Gly Asn Asn Glu Trp Arg Asn Ile Asp Ile Ala Lys Leu Ile
                245                 250                 255

Leu Lys Leu Leu Gly Lys Pro Glu Asn Leu Ile Gln Phe Val Ala Asp
            260                 265                 270

Arg Pro Gly His Asp Arg Arg Tyr Ala Ile Asp Ser Ser Lys Ile Gln
        275                 280                 285

Lys Glu Leu Gly Trp Lys Val Glu Tyr Lys Phe Asp Glu Ala Ile Arg
    290                 295                 300

Lys Thr Ile Glu Trp Tyr Lys Asn Glu Phe Phe Lys Gly Glu
305                 310                 315


<210> SEQ ID NO 44
<211> LENGTH: 2172
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 44

atgagaagta taaaccaatt aaatattggt ctaacagaga caaagtacgg caaattctat     60

tatttaaaga gtgatgcaat aataggcaaa tctttagaaa tatacggaga atgggcggta    120

ccagaaatag agttgttaac aagttttata gaagaagggg atatagtaat tgatgtaggc    180

gcatacatag gaactcatac tattccattt gctcaaaagt tgaatggaaa aggttttgta    240

attgcattcg aaccccaaga gattatattc aatatactag caaagaatat aaggacaaat    300

aacgcagata atgtgcaaat atttaacaaa gctgttttag ataaaaatac aataacctac    360

atagaaactt ttaactacaa tgaaactaat aattttggca gtgctaaaat aataactgat    420

gatattgaaa ctcaaggaat tatcaaaaaa attgaagcaa tcactattga tagtttagaa    480

ttgaataatt gtaaattaat aaaaattgac gtagaaggac aagaagaatt tgtattgaga    540

ggctcagaaa agacaataaa aagttttatg ccaattattt attttgaaag taatgagctt    600

gagaaaacat ggaatagtat ttgcttagta aagaaatggg gatacgatac tttcttgttt    660

agatttccag catttaatcc atttaacatt aaaggtacac aagataatat ttttggttat    720

gcatgtgaaa caggtatttt ggggattcat aaatcaaaac tgagtactta tggaaatgat    780

actttgaaat ggcgaaaatt aaattatcta tttgaaatta ccacgcttga cgatttagct    840

tttttattaa tagaaacttc tcgtggacat gatcccatta tacctgatat taaatataat    900

gaaatcagaa acataaagtt tcttcaaaat tatataattg aggtggtaaa agaaattgaa    960

aaaaaagata aactattaca aacaattaaa gaagatttaa aaataaaaga acaacactta   1020

caaagtattt attgttcgga aggttggaaa ctgttaacca agtgttataa aataagagat   1080

aagatattcc caccgaatag taagagaaga gaattagcaa agtttatttt atttattatg   1140

aaaaaactga atttaaaatt tgttcaggaa ataataagat ttgcaaggat gtatggtatt   1200

aaagctcttt ataatcgaag tcgaattgct ttgcactcta tgaaaagtta ttcgagtaac   1260

caagaaaatt attttgttcc tgaacaattt gaaagagatt atcttttaag tgaaatcagt   1320

gttgatataa ttgtacctat ttataatgct tatgaggatt taaagaggtg tgtagaaagt   1380

attcttaagc acactgattt gaaaaaacat agattggtat taattaatga ttgttcgact   1440

gatgagagaa tatattcttt cttaaaaaat cttgagaacg aaagaaccga agaaaattta   1500
```

-continued

```
cttgttattc ataataagga gaatttgggt tttgtaaaaa cagttaataa aggaattagt   1560

ttgagtaata aagatgttat tattcttaac tctgatacaa ttgtaactgc gagatgggta   1620

gaaaagctta tcagagcagc gtattcaagg tctaatgttg caacagttac accttttttca   1680

aataatgcca caatttgttc attaccaata atgttaaaag ataattcgct tcctttggat   1740

tgggatattg attattttgc taaagttgta gataggattt cacttttaaa atatcctgaa   1800

attccaacag cagttggatt ttgtatgtat attaaaagag aagttattga aaaaatagga   1860

atgtttaacg aagagaaatt tggcaaaggt tatggagaag aaaatgattt ttgtatgcga   1920

gcgctaaacg aagggtatgt aaatatttta tgtgataatt tatttattta tcacaaagga   1980

agtcaaagct ttacggaaga agttaaaagg aaaagagaaa tggagagctt aaaggttatc   2040

aatcaattac atccatttta ctctgagatg gttaaggcct ttatagaaaa aaatccttta   2100

aaatattatc attctacact tgaagaaatc atgtcccttt ataacttatt aaaaagtgag   2160

gagaaaaagt ga                                                       2172
```

```
<210> SEQ ID NO 45
<211> LENGTH: 723
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 45

Met Arg Ser Ile Asn Gln Leu Asn Ile Gly Leu Thr Glu Thr Lys Tyr
1               5                   10                  15

Gly Lys Phe Tyr Tyr Leu Lys Ser Asp Ala Ile Ile Gly Lys Ser Leu
            20                  25                  30

Glu Ile Tyr Gly Glu Trp Ala Val Pro Glu Ile Glu Leu Leu Thr Ser
        35                  40                  45

Phe Ile Glu Glu Gly Asp Ile Val Ile Asp Val Gly Ala Tyr Ile Gly
    50                  55                  60

Thr His Thr Ile Pro Phe Ala Gln Lys Leu Asn Gly Lys Gly Phe Val
65                  70                  75                  80

Ile Ala Phe Glu Pro Gln Glu Ile Ile Phe Asn Ile Leu Ala Lys Asn
                85                  90                  95

Ile Arg Thr Asn Asn Ala Asp Asn Val Gln Ile Phe Asn Lys Ala Val
            100                 105                 110

Leu Asp Lys Asn Thr Ile Thr Tyr Ile Glu Thr Phe Asn Tyr Asn Glu
        115                 120                 125

Thr Asn Asn Phe Gly Ser Ala Lys Ile Ile Thr Asp Asp Ile Glu Thr
    130                 135                 140

Gln Gly Ile Ile Lys Lys Ile Glu Ala Ile Thr Ile Asp Ser Leu Glu
145                 150                 155                 160

Leu Asn Asn Cys Lys Leu Ile Lys Ile Asp Val Glu Gly Gln Glu Glu
                165                 170                 175

Phe Val Leu Arg Gly Ser Glu Lys Thr Ile Lys Ser Phe Met Pro Ile
            180                 185                 190

Ile Tyr Phe Glu Ser Asn Glu Leu Glu Lys Thr Trp Asn Ser Ile Cys
        195                 200                 205

Leu Val Lys Lys Trp Gly Tyr Asp Thr Phe Leu Phe Arg Phe Pro Ala
    210                 215                 220

Phe Asn Pro Phe Asn Ile Lys Gly Thr Gln Asp Asn Ile Phe Gly Tyr
225                 230                 235                 240
```

-continued

```
Ala Cys Glu Thr Gly Ile Leu Gly Ile His Lys Ser Lys Leu Ser Thr
                245                 250                 255

Tyr Gly Asn Asp Thr Leu Lys Trp Arg Lys Leu Asn Tyr Leu Phe Glu
            260                 265                 270

Ile Thr Thr Leu Asp Asp Leu Ala Phe Leu Leu Ile Glu Thr Ser Arg
        275                 280                 285

Gly His Asp Pro Ile Ile Pro Asp Ile Lys Tyr Asn Glu Ile Arg Asn
    290                 295                 300

Ile Lys Phe Leu Gln Asn Tyr Ile Ile Glu Val Val Lys Glu Ile Glu
305                 310                 315                 320

Lys Lys Asp Lys Leu Leu Gln Thr Ile Lys Glu Asp Leu Lys Ile Lys
                325                 330                 335

Glu Gln His Leu Gln Ser Ile Tyr Cys Ser Glu Gly Trp Lys Leu Leu
            340                 345                 350

Thr Lys Cys Tyr Lys Ile Arg Asp Lys Ile Phe Pro Pro Asn Ser Lys
        355                 360                 365

Arg Arg Glu Leu Ala Lys Phe Ile Leu Phe Ile Met Lys Lys Leu Asn
    370                 375                 380

Leu Lys Phe Val Gln Glu Ile Ile Arg Phe Ala Arg Met Tyr Gly Ile
385                 390                 395                 400

Lys Ala Leu Tyr Asn Arg Ser Arg Ile Ala Leu His Ser Met Lys Ser
                405                 410                 415

Tyr Ser Ser Asn Gln Glu Asn Tyr Phe Val Pro Glu Gln Phe Glu Arg
            420                 425                 430

Asp Tyr Leu Leu Ser Glu Ile Ser Val Asp Ile Ile Val Pro Ile Tyr
        435                 440                 445

Asn Ala Tyr Glu Asp Leu Lys Arg Cys Val Glu Ser Ile Leu Lys His
    450                 455                 460

Thr Asp Leu Lys Lys His Arg Leu Val Leu Ile Asn Asp Cys Ser Thr
465                 470                 475                 480

Asp Glu Arg Ile Tyr Ser Phe Leu Lys Asn Leu Glu Asn Glu Arg Thr
                485                 490                 495

Glu Glu Asn Leu Leu Val Ile His Asn Lys Glu Asn Leu Gly Phe Val
            500                 505                 510

Lys Thr Val Asn Lys Gly Ile Ser Leu Ser Asn Lys Asp Val Ile Ile
        515                 520                 525

Leu Asn Ser Asp Thr Ile Val Thr Ala Arg Trp Val Glu Lys Leu Ile
    530                 535                 540

Arg Ala Ala Tyr Ser Arg Ser Asn Val Ala Thr Val Thr Pro Phe Ser
545                 550                 555                 560

Asn Asn Ala Thr Ile Cys Ser Leu Pro Ile Met Leu Lys Asp Asn Ser
                565                 570                 575

Leu Pro Leu Asp Trp Asp Ile Asp Tyr Phe Ala Lys Val Val Asp Arg
            580                 585                 590

Ile Ser Leu Leu Lys Tyr Pro Glu Ile Pro Thr Ala Val Gly Phe Cys
        595                 600                 605

Met Tyr Ile Lys Arg Glu Val Ile Glu Lys Ile Gly Met Phe Asn Glu
    610                 615                 620

Glu Lys Phe Gly Lys Gly Tyr Gly Glu Glu Asn Asp Phe Cys Met Arg
625                 630                 635                 640
```

-continued

```
Ala Leu Asn Glu Gly Tyr Val Asn Ile Leu Cys Asp Asn Leu Phe Ile
                645                 650                 655

Tyr His Lys Gly Ser Gln Ser Phe Thr Glu Glu Val Lys Arg Lys Arg
            660                 665                 670

Glu Met Glu Ser Leu Lys Val Ile Asn Gln Leu His Pro Phe Tyr Ser
        675                 680                 685

Glu Met Val Lys Ala Phe Ile Glu Lys Asn Pro Leu Lys Tyr Tyr His
    690                 695                 700

Ser Thr Leu Glu Glu Ile Met Ser Leu Tyr Asn Leu Leu Lys Ser Glu
705                 710                 715                 720

Glu Lys Lys


<210> SEQ ID NO 46
<211> LENGTH: 1242
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 46

atgagcacaa ttgtttattt agtccattca atacccaaat atgaaaagag tggaacccca     60

atagctgctt ggagagtggc aaagggtgta aaagaaaaat ataatcaaaa tgtagcattt    120

ataattcctt cacctgatgg agaagaggga aaagaaaagg ttgacgatat tttggtttac    180

aaagtaaaaa gaatagattg gcatgaaaat ttttttcatg attttgatat agatagagaa    240

acgtatatta gaaaaataaa gaatatttta aaagaagtta attgtaatat tttgcatatt    300

tataatttag tatttagctc ttatcaggta atgaaactta ggaaagaagg aatcagaatt    360

gttcgtacga taacacatac tgaggatatt tgttttaatg ttgatccttt tgttaaagtt    420

ggtgataaaa ttgagatatg tagtgggcct gatccgatag caaaatgcgc ttgtcattat    480

aaacagatgt atggtggtaa taatttgatg gagtttataa ttaaaaagat atctaagcat    540

tttacgagcg tagaattatt atattctaat ttttgtgata taataacctt tactaatgaa    600

gaatttgcta aatattttac caactacgta aatattccaa gagatgtaat ccgaataata    660

ccacatggag tagaaaataa attagaaaaa tatatattgc cgaatatgcc aaaaaacgaa    720

ggatttagat ttttatatct tggaggagat aattttagaa aaggatttgt tatattagat    780

aatgcgttaa attctttaaa tggggagtta ttcaacaaga ttaaggaaat aactatagtg    840

ggtaaaacaa caaaggaatt tagagaaagg tttaataatg ataaatatat gtttaagggg    900

gtattgccag aagaagaatt atataacgaa ataagtaatg ctgatcttgt gatattgccc    960

acatttttcg aaacatataa catatcttta agagaagcta tcaagttagg aaaacctgtt   1020

ataactacta aaacttttgg ttcgaatatt gtagtggatg gatataatgg gtttagattt   1080

gatattggtg atagcttaca attaaaaaac ataatagaat taatattaag caatccccaa   1140

atactggtag atatgagcaa aaattgttta aatactcata taactgatat tgaagaagaa   1200

ataaagttat ttatgaaagt ttacaatgaa ttaaaagatt ag                      1242


<210> SEQ ID NO 47
<211> LENGTH: 413
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 47

Met Ser Thr Ile Val Tyr Leu Val His Ser Ile Pro Lys Tyr Glu Lys
```

-continued

```
1               5                   10                  15

Ser Gly Thr Pro Ile Ala Ala Trp Arg Val Ala Lys Gly Val Lys Glu
            20                  25                  30

Lys Tyr Asn Gln Asn Val Ala Phe Ile Ile Pro Ser Pro Asp Gly Glu
        35                  40                  45

Glu Gly Lys Glu Lys Val Asp Asp Ile Leu Val Tyr Lys Val Lys Arg
    50                  55                  60

Ile Asp Trp His Glu Asn Phe Phe His Asp Phe Asp Ile Asp Arg Glu
65                  70                  75                  80

Thr Tyr Ile Arg Lys Ile Lys Asn Ile Leu Lys Glu Val Asn Cys Asn
                85                  90                  95

Ile Leu His Ile Tyr Asn Leu Val Phe Ser Ser Tyr Gln Val Met Lys
            100                 105                 110

Leu Arg Lys Glu Gly Ile Arg Ile Val Arg Thr Ile Thr His Thr Glu
        115                 120                 125

Asp Ile Cys Phe Asn Val Asp Pro Phe Val Lys Val Gly Asp Lys Ile
    130                 135                 140

Glu Ile Cys Ser Gly Pro Asp Pro Ile Ala Lys Cys Ala Cys His Tyr
145                 150                 155                 160

Lys Gln Met Tyr Gly Gly Asn Asn Leu Met Glu Phe Ile Ile Lys Lys
                165                 170                 175

Ile Ser Lys His Phe Thr Ser Val Glu Leu Leu Tyr Ser Asn Phe Cys
            180                 185                 190

Asp Ile Ile Thr Phe Thr Asn Glu Glu Phe Ala Lys Tyr Phe Thr Asn
        195                 200                 205

Tyr Val Asn Ile Pro Arg Asp Val Ile Arg Ile Ile Pro His Gly Val
    210                 215                 220

Glu Asn Lys Leu Glu Lys Tyr Ile Leu Pro Asn Met Pro Lys Asn Glu
225                 230                 235                 240

Gly Phe Arg Phe Leu Tyr Leu Gly Gly Asp Asn Phe Arg Lys Gly Phe
                245                 250                 255

Val Ile Leu Asp Asn Ala Leu Asn Ser Leu Asn Gly Glu Leu Phe Asn
            260                 265                 270

Lys Ile Lys Glu Ile Thr Ile Val Gly Lys Thr Thr Lys Glu Phe Arg
        275                 280                 285

Glu Arg Phe Asn Asn Asp Lys Tyr Met Phe Lys Gly Val Leu Pro Glu
    290                 295                 300

Glu Glu Leu Tyr Asn Glu Ile Ser Asn Ala Asp Leu Val Ile Leu Pro
305                 310                 315                 320

Thr Phe Phe Glu Thr Tyr Asn Ile Ser Leu Arg Glu Ala Ile Lys Leu
                325                 330                 335

Gly Lys Pro Val Ile Thr Thr Lys Thr Phe Gly Ser Asn Ile Val Val
            340                 345                 350

Asp Gly Tyr Asn Gly Phe Arg Phe Asp Ile Gly Asp Ser Leu Gln Leu
        355                 360                 365

Lys Asn Ile Ile Glu Leu Ile Leu Ser Asn Pro Gln Ile Leu Val Asp
    370                 375                 380

Met Ser Lys Asn Cys Leu Asn Thr His Ile Thr Asp Ile Glu Glu Glu
385                 390                 395                 400

Ile Lys Leu Phe Met Lys Val Tyr Asn Glu Leu Lys Asp
                405                 410
```

-continued

<210> SEQ ID NO 48
<211> LENGTH: 1008
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 48

atgagaaatt taacagttat aatacctaca tataatcaaa aagatttatt agaaagagca 60

ataaattcgt taatatctaa gtgcaatgac gaaataaata tatatgtatt agtcaacaat 120

acggaaagta attacatcgt aagaaataaa aattatgcaa atctacatgt tgaatatttg 180

aacacaaact gtggtttttg taaagcagtc aattatggat tacgattaat aaaaaaatca 240

agatttattt ttcttttaaa cgatgatact gaagtaataa atcagataga tatagataat 300

attattcatg aattgattga aaaaggcaac atttttttcta tatcgttgaa aatgctgaag 360

ggtaattatc caaatctttt agacgacgca ggtgatatgt acactatctt aggttggcag 420

tttaaaagag gcaatggtct tccaaaagaa ctttatgata gaccgtgtga aattatttcc 480

gcatgtggtg gtgctgcaat ctataacaaa aaaattcttg atgagatagg ttacttcgat 540

gaagattttt ttgcatatct tgaggatgta gatttaggtt tgagagcgct catgagggga 600

tataaaaatt tatattatcc ttacataagc gtattgcatg ttggaagtgc gacaacagga 660

ggaaaatata acgatattac tatcagactt acagcaagaa actcgatata tgttatatac 720

aaaaatcttc ccttacccct tttaataatt aattttccct ttattttatt gggatactta 780

atcaaattta tattctttgc taaaaaagga aaaggaaaaa tttacataag tggagttctt 840

gaaggactaa aaaatttgcc caaatttaaa gaaaaaagaa gagaaaatat gagaaaaaag 900

aaaatttcta acataaagct tgaatggatt cttattaaag ctacatttga atattttcac 960

caatatataa aaagagcatt ttacacttta agaggtgcaa agaaatga 1008

<210> SEQ ID NO 49
<211> LENGTH: 335
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 49

Met Arg Asn Leu Thr Val Ile Ile Pro Thr Tyr Asn Gln Lys Asp Leu
1 5 10 15

Leu Glu Arg Ala Ile Asn Ser Leu Ile Ser Lys Cys Asn Asp Glu Ile
20 25 30

Asn Ile Tyr Val Leu Val Asn Asn Thr Glu Ser Asn Tyr Ile Val Arg
35 40 45

Asn Lys Asn Tyr Ala Asn Leu His Val Glu Tyr Leu Asn Thr Asn Cys
50 55 60

Gly Phe Cys Lys Ala Val Asn Tyr Gly Leu Arg Leu Ile Lys Lys Ser
65 70 75 80

Arg Phe Ile Phe Leu Leu Asn Asp Asp Thr Glu Val Ile Asn Gln Ile
85 90 95

Asp Ile Asp Asn Ile Ile His Glu Leu Ile Glu Lys Gly Asn Ile Phe
100 105 110

Ser Ile Ser Leu Lys Met Leu Lys Gly Asn Tyr Pro Asn Leu Leu Asp
115 120 125

Asp Ala Gly Asp Met Tyr Thr Ile Leu Gly Trp Gln Phe Lys Arg Gly

-continued

```
    130                 135                 140

Asn Gly Leu Pro Lys Glu Leu Tyr Asp Arg Pro Cys Glu Ile Ile Ser
145                 150                 155                 160

Ala Cys Gly Gly Ala Ala Ile Tyr Asn Lys Lys Ile Leu Asp Glu Ile
                165                 170                 175

Gly Tyr Phe Asp Glu Asp Phe Phe Ala Tyr Leu Glu Asp Val Asp Leu
            180                 185                 190

Gly Leu Arg Ala Leu Met Arg Gly Tyr Lys Asn Leu Tyr Tyr Pro Tyr
        195                 200                 205

Ile Ser Val Leu His Val Gly Ser Ala Thr Thr Gly Gly Lys Tyr Asn
    210                 215                 220

Asp Ile Thr Ile Arg Leu Thr Ala Arg Asn Ser Ile Tyr Val Ile Tyr
225                 230                 235                 240

Lys Asn Leu Pro Leu Pro Leu Leu Ile Ile Asn Phe Pro Phe Ile Leu
                245                 250                 255

Leu Gly Tyr Leu Ile Lys Phe Ile Phe Phe Ala Lys Lys Gly Lys Gly
            260                 265                 270

Lys Ile Tyr Ile Ser Gly Val Leu Glu Gly Leu Lys Asn Leu Pro Lys
        275                 280                 285

Phe Lys Glu Lys Arg Arg Glu Asn Met Arg Lys Lys Lys Ile Ser Asn
    290                 295                 300

Ile Lys Leu Glu Trp Ile Leu Ile Lys Ala Thr Phe Glu Tyr Phe His
305                 310                 315                 320

Gln Tyr Ile Lys Arg Ala Phe Tyr Thr Leu Arg Gly Ala Lys Lys
                325                 330                 335


<210> SEQ ID NO 50
<211> LENGTH: 846
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 50

atgatactaa tcacaggttc caaaggccaa cttggaagcg aatttataaa gcagtttgaa      60

aataaatata atgtgaaagg tattgacatt gaacaagttg atattactga tcttgacagt     120

acagttagtt atatttcagc cacaaagccc aacatcatca tccactgtgc ggcttacacc     180

aatgtagatg ggtgtgagag tgacaaagac actgctttta aagtaaatgc cattgggaca     240

cgaaatgttg caatggctgc agaaaaagtt ggtgcaaagc ttgtatatat atctaccgat     300

tatgtatttg acggtgaaaa agagaagcca tataatgaat ttgatagacc aaatccgata     360

agcatatatg gcctttctaa gcttgcggga gaggaatttg taaaaacttt ttgtagcaga     420

tattttatag tgaggattgc ctggctgtat ggtgaaaatg gtaacaactt tgtaaaaaca     480

attgtaaaac ttgctaaaga aaaaggcgag attgatgttg ttaatgatca gagaggcaat     540

cccacattta ccaaagatgt tgttcaagct gtagaagtaa taatgaactc agagaaatat     600

ggaacatacc atgtaacaaa tgaaggcata acttcttggt acgattttgc atataagatt     660

gtaagcacgt ttgggataga ctgcaaagtc aatccaacta caagtgataa atttattcgc     720

ccggcaaaac gtcctaagaa ctcagcactt gacaagatga tgttaagact tgaatttgga     780

tataaaatga gacattggga agaggctttt gaagagtttg ctatgctgat gaaaggaaga     840

atatag                                                                846
```

-continued

<210> SEQ ID NO 51
<211> LENGTH: 281
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 51

Met Ile Leu Ile Thr Gly Ser Lys Gly Gln Leu Gly Ser Glu Phe Ile
1 5 10 15

Lys Gln Phe Glu Asn Lys Tyr Asn Val Lys Gly Ile Asp Ile Glu Gln
20 25 30

Val Asp Ile Thr Asp Leu Asp Ser Thr Val Ser Tyr Ile Ser Ala Thr
35 40 45

Lys Pro Asn Ile Ile Ile His Cys Ala Ala Tyr Thr Asn Val Asp Gly
50 55 60

Cys Glu Ser Asp Lys Asp Thr Ala Phe Lys Val Asn Ala Ile Gly Thr
65 70 75 80

Arg Asn Val Ala Met Ala Ala Glu Lys Val Gly Ala Lys Leu Val Tyr
85 90 95

Ile Ser Thr Asp Tyr Val Phe Asp Gly Glu Lys Glu Lys Pro Tyr Asn
100 105 110

Glu Phe Asp Arg Pro Asn Pro Ile Ser Ile Tyr Gly Leu Ser Lys Leu
115 120 125

Ala Gly Glu Glu Phe Val Lys Thr Phe Cys Ser Arg Tyr Phe Ile Val
130 135 140

Arg Ile Ala Trp Leu Tyr Gly Glu Asn Gly Asn Asn Phe Val Lys Thr
145 150 155 160

Ile Val Lys Leu Ala Lys Glu Lys Gly Glu Ile Asp Val Val Asn Asp
165 170 175

Gln Arg Gly Asn Pro Thr Phe Thr Lys Asp Val Val Gln Ala Val Glu
180 185 190

Val Ile Met Asn Ser Glu Lys Tyr Gly Thr Tyr His Val Thr Asn Glu
195 200 205

Gly Ile Thr Ser Trp Tyr Asp Phe Ala Tyr Lys Ile Val Ser Thr Phe
210 215 220

Gly Ile Asp Cys Lys Val Asn Pro Thr Thr Ser Asp Lys Phe Ile Arg
225 230 235 240

Pro Ala Lys Arg Pro Lys Asn Ser Ala Leu Asp Lys Met Met Leu Arg
245 250 255

Leu Glu Phe Gly Tyr Lys Met Arg His Trp Glu Glu Ala Phe Glu Glu
260 265 270

Phe Ala Met Leu Met Lys Gly Arg Ile
275 280

<210> SEQ ID NO 52
<211> LENGTH: 897
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 52

atggatttgt ctataattat agttaattac aataccagga atttgctcag aaaaactctt 60

gagtcaatat ataagaatcc tacctataga gaatttgaaa taattgtggt agacaatgca 120

tcaagtgatg gcagccaaga gatggtaaaa aaagagtttc caaatgttat tttgattaaa 180

-continued

```
aataaacaaa atttaggatt tgccaaagcc aacaatatag ggataagaat tgcaaaggga    240

aagtacatat tgcttttaaa ttcagataca gaagttttaa ggggtacact tgatagttgt    300

atagattttt tggagaaaga tgaagctaaa gaaattggaa ttcttggatg taaagtggtg    360

cttccagatg gcaagcttga tttagcatgt agaagagggt ttcctactcc taagaattcg    420

ttttttaaga tatttgggct tgcgaagctg tttcctaaaa gtcgattttt tgcaggttac    480

aatcttacat atcttgatga aaatcaatct tatgaagttg attcagttgt tggggcgttt    540

atgttaataa ggcgcgaagt tatagataaa ataggtttac ttgatgagga ttatttcatg    600

tttggcgagg acatagattt ttgttttaga gccaaacaaa atggatttaa agtttactac    660

tatgctgatg caaaaatcat tcatcacaaa agaggttctg gcaggaattt gaaggttttg    720

tcagcttttt atgactcaat gtggatattt tataaaaaac attactacaa cagataccca    780

aaaaccttag ccttattaat atttataaca attaagttga ttaaatcatt aaagcttgcc    840

catgcgaaag tacggaactt tcctattaga aagggaaaaa aaccatgcat agcctaa       897
```

```
<210> SEQ ID NO 53
<211> LENGTH: 298
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 53

Met Asp Leu Ser Ile Ile Ile Val Asn Tyr Asn Thr Arg Asn Leu Leu
1               5                   10                  15

Arg Lys Thr Leu Glu Ser Ile Tyr Lys Asn Pro Thr Tyr Arg Glu Phe
            20                  25                  30

Glu Ile Ile Val Val Asp Asn Ala Ser Ser Asp Gly Ser Gln Glu Met
        35                  40                  45

Val Lys Lys Glu Phe Pro Asn Val Ile Leu Ile Lys Asn Lys Gln Asn
    50                  55                  60

Leu Gly Phe Ala Lys Ala Asn Asn Ile Gly Ile Arg Ile Ala Lys Gly
65                  70                  75                  80

Lys Tyr Ile Leu Leu Leu Asn Ser Asp Thr Glu Val Leu Arg Gly Thr
                85                  90                  95

Leu Asp Ser Cys Ile Asp Phe Leu Glu Lys Asp Glu Ala Lys Glu Ile
            100                 105                 110

Gly Ile Leu Gly Cys Lys Val Val Leu Pro Asp Gly Lys Leu Asp Leu
        115                 120                 125

Ala Cys Arg Arg Gly Phe Pro Thr Pro Lys Asn Ser Phe Phe Lys Ile
    130                 135                 140

Phe Gly Leu Ala Lys Leu Phe Pro Lys Ser Arg Phe Phe Ala Gly Tyr
145                 150                 155                 160

Asn Leu Thr Tyr Leu Asp Glu Asn Gln Ser Tyr Glu Val Asp Ser Val
                165                 170                 175

Val Gly Ala Phe Met Leu Ile Arg Arg Glu Val Ile Asp Lys Ile Gly
            180                 185                 190

Leu Leu Asp Glu Asp Tyr Phe Met Phe Gly Glu Asp Ile Asp Phe Cys
        195                 200                 205

Phe Arg Ala Lys Gln Asn Gly Phe Lys Val Tyr Tyr Tyr Ala Asp Ala
    210                 215                 220

Lys Ile Ile His His Lys Arg Gly Ser Gly Arg Asn Leu Lys Val Leu
225                 230                 235                 240
```

-continued

```
Ser Ala Phe Tyr Asp Ser Met Trp Ile Phe Tyr Lys Lys His Tyr Tyr
                245                 250                 255

Asn Arg Tyr Pro Lys Thr Leu Ala Leu Leu Ile Phe Ile Thr Ile Lys
            260                 265                 270

Leu Ile Lys Ser Leu Lys Leu Ala His Ala Lys Val Arg Asn Phe Pro
        275                 280                 285

Ile Arg Lys Gly Lys Lys Pro Cys Ile Ala
    290                 295


<210> SEQ ID NO 54
<211> LENGTH: 5133
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 54

ttgaattcaa agaaagcttt tagactgcta tcatgggtag tgataatatc ttttgtattg     60

ggttttataa acccagctgt ttttgcaaaa ggttttaaag acacatcaaa tcactgggca    120

aaagatgtaa ttgagagatg ggcaaataca tacaatgtgg caaatggtta ttctgatggc    180

acttttaaac cctcaaatgc tattacaagg gctgaatttg ctcagcttgt gagcagagta    240

attggtgagg ctttagtcag gtctgaaatt aactttaagg atgtaaaaga aaatgattgg    300

ttttactcag ccgtcaaaaa tcttgctgat tatataagtg ggtatccaga tggtacattt    360

aaaccaaaga acagtataac aagagaagaa gcagcatgta ttttggcaaa ggtatttggt    420

attgacaaat cacagagcaa tgttctttcg aagttttctg attataaaca agtgtcagag    480

tgggcaaaag agtatttagc agctatggtt gaaaatggat atattaatgg ctataaagat    540

aaaacattaa gacctaaaaa ttacattaca aaagctgagg cactgactat tttagataac    600

attgttgggc ttttgttatc aaaaaagggt atgtatgtag gaagagaagt caagggaaat    660

ttaattataa gtgcacctta tgttagtgtg tcaggattta atgtaagtaa aaattgcttg    720

ataacagaag gagtaaaaga tggaaatgtt acaattacta atacgcaggt agacggtaat    780

atcattgttc gtggcggagg agaacacagt gtcgttttga aaaatgtcaa agcttctgca    840

gttatagttg tgaacaaaca agcagctaca aaagtaaata ttagtggaaa ttcaaagata    900

gataaggttg taattgaaaa gccagctaat attttagttg agaagactgc agaagtgaac    960

acagtggata tcagagctga cagaacaatt tttaaggcag aaggcaaaat agagaggata   1020

gcagtaaatg caaaagatgt aaaagtaaat gataaggagg tccagcaagg cgctatacta   1080

acaattacta ataccaatga aaatgagaaa aagactcaat ctacaaatac caatacttcg   1140

caggagtcaa cacaaaatga aactacgtct caaacatcta ataatacttc aagtgtgcaa   1200

acaaatactg gaacatcaac tggaaatata tcttcaggca cttcaactgg tggctataca   1260

ggtggaggaa gtgtatatag tggcggttct tcgggttctg gaacaagtgt ccaatatgga   1320

cctgtaagta gagtagaggt tgacaaagaa accatcccta ttggtcagaa cgtccaagta   1380

acagttacag tgaaggatgc agcggggaat ttattgccaa acaaggttgt aaggattgaa   1440

ggacaggctg cttatacaaa ttccttaggt agtgcaacat ttactttatc agttcaggat   1500

agcaaagaaa tagttataaa tgtagatggc aatgactatt atggtctttt gtacgcaatt   1560

aaaccaactg aaggagtttt gacatttaag ttaaagacgg caagtggtaa ctttttaagt   1620

ggatttaatg tgaaactagt aaatcaagct aagcaatttt ctgagttaaa gtctgccaca   1680
```

-continued

```
gctgaacaag tgtcattcat tgtaccagca gttgatggat acaaagcttt aatatggtca   1740
tatgatagtc aaaatgggct tatatatacc atcttagata atctgtcagt tggaaacggt   1800
attgtgagac ttgttgcaga cacaacatta ccaaattatg tacaagctac actagatttt   1860
agcataaata accaacccct caataactat gagttttcaa ttatcaatag ctcaaaatct   1920
aatacttttt caattgatga tgtaaagcta aatatttctt caaatcagct taaaattaca   1980
gcagatgcgg gcagttattc aatcaaggta gcaaaagata ctcaggatgg taagctatat   2040
tttatgagag attttaatct ccaacagtca ggacaagctt tttcatttaa tttcagctct   2100
tttaaaagag ttatatttaa tttttctggt attggcactt tacaaaagag tgtatctgta   2160
aaattaaatg gtagctggtt tgaactctcc ggggaaaatg atatatatct gcaaagagga   2220
gtttataacc tggaagaagt ttttataaag acatacgatg agaacagtaa agaggttgat   2280
tattcatata aggtgccagc aggcagtagt ccaattgtag ttgatgcgac aggaagtaac   2340
gaagagtgct ctgtgaatgt agacctttca atagataatg tctcaagcag tgtatatgct   2400
actaatatag atggaaatga atttgcttct ataaaggcag gttcacttat agattttgta   2460
gtggttttaa agacgaaaag tgggctgaac ttagctttgt taaagccatc tgaaaagatg   2520
tcagatacag caccattgtt taaagcaagt gatattgtgg catacatatc aagtggaaat   2580
gtacagaata atattgaact tctttgtaca tcggtagaag ataataatat ttcgcatgta   2640
ctcggatatt tgccaaggaa tcttcaagat ggtactgtta gcgtagaatt tacggcagaa   2700
ataggtttgc tttatggaag ctctattaca agcagtattt cgctttctat agacaatcaa   2760
aatggtgaca gtagaggtta tgtatatcca caaaatacaa ctacaagtgc tgtatacttt   2820
gtatgtgata aaattgacag tggagcaagt gactttttag cacttagtct tccggcaagt   2880
gaaacttgtg acgtcaatct tattgccaac aaaatatatt ttgaactcat gacagaaaga   2940
ccgcttactt ttgaatacgg attgaatgga atgtttggac tgagctttaa tctcaccagt   3000
gaaaacctct acatgatatg gcctgtttat cttttcagca aagacgaggg tcttacaaga   3060
agacaagctg ttgaacaaaa gtccttgcaa atagtatcta cagttgtaga ccaaacttac   3120
aatgactatg ataaagtttt aggtttgcat gactggcttg ttttgcacac ccagtacgat   3180
ttggaagggt atcttaacaa caacatacct tatgagtcac atacagctta tggtgcacta   3240
attaatggta ttgcagtttg caacggttat gcaacagcta tgcttgcgct tttagaggat   3300
gcaggtatag actcaaaaga gatctatggt atggcaggtg taggtaattc aaaagaagcc   3360
catgcatgga acatggtaag cttagaaaat aattggtatc atttagatgc aacatgggat   3420
gacccagact ggggcaatta tgttgaacat gccttttca atgttccaga tagtaaaatt   3480
gagttaacgc atgattggga aagaaacctt tatcctgctg ctacagctat agattacagc   3540
tatggtaact attacaatgt agaaaccatt ccacaggatg tttattacaa cgaagacaat   3600
attgtaacaa tagttgtcaa aaactataaa ggtgaactac aagccaataa acttattaca   3660
ataacgaaat ttaaccaagc aggagttgaa gatattgtat ttactggcta tacttctgct   3720
gatggtacaa taacattcaa tttaaagcct actgagatga tatcttacca gatttatata   3780
acctctttta tggaagacaa aggatacatt aaagttgttg aaaaactgaa gccggtgaca   3840
ttaagtttaa atattgacaa tcaaccaatt gatgaatttt atatctcaga tggtttgaac   3900
ttcttgaaag tcacgggtgg tcagcgacag gttggactta gcaggtggcg cgagaatgat   3960
```

-continued

```
ctaattttct atggtttagg ttttgtaata aagaaaagtt taacatttga cagctatgac   4020

ccgatccaac ttacagtttg taacgatgca tacgattctt atgtactttt aagtgtgtat   4080

aacagtgctt atgcagttgt cggtgcagac gtttatgtga ttgacaaaga taccttcaag   4140

gagctgtttg taggcagcac agactcagaa ggaaatcttc cgcttgttct tacaaacgga   4200

gaatacattt taaaagttgc taattataat tcagaaacaa attctcaaga tttttactat   4260

tcgattttac aggtaacaca agatagtgta acttcaattg atttgtctca gtttaatcag   4320

attcaatata taccaggatt taccgaaggc ggcgttggaa tgggagacaa gctgttactt   4380

ggtttcgact ttgactatga tggaaaattt gttgtgtcat cgatggggca ctcgagaaag   4440

gcttattttg caccgggtga ttatggttgg gtatggattt tggcccataa cccgtatgat   4500

atagaagata atgttgctac atcattagtg tataaacctt tgcaaaaact aattattcca   4560

gaaggtacgc ttcagcaata tactgttgat ttgagagttg acagaagcaa aatagtcttg   4620

aactgtgaac gtcgaccaga aggaagctgg gattttcttc caacaacgct tgatatggta   4680

tatgaaaagg atgacattat aattgatatt tatattccaa ccaatagtca gagctacgtt   4740

gtgggtgggg atactgtacc tcagcagctt atagaaaata atggagatat gcatattatg   4800

tggatgggta ttgctggcat ttatccaatt agaacgtata ctatttcaaa tagtactcca   4860

tatgaattgt ttacaggaca tcttggtttt attgacaata ctaatagttg gcgtataatt   4920

gtacttgatg tagtaggtga tctaaattca acatttagaa tagtattacc actatcaccg   4980

tatgacattg atgttattat cgacaaagaa attacaattt acccacttac tgcaggatca   5040

ggtggtttgc gaaaagagtc atatattcaa aaggctaaga tttacaatca aataatcaac   5100

acacagctta ttaaaaaaga gaataaaaaa taa                                5133
```

```
<210> SEQ ID NO 55
<211> LENGTH: 1710
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 55

Leu Asn Ser Lys Lys Ala Phe Arg Leu Leu Ser Trp Val Val Ile Ile
1               5                   10                  15

Ser Phe Val Leu Gly Phe Ile Asn Pro Ala Val Phe Ala Lys Gly Phe
            20                  25                  30

Lys Asp Thr Ser Asn His Trp Ala Lys Asp Val Ile Glu Arg Trp Ala
        35                  40                  45

Asn Thr Tyr Asn Val Ala Asn Gly Tyr Ser Asp Gly Thr Phe Lys Pro
    50                  55                  60

Ser Asn Ala Ile Thr Arg Ala Glu Phe Ala Gln Leu Val Ser Arg Val
65                  70                  75                  80

Ile Gly Glu Ala Leu Val Arg Ser Glu Ile Asn Phe Lys Asp Val Lys
                85                  90                  95

Glu Asn Asp Trp Phe Tyr Ser Ala Val Lys Asn Leu Ala Asp Tyr Ile
            100                 105                 110

Ser Gly Tyr Pro Asp Gly Thr Phe Lys Pro Lys Asn Ser Ile Thr Arg
        115                 120                 125

Glu Glu Ala Ala Cys Ile Leu Ala Lys Val Phe Gly Ile Asp Lys Ser
    130                 135                 140
```

-continued

```
Gln Ser Asn Val Leu Ser Lys Phe Ser Asp Tyr Lys Gln Val Ser Glu
145                 150                 155                 160

Trp Ala Lys Glu Tyr Leu Ala Ala Met Val Glu Asn Gly Tyr Ile Asn
                165                 170                 175

Gly Tyr Lys Asp Lys Thr Leu Arg Pro Lys Asn Tyr Ile Thr Lys Ala
            180                 185                 190

Glu Ala Leu Thr Ile Leu Asp Asn Ile Val Gly Leu Leu Leu Ser Lys
        195                 200                 205

Lys Gly Met Tyr Val Gly Arg Glu Val Lys Gly Asn Leu Ile Ile Ser
    210                 215                 220

Ala Pro Tyr Val Ser Val Ser Gly Phe Asn Val Ser Lys Asn Cys Leu
225                 230                 235                 240

Ile Thr Glu Gly Val Lys Asp Gly Asn Val Thr Ile Thr Asn Thr Gln
                245                 250                 255

Val Asp Gly Asn Ile Ile Val Arg Gly Gly Gly Glu His Ser Val Val
            260                 265                 270

Leu Lys Asn Val Lys Ala Ser Ala Val Ile Val Val Asn Lys Gln Ala
        275                 280                 285

Ala Thr Lys Val Asn Ile Ser Gly Asn Ser Lys Ile Asp Lys Val Val
    290                 295                 300

Ile Glu Lys Pro Ala Asn Ile Leu Val Glu Lys Thr Ala Glu Val Asn
305                 310                 315                 320

Thr Val Asp Ile Arg Ala Asp Arg Thr Ile Phe Lys Ala Glu Gly Lys
                325                 330                 335

Ile Glu Arg Ile Ala Val Asn Ala Lys Asp Val Lys Val Asn Asp Lys
            340                 345                 350

Glu Val Gln Gln Gly Ala Ile Leu Thr Ile Thr Asn Thr Asn Glu Asn
        355                 360                 365

Glu Lys Lys Thr Gln Ser Thr Asn Thr Asn Thr Ser Gln Glu Ser Thr
    370                 375                 380

Gln Asn Glu Thr Thr Ser Gln Thr Ser Asn Asn Thr Ser Ser Val Gln
385                 390                 395                 400

Thr Asn Thr Gly Thr Ser Thr Gly Asn Ile Ser Ser Gly Thr Ser Thr
                405                 410                 415

Gly Gly Tyr Thr Gly Gly Gly Ser Val Tyr Ser Gly Gly Ser Ser Gly
            420                 425                 430

Ser Gly Thr Ser Val Gln Tyr Gly Pro Val Ser Arg Val Glu Val Asp
        435                 440                 445

Lys Glu Thr Ile Pro Ile Gly Gln Asn Val Gln Val Thr Val Thr Val
    450                 455                 460

Lys Asp Ala Ala Gly Asn Leu Leu Pro Asn Lys Val Val Arg Ile Glu
465                 470                 475                 480

Gly Gln Ala Ala Tyr Thr Asn Ser Leu Gly Ser Ala Thr Phe Thr Leu
                485                 490                 495

Ser Val Gln Asp Ser Lys Glu Ile Val Ile Asn Val Asp Gly Asn Asp
            500                 505                 510

Tyr Tyr Gly Leu Leu Tyr Ala Ile Lys Pro Thr Glu Gly Val Leu Thr
        515                 520                 525

Phe Lys Leu Lys Thr Ala Ser Gly Asn Phe Leu Ser Gly Phe Asn Val
    530                 535                 540

Lys Leu Val Asn Gln Ala Lys Gln Phe Ser Glu Leu Lys Ser Ala Thr
```

-continued

```
545                 550                 555                 560

Ala Glu Gln Val Ser Phe Ile Val Pro Ala Val Asp Gly Tyr Lys Ala
                565                 570                 575

Leu Ile Trp Ser Tyr Asp Ser Gln Asn Gly Leu Ile Tyr Thr Ile Leu
            580                 585                 590

Asp Asn Leu Ser Val Gly Asn Gly Ile Val Arg Leu Val Ala Asp Thr
        595                 600                 605

Thr Leu Pro Asn Tyr Val Gln Ala Thr Leu Asp Phe Ser Ile Asn Asn
    610                 615                 620

Gln Pro Leu Asn Asn Tyr Glu Phe Ser Ile Ile Asn Ser Ser Lys Ser
625                 630                 635                 640

Asn Thr Phe Ser Ile Asp Asp Val Lys Leu Asn Ile Ser Ser Asn Gln
                645                 650                 655

Leu Lys Ile Thr Ala Asp Ala Gly Ser Tyr Ser Ile Lys Val Ala Lys
            660                 665                 670

Asp Thr Gln Asp Gly Lys Leu Tyr Phe Met Arg Asp Phe Asn Leu Gln
        675                 680                 685

Gln Ser Gly Gln Ala Phe Ser Phe Asn Phe Ser Ser Phe Lys Arg Val
    690                 695                 700

Ile Phe Asn Phe Ser Gly Ile Gly Thr Leu Gln Lys Ser Val Ser Val
705                 710                 715                 720

Lys Leu Asn Gly Ser Trp Phe Glu Leu Ser Gly Glu Asn Asp Ile Tyr
                725                 730                 735

Leu Gln Arg Gly Val Tyr Asn Leu Glu Glu Val Phe Ile Lys Thr Tyr
            740                 745                 750

Asp Glu Asn Ser Lys Glu Val Asp Tyr Ser Tyr Lys Val Pro Ala Gly
        755                 760                 765

Ser Ser Pro Ile Val Val Asp Ala Thr Gly Ser Asn Glu Glu Cys Ser
    770                 775                 780

Val Asn Val Asp Leu Ser Ile Asp Asn Val Ser Ser Ser Val Tyr Ala
785                 790                 795                 800

Thr Asn Ile Asp Gly Asn Glu Phe Ala Ser Ile Lys Ala Gly Ser Leu
                805                 810                 815

Ile Asp Phe Val Val Val Leu Lys Thr Lys Ser Gly Leu Asn Leu Ala
            820                 825                 830

Leu Leu Lys Pro Ser Glu Lys Met Ser Asp Thr Ala Pro Leu Phe Lys
        835                 840                 845

Ala Ser Asp Ile Val Ala Tyr Ile Ser Ser Gly Asn Val Gln Asn Asn
    850                 855                 860

Ile Glu Leu Leu Cys Thr Ser Val Glu Asp Asn Asn Ile Ser His Val
865                 870                 875                 880

Leu Gly Tyr Leu Pro Arg Asn Leu Gln Asp Gly Thr Val Ser Val Glu
                885                 890                 895

Phe Thr Ala Glu Ile Gly Leu Leu Tyr Gly Ser Ser Ile Thr Ser Ser
            900                 905                 910

Ile Ser Leu Ser Ile Asp Asn Gln Asn Gly Asp Ser Arg Gly Tyr Val
        915                 920                 925

Tyr Pro Gln Asn Thr Thr Thr Ser Ala Val Tyr Phe Val Cys Asp Lys
    930                 935                 940

Ile Asp Ser Gly Ala Ser Asp Phe Leu Ala Leu Ser Leu Pro Ala Ser
945                 950                 955                 960
```

-continued

```
Glu Thr Cys Asp Val Asn Leu Ile Ala Asn Lys Ile Tyr Phe Glu Leu
                965                 970                 975

Met Thr Glu Arg Pro Leu Thr Phe Glu Tyr Gly Leu Asn Gly Met Phe
            980                 985                 990

Gly Leu Ser Phe Asn Leu Thr Ser  Glu Asn Leu Tyr Met  Ile Trp Pro
        995                 1000                 1005

Val Tyr  Leu Phe Ser Lys Asp  Glu Gly Leu Thr Arg  Arg Gln Ala
    1010                 1015                 1020

Val Glu  Gln Lys Ser Leu Gln  Ile Val Ser Thr Val  Val Asp Gln
1025                 1030                 1035

Thr Tyr  Asn Asp Tyr Asp Lys  Val Leu Gly Leu His  Asp Trp Leu
1040                 1045                 1050

Val Leu  His Thr Gln Tyr Asp  Leu Glu Gly Tyr Leu  Asn Asn Asn
1055                 1060                 1065

Ile Pro  Tyr Glu Ser His Thr  Ala Tyr Gly Ala Leu  Ile Asn Gly
1070                 1075                 1080

Ile Ala  Val Cys Asn Gly Tyr  Ala Thr Ala Met Leu  Ala Leu Leu
1085                 1090                 1095

Glu Asp  Ala Gly Ile Asp Ser  Lys Glu Ile Tyr Gly  Met Ala Gly
1100                 1105                 1110

Val Gly  Asn Ser Lys Glu Ala  His Ala Trp Asn Met  Val Ser Leu
1115                 1120                 1125

Glu Asn  Asn Trp Tyr His Leu  Asp Ala Thr Trp Asp  Asp Pro Asp
1130                 1135                 1140

Trp Gly  Asn Tyr Val Glu His  Ala Phe Phe Asn Val  Pro Asp Ser
1145                 1150                 1155

Lys Ile  Glu Leu Thr His Asp  Trp Glu Arg Asn Leu  Tyr Pro Ala
1160                 1165                 1170

Ala Thr  Ala Ile Asp Tyr Ser  Tyr Gly Asn Tyr Tyr  Asn Val Glu
1175                 1180                 1185

Thr Ile  Pro Gln Asp Val Tyr  Tyr Asn Glu Asp Asn  Ile Val Thr
1190                 1195                 1200

Ile Val  Val Lys Asn Tyr Lys  Gly Glu Leu Gln Ala  Asn Lys Leu
1205                 1210                 1215

Ile Thr  Ile Thr Lys Phe Asn  Gln Ala Gly Val Glu  Asp Ile Val
1220                 1225                 1230

Phe Thr  Gly Tyr Thr Ser Ala  Asp Gly Thr Ile Thr  Phe Asn Leu
1235                 1240                 1245

Lys Pro  Thr Glu Met Ile Ser  Tyr Gln Ile Tyr Ile  Thr Ser Phe
1250                 1255                 1260

Met Glu  Asp Lys Gly Tyr Ile  Lys Val Val Glu Lys  Leu Lys Pro
1265                 1270                 1275

Val Thr  Leu Ser Leu Asn Ile  Asp Asn Gln Pro Ile  Asp Glu Phe
1280                 1285                 1290

Tyr Ile  Ser Asp Gly Leu Asn  Phe Leu Lys Val Thr  Gly Gly Gln
1295                 1300                 1305

Arg Gln  Val Gly Leu Ser Arg  Trp Arg Glu Asn Asp  Leu Ile Phe
1310                 1315                 1320

Tyr Gly  Leu Gly Phe Val Ile  Lys Lys Ser Leu Thr  Phe Asp Ser
1325                 1330                 1335
```

-continued

```
Tyr Asp  Pro Ile Gln Leu Thr  Val Cys Asn Asp Ala  Tyr Asp Ser
1340                 1345                 1350

Tyr Val  Leu Leu Ser Val Tyr  Asn Ser Ala Tyr Ala  Val Val Gly
1355                 1360                 1365

Ala Asp  Val Tyr Val Ile Asp  Lys Asp Thr Phe Lys  Glu Leu Phe
1370                 1375                 1380

Val Gly  Ser Thr Asp Ser Glu  Gly Asn Leu Pro Leu  Val Leu Thr
1385                 1390                 1395

Asn Gly  Glu Tyr Ile Leu Lys  Val Ala Asn Tyr Asn  Ser Glu Thr
1400                 1405                 1410

Asn Ser  Gln Asp Phe Tyr Tyr  Ser Ile Leu Gln Val  Thr Gln Asp
1415                 1420                 1425

Ser Val  Thr Ser Ile Asp Leu  Ser Gln Phe Asn Gln  Ile Gln Tyr
1430                 1435                 1440

Ile Pro  Gly Phe Thr Glu Gly  Gly Val Gly Met Gly  Asp Lys Leu
1445                 1450                 1455

Leu Leu  Gly Phe Asp Phe Asp  Tyr Asp Gly Lys Phe  Val Val Ser
1460                 1465                 1470

Ser Met  Gly His Ser Arg Lys  Ala Tyr Phe Ala Pro  Gly Asp Tyr
1475                 1480                 1485

Gly Trp  Val Trp Ile Leu Ala  His Asn Pro Tyr Asp  Ile Glu Asp
1490                 1495                 1500

Asn Val  Ala Thr Ser Leu Val  Tyr Lys Pro Leu Gln  Lys Leu Ile
1505                 1510                 1515

Ile Pro  Glu Gly Thr Leu Gln  Gln Tyr Thr Val Asp  Leu Arg Val
1520                 1525                 1530

Asp Arg  Ser Lys Ile Val Leu  Asn Cys Glu Arg Arg  Pro Glu Gly
1535                 1540                 1545

Ser Trp  Asp Phe Leu Pro Thr  Thr Leu Asp Met Val  Tyr Glu Lys
1550                 1555                 1560

Asp Asp  Ile Ile Ile Asp Ile  Tyr Ile Pro Thr Asn  Ser Gln Ser
1565                 1570                 1575

Tyr Val  Val Gly Gly Asp Thr  Val Pro Gln Gln Leu  Ile Glu Asn
1580                 1585                 1590

Asn Gly  Asp Met His Ile Met  Trp Met Gly Ile Ala  Gly Ile Tyr
1595                 1600                 1605

Pro Ile  Arg Thr Tyr Thr Ile  Ser Asn Ser Thr Pro  Tyr Glu Leu
1610                 1615                 1620

Phe Thr  Gly His Leu Gly Phe  Ile Asp Asn Thr Asn  Ser Trp Arg
1625                 1630                 1635

Ile Ile  Val Leu Asp Val Val  Gly Asp Leu Asn Ser  Thr Phe Arg
1640                 1645                 1650

Ile Val  Leu Pro Leu Ser Pro  Tyr Asp Ile Asp Val  Ile Ile Asp
1655                 1660                 1665

Lys Glu  Ile Thr Ile Tyr Pro  Leu Thr Ala Gly Ser  Gly Gly Leu
1670                 1675                 1680

Arg Lys  Glu Ser Tyr Ile Gln  Lys Ala Lys Ile Tyr  Asn Gln Ile
1685                 1690                 1695

Ile Asn  Thr Gln Leu Ile Lys  Lys Glu Asn Lys Lys
1700                 1705                 1710
```

-continued

<210> SEQ ID NO 56
<211> LENGTH: 660
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 56

gtgacagact tatatataac cattttattt gtctcgataa ttatagcatt tgggattgaa 60

attttacaaa gtaaaacctt tcataaaagt tgtttaacag gagacaaaaa tttatttctg 120

ctcataggga tgactttttt aggaattcat ttgtttataa catttccggg catgatttcg 180

ataaagatat ggatgatgct aaatcttatt gtattgtgga tcttgtcata ttacagagaa 240

actataaaaa aagaaatatt attcaagaat atttctaaac ttttaaaatc tcaaagaaga 300

aaaatatttt ttataagtag catattgtgt tacataattg tctgtcttct tttaaattat 360

tacatagttt acaagttcat atattcatct gcagtttcaa aattaacaat aacattcact 420

atttcctttc ttatactaac agaagctaaa aatatagctt gttttattac aattttaaaa 480

ttttattgca tatctttggt aacgatttgg ggagaaaatt tagtgaaata tgagggatgg 540

cttattggag caacaagaga ttattacata ttaaaagaaa aattttcggg gaaattaata 600

acagtaaaaa aagatatagt taatcaaata gtgataatgg gaaaagtttt tgaaagataa 660

<210> SEQ ID NO 57
<211> LENGTH: 219
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 57

Val Thr Asp Leu Tyr Ile Thr Ile Leu Phe Val Ser Ile Ile Ile Ala
1 5 10 15

Phe Gly Ile Glu Ile Leu Gln Ser Lys Thr Phe His Lys Ser Cys Leu
20 25 30

Thr Gly Asp Lys Asn Leu Phe Leu Leu Ile Gly Met Thr Phe Leu Gly
35 40 45

Ile His Leu Phe Ile Thr Phe Pro Gly Met Ile Ser Ile Lys Ile Trp
50 55 60

Met Met Leu Asn Leu Ile Val Leu Trp Ile Leu Ser Tyr Tyr Arg Glu
65 70 75 80

Thr Ile Lys Lys Glu Ile Leu Phe Lys Asn Ile Ser Lys Leu Leu Lys
85 90 95

Ser Gln Arg Arg Lys Ile Phe Phe Ile Ser Ser Ile Leu Cys Tyr Ile
100 105 110

Ile Val Cys Leu Leu Leu Asn Tyr Tyr Ile Val Tyr Lys Phe Ile Tyr
115 120 125

Ser Ser Ala Val Ser Lys Leu Thr Ile Thr Phe Thr Ile Ser Phe Leu
130 135 140

Ile Leu Thr Glu Ala Lys Asn Ile Ala Cys Phe Ile Thr Ile Leu Lys
145 150 155 160

Phe Tyr Cys Ile Ser Leu Val Thr Ile Trp Gly Glu Asn Leu Val Lys
165 170 175

Tyr Glu Gly Trp Leu Ile Gly Ala Thr Arg Asp Tyr Tyr Ile Leu Lys
180 185 190

Glu Lys Phe Ser Gly Lys Leu Ile Thr Val Lys Lys Asp Ile Val Asn
195 200 205

-continued

```
Gln Ile Val Ile Met Gly Lys Val Phe Glu Arg
    210                 215


<210> SEQ ID NO 58
<211> LENGTH: 1074
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 58

atgaggttta aaaagttttt aaaagttttg atagctgttt taatgtgctt tatgcttgga     60

aatccgtttt atgcgcaagc tgcaataacc ctcacatcca atgctagtgg aacttatgat    120

ggctattact atgaattgtg gaaggattct ggaaatacaa caatgacagt tgacacagga    180

ggaagattta gctgtcaatg gagtaacatt aacaatgcac ttttcagaac aggtaaaaaa    240

ttcagtacag catggaacca gctcggaaca gtgaaaatca cctactctgc tacctacaat    300

ccaaatggta attcatatct atgtatctat ggatggtcaa gaaatccact ggttgaattt    360

tacattgtcg aaagctgggg cacatggcgt ccgcccgggg caacatcatt gggtactgta    420

acgattgatg gaggaacata cgatatttac aagacaactc gtgtaaatca accatctata    480

gaagggacaa ctacatttga tcagtattgg agtgttagaa catcaaaaag aacaagtggc    540

actgttactg tgactgatca ttttaaagca tgggctgcaa aaggcctgaa cttgggtaca    600

attgaccaga ttacactctg tgtagaaggt taccagagca gtggttcagc taatataaca    660

caaaatacat tttctataac aagtgcctct tcaggtggaa caacacccac taccacaaag    720

gtagagtgtg aaaatatgtc actgagtggg ccgtatgcat caaaaattac aagtccattt    780

tacggcatgg ctctttatgc aaacggagat aaagcaacaa caaatataaa cttttcagca    840

agccgtaact atacttttaa attacgagga tgtggaaata acaataattt agcatcagtt    900

gatttactaa tagatgggaa gaaagtaggt tcgttctatt atcgggggac atatccttgg    960

gaagctccta tagagaatgt gtatgtgagt gcaggttcgc acaaagtgga aattgtagtt   1020

tctgctgata atggtacatg ggatgtttat gcagattatt tgttaataca atga         1074


<210> SEQ ID NO 59
<211> LENGTH: 357
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 59

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

Phe Met Leu Gly Asn Pro Phe Tyr Ala Gln Ala Ala Ile Thr Leu Thr
            20                  25                  30

Ser Asn Ala Ser Gly Thr Tyr Asp Gly Tyr Tyr Tyr Glu Leu Trp Lys
        35                  40                  45

Asp Ser Gly Asn Thr Thr Met Thr Val Asp Thr Gly Gly Arg Phe Ser
    50                  55                  60

Cys Gln Trp Ser Asn Ile Asn Asn Ala Leu Phe Arg Thr Gly Lys Lys
65                  70                  75                  80

Phe Ser Thr Ala Trp Asn Gln Leu Gly Thr Val Lys Ile Thr Tyr Ser
                85                  90                  95

Ala Thr Tyr Asn Pro Asn Gly Asn Ser Tyr Leu Cys Ile Tyr Gly Trp
            100                 105                 110
```

-continued

```
Ser Arg Asn Pro Leu Val Glu Phe Tyr Ile Val Glu Ser Trp Gly Thr
        115                 120                 125

Trp Arg Pro Pro Gly Ala Thr Ser Leu Gly Thr Val Thr Ile Asp Gly
    130                 135                 140

Gly Thr Tyr Asp Ile Tyr Lys Thr Thr Arg Val Asn Gln Pro Ser Ile
145                 150                 155                 160

Glu Gly Thr Thr Thr Phe Asp Gln Tyr Trp Ser Val Arg Thr Ser Lys
                165                 170                 175

Arg Thr Ser Gly Thr Val Thr Val Thr Asp His Phe Lys Ala Trp Ala
            180                 185                 190

Ala Lys Gly Leu Asn Leu Gly Thr Ile Asp Gln Ile Thr Leu Cys Val
        195                 200                 205

Glu Gly Tyr Gln Ser Ser Gly Ser Ala Asn Ile Thr Gln Asn Thr Phe
    210                 215                 220

Ser Ile Thr Ser Ala Ser Ser Gly Gly Thr Thr Pro Thr Thr Thr Lys
225                 230                 235                 240

Val Glu Cys Glu Asn Met Ser Leu Ser Gly Pro Tyr Ala Ser Lys Ile
                245                 250                 255

Thr Ser Pro Phe Tyr Gly Met Ala Leu Tyr Ala Asn Gly Asp Lys Ala
            260                 265                 270

Thr Thr Asn Ile Asn Phe Ser Ala Ser Arg Asn Tyr Thr Phe Lys Leu
        275                 280                 285

Arg Gly Cys Gly Asn Asn Asn Asn Leu Ala Ser Val Asp Leu Leu Ile
    290                 295                 300

Asp Gly Lys Lys Val Gly Ser Phe Tyr Tyr Arg Gly Thr Tyr Pro Trp
305                 310                 315                 320

Glu Ala Pro Ile Glu Asn Val Tyr Val Ser Ala Gly Ser His Lys Val
                325                 330                 335

Glu Ile Val Val Ser Ala Asp Asn Gly Thr Trp Asp Val Tyr Ala Asp
            340                 345                 350

Tyr Leu Leu Ile Gln
        355


<210> SEQ ID NO 60
<211> LENGTH: 363
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 60

atgtttggcg gaagtacaat atccaaatat gttgaaggtt atacaacatt aaagaatatc     60

tatggaacag aaatcaattc ttcgaccatt caatcttgtg ataggaatat tggtcaggct    120

acaatttatg taagccctta tgtaactgca tttacacata ctttaactca gcgaagaggt    180

attttaatgc atgagatggg acacgctatg ggcttaagac atcctaatta ttcggactca    240

agaaatagct attcagctga tagttatgga agtattatgg attacagtta taccgaagaa    300

tacccaacta tacatgatat atgtgacatt gaaataatgt atggttttaa ctgttacaat    360

taa                                                                  363


<210> SEQ ID NO 61
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725
```

-continued

<400> SEQUENCE: 61

Met Phe Gly Gly Ser Thr Ile Ser Lys Tyr Val Glu Gly Tyr Thr Thr
1 5 10 15

Leu Lys Asn Ile Tyr Gly Thr Glu Ile Asn Ser Ser Thr Ile Gln Ser
20 25 30

Cys Asp Arg Asn Ile Gly Gln Ala Thr Ile Tyr Val Ser Pro Tyr Val
35 40 45

Thr Ala Phe Thr His Thr Leu Thr Gln Arg Arg Gly Ile Leu Met His
50 55 60

Glu Met Gly His Ala Met Gly Leu Arg His Pro Asn Tyr Ser Asp Ser
65 70 75 80

Arg Asn Ser Tyr Ser Ala Asp Ser Tyr Gly Ser Ile Met Asp Tyr Ser
85 90 95

Tyr Thr Glu Glu Tyr Pro Thr Ile His Asp Ile Cys Asp Ile Glu Ile
100 105 110

Met Tyr Gly Phe Asn Cys Tyr Asn
115 120

<210> SEQ ID NO 62
<211> LENGTH: 930
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 62

atgtttaaag atttcttaaa agactcattt ttcttcacaa aaagatacta tggatacata 60

tttggtctat tggttatatc atttattctt ggtactttgg cagttgtggt tttactaata 120

cttggtattt tattggctat gttaatgggt gtggatatgt ctacattagg tttattaaat 180

gataaaaatg tttttccatt tttaagtagc aaatttattt tctatatgat tttaatgatt 240

ctacttttta taatgtgggt attgttagta ttagctttta tccaatatcc acttgtaaag 300

acctttattg aaataacacg agaaaaggat atatacaaaa aaccttttga aattttcttt 360

gccggaataa aagaaaagaa tctgatgatg ggcttgaaaa tagttggact gggattttta 420

ttaacactta tagttgggtc aatccttttt ataggtatta taggtatagc ttttagtacc 480

aatactgtcg ataatttagc acgctttgct ttgattataa ttggtattgt tggtatagta 540

ttaggtgttt acctgttatt aaggttaatg tttgctaatg cagcgttggt ggataaagat 600

ataggagtta tagaaagtat taaagagagt ttaaagctta caaaggggaa aatgggattt 660

gtaatagcag cttttgttta tagtattttg gtatccatag ttttccaaat accagtttat 720

attgttgata cattatttaa aacagaaaca tctcaagaga gtatattatt tattatttta 780

agtttgatag gatttgtttt ttaccttata gcagtgcctt actttattgt gttacagtat 840

ctaccgtata ataccttgaa aagccatgtt gaaaatgcaa accatataga tagccagttt 900

tacggaactt ataataatat tattgggtaa 930

<210> SEQ ID NO 63
<211> LENGTH: 309
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 63

Met Phe Lys Asp Phe Leu Lys Asp Ser Phe Phe Phe Thr Lys Arg Tyr
1 5 10 15

-continued

```
Tyr Gly Tyr Ile Phe Gly Leu Leu Val Ile Ser Phe Ile Leu Gly Thr
            20                  25                  30

Leu Ala Val Val Val Leu Leu Ile Leu Gly Ile Leu Leu Ala Met Leu
        35                  40                  45

Met Gly Val Asp Met Ser Thr Leu Gly Leu Leu Asn Asp Lys Asn Val
    50                  55                  60

Phe Pro Phe Leu Ser Ser Lys Phe Ile Phe Tyr Met Ile Leu Met Ile
65                  70                  75                  80

Leu Leu Phe Ile Met Trp Val Leu Leu Val Leu Ala Phe Ile Gln Tyr
                85                  90                  95

Pro Leu Val Lys Thr Phe Ile Glu Ile Thr Arg Glu Lys Asp Ile Tyr
            100                 105                 110

Lys Lys Pro Phe Glu Ile Phe Phe Ala Gly Ile Lys Glu Lys Asn Leu
        115                 120                 125

Met Met Gly Leu Lys Ile Val Gly Leu Gly Phe Leu Leu Thr Leu Ile
    130                 135                 140

Val Gly Ser Ile Leu Phe Ile Gly Ile Ile Gly Ile Ala Phe Ser Thr
145                 150                 155                 160

Asn Thr Val Asp Asn Leu Ala Arg Phe Ala Leu Ile Ile Ile Gly Ile
                165                 170                 175

Val Gly Ile Val Leu Gly Val Tyr Leu Leu Leu Arg Leu Met Phe Ala
            180                 185                 190

Asn Ala Ala Leu Val Asp Lys Asp Ile Gly Val Ile Glu Ser Ile Lys
        195                 200                 205

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

Phe Val Tyr Ser Ile Leu Val Ser Ile Val Phe Gln Ile Pro Val Tyr
225                 230                 235                 240

Ile Val Asp Thr Leu Phe Lys Thr Glu Thr Ser Gln Glu Ser Ile Leu
                245                 250                 255

Phe Ile Ile Leu Ser Leu Ile Gly Phe Val Phe Tyr Leu Ile Ala Val
            260                 265                 270

Pro Tyr Phe Ile Val Leu Gln Tyr Leu Pro Tyr Asn Thr Leu Lys Ser
        275                 280                 285

His Val Glu Asn Ala Asn His Ile Asp Ser Gln Phe Tyr Gly Thr Tyr
    290                 295                 300

Asn Asn Ile Ile Gly
305


<210> SEQ ID NO 64
<211> LENGTH: 807
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 64

atgagtgcaa aagaatttcc aaaaaacgaa ctaataagcg ttataattcc ctgttataac      60

gaggcacaaa atattggaca aacactcaaa gaaatttatg attatttaga tgagtttgta     120

cctaactatg aagtaattgt tgtagttgaa aaaagcacag acaatacttt ggaagtaata     180

aactcaagaa aaaatcaaaa aactatcgta ttagagaata caaaaaaata tggcaaggga     240

tatagtttaa aaagaggtat atattttgca aaaggacaat acatacttac ctgcgatgcg     300
```

-continued

```
gaccttcctg tggacataaa aaaatatttt cttcctatgt tagaactatt aaagagagat    360

gaaaaggttg ctgcagtttt tgccactgcc ttagctataa aaacatgcag gaaagaaaga    420

ggctttgtaa gaagtattgt ttcactgata ttttttgttt ttagacagtt gttattacag    480

tttcctgtaa gtgatacaca attaggtttt aaattgttta gagcagatgt agcaaaaaag    540

ttctgtcaaa aagttaatga gaatgggttt ttgtttgact tgatattaac agatttgatg    600

ttaaatgaag gatatcaaat tgaagaagta aatgtaaaag ttgtagaaag gaaaatcaag    660

tcttctgttt ctgtcgctga aataataaaa actacttata agttttgcaa atatatattg    720

tttacgcgtt caaaattact tagaaaatgt acagataaag ttatgataaa tgataaaagt    780

aaaagcataa aagctacaag tatttga                                        807
```

```
<210> SEQ ID NO 65
<211> LENGTH: 268
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 65

Met Ser Ala Lys Glu Phe Pro Lys Asn Glu Leu Ile Ser Val Ile Ile
1               5                   10                  15

Pro Cys Tyr Asn Glu Ala Gln Asn Ile Gly Gln Thr Leu Lys Glu Ile
            20                  25                  30

Tyr Asp Tyr Leu Asp Glu Phe Val Pro Asn Tyr Glu Val Ile Val Val
        35                  40                  45

Val Glu Lys Ser Thr Asp Asn Thr Leu Glu Val Ile Asn Ser Arg Lys
    50                  55                  60

Asn Gln Lys Thr Ile Val Leu Glu Asn Thr Lys Lys Tyr Gly Lys Gly
65                  70                  75                  80

Tyr Ser Leu Lys Arg Gly Ile Tyr Phe Ala Lys Gly Gln Tyr Ile Leu
                85                  90                  95

Thr Cys Asp Ala Asp Leu Pro Val Asp Ile Lys Lys Tyr Phe Leu Pro
            100                 105                 110

Met Leu Glu Leu Leu Lys Arg Asp Glu Lys Val Ala Ala Val Phe Ala
        115                 120                 125

Thr Ala Leu Ala Ile Lys Thr Cys Arg Lys Glu Arg Gly Phe Val Arg
    130                 135                 140

Ser Ile Val Ser Leu Ile Phe Phe Val Phe Arg Gln Leu Leu Leu Gln
145                 150                 155                 160

Phe Pro Val Ser Asp Thr Gln Leu Gly Phe Lys Leu Phe Arg Ala Asp
                165                 170                 175

Val Ala Lys Lys Phe Cys Gln Lys Val Asn Glu Asn Gly Phe Leu Phe
            180                 185                 190

Asp Leu Ile Leu Thr Asp Leu Met Leu Asn Glu Gly Tyr Gln Ile Glu
        195                 200                 205

Glu Val Asn Val Lys Val Val Glu Arg Lys Ile Lys Ser Ser Val Ser
    210                 215                 220

Val Ala Glu Ile Ile Lys Thr Thr Tyr Lys Phe Cys Lys Tyr Ile Leu
225                 230                 235                 240

Phe Thr Arg Ser Lys Leu Leu Arg Lys Cys Thr Asp Lys Val Met Ile
                245                 250                 255

Asn Asp Lys Ser Lys Ser Ile Lys Ala Thr Ser Ile
            260                 265
```

-continued

<210> SEQ ID NO 66
<211> LENGTH: 1116
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 66

gtgaacgaac gaaataaaga aattgcatat attattcttg acttattcgt tgctgccgtt 60

ccatttttgt atattacata cctttacaga ttaggagtta cagaattaga aagtttaaaa 120

gaagttgtat atgctaattc ccttttaagt acactttttt atatttctaa catagttatg 180

tttctcaaaa tagaatataa tattagcttc tacgcaactg caaaagtaat aagtgtaatt 240

tcattaattt taggaattat attgcaagtc ttatatccat cttatggatg ggcaataagt 300

atacttttat taatggacag ttttgtagtt tttctcagag gaactatgtt gagagagaaa 360

aattacatga aagcggcttt aattaatttt ctgttgcaca caagtagatt tgggatatta 420

atactgtttc ctttagatgg tcttttgaca agatggattg tatctaatat cggtggatat 480

ttgctggtat tggcaatgta tttatctttg ataaagtgga tttttgaaag agaagataaa 540

aatagcatta tgagtttaaa catagatgaa attagaaaaa gagcttatta tttacttttt 600

gtaatgataa tttttgtaat tcaaaatatt gatctattag cactaaaggg aatgaatttt 660

tataacttgc ttgttttgtc aaggccgtgg ggcttagttg cttatgttgt tacagtatca 720

attggaaatt tagtcttgat taataaaaaa acatatccct attggattta tttaactgta 780

gcatgggctg gaattttcat tggatatatt tctttgggta aacacataaa ttactatctt 840

ttcaaaattc catattcgtt tgatttaagt atattaccaa ttttatttta ccttttaatg 900

gctacgatat tatttacaaa ttacttaaat tatgcaaaaa agaagcagtt aataattgaa 960

cctttgtgct tattcctatt aattgaagtc ttaaaaaata aaatctctct aatgagagaa 1020

ttgaaatttg aggtattaat tttaatttta acatttttga taattgttca cttgttttta 1080

atttttttat attcaactta taacaaatgt atgtag 1116

<210> SEQ ID NO 67
<211> LENGTH: 371
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 67

Val Asn Glu Arg Asn Lys Glu Ile Ala Tyr Ile Ile Leu Asp Leu Phe
1 5 10 15

Val Ala Ala Val Pro Phe Leu Tyr Ile Thr Tyr Leu Tyr Arg Leu Gly
20 25 30

Val Thr Glu Leu Glu Ser Leu Lys Glu Val Val Tyr Ala Asn Ser Leu
35 40 45

Leu Ser Thr Leu Phe Tyr Ile Ser Asn Ile Val Met Phe Leu Lys Ile
50 55 60

Glu Tyr Asn Ile Ser Phe Tyr Ala Thr Ala Lys Val Ile Ser Val Ile
65 70 75 80

Ser Leu Ile Leu Gly Ile Ile Leu Gln Val Leu Tyr Pro Ser Tyr Gly
85 90 95

Trp Ala Ile Ser Ile Leu Leu Leu Met Asp Ser Phe Val Val Phe Leu
100 105 110

-continued

```
Arg Gly Thr Met Leu Arg Glu Lys Asn Tyr Met Lys Ala Ala Leu Ile
        115                 120                 125

Asn Phe Leu Leu His Thr Ser Arg Phe Gly Ile Leu Ile Leu Phe Pro
    130                 135                 140

Leu Asp Gly Leu Leu Thr Arg Trp Ile Val Ser Asn Ile Gly Gly Tyr
145                 150                 155                 160

Leu Leu Val Leu Ala Met Tyr Leu Ser Leu Ile Lys Trp Ile Phe Glu
                165                 170                 175

Arg Glu Asp Lys Asn Ser Ile Met Ser Leu Asn Ile Asp Glu Ile Arg
            180                 185                 190

Lys Arg Ala Tyr Tyr Leu Leu Phe Val Met Ile Ile Phe Val Ile Gln
        195                 200                 205

Asn Ile Asp Leu Leu Ala Leu Lys Gly Met Asn Phe Tyr Asn Leu Leu
    210                 215                 220

Val Leu Ser Arg Pro Trp Gly Leu Val Ala Tyr Val Val Thr Val Ser
225                 230                 235                 240

Ile Gly Asn Leu Val Leu Ile Asn Lys Lys Thr Tyr Pro Tyr Trp Ile
                245                 250                 255

Tyr Leu Thr Val Ala Trp Ala Gly Ile Phe Ile Gly Tyr Ile Ser Leu
            260                 265                 270

Gly Lys His Ile Asn Tyr Tyr Leu Phe Lys Ile Pro Tyr Ser Phe Asp
        275                 280                 285

Leu Ser Ile Leu Pro Ile Leu Phe Tyr Leu Leu Met Ala Thr Ile Leu
    290                 295                 300

Phe Thr Asn Tyr Leu Asn Tyr Ala Lys Lys Lys Gln Leu Ile Ile Glu
305                 310                 315                 320

Pro Leu Cys Leu Phe Leu Leu Ile Glu Val Leu Lys Asn Lys Ile Ser
                325                 330                 335

Leu Met Arg Glu Leu Lys Phe Glu Val Leu Ile Leu Ile Leu Thr Phe
            340                 345                 350

Leu Ile Ile Val His Leu Phe Leu Ile Phe Leu Tyr Ser Thr Tyr Asn
        355                 360                 365

Lys Cys Met
    370
```

<210> SEQ ID NO 68
<211> LENGTH: 1080
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 68

```
atgaagacta ttttagtgtt aacgcaaagg gatatatacc ataaaaaagc tggtggtgca     60

gaaagatatt tgtttaatgt gctaaaaggg ttgagtgatc attacaaaat cacttgtctg    120

tgtcaaaatg atggtactca gaaagattat gagatatatg ataatataac gtttattcga    180

attaaaacta atttaattag cctaattttt aaagcaatgt tctattataa aagaaataag    240

gaaaacattg acttaataat tgaccataca aatacacatc agtttttac tttttatac      300

gttccgaaaa ataagagact attaatagtt caccagcttg cacttgagat ttgggaatat    360

tattttccaa agtacattgg aaaagcgcta aaattacttg aaaaacttct ttggcgtcta    420

tcatccggaa tggcagtaac agttagccgg tcaactaagg aggatttaca gaggtttggt    480

tttaaagaag aatttatatg gattgtaaaa aattcattaa agcataaata tacttcactt    540
```

-continued

```
ccacatatgg agaaagaaga ttttcttgtt agtgtcggga gattagttcc atataaaaga    600

tttgaagatg caatttattt agcaaaaaaa gttaacaaaa aaatatttat tataggagaa    660

ggtcaggaaa attataaaag aaaattaaga aattatgcaa aaaaaattaa cgcagatgtg    720

atttttactg gatatatttc tgaagagaaa aaacaagaca tagtagagaa agcctatatg    780

cacatttttc catctataag agaaggttgg ggacttgtaa taagcgaagc agctaactta    840

gggacacctt ccttagtata tcctgttcct ggctgcctgg atgcagtaaa ttacggaaaa    900

gctggttttg taacaaagag aataggcaaa gaacatttgt tggaaagatt tctaactatt    960

aatagggaag aatatgaaag gatgagaatt tgtgcatttg aatttacttc acagctaaac   1020

tataataaac aatgtgaaga gtttcaacag gttatacaaa gtataataga aaatacatag   1080
```

<210> SEQ ID NO 69
<211> LENGTH: 359
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 69

```
Met Lys Thr Ile Leu Val Leu Thr Gln Arg Asp Ile Tyr His Lys Lys
1               5                   10                  15

Ala Gly Gly Ala Glu Arg Tyr Leu Phe Asn Val Leu Lys Gly Leu Ser
            20                  25                  30

Asp His Tyr Lys Ile Thr Cys Leu Cys Gln Asn Asp Gly Thr Gln Lys
        35                  40                  45

Asp Tyr Glu Ile Tyr Asp Asn Ile Thr Phe Ile Arg Ile Lys Thr Asn
    50                  55                  60

Leu Ile Ser Leu Ile Phe Lys Ala Met Phe Tyr Tyr Lys Arg Asn Lys
65                  70                  75                  80

Glu Asn Ile Asp Leu Ile Ile Asp His Thr Asn Thr His Gln Phe Phe
                85                  90                  95

Thr Phe Leu Tyr Val Pro Lys Asn Lys Arg Leu Leu Ile Val His Gln
            100                 105                 110

Leu Ala Leu Glu Ile Trp Glu Tyr Tyr Phe Pro Lys Tyr Ile Gly Lys
        115                 120                 125

Ala Leu Lys Leu Leu Glu Lys Leu Leu Trp Arg Leu Ser Ser Gly Met
    130                 135                 140

Ala Val Thr Val Ser Arg Ser Thr Lys Glu Asp Leu Gln Arg Phe Gly
145                 150                 155                 160

Phe Lys Glu Glu Phe Ile Trp Ile Val Lys Asn Ser Leu Lys His Lys
                165                 170                 175

Tyr Thr Ser Leu Pro His Met Glu Lys Glu Asp Phe Leu Val Ser Val
            180                 185                 190

Gly Arg Leu Val Pro Tyr Lys Arg Phe Glu Asp Ala Ile Tyr Leu Ala
        195                 200                 205

Lys Lys Val Asn Lys Lys Ile Phe Ile Ile Gly Glu Gly Gln Glu Asn
    210                 215                 220

Tyr Lys Arg Lys Leu Arg Asn Tyr Ala Lys Lys Ile Asn Ala Asp Val
225                 230                 235                 240

Ile Phe Thr Gly Tyr Ile Ser Glu Glu Lys Lys Gln Asp Ile Val Glu
                245                 250                 255

Lys Ala Tyr Met His Ile Phe Pro Ser Ile Arg Glu Gly Trp Gly Leu
```

-continued

```
            260                 265                 270

Val Ile Ser Glu Ala Ala Asn Leu Gly Thr Pro Ser Leu Val Tyr Pro
        275                 280                 285

Val Pro Gly Cys Leu Asp Ala Val Asn Tyr Gly Lys Ala Gly Phe Val
    290                 295                 300

Thr Lys Arg Ile Gly Lys Glu His Leu Leu Glu Arg Phe Leu Thr Ile
305                 310                 315                 320

Asn Arg Glu Glu Tyr Glu Arg Met Arg Ile Cys Ala Phe Glu Phe Thr
                325                 330                 335

Ser Gln Leu Asn Tyr Asn Lys Gln Cys Glu Glu Phe Gln Gln Val Ile
            340                 345                 350

Gln Ser Ile Ile Glu Asn Thr
        355


<210> SEQ ID NO 70
<211> LENGTH: 852
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 70

atgaagctgg gaattttgat aacaacatat aatgatgact atattatttt aagatgctta     60

gattctatct ataaccaact tgatgagata gattttccaa tttatgttgt atgtgttgat    120

gatggttcag atttgcctct tacatatcct cattttgata ttctaagaac tgaacatagg    180

ggaagaagct atgcgagaat tgagggactg aaaaaaattt tagctgaaaa ttgcacacat    240

ttcttatttt tagatagcga tatggtcctt ccaccaggct ttctcaaaaa gttgaagaca    300

gtagttgaaa attacgatag tgatgctttc attatccctg aagtggcttt tagtagttat    360

aacaattttt ggacaaaagt taaggtcttt gaaagaaatt tatatagagt aagttattgc    420

aaagaaagtg gaaatattga agctgccaga ttatggaaaa caaatgcatt tccgggtttt    480

gttgaaggat tagaagcatt tgaagagatt cagccaacaa tattgggtgt taaaaaaggt    540

ttaaagattt taaaaatcca agagattttt attcttcatg atgaaaagaa ggtaactttc    600

aaagacttaa taagaaaaaa gaatacttac ttttgttgta tgattggttc tgaaaaatgt    660

tcaaagtggg agataatgaa aaggtattat ttctttcgtc cccatcttta ccataaggaa    720

aatttaaaga aatacataag acatcccatt ttagctattg gagttgtgtt tatgtacctt    780

gtattgacat tgaattttat atggacgagt atatcaataa atttaattag aaaaggattg    840

atggaaaaat ga                                                        852


<210> SEQ ID NO 71
<211> LENGTH: 283
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 71

Met Lys Leu Gly Ile Leu Ile Thr Thr Tyr Asn Asp Asp Tyr Ile Ile
1               5                   10                  15

Leu Arg Cys Leu Asp Ser Ile Tyr Asn Gln Leu Asp Glu Ile Asp Phe
            20                  25                  30

Pro Ile Tyr Val Val Cys Val Asp Asp Gly Ser Asp Leu Pro Leu Thr
        35                  40                  45

Tyr Pro His Phe Asp Ile Leu Arg Thr Glu His Arg Gly Arg Ser Tyr
```

-continued

```
    50                  55                  60

Ala Arg Ile Glu Gly Leu Lys Lys Ile Leu Ala Glu Asn Cys Thr His
65                  70                  75                  80

Phe Leu Phe Leu Asp Ser Asp Met Val Leu Pro Pro Gly Phe Leu Lys
                85                  90                  95

Lys Leu Lys Thr Val Val Glu Asn Tyr Asp Ser Asp Ala Phe Ile Ile
            100                 105                 110

Pro Glu Val Ala Phe Ser Ser Tyr Asn Asn Phe Trp Thr Lys Val Lys
        115                 120                 125

Val Phe Glu Arg Asn Leu Tyr Arg Val Ser Tyr Cys Lys Glu Ser Gly
    130                 135                 140

Asn Ile Glu Ala Ala Arg Leu Trp Lys Thr Asn Ala Phe Pro Gly Phe
145                 150                 155                 160

Val Glu Gly Leu Glu Ala Phe Glu Glu Ile Gln Pro Thr Ile Leu Gly
                165                 170                 175

Val Lys Lys Gly Leu Lys Ile Leu Lys Ile Gln Glu Ile Phe Ile Leu
            180                 185                 190

His Asp Glu Lys Lys Val Thr Phe Lys Asp Leu Ile Arg Lys Lys Asn
        195                 200                 205

Thr Tyr Phe Cys Cys Met Ile Gly Ser Glu Lys Cys Ser Lys Trp Glu
    210                 215                 220

Ile Met Lys Arg Tyr Tyr Phe Phe Arg Pro His Leu Tyr His Lys Glu
225                 230                 235                 240

Asn Leu Lys Lys Tyr Ile Arg His Pro Ile Leu Ala Ile Gly Val Val
                245                 250                 255

Phe Met Tyr Leu Val Leu Thr Leu Asn Phe Ile Trp Thr Ser Ile Ser
            260                 265                 270

Ile Asn Leu Ile Arg Lys Gly Leu Met Glu Lys
        275                 280


<210> SEQ ID NO 72
<211> LENGTH: 237
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 72

atgagaagaa gggggagaaa tgttctatat aaggtaagac gaagggaaaa ttatgcaaaa      60

atagttataa tagctgtatt aatagggtgc cttattttat taggatggca tacggttgta     120

aagtcaagta attggtttaa atcaattatt atcgagaaca tggactacaa ctatggaaaa     180

gattcaaata gtatccagcc ttatgcacct aaaaaaacac atcctcgagg agaataa        237


<210> SEQ ID NO 73
<211> LENGTH: 78
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 73

Met Arg Arg Arg Gly Arg Asn Val Leu Tyr Lys Val Arg Arg Arg Glu
1               5                   10                  15

Asn Tyr Ala Lys Ile Val Ile Ile Ala Val Leu Ile Gly Cys Leu Ile
            20                  25                  30

Leu Leu Gly Trp His Thr Val Val Lys Ser Ser Asn Trp Phe Lys Ser
        35                  40                  45
```

-continued

```
Ile Ile Ile Glu Asn Met Asp Tyr Asn Tyr Gly Lys Asp Ser Asn Ser
    50                  55                  60

Ile Gln Pro Tyr Ala Pro Lys Lys Thr His Pro Arg Gly Glu
65                  70                  75


<210> SEQ ID NO 74
<211> LENGTH: 4224
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 74

atgaaacaag aagagatatt gcaaagacaa aggaagatag caaagactat ttcgttgatt     60

gctgtagtag ttataatcct catagataca aaaaatattt ggttgaataa tggctttatt    120

attttcagtg atcttgactt tagtggaatt aatgataaaa ggtatattga gagaatatgg    180

ggtgttttta atccgcattt ttcttcaacc aatttttta atttatctcg tttgtttttt    240

atatttcctt tttatgccct aaactttttg gtctcttctt tttacaatca agcaatgctt    300

aagttgataa ttttgtcttg tttactaata tcaggaatag gtatgttttt attatgtgag    360

tatctattga taaaaacctt cagaggtttt cctgattatg tgcattattt tggattaata    420

ataccggctg tgtattatgc attaaatcca tgggttatat tcagaattca gcatatattt    480

cttttgagcg ggtacagtgt atttccgctt ataataaaag aatttttgag gttgtatgaa    540

ataaagattt ttgactttca aaaagatatt gaagaatatc aacaaagaag aagagttccg    600

aaagagcata ttattataga tgaaaaaata aatcttaagg acttcttcaa aatagttttg    660

tactccatga ttggttctgc agctattcac tatttttct tttatgtcat tttccttacc    720

atattattta taggcattat gccaagaata tggaaattaa ataggtattc taaagttgta    780

cttttgtttt atctaaaaaa gtacttttta atgatttttt tgatgatagt tggaaatatg    840

tattggtttt tgacatattc tatagcacta ttatttgtta atattgaacc tcaaaatgta    900

aatgtggttg atacaattca attattttca cgaaatagct ctgtacagaa tgtactctat    960

ttaatttcct attggcttcc atttttaac ttagaaatgt ttttagacaa aatgttttgg   1020

gttgcaggag gagtcttcct gcttataatt gcatacataa tattttatcg ctttgggtgg   1080

catttttatg taagactttt caccctagta acaagtctgg taatcttact ctctttagga   1140

acaaacgttg acccccttagc tgatgtatat gttaaagtag taacaaacgt tccgataata   1200

ggtcatctct ttagagaccc taacaaaatt gtaggtgtac tggcttgttt tttagctata   1260

atcctgggtt ttggagtaga cagaatagta tttttcctgt taaatgcagg ctttggcaga   1320

aaatttgcag tagcttttgt tttagtcatg ttgctttgtt tctatttcta tcaaagacca   1380

gttaaaattt tctacatcga taattattac aaaggtgtgc caattccaaa agaatataaa   1440

gaaataagta aatattatac aaaagatggt aaaatattat ggatacctac tatggacaac   1500

atggtacttt caaatggcat ttcggcttat agatggaata taaacaagga aatgcctggt   1560

tcaatgaaag cggctgggga ttttcatgta tatgcatctt ctaaaaatac aatatttcaa   1620

catgaaaata atgtagggac agtttcttat ttttattctt atttacagca tctattggat   1680

aaaggtggaa ctgacaaaat aggcaggtta ttaaaactta caggtttcaa tcaagtggct   1740

tatcacagag atgtccaaaa tcaagaagag agacaaaagt ttaatctttt tgttctggaa   1800

agacagagag gaattgttct gagaaagaaa ttaggattta ttcaccttta caatgtaaag   1860
```

-continued

```
aagaattctg ataacgctaa ggtggacatt tacaatacca agggattaaa ccattttgtt   1920
cagctctttg actttgaaaa tatactggga tcaaaaggat ataatataat ctggtctcag   1980
ggtaaaaggg agaaatttga tcttaacaag gtaaatttag tggtaggcga taataaattt   2040
gatatttttc aacaaaatat acctgagaaa tatataattt cattatttga caagataaat   2100
acaggaaatc catatgtagg ttgggctaag tcaatgtgca aggagtcaga ttggttttgg   2160
attcacaaag taaacggttt aaatcgtctg ccatgggatt atgattatgg taaaggattt   2220
atattcacat atacacccta taaagtggaa cttcctccat ataagttaaa taagaatgtt   2280
ggggaggtta tactcaactc caatgatatt gtaaaaacag atttttttga aattgacaaa   2340
gaggagacaa attttaaact tgagattaca aaagatgtaa tgggcaatag catcttagca   2400
gcggaagttg ctcccaaaaa atttattggt aatgttttct ggaaaattgc aaggtcaaaa   2460
cttatgtcga taaaacctgg gtttgtaaat cttcgggcag tggtttccgg tgtaaatgct   2520
ggaaaagttc attttaaact caaatttttc aatcaaaaca tgaatgaatt gggagtggtc   2580
tatggtaatg ttccgggaga aatagctgaa tttaataaag cacttataac tgctaatgct   2640
gttgtaccgc ctgcaacaaa gtatatgagg gtggaacttt ggctacttga agatcaaaaa   2700
acaccagtat atttttggat tcatgacatt gagttaagat atttttcaag catttcatct   2760
aatgaactga gaataaaaat tccttcatat attaaaggag aataccatct ttacgtgaga   2820
gctctggtaa atgagaatgg tgggaaggtc aagatcaatg ataccaaaat aaacttaaaa   2880
ggtgatgaga accagtttaa atggttgtat gttggtaaat ataccagaga tactattgtt   2940
ataaaaccag aagaagggtt tgtactattg aaccttcttt gtgctttacc tgattggtta   3000
tataaaaaaa ttcaaaacga agatttgaag agtaaacaag caatagtatt gtttagcaat   3060
gaatttgcat taaaagaagg ttttaaagtt aaaaaaatgg atgattacat tattcatcca   3120
aattttgttg atagtctttt agacgtaata agtgatggga tttatactaa aaaaatagat   3180
atattaagaa gtgatcacta taaactatat tttataggaa acgttacaag tcaagttaca   3240
ataacattga agaataaaga cgataaagtt gtaagatttg gatatttgaa atttttttaaa   3300
ctaattaata ataactttaa tggtataaga ttttatgttt ctaacaagcc ttactcgtac   3360
tttttaaata ttgaaaaaga agaaaatcca aaatggcgtt cacaattgta ttatattgac   3420
cttggttttt tgaaaaaagg gcaatatact ttagagataa agtttgaaaa tggaataaaa   3480
tctttaacag aacctaatga tattcatgtt ttacgtccag aagaggttaa agttaattta   3540
aaagttgacg acactcttta tatattaatt tcaaacattc ttcaagaaaa tgtgacgact   3600
caaaagaaag aaaataaaat ttttagcagc aagtattcag acatatattc tgataaagta   3660
aaatactggc tcatctatgg aacaaaacag gtagagtcaa aaaagaatga gttatattac   3720
gtgaactttg atattgaaac gaaaggtttg gcagaggtta gtgctaagat tttgtttatg   3780
gataagaata aagtacttta caatagccaa tatattgaca tatttaacaa taaaggacaa   3840
agtatatttt cagctaaaaa aaatggcttt atccaaatag tatttttttgt gcgtagtaac   3900
ggtaaagaag atggtgtgtt tgaattaaag caaattagat tactagagat aagtaaaatg   3960
aacaaatttt catcaatagt aattttacca agtagtatac aaaaaggtag taaattgaat   4020
atactgaatg aaacctataa tcctctatgg gaaaacaatg gtgaaaaagc gaaacaagta   4080
aacattttct taaatggttt ttacactcaa agcaacaaat atgtatttgg taaaaaaata   4140
```

-continued

```
tggatagctt atatatttgg tttactaata tcaattttat atattacagt aaatttgtgg   4200

ttgctattaa gaagaaaaaa ataa                                          4224


<210> SEQ ID NO 75
<211> LENGTH: 1407
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 75

Met Lys Gln Glu Glu Ile Leu Gln Arg Gln Arg Lys Ile Ala Lys Thr
1               5                   10                  15

Ile Ser Leu Ile Ala Val Val Val Ile Ile Leu Ile Asp Thr Lys Asn
            20                  25                  30

Ile Trp Leu Asn Asn Gly Phe Ile Ile Phe Ser Asp Leu Asp Phe Ser
        35                  40                  45

Gly Ile Asn Asp Lys Arg Tyr Ile Glu Arg Ile Trp Gly Val Phe Asn
    50                  55                  60

Pro His Phe Ser Ser Thr Asn Phe Phe Asn Leu Ser Arg Leu Phe Phe
65                  70                  75                  80

Ile Phe Pro Phe Tyr Ala Leu Asn Phe Leu Val Ser Ser Phe Tyr Asn
                85                  90                  95

Gln Ala Met Leu Lys Leu Ile Ile Leu Ser Cys Leu Leu Ile Ser Gly
            100                 105                 110

Ile Gly Met Phe Leu Leu Cys Glu Tyr Leu Leu Ile Lys Thr Phe Arg
        115                 120                 125

Gly Phe Pro Asp Tyr Val His Tyr Phe Gly Leu Ile Ile Pro Ala Val
    130                 135                 140

Tyr Tyr Ala Leu Asn Pro Trp Val Ile Phe Arg Ile Gln His Ile Phe
145                 150                 155                 160

Leu Leu Ser Gly Tyr Ser Val Phe Pro Leu Ile Ile Lys Glu Phe Leu
                165                 170                 175

Arg Leu Tyr Glu Ile Lys Ile Phe Asp Phe Gln Lys Asp Ile Glu Glu
            180                 185                 190

Tyr Gln Gln Arg Arg Arg Val Pro Lys Glu His Ile Ile Ile Asp Glu
        195                 200                 205

Lys Ile Asn Leu Lys Asp Phe Phe Lys Ile Val Leu Tyr Ser Met Ile
    210                 215                 220

Gly Ser Ala Ala Ile His Tyr Phe Phe Phe Tyr Val Ile Phe Leu Thr
225                 230                 235                 240

Ile Leu Phe Ile Gly Ile Met Pro Arg Ile Trp Lys Leu Asn Arg Tyr
                245                 250                 255

Ser Lys Val Val Leu Leu Phe Tyr Leu Lys Lys Tyr Phe Leu Met Ile
            260                 265                 270

Phe Leu Met Ile Val Gly Asn Met Tyr Trp Phe Leu Thr Tyr Ser Ile
        275                 280                 285

Ala Leu Leu Phe Val Asn Ile Glu Pro Gln Asn Val Asn Val Val Asp
    290                 295                 300

Thr Ile Gln Leu Phe Ser Arg Asn Ser Ser Val Gln Asn Val Leu Tyr
305                 310                 315                 320

Leu Ile Ser Tyr Trp Leu Pro Phe Phe Asn Leu Glu Met Phe Leu Asp
                325                 330                 335
```

-continued

```
Lys Met Phe Trp Val Ala Gly Gly Val Phe Leu Leu Ile Ile Ala Tyr
            340                 345                 350

Ile Ile Phe Tyr Arg Phe Gly Trp His Phe Tyr Val Arg Leu Phe Thr
        355                 360                 365

Leu Val Thr Ser Leu Val Ile Leu Leu Ser Leu Gly Thr Asn Val Asp
    370                 375                 380

Pro Leu Ala Asp Val Tyr Val Lys Val Val Thr Asn Val Pro Ile Ile
385                 390                 395                 400

Gly His Leu Phe Arg Asp Pro Asn Lys Ile Val Gly Val Leu Ala Cys
                405                 410                 415

Phe Leu Ala Ile Ile Leu Gly Phe Gly Val Asp Arg Ile Val Phe Phe
            420                 425                 430

Leu Leu Asn Ala Gly Phe Gly Arg Lys Phe Ala Val Ala Phe Val Leu
        435                 440                 445

Val Met Leu Leu Cys Phe Tyr Phe Tyr Gln Arg Pro Val Lys Ile Phe
    450                 455                 460

Tyr Ile Asp Asn Tyr Tyr Lys Gly Val Pro Ile Pro Lys Glu Tyr Lys
465                 470                 475                 480

Glu Ile Ser Lys Tyr Tyr Thr Lys Asp Gly Lys Ile Leu Trp Ile Pro
                485                 490                 495

Thr Met Asp Asn Met Val Leu Ser Asn Gly Ile Ser Ala Tyr Arg Trp
            500                 505                 510

Asn Ile Asn Lys Glu Met Pro Gly Ser Met Lys Ala Ala Gly Asp Phe
        515                 520                 525

His Val Tyr Ala Ser Ser Lys Asn Thr Ile Phe Gln His Glu Asn Asn
    530                 535                 540

Val Gly Thr Val Ser Tyr Phe Tyr Ser Tyr Leu Gln His Leu Leu Asp
545                 550                 555                 560

Lys Gly Gly Thr Asp Lys Ile Gly Arg Leu Leu Lys Leu Thr Gly Phe
                565                 570                 575

Asn Gln Val Ala Tyr His Arg Asp Val Gln Asn Gln Glu Glu Arg Gln
            580                 585                 590

Lys Phe Asn Leu Phe Val Leu Glu Arg Gln Arg Gly Ile Val Leu Arg
        595                 600                 605

Lys Lys Leu Gly Phe Ile His Leu Tyr Asn Val Lys Lys Asn Ser Asp
    610                 615                 620

Asn Ala Lys Val Asp Ile Tyr Asn Thr Lys Gly Leu Asn His Phe Val
625                 630                 635                 640

Gln Leu Phe Asp Phe Glu Asn Ile Leu Gly Ser Lys Gly Tyr Asn Ile
                645                 650                 655

Ile Trp Ser Gln Gly Lys Arg Glu Lys Phe Asp Leu Asn Lys Val Asn
            660                 665                 670

Leu Val Val Gly Asp Asn Lys Phe Asp Ile Phe Gln Gln Asn Ile Pro
        675                 680                 685

Glu Lys Tyr Ile Ile Ser Leu Phe Asp Lys Ile Asn Thr Gly Asn Pro
    690                 695                 700

Tyr Val Gly Trp Ala Lys Ser Met Cys Lys Glu Ser Asp Trp Phe Trp
705                 710                 715                 720

Ile His Lys Val Asn Gly Leu Asn Arg Leu Pro Trp Asp Tyr Asp Tyr
                725                 730                 735

Gly Lys Gly Phe Ile Phe Thr Tyr Thr Pro Tyr Lys Val Glu Leu Pro
```

-continued

```
            740                 745                 750

Pro Tyr Lys Leu Asn Lys Asn Val Gly Glu Val Ile Leu Asn Ser Asn
        755                 760                 765

Asp Ile Val Lys Thr Asp Phe Phe Glu Ile Asp Lys Glu Glu Thr Asn
    770                 775                 780

Phe Lys Leu Glu Ile Thr Lys Asp Val Met Gly Asn Ser Ile Leu Ala
785                 790                 795                 800

Ala Glu Val Ala Pro Lys Lys Phe Ile Gly Asn Val Phe Trp Lys Ile
                805                 810                 815

Ala Arg Ser Lys Leu Met Ser Ile Lys Pro Gly Phe Val Asn Leu Arg
            820                 825                 830

Ala Val Val Ser Gly Val Asn Ala Gly Lys Val His Phe Lys Leu Lys
        835                 840                 845

Phe Phe Asn Gln Asn Met Asn Glu Leu Gly Val Val Tyr Gly Asn Val
    850                 855                 860

Pro Gly Glu Ile Ala Glu Phe Asn Lys Ala Leu Ile Thr Ala Asn Ala
865                 870                 875                 880

Val Val Pro Pro Ala Thr Lys Tyr Met Arg Val Glu Leu Trp Leu Leu
                885                 890                 895

Glu Asp Gln Lys Thr Pro Val Tyr Phe Trp Ile His Asp Ile Glu Leu
            900                 905                 910

Arg Tyr Phe Ser Ser Ile Ser Ser Asn Glu Leu Arg Ile Lys Ile Pro
        915                 920                 925

Ser Tyr Ile Lys Gly Glu Tyr His Leu Tyr Val Arg Ala Leu Val Asn
    930                 935                 940

Glu Asn Gly Gly Lys Val Lys Ile Asn Asp Thr Lys Ile Asn Leu Lys
945                 950                 955                 960

Gly Asp Glu Asn Gln Phe Lys Trp Leu Tyr Val Gly Lys Tyr Thr Arg
                965                 970                 975

Asp Thr Ile Val Ile Lys Pro Glu Glu Gly Phe Val Leu Leu Asn Leu
            980                 985                 990

Leu Cys Ala Leu Pro Asp Trp Leu  Tyr Lys Lys Ile Gln  Asn Glu Asp
        995                 1000                1005

Leu Lys  Ser Lys Gln Ala Ile  Val Leu Phe Ser Asn  Glu Phe Ala
    1010                1015                1020

Leu Lys  Glu Gly Phe Lys Val  Lys Lys Met Asp Asp  Tyr Ile Ile
    1025                1030                1035

His Pro  Asn Phe Val Asp Ser  Leu Leu Asp Val Ile  Ser Asp Gly
    1040                1045                1050

Ile Tyr  Thr Lys Lys Ile Asp  Ile Leu Arg Ser Asp  His Tyr Lys
    1055                1060                1065

Leu Tyr  Phe Ile Gly Asn Val  Thr Ser Gln Val Thr  Ile Thr Leu
    1070                1075                1080

Lys Asn  Lys Asp Asp Lys Val  Val Arg Phe Gly Tyr  Leu Lys Phe
    1085                1090                1095

Phe Lys  Leu Ile Asn Asn Asn  Phe Asn Gly Ile Arg  Phe Tyr Val
    1100                1105                1110

Ser Asn  Lys Pro Tyr Ser Tyr  Phe Leu Asn Ile Glu  Lys Glu Glu
    1115                1120                1125

Asn Pro  Lys Trp Arg Ser Gln  Leu Tyr Tyr Ile Asp  Leu Gly Phe
    1130                1135                1140
```

-continued

```
Leu Lys  Lys Gly Gln Tyr Thr  Leu Glu Ile Lys Phe  Glu Asn Gly
    1145                 1150                 1155

Ile Lys  Ser Leu Thr Glu Pro  Asn Asp Ile His Val  Leu Arg Pro
    1160                 1165                 1170

Glu Glu  Val Lys Val Asn Leu  Lys Val Asp Asp Thr  Leu Tyr Ile
    1175                 1180                 1185

Leu Ile  Ser Asn Ile Leu Gln  Glu Asn Val Thr Thr  Gln Lys Lys
    1190                 1195                 1200

Glu Asn  Lys Ile Phe Ser Ser  Lys Tyr Ser Asp Ile  Tyr Ser Asp
    1205                 1210                 1215

Lys Val  Lys Tyr Trp Leu Ile  Tyr Gly Thr Lys Gln  Val Glu Ser
    1220                 1225                 1230

Lys Lys  Asn Glu Leu Tyr Tyr  Val Asn Phe Asp Ile  Glu Thr Lys
    1235                 1240                 1245

Gly Leu  Ala Glu Val Ser Ala  Lys Ile Leu Phe Met  Asp Lys Asn
    1250                 1255                 1260

Lys Val  Leu Tyr Asn Ser Gln  Tyr Ile Asp Ile Phe  Asn Asn Lys
    1265                 1270                 1275

Gly Gln  Ser Ile Phe Ser Ala  Lys Lys Asn Gly Phe  Ile Gln Ile
    1280                 1285                 1290

Val Phe  Phe Val Arg Ser Asn  Gly Lys Glu Asp Gly  Val Phe Glu
    1295                 1300                 1305

Leu Lys  Gln Ile Arg Leu Leu  Glu Ile Ser Lys Met  Asn Lys Phe
    1310                 1315                 1320

Ser Ser  Ile Val Ile Leu Pro  Ser Ser Ile Gln Lys  Gly Ser Lys
    1325                 1330                 1335

Leu Asn  Ile Leu Asn Glu Thr  Tyr Asn Pro Leu Trp  Glu Asn Asn
    1340                 1345                 1350

Gly Glu  Lys Ala Lys Gln Val  Asn Ile Phe Leu Asn  Gly Phe Tyr
    1355                 1360                 1365

Thr Gln  Ser Asn Lys Tyr Val  Phe Gly Lys Lys Ile  Trp Ile Ala
    1370                 1375                 1380

Tyr Ile  Phe Gly Leu Leu Ile  Ser Ile Leu Tyr Ile  Thr Val Asn
    1385                 1390                 1395

Leu Trp  Leu Leu Leu Arg Arg  Lys Lys
    1400                 1405


<210> SEQ ID NO 76
<211> LENGTH: 405
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 76

atgccaggta aaaaaagatc tgttgtgggt gtgttgatat tctcaataat cactttgggt      60

atctattact tatactggat ttatgtaacc tcaaaagaaa cccaacggta cttagaaaag     120

aatactacaa gcccaggtct tgaactttta ctatgtataa ttacatgtgg tctttattgg     180

ttttattgga tttttaaata tagcaaaatt gctgttgaat gccaacaaaa agctggtttg     240

cctacagaag ataatgctgt tataaacctt attcttagca tcattggact tggaattata     300

agttctatga ttctccaatc aagtttaaac aaagtttggg aatttgaaag ttctaaaaat     360

agtccagtta tttcaacaac taatacttct gaaaacaata agtaa                     405
```

-continued

<210> SEQ ID NO 77
<211> LENGTH: 134
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 77

Met Pro Gly Lys Lys Arg Ser Val Val Gly Val Leu Ile Phe Ser Ile
1 5 10 15

Ile Thr Leu Gly Ile Tyr Tyr Leu Tyr Trp Ile Tyr Val Thr Ser Lys
20 25 30

Glu Thr Gln Arg Tyr Leu Glu Lys Asn Thr Thr Ser Pro Gly Leu Glu
35 40 45

Leu Leu Leu Cys Ile Ile Thr Cys Gly Leu Tyr Trp Phe Tyr Trp Ile
50 55 60

Phe Lys Tyr Ser Lys Ile Ala Val Glu Cys Gln Gln Lys Ala Gly Leu
65 70 75 80

Pro Thr Glu Asp Asn Ala Val Ile Asn Leu Ile Leu Ser Ile Ile Gly
85 90 95

Leu Gly Ile Ile Ser Ser Met Ile Leu Gln Ser Ser Leu Asn Lys Val
100 105 110

Trp Glu Phe Glu Ser Ser Lys Asn Ser Pro Val Ile Ser Thr Thr Asn
115 120 125

Thr Ser Glu Asn Asn Lys
130

<210> SEQ ID NO 78
<211> LENGTH: 762
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 78

atgatagtta aaaatgtgtt gattgcagat gaaaatgaat acataagaaa agcaataatt 60

gaaaagtttg aaaattcttt tgattcagta catcactttg tatttttga ggctgctgat 120

ggtgaagaag ctttgaacaa aattggtgag aacagtattc atgttgcaat aattgataca 180

aacttgccta aaattagtgg atttgaagtt ttgagaacta tcaaaaaaag ttctgttaaa 240

gcttacattc cagtaatttt gctcagtagc aatattcatc gtgcaacaag agcaaaggca 300

tacgagcttg gagctattgg agttattcca aagccttttt caactttaga agtttacaat 360

atggtaaggt cacttttgta cactcaagat gaatatttac atattaatga agttgtacat 420

cttatttctt ttttgaatga aatggcaaat gaacatatca ttgtagttcc caaaattata 480

gatgaatttt tatcagaata ttttgtttca tatcacttgt tggctttaaa agatactaca 540

aaagaggtga tttataacag aggttttgat gaatatgaaa ttaaaaaggt tattggctat 600

ttaaatgatg agaacagaat aaattcaagc tgttattcga ttataaaact taacaataag 660

gaagaaactt attactttat attcaaaatt ctgaatgata ataaaatatt aattcttcta 720

aagaaagttt tagaactttg gagtgaatta aatgatagat aa 762

<210> SEQ ID NO 79
<211> LENGTH: 253
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

-continued

<400> SEQUENCE: 79

Met Ile Val Lys Asn Val Leu Ile Ala Asp Glu Asn Glu Tyr Ile Arg
1 5 10 15

Lys Ala Ile Ile Glu Lys Phe Glu Asn Ser Phe Asp Ser Val His His
20 25 30

Phe Val Phe Phe Glu Ala Ala Asp Gly Glu Glu Ala Leu Asn Lys Ile
35 40 45

Gly Glu Asn Ser Ile His Val Ala Ile Ile Asp Thr Asn Leu Pro Lys
50 55 60

Ile Ser Gly Phe Glu Val Leu Arg Thr Ile Lys Lys Ser Ser Val Lys
65 70 75 80

Ala Tyr Ile Pro Val Ile Leu Leu Ser Ser Asn Ile His Arg Ala Thr
85 90 95

Arg Ala Lys Ala Tyr Glu Leu Gly Ala Ile Gly Val Ile Pro Lys Pro
100 105 110

Phe Ser Thr Leu Glu Val Tyr Asn Met Val Arg Ser Leu Leu Tyr Thr
115 120 125

Gln Asp Glu Tyr Leu His Ile Asn Glu Val Val His Leu Ile Ser Phe
130 135 140

Leu Asn Glu Met Ala Asn Glu His Ile Ile Val Val Pro Lys Ile Ile
145 150 155 160

Asp Glu Phe Leu Ser Glu Tyr Phe Val Ser Tyr His Leu Leu Ala Leu
165 170 175

Lys Asp Thr Thr Lys Glu Val Ile Tyr Asn Arg Gly Phe Asp Glu Tyr
180 185 190

Glu Ile Lys Lys Val Ile Gly Tyr Leu Asn Asp Glu Asn Arg Ile Asn
195 200 205

Ser Ser Cys Tyr Ser Ile Ile Lys Leu Asn Asn Lys Glu Glu Thr Tyr
210 215 220

Tyr Phe Ile Phe Lys Ile Leu Asn Asp Asn Lys Ile Leu Ile Leu Leu
225 230 235 240

Lys Lys Val Leu Glu Leu Trp Ser Glu Leu Asn Asp Arg
245 250

<210> SEQ ID NO 80
<211> LENGTH: 1233
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 80

atgatagata atataaattt tgttatcttg tgtattatta ttgttattct tttagttata 60

acttcgggag ttgtttttt atacaaagtt ataaatagta tccgtcaaga aaaaataaag 120

cagattttta aaaattacag taaaaatgtc tatgatgtga taacaggaaa gaaaaatgtt 180

attgattctt caaacattta cattctgagt gatgttataa atatatatta ttcatgggta 240

catggagaag aaaaaaaacg gctatatata gctttgaaga aattgaattt ttttgacatt 300

gctatcaaga tggttcaaaa agggaacaaa gttcgaaaac ttaggtttgc aaaagttatt 360

agtatagtgg gcgaagaaga tgagctaaaa acgcttttaa aaataagcat taaagagccg 420

tatctgatag atacaactgc tgaagccatt ttcaaaaaca tagatgagat acagaattta 480

agtgttttta aaccttattt aaagactatt tttctaaaca ttgataaata tccagattcg 540

-continued

```
gttagaaaaa gaatagaatt ttttacagtt tatggtggcg aaaaaataaa agatataatt    600

ttatacgtga taaaagaaaa cccgtcggat aaagttctga tatcatgttt gaatatattt    660

tcagagattg cggcattaga agatttagaa cacattgatt ttcttataaa tcatccatca    720

ccagaagtca gatcagcttt ctgcagagtg attgaaaaaa taggatgtag aaactgcaaa    780

gaaaaacttg aaactctgat aaaaaacgaa aatataaatt ttgtgaagtt acgagcctta    840

agggcgttga gcaatgtttc atctaagggt tctttaaaat atttacttgc ctcccttgaa    900

gatgactggt tctacatgag agactttgca agaaagatgc tatctgagtt tggaccagtt    960

attttaaatg atcttctaaa attttactat acaacaaatg ataaatttgc aagggacaaa   1020

ataagagagg ttttttatag tccagtgaat tttgattata ttatcaagag tgctttgaat   1080

tatagaactg aacaggagaa agatttagct ttagaaataa ttaaaatact taaatcgtct   1140

aacccccact gtttttatca aagaatgaaa gatgaaggat ttgaatatct tataaccgag   1200

aaagaggtag tgaataacga atgcgacata taa                                1233
```

```
<210> SEQ ID NO 81
<211> LENGTH: 410
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 81

Met Ile Asp Asn Ile Asn Phe Val Ile Leu Cys Ile Ile Ile Val Ile
1               5                   10                  15

Leu Leu Val Ile Thr Ser Gly Val Val Phe Leu Tyr Lys Val Ile Asn
            20                  25                  30

Ser Ile Arg Gln Glu Lys Ile Lys Gln Ile Phe Lys Asn Tyr Ser Lys
        35                  40                  45

Asn Val Tyr Asp Val Ile Thr Gly Lys Lys Asn Val Ile Asp Ser Ser
    50                  55                  60

Asn Ile Tyr Ile Leu Ser Asp Val Ile Asn Ile Tyr Tyr Ser Trp Val
65                  70                  75                  80

His Gly Glu Glu Lys Lys Arg Leu Tyr Ile Ala Leu Lys Lys Leu Asn
                85                  90                  95

Phe Phe Asp Ile Ala Ile Lys Met Val Gln Lys Gly Asn Lys Val Arg
            100                 105                 110

Lys Leu Arg Phe Ala Lys Val Ile Ser Ile Val Gly Glu Glu Asp Glu
        115                 120                 125

Leu Lys Thr Leu Leu Lys Ile Ser Ile Lys Glu Pro Tyr Leu Ile Asp
    130                 135                 140

Thr Thr Ala Glu Ala Ile Phe Lys Asn Ile Asp Glu Ile Gln Asn Leu
145                 150                 155                 160

Ser Val Phe Lys Pro Tyr Leu Lys Thr Ile Phe Leu Asn Ile Asp Lys
                165                 170                 175

Tyr Pro Asp Ser Val Arg Lys Arg Ile Glu Phe Phe Thr Val Tyr Gly
            180                 185                 190

Gly Glu Lys Ile Lys Asp Ile Ile Leu Tyr Val Ile Lys Glu Asn Pro
        195                 200                 205

Ser Asp Lys Val Leu Ile Ser Cys Leu Asn Ile Phe Ser Glu Ile Ala
    210                 215                 220

Ala Leu Glu Asp Leu Glu His Ile Asp Phe Leu Ile Asn His Pro Ser
225                 230                 235                 240
```

-continued

```
Pro Glu Val Arg Ser Ala Phe Cys Arg Val Ile Glu Lys Ile Gly Cys
                245                 250                 255

Arg Asn Cys Lys Glu Lys Leu Glu Thr Leu Ile Lys Asn Glu Asn Ile
            260                 265                 270

Asn Phe Val Lys Leu Arg Ala Leu Arg Ala Leu Ser Asn Val Ser Ser
        275                 280                 285

Lys Gly Ser Leu Lys Tyr Leu Leu Ala Ser Leu Glu Asp Asp Trp Phe
    290                 295                 300

Tyr Met Arg Asp Phe Ala Arg Lys Met Leu Ser Glu Phe Gly Pro Val
305                 310                 315                 320

Ile Leu Asn Asp Leu Leu Lys Phe Tyr Tyr Thr Thr Asn Asp Lys Phe
                325                 330                 335

Ala Arg Asp Lys Ile Arg Glu Val Phe Tyr Ser Pro Val Asn Phe Asp
            340                 345                 350

Tyr Ile Ile Lys Ser Ala Leu Asn Tyr Arg Thr Glu Gln Glu Lys Asp
        355                 360                 365

Leu Ala Leu Glu Ile Ile Lys Ile Leu Lys Ser Ser Asn Pro His Cys
    370                 375                 380

Phe Tyr Gln Arg Met Lys Asp Glu Gly Phe Glu Tyr Leu Ile Thr Glu
385                 390                 395                 400

Lys Glu Val Val Asn Asn Glu Cys Asp Ile
                405                 410


<210> SEQ ID NO 82
<211> LENGTH: 1395
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 82

atgcgacata taataaactg gttttcatgg tttgtatctt attatgtttt agttttgaat     60

actgtttatg ccattttaat tttaatatct ctttttggta ttgttagtta ctggagaaat    120

aagataaaag gaaggattgt ggaggttgtc tcatcagact ttgcactccc tgtatcactg    180

ctggtgccgg catacaatga agaagaaaca atagcaaaat ctgtaaaatc ttttttgcag    240

attgaatatc cagaatatga agtagtagta attaatgatg ggtcaaaaga tgggacattg    300

gatgtattaa aaaacgaatt tgacctttac attgtagata ggaaatttag aaagattttg    360

tcaacaaaag agataaaggt catatactat tccaaaaagt attcaaattt aatcgtagtg    420

gataaagaaa atgggggaaa agctgatgct ctcaatgcag ggataaacgt atgtacatac    480

ccatatgttt gttcacttga tgccgattca attttagaaa gagattcaat agccaaggtt    540

atgcaaccat tttttgataa cccttatgaa gttgtagcca caacgggtat tgttaggatt    600

gtaaatggaa ccgaattgga ttcctttggc aatataaaaa aattaaagct accaagttca    660

agtcttgcaa gatttcagat aatagaatac ttgagagctt ttttgggtgc gagaaaaggt    720

ctttctatga taggaagtct tgttattgca tctggagcat ttgcagcatt taataaaaat    780

gctattataa aggttggagg gttttctgat agaactgttg gcgaggatat ggagattgtt    840

gtaaagctca gaaaaaactc ttataaagaa ggcgcactgg gaagagttga atttgtacca    900

gacccgattg tatggactca atgtccagag actttaaaag acctttcaaa acaaagaaga    960

agatggcaaa gaggtctttg ccaagttatt ttcatgcaca aagatattct atttaatcct   1020
```

-continued

```
aagtatggca tattaggtct ttttgctatg ccatatcagc ttatgtttga actattgggg   1080

ccgtttgtgg agatgttggg ttatattttt atacctatat cgtattttgc tcacataatc   1140

aatttagaag tggccttatt tttctttgca gttgagataa tgtacggaat agttatttcg   1200

attttagcag ttcttcttgg agaattttct gatagaaaat atgaagggtg gagagagttc   1260

ggtatacttg tattgtttgc gatattagaa aattttggct acagacagat gacaatgtta   1320

ttcagaattg ttggtacatt tgaagctata cttagaaaga aaggttgggc aaagcctgag   1380

agaaagaagt tataa                                                    1395
```

```
<210> SEQ ID NO 83
<211> LENGTH: 464
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 83

Met Arg His Ile Ile Asn Trp Phe Ser Trp Phe Val Ser Tyr Tyr Val
1               5                   10                  15

Leu Val Leu Asn Thr Val Tyr Ala Ile Leu Ile Leu Ile Ser Leu Phe
            20                  25                  30

Gly Ile Val Ser Tyr Trp Arg Asn Lys Ile Lys Gly Arg Ile Val Glu
        35                  40                  45

Val Val Ser Ser Asp Phe Ala Leu Pro Val Ser Leu Leu Val Pro Ala
    50                  55                  60

Tyr Asn Glu Glu Glu Thr Ile Ala Lys Ser Val Lys Ser Phe Leu Gln
65                  70                  75                  80

Ile Glu Tyr Pro Glu Tyr Glu Val Val Val Ile Asn Asp Gly Ser Lys
                85                  90                  95

Asp Gly Thr Leu Asp Val Leu Lys Asn Glu Phe Asp Leu Tyr Ile Val
            100                 105                 110

Asp Arg Lys Phe Arg Lys Ile Leu Ser Thr Lys Glu Ile Lys Val Ile
        115                 120                 125

Tyr Tyr Ser Lys Lys Tyr Ser Asn Leu Ile Val Val Asp Lys Glu Asn
    130                 135                 140

Gly Gly Lys Ala Asp Ala Leu Asn Ala Gly Ile Asn Val Cys Thr Tyr
145                 150                 155                 160

Pro Tyr Val Cys Ser Leu Asp Ala Asp Ser Ile Leu Glu Arg Asp Ser
                165                 170                 175

Ile Ala Lys Val Met Gln Pro Phe Phe Asp Asn Pro Tyr Glu Val Val
            180                 185                 190

Ala Thr Thr Gly Ile Val Arg Ile Val Asn Gly Thr Glu Leu Asp Ser
        195                 200                 205

Phe Gly Asn Ile Lys Lys Leu Lys Leu Pro Ser Ser Ser Leu Ala Arg
    210                 215                 220

Phe Gln Ile Ile Glu Tyr Leu Arg Ala Phe Leu Gly Ala Arg Lys Gly
225                 230                 235                 240

Leu Ser Met Ile Gly Ser Leu Val Ile Ala Ser Gly Ala Phe Ala Ala
                245                 250                 255

Phe Asn Lys Asn Ala Ile Ile Lys Val Gly Gly Phe Ser Asp Arg Thr
            260                 265                 270

Val Gly Glu Asp Met Glu Ile Val Val Lys Leu Arg Lys Asn Ser Tyr
        275                 280                 285
```

-continued

```
Lys Glu Gly Ala Leu Gly Arg Val Glu Phe Val Pro Asp Pro Ile Val
    290                 295                 300

Trp Thr Gln Cys Pro Glu Thr Leu Lys Asp Leu Ser Lys Gln Arg Arg
305                 310                 315                 320

Arg Trp Gln Arg Gly Leu Cys Gln Val Ile Phe Met His Lys Asp Ile
                325                 330                 335

Leu Phe Asn Pro Lys Tyr Gly Ile Leu Gly Leu Phe Ala Met Pro Tyr
            340                 345                 350

Gln Leu Met Phe Glu Leu Leu Gly Pro Phe Val Glu Met Leu Gly Tyr
        355                 360                 365

Ile Phe Ile Pro Ile Ser Tyr Phe Ala His Ile Ile Asn Leu Glu Val
    370                 375                 380

Ala Leu Phe Phe Phe Ala Val Glu Ile Met Tyr Gly Ile Val Ile Ser
385                 390                 395                 400

Ile Leu Ala Val Leu Leu Gly Glu Phe Ser Asp Arg Lys Tyr Glu Gly
                405                 410                 415

Trp Arg Glu Phe Gly Ile Leu Val Leu Phe Ala Ile Leu Glu Asn Phe
            420                 425                 430

Gly Tyr Arg Gln Met Thr Met Leu Phe Arg Ile Val Gly Thr Phe Glu
        435                 440                 445

Ala Ile Leu Arg Lys Lys Gly Trp Ala Lys Pro Glu Arg Lys Lys Leu
    450                 455                 460


<210> SEQ ID NO 84
<211> LENGTH: 1005
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 84

atgttctatc ggcaggaccc agatatcgaa aagaatatag tgagtgttca aagattagat      60

attacaacag aggtagggca aaagccacag cttccgccaa ctgtcatggt ggtttatgat     120

gatgatacat cctctgaggc aaatgttaca tgggaagcta ttgatgagag taaatactca     180

gcgattggtg aatttactgt gactggtact gtagaaggca caactctaaa agcatatgcc     240

aaagtcaagg ttatcaacag caaaaacctt gtaaaaaatt acagttttga agaaggacta     300

aagtattgga taagcgaagg aaatacaaat gcagtcaaaa ctgaaagtgg tgggcattct     360

gggacagcgc agcttactca ctggagtgac aaaccttaca aggtggtcac atatcaaaca     420

attgagaata taccaaatgg tatatacatt ttcagagcct ttgcaaatgg aggaggcggt     480

cagaatgcta actacctgtt tgtaaaagac tatggtggag aagaattgag aataaatatg     540

ccaactacat gggtcgctga gtggcacaga atggtgattg caaacattaa ggttacgacc     600

ggacgtgtta caatagggct ttattcagat tcagagaatg caggaggtac atggtgcaac     660

cttgacgatg ttgaattttt caaggttgct gacagtgaat ttgaaatttt aaatgcaaat     720

ctgcaaaagg ataaaggaat tgttgcgagt gttacagtga aacaatcaga aggtatccag     780

catgagggga aagaggccat tgtatttgag ctacttaaag ggacaactcc tgtttcaatt     840

gttgctattg aaaaagatat aatagatgct gaagatttca aggcatactt taatgtaaat     900

gattatctta gccctgatta cagggtaaaa gtatttgtat ttgacaaatt tgacacaagc     960

ctttctgttc caaatagtct tgctgagcct gttgagcttc gatga                    1005
```

-continued

<210> SEQ ID NO 85
<211> LENGTH: 334
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 85

Met Phe Tyr Arg Gln Asp Pro Asp Ile Glu Lys Asn Ile Val Ser Val
1 5 10 15

Gln Arg Leu Asp Ile Thr Thr Glu Val Gly Gln Lys Pro Gln Leu Pro
20 25 30

Pro Thr Val Met Val Val Tyr Asp Asp Asp Thr Ser Ser Glu Ala Asn
35 40 45

Val Thr Trp Glu Ala Ile Asp Glu Ser Lys Tyr Ser Ala Ile Gly Glu
50 55 60

Phe Thr Val Thr Gly Thr Val Glu Gly Thr Thr Leu Lys Ala Tyr Ala
65 70 75 80

Lys Val Lys Val Ile Asn Ser Lys Asn Leu Val Lys Asn Tyr Ser Phe
85 90 95

Glu Glu Gly Leu Lys Tyr Trp Ile Ser Glu Gly Asn Thr Asn Ala Val
100 105 110

Lys Thr Glu Ser Gly Gly His Ser Gly Thr Ala Gln Leu Thr His Trp
115 120 125

Ser Asp Lys Pro Tyr Lys Val Val Thr Tyr Gln Thr Ile Glu Asn Ile
130 135 140

Pro Asn Gly Ile Tyr Ile Phe Arg Ala Phe Ala Asn Gly Gly Gly Gly
145 150 155 160

Gln Asn Ala Asn Tyr Leu Phe Val Lys Asp Tyr Gly Gly Glu Glu Leu
165 170 175

Arg Ile Asn Met Pro Thr Thr Trp Val Ala Glu Trp His Arg Met Val
180 185 190

Ile Ala Asn Ile Lys Val Thr Thr Gly Arg Val Thr Ile Gly Leu Tyr
195 200 205

Ser Asp Ser Glu Asn Ala Gly Gly Thr Trp Cys Asn Leu Asp Asp Val
210 215 220

Glu Phe Phe Lys Val Ala Asp Ser Glu Phe Glu Ile Leu Asn Ala Asn
225 230 235 240

Leu Gln Lys Asp Lys Gly Ile Val Ala Ser Val Thr Val Lys Gln Ser
245 250 255

Glu Gly Ile Gln His Glu Gly Lys Glu Ala Ile Val Phe Glu Leu Leu
260 265 270

Lys Gly Thr Thr Pro Val Ser Ile Val Ala Ile Glu Lys Asp Ile Ile
275 280 285

Asp Ala Glu Asp Phe Lys Ala Tyr Phe Asn Val Asn Asp Tyr Leu Ser
290 295 300

Pro Asp Tyr Arg Val Lys Val Phe Val Phe Asp Lys Phe Asp Thr Ser
305 310 315 320

Leu Ser Val Pro Asn Ser Leu Ala Glu Pro Val Glu Leu Arg
325 330

<210> SEQ ID NO 86
<211> LENGTH: 1644
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

-continued

<400> SEQUENCE: 86

atgaaaaaaa gaacaaggat agcttcaagt ctattgcttg tttttgtgtt tttgttctct 60

ttattgtctg tgggatattc tgcccaggaa gttattttaa attccatacc tgataaaaat 120

ccaggtgaag acgttgtaat ttctgggcaa acgatgtttg atgagattgt tatcaaagtg 180

ttgagaccaa attctaccat actttatata aatacagtga aggggaaaaa cttcagtgac 240

aaatttactt taccagcaga cttgcctgag ggtacttaca cagttgttac tggaaaaggt 300

tctattgtcg atataaaaac tttcaatgta gtaaggaaac aaccatcttt acctaattcg 360

tttcctgaga ttttaatccc taatttcaaa gaagctgaaa aacaggcagg tgtacagcca 420

tctcaaaagg aagatattaa acaaactgat aaagctgctt tgcctgaaat tacaattgat 480

aatagtggga taattaggcc caaaatatat cagctgtcag agggaagatt agacgttaga 540

ttagatgagt ctttagtaca aaaagcactg cagatgcagg tgtcaaaaaa caaaatttta 600

acgttggatt taaaaaacag caagccttta gtacagtaca atattgtgct tccatcgaaa 660

gtatttacag cttcttttga tgtgcaggaa gttttagttg atacagatga attagatttg 720

aaacttccag ctgacctatt aaaggaaaag aacatagaca agaacaagca aatagagata 780

ttgataaaaa aagtagaaaa ttctaatata ccttcagaga ttcgtgcggc agttgggaaa 840

agaccaatat atgaaattgg gttttatcaa gatttgaaaa aggtgaatgt tgaaagacct 900

acaaattcta taaaaatatc tattgcttat actcctggaa tagaagagct gagcggtatt 960

gaaaacttgg ttttgtttca tataaaagag gacgggaaaa cagagatttt atcaaatagt 1020

aaatatgtaa ggacatcgaa atgtgtgatt gcaaatgtga atggttttag caagtttgca 1080

atagggtatg ttaccaaaaa atttgacgac ttaaaaaatt attcgtgggc tgaaaaagct 1140

gtgtcttccc tggccgcaag aaatataata agtggtgttg acaaaagtag ttttagacct 1200

caagagtata taaagcgtgg agagtttgtc aagtggctgg taaatgctct tggccttgat 1260

gctcagtatt cttcgaactt tgaggatgtc aaaaaagata gcagctactg gcgtgaagtt 1320

gcaattgcaa aagcacttgg cattgcaaaa ggttatgcga ataagtttaa accagaagat 1380

tatataacaa ggcaagatat gatggttctt gttaaaagag cattggaggt tgcaaacaaa 1440

ccaattgtaa aaacaaaaag tagccttgca actaaatttt ctgattcatc tgatatatct 1500

ttgtatgctc aagatagtat ttctattttg gttgcaaatg atttaataaa aggaaatagc 1560

aaaaatcaaa tactgcctcg aaagtttgca actcgtgcag aagctgctca gcttttgttc 1620

aggatatttt tcaaatggga atag 1644

<210> SEQ ID NO 87
<211> LENGTH: 547
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 87

Met Lys Lys Arg Thr Arg Ile Ala Ser Ser Leu Leu Leu Val Phe Val
1 5 10 15

Phe Leu Phe Ser Leu Leu Ser Val Gly Tyr Ser Ala Gln Glu Val Ile
20 25 30

Leu Asn Ser Ile Pro Asp Lys Asn Pro Gly Glu Asp Val Val Ile Ser
35 40 45

Gly Gln Thr Met Phe Asp Glu Ile Val Ile Lys Val Leu Arg Pro Asn

-continued

```
    50                  55                  60

Ser Thr Ile Leu Tyr Ile Asn Thr Val Lys Gly Lys Asn Phe Ser Asp
65                  70                  75                  80

Lys Phe Thr Leu Pro Ala Asp Leu Pro Glu Gly Thr Tyr Thr Val Val
                85                  90                  95

Thr Gly Lys Gly Ser Ile Val Asp Ile Lys Thr Phe Asn Val Val Arg
            100                 105                 110

Lys Gln Pro Ser Leu Pro Asn Ser Phe Pro Glu Ile Leu Ile Pro Asn
        115                 120                 125

Phe Lys Glu Ala Glu Lys Gln Ala Gly Val Gln Pro Ser Gln Lys Glu
    130                 135                 140

Asp Ile Lys Gln Thr Asp Lys Ala Ala Leu Pro Glu Ile Thr Ile Asp
145                 150                 155                 160

Asn Ser Gly Ile Ile Arg Pro Lys Ile Tyr Gln Leu Ser Glu Gly Arg
                165                 170                 175

Leu Asp Val Arg Leu Asp Glu Ser Leu Val Gln Lys Ala Leu Gln Met
            180                 185                 190

Gln Val Ser Lys Asn Lys Ile Leu Thr Leu Asp Leu Lys Asn Ser Lys
        195                 200                 205

Pro Leu Val Gln Tyr Asn Ile Val Leu Pro Ser Lys Val Phe Thr Ala
    210                 215                 220

Ser Phe Asp Val Gln Glu Val Leu Val Asp Thr Asp Glu Leu Asp Leu
225                 230                 235                 240

Lys Leu Pro Ala Asp Leu Leu Lys Glu Lys Asn Ile Asp Lys Asn Lys
                245                 250                 255

Gln Ile Glu Ile Leu Ile Lys Lys Val Glu Asn Ser Asn Ile Pro Ser
            260                 265                 270

Glu Ile Arg Ala Ala Val Gly Lys Arg Pro Ile Tyr Glu Ile Gly Phe
        275                 280                 285

Tyr Gln Asp Leu Lys Lys Val Asn Val Glu Arg Pro Thr Asn Ser Ile
    290                 295                 300

Lys Ile Ser Ile Ala Tyr Thr Pro Gly Ile Glu Glu Leu Ser Gly Ile
305                 310                 315                 320

Glu Asn Leu Val Leu Phe His Ile Lys Glu Asp Gly Lys Thr Glu Ile
                325                 330                 335

Leu Ser Asn Ser Lys Tyr Val Arg Thr Ser Lys Cys Val Ile Ala Asn
            340                 345                 350

Val Asn Gly Phe Ser Lys Phe Ala Ile Gly Tyr Val Thr Lys Lys Phe
        355                 360                 365

Asp Asp Leu Lys Asn Tyr Ser Trp Ala Glu Lys Ala Val Ser Ser Leu
    370                 375                 380

Ala Ala Arg Asn Ile Ile Ser Gly Val Asp Lys Ser Ser Phe Arg Pro
385                 390                 395                 400

Gln Glu Tyr Ile Lys Arg Gly Glu Phe Val Lys Trp Leu Val Asn Ala
                405                 410                 415

Leu Gly Leu Asp Ala Gln Tyr Ser Ser Asn Phe Glu Asp Val Lys Lys
            420                 425                 430

Asp Ser Ser Tyr Trp Arg Glu Val Ala Ile Ala Lys Ala Leu Gly Ile
        435                 440                 445

Ala Lys Gly Tyr Ala Asn Lys Phe Lys Pro Glu Asp Tyr Ile Thr Arg
    450                 455                 460
```

-continued

```
Gln Asp Met Met Val Leu Val Lys Arg Ala Leu Glu Val Ala Asn Lys
465                 470                 475                 480

Pro Ile Val Lys Thr Lys Ser Ser Leu Ala Thr Lys Phe Ser Asp Ser
                485                 490                 495

Ser Asp Ile Ser Leu Tyr Ala Gln Asp Ser Ile Ser Ile Leu Val Ala
            500                 505                 510

Asn Asp Leu Ile Lys Gly Asn Ser Lys Asn Gln Ile Leu Pro Arg Lys
        515                 520                 525

Phe Ala Thr Arg Ala Glu Ala Ala Gln Leu Leu Phe Arg Ile Phe Phe
    530                 535                 540

Lys Trp Glu
545


<210> SEQ ID NO 88
<211> LENGTH: 2502
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 88

atgttaacta ataaaaagct tagggagtat gtatattttg tatcgattat ttttttcatg     60

actattatat tgttaaatgg acagataaat gtgagcaaaa gtaatttagc aatggcggca    120

acaggaagtc aaattgagaa gttaaataga ggattaattg caataaaagt aaccaacgga    180

gtatttttga gttggagaat gtttggttca gatttagcga atatagggtt taatatttac    240

agaaatggag tgaagataac taatacgccg attcagaaca gcacaaacta tgttgatact    300

gctgggacag ctgcttcgaa gtattatgta aaagcagtga taaatggtgt agaggtagag    360

caatctgaag aagtaagtgt gcttagtagc aattatattg aaataagatt aaataaacca    420

gcgaattctc cactcggtgc ttcatattcg ccaaatgatg caagtgttgg cgatttagat    480

ggtgatgggg aatatgagat tgttctgaaa tgggatccaa gtgattcaaa ggataactca    540

caatctggat acacaagcaa tgtgtattta gacgcttaca aattaaatgg caagttttta    600

tggagaattg atttgggtag gaatataaga gcaggagcac attatacaca atttatagta    660

tatgatttag atggtgacgg aaaagctgaa gttgcatgta aaacagctga tggaacaata    720

gatggacaag ggaatgtgat aggtgatcca aatgctgatt ggaggaattc ctctggttat    780

attttatcag ggcctgaata tttgactata tttgaaggtg cgacaggacg agcaataaag    840

acggttaatt atattccgcc acgagggaat gtttcatcat ggggtgattc ttatggaaac    900

agggttgaca ggtttttagc agcagtagct tatttggatg ggaatagacc tagcttaatt    960

atgtgccgag gatattacac aaaaacatat atagttgctt ggaattggcg aaatggtgag   1020

ttaacaaagt tatggcaatt tgacacagga gagattagag atggatatag agatgattac   1080

gaaggacaag gaaatcacaa tttgagtgtg gctgatgttg acaatgatgg taaagatgag   1140

ataatatatg gtgctatggt agtagatgat aatggagcac cattatattc aactaaatta   1200

ggtcatggtg atgcaatgca tgtgacagat attgatccag acaggccagg attagaagtt   1260

tggcagtgtc atgaaggaag tacaggagcg agtttaaggg atgcacgaac tggacagata   1320

ttagtgagag ttttaacatc tggagatagt ggacgtgctt tgacggcaga tattaacccg   1380

cgatatagag gattagaaat gtgggcggca ggtggaataa gtgttagaga ttgtagaggt   1440

aatgtaatca gtaatgcgac accaccaatt aattttgcaa tatggtggga tggagattta   1500
```

-continued

```
ggtagagaat tgttggataa tgtatatatt tataaatggg attataacaa taataggagt   1560

aatactatat tcacagcaag tggatgttca tcaaacaatg gcacaaaagc aacaccttgc   1620

ttgagtgcag atatacttgg ggattggcga gaagaggtca tatttaggac tagtgataac   1680

aatgcaatca ggatatatac aaccaccaca ttgacagatt ataagatacc tacgcttatg   1740

cataacaggc aatatagggt gtctatagca tggcagaacg ttgcatataa tcaaccacct   1800

cacgtaagtt tttatttagg gtatgagact aatgtgaata acatatatca atattttgaa   1860

ggttatgggc aacaaccaat tgttacaccg tcgctaaccc cgacaaaaac accaacgcct   1920

acatcaactc cattgccaac atcaactgca acatctacgc caactccaac agcaacagca   1980

acaccaacac cgacaccaac agcaacacca acaccgacgc cgagcagcac acctgtagca   2040

ggtggacaga taaaggtatt gtatgctaac aaggagacaa atagcacaac aaacacgata   2100

aggccatggt tgaaggtagt gaacactgga agtagcagca tagatttgag cagggtaacg   2160

ataaggtact ggtacacggt agatggggac aaggcacaga gtgcgatatc agactgggca   2220

cagataggag caagcaatgt gacattcaag tttgtgaagc tgagcagtag cgtaagtgga   2280

gcggactatt atttagagat aggatttaag agtggagctg ggcagttgca ggctggtaaa   2340

gacacagggg agatacagat aaggttcaac aagagtgatt ggagcaatta caatcaaggg   2400

aatgactggt catggatgca gagcatgacg aattatggag agaatgtgaa ggtaacagcg   2460

tatgtagatg gggtgctggt atgggggcaa gaaccaaaat aa                      2502
```

<210> SEQ ID NO 89
<211> LENGTH: 833
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 89

```
Met Leu Thr Asn Lys Lys Leu Arg Glu Tyr Val Tyr Phe Val Ser Ile
1               5                   10                  15

Ile Phe Phe Met Thr Ile Ile Leu Leu Asn Gly Gln Ile Asn Val Ser
            20                  25                  30

Lys Ser Asn Leu Ala Met Ala Ala Thr Gly Ser Gln Ile Glu Lys Leu
        35                  40                  45

Asn Arg Gly Leu Ile Ala Ile Lys Val Thr Asn Gly Val Phe Leu Ser
    50                  55                  60

Trp Arg Met Phe Gly Ser Asp Leu Ala Asn Ile Gly Phe Asn Ile Tyr
65                  70                  75                  80

Arg Asn Gly Val Lys Ile Thr Asn Thr Pro Ile Gln Asn Ser Thr Asn
                85                  90                  95

Tyr Val Asp Thr Ala Gly Thr Ala Ala Ser Lys Tyr Tyr Val Lys Ala
            100                 105                 110

Val Ile Asn Gly Val Glu Val Glu Gln Ser Glu Glu Val Ser Val Leu
        115                 120                 125

Ser Ser Asn Tyr Ile Glu Ile Arg Leu Asn Lys Pro Ala Asn Ser Pro
    130                 135                 140

Leu Gly Ala Ser Tyr Ser Pro Asn Asp Ala Ser Val Gly Asp Leu Asp
145                 150                 155                 160

Gly Asp Gly Glu Tyr Glu Ile Val Leu Lys Trp Asp Pro Ser Asp Ser
                165                 170                 175
```

-continued

```
Lys Asp Asn Ser Gln Ser Gly Tyr Thr Ser Asn Val Tyr Leu Asp Ala
            180                 185                 190

Tyr Lys Leu Asn Gly Lys Phe Leu Trp Arg Ile Asp Leu Gly Arg Asn
        195                 200                 205

Ile Arg Ala Gly Ala His Tyr Thr Gln Phe Ile Val Tyr Asp Leu Asp
    210                 215                 220

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

Asp Gly Gln Gly Asn Val Ile Gly Asp Pro Asn Ala Asp Trp Arg Asn
                245                 250                 255

Ser Ser Gly Tyr Ile Leu Ser Gly Pro Glu Tyr Leu Thr Ile Phe Glu
            260                 265                 270

Gly Ala Thr Gly Arg Ala Ile Lys Thr Val Asn Tyr Ile Pro Pro Arg
        275                 280                 285

Gly Asn Val Ser Ser Trp Gly Asp Ser Tyr Gly Asn Arg Val Asp Arg
    290                 295                 300

Phe Leu Ala Ala Val Ala Tyr Leu Asp Gly Asn Arg Pro Ser Leu Ile
305                 310                 315                 320

Met Cys Arg Gly Tyr Tyr Thr Lys Thr Tyr Ile Val Ala Trp Asn Trp
                325                 330                 335

Arg Asn Gly Glu Leu Thr Lys Leu Trp Gln Phe Asp Thr Gly Glu Ile
            340                 345                 350

Arg Asp Gly Tyr Arg Asp Asp Tyr Glu Gly Gln Gly Asn His Asn Leu
        355                 360                 365

Ser Val Ala Asp Val Asp Asn Asp Gly Lys Asp Glu Ile Ile Tyr Gly
    370                 375                 380

Ala Met Val Val Asp Asp Asn Gly Ala Pro Leu Tyr Ser Thr Lys Leu
385                 390                 395                 400

Gly His Gly Asp Ala Met His Val Thr Asp Ile Asp Pro Asp Arg Pro
                405                 410                 415

Gly Leu Glu Val Trp Gln Cys His Glu Gly Ser Thr Gly Ala Ser Leu
            420                 425                 430

Arg Asp Ala Arg Thr Gly Gln Ile Leu Val Arg Val Leu Thr Ser Gly
        435                 440                 445

Asp Ser Gly Arg Ala Leu Thr Ala Asp Ile Asn Pro Arg Tyr Arg Gly
    450                 455                 460

Leu Glu Met Trp Ala Ala Gly Gly Ile Ser Val Arg Asp Cys Arg Gly
465                 470                 475                 480

Asn Val Ile Ser Asn Ala Thr Pro Pro Ile Asn Phe Ala Ile Trp Trp
                485                 490                 495

Asp Gly Asp Leu Gly Arg Glu Leu Leu Asp Asn Val Tyr Ile Tyr Lys
            500                 505                 510

Trp Asp Tyr Asn Asn Asn Arg Ser Asn Thr Ile Phe Thr Ala Ser Gly
        515                 520                 525

Cys Ser Ser Asn Asn Gly Thr Lys Ala Thr Pro Cys Leu Ser Ala Asp
    530                 535                 540

Ile Leu Gly Asp Trp Arg Glu Glu Val Ile Phe Arg Thr Ser Asp Asn
545                 550                 555                 560

Asn Ala Ile Arg Ile Tyr Thr Thr Thr Thr Leu Thr Asp Tyr Lys Ile
                565                 570                 575

Pro Thr Leu Met His Asn Arg Gln Tyr Arg Val Ser Ile Ala Trp Gln
```

-continued

```
            580                 585                 590

Asn Val Ala Tyr Asn Gln Pro Pro His Val Ser Phe Tyr Leu Gly Tyr
        595                 600                 605

Glu Thr Asn Val Asn Asn Ile Tyr Gln Tyr Phe Glu Gly Tyr Gly Gln
    610                 615                 620

Gln Pro Ile Val Thr Pro Ser Leu Thr Pro Thr Lys Thr Pro Thr Pro
625                 630                 635                 640

Thr Ser Thr Pro Leu Pro Thr Ser Thr Ala Thr Ser Thr Pro Thr Pro
                645                 650                 655

Thr Ala Thr Ala Thr Pro Thr Pro Thr Pro Thr Ala Thr Pro Thr Pro
            660                 665                 670

Thr Pro Ser Ser Thr Pro Val Ala Gly Gly Gln Ile Lys Val Leu Tyr
        675                 680                 685

Ala Asn Lys Glu Thr Asn Ser Thr Thr Asn Thr Ile Arg Pro Trp Leu
    690                 695                 700

Lys Val Val Asn Thr Gly Ser Ser Ser Ile Asp Leu Ser Arg Val Thr
705                 710                 715                 720

Ile Arg Tyr Trp Tyr Thr Val Asp Gly Asp Lys Ala Gln Ser Ala Ile
                725                 730                 735

Ser Asp Trp Ala Gln Ile Gly Ala Ser Asn Val Thr Phe Lys Phe Val
            740                 745                 750

Lys Leu Ser Ser Ser Val Ser Gly Ala Asp Tyr Tyr Leu Glu Ile Gly
        755                 760                 765

Phe Lys Ser Gly Ala Gly Gln Leu Gln Ala Gly Lys Asp Thr Gly Glu
    770                 775                 780

Ile Gln Ile Arg Phe Asn Lys Ser Asp Trp Ser Asn Tyr Asn Gln Gly
785                 790                 795                 800

Asn Asp Trp Ser Trp Met Gln Ser Met Thr Asn Tyr Gly Glu Asn Val
                805                 810                 815

Lys Val Thr Ala Tyr Val Asp Gly Val Leu Val Trp Gly Gln Glu Pro
            820                 825                 830

Lys


<210> SEQ ID NO 90
<211> LENGTH: 1383
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 90

atgataaaat ctaagaataa aaaagaggag gtttgggtga tgagtaacag gaagatttta      60

gccattgtag tcagtttgat aatggttgtt tcattgttta cagggattgg gttgcgtaat     120

gaagttgcaa aggcagcgac acttttaaca gatgattttg aagatggcaa cagagatgga     180

tggtcgacat cgaacggtag ttggagtgta gtagtggatg ggagcaaggt tttaaagcag     240

gctagcacag gttctgaggc gagagcatat actggttcat ctgattggag tgattataca     300

gttgaagcga aagttaaagt attaaatgtg aaggattcga gttcaggtgc gggagtgata     360

gtgagatata aaaactcagg taacttttat gcgttggtgc taaggggttc aaagatagaa     420

atagggaaga aattaaacag taactggagt acattggcgt tcaagtcatt tacgttggat     480

caggatacct ggtataatgt gaaattagaa gtaaatggga gcaagttagt tggatatgtt     540

aatgggagtc aagtattaag tgcaagtgat ttatcgatta cgacaggaaa agcaggttta     600
```

-continued

```
atagctgaca ggtgtgttgc tgaatttgat gatgttgttg taaattcaag tgtgagcggt    660

acagcaccta ctccgacacc aacaccgact tcatcagtga caccaacacc gacatcgact    720

ccaacgccaa ccaaaacacc tactccaact tccacaccag taccaacaca gaccccagca    780

gtaacaccga cgccgacccc aaatacgggt ggtgttttag ttattacaga tacaataatt    840

gtaaaatccg gtcaaacata tgatggtaaa ggaataaaaa taatagctca aggaatgggt    900

gacggaagtc aatctgaaaa tcaaaagccc atatttaaac ttgaaaaagg ggcaaatttg    960

aaaaatgtaa taattggagc gccaggttgt gacgggatac attgttatgg tgataatgtg   1020

gttgaaaatg ttgtatggga agatgttgga gaggatgcgt tgactgtaaa aagtgagggg   1080

gtagtggaag ttattggtgg ttcagcaaaa gaagctgctg acaaggtgtt ccaacttaat   1140

gcaccgtgta cattcaaagt aaaaaacttc acagctacaa atataggaaa gcttgtaaga   1200

caaaatggta atactacttt caaagtagtt atttatcttg aagatgtaac attaaacaat   1260

gtaaaaagct gtgttgcaaa atctgatagt ccagtatcag aactgtggta tcataacttg   1320

aatgtaaaca attgtaaaac attatttgaa tttccgtccc aatcacagat acatcaatac   1380

taa                                                                 1383
```

```
<210> SEQ ID NO 91
<211> LENGTH: 460
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 91

Met Ile Lys Ser Lys Asn Lys Lys Glu Glu Val Trp Val Met Ser Asn
1               5                   10                  15

Arg Lys Ile Leu Ala Ile Val Val Ser Leu Ile Met Val Val Ser Leu
            20                  25                  30

Phe Thr Gly Ile Gly Leu Arg Asn Glu Val Ala Lys Ala Ala Thr Leu
        35                  40                  45

Leu Thr Asp Asp Phe Glu Asp Gly Asn Arg Asp Gly Trp Ser Thr Ser
    50                  55                  60

Asn Gly Ser Trp Ser Val Val Val Asp Gly Ser Lys Val Leu Lys Gln
65                  70                  75                  80

Ala Ser Thr Gly Ser Glu Ala Arg Ala Tyr Thr Gly Ser Ser Asp Trp
                85                  90                  95

Ser Asp Tyr Thr Val Glu Ala Lys Val Lys Val Leu Asn Val Lys Asp
            100                 105                 110

Ser Ser Ser Gly Ala Gly Val Ile Val Arg Tyr Lys Asn Ser Gly Asn
        115                 120                 125

Phe Tyr Ala Leu Val Leu Arg Gly Ser Lys Ile Glu Ile Gly Lys Lys
    130                 135                 140

Leu Asn Ser Asn Trp Ser Thr Leu Ala Phe Lys Ser Phe Thr Leu Asp
145                 150                 155                 160

Gln Asp Thr Trp Tyr Asn Val Lys Leu Glu Val Asn Gly Ser Lys Leu
                165                 170                 175

Val Gly Tyr Val Asn Gly Ser Gln Val Leu Ser Ala Ser Asp Leu Ser
            180                 185                 190

Ile Thr Thr Gly Lys Ala Gly Leu Ile Ala Asp Arg Cys Val Ala Glu
        195                 200                 205
```

-continued

```
Phe Asp Asp Val Val Val Asn Ser Ser Val Ser Gly Thr Ala Pro Thr
    210                 215                 220

Pro Thr Pro Thr Pro Thr Ser Ser Val Thr Pro Thr Pro Thr Ser Thr
225                 230                 235                 240

Pro Thr Pro Thr Lys Thr Pro Thr Pro Thr Ser Thr Pro Val Pro Thr
                245                 250                 255

Gln Thr Pro Ala Val Thr Pro Thr Pro Thr Pro Asn Thr Gly Gly Val
            260                 265                 270

Leu Val Ile Thr Asp Thr Ile Ile Val Lys Ser Gly Gln Thr Tyr Asp
        275                 280                 285

Gly Lys Gly Ile Lys Ile Ile Ala Gln Gly Met Gly Asp Gly Ser Gln
    290                 295                 300

Ser Glu Asn Gln Lys Pro Ile Phe Lys Leu Glu Lys Gly Ala Asn Leu
305                 310                 315                 320

Lys Asn Val Ile Ile Gly Ala Pro Gly Cys Asp Gly Ile His Cys Tyr
                325                 330                 335

Gly Asp Asn Val Val Glu Asn Val Val Trp Glu Asp Val Gly Glu Asp
            340                 345                 350

Ala Leu Thr Val Lys Ser Glu Gly Val Val Glu Val Ile Gly Gly Ser
        355                 360                 365

Ala Lys Glu Ala Ala Asp Lys Val Phe Gln Leu Asn Ala Pro Cys Thr
    370                 375                 380

Phe Lys Val Lys Asn Phe Thr Ala Thr Asn Ile Gly Lys Leu Val Arg
385                 390                 395                 400

Gln Asn Gly Asn Thr Thr Phe Lys Val Val Ile Tyr Leu Glu Asp Val
                405                 410                 415

Thr Leu Asn Asn Val Lys Ser Cys Val Ala Lys Ser Asp Ser Pro Val
            420                 425                 430

Ser Glu Leu Trp Tyr His Asn Leu Asn Val Asn Asn Cys Lys Thr Leu
        435                 440                 445

Phe Glu Phe Pro Ser Gln Ser Gln Ile His Gln Tyr
    450                 455                 460


<210> SEQ ID NO 92
<211> LENGTH: 1962
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 92

atgagtaaca ggaagatttt agccattgta gtcagtttga taatggttgt ttcattgttt     60

acagggattg ggttgcgtaa tgaagttgca aaggcagcga cacttttaac agatgatttt    120

gaagatggca acagagatgg atggtcgaca tcgaacggta gttggagtgt agtagtggat    180

gggagcaagg ttttaaagca ggctagcaca ggttctgagg cgagagcata tactggttca    240

tctgattgga gtgattatac agttgaagcg aaagttaaag tattaaatgt gaaggattcg    300

agttcaggtg cgggagtgat agtgagatat aaaaactcag gtaactttta tgcgttggtg    360

ctaaggggtt caaagataga aatagggaag aaattaaaca gtaactggag tacattggcg    420

ttcaagtcat ttacgttgga tcaggatacc tggtataatg tgaaattaga agtaaatggg    480

agcaagttag ttggatatgt taatgggagt caagtattaa gtgcaagtga tttatcgatt    540

acgacaggaa aagcaggttt aatagctgac aggtgtgttg ctgaatttga tgatgttgtt    600
```

-continued

```
gtaaattcaa gtgtgagcgg tacagcacct actccgacac caacaccgac ttcatcagtg    660

acaccaacac cgacatcgac tccaacgcca accaaaacac ctactccaac ttccacacca    720

gtaccaacac agaccccagc agtaacaccg acgccgaccc caacgccgac gacagttcca    780

acacctgccc cgacacctgt acctggggtg aatgctattt atgtggcacc aagtgggagc    840

tcagataatc ctggtaccat tgatcgacct actacattag aaaaagcaat cacgatagta    900

caacctgggc agataatcta catgagaggt gggacgtata agtattctgc gcagatcaca    960

attgaaagaa ataatagtgg tacaagcaat gcaagaaaat gtatttatgc atttccaaat   1020

gaaaggccaa tattggactt ttcatctcaa acatatggga gtgtggactc aaatccaaga   1080

ggattacaga ttaatgggaa ctattggcac ataaaaggat tagaagtcat gggagctgcg   1140

gacaacggaa tcttcgtagg aggcagctac aatataattg aacaatgtga aattcatcat   1200

aacagagatt caggtttgca gataagcagg tatataagtt ctgcaaccag agatgagtgg   1260

ccaagttata acttgatatt gaattgtaca tcacatgaca atatggatcc agataacggt   1320

gaagatgcag atggttttgc atgcaaacta acagcaggac caggaaatgt attccgaggt   1380

tgtgtagcgt actacaatgt tgatgatggt tgggatttat acacaaaaag tgagacagga   1440

gctattggtg aagtattaat tgaggattgt gtggcatatg gtcacgggca aacatcaacc   1500

gggagtgcca catctagcag tgatggaaat ggctttaagc taggaggcag taatataaag   1560

gtcaatcata cagtgagaag atgtatagca tttaataaca acaaacatgg atttacttat   1620

aatagtaatc cgggtagcat aacagtggaa aattgtacgg gctataataa cggtttaaag   1680

gtaagtggaa ggaactttta ttttgaagaa ggtacacacg tgttgaagaa ttgtttatcc   1740

tacaaagaga gtgcatcgag tgatttggta agtggaacga taattaattg tgttttgtgg   1800

agtaataggc aagcaataaa gctaaatggt caactggtaa ccgataatga cttttacagc   1860

ttaacaccaa ccataacaag gaatagtgat gggggtttaa acttaggaga ctttttaaag   1920

ccaaagcctg gtagtggttt agaaggaata ggagcaaggt aa                      1962
```

<210> SEQ ID NO 93
<211> LENGTH: 653
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 93

```
Met Ser Asn Arg Lys Ile Leu Ala Ile Val Val Ser Leu Ile Met Val
1               5                   10                  15

Val Ser Leu Phe Thr Gly Ile Gly Leu Arg Asn Glu Val Ala Lys Ala
            20                  25                  30

Ala Thr Leu Leu Thr Asp Asp Phe Glu Asp Gly Asn Arg Asp Gly Trp
        35                  40                  45

Ser Thr Ser Asn Gly Ser Trp Ser Val Val Val Asp Gly Ser Lys Val
    50                  55                  60

Leu Lys Gln Ala Ser Thr Gly Ser Glu Ala Arg Ala Tyr Thr Gly Ser
65                  70                  75                  80

Ser Asp Trp Ser Asp Tyr Thr Val Glu Ala Lys Val Lys Val Leu Asn
                85                  90                  95

Val Lys Asp Ser Ser Ser Gly Ala Gly Val Ile Val Arg Tyr Lys Asn
            100                 105                 110

Ser Gly Asn Phe Tyr Ala Leu Val Leu Arg Gly Ser Lys Ile Glu Ile
```

-continued

```
        115                 120                 125

Gly Lys Lys Leu Asn Ser Asn Trp Ser Thr Leu Ala Phe Lys Ser Phe
    130                 135                 140

Thr Leu Asp Gln Asp Thr Trp Tyr Asn Val Lys Leu Glu Val Asn Gly
145                 150                 155                 160

Ser Lys Leu Val Gly Tyr Val Asn Gly Ser Gln Val Leu Ser Ala Ser
                165                 170                 175

Asp Leu Ser Ile Thr Thr Gly Lys Ala Gly Leu Ile Ala Asp Arg Cys
            180                 185                 190

Val Ala Glu Phe Asp Asp Val Val Val Asn Ser Ser Val Ser Gly Thr
        195                 200                 205

Ala Pro Thr Pro Thr Pro Thr Pro Thr Ser Ser Val Thr Pro Thr Pro
    210                 215                 220

Thr Ser Thr Pro Thr Pro Thr Lys Thr Pro Thr Pro Thr Ser Thr Pro
225                 230                 235                 240

Val Pro Thr Gln Thr Pro Ala Val Thr Pro Thr Pro Thr Pro Thr Pro
                245                 250                 255

Thr Thr Val Pro Thr Pro Ala Pro Thr Pro Val Pro Gly Val Asn Ala
            260                 265                 270

Ile Tyr Val Ala Pro Ser Gly Ser Ser Asp Asn Pro Gly Thr Ile Asp
        275                 280                 285

Arg Pro Thr Thr Leu Glu Lys Ala Ile Thr Ile Val Gln Pro Gly Gln
    290                 295                 300

Ile Ile Tyr Met Arg Gly Gly Thr Tyr Lys Tyr Ser Ala Gln Ile Thr
305                 310                 315                 320

Ile Glu Arg Asn Asn Ser Gly Thr Ser Asn Ala Arg Lys Cys Ile Tyr
                325                 330                 335

Ala Phe Pro Asn Glu Arg Pro Ile Leu Asp Phe Ser Ser Gln Thr Tyr
            340                 345                 350

Gly Ser Val Asp Ser Asn Pro Arg Gly Leu Gln Ile Asn Gly Asn Tyr
        355                 360                 365

Trp His Ile Lys Gly Leu Glu Val Met Gly Ala Ala Asp Asn Gly Ile
    370                 375                 380

Phe Val Gly Gly Ser Tyr Asn Ile Ile Glu Gln Cys Glu Ile His His
385                 390                 395                 400

Asn Arg Asp Ser Gly Leu Gln Ile Ser Arg Tyr Ile Ser Ser Ala Thr
                405                 410                 415

Arg Asp Glu Trp Pro Ser Tyr Asn Leu Ile Leu Asn Cys Thr Ser His
            420                 425                 430

Asp Asn Met Asp Pro Asp Asn Gly Glu Asp Ala Asp Gly Phe Ala Cys
        435                 440                 445

Lys Leu Thr Ala Gly Pro Gly Asn Val Phe Arg Gly Cys Val Ala Tyr
    450                 455                 460

Tyr Asn Val Asp Asp Gly Trp Asp Leu Tyr Thr Lys Ser Glu Thr Gly
465                 470                 475                 480

Ala Ile Gly Glu Val Leu Ile Glu Asp Cys Val Ala Tyr Gly His Gly
                485                 490                 495

Gln Thr Ser Thr Gly Ser Ala Thr Ser Ser Ser Asp Gly Asn Gly Phe
            500                 505                 510

Lys Leu Gly Gly Ser Asn Ile Lys Val Asn His Thr Val Arg Arg Cys
        515                 520                 525
```

-continued

```
Ile Ala Phe Asn Asn Asn Lys His Gly Phe Thr Tyr Asn Ser Asn Pro
    530                 535                 540

Gly Ser Ile Thr Val Glu Asn Cys Thr Gly Tyr Asn Asn Gly Leu Lys
545                 550                 555                 560

Val Ser Gly Arg Asn Phe Tyr Phe Glu Glu Gly Thr His Val Leu Lys
                565                 570                 575

Asn Cys Leu Ser Tyr Lys Glu Ser Ala Ser Ser Asp Leu Val Ser Gly
            580                 585                 590

Thr Ile Ile Asn Cys Val Leu Trp Ser Asn Arg Gln Ala Ile Lys Leu
        595                 600                 605

Asn Gly Gln Leu Val Thr Asp Asn Asp Phe Tyr Ser Leu Thr Pro Thr
    610                 615                 620

Ile Thr Arg Asn Ser Asp Gly Gly Leu Asn Leu Gly Asp Phe Leu Lys
625                 630                 635                 640

Pro Lys Pro Gly Ser Gly Leu Glu Gly Ile Gly Ala Arg
                645                 650


<210> SEQ ID NO 94
<211> LENGTH: 2382
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 94

atgggaagaa ccatatgtat taaaaaaagt gcaaatatgt tcttaatatt tttagttatg     60

gtttttcttt ttcattttct aatgtcagtt ttcttattat atatgactaa agaacttatt    120

ttgagagaca taattaatga taagtcactt caattagata aacttcgcga aaaaatagac    180

aatgaactaa agaaaatgaa tgagatagtt acacgtattc ttactgatga tgaattaaaa    240

tgggtaggga atctaagggg gttaagagaa aattctttag atttatggga gtactttgaa    300

tattataaac actttaaaga cataggcatg attaatggag agttaaaacc tatcattgta    360

ctttttttac gcgagggaga aatagtatat tttgcttccg aagaatttgg ctctttttt    420

acatttggat ttgaaaattt ttgtaattac ttttcacctg acagaatgaa ctgccgcgtt    480

tggctcaata aaatattcga taaaaacaac agtaaatgca aaaataagtt aatatcacaa    540

gaatattcaa ttagtggcca aagaatattg gcattgcatg aaacatatta ctttccattt    600

gataacattg gaaatcaatt ggcaattctt ctggtaatta ttgatgttgc taaattttgt    660

aagctactaa aggaaaacaa attaaacagc caagattcaa taattttaat ttatgatagg    720

tgtgagaaaa aaatattaac ttcaaataaa gcagatatta ataatacaat aaataacata    780

ttagaaaaac ttagcaaaaa taagaaatat atttctaatc tgtataatgt aataacaata    840

cgaaataaaa agtatatatt tttacggttg gcttcaaatg tatatgattg ggattatgtt    900

tatctagtac catataacag cataaaaagt gaaatataca tttcaagaac tgtttataca    960

ttgtatctaa cggagtttac acttttcata attattataa gtttatatgc tatttatatc   1020

aaattatata aagtaaggac atataaaata actaagggct taaaaaatgt aaacgaggag   1080

aatgagaata aaatacttac ttttaggtgc ataaacaagt taaaggagaa agaccgtagg   1140

ttaatacaca ctaaaattac aagttacagg attcttaata aaataaccag taaacaagaa   1200

ttattaaacg atatgataat agaaaagctt atttatggtt ggtcttatag caaaggggca   1260

atagaagaaa aaatccaaag tataggctta aaaattggag gtaaaaaatt tttagttgcc   1320
```

-continued

```
ataatcaaaa tgctctctct taccaaacaa gttagaacaa gagatataat aaaaaatgag   1380

ttagaaagta ttagggttga ctcaaaaatt aacttggtat tttattatgt ttaccaactt   1440

gcagatagtg atcttgcttt aattttagca tttgatgaag atgaagatga aaaagtaagt   1500

caaaatgtaa atctatggct gagattgata ggtgatagat tagtaattaa aatatgtcac   1560

aaatatttaa ttgcagttgg tcgtatagtt aataacattg acgaatgtag aatttcattt   1620

ttagatgcaa aagaaattat tgaagctaat aaattaatca caacagagaa cgttgagaat   1680

ggcatactgt ggtattataa tgttgttaag aaagataata atatatggta tccaatagaa   1740

atagaagaga gattaatgtt gctagttaat cttgggaaaa tatctgagat tgaagaaatt   1800

ctacacctat tatattataa aaactttgaa gaaaaagata tatcactgag tttaaagtat   1860

gtgctaatta gtgagttaat tggaacaatt attaaaattg caaatatgac aaaagtaaat   1920

attgacaatc tatttgacat taagaacttt atttttattg aagaaaatga ctttgacagg   1980

ttatttgaag atataaagca agtatttatt aatataacag aacaaatcaa gagaaacagg   2040

aaggcttcta aggaaaaatt aatagaggag atattagaat ttataaacaa aaacttattt   2100

aatccaaaca tgagtatatc ccttgtcgct gaaagattta atttatctga atcttatttt   2160

tctaatattt ttaaaaatgc agtaggtata aagttcagtg attatgtaga aaagttaaga   2220

atagaagaag catataaatt gataaagcaa aaaaagtgga atttagacga tattagtaaa   2280

atggtaggat atactaacat taaaaccttt agaagagctt ttaaaagggt aaaaggttgt   2340

ttacccagtg aaattttaaa cataaatgaa catgatatat ag                      2382
```

```
<210> SEQ ID NO 95
<211> LENGTH: 793
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 95

Met Gly Arg Thr Ile Cys Ile Lys Lys Ser Ala Asn Met Phe Leu Ile
1               5                   10                  15

Phe Leu Val Met Val Phe Leu Phe His Phe Leu Met Ser Val Phe Leu
            20                  25                  30

Leu Tyr Met Thr Lys Glu Leu Ile Leu Arg Asp Ile Ile Asn Asp Lys
        35                  40                  45

Ser Leu Gln Leu Asp Lys Leu Arg Glu Lys Ile Asp Asn Glu Leu Lys
    50                  55                  60

Lys Met Asn Glu Ile Val Thr Arg Ile Leu Thr Asp Asp Glu Leu Lys
65                  70                  75                  80

Trp Val Gly Asn Leu Arg Gly Leu Arg Glu Asn Ser Leu Asp Leu Trp
                85                  90                  95

Glu Tyr Phe Glu Tyr Tyr Lys His Phe Lys Asp Ile Gly Met Ile Asn
            100                 105                 110

Gly Glu Leu Lys Pro Ile Ile Val Leu Phe Leu Arg Glu Gly Glu Ile
        115                 120                 125

Val Tyr Phe Ala Ser Glu Glu Phe Gly Ser Phe Phe Thr Phe Gly Phe
    130                 135                 140

Glu Asn Phe Cys Asn Tyr Phe Ser Pro Asp Arg Met Asn Cys Arg Val
145                 150                 155                 160

Trp Leu Asn Lys Ile Phe Asp Lys Asn Asn Ser Lys Cys Lys Asn Lys
```

-continued

```
                165                 170                 175

Leu Ile Ser Gln Glu Tyr Ser Ile Ser Gly Gln Arg Ile Leu Ala Leu
            180                 185                 190

His Glu Thr Tyr Tyr Phe Pro Phe Asp Asn Ile Gly Asn Gln Leu Ala
        195                 200                 205

Ile Leu Leu Val Ile Ile Asp Val Ala Lys Phe Cys Lys Leu Leu Lys
    210                 215                 220

Glu Asn Lys Leu Asn Ser Gln Asp Ser Ile Ile Leu Ile Tyr Asp Arg
225                 230                 235                 240

Cys Glu Lys Lys Ile Leu Thr Ser Asn Lys Ala Asp Ile Asn Asn Thr
                245                 250                 255

Ile Asn Asn Ile Leu Glu Lys Leu Ser Lys Asn Lys Lys Tyr Ile Ser
            260                 265                 270

Asn Leu Tyr Asn Val Ile Thr Ile Arg Asn Lys Lys Tyr Ile Phe Leu
        275                 280                 285

Arg Leu Ala Ser Asn Val Tyr Asp Trp Asp Tyr Val Tyr Leu Val Pro
    290                 295                 300

Tyr Asn Ser Ile Lys Ser Glu Ile Tyr Ile Ser Arg Thr Val Tyr Thr
305                 310                 315                 320

Leu Tyr Leu Thr Glu Phe Thr Leu Phe Ile Ile Ile Ile Ser Leu Tyr
                325                 330                 335

Ala Ile Tyr Ile Lys Leu Tyr Lys Val Arg Thr Tyr Lys Ile Thr Lys
            340                 345                 350

Gly Leu Lys Asn Val Asn Glu Glu Asn Glu Asn Lys Ile Leu Thr Phe
        355                 360                 365

Arg Cys Ile Asn Lys Leu Lys Glu Lys Asp Arg Arg Leu Ile His Thr
    370                 375                 380

Lys Ile Thr Ser Tyr Arg Ile Leu Asn Lys Ile Thr Ser Lys Gln Glu
385                 390                 395                 400

Leu Leu Asn Asp Met Ile Ile Glu Lys Leu Ile Tyr Gly Trp Ser Tyr
                405                 410                 415

Ser Lys Gly Ala Ile Glu Glu Lys Ile Gln Ser Ile Gly Leu Lys Ile
            420                 425                 430

Gly Gly Lys Lys Phe Leu Val Ala Ile Ile Lys Met Leu Ser Leu Thr
        435                 440                 445

Lys Gln Val Arg Thr Arg Asp Ile Ile Lys Asn Glu Leu Glu Ser Ile
    450                 455                 460

Arg Val Asp Ser Lys Ile Asn Leu Val Phe Tyr Tyr Val Tyr Gln Leu
465                 470                 475                 480

Ala Asp Ser Asp Leu Ala Leu Ile Leu Ala Phe Asp Glu Asp Glu Asp
                485                 490                 495

Glu Lys Val Ser Gln Asn Val Asn Leu Trp Leu Arg Leu Ile Gly Asp
            500                 505                 510

Arg Leu Val Ile Lys Ile Cys His Lys Tyr Leu Ile Ala Val Gly Arg
        515                 520                 525

Ile Val Asn Asn Ile Asp Glu Cys Arg Ile Ser Phe Leu Asp Ala Lys
    530                 535                 540

Glu Ile Ile Glu Ala Asn Lys Leu Ile Thr Thr Glu Asn Val Glu Asn
545                 550                 555                 560

Gly Ile Leu Trp Tyr Tyr Asn Val Val Lys Lys Asp Asn Asn Ile Trp
                565                 570                 575
```

-continued

```
Tyr Pro Ile Glu Ile Glu Glu Arg Leu Met Leu Leu Val Asn Leu Gly
            580                 585                 590

Lys Ile Ser Glu Ile Glu Glu Ile Leu His Leu Leu Tyr Tyr Lys Asn
        595                 600                 605

Phe Glu Glu Lys Asp Ile Ser Leu Ser Leu Lys Tyr Val Leu Ile Ser
    610                 615                 620

Glu Leu Ile Gly Thr Ile Ile Lys Ile Ala Asn Met Thr Lys Val Asn
625                 630                 635                 640

Ile Asp Asn Leu Phe Asp Ile Lys Asn Phe Ile Phe Ile Glu Glu Asn
                645                 650                 655

Asp Phe Asp Arg Leu Phe Glu Asp Ile Lys Gln Val Phe Ile Asn Ile
            660                 665                 670

Thr Glu Gln Ile Lys Arg Asn Arg Lys Ala Ser Lys Glu Lys Leu Ile
        675                 680                 685

Glu Glu Ile Leu Glu Phe Ile Asn Lys Asn Leu Phe Asn Pro Asn Met
    690                 695                 700

Ser Ile Ser Leu Val Ala Glu Arg Phe Asn Leu Ser Glu Ser Tyr Phe
705                 710                 715                 720

Ser Asn Ile Phe Lys Asn Ala Val Gly Ile Lys Phe Ser Asp Tyr Val
                725                 730                 735

Glu Lys Leu Arg Ile Glu Glu Ala Tyr Lys Leu Ile Lys Gln Lys Lys
            740                 745                 750

Trp Asn Leu Asp Asp Ile Ser Lys Met Val Gly Tyr Thr Asn Ile Lys
        755                 760                 765

Thr Phe Arg Arg Ala Phe Lys Arg Val Lys Gly Cys Leu Pro Ser Glu
    770                 775                 780

Ile Leu Asn Ile Asn Glu His Asp Ile
785                 790


<210> SEQ ID NO 96
<211> LENGTH: 1818
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 96

gtgggtgatg taaattctgt caattccgct ttaaaaaact tcaatattgt agttgcttct     60

gccaacaaca aagctgtgga aaatgtcacc aaagaaatac cagttctaaa ttcagttgac    120

caatcctgtc ttgaaaaata tgaactccat tatttcaaag atggagccga actggtatac    180

gactacagag gggaaaacaa tgaaagcagt gagaaaactg aatctttagt atcagttgaa    240

aatgggcagc aatgctgggc attgatttca gcagttttgg gtaaaaaaga aaacagagaa    300

aagtttttca gtgctctgga aaagtatatt aatgaattat tttcttctat accgccagtc    360

aagtgggaac agtgtaaaag aagtttcaat catgtcttca agaagtttag acatgtccag    420

aaactttaca ggactatgga actttttctg aaactcgaag aaaacaggtt taaatttacc    480

cgtatgatag ctgctgcggg gaagatttgc ctaaaaaaat tttttgccag atattttccc    540

gatgacatga aacttccttc ctcagagttc tggtcccaaa atgagcacga aattcacaaa    600

tcatctccat ggatgagcag gtatctcaat gatatacgca atgaagtttt tgcaagagct    660

cttcagttac atcaggcagc tattgctgca aacaaagaga agttcgggga gaatttagaa    720

aaattcatta aatatatgag aaaaaacgaa tcactgcctg cagaaaaggc aaaagaactg    780
```

-continued

```
tggcatactt tctttatgat tgttccagtt gtctcaacaa catttgcgtc gctgagcaat    840

atgtttaaag atgttgacga cgaaatcatt gactggctga ttgttgatga ggctgggcag    900

gcactacccc agcattttat cggcgcactc ctcagaagta aaagagcaat aatagtggga    960

gaccctctac aaattccacc tgttgttaaa ataccaccat ttgttattaa tgatgtgttc   1020

aaagcatatg ggattttcaa atggagacag gaagacagta attccagtac accacgtata   1080

actgaaactg attccgtaca aattgtagcc gatagagcca acaaattcgg tgcaaagatt   1140

ggcgatatgt gggttgggtg tcctctgaga gtacataagc gttgtacgga accaatgttt   1200

acagtagcga atagaattgc ttataaaaac cttatgattt ttgatgtgaa taaacctgaa   1260

ggattgcaaa ctgtttttaa agatagcttc tgggtagatg taaagggtaa gtgtgtttac   1320

aagcattatg taagagaaca gggaaaagta gtgaaagcta tcgtacaaga atttttgcag   1380

cgaaccttga caactcaaaa tgaggttaaa ttaacagaag agctttttat catttcgcct   1440

tttaaagcag tcaaaagtag tatatcttct atactgaaaa gaatgcctct gtataacaca   1500

aatttaaaga aagaagaatg ggagaatata gttaatgaaa tagttggcac aattcacagt   1560

tttcagggca agcaagccaa caatgtaatc atatgccttg gtgccgatga aagtaatgaa   1620

ggagctgtaa gatgggcatc ttcagaaccg aatattttaa atgtggcgct gaccagagct   1680

aaatatagag tcatagtaat aggtgataaa gacctgtggg gaaagcataa atattttgat   1740

acactcttag aagaactagg tgaaaaagtg attgagtata ccacagaaaa agatctggtc   1800

aacaagattt ttgcttag                                                 1818
```

```
<210> SEQ ID NO 97
<211> LENGTH: 605
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 97

Val Gly Asp Val Asn Ser Val Asn Ser Ala Leu Lys Asn Phe Asn Ile
1               5                   10                  15

Val Val Ala Ser Ala Asn Asn Lys Ala Val Glu Asn Val Thr Lys Glu
            20                  25                  30

Ile Pro Val Leu Asn Ser Val Asp Gln Ser Cys Leu Glu Lys Tyr Glu
        35                  40                  45

Leu His Tyr Phe Lys Asp Gly Ala Glu Leu Val Tyr Asp Tyr Arg Gly
    50                  55                  60

Glu Asn Asn Glu Ser Ser Glu Lys Thr Glu Ser Leu Val Ser Val Glu
65                  70                  75                  80

Asn Gly Gln Gln Cys Trp Ala Leu Ile Ser Ala Val Leu Gly Lys Lys
                85                  90                  95

Glu Asn Arg Glu Lys Phe Phe Ser Ala Leu Glu Lys Tyr Ile Asn Glu
            100                 105                 110

Leu Phe Ser Ser Ile Pro Pro Val Lys Trp Glu Gln Cys Lys Arg Ser
        115                 120                 125

Phe Asn His Val Phe Lys Lys Phe Arg His Val Gln Lys Leu Tyr Arg
    130                 135                 140

Thr Met Glu Leu Phe Leu Lys Leu Glu Glu Asn Arg Phe Lys Phe Thr
145                 150                 155                 160

Arg Met Ile Ala Ala Ala Gly Lys Ile Cys Leu Lys Lys Phe Phe Ala
```

-continued

```
                165                 170                 175

Arg Tyr Phe Pro Asp Asp Met Lys Leu Pro Ser Ser Glu Phe Trp Ser
            180                 185                 190

Gln Asn Glu His Glu Ile His Lys Ser Ser Pro Trp Met Ser Arg Tyr
        195                 200                 205

Leu Asn Asp Ile Arg Asn Glu Val Phe Ala Arg Ala Leu Gln Leu His
    210                 215                 220

Gln Ala Ala Ile Ala Ala Asn Lys Glu Lys Phe Gly Glu Asn Leu Glu
225                 230                 235                 240

Lys Phe Ile Lys Tyr Met Arg Lys Asn Glu Ser Leu Pro Ala Glu Lys
                245                 250                 255

Ala Lys Glu Leu Trp His Thr Phe Phe Met Ile Val Pro Val Val Ser
            260                 265                 270

Thr Thr Phe Ala Ser Leu Ser Asn Met Phe Lys Asp Val Asp Asp Glu
        275                 280                 285

Ile Ile Asp Trp Leu Ile Val Asp Glu Ala Gly Gln Ala Leu Pro Gln
    290                 295                 300

His Phe Ile Gly Ala Leu Leu Arg Ser Lys Arg Ala Ile Ile Val Gly
305                 310                 315                 320

Asp Pro Leu Gln Ile Pro Pro Val Val Lys Ile Pro Pro Phe Val Ile
                325                 330                 335

Asn Asp Val Phe Lys Ala Tyr Gly Ile Phe Lys Trp Arg Gln Glu Asp
            340                 345                 350

Ser Asn Ser Ser Thr Pro Arg Ile Thr Glu Thr Asp Ser Val Gln Ile
        355                 360                 365

Val Ala Asp Arg Ala Asn Lys Phe Gly Ala Lys Ile Gly Asp Met Trp
    370                 375                 380

Val Gly Cys Pro Leu Arg Val His Lys Arg Cys Thr Glu Pro Met Phe
385                 390                 395                 400

Thr Val Ala Asn Arg Ile Ala Tyr Lys Asn Leu Met Ile Phe Asp Val
                405                 410                 415

Asn Lys Pro Glu Gly Leu Gln Thr Val Phe Lys Asp Ser Phe Trp Val
            420                 425                 430

Asp Val Lys Gly Lys Cys Val Tyr Lys His Tyr Val Arg Glu Gln Gly
        435                 440                 445

Lys Val Val Lys Ala Ile Val Gln Glu Phe Leu Gln Arg Thr Leu Thr
    450                 455                 460

Thr Gln Asn Glu Val Lys Leu Thr Glu Glu Leu Phe Ile Ile Ser Pro
465                 470                 475                 480

Phe Lys Ala Val Lys Ser Ser Ile Ser Ser Ile Leu Lys Arg Met Pro
                485                 490                 495

Leu Tyr Asn Thr Asn Leu Lys Lys Glu Glu Trp Glu Asn Ile Val Asn
            500                 505                 510

Glu Ile Val Gly Thr Ile His Ser Phe Gln Gly Lys Gln Ala Asn Asn
        515                 520                 525

Val Ile Ile Cys Leu Gly Ala Asp Glu Ser Asn Glu Gly Ala Val Arg
    530                 535                 540

Trp Ala Ser Ser Glu Pro Asn Ile Leu Asn Val Ala Leu Thr Arg Ala
545                 550                 555                 560

Lys Tyr Arg Val Ile Val Ile Gly Asp Lys Asp Leu Trp Gly Lys His
                565                 570                 575
```

-continued

```
Lys Tyr Phe Asp Thr Leu Leu Glu Glu Leu Gly Glu Lys Val Ile Glu
            580                 585                 590

Tyr Thr Thr Glu Lys Asp Leu Val Asn Lys Ile Phe Ala
        595                 600                 605


<210> SEQ ID NO 98
<211> LENGTH: 330
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 98

ttgaaagata cttttgaaca gttaattaaa aatcagaaaa atattagaga aatctatatg     60

aatttactcc agaaattttg ttacgatatt tatccagtca ggaacaaaga atttacaaat    120

tcattcggtg ttacatggct taattccaaa aaaggcaaac aacacgaaca aagaagaaaa    180

gtcgagatag aaaacgaaat agaggaagaa attgaaacgc cactgcaaga ccttttttgcc   240

agtttttatc taaaagacat tgaagaagca cttaaacgcc ttgagaatga taatctaaac    300

cctgtcttgt ttcgcttctt tactgagtaa                                     330


<210> SEQ ID NO 99
<211> LENGTH: 109
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 99

Leu Lys Asp Thr Phe Glu Gln Leu Ile Lys Asn Gln Lys Asn Ile Arg
1               5                   10                  15

Glu Ile Tyr Met Asn Leu Leu Gln Lys Phe Cys Tyr Asp Ile Tyr Pro
            20                  25                  30

Val Arg Asn Lys Glu Phe Thr Asn Ser Phe Gly Val Thr Trp Leu Asn
        35                  40                  45

Ser Lys Lys Gly Lys Gln His Glu Gln Arg Arg Lys Val Glu Ile Glu
    50                  55                  60

Asn Glu Ile Glu Glu Glu Ile Glu Thr Pro Leu Gln Asp Leu Phe Ala
65                  70                  75                  80

Ser Phe Tyr Leu Lys Asp Ile Glu Glu Ala Leu Lys Arg Leu Glu Asn
                85                  90                  95

Asp Asn Leu Asn Pro Val Leu Phe Arg Phe Phe Thr Glu
            100                 105


<210> SEQ ID NO 100
<211> LENGTH: 474
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 100

atgataaata caattgcaaa agatgagatg aaaagaatac tgctggcata ccagaaactt     60

gaattgttgc agcctgctcc tattcccaat ggagttttta taaacgactt tactgaaatc    120

tggagtccta actatttcac agattgcaga gtctatcttg gaatttatga ttaccaggaa    180

gtaaaagagg cttttgaaaa aatttataac cagcaatatg accatcccga aaaagaagat    240

gaaaaacaaa acagaaaaca aaaactttac tggttctatt tcgatattga tgagaacggg    300

aaatataaag atgacagttt aaaaatttct tcaactttgt gggctcttaa acaactggaa    360
```

-continued

```
ctggaaaata caaacaaaag aaggctgtca gaaactgcgg agaaaaaaga tgaagaaaag    420

caaactgaag attatgatct tggtaaaaat ttacagagtt tgctcaaaac ttga         474


<210> SEQ ID NO 101
<211> LENGTH: 157
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 101

Met Ile Asn Thr Ile Ala Lys Asp Glu Met Lys Arg Ile Leu Leu Ala
1               5                   10                  15

Tyr Gln Lys Leu Glu Leu Leu Gln Pro Ala Pro Ile Pro Asn Gly Val
            20                  25                  30

Phe Ile Asn Asp Phe Thr Glu Ile Trp Ser Pro Asn Tyr Phe Thr Asp
        35                  40                  45

Cys Arg Val Tyr Leu Gly Ile Tyr Asp Tyr Gln Glu Val Lys Glu Ala
    50                  55                  60

Phe Glu Lys Ile Tyr Asn Gln Gln Tyr Asp His Pro Glu Lys Glu Asp
65                  70                  75                  80

Glu Lys Gln Asn Arg Lys Gln Lys Leu Tyr Trp Phe Tyr Phe Asp Ile
                85                  90                  95

Asp Glu Asn Gly Lys Tyr Lys Asp Asp Ser Leu Lys Ile Ser Ser Thr
            100                 105                 110

Leu Trp Ala Leu Lys Gln Leu Glu Leu Glu Asn Thr Asn Lys Arg Arg
        115                 120                 125

Leu Ser Glu Thr Ala Glu Lys Lys Asp Glu Glu Lys Gln Thr Glu Asp
    130                 135                 140

Tyr Asp Leu Gly Lys Asn Leu Gln Ser Leu Leu Lys Thr
145                 150                 155


<210> SEQ ID NO 102
<211> LENGTH: 186
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 102

atggctgact ataagaaagc atatacatta cgaatagatg aaactttgct tgataagatt     60

agagtaatcg ccgagaaaaa taaacgctca attaatgctc aaattgaata tttgattcaa    120

cagtgtgtcg aagaatttga agctgagcac ggtgagataa aaattgaaga agaaaatact    180

gaatga                                                               186


<210> SEQ ID NO 103
<211> LENGTH: 61
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 103

Met Ala Asp Tyr Lys Lys Ala Tyr Thr Leu Arg Ile Asp Glu Thr Leu
1               5                   10                  15

Leu Asp Lys Ile Arg Val Ile Ala Glu Lys Asn Lys Arg Ser Ile Asn
            20                  25                  30

Ala Gln Ile Glu Tyr Leu Ile Gln Gln Cys Val Glu Glu Phe Glu Ala
        35                  40                  45

Glu His Gly Glu Ile Lys Ile Glu Glu Glu Asn Thr Glu
```

-continued

```
    50                  55                  60


<210> SEQ ID NO 104
<211> LENGTH: 843
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 104

atggatatat ttaccttaat aagaacaaaa tataatactt tatctcaaac tcaaaagaaa     60

attgcagatt ggattttaga acatggaaat gaaattattt taatgtcaat aggcgaaatc    120

gcagcaaagt gttcaacaag tgaagctact gtaatgagat ttttaagaaa acttggtttt    180

gattcatatc agctctttaa ggtaaaagtg gctcaagata ttgttgatgt gactccccga    240

gccgtttatg aggatgtaag tagtgaagat agtgttaatg agataaaaca gaaagtaatt    300

caatctacaa ttgatgctat cagagatata gataagttaa ttgaagatag cacaatagaa    360

aaagctatag agataatggc taatgcaaaa agaatctttt tctttggagt gggagcatca    420

ggagcaatag caaaggacgc atttcacaag tttttaagat taggcataaa taccatatat    480

tgtagtgatt ctcatattat gagcataatg tgtagccata tgacagaaaa agatgccgtt    540

ttggcgattt cacactcagg agaaagtcga gaaataattg atgcaataga acttgcaaaa    600

gaaaacaaag caaaagtaat ttcgtttaca agttacccga attctacctt ggcaaaactt    660

tcagatgtgg tttttctaag tgctaccaaa gagactaaat tcagatctga cgcgatggtt    720

tcaagaattg tgcaatgtgt gattatagac attctctatg tttctctagt tttaaagtta    780

ggaagtgaag caattgtaaa tgtgaataaa tcaagattag cagtcgcaaa aaagaaaaaa    840

taa                                                                  843


<210> SEQ ID NO 105
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 105

Met Asp Ile Phe Thr Leu Ile Arg Thr Lys Tyr Asn Thr Leu Ser Gln
1               5                   10                  15

Thr Gln Lys Lys Ile Ala Asp Trp Ile Leu Glu His Gly Asn Glu Ile
            20                  25                  30

Ile Leu Met Ser Ile Gly Glu Ile Ala Ala Lys Cys Ser Thr Ser Glu
        35                  40                  45

Ala Thr Val Met Arg Phe Leu Arg Lys Leu Gly Phe Asp Ser Tyr Gln
    50                  55                  60

Leu Phe Lys Val Lys Val Ala Gln Asp Ile Val Asp Val Thr Pro Arg
65                  70                  75                  80

Ala Val Tyr Glu Asp Val Ser Ser Glu Asp Ser Val Asn Glu Ile Lys
                85                  90                  95

Gln Lys Val Ile Gln Ser Thr Ile Asp Ala Ile Arg Asp Ile Asp Lys
            100                 105                 110

Leu Ile Glu Asp Ser Thr Ile Glu Lys Ala Ile Glu Ile Met Ala Asn
        115                 120                 125

Ala Lys Arg Ile Phe Phe Phe Gly Val Gly Ala Ser Gly Ala Ile Ala
    130                 135                 140

Lys Asp Ala Phe His Lys Phe Leu Arg Leu Gly Ile Asn Thr Ile Tyr
```

-continued

```
145                 150                 155                 160

Cys Ser Asp Ser His Ile Met Ser Ile Met Cys Ser His Met Thr Glu
                165                 170                 175

Lys Asp Ala Val Leu Ala Ile Ser His Ser Gly Glu Ser Arg Glu Ile
            180                 185                 190

Ile Asp Ala Ile Glu Leu Ala Lys Glu Asn Lys Ala Lys Val Ile Ser
        195                 200                 205

Phe Thr Ser Tyr Pro Asn Ser Thr Leu Ala Lys Leu Ser Asp Val Val
    210                 215                 220

Phe Leu Ser Ala Thr Lys Glu Thr Lys Phe Arg Ser Asp Ala Met Val
225                 230                 235                 240

Ser Arg Ile Val Gln Cys Val Ile Ile Asp Ile Leu Tyr Val Ser Leu
                245                 250                 255

Val Leu Lys Leu Gly Ser Glu Ala Ile Val Asn Val Asn Lys Ser Arg
            260                 265                 270

Leu Ala Val Ala Lys Lys Lys Lys
        275                 280


<210> SEQ ID NO 106
<211> LENGTH: 1098
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 106

atgagtaaaa aaatgaaaat aatggtaatt ggagacgcaa tgataccggg taaagatttt      60

gaatcagcag ctaaaaaata tttatctgat tatgtggaag aaataattac aggagattgg     120

gaaaataatt gggacaattt acaaagaaga agattggaag tagaaaagaa agggcccgag     180

attgaggaag tagttccttt aataaaggaa aaagggcaag atgtttcaat gttgtttggt     240

ttatttgttc ccatttccaa agaaacattc aattacttgc caaaggtaaa gattattggg     300

gtttcgcgag caggcttaga aaatgtaaac gtaaaagaag caacccaacg aggagtttta     360

gtgttcaatg tccagggaag aaatgcagaa gctgtttctg actttgcaat aggtttgctt     420

ttggcagaat gtagaaacat tgcgagagcc cactatgcaa taaagaatgg ccagtggcgg     480

aaagaatttt ctaattctga ttggattccg gaactaaaag gcaaaacagt tggtattatt     540

ggttttggat atattggtag actggtagca aaaaaactct ctggatttga agttagaaga     600

cttgtgtacg atccttatgt aagtgaagag gaaattagag aatgcggatg tataccagta     660

gacaaagaga ctttgttcaa agaaagtgat tttattactc tccatgcacg cctcacagaa     720

gagaataaaa atttggttgg caaatatgag atttcattga tgaaaccaac agcatacatt     780

attaacactg cacgggcagg tctaattgat aaagaagcat taatagaggc tctaaagaca     840

aagagaatag caggagcagc actggatgtg ttctgggaag aacctattcc ttcggacagt     900

gagttgttag aattggacaa tgttactctt acaagtcatt tagcaggaac aaccaaagaa     960

gcacttacaa gatcacctga gcttttaatg gaggatgtca agaagtttat tgaagggcag    1020

aaagcaagat ttattgtgaa tccagaggtt ttggaaaacc aagagttcaa gaaatggctg    1080

gagggtgtga agaaatga                                                  1098


<210> SEQ ID NO 107
<211> LENGTH: 365
<212> TYPE: PRT
```

-continued

<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 107

```
Met Ser Lys Lys Met Lys Ile Met Val Ile Gly Asp Ala Met Ile Pro
1               5                   10                  15

Gly Lys Asp Phe Glu Ser Ala Ala Lys Lys Tyr Leu Ser Asp Tyr Val
            20                  25                  30

Glu Glu Ile Ile Thr Gly Asp Trp Glu Asn Asn Trp Asp Asn Leu Gln
        35                  40                  45

Arg Arg Arg Leu Glu Val Glu Lys Lys Gly Pro Glu Ile Glu Glu Val
    50                  55                  60

Val Pro Leu Ile Lys Glu Lys Gly Gln Asp Val Ser Met Leu Phe Gly
65                  70                  75                  80

Leu Phe Val Pro Ile Ser Lys Glu Thr Phe Asn Tyr Leu Pro Lys Val
                85                  90                  95

Lys Ile Ile Gly Val Ser Arg Ala Gly Leu Glu Asn Val Asn Val Lys
            100                 105                 110

Glu Ala Thr Gln Arg Gly Val Leu Val Phe Asn Val Gln Gly Arg Asn
        115                 120                 125

Ala Glu Ala Val Ser Asp Phe Ala Ile Gly Leu Leu Leu Ala Glu Cys
    130                 135                 140

Arg Asn Ile Ala Arg Ala His Tyr Ala Ile Lys Asn Gly Gln Trp Arg
145                 150                 155                 160

Lys Glu Phe Ser Asn Ser Asp Trp Ile Pro Glu Leu Lys Gly Lys Thr
                165                 170                 175

Val Gly Ile Ile Gly Phe Gly Tyr Ile Gly Arg Leu Val Ala Lys Lys
            180                 185                 190

Leu Ser Gly Phe Glu Val Arg Arg Leu Val Tyr Asp Pro Tyr Val Ser
        195                 200                 205

Glu Glu Glu Ile Arg Glu Cys Gly Cys Ile Pro Val Asp Lys Glu Thr
    210                 215                 220

Leu Phe Lys Glu Ser Asp Phe Ile Thr Leu His Ala Arg Leu Thr Glu
225                 230                 235                 240

Glu Asn Lys Asn Leu Val Gly Lys Tyr Glu Ile Ser Leu Met Lys Pro
                245                 250                 255

Thr Ala Tyr Ile Ile Asn Thr Ala Arg Ala Gly Leu Ile Asp Lys Glu
            260                 265                 270

Ala Leu Ile Glu Ala Leu Lys Thr Lys Arg Ile Ala Gly Ala Ala Leu
        275                 280                 285

Asp Val Phe Trp Glu Glu Pro Ile Pro Ser Asp Ser Glu Leu Leu Glu
    290                 295                 300

Leu Asp Asn Val Thr Leu Thr Ser His Leu Ala Gly Thr Thr Lys Glu
305                 310                 315                 320

Ala Leu Thr Arg Ser Pro Glu Leu Leu Met Glu Asp Val Lys Lys Phe
                325                 330                 335

Ile Glu Gly Gln Lys Ala Arg Phe Ile Val Asn Pro Glu Val Leu Glu
            340                 345                 350

Asn Gln Glu Phe Lys Lys Trp Leu Glu Gly Val Lys Lys
        355                 360                 365
```

<210> SEQ ID NO 108
<211> LENGTH: 3120

-continued

<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 108

atgaacatca actttaaaat aaaaccattc tggttttgga atgggaaaat ggaaaatgat 60

gaaatagcag atcaaatagc tcagatgcat gaaaaaggaa ttggtggatt cttcattcat 120

ccacggcaag gtctcgaaat accctatctt tctcacgagt ggtttgaaaa ggtttctgtt 180

gcaattgaat gtgcaaaaaa atataacatg gaagtgtggc tttatgatga atatccttat 240

ccaagtggaa tttcagctgg tgaagtagtt gttcagcatc ctgaatatca agcttttata 300

ttagattata aagtgtttga ggcaaaagat aatgaagaaa tttgcattga aattcccatg 360

tgtgaagtgt tattagcaag agcttataga ataaggaaca atattatagt atggaatgaa 420

tacatagatt tgattgatta tattggtgta atctacaaag aacacattta ccaagaaagt 480

gggcttacgt tttacaatag gaaaagatac tttgttggag atggggcaaa agcacttaaa 540

tggaaagtac caaaaggaaa gtggaagata tttttgtttt atcagtatcc attaaaaaat 600

tttaaatatt tcgggacatt tattgatcct ctaaataaag atgctgtaag actgtttatt 660

caaaccactc atgaaaaata taaaaaatac ttgggtcatg agtttggaaa aacaattaaa 720

ggaattttta cagatgaaac agctccagtt gctggcaaac ttccttggtc aaaattgctt 780

cctaagcttt ttgagcaaac atatggggaa aatcttattg aaaaattacc tcaaattatt 840

tgcacagaca tatttgatac agccggttca aaaattaggt atcagttttg gaaacttgtt 900

gtggacacat ttattgagag ctacgacaaa caaattcttg aatggtgtca tcagaacaat 960

cttttgtatg tgggcgaaaa gccaattttg agaagctcac aattagcttt tatggacatt 1020

cccggaatag atgcagggca ccaaaaggct ggtgacattc cacaggttgt atctgaaaac 1080

tacagagcaa atccaaagat agcatcatct gccgctcact tttataaaaa ggaaagggtt 1140

ttgtgtgaat gctttcacag tattggctgg agcatgacaa tgcaagatat gaaatggata 1200

tttgactggc taatattgca gggaatagat atgtttgtcc cccatgcctt ttattatagt 1260

gcagatggac tcaaaaaaca cgatgcccca ccttcagcct ttttccaaat gccttggtgg 1320

aaacatcaaa aaatattgtc tgagtatgta gaaaatgtaa ctaaaatgct taaaaattgt 1380

aaaagaaaag ttgatgtact tatagtagat ccaattacaa gccagtggac ctgttttaac 1440

gacaaagaag taaaagagaa gatttcgatg gatttctgta gaattcagca aattctttta 1500

gaagaaaatg tagactatta tgtgattgac cagtcattag taggaagttt ggaatgcagg 1560

aatcagaaga tttattatga caatgaaaaa tttgaattat tgatcattcc acctgtgacc 1620

aatttagaaa aagaggctta catgaagata aaagatctaa tattgaaagg atgtaaagtt 1680

gtatttattg gctgtttgcc attccaaacc atcgaagatt ttgatgttgc caaagatatt 1740

agcaattttc ttggagtaaa ttctatggac atcgcaaaag catataaaac aggttctaaa 1800

ttgaacaata cagtttttct taacagttgt atcttcattg gtaatataga agatttagta 1860

gcaaaaattg acaagatttg taaaaagcct gtaagtatat catatgaatc ttctaatgac 1920

cgtggtattc tctgtgctta ctttgaggac gcagaacacg actttttatt tatgattaac 1980

ccaactaatg aaaagaagat ctgcaaagta tacttgcggt acaggcctga tgaaataaac 2040

aaaatttatt cggttccatt gacatcgcaa gaacttgata aggaaataaa tttcgaaaat 2100

tctttagata aaaagcaaat aactttttcc atgaattttg aacctttca atcgtattta 2160

-continued

```
attaaattag aaaaaagttt tgttaagaaa aacgaccata aaaatgatgt tgaaaagaga   2220

gtttttgagt acaaaatccc tcttgcaact gtttgggaat tttcaattga gagtttaaat   2280

cctttaaggc tcggcagatg gaatttaaag ttgattttca acaacgaaaa tgaacagtat   2340

tcaatatctt caaaaattcc agttacacca aaaccaataa ttgatcagat agaagaggca   2400

aaaattccaa taccactaaa aacaaaaagt ttctttggat gtccaaaaga aataacattg   2460

ccaagctttg aggcaatata cactacttca ttttttatag atgcagcaag ccaaaaattc   2520

tggcttgtaa ttgaagatga aggtataaaa ggtgaatggg ttgtgctttt gaacaatcat   2580

actattttac caagagattt tgtacttaaa agattttatt ctcatactaa tttggcatat   2640

gatattagca acttaattaa actgggggaa aatcagcttt gtgtatgtgt aaagattagc   2700

aggtcttttg atggacttct tacacccatt tatatcttca gcacggcagg tgtattcaaa   2760

gttgatgata gctggcatat agacaaactt ccaactcaag gctgctttgg taaagacctt   2820

gaaaatggca ttccttttta tgctggtttt ataaagtacg aaaaagaagt tcaaatgcca   2880

tcttttgaaa atggttttgt ggaattcttt attgaagata acataaatca gtgtgtaagt   2940

ctttatataa atgatgaatt tataggtaca aggtgctggc agccttatag atggaaggta   3000

gattctgatt tactttcttc aaagaaggta aagctcacac ttgaagtatc aacttcgagc   3060

ctgcagctgt tcgaaggtga agttattgaa ccaataacac ataaaattaa gacaatataa   3120
```

```
<210> SEQ ID NO 109
<211> LENGTH: 1039
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 109

Met Asn Ile Asn Phe Lys Ile Lys Pro Phe Trp Phe Trp Asn Gly Lys
1               5                   10                  15

Met Glu Asn Asp Glu Ile Ala Asp Gln Ile Ala Gln Met His Glu Lys
            20                  25                  30

Gly Ile Gly Gly Phe Phe Ile His Pro Arg Gln Gly Leu Glu Ile Pro
        35                  40                  45

Tyr Leu Ser His Glu Trp Phe Glu Lys Val Ser Val Ala Ile Glu Cys
    50                  55                  60

Ala Lys Lys Tyr Asn Met Glu Val Trp Leu Tyr Asp Glu Tyr Pro Tyr
65                  70                  75                  80

Pro Ser Gly Ile Ser Ala Gly Glu Val Val Val Gln His Pro Glu Tyr
                85                  90                  95

Gln Ala Phe Ile Leu Asp Tyr Lys Val Phe Glu Ala Lys Asp Asn Glu
            100                 105                 110

Glu Ile Cys Ile Glu Ile Pro Met Cys Glu Val Leu Leu Ala Arg Ala
        115                 120                 125

Tyr Arg Ile Arg Asn Asn Ile Ile Val Trp Asn Glu Tyr Ile Asp Leu
    130                 135                 140

Ile Asp Tyr Ile Gly Val Ile Tyr Lys Glu His Ile Tyr Gln Glu Ser
145                 150                 155                 160

Gly Leu Thr Phe Tyr Asn Arg Lys Arg Tyr Phe Val Gly Asp Gly Ala
                165                 170                 175

Lys Ala Leu Lys Trp Lys Val Pro Lys Gly Lys Trp Lys Ile Phe Leu
            180                 185                 190
```

-continued

```
Phe Tyr Gln Tyr Pro Leu Lys Asn Phe Lys Tyr Phe Gly Thr Phe Ile
        195                 200                 205

Asp Pro Leu Asn Lys Asp Ala Val Arg Leu Phe Ile Gln Thr Thr His
    210                 215                 220

Glu Lys Tyr Lys Lys Tyr Leu Gly His Glu Phe Gly Lys Thr Ile Lys
225                 230                 235                 240

Gly Ile Phe Thr Asp Glu Thr Ala Pro Val Ala Gly Lys Leu Pro Trp
                245                 250                 255

Ser Lys Leu Leu Pro Lys Leu Phe Glu Gln Thr Tyr Gly Glu Asn Leu
            260                 265                 270

Ile Glu Lys Leu Pro Gln Ile Ile Cys Thr Asp Ile Phe Asp Thr Ala
        275                 280                 285

Gly Ser Lys Ile Arg Tyr Gln Phe Trp Lys Leu Val Val Asp Thr Phe
    290                 295                 300

Ile Glu Ser Tyr Asp Lys Gln Ile Leu Glu Trp Cys His Gln Asn Asn
305                 310                 315                 320

Leu Leu Tyr Val Gly Glu Lys Pro Ile Leu Arg Ser Ser Gln Leu Ala
                325                 330                 335

Phe Met Asp Ile Pro Gly Ile Asp Ala Gly His Gln Lys Ala Gly Asp
            340                 345                 350

Ile Pro Gln Val Val Ser Glu Asn Tyr Arg Ala Asn Pro Lys Ile Ala
        355                 360                 365

Ser Ser Ala Ala His Phe Tyr Lys Lys Glu Arg Val Leu Cys Glu Cys
    370                 375                 380

Phe His Ser Ile Gly Trp Ser Met Thr Met Gln Asp Met Lys Trp Ile
385                 390                 395                 400

Phe Asp Trp Leu Ile Leu Gln Gly Ile Asp Met Phe Val Pro His Ala
                405                 410                 415

Phe Tyr Tyr Ser Ala Asp Gly Leu Lys Lys His Asp Ala Pro Pro Ser
            420                 425                 430

Ala Phe Phe Gln Met Pro Trp Trp Lys His Gln Lys Ile Leu Ser Glu
        435                 440                 445

Tyr Val Glu Asn Val Thr Lys Met Leu Lys Asn Cys Lys Arg Lys Val
    450                 455                 460

Asp Val Leu Ile Val Asp Pro Ile Thr Ser Gln Trp Thr Cys Phe Asn
465                 470                 475                 480

Asp Lys Glu Val Lys Glu Lys Ile Ser Met Asp Phe Cys Arg Ile Gln
                485                 490                 495

Gln Ile Leu Leu Glu Glu Asn Val Asp Tyr Tyr Val Ile Asp Gln Ser
            500                 505                 510

Leu Val Gly Ser Leu Glu Cys Arg Asn Gln Lys Ile Tyr Tyr Asp Asn
        515                 520                 525

Glu Lys Phe Glu Leu Leu Ile Ile Pro Pro Val Thr Asn Leu Glu Lys
    530                 535                 540

Glu Ala Tyr Met Lys Ile Lys Asp Leu Ile Leu Lys Gly Cys Lys Val
545                 550                 555                 560

Val Phe Ile Gly Cys Leu Pro Phe Gln Thr Ile Glu Asp Phe Asp Val
                565                 570                 575

Ala Lys Asp Ile Ser Asn Phe Leu Gly Val Asn Ser Met Asp Ile Ala
            580                 585                 590
```

-continued

```
Lys Ala Tyr Lys Thr Gly Ser Lys Leu Asn Asn Thr Val Phe Leu Asn
        595                 600                 605

Ser Cys Ile Phe Ile Gly Asn Ile Glu Asp Leu Val Ala Lys Ile Asp
    610                 615                 620

Lys Ile Cys Lys Lys Pro Val Ser Ile Ser Tyr Glu Ser Ser Asn Asp
625                 630                 635                 640

Arg Gly Ile Leu Cys Ala Tyr Phe Glu Asp Ala Glu His Asp Phe Leu
                645                 650                 655

Phe Met Ile Asn Pro Thr Asn Glu Lys Lys Ile Cys Lys Val Tyr Leu
            660                 665                 670

Arg Tyr Arg Pro Asp Glu Ile Asn Lys Ile Tyr Ser Val Pro Leu Thr
        675                 680                 685

Ser Gln Glu Leu Asp Lys Glu Ile Asn Phe Glu Asn Ser Leu Asp Lys
    690                 695                 700

Lys Gln Ile Thr Phe Ser Met Asn Phe Glu Pro Phe Gln Ser Tyr Leu
705                 710                 715                 720

Ile Lys Leu Glu Lys Ser Phe Val Lys Lys Asn Asp His Lys Asn Asp
                725                 730                 735

Val Glu Lys Arg Val Phe Glu Tyr Lys Ile Pro Leu Ala Thr Val Trp
            740                 745                 750

Glu Phe Ser Ile Glu Ser Leu Asn Pro Leu Arg Leu Gly Arg Trp Asn
        755                 760                 765

Leu Lys Leu Ile Phe Asn Asn Glu Asn Glu Gln Tyr Ser Ile Ser Ser
    770                 775                 780

Lys Ile Pro Val Thr Pro Lys Pro Ile Ile Asp Gln Ile Glu Glu Ala
785                 790                 795                 800

Lys Ile Pro Ile Pro Leu Lys Thr Lys Ser Phe Phe Gly Cys Pro Lys
                805                 810                 815

Glu Ile Thr Leu Pro Ser Phe Glu Ala Ile Tyr Thr Thr Ser Phe Phe
            820                 825                 830

Ile Asp Ala Ala Ser Gln Lys Phe Trp Leu Val Ile Glu Asp Glu Gly
        835                 840                 845

Ile Lys Gly Glu Trp Val Val Leu Leu Asn Asn His Thr Ile Leu Pro
    850                 855                 860

Arg Asp Phe Val Leu Lys Arg Phe Tyr Ser His Thr Asn Leu Ala Tyr
865                 870                 875                 880

Asp Ile Ser Asn Leu Ile Lys Leu Gly Glu Asn Gln Leu Cys Val Cys
                885                 890                 895

Val Lys Ile Ser Arg Ser Phe Asp Gly Leu Leu Thr Pro Ile Tyr Ile
            900                 905                 910

Phe Ser Thr Ala Gly Val Phe Lys Val Asp Asp Ser Trp His Ile Asp
        915                 920                 925

Lys Leu Pro Thr Gln Gly Cys Phe Gly Lys Asp Leu Glu Asn Gly Ile
    930                 935                 940

Pro Phe Tyr Ala Gly Phe Ile Lys Tyr Glu Lys Glu Val Gln Met Pro
945                 950                 955                 960

Ser Phe Glu Asn Gly Phe Val Glu Phe Phe Ile Glu Asp Asn Ile Asn
                965                 970                 975

Gln Cys Val Ser Leu Tyr Ile Asn Asp Glu Phe Ile Gly Thr Arg Cys
            980                 985                 990

Trp Gln Pro Tyr Arg Trp Lys Val  Asp Ser Asp Leu Leu  Ser Ser Lys
```

-continued

```
        995                 1000                1005

Lys Val  Lys Leu Thr Leu Glu  Val Ser Thr Ser Ser  Leu Gln Leu
    1010                1015                1020

Phe Glu  Gly Glu Val Ile Glu  Pro Ile Thr His Lys  Ile Lys Thr
    1025                1030                1035

Ile


<210> SEQ ID NO 110
<211> LENGTH: 726
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 110

atgatgcgaa ttgtagaaaa aattagacaa aaaaatgaaa attatcactc cggtccaata     60

cctacaattg ccttttggg tgacagtata acccacggtt gttttgaggt cattgaagga    120

acaaaaagaa cattagaagt tgtgtgcgat tttgaagctg tctatcataa tcaatttaaa    180

aaaatcttat caatgatgtt tccatttgct caaattaata ttgtcaatgc aggcataagt    240

ggtgatacag cacagggcgg cttacaaaga cttgaaagag atgttttaag atttaaccca    300

gaccttgtgg tagtttgtta tggattaaat gatagtaaca aaggaaaaga gtatttgaat    360

gaatatttag atggtttagc agggatattt atagaactta aaaaacacga tattgaagta    420

atttttttaa ctcccaatat gaaaaatact tatatttcac cagctataaa gagcttgcct    480

ttgattgaga tggcaaaatt aaatatggaa agtcaaatca atggaacgtt agatctttac    540

atggattcag ctaaggagct ttgcaaaaag gaaaaggttg tagtttgtga ttgttatgaa    600

aaatggaaaa agctttatca cagtggcgtt gacacaacaa atcttctttc aaattttata    660

aatcacccca atcgcccgat gcataagctc tttgcatggt cattatttga gacaattatg    720

ttttaa                                                               726


<210> SEQ ID NO 111
<211> LENGTH: 241
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 111

Met Met Arg Ile Val Glu Lys Ile Arg Gln Lys Asn Glu Asn Tyr His
1               5                   10                  15

Ser Gly Pro Ile Pro Thr Ile Ala Phe Leu Gly Asp Ser Ile Thr His
            20                  25                  30

Gly Cys Phe Glu Val Ile Glu Gly Thr Lys Arg Thr Leu Glu Val Val
        35                  40                  45

Cys Asp Phe Glu Ala Val Tyr His Asn Gln Phe Lys Lys Ile Leu Ser
    50                  55                  60

Met Met Phe Pro Phe Ala Gln Ile Asn Ile Val Asn Ala Gly Ile Ser
65                  70                  75                  80

Gly Asp Thr Ala Gln Gly Gly Leu Gln Arg Leu Glu Arg Asp Val Leu
                85                  90                  95

Arg Phe Asn Pro Asp Leu Val Val Val Cys Tyr Gly Leu Asn Asp Ser
            100                 105                 110

Asn Lys Gly Lys Glu Tyr Leu Asn Glu Tyr Leu Asp Gly Leu Ala Gly
        115                 120                 125
```

-continued

```
Ile Phe Ile Glu Leu Lys Lys His Asp Ile Glu Val Ile Phe Leu Thr
    130                 135                 140

Pro Asn Met Lys Asn Thr Tyr Ile Ser Pro Ala Ile Lys Ser Leu Pro
145                 150                 155                 160

Leu Ile Glu Met Ala Lys Leu Asn Met Glu Ser Gln Ile Asn Gly Thr
                165                 170                 175

Leu Asp Leu Tyr Met Asp Ser Ala Lys Glu Leu Cys Lys Lys Glu Lys
            180                 185                 190

Val Val Val Cys Asp Cys Tyr Glu Lys Trp Lys Lys Leu Tyr His Ser
        195                 200                 205

Gly Val Asp Thr Thr Asn Leu Leu Ser Asn Phe Ile Asn His Pro Asn
    210                 215                 220

Arg Pro Met His Lys Leu Phe Ala Trp Ser Leu Phe Glu Thr Ile Met
225                 230                 235                 240

Phe


<210> SEQ ID NO 112
<211> LENGTH: 990
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 112

atgaaaattt gcatagtagg cagtagtgga cactatgtat atgctttaag aggaataaaa     60

gaagaccctc atgcccaaat tgtgggaatc tctcctggat gtgaaggaga gaatattgaa    120

aggttacatt ctcaagtaaa tgaaatggga ttcacacctg tggtttatag caatcctata    180

aggatgtttg aagatctcaa acctgacatt gctgtgatta atacattttt ttataaaaat    240

tctgagcttg caattgaggc tatgaaaaga ggaatccacg tatatatgga aaagcctgtt    300

gcactatcaa tagaaaaact tgaagaacta aagagtgtgt ggaggcaaac aaaagtaaaa    360

ctctcatcaa tgctgggatt gcgctataca ccccattttt ggactgctta taaacttata    420

aatgaaaaca agataggtag aataagactg atacatgccc aaaaatctta taaacttgga    480

actcgacctg acttttataa acatagaaga acatatggcg gaacaattcc ctgggttggc    540

attcatgcta ttgattggat ttattggcta agtggcaaga aatttaaatc ggtctttgca    600

ggacattcaa aactttataa taatgatcat ggtgagcttg aatctactgc tttttgtagt    660

tttgtaatgg aagatgagat ttttgcaacg gtgaacattg actatctgcg tcctgctact    720

gcccctactc atgatgatga tagaattaga attgtgggaa caagaggaat ttttgaagtt    780

ttaaatggaa aagttttctt gctaaatgat accactaaag agatctcaga agtctcttta    840

gaaaaaccac ctattgtgtt tttagatttc ttaaatgagg taagaggtac agataagtgc    900

ttagttagta gcgaggatag cttttatgta acctttgctt cgcttttagc aaggcagtct    960

gctgatgagg ataaggtaat tgaattttaa                                     990


<210> SEQ ID NO 113
<211> LENGTH: 329
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 113

Met Lys Ile Cys Ile Val Gly Ser Ser Gly His Tyr Val Tyr Ala Leu
1               5                   10                  15
```

-continued

```
Arg Gly Ile Lys Glu Asp Pro His Ala Gln Ile Val Gly Ile Ser Pro
            20                  25                  30

Gly Cys Glu Gly Glu Asn Ile Glu Arg Leu His Ser Gln Val Asn Glu
        35                  40                  45

Met Gly Phe Thr Pro Val Val Tyr Ser Asn Pro Ile Arg Met Phe Glu
    50                  55                  60

Asp Leu Lys Pro Asp Ile Ala Val Ile Asn Thr Phe Phe Tyr Lys Asn
65                  70                  75                  80

Ser Glu Leu Ala Ile Glu Ala Met Lys Arg Gly Ile His Val Tyr Met
                85                  90                  95

Glu Lys Pro Val Ala Leu Ser Ile Glu Lys Leu Glu Glu Leu Lys Ser
            100                 105                 110

Val Trp Arg Gln Thr Lys Val Lys Leu Ser Ser Met Leu Gly Leu Arg
        115                 120                 125

Tyr Thr Pro His Phe Trp Thr Ala Tyr Lys Leu Ile Asn Glu Asn Lys
    130                 135                 140

Ile Gly Arg Ile Arg Leu Ile His Ala Gln Lys Ser Tyr Lys Leu Gly
145                 150                 155                 160

Thr Arg Pro Asp Phe Tyr Lys His Arg Arg Thr Tyr Gly Gly Thr Ile
                165                 170                 175

Pro Trp Val Gly Ile His Ala Ile Asp Trp Ile Tyr Trp Leu Ser Gly
            180                 185                 190

Lys Lys Phe Lys Ser Val Phe Ala Gly His Ser Lys Leu Tyr Asn Asn
        195                 200                 205

Asp His Gly Glu Leu Glu Ser Thr Ala Phe Cys Ser Phe Val Met Glu
    210                 215                 220

Asp Glu Ile Phe Ala Thr Val Asn Ile Asp Tyr Leu Arg Pro Ala Thr
225                 230                 235                 240

Ala Pro Thr His Asp Asp Asp Arg Ile Arg Ile Val Gly Thr Arg Gly
                245                 250                 255

Ile Phe Glu Val Leu Asn Gly Lys Val Phe Leu Leu Asn Asp Thr Thr
            260                 265                 270

Lys Glu Ile Ser Glu Val Ser Leu Glu Lys Pro Pro Ile Val Phe Leu
        275                 280                 285

Asp Phe Leu Asn Glu Val Arg Gly Thr Asp Lys Cys Leu Val Ser Ser
    290                 295                 300

Glu Asp Ser Phe Tyr Val Thr Phe Ala Ser Leu Leu Ala Arg Gln Ser
305                 310                 315                 320

Ala Asp Glu Asp Lys Val Ile Glu Phe
                325


<210> SEQ ID NO 114
<211> LENGTH: 843
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 114

gtgagaagaa caaagacaat gaaagcatta ttgtttcatc tgctaacaat tctatttggt      60

tacataatgc tgtatccact tctttggatg tttttcagtt ccttcaaaga aaatagtgaa     120

atttttctaa atgcgcacga attgctaccc aaaaagtggc ttttcaaaaa ttatattgat     180

ggatggagag gttttgcggg atacccattt tcagtttttt ttaaaaattc atttattgtt     240
```

-continued

```
actattattg gtactgttgg tgctgtcatt tcctcagcta ttgtagctta tggttttgca    300

agatgtaagt tcaaaggcaa aggattttgg tttggttgta tgattataac tatgctttta    360

ccatatcagg ttgttatgat tccgcaatac atcatgtttc aaaagatggg atgggtaaat    420

acgttcaagc ctttacttgt tccagctttt ttgggtcagc catttttat ctttttaatg    480

attcaattta taagaggaat tccaaacgaa ttagatgaag cggctaaaat tgatggttgt    540

agtaaatatt caatttttac aagaattatt ttgccattaa tttctccggc tttaattaca    600

tcagcaatct tttcattttt gtggcgttgg gatgactttt tagggcctct tttgtacctc    660

agcaaacctg agttgtatac agtatcttta gggttgagaa tgttctctga cccgacagca    720

gtttctaact ggggagctgc ctttgctatg gcaacattat cgctcgttcc ctcatttata    780

atatttatat tcttccaacg ctatctagtg gaaggaattg ttactacagg tttaaagggt    840

taa                                                                  843
```

```
<210> SEQ ID NO 115
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 115

Val Arg Arg Thr Lys Thr Met Lys Ala Leu Leu Phe His Leu Leu Thr
1               5                   10                  15

Ile Leu Phe Gly Tyr Ile Met Leu Tyr Pro Leu Leu Trp Met Phe Phe
            20                  25                  30

Ser Ser Phe Lys Glu Asn Ser Glu Ile Phe Leu Asn Ala His Glu Leu
        35                  40                  45

Leu Pro Lys Lys Trp Leu Phe Lys Asn Tyr Ile Asp Gly Trp Arg Gly
    50                  55                  60

Phe Ala Gly Tyr Pro Phe Ser Val Phe Phe Lys Asn Ser Phe Ile Val
65                  70                  75                  80

Thr Ile Ile Gly Thr Val Gly Ala Val Ile Ser Ser Ala Ile Val Ala
                85                  90                  95

Tyr Gly Phe Ala Arg Cys Lys Phe Lys Gly Lys Gly Phe Trp Phe Gly
            100                 105                 110

Cys Met Ile Ile Thr Met Leu Leu Pro Tyr Gln Val Val Met Ile Pro
        115                 120                 125

Gln Tyr Ile Met Phe Gln Lys Met Gly Trp Val Asn Thr Phe Lys Pro
    130                 135                 140

Leu Leu Val Pro Ala Phe Leu Gly Gln Pro Phe Phe Ile Phe Leu Met
145                 150                 155                 160

Ile Gln Phe Ile Arg Gly Ile Pro Asn Glu Leu Asp Glu Ala Ala Lys
                165                 170                 175

Ile Asp Gly Cys Ser Lys Tyr Ser Ile Phe Thr Arg Ile Ile Leu Pro
            180                 185                 190

Leu Ile Ser Pro Ala Leu Ile Thr Ser Ala Ile Phe Ser Phe Leu Trp
        195                 200                 205

Arg Trp Asp Asp Phe Leu Gly Pro Leu Leu Tyr Leu Ser Lys Pro Glu
    210                 215                 220

Leu Tyr Thr Val Ser Leu Gly Leu Arg Met Phe Ser Asp Pro Thr Ala
225                 230                 235                 240

Val Ser Asn Trp Gly Ala Ala Phe Ala Met Ala Thr Leu Ser Leu Val
```

-continued

```
                245                 250                 255

Pro Ser Phe Ile Ile Phe Ile Phe Phe Gln Arg Tyr Leu Val Glu Gly
            260                 265                 270

Ile Val Thr Thr Gly Leu Lys Gly
        275                 280


<210> SEQ ID NO 116
<211> LENGTH: 927
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 116

atgtcgaata ttaatcgaaa aaagtctttc agagaacttt tgatgagttc agaaaatgtt      60

gctggttatg tttttatatc tccctggctt attggctttt ttgtgttcac tttgattcct     120

attgcagcaa ccttttattt gtcttttact caatatgatt tattatcatc tcctaaattt     180

gtaggattac aaaattatgt acaaatgttt aaagaagatc cgttattttg gaaatcaatg     240

tcagtaactt ttttctatgt gtttgtaact gtgccattaa agctggcttt tgcattgctt     300

cttgcccttt ggctttctta caaaagcaga ctaacaccat tttacagggc tgtatactat     360

gttccttcta tgatgggtgg cagtgtggct gtggcagtgc tttggcaaag actttttaca     420

agtgatggtg ttataaattc aatattgaaa ctatttggaa ttcaaagtga gacttcatgg     480

ataggaaatc caagaactgc tatatggacg ttgatattac ttgcggtttg gcaatttggt     540

tcgccgatgt tgatattttt agcaggttta aagcaaatac cagaaagcta ttatgaagca     600

gctattattg atggagcaaa tagctggcaa aagtttgtta aaataacctt gccgatgctc     660

acaccaataa tatttttcaa cttgattatg cagatgatag gaagctttat gacttttact     720

caaggattca ttattacaaa tggcggccct gtgaacagca cactctttta cgctatttac     780

ctctacagaa gagcattcca attttatgac atgggctaca gctgtgctat gtcgtgggta     840

atgcttatta tcattggaat actcacagct tttatattca aatcatctac attttgggta     900

tattatgagt ccaaggaagg tgaataa                                         927


<210> SEQ ID NO 117
<211> LENGTH: 308
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 117

Met Ser Asn Ile Asn Arg Lys Lys Ser Phe Arg Glu Leu Leu Met Ser
1               5                   10                  15

Ser Glu Asn Val Ala Gly Tyr Val Phe Ile Ser Pro Trp Leu Ile Gly
            20                  25                  30

Phe Phe Val Phe Thr Leu Ile Pro Ile Ala Ala Thr Phe Tyr Leu Ser
        35                  40                  45

Phe Thr Gln Tyr Asp Leu Leu Ser Ser Pro Lys Phe Val Gly Leu Gln
    50                  55                  60

Asn Tyr Val Gln Met Phe Lys Glu Asp Pro Leu Phe Trp Lys Ser Met
65                  70                  75                  80

Ser Val Thr Phe Phe Tyr Val Phe Val Thr Val Pro Leu Lys Leu Ala
                85                  90                  95

Phe Ala Leu Leu Leu Ala Leu Trp Leu Ser Tyr Lys Ser Arg Leu Thr
            100                 105                 110
```

-continued

```
Pro Phe Tyr Arg Ala Val Tyr Tyr Val Pro Ser Met Met Gly Gly Ser
        115                 120                 125

Val Ala Val Ala Val Leu Trp Gln Arg Leu Phe Thr Ser Asp Gly Val
    130                 135                 140

Ile Asn Ser Ile Leu Lys Leu Phe Gly Ile Gln Ser Glu Thr Ser Trp
145                 150                 155                 160

Ile Gly Asn Pro Arg Thr Ala Ile Trp Thr Leu Ile Leu Leu Ala Val
                165                 170                 175

Trp Gln Phe Gly Ser Pro Met Leu Ile Phe Leu Ala Gly Leu Lys Gln
            180                 185                 190

Ile Pro Glu Ser Tyr Tyr Glu Ala Ala Ile Ile Asp Gly Ala Asn Ser
        195                 200                 205

Trp Gln Lys Phe Val Lys Ile Thr Leu Pro Met Leu Thr Pro Ile Ile
    210                 215                 220

Phe Phe Asn Leu Ile Met Gln Met Ile Gly Ser Phe Met Thr Phe Thr
225                 230                 235                 240

Gln Gly Phe Ile Ile Thr Asn Gly Gly Pro Val Asn Ser Thr Leu Phe
                245                 250                 255

Tyr Ala Ile Tyr Leu Tyr Arg Arg Ala Phe Gln Phe Tyr Asp Met Gly
            260                 265                 270

Tyr Ser Cys Ala Met Ser Trp Val Met Leu Ile Ile Ile Gly Ile Leu
        275                 280                 285

Thr Ala Phe Ile Phe Lys Ser Ser Thr Phe Trp Val Tyr Tyr Glu Ser
    290                 295                 300

Lys Glu Gly Glu
305


<210> SEQ ID NO 118
<211> LENGTH: 1305
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 118

atgaaaagaa tattgtgcat tggaactatt ataatatttt tattttcaat ccttatattt     60

ccttttacta aaccttctga aaaggctttt ggttcttctc agtcaatcaa actcagagtt    120

gcgtggtggg gaagccaaac gcgtcatgac agaactcaaa aggtgcttga actataccga    180

gcaaaggtta atcgaaaagt aagttttgta acagaatttg gaagctggtc gggatattgg    240

gataaactta caactcaagc agctgcaaag aatttgcctg acattattca aatggactat    300

atgtatttag ctcaatatgt tcaaaaagga ttgttggcag atttaacacc ttacacaaaa    360

aatggaatat tgaatctgaa agatgtaagt gaagcaagta taaaaagtgg atcagttggt    420

ggaaggattt atgccataag tttagggact aacgctttag caataattta tgaccctgct    480

gtggctcaaa aagcaggtgt taagatacct gaagatggca actggacctg gaatgactac    540

aaagagatta tcaaaaaagt ttttcaaaaa accaaaatca gagcagactt ggcattaaca    600

gctgacccaa aattcttact tgaatactat gtaaggcagc aaggcaaaag cttatataaa    660

cacgatggaa cgggtttggg atttactcaa gaaaaatttg ttattgatgt atttaatatc    720

aatttagagt tattgaaagg cggatataca gcaaaacctg atgaagtttc agcaacctcc    780

acaattgaag acagcttatt tgtgaaaggc aagacatgga taggatggac atggagcaac    840
```

-continued

```
atgtttgttg caacagcaaa tgctgcaaaa cgcccactgg cattagctct tccacccaag    900

gggggtataa gaccagggtt gtatttaaaa ccttctcagt tcttctctat tgctgcaacc    960

tccaaatata aaactgaagc agctaaagtt ataaacttct ttacaaatag catagaggcg   1020

aataatattt tgcttgctga aagaggtgtg cctatttcag caaaagtaag agaaggtatc   1080

aagaatgcag tcgaacctgc agtgagacag acatttgatt atatagccct cgctgaaaag   1140

aattgttcac ctattgatcc accagatcca ccaggcggaa ctgaagtagg gcagacgttt   1200

aaagatcttt acgaccaagt cttgtatggt cagataaaac ctgaaacagc ggcaaagatg   1260

tttatgcaaa aagcaaatca aatacttgcc aaaaacaaga aataa                   1305
```

<210> SEQ ID NO 119
<211> LENGTH: 434
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 119

```
Met Lys Arg Ile Leu Cys Ile Gly Thr Ile Ile Ile Phe Leu Phe Ser
1               5                   10                  15

Ile Leu Ile Phe Pro Phe Thr Lys Pro Ser Glu Lys Ala Phe Gly Ser
            20                  25                  30

Ser Gln Ser Ile Lys Leu Arg Val Ala Trp Trp Gly Ser Gln Thr Arg
        35                  40                  45

His Asp Arg Thr Gln Lys Val Leu Glu Leu Tyr Arg Ala Lys Val Asn
    50                  55                  60

Arg Lys Val Ser Phe Val Thr Glu Phe Gly Ser Trp Ser Gly Tyr Trp
65                  70                  75                  80

Asp Lys Leu Thr Thr Gln Ala Ala Ala Lys Asn Leu Pro Asp Ile Ile
                85                  90                  95

Gln Met Asp Tyr Met Tyr Leu Ala Gln Tyr Val Gln Lys Gly Leu Leu
            100                 105                 110

Ala Asp Leu Thr Pro Tyr Thr Lys Asn Gly Ile Leu Asn Leu Lys Asp
        115                 120                 125

Val Ser Glu Ala Ser Ile Lys Ser Gly Ser Val Gly Gly Arg Ile Tyr
    130                 135                 140

Ala Ile Ser Leu Gly Thr Asn Ala Leu Ala Ile Ile Tyr Asp Pro Ala
145                 150                 155                 160

Val Ala Gln Lys Ala Gly Val Lys Ile Pro Glu Asp Gly Asn Trp Thr
                165                 170                 175

Trp Asn Asp Tyr Lys Glu Ile Ile Lys Lys Val Phe Gln Lys Thr Lys
            180                 185                 190

Ile Arg Ala Asp Leu Ala Leu Thr Ala Asp Pro Lys Phe Leu Leu Glu
        195                 200                 205

Tyr Tyr Val Arg Gln Gln Gly Lys Ser Leu Tyr Lys His Asp Gly Thr
    210                 215                 220

Gly Leu Gly Phe Thr Gln Glu Lys Phe Val Ile Asp Val Phe Asn Ile
225                 230                 235                 240

Asn Leu Glu Leu Leu Lys Gly Gly Tyr Thr Ala Lys Pro Asp Glu Val
                245                 250                 255

Ser Ala Thr Ser Thr Ile Glu Asp Ser Leu Phe Val Lys Gly Lys Thr
            260                 265                 270

Trp Ile Gly Trp Thr Trp Ser Asn Met Phe Val Ala Thr Ala Asn Ala
```

-continued

```
        275                 280                 285

Ala Lys Arg Pro Leu Ala Leu Ala Leu Pro Pro Lys Gly Gly Ile Arg
    290                 295                 300

Pro Gly Leu Tyr Leu Lys Pro Ser Gln Phe Phe Ser Ile Ala Ala Thr
305                 310                 315                 320

Ser Lys Tyr Lys Thr Glu Ala Ala Lys Val Ile Asn Phe Phe Thr Asn
                325                 330                 335

Ser Ile Glu Ala Asn Asn Ile Leu Leu Ala Glu Arg Gly Val Pro Ile
            340                 345                 350

Ser Ala Lys Val Arg Glu Gly Ile Lys Asn Ala Val Glu Pro Ala Val
        355                 360                 365

Arg Gln Thr Phe Asp Tyr Ile Ala Leu Ala Glu Lys Asn Cys Ser Pro
    370                 375                 380

Ile Asp Pro Pro Asp Pro Pro Gly Gly Thr Glu Val Gly Gln Thr Phe
385                 390                 395                 400

Lys Asp Leu Tyr Asp Gln Val Leu Tyr Gly Gln Ile Lys Pro Glu Thr
                405                 410                 415

Ala Ala Lys Met Phe Met Gln Lys Ala Asn Gln Ile Leu Ala Lys Asn
            420                 425                 430

Lys Lys


<210> SEQ ID NO 120
<211> LENGTH: 1212
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 120

ttgaatattg caaagattgg tcttattgga attagtggtt ttgggagtat acacttgcgg     60

tcaatagaac agcttcaagg gaagatggtt gacttgagag caattgttgc aacaagctac    120

gaaaaaaata aagaagtgat tgatagattg gcttctcgag gtgttgagta ttatcaggat    180

tatagattaa tgcttgaaaa tcataaagac ttagattttg ttgccatctc aacgcccatt    240

catttacatg ctccaatggc aattgatgca atggaaagag gttttaatgt cctgcttgaa    300

aagccgcctg ctgtgacaat tcaggatatt gacgctataa ttgagacgaa gagaaaaaca    360

aaaagggttt gtagcgtgaa ctttcaaaac acctctggta aagcatttag aaaacttctt    420

gagtatataa gagaaggcag gcttggcagg ataaaatcaa ttattggtgt tggacgttgg    480

aaaagggatg aaagctatta tcaaagaaat gcgtgggctg ggaagctaat tgttgatggt    540

aactatgtct tggatggaac aataaacaat cctcttgcac atcttttgaa caatgaattg    600

attattgctg agacctcaga agaaaatgga ggtgtacctc agaaagttac tgcagaactt    660

tatcacggtc ataaaattga aggtgaagac acagcgtgtg tgagaatcat cacaaaaaca    720

ggaatcgagg tctattttta ttcaacgtta tgtaacaggg aagaagagtc accctatatc    780

ataatagaag ctgaaaaagc aagagcttat tggacatttg caaataagtt taagatagaa    840

tattttgatg gcaatacaga agagtttgat ggtggacgcg aagatttgtt tgtgaatatg    900

tacattaata tggtggagca tctgtttgaa ggcaaacaac tttactgtcc tctggaagtg    960

actcgcaatt ttgttttggc atccaacggg gctttcgaat cgtcagggtg tatttatgat   1020

attccagacg aatatttaga gattagcaat gaaaatggta agatatatac ctatatcaaa   1080

aatataaaag aaatcataga tgaggcagct gagaatagaa aactattttc tgaaatagga   1140
```

-continued

```
gtaccgtggg cgaaacaaac agaagagttt gatttaattg attactgttg ttttagtatg   1200

tttaaaaggt aa                                                       1212


<210> SEQ ID NO 121
<211> LENGTH: 403
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 121

Leu Asn Ile Ala Lys Ile Gly Leu Ile Gly Ile Ser Gly Phe Gly Ser
1               5                   10                  15

Ile His Leu Arg Ser Ile Glu Gln Leu Gln Gly Lys Met Val Asp Leu
            20                  25                  30

Arg Ala Ile Val Ala Thr Ser Tyr Glu Lys Asn Lys Glu Val Ile Asp
        35                  40                  45

Arg Leu Ala Ser Arg Gly Val Glu Tyr Tyr Gln Asp Tyr Arg Leu Met
    50                  55                  60

Leu Glu Asn His Lys Asp Leu Asp Phe Val Ala Ile Ser Thr Pro Ile
65                  70                  75                  80

His Leu His Ala Pro Met Ala Ile Asp Ala Met Glu Arg Gly Phe Asn
                85                  90                  95

Val Leu Leu Glu Lys Pro Pro Ala Val Thr Ile Gln Asp Ile Asp Ala
            100                 105                 110

Ile Ile Glu Thr Lys Arg Lys Thr Lys Arg Val Cys Ser Val Asn Phe
        115                 120                 125

Gln Asn Thr Ser Gly Lys Ala Phe Arg Lys Leu Leu Glu Tyr Ile Arg
    130                 135                 140

Glu Gly Arg Leu Gly Arg Ile Lys Ser Ile Ile Gly Val Gly Arg Trp
145                 150                 155                 160

Lys Arg Asp Glu Ser Tyr Tyr Gln Arg Asn Ala Trp Ala Gly Lys Leu
                165                 170                 175

Ile Val Asp Gly Asn Tyr Val Leu Asp Gly Thr Ile Asn Asn Pro Leu
            180                 185                 190

Ala His Leu Leu Asn Asn Glu Leu Ile Ile Ala Glu Thr Ser Glu Glu
        195                 200                 205

Asn Gly Gly Val Pro Gln Lys Val Thr Ala Glu Leu Tyr His Gly His
    210                 215                 220

Lys Ile Glu Gly Glu Asp Thr Ala Cys Val Arg Ile Ile Thr Lys Thr
225                 230                 235                 240

Gly Ile Glu Val Tyr Phe Tyr Ser Thr Leu Cys Asn Arg Glu Glu Glu
                245                 250                 255

Ser Pro Tyr Ile Ile Ile Glu Ala Glu Lys Ala Arg Ala Tyr Trp Thr
            260                 265                 270

Phe Ala Asn Lys Phe Lys Ile Glu Tyr Phe Asp Gly Asn Thr Glu Glu
        275                 280                 285

Phe Asp Gly Gly Arg Glu Asp Leu Phe Val Asn Met Tyr Ile Asn Met
    290                 295                 300

Val Glu His Leu Phe Glu Gly Lys Gln Leu Tyr Cys Pro Leu Glu Val
305                 310                 315                 320

Thr Arg Asn Phe Val Leu Ala Ser Asn Gly Ala Phe Glu Ser Ser Gly
                325                 330                 335
```

-continued

```
Cys Ile Tyr Asp Ile Pro Asp Glu Tyr Leu Glu Ile Ser Asn Glu Asn
            340                 345                 350

Gly Lys Ile Tyr Thr Tyr Ile Lys Asn Ile Lys Glu Ile Ile Asp Glu
        355                 360                 365

Ala Ala Glu Asn Arg Lys Leu Phe Ser Glu Ile Gly Val Pro Trp Ala
    370                 375                 380

Lys Gln Thr Glu Glu Phe Asp Leu Ile Asp Tyr Cys Cys Phe Ser Met
385                 390                 395                 400

Phe Lys Arg


<210> SEQ ID NO 122
<211> LENGTH: 1089
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 122

atgaaagctt atgcaatggt attagaagaa tttaacaagc cgttaaaagc aaaagagttt     60

gaactaatca agccttctga tggtgaactt cttgttaaaa ttgaagcggc gggtgtttgt    120

ggatctgatg tgcatatgtt cagaggtaat gacccgcgta caaaacttcc catgatttta    180

gggcatgaag gtgttggacg tgtgtatgct atttcaggtc agtggcgtga tataaatggc    240

gagaaaattc aagagggaga tttgataatt tgggacaggg gtgttgtgtg tggtaggtgc    300

tacttttgtg ctgtcaaaaa agaaagctat ctgtgtccgc acagatggac atatgggata    360

agcgttagct gcgcagagcc tccgcatttg agaggctgct attcggagta catttatctt    420

cacaaagata cgaaagtgat aaaaataaaa gagaatgttg atccagaaat tttagtatct    480

gcctcatgtt ctggtgcaac gtgtgctcat gcttttgaca ttgtttcacc tgattttggt    540

gacagtgtcc taattcaagg gccaggtcct atagggcttt atgcaatcat ttttgcaaaa    600

cttagaggag cacgaaatat aattgtgatt ggtggcacaa aagaaagact taaaatgtgt    660

gaagaatttg gggcaacgca tgtgcttgat agaaattcaa ctacagcttg ccaaagacag    720

gaaataataa tggatatcac aaatgggcgt ggagtcgatt tggcaattga agctgtggga    780

catccatcag cagtaagtga gggaataaaa cttgttcgaa atggtggaag ctacttatca    840

cttggttttg gtgacccaaa cggcagcgtt acactcgatt gttactatga tattgtgaga    900

aaaaatttaa gatatcaagg ggtatgggtc agcgatacaa aacatttata tatggcagtg    960

aatgttgtgc tccagaacag ggaacttttc aaaaagatga ttacaaatgt ttataagttg   1020

actgatgcga caaaagctct tgaggatatg gaaaacaaaa atacaataaa atctgttcta   1080

aagccttga                                                           1089


<210> SEQ ID NO 123
<211> LENGTH: 362
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 123

Met Lys Ala Tyr Ala Met Val Leu Glu Glu Phe Asn Lys Pro Leu Lys
1               5                   10                  15

Ala Lys Glu Phe Glu Leu Ile Lys Pro Ser Asp Gly Glu Leu Leu Val
            20                  25                  30

Lys Ile Glu Ala Ala Gly Val Cys Gly Ser Asp Val His Met Phe Arg
        35                  40                  45
```

-continued

```
Gly Asn Asp Pro Arg Thr Lys Leu Pro Met Ile Leu Gly His Glu Gly
    50                  55                  60

Val Gly Arg Val Tyr Ala Ile Ser Gly Gln Trp Arg Asp Ile Asn Gly
65                  70                  75                  80

Glu Lys Ile Gln Glu Gly Asp Leu Ile Ile Trp Asp Arg Gly Val Val
                85                  90                  95

Cys Gly Arg Cys Tyr Phe Cys Ala Val Lys Lys Glu Ser Tyr Leu Cys
            100                 105                 110

Pro His Arg Trp Thr Tyr Gly Ile Ser Val Ser Cys Ala Glu Pro Pro
        115                 120                 125

His Leu Arg Gly Cys Tyr Ser Glu Tyr Ile Tyr Leu His Lys Asp Thr
    130                 135                 140

Lys Val Ile Lys Ile Lys Glu Asn Val Asp Pro Glu Ile Leu Val Ser
145                 150                 155                 160

Ala Ser Cys Ser Gly Ala Thr Cys Ala His Ala Phe Asp Ile Val Ser
                165                 170                 175

Pro Asp Phe Gly Asp Ser Val Leu Ile Gln Gly Pro Gly Pro Ile Gly
            180                 185                 190

Leu Tyr Ala Ile Ile Phe Ala Lys Leu Arg Gly Ala Arg Asn Ile Ile
        195                 200                 205

Val Ile Gly Gly Thr Lys Glu Arg Leu Lys Met Cys Glu Glu Phe Gly
    210                 215                 220

Ala Thr His Val Leu Asp Arg Asn Ser Thr Thr Ala Cys Gln Arg Gln
225                 230                 235                 240

Glu Ile Ile Met Asp Ile Thr Asn Gly Arg Gly Val Asp Leu Ala Ile
                245                 250                 255

Glu Ala Val Gly His Pro Ser Ala Val Ser Glu Gly Ile Lys Leu Val
            260                 265                 270

Arg Asn Gly Gly Ser Tyr Leu Ser Leu Gly Phe Gly Asp Pro Asn Gly
        275                 280                 285

Ser Val Thr Leu Asp Cys Tyr Tyr Asp Ile Val Arg Lys Asn Leu Arg
    290                 295                 300

Tyr Gln Gly Val Trp Val Ser Asp Thr Lys His Leu Tyr Met Ala Val
305                 310                 315                 320

Asn Val Val Leu Gln Asn Arg Glu Leu Phe Lys Lys Met Ile Thr Asn
                325                 330                 335

Val Tyr Lys Leu Thr Asp Ala Thr Lys Ala Leu Glu Asp Met Glu Asn
            340                 345                 350

Lys Asn Thr Ile Lys Ser Val Leu Lys Pro
        355                 360


<210> SEQ ID NO 124
<211> LENGTH: 1806
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 124

gtgatataca aagcagatgt gctggtagta ggtagtggcg gagcaggttt gagagctgca      60

attgcagcgt gtgaaagagc atatgagagc cgaaaaagaa taaaagtttt gctggcaagt     120

aaaggaaaaa taggaagttg tggtactaca gctcttgcat actctgatag aatggcattc     180

catgtcacac ttcccacaac agagcctaaa ggtgaagata actggaaata tcatgcaaaa     240
```

-continued

```
gatatctatg agattggcgg acttgtttct gattatgact tggctgagat tttagcaaaa    300
aactcagccg atgcttattt ttacttagat agtctcggtg ttccgtttgt aaaagaaaat    360
ggcgtacctg ctcaatttgt gacagacggt tctatatatg cgcgtgcatg ttttacagga    420
cctgatactg ctgttcaaat agaaaaggct ttgattcgaa agcttggtga gatgaaggat    480
attgaagttt tagaagatgt gatgataagt gatttgattg ttgtgaataa caaagtttgc    540
ggagcaattg cttttaaagg gaatcaaaat atcataatac ttgcaaaagc cattgtttta    600
gcaacaggtg gggcaggaag tatttataaa agcaatgtgt ttccaccacg catgacaggt    660
gatgggtatg caatggcact tcgtgcaggg gctttgcttg tgaatatgga gtttattcaa    720
ataggtcttt catcccccaa aacaaaactt gcgtgttcag gaagtataat gagatgtgtc    780
cccaggtttg taaatgaaaa aggagaagaa tttttattaa attatcctat cgcgtacaat    840
gatgtatttg aaaaaggtgc aacgtggcca ataagctacg agcacaagac atgtttgata    900
gacattgcag tgttcagaga gattgcccgt gggggtaaag tgtttttaga ctttactcaa    960
aatccaaaag gttttgaatt taaacaccta agagaagact taaaacaaag atattactca   1020
gaagttaaaa atcttacaaa taaaagatct actccatacg aaagactttg tgaaataaat   1080
cctcagacag ttgagtggtt tttgaaaaga ggaattgacc ttagaaacca aatgttagaa   1140
attgcgccat caattcagca tttccagggt ggtgtaaaaa ttagagaaaa tgcaaataca   1200
gcaattagtg gcttgtatgc ctgtggcgag tgtgcaggcg gacagcatgg agcaaacaga   1260
ccaggcggaa atgcactttt ggatactcag gtttttggta agatagcagg ggaaagtagc   1320
tttgaatttg cctcaaacac ttcgattgat gaagaatctg caatttctga agcaaaccgc   1380
ctgtttgaaa gctataaaat ctatatagct gaagatggca ttgatttgga gaatgctata   1440
tcagaactaa atagggttat ggatttgtac gcaagtgttg tgagacatca ggatggactt   1500
caaaaagcac ttatgaaaat tgaggagttg aaaacaagaa aaatcaagcc tgttgagtat   1560
gaatatcttt tagagctaaa aaacatgctt ttgtgtgcag aggcagtggt aaaaagttgc   1620
attttaagag atgaaagcag aggaccgcac ttgatgttcg aaaattacag cgatttgtgg   1680
ccaaagccaa gagatgaaag atacaacata tatcacgtat gcaagctaaa caaagagaca   1740
aatcaagttg aagtctttcc tatggaacca gtaaaacccg aaaccttagg gggcaaagta   1800
aaatga                                                              1806
```

<210> SEQ ID NO 125
<211> LENGTH: 601
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 125

```
Val Ile Tyr Lys Ala Asp Val Leu Val Val Gly Ser Gly Gly Ala Gly
1               5                   10                  15

Leu Arg Ala Ala Ile Ala Ala Cys Glu Arg Ala Tyr Glu Ser Arg Lys
            20                  25                  30

Arg Ile Lys Val Leu Leu Ala Ser Lys Gly Lys Ile Gly Ser Cys Gly
        35                  40                  45

Thr Thr Ala Leu Ala Tyr Ser Asp Arg Met Ala Phe His Val Thr Leu
    50                  55                  60

Pro Thr Thr Glu Pro Lys Gly Glu Asp Asn Trp Lys Tyr His Ala Lys
```

-continued

```
65                  70                  75                  80

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

Ile Leu Ala Lys Asn Ser Ala Asp Ala Tyr Phe Tyr Leu Asp Ser Leu
            100                 105                 110

Gly Val Pro Phe Val Lys Glu Asn Gly Val Pro Ala Gln Phe Val Thr
        115                 120                 125

Asp Gly Ser Ile Tyr Ala Arg Ala Cys Phe Thr Gly Pro Asp Thr Ala
    130                 135                 140

Val Gln Ile Glu Lys Ala Leu Ile Arg Lys Leu Gly Glu Met Lys Asp
145                 150                 155                 160

Ile Glu Val Leu Glu Asp Val Met Ile Ser Asp Leu Ile Val Val Asn
                165                 170                 175

Asn Lys Val Cys Gly Ala Ile Ala Phe Lys Gly Asn Gln Asn Ile Ile
            180                 185                 190

Ile Leu Ala Lys Ala Ile Val Leu Ala Thr Gly Gly Ala Gly Ser Ile
        195                 200                 205

Tyr Lys Ser Asn Val Phe Pro Pro Arg Met Thr Gly Asp Gly Tyr Ala
    210                 215                 220

Met Ala Leu Arg Ala Gly Ala Leu Leu Val Asn Met Glu Phe Ile Gln
225                 230                 235                 240

Ile Gly Leu Ser Ser Pro Lys Thr Lys Leu Ala Cys Ser Gly Ser Ile
                245                 250                 255

Met Arg Cys Val Pro Arg Phe Val Asn Glu Lys Gly Glu Glu Phe Leu
            260                 265                 270

Leu Asn Tyr Pro Ile Ala Tyr Asn Asp Val Phe Glu Lys Gly Ala Thr
        275                 280                 285

Trp Pro Ile Ser Tyr Glu His Lys Thr Cys Leu Ile Asp Ile Ala Val
    290                 295                 300

Phe Arg Glu Ile Ala Arg Gly Gly Lys Val Phe Leu Asp Phe Thr Gln
305                 310                 315                 320

Asn Pro Lys Gly Phe Glu Phe Lys His Leu Arg Glu Asp Leu Lys Gln
                325                 330                 335

Arg Tyr Tyr Ser Glu Val Lys Asn Leu Thr Asn Lys Arg Ser Thr Pro
            340                 345                 350

Tyr Glu Arg Leu Cys Glu Ile Asn Pro Gln Thr Val Glu Trp Phe Leu
        355                 360                 365

Lys Arg Gly Ile Asp Leu Arg Asn Gln Met Leu Glu Ile Ala Pro Ser
    370                 375                 380

Ile Gln His Phe Gln Gly Gly Val Lys Ile Arg Glu Asn Ala Asn Thr
385                 390                 395                 400

Ala Ile Ser Gly Leu Tyr Ala Cys Gly Glu Cys Ala Gly Gly Gln His
                405                 410                 415

Gly Ala Asn Arg Pro Gly Gly Asn Ala Leu Leu Asp Thr Gln Val Phe
            420                 425                 430

Gly Lys Ile Ala Gly Glu Ser Ser Phe Glu Phe Ala Ser Asn Thr Ser
        435                 440                 445

Ile Asp Glu Glu Ser Ala Ile Ser Glu Ala Asn Arg Leu Phe Glu Ser
    450                 455                 460

Tyr Lys Ile Tyr Ile Ala Glu Asp Gly Ile Asp Leu Glu Asn Ala Ile
465                 470                 475                 480
```

-continued

```
Ser Glu Leu Asn Arg Val Met Asp Leu Tyr Ala Ser Val Val Arg His
                485                 490                 495

Gln Asp Gly Leu Gln Lys Ala Leu Met Lys Ile Glu Glu Leu Lys Thr
            500                 505                 510

Arg Lys Ile Lys Pro Val Glu Tyr Glu Tyr Leu Leu Glu Leu Lys Asn
        515                 520                 525

Met Leu Leu Cys Ala Glu Ala Val Val Lys Ser Cys Ile Leu Arg Asp
    530                 535                 540

Glu Ser Arg Gly Pro His Leu Met Phe Glu Asn Tyr Ser Asp Leu Trp
545                 550                 555                 560

Pro Lys Pro Arg Asp Glu Arg Tyr Asn Ile Tyr His Val Cys Lys Leu
                565                 570                 575

Asn Lys Glu Thr Asn Gln Val Glu Val Phe Pro Met Glu Pro Val Lys
            580                 585                 590

Pro Glu Thr Leu Gly Gly Lys Val Lys
        595                 600


<210> SEQ ID NO 126
<211> LENGTH: 1146
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 126

atgccaaacc tatcaacgac gtatgcaaag ctgaatttaa gaacacctgt aattgttgca      60

tctgctggca ttactggaac tgtggagagg cttcaaagat gcgaagaaaa cggtgctggg     120

gcagttgtga caaaaagtct ttttcaaaag gaaatatgca gaattgcacc cactccacgg     180

tttaaaatag tcaagcatga aaacacgttt acgctttact catatgaaca ggcaagcgaa     240

tttaaccctc aagagtatgc tgaatttata ttcaaagcaa aacaaaagct aagcattcca     300

gttattgcga gtataaactg ctacacagat gatgcatggc ttgagtatag caagcttatg     360

gagcaggcag gggctgatgc gatagagcta aacctttcat gtcctcacgg tgtgcatata     420

atgtctggta tggatgtaat tgaagagatg gtcaacacaa caaaacttgt caaaagcaat     480

gttaagatac cagtgatacc caaaatgact cctcaatcta caaatccggg atctgatgcc     540

ttaagactcg acagtgcagg agcagacggg cttgtaatgt tcaatagatt tacagggctt     600

gacattgata ttgagaaaga agcacccatt ttgcacggcg gttatgcagg gcatggtggt     660

ccgtgggcaa ttatgtatgg tttgaggtgg ataagcgctg tatcgccaaa agtgaaatgt     720

agtatcagtg caagcggcgg tgccatgaat ggagaagatg ttgtcaaata catattggca     780

ggtgcgtcgg ctgttcaagt ttgcacaact gttattttga atggctatgg ggttataaaa     840

aagataaaca agtatttaga agagtacatg gagagaaaag gttacaacac aattgatgat     900

tttaaaggaa aggtgtgcag tagaattctt gacatggact ctgttgacag aacgcactgg     960

gctgttgcaa ggattgacaa agaaaaatgc acatcttgtg gcaagtgctt cacagtttgc    1020

atatatgatg caattgaaaa ggatgatgga aagtttaaag taaatcaaaa ctgcgatggc    1080

tgcggacttt gtgcagaact gtgcccagcc aaggcaatct taatggtaag aagaggtgaa    1140

gtttaa                                                               1146


<210> SEQ ID NO 127
<211> LENGTH: 381
```

-continued

<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 127

Met Pro Asn Leu Ser Thr Thr Tyr Ala Lys Leu Asn Leu Arg Thr Pro
1 5 10 15

Val Ile Val Ala Ser Ala Gly Ile Thr Gly Thr Val Glu Arg Leu Gln
20 25 30

Arg Cys Glu Glu Asn Gly Ala Gly Ala Val Val Thr Lys Ser Leu Phe
35 40 45

Gln Lys Glu Ile Cys Arg Ile Ala Pro Thr Pro Arg Phe Lys Ile Val
50 55 60

Lys His Glu Asn Thr Phe Thr Leu Tyr Ser Tyr Glu Gln Ala Ser Glu
65 70 75 80

Phe Asn Pro Gln Glu Tyr Ala Glu Phe Ile Phe Lys Ala Lys Gln Lys
85 90 95

Leu Ser Ile Pro Val Ile Ala Ser Ile Asn Cys Tyr Thr Asp Asp Ala
100 105 110

Trp Leu Glu Tyr Ser Lys Leu Met Glu Gln Ala Gly Ala Asp Ala Ile
115 120 125

Glu Leu Asn Leu Ser Cys Pro His Gly Val His Ile Met Ser Gly Met
130 135 140

Asp Val Ile Glu Glu Met Val Asn Thr Thr Lys Leu Val Lys Ser Asn
145 150 155 160

Val Lys Ile Pro Val Ile Pro Lys Met Thr Pro Gln Ser Thr Asn Pro
165 170 175

Gly Ser Asp Ala Leu Arg Leu Asp Ser Ala Gly Ala Asp Gly Leu Val
180 185 190

Met Phe Asn Arg Phe Thr Gly Leu Asp Ile Asp Ile Glu Lys Glu Ala
195 200 205

Pro Ile Leu His Gly Gly Tyr Ala Gly His Gly Gly Pro Trp Ala Ile
210 215 220

Met Tyr Gly Leu Arg Trp Ile Ser Ala Val Ser Pro Lys Val Lys Cys
225 230 235 240

Ser Ile Ser Ala Ser Gly Gly Ala Met Asn Gly Glu Asp Val Val Lys
245 250 255

Tyr Ile Leu Ala Gly Ala Ser Ala Val Gln Val Cys Thr Thr Val Ile
260 265 270

Leu Asn Gly Tyr Gly Val Ile Lys Lys Ile Asn Lys Tyr Leu Glu Glu
275 280 285

Tyr Met Glu Arg Lys Gly Tyr Asn Thr Ile Asp Asp Phe Lys Gly Lys
290 295 300

Val Cys Ser Arg Ile Leu Asp Met Asp Ser Val Asp Arg Thr His Trp
305 310 315 320

Ala Val Ala Arg Ile Asp Lys Glu Lys Cys Thr Ser Cys Gly Lys Cys
325 330 335

Phe Thr Val Cys Ile Tyr Asp Ala Ile Glu Lys Asp Asp Gly Lys Phe
340 345 350

Lys Val Asn Gln Asn Cys Asp Gly Cys Gly Leu Cys Ala Glu Leu Cys
355 360 365

Pro Ala Lys Ala Ile Leu Met Val Arg Arg Gly Glu Val
370 375 380

-continued

<210> SEQ ID NO 128
<211> LENGTH: 1539
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 128

ttgtacatca ttatattcca ctttcagtac atttttattt tatcattaac accttttgaa 60

gttgagatta ataaatttgc ccaaattggc tgtgagaata gacttacaat tgttgtaaac 120

aacatcttgg actggagttg tcttccacca gggtttataa gggaatacaa tgacccaatg 180

catccagaag ggtataaaac tcaggaatat ctttttgact ttttcaacta ttcaggtatt 240

cacagaccag ttttgctcta caccacttcc aaaacatata ttgaggatat taagattgaa 300

acccagattg agggtcaaaa gggtatagtt tgctttaagg tggctgtaag tggcgaaaaa 360

aaggatgaat gtcagatagc agtagctttg tatgacaaag atggaaagca aatagcaaag 420

gtcgaagggc cagagggtat gatagaggtt ggagatgcga tattttggga gccttcaaat 480

ccatatcttt acaaactaaa tgtaacttta atacacgatg aaaaggtggt agatgaatat 540

tatcttcctg tgggaataag gacagttgag gtaaaaggca aaagactttt cctaaatggt 600

aagccagtgt atcttaaagg tttggcaaag catgaagaca gtgatataag gggcaaggga 660

tacgaccctg tgatagctgt gaaagatttc aacctcctaa aatggatagg agcaaactca 720

ttcagaacat cacattatcc ttacgcagaa gagattttaa acttggcaga cgagtatggt 780

tttttggtaa ttgacgaggc accagctgtt ggcatgaatt tctttaacaa aaacgaaaaa 840

gtgtttaccg cggagagagt aaaccaaaag acattagaac atcacttaga agttataaga 900

caacttattg caagggataa aaaccatcca agtgtgatta tgtggagtgt ggcaaatgag 960

gctgcaacat atgaagatgg ggcatatgaa tatttcaaaa gagtaataga tgaggtgaga 1020

aagcttgacc cgacaagacc ggtgacgctg gttgaatcct cttttccaga tgagaccaaa 1080

gtgggaagtc ttgttgatgt tatatgtgta aacaggtact attcatggta ttctgatcct 1140

ggcagactgg atttgataga gttccagctt gaaaaggagc tgaaaaggtg gtttgagctt 1200

tatcaaaaac cagtgataat aacagagtat ggggcagata caattgcagg atttcattca 1260

agtcctccaa tgatgttttc tgaggaatat cagtgtgaga tgcttgaaag atatcatagg 1320

gtgtttgaca ggctggattt tgtgataggc gaacacatat ggaactttgc agactttgca 1380

acaaaacaag aggttcgaag gattatgggc aacaggaaag gaatctttac aaggcaaaga 1440

cagccaaaag ccgcagcttt cttgctcaaa aaaagatggc aaaattcaga gcacaaaagg 1500

ctggaggaaa atgtttcaga agataaaaca cgtaattaa 1539

<210> SEQ ID NO 129
<211> LENGTH: 512
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 129

Leu Tyr Ile Ile Ile Phe His Phe Gln Tyr Ile Phe Ile Leu Ser Leu
1 5 10 15

Thr Pro Phe Glu Val Glu Ile Asn Lys Phe Ala Gln Ile Gly Cys Glu
20 25 30

Asn Arg Leu Thr Ile Val Val Asn Asn Ile Leu Asp Trp Ser Cys Leu

-continued

```
        35                  40                  45

Pro Pro Gly Phe Ile Arg Glu Tyr Asn Asp Pro Met His Pro Glu Gly
    50                  55                  60

Tyr Lys Thr Gln Glu Tyr Leu Phe Asp Phe Phe Asn Tyr Ser Gly Ile
65                  70                  75                  80

His Arg Pro Val Leu Leu Tyr Thr Thr Ser Lys Thr Tyr Ile Glu Asp
                85                  90                  95

Ile Lys Ile Glu Thr Gln Ile Glu Gly Gln Lys Gly Ile Val Cys Phe
            100                 105                 110

Lys Val Ala Val Ser Gly Glu Lys Lys Asp Glu Cys Gln Ile Ala Val
        115                 120                 125

Ala Leu Tyr Asp Lys Asp Gly Lys Gln Ile Ala Lys Val Glu Gly Pro
    130                 135                 140

Glu Gly Met Ile Glu Val Gly Asp Ala Ile Phe Trp Glu Pro Ser Asn
145                 150                 155                 160

Pro Tyr Leu Tyr Lys Leu Asn Val Thr Leu Ile His Asp Glu Lys Val
                165                 170                 175

Val Asp Glu Tyr Tyr Leu Pro Val Gly Ile Arg Thr Val Glu Val Lys
            180                 185                 190

Gly Lys Arg Leu Phe Leu Asn Gly Lys Pro Val Tyr Leu Lys Gly Leu
        195                 200                 205

Ala Lys His Glu Asp Ser Asp Ile Arg Gly Lys Gly Tyr Asp Pro Val
    210                 215                 220

Ile Ala Val Lys Asp Phe Asn Leu Leu Lys Trp Ile Gly Ala Asn Ser
225                 230                 235                 240

Phe Arg Thr Ser His Tyr Pro Tyr Ala Glu Glu Ile Leu Asn Leu Ala
                245                 250                 255

Asp Glu Tyr Gly Phe Leu Val Ile Asp Glu Ala Pro Ala Val Gly Met
            260                 265                 270

Asn Phe Phe Asn Lys Asn Glu Lys Val Phe Thr Ala Glu Arg Val Asn
        275                 280                 285

Gln Lys Thr Leu Glu His His Leu Glu Val Ile Arg Gln Leu Ile Ala
    290                 295                 300

Arg Asp Lys Asn His Pro Ser Val Ile Met Trp Ser Val Ala Asn Glu
305                 310                 315                 320

Ala Ala Thr Tyr Glu Asp Gly Ala Tyr Glu Tyr Phe Lys Arg Val Ile
                325                 330                 335

Asp Glu Val Arg Lys Leu Asp Pro Thr Arg Pro Val Thr Leu Val Glu
            340                 345                 350

Ser Ser Phe Pro Asp Glu Thr Lys Val Gly Ser Leu Val Asp Val Ile
        355                 360                 365

Cys Val Asn Arg Tyr Tyr Ser Trp Tyr Ser Asp Pro Gly Arg Leu Asp
    370                 375                 380

Leu Ile Glu Phe Gln Leu Glu Lys Glu Leu Lys Arg Trp Phe Glu Leu
385                 390                 395                 400

Tyr Gln Lys Pro Val Ile Ile Thr Glu Tyr Gly Ala Asp Thr Ile Ala
                405                 410                 415

Gly Phe His Ser Ser Pro Pro Met Met Phe Ser Glu Glu Tyr Gln Cys
            420                 425                 430

Glu Met Leu Glu Arg Tyr His Arg Val Phe Asp Arg Leu Asp Phe Val
        435                 440                 445
```

-continued

```
Ile Gly Glu His Ile Trp Asn Phe Ala Asp Phe Ala Thr Lys Gln Glu
    450                 455                 460

Val Arg Arg Ile Met Gly Asn Arg Lys Gly Ile Phe Thr Arg Gln Arg
465                 470                 475                 480

Gln Pro Lys Ala Ala Ala Phe Leu Leu Lys Lys Arg Trp Gln Asn Ser
                485                 490                 495

Glu His Lys Arg Leu Glu Glu Asn Val Ser Glu Asp Lys Thr Arg Asn
            500                 505                 510


<210> SEQ ID NO 130
<211> LENGTH: 1203
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 130

atgcagaact taacagcaca tttaaacatg ataagggcga tgaaaatgaa acctaacttt     60

tcagaacttg caagaatata tgggatggat agaagaacag ttaaaaaata ttatgagggt    120

tatgaaggaa aacctaagaa tagaaataaa ccaagtaaat tggacaaata ctatgatgag    180

ataaaatcaa agcttgctat caaaggagtt acagtcaagg gtgtttatga gtatttaaaa    240

tcaaaagatg agacaatagg aacatattca aacttcaata agtatgttaa gaaaaaagga    300

ttaaagccag agaagaaaat aaaaggtcac ccaagatttg agacagatcc aggtgagcaa    360

gcgcaagttg attggaaaga gaatataaag cttgtctcaa gaaatggaga ggagtttatc    420

attaatgttc ttgattttaa attaggttat tcaaggtatt gctgctttga gataaacagg    480

acaaaaactc aagaagaatt aatagaaact ctaataagaa tattcaaaga tataggcgga    540

gtaccgagag agattttatt tgacaataca gcagcagttg ttgatataac aggtgagaaa    600

attaaagtaa attcaagatt taaaagtttt gcaaaagact ttgggtttga agtaaaactg    660

tgcaaaccaa gacattcgta cacaaaagga aaagttgaag cagcaaacaa gtttatagat    720

tggatactgc catatcaggg tgaatttgaa acagaagagg acttagtaag gataataaaa    780

gagataaacg caaaggtcaa tatgcagcca aatcaaacaa ctcaagttcc acctgctctt    840

ctgtttcaaa aagaaaaaga gtatttacaa cccttgccag acaaaaggtt aatagacagt    900

tacctaaatt cctacaagtc agttaaagtc caaaaggact ctctgattta ctacaaggga    960

agtaaatact ctgttccacc cgaatacata ggaaagacag tccaagtaaa ggaggtggaa   1020

aacaaaattt atatttatta taacacaaac ctgttaagga tacatgttat tgatgaaaaa   1080

aatatcaatt atcacgatga agattacaag cagctaatgc taatgagagt tggtcaaaga   1140

gaagagctta acaagatatg tgaggaaaac ctaaagaaat ttgataatct gttgaaaacc   1200

taa                                                                 1203


<210> SEQ ID NO 131
<211> LENGTH: 400
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 131

Met Gln Asn Leu Thr Ala His Leu Asn Met Ile Arg Ala Met Lys Met
1               5                   10                  15

Lys Pro Asn Phe Ser Glu Leu Ala Arg Ile Tyr Gly Met Asp Arg Arg
            20                  25                  30
```

-continued

```
Thr Val Lys Lys Tyr Tyr Glu Gly Tyr Glu Gly Lys Pro Lys Asn Arg
        35                  40                  45

Asn Lys Pro Ser Lys Leu Asp Lys Tyr Tyr Asp Glu Ile Lys Ser Lys
    50                  55                  60

Leu Ala Ile Lys Gly Val Thr Val Lys Gly Val Tyr Glu Tyr Leu Lys
65                  70                  75                  80

Ser Lys Asp Glu Thr Ile Gly Thr Tyr Ser Asn Phe Asn Lys Tyr Val
                85                  90                  95

Lys Lys Lys Gly Leu Lys Pro Glu Lys Lys Ile Lys Gly His Pro Arg
            100                 105                 110

Phe Glu Thr Asp Pro Gly Glu Gln Ala Gln Val Asp Trp Lys Glu Asn
        115                 120                 125

Ile Lys Leu Val Ser Arg Asn Gly Glu Glu Phe Ile Ile Asn Val Leu
    130                 135                 140

Asp Phe Lys Leu Gly Tyr Ser Arg Tyr Cys Cys Phe Glu Ile Asn Arg
145                 150                 155                 160

Thr Lys Thr Gln Glu Glu Leu Ile Glu Thr Leu Ile Arg Ile Phe Lys
                165                 170                 175

Asp Ile Gly Gly Val Pro Arg Glu Ile Leu Phe Asp Asn Thr Ala Ala
            180                 185                 190

Val Val Asp Ile Thr Gly Glu Lys Ile Lys Val Asn Ser Arg Phe Lys
        195                 200                 205

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

His Ser Tyr Thr Lys Gly Lys Val Glu Ala Ala Asn Lys Phe Ile Asp
225                 230                 235                 240

Trp Ile Leu Pro Tyr Gln Gly Glu Phe Glu Thr Glu Glu Asp Leu Val
                245                 250                 255

Arg Ile Ile Lys Glu Ile Asn Ala Lys Val Asn Met Gln Pro Asn Gln
            260                 265                 270

Thr Thr Gln Val Pro Pro Ala Leu Leu Phe Gln Lys Glu Lys Glu Tyr
        275                 280                 285

Leu Gln Pro Leu Pro Asp Lys Arg Leu Ile Asp Ser Tyr Leu Asn Ser
    290                 295                 300

Tyr Lys Ser Val Lys Val Gln Lys Asp Ser Leu Ile Tyr Tyr Lys Gly
305                 310                 315                 320

Ser Lys Tyr Ser Val Pro Pro Glu Tyr Ile Gly Lys Thr Val Gln Val
                325                 330                 335

Lys Glu Val Glu Asn Lys Ile Tyr Ile Tyr Tyr Asn Thr Asn Leu Leu
            340                 345                 350

Arg Ile His Val Ile Asp Glu Lys Asn Ile Asn Tyr His Asp Glu Asp
        355                 360                 365

Tyr Lys Gln Leu Met Leu Met Arg Val Gly Gln Arg Glu Glu Leu Asn
    370                 375                 380

Lys Ile Cys Glu Glu Asn Leu Lys Lys Phe Asp Asn Leu Leu Lys Thr
385                 390                 395                 400
```

<210> SEQ ID NO 132
<211> LENGTH: 765
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

-continued

<400> SEQUENCE: 132

atgagcaact atgtgaaact acttaacaac ttagaagagt taggtcttca caacataaag 60

aataaccttg acaaatactt agatttagtg gcaagcggag aaaaaagtat gacagatgca 120

ttatatgaac ttagtaattt agagataaaa gccaaagaag aaagggcgat attaggatgt 180

gtgagggtgg caaatttccc atttataaaa ggtatagaag attttgattt ttcatttcag 240

ccaagtataa acaagcaaca gataatggac ttaatgagtt tgagattttt agagggtaat 300

gaaaatatac tatttgtcgg aacaccaggg gtagggaaaa cgcatctagc cacagcaata 360

ggtatagagt gtgcaaaacg aaggtattca acatatttta tacattttca agagttaata 420

gcccagctaa agaaagcatt attggagaac agattagagt acagacttaa gcatttttcg 480

aaatacaaag ttttaataat agatgagata ggttatttgc caatagacaa tgatggagca 540

aatttatttt tccagctgat atcgagcaga tatgagaaga gcagtacaat aataacaact 600

aatgttgtat tctcagaatg gggagagata tttggtggag cgacaatagc aaatgcaatt 660

ttagataggc tactgcatca ttcttacgtg attttcataa aaggtccttc atacagatta 720

cagtcaaaaa cagcatattt tagcaataca aaccagcaaa gttaa 765

<210> SEQ ID NO 133
<211> LENGTH: 254
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 133

Met Ser Asn Tyr Val Lys Leu Leu Asn Asn Leu Glu Glu Leu Gly Leu
1 5 10 15

His Asn Ile Lys Asn Asn Leu Asp Lys Tyr Leu Asp Leu Val Ala Ser
20 25 30

Gly Glu Lys Ser Met Thr Asp Ala Leu Tyr Glu Leu Ser Asn Leu Glu
35 40 45

Ile Lys Ala Lys Glu Glu Arg Ala Ile Leu Gly Cys Val Arg Val Ala
50 55 60

Asn Phe Pro Phe Ile Lys Gly Ile Glu Asp Phe Asp Phe Ser Phe Gln
65 70 75 80

Pro Ser Ile Asn Lys Gln Gln Ile Met Asp Leu Met Ser Leu Arg Phe
85 90 95

Leu Glu Gly Asn Glu Asn Ile Leu Phe Val Gly Thr Pro Gly Val Gly
100 105 110

Lys Thr His Leu Ala Thr Ala Ile Gly Ile Glu Cys Ala Lys Arg Arg
115 120 125

Tyr Ser Thr Tyr Phe Ile His Phe Gln Glu Leu Ile Ala Gln Leu Lys
130 135 140

Lys Ala Leu Leu Glu Asn Arg Leu Glu Tyr Arg Leu Lys His Phe Ser
145 150 155 160

Lys Tyr Lys Val Leu Ile Ile Asp Glu Ile Gly Tyr Leu Pro Ile Asp
165 170 175

Asn Asp Gly Ala Asn Leu Phe Phe Gln Leu Ile Ser Ser Arg Tyr Glu
180 185 190

Lys Ser Ser Thr Ile Ile Thr Thr Asn Val Val Phe Ser Glu Trp Gly
195 200 205

Glu Ile Phe Gly Gly Ala Thr Ile Ala Asn Ala Ile Leu Asp Arg Leu

-continued

```
    210                 215                 220

Leu His His Ser Tyr Val Ile Phe Ile Lys Gly Pro Ser Tyr Arg Leu
225                 230                 235                 240

Gln Ser Lys Thr Ala Tyr Phe Ser Asn Thr Asn Gln Gln Ser
                245                 250


<210> SEQ ID NO 134
<211> LENGTH: 399
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 134

atgctttatc cacgtgagac aagaacaaga gagataaaag acttaagcgg tatatggagg      60

tttaaagttg acagggacaa taaaggctat gaacagaggt ggtttgaaaa accccttgaa     120

gatgctattt tgatgcctgt gccgtcaagc tacaatgaca taacccagga tgagagcata     180

agagatcaca taggtgatgt gtggtatgaa aggacatttt acatacctga aggttggaag     240

gacaaaagaa tagttttgag agttggaagt gctactcatc acgcaagagt atttttaaat     300

ggtaaggaaa tagcccagaa caaaggtggg tttttgcctt ttgtaaatgt caatatcaaa     360

atgaacaaaa aatcgaaaat aaaaatgtac aaaaaataa                            399


<210> SEQ ID NO 135
<211> LENGTH: 132
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 135

Met Leu Tyr Pro Arg Glu Thr Arg Thr Arg Glu Ile Lys Asp Leu Ser
1               5                   10                  15

Gly Ile Trp Arg Phe Lys Val Asp Arg Asp Asn Lys Gly Tyr Glu Gln
            20                  25                  30

Arg Trp Phe Glu Lys Pro Leu Glu Asp Ala Ile Leu Met Pro Val Pro
        35                  40                  45

Ser Ser Tyr Asn Asp Ile Thr Gln Asp Glu Ser Ile Arg Asp His Ile
    50                  55                  60

Gly Asp Val Trp Tyr Glu Arg Thr Phe Tyr Ile Pro Glu Gly Trp Lys
65                  70                  75                  80

Asp Lys Arg Ile Val Leu Arg Val Gly Ser Ala Thr His His Ala Arg
                85                  90                  95

Val Phe Leu Asn Gly Lys Glu Ile Ala Gln Asn Lys Gly Gly Phe Leu
            100                 105                 110

Pro Phe Val Asn Val Asn Ile Lys Met Asn Lys Lys Ser Lys Ile Lys
        115                 120                 125

Met Tyr Lys Lys
    130


<210> SEQ ID NO 136
<211> LENGTH: 2598
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 136

atgcaagaaa taaagctaaa atggcttgtc aaaccgcagg taggcaatat tggtgcaacg      60

ttttgtattc catggaaaaa agctcagctc tggaatacag aaaatgttat tgtcagaaac     120
```

-continued

```
aaggatggaa gtattattcc tactcaaaac tgggcacttt catactgggg tgatgggtcg    180
gtaaaggtga gtagccacgc tgctttcttt ggcaccaatt tacctgatga gctctttgca    240
tgtattgaag aagataccgt aacaaaattc gaaacaatgg ttgaagttta tgaagacaag    300
gagcatatta gcgtgacaac aggcaagctt gagtgcaaaa ttagcaagac cggagataaa    360
attattgaat atttaaaagt aggacaaaag attatttgct ctaatctcac gcaaattctt    420
attacagaaa gaatttctaa ttttacagga tataagacta aagttgaaga agaatacatt    480
cctcaaatat taggtgcaaa agtagagtct cacggacctt tgagatgtgt tataagggtg    540
tggggaagac acttgcggaa taccaagatg acatttgatg gctctttact caaacaaggg    600
tggttgcctt ttgagttgag attatatttc tatgcaaatt cagataaaat taagattgtt    660
catacattta tttacaacgg aaatccccat caagatttta taaaaggatt gggcttaaaa    720
tttgatcttc ctttgcactc tcctctgtac aatagatttg taagatttgg aggggacagt    780
ggactttttt gtgagtcacc aaaaaatttg atagatgtgt acagaaaaga gagatacaaa    840
ataatgtttc aaaatcaggt caatggacaa ccagtagatt ttgatattga cgaggaccac    900
cattatatta gggagattaa caatactcca gagtgggatg aatttattct ttttcaagat    960
tcttccgagc actatgttat taaaaaacgt acatcttgcc agtgcagttt tgtaaaagca   1020
acagaaggca gaagatctat ggggtttgca tatattggtg atgagaatgg tgggcttgga   1080
gtagctttaa aggatttttg gcaaaaatat ccttcaggat ttgaggcaaa agatttaact   1140
tcttctgagg caaatttaat agtttggctt tggccacctt atgctgaagt tatggatttg   1200
agacactatg attttggtac atacacatct tcttcatatg aagggtttga tgagtataga   1260
agtactccct atggtatagc taatacaaat gagctgtggc tgtactgttt tgattatgct   1320
ccaacaagtg agcagctttt ggaatatgca aaaatgcaag gattttctcc tcttcttgta   1380
tgtagccctg aaaggtataa agagacagag atatttggag gattaaattt accggataat   1440
agcaactcca agaaagaaaa gattgaaaga atactgacag ctattgttga tttttatcta   1500
aaggaagttg agagaagaaa gtggtacggt ttttgggact atggtgattt tatgcatagc   1560
tatgatatgg taagacatat gtggaaatat gatattggtg gatatgcttg gcagaacaca   1620
gagcttgtgc caaacatttg gctttggctt atgtttttaa gaacaggacg ttttgatata   1680
tttaaaatgg ctgaggcaat gacaagacat acatcagaag tagatgtgta tcatcttggg   1740
gaatacaagg gactgggttc aaggcacaat gttgttcatt ggggatgtgg ttgcaaagaa   1800
gtgagaatta gcatggcata ccttcacagg tattactatt atcttactgc cgatgaaaga   1860
atagggcagc ttatggatga tgtaaaagat attgacaaac aaattatgca catggaccct   1920
atgcgtgcgt attttacgaa tgatcctgaa aatagagtac acataagggt tggaccggat   1980
gttatgacct tttgcgccaa ttggtttgta agatgggaaa ggcacaagga agaaatttac   2040
aaacaaaaga taataaagat actggacttt ttgaagaaaa atccagcagc atttatttcg   2100
ggtggggtat ttgactatga tcctgaaaaa acccagttaa aacctgttga atacactggt   2160
ggttcaaatt ttgtattttg ttttggaaat acacttgttt ggcttgagat agccaagaat   2220
tttgaagata aagaatttga agaattgtgt gctcagcagg gccttttta tactgaattt   2280
aaagaaaaca aaaatgagat tttaaaaagc tggggtgttc caagctttgg atttaagttg   2340
aacatgctca atattgggat ggctgccttt gctgcaatga agaaaaatat tcctgaattg   2400
```

-continued

```
aaaaaagaaa tatggcagat gtttttagac cataacaaaa atccatggct aaaatttatt   2460

tgtgatggag gtataaattt acaattggca actgtacctg cgcctgttgc tgagattcct   2520

tttatttcaa cgaatatagc ttctcaatgg tcattaaacg cattgcttgc gctagaattt   2580

attggagatg agctatga                                                 2598


<210> SEQ ID NO 137
<211> LENGTH: 865
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 137

Met Gln Glu Ile Lys Leu Lys Trp Leu Val Lys Pro Gln Val Gly Asn
1               5                   10                  15

Ile Gly Ala Thr Phe Cys Ile Pro Trp Lys Lys Ala Gln Leu Trp Asn
            20                  25                  30

Thr Glu Asn Val Ile Val Arg Asn Lys Asp Gly Ser Ile Ile Pro Thr
        35                  40                  45

Gln Asn Trp Ala Leu Ser Tyr Trp Gly Asp Gly Ser Val Lys Val Ser
    50                  55                  60

Ser His Ala Ala Phe Phe Gly Thr Asn Leu Pro Asp Glu Leu Phe Ala
65                  70                  75                  80

Cys Ile Glu Glu Asp Thr Val Thr Lys Phe Glu Thr Met Val Glu Val
                85                  90                  95

Tyr Glu Asp Lys Glu His Ile Ser Val Thr Thr Gly Lys Leu Glu Cys
            100                 105                 110

Lys Ile Ser Lys Thr Gly Asp Lys Ile Ile Glu Tyr Leu Lys Val Gly
        115                 120                 125

Gln Lys Ile Ile Cys Ser Asn Leu Thr Gln Ile Leu Ile Thr Glu Arg
    130                 135                 140

Ile Ser Asn Phe Thr Gly Tyr Lys Thr Lys Val Glu Glu Glu Tyr Ile
145                 150                 155                 160

Pro Gln Ile Leu Gly Ala Lys Val Glu Ser His Gly Pro Leu Arg Cys
                165                 170                 175

Val Ile Arg Val Trp Gly Arg His Leu Arg Asn Thr Lys Met Thr Phe
            180                 185                 190

Asp Gly Ser Leu Leu Lys Gln Gly Trp Leu Pro Phe Glu Leu Arg Leu
        195                 200                 205

Tyr Phe Tyr Ala Asn Ser Asp Lys Ile Lys Ile Val His Thr Phe Ile
    210                 215                 220

Tyr Asn Gly Asn Pro His Gln Asp Phe Ile Lys Gly Leu Gly Leu Lys
225                 230                 235                 240

Phe Asp Leu Pro Leu His Ser Pro Leu Tyr Asn Arg Phe Val Arg Phe
                245                 250                 255

Gly Gly Asp Ser Gly Leu Phe Cys Glu Ser Pro Lys Asn Leu Ile Asp
            260                 265                 270

Val Tyr Arg Lys Glu Arg Tyr Lys Ile Met Phe Gln Asn Gln Val Asn
        275                 280                 285

Gly Gln Pro Val Asp Phe Asp Ile Asp Glu Asp His His Tyr Ile Arg
    290                 295                 300

Glu Ile Asn Asn Thr Pro Glu Trp Asp Glu Phe Ile Leu Phe Gln Asp
305                 310                 315                 320
```

-continued

```
Ser Ser Glu His Tyr Val Ile Lys Lys Arg Thr Ser Cys Gln Cys Ser
                325                 330                 335

Phe Val Lys Ala Thr Glu Gly Arg Arg Ser Met Gly Phe Ala Tyr Ile
            340                 345                 350

Gly Asp Glu Asn Gly Gly Leu Gly Val Ala Leu Lys Asp Phe Trp Gln
        355                 360                 365

Lys Tyr Pro Ser Gly Phe Glu Ala Lys Asp Leu Thr Ser Ser Glu Ala
    370                 375                 380

Asn Leu Ile Val Trp Leu Trp Pro Pro Tyr Ala Glu Val Met Asp Leu
385                 390                 395                 400

Arg His Tyr Asp Phe Gly Thr Tyr Thr Ser Ser Ser Tyr Glu Gly Phe
                405                 410                 415

Asp Glu Tyr Arg Ser Thr Pro Tyr Gly Ile Ala Asn Thr Asn Glu Leu
            420                 425                 430

Trp Leu Tyr Cys Phe Asp Tyr Ala Pro Thr Ser Glu Gln Leu Leu Glu
        435                 440                 445

Tyr Ala Lys Met Gln Gly Phe Ser Pro Leu Leu Val Cys Ser Pro Glu
    450                 455                 460

Arg Tyr Lys Glu Thr Glu Ile Phe Gly Gly Leu Asn Leu Pro Asp Asn
465                 470                 475                 480

Ser Asn Ser Lys Lys Glu Lys Ile Glu Arg Ile Leu Thr Ala Ile Val
                485                 490                 495

Asp Phe Tyr Leu Lys Glu Val Glu Arg Arg Lys Trp Tyr Gly Phe Trp
            500                 505                 510

Asp Tyr Gly Asp Phe Met His Ser Tyr Asp Met Val Arg His Met Trp
        515                 520                 525

Lys Tyr Asp Ile Gly Gly Tyr Ala Trp Gln Asn Thr Glu Leu Val Pro
    530                 535                 540

Asn Ile Trp Leu Trp Leu Met Phe Leu Arg Thr Gly Arg Phe Asp Ile
545                 550                 555                 560

Phe Lys Met Ala Glu Ala Met Thr Arg His Thr Ser Glu Val Asp Val
                565                 570                 575

Tyr His Leu Gly Glu Tyr Lys Gly Leu Gly Ser Arg His Asn Val Val
            580                 585                 590

His Trp Gly Cys Gly Cys Lys Glu Val Arg Ile Ser Met Ala Tyr Leu
        595                 600                 605

His Arg Tyr Tyr Tyr Tyr Leu Thr Ala Asp Glu Arg Ile Gly Gln Leu
    610                 615                 620

Met Asp Asp Val Lys Asp Ile Asp Lys Gln Ile Met His Met Asp Pro
625                 630                 635                 640

Met Arg Ala Tyr Phe Thr Asn Asp Pro Glu Asn Arg Val His Ile Arg
                645                 650                 655

Val Gly Pro Asp Val Met Thr Phe Cys Ala Asn Trp Phe Val Arg Trp
            660                 665                 670

Glu Arg His Lys Glu Glu Ile Tyr Lys Gln Lys Ile Ile Lys Ile Leu
        675                 680                 685

Asp Phe Leu Lys Lys Asn Pro Ala Ala Phe Ile Ser Gly Gly Val Phe
    690                 695                 700

Asp Tyr Asp Pro Glu Lys Thr Gln Leu Lys Pro Val Glu Tyr Thr Gly
705                 710                 715                 720
```

-continued

```
Gly Ser Asn Phe Val Phe Cys Phe Gly Asn Thr Leu Val Trp Leu Glu
                725                 730                 735

Ile Ala Lys Asn Phe Glu Asp Lys Glu Phe Glu Glu Leu Cys Ala Gln
            740                 745                 750

Gln Gly Leu Phe Tyr Thr Glu Phe Lys Glu Asn Lys Asn Glu Ile Leu
        755                 760                 765

Lys Ser Trp Gly Val Pro Ser Phe Gly Phe Lys Leu Asn Met Leu Asn
    770                 775                 780

Ile Gly Met Ala Ala Phe Ala Ala Met Lys Lys Asn Ile Pro Glu Leu
785                 790                 795                 800

Lys Lys Glu Ile Trp Gln Met Phe Leu Asp His Asn Lys Asn Pro Trp
                805                 810                 815

Leu Lys Phe Ile Cys Asp Gly Gly Ile Asn Leu Gln Leu Ala Thr Val
            820                 825                 830

Pro Ala Pro Val Ala Glu Ile Pro Phe Ile Ser Thr Asn Ile Ala Ser
        835                 840                 845

Gln Trp Ser Leu Asn Ala Leu Leu Ala Leu Glu Phe Ile Gly Asp Glu
    850                 855                 860

Leu
865


<210> SEQ ID NO 138
<211> LENGTH: 1569
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 138

atgagtaaag ttaagaagat tttgaggtta tttgtagtta gtatctgtat tgcaggtttg      60

ttgttaacat caataggagt attgggtgca tcaaaatcaa agtattcgtt taaactcaca     120

atcatggctc agtattttgg gacagagcct gcgccatcaa atagtccagt aattttaaaa     180

gcagaacagt acctaaaaac tgacctggag tttacatggg tacctgccga tggttacaat     240

gacaaattaa acattatgtt agcaagtgga aatcttccaa tggtggttta tgtaccagga     300

aaaactgcgt ctataattgg tgcatgtaaa gcaggagcat tctgggaact tggaccgtat     360

ataaagcaat ataagaattt aaaagcaatc ccggatatag ttctgtggaa ctcttccatt     420

gatggtaaaa tatatggtat tccacgttca aggaccttgg gaagaaatgg tattgtttat     480

agaaaagact gggctaaaaa tgttggtatc actaagcttg aaactattga tgatttgtac     540

aacatgttaa agaagtttac ttacaatgat ccagataaaa atggaaaaaa tgacacatat     600

ggaatgattg tttgcaacta caatggacca ttctatatta ctctcacatg gtttggtggt     660

ccgaatggat ggggattgaa caaaaacgga cagttggtgc caactttctt gacaaatgca     720

tatatagaaa atttaaaatt ctggcgaaag atgtatcaag aaaagctgtt taaccatgac     780

ttcccgagtg ttccaggtgc aagatgggag gattattact cacagggtaa aggtggagta     840

aagattgacg ttatagactc tgcaaacaga atttacaatg gcctcaagaa caacggtatc     900

attccaaaag atgcaaagga tacagatata atggacgttg ttgtgtctgt gaaaggtaaa     960

tatggtctaa gaaatatgcc gacttctggt tatgcggggt atctgatggt ttcaaagaca    1020

agtgtaaaag atatgaatac attcaagaaa gtaatgtcga tattagataa gtttggcgac    1080

agaacaatgc aggatctgtt tggttatgga ctaccgaaca gacattacaa acttgtagat    1140
```

-continued

```
ggtaaaatag atcctattcc aaacttgcca gcagatttgt caagagaaat aagcggactt   1200

aaccaggttc tacattttta tcctgcaaac ggtggaacac caagatatat gactccactt   1260

ttgcagctgc aggcagatat gcaagctttg aatgaaaagc tcaatattct tgtgccaaat   1320

ccggcagaag cattgatagg tatgtcagat acttatatca aacgcggtgt cacacttgat   1380

aatatgatag aggatgcgcg cgttaaatat attactggtc agctcaatga ccagggcttt   1440

aagaaagttc ttgataactg gagaaggcag ggaggagacc agattataaa agaagtaaat   1500

gcactgtatc gtaaatataa aaagaatatt ccttacaaag aagacatata taagattctc   1560

aatccttag                                                           1569
```

<210> SEQ ID NO 139
<211> LENGTH: 522
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 139

```
Met Ser Lys Val Lys Lys Ile Leu Arg Leu Phe Val Val Ser Ile Cys
1               5                   10                  15

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

Ser Lys Tyr Ser Phe Lys Leu Thr Ile Met Ala Gln Tyr Phe Gly Thr
        35                  40                  45

Glu Pro Ala Pro Ser Asn Ser Pro Val Ile Leu Lys Ala Glu Gln Tyr
    50                  55                  60

Leu Lys Thr Asp Leu Glu Phe Thr Trp Val Pro Ala Asp Gly Tyr Asn
65                  70                  75                  80

Asp Lys Leu Asn Ile Met Leu Ala Ser Gly Asn Leu Pro Met Val Val
                85                  90                  95

Tyr Val Pro Gly Lys Thr Ala Ser Ile Ile Gly Ala Cys Lys Ala Gly
            100                 105                 110

Ala Phe Trp Glu Leu Gly Pro Tyr Ile Lys Gln Tyr Lys Asn Leu Lys
        115                 120                 125

Ala Ile Pro Asp Ile Val Leu Trp Asn Ser Ser Ile Asp Gly Lys Ile
    130                 135                 140

Tyr Gly Ile Pro Arg Ser Arg Thr Leu Gly Arg Asn Gly Ile Val Tyr
145                 150                 155                 160

Arg Lys Asp Trp Ala Lys Asn Val Gly Ile Thr Lys Leu Glu Thr Ile
                165                 170                 175

Asp Asp Leu Tyr Asn Met Leu Lys Lys Phe Thr Tyr Asn Asp Pro Asp
            180                 185                 190

Lys Asn Gly Lys Asn Asp Thr Tyr Gly Met Ile Val Cys Asn Tyr Asn
        195                 200                 205

Gly Pro Phe Tyr Ile Thr Leu Thr Trp Phe Gly Gly Pro Asn Gly Trp
    210                 215                 220

Gly Leu Asn Lys Asn Gly Gln Leu Val Pro Thr Phe Leu Thr Asn Ala
225                 230                 235                 240

Tyr Ile Glu Asn Leu Lys Phe Trp Arg Lys Met Tyr Gln Glu Lys Leu
                245                 250                 255

Phe Asn His Asp Phe Pro Ser Val Pro Gly Ala Arg Trp Glu Asp Tyr
            260                 265                 270

Tyr Ser Gln Gly Lys Gly Gly Val Lys Ile Asp Val Ile Asp Ser Ala
```

-continued

```
        275                 280                 285

Asn Arg Ile Tyr Asn Gly Leu Lys Asn Asn Gly Ile Ile Pro Lys Asp
    290                 295                 300

Ala Lys Asp Thr Asp Ile Met Asp Val Val Val Ser Val Lys Gly Lys
305                 310                 315                 320

Tyr Gly Leu Arg Asn Met Pro Thr Ser Gly Tyr Ala Gly Tyr Leu Met
                325                 330                 335

Val Ser Lys Thr Ser Val Lys Asp Met Asn Thr Phe Lys Lys Val Met
            340                 345                 350

Ser Ile Leu Asp Lys Phe Gly Asp Arg Thr Met Gln Asp Leu Phe Gly
        355                 360                 365

Tyr Gly Leu Pro Asn Arg His Tyr Lys Leu Val Asp Gly Lys Ile Asp
    370                 375                 380

Pro Ile Pro Asn Leu Pro Ala Asp Leu Ser Arg Glu Ile Ser Gly Leu
385                 390                 395                 400

Asn Gln Val Leu His Phe Tyr Pro Ala Asn Gly Gly Thr Pro Arg Tyr
                405                 410                 415

Met Thr Pro Leu Leu Gln Leu Gln Ala Asp Met Gln Ala Leu Asn Glu
            420                 425                 430

Lys Leu Asn Ile Leu Val Pro Asn Pro Ala Glu Ala Leu Ile Gly Met
        435                 440                 445

Ser Asp Thr Tyr Ile Lys Arg Gly Val Thr Leu Asp Asn Met Ile Glu
    450                 455                 460

Asp Ala Arg Val Lys Tyr Ile Thr Gly Gln Leu Asn Asp Gln Gly Phe
465                 470                 475                 480

Lys Lys Val Leu Asp Asn Trp Arg Arg Gln Gly Gly Asp Gln Ile Ile
                485                 490                 495

Lys Glu Val Asn Ala Leu Tyr Arg Lys Tyr Lys Lys Asn Ile Pro Tyr
            500                 505                 510

Lys Glu Asp Ile Tyr Lys Ile Leu Asn Pro
        515                 520


<210> SEQ ID NO 140
<211> LENGTH: 879
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 140

atgagacaga ataagacagt agcaagtact atttttgatg tttttaatca tatattcctt     60

ggaatatggg cgataataac tgttttacct ttcctatacg ttttggctgc atcttttgct    120

ccagattcag aaataaaaac gaggacattc tttataattc ctcataatcc tactcttctc    180

acctataaat ttatttttgc ttcaaactat tttcttcgta gtatgctcaa cagtgtgatt    240

atcacggtag gaggtacctt ggtaaatctg tttttcacat ttacaatggc atatgctctt    300

tcaaaaaagc actttatagg cagaagtatt gtcttaaatg gcataatatt tacaatgctt    360

tttggtggag gaatgattcc tacatatttg ctggtaaaaa gcctgggtct cttaaattct    420

tattgggcac tatggttgcc aggtgcaatt agtcccttta acttctttgt tgtcaaaaac    480

tttttccagg aaatgccaca ggatttagag gatgcagcaa gaattgatgg atgtacagaa    540

gctcaagtgt tgtggaagat aatacttcct ctttcaaagc ctattattgc gacatttgct    600

ttgttctatg gagtggggca ttggaattca tggtttggag cgcttttgta tataaacgat    660
```

-continued

```
gctgagaaat ggcctgtgca gctaatttta aggcaaatag taatgttgtc aaccacgctt    720

gcttcggacc taacacaatt tgatcctaat ttccaacctc cacaggagtc gctgaagatg    780

gcaataattg tcgtagcaac tttaccaata atgcttttgt atccatggct tcagaagtac    840

ttcataaaag gtatgtttat tggttcattg aaggagtga                           879


<210> SEQ ID NO 141
<211> LENGTH: 292
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 141

Met Arg Gln Asn Lys Thr Val Ala Ser Thr Ile Phe Asp Val Phe Asn
1               5                   10                  15

His Ile Phe Leu Gly Ile Trp Ala Ile Ile Thr Val Leu Pro Phe Leu
            20                  25                  30

Tyr Val Leu Ala Ala Ser Phe Ala Pro Asp Ser Glu Ile Lys Thr Arg
        35                  40                  45

Thr Phe Phe Ile Ile Pro His Asn Pro Thr Leu Leu Thr Tyr Lys Phe
    50                  55                  60

Ile Phe Ala Ser Asn Tyr Phe Leu Arg Ser Met Leu Asn Ser Val Ile
65                  70                  75                  80

Ile Thr Val Gly Gly Thr Leu Val Asn Leu Phe Phe Thr Phe Thr Met
                85                  90                  95

Ala Tyr Ala Leu Ser Lys Lys His Phe Ile Gly Arg Ser Ile Val Leu
            100                 105                 110

Asn Gly Ile Ile Phe Thr Met Leu Phe Gly Gly Gly Met Ile Pro Thr
        115                 120                 125

Tyr Leu Leu Val Lys Ser Leu Gly Leu Leu Asn Ser Tyr Trp Ala Leu
    130                 135                 140

Trp Leu Pro Gly Ala Ile Ser Pro Phe Asn Phe Phe Val Val Lys Asn
145                 150                 155                 160

Phe Phe Gln Glu Met Pro Gln Asp Leu Glu Asp Ala Ala Arg Ile Asp
                165                 170                 175

Gly Cys Thr Glu Ala Gln Val Leu Trp Lys Ile Ile Leu Pro Leu Ser
            180                 185                 190

Lys Pro Ile Ile Ala Thr Phe Ala Leu Phe Tyr Gly Val Gly His Trp
        195                 200                 205

Asn Ser Trp Phe Gly Ala Leu Leu Tyr Ile Asn Asp Ala Glu Lys Trp
    210                 215                 220

Pro Val Gln Leu Ile Leu Arg Gln Ile Val Met Leu Ser Thr Thr Leu
225                 230                 235                 240

Ala Ser Asp Leu Thr Gln Phe Asp Pro Asn Phe Gln Pro Pro Gln Glu
                245                 250                 255

Ser Leu Lys Met Ala Ile Ile Val Val Ala Thr Leu Pro Ile Met Leu
            260                 265                 270

Leu Tyr Pro Trp Leu Gln Lys Tyr Phe Ile Lys Gly Met Phe Ile Gly
        275                 280                 285

Ser Leu Lys Glu
    290


<210> SEQ ID NO 142
```

-continued

<211> LENGTH: 966
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 142

atgaaaaaca atattatcaa gggaccaaaa aagaagggag agaatgattt aatgacagct 60

accacttcta tatggaaaag actgaaaaaa gacaaatggc tgtatatctt agctctaccg 120

ggtattttgt acttcatcat ttttaggtac attccaatgt ttggtatagt agttgccttt 180

caagatttta atccgttttt agggttctgg aaaagtccgt gggtaggttt tgagcatttt 240

aaaacacttt tcaccgaccc tgattttccg atgctgttta gaaatacact gttaatttca 300

ttttacaata tacttttcta tttccctgtg ccaattattt tagcgctact gatcaacgag 360

gtgcgaaatc aggtttataa gagaattgtc cagacgtgtg tttatgttcc tcactttgtt 420

tcaatggtgg tcatagcaag tattacatac gtgcttctct caagtgaggc aggggttata 480

aataatattc tgtatagctt aacaggtaaa aaaattgaat ttttgacaga tccaagatgg 540

ttcagacctc ttataataat tcagagcata tggaaagaag caggatgggg aacaataatc 600

ttcttggcag cactatcaaa tgttgaccca acattatatg aggctgctat tgtggatggg 660

gcaacgaggt ggcagcagac gtggcatatc acaataccct caattatgag tacggtaatc 720

atacttttca ttttgcgatt gggacatctt cttgatactg gctttgaaca gatatttttg 780

atgaaaaatc ctattaacag gtcggtagca gaagtgtttg acacatatgt ttatcaagta 840

ggggttaccc aaggagcgta cagttacagt acagcggttg gtctttttaa gtctgttgtt 900

ggactgattt tgattcaggt ttctaactat ctgtccaaga aatttactga aacttcattg 960

ttctaa 966

<210> SEQ ID NO 143
<211> LENGTH: 321
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 143

Met Lys Asn Asn Ile Ile Lys Gly Pro Lys Lys Lys Gly Glu Asn Asp
1 5 10 15

Leu Met Thr Ala Thr Thr Ser Ile Trp Lys Arg Leu Lys Lys Asp Lys
20 25 30

Trp Leu Tyr Ile Leu Ala Leu Pro Gly Ile Leu Tyr Phe Ile Ile Phe
35 40 45

Arg Tyr Ile Pro Met Phe Gly Ile Val Val Ala Phe Gln Asp Phe Asn
50 55 60

Pro Phe Leu Gly Phe Trp Lys Ser Pro Trp Val Gly Phe Glu His Phe
65 70 75 80

Lys Thr Leu Phe Thr Asp Pro Asp Phe Pro Met Leu Phe Arg Asn Thr
85 90 95

Leu Leu Ile Ser Phe Tyr Asn Ile Leu Phe Tyr Phe Pro Val Pro Ile
100 105 110

Ile Leu Ala Leu Leu Ile Asn Glu Val Arg Asn Gln Val Tyr Lys Arg
115 120 125

Ile Val Gln Thr Cys Val Tyr Val Pro His Phe Val Ser Met Val Val
130 135 140

Ile Ala Ser Ile Thr Tyr Val Leu Leu Ser Ser Glu Ala Gly Val Ile

-continued

```
145                 150                 155                 160

Asn Asn Ile Leu Tyr Ser Leu Thr Gly Lys Lys Ile Glu Phe Leu Thr
                165                 170                 175

Asp Pro Arg Trp Phe Arg Pro Leu Ile Ile Ile Gln Ser Ile Trp Lys
            180                 185                 190

Glu Ala Gly Trp Gly Thr Ile Ile Phe Leu Ala Ala Leu Ser Asn Val
        195                 200                 205

Asp Pro Thr Leu Tyr Glu Ala Ala Ile Val Asp Gly Ala Thr Arg Trp
    210                 215                 220

Gln Gln Thr Trp His Ile Thr Ile Pro Ser Ile Met Ser Thr Val Ile
225                 230                 235                 240

Ile Leu Phe Ile Leu Arg Leu Gly His Leu Leu Asp Thr Gly Phe Glu
                245                 250                 255

Gln Ile Phe Leu Met Lys Asn Pro Ile Asn Arg Ser Val Ala Glu Val
            260                 265                 270

Phe Asp Thr Tyr Val Tyr Gln Val Gly Val Thr Gln Gly Ala Tyr Ser
        275                 280                 285

Tyr Ser Thr Ala Val Gly Leu Phe Lys Ser Val Val Gly Leu Ile Leu
    290                 295                 300

Ile Gln Val Ser Asn Tyr Leu Ser Lys Lys Phe Thr Glu Thr Ser Leu
305                 310                 315                 320

Phe


<210> SEQ ID NO 144
<211> LENGTH: 2535
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 144

atgaatgaga gtataaaatt gaattctctt tttgatgaaa tcaacaaccc tccagctact      60

gccagtatta ttcattggtg gatttttct gatgagatga acgagaacag aataaatgct     120

gagcttgatt atatttcaaa tcttggcttt aagcaagtat taattgcagt aggacacaat     180

gtttcgccta aatatttgac acatggttgg tttgaaatgg taaaatttgc agttctccaa     240

gctaaaaaaa gaggagttaa agtgtggatt gccgatgaag ggacatatcc aagtggcttt     300

gctggcgaaa cttttaataa gaagtatcct cacaaaagga tgaaggctat tattgttgag     360

aaggagttta ttattgaggg taatttatgt gaagttgaac ctcactctgg tacaattggg     420

attttggcca aagacatgaa ccagaataaa tactttgctt ttgaaaagct tgaatttagt     480

agcggatttt tatacttgcc ctatcattcg acttggcaaa taaaagtaat atcttcagct     540

tacaggacat ctccaacaag atacgttcac catccaacag gtgcaaaaga tactacattt     600

tcactttgtg attatcttga ctatgaagct gtcaatctat tcataagtga ggtatatgaa     660

aaatataaag cttatatggg aaatgaattt ggaaagacaa taattggatt tttcgctgac     720

gaacctgatt attctatttc tggactacca tatacggata atatatttga tatattttac     780

aatgaacaag gatacgacgt taaaaagtac ataccgtatt tctttaaaga gcaattagat     840

gaaaaaataa aaagagtaaa agcagattac tgggatgtat ggagcaatat ttttacaaat     900

actttcttta agcagatcta caaatggtgt gaggcaaatg gcctcaaatt tgtagtacat     960

ctaaatcatg aagatatgat agaacacctt accaaatctg aaggacagtt cttttcgcat    1020
```

-continued

```
atgaagtatg ttcatattcc agcaattgat gtaatttgga gacaaatctg gtatgacaaa   1080

aaagcaatat tccctaaata cgcttcttct gtttctcata ttaaaaatat tgctcagacc   1140

ttttcagaga gttttgcagt atatggacaa ggtatatctg ttgagcaaat caaatgggta   1200

gttgattacc agtttgcaat ggacataaat ctatttttga cctcaatctt caagtatctt   1260

tatgaccatc cgcaaaatta tttctttcca gaggtaatta agtatattaa taccatttca   1320

tatcttctct atgtaagcac cccttgtaca aaggttctgg tttactttcc tacaccggat   1380

ctgtgggcag gtgaaaatat gtctgcttca aaagcaatgg aaattggcaa tgcactttta   1440

gagaaccaga ttgattttga tttttttgac cattctcttt tagaatatct ggaaattaaa   1500

aaccatagaa tatacgctaa caatagaaaa gaatacgaca ttgttattct tccgcctata   1560

aagtatttgc cacaagatct gttcagattt ttaaagcttt tctcaagcaa aggagggaag   1620

attattttct tcgagaactc tcctttgttt gtttataaca aaacctttac atcgtttttc   1680

cactttgtag atagagaaat aggtgtggtt gttgaaagta tcgagcagct ttcaaaaatg   1740

gttgaaaaag atgtcactgt tgtagacagc aaagatgtta gagttcttca taaaagaata   1800

gaaggcaata atctgatttt tctcttcaat gtttcaggta cttcattttt gggtaagata   1860

atattaaaat tttctaagaa aaatgtatat atatgggatc atatacagaa taaattttta   1920

atggtttcaa atatcaaaag taataaaaaa aacatacaat tagaactcta tatacatcca   1980

tatcagactt tggttttaat agcaagtgat gagtatgtag atggaattca aaaaacaaca   2040

ctgcttggaa gcttaccgag aacagtcttg gaattaaacg ataactggga aattcatttt   2100

gataaagatt ttgttttgtt ttcagattta aaagattggc aaagcttggg ctttggtgac   2160

tattctggca gtgtagttta tagaaaaata ttttcgtttt ctcatgatga ctttattaaa   2220

aataaacatc ttttcctcaa ctgccccaat gtaaagtact ctgcaaaggt ttggttaaat   2280

aaaagatatc ttggtgtaag agctttttcg ccttttatgt gggatataac agaggcattg   2340

aaaattggtg agaatgaact tgtgattgaa gttcaaaaca cccctgcagc agctctactt   2400

ggaacacaag aaaaattgga aaaattaaga aaagaggcag agaagaactt ttatctttct   2460

atttctctaa aatttgacct ggaaatggtc caatcaggat tgttgcctcc agttgctatt   2520

gtttctttag aatga                                                    2535
```

<210> SEQ ID NO 145
<211> LENGTH: 844
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 145

```
Met Asn Glu Ser Ile Lys Leu Asn Ser Leu Phe Asp Glu Ile Asn Asn
1               5                   10                  15

Pro Pro Ala Thr Ala Ser Ile Ile His Trp Trp Ile Phe Ser Asp Glu
            20                  25                  30

Met Asn Glu Asn Arg Ile Asn Ala Glu Leu Asp Tyr Ile Ser Asn Leu
        35                  40                  45

Gly Phe Lys Gln Val Leu Ile Ala Val Gly His Asn Val Ser Pro Lys
    50                  55                  60

Tyr Leu Thr His Gly Trp Phe Glu Met Val Lys Phe Ala Val Leu Gln
65                  70                  75                  80

Ala Lys Lys Arg Gly Val Lys Val Trp Ile Ala Asp Glu Gly Thr Tyr
```

-continued

```
                85                  90                  95

Pro Ser Gly Phe Ala Gly Glu Thr Phe Asn Lys Lys Tyr Pro His Lys
            100                 105                 110

Arg Met Lys Ala Ile Ile Val Glu Lys Glu Phe Ile Ile Glu Gly Asn
        115                 120                 125

Leu Cys Glu Val Glu Pro His Ser Gly Thr Ile Gly Ile Leu Ala Lys
    130                 135                 140

Asp Met Asn Gln Asn Lys Tyr Phe Ala Phe Glu Lys Leu Glu Phe Ser
145                 150                 155                 160

Ser Gly Phe Leu Tyr Leu Pro Tyr His Ser Thr Trp Gln Ile Lys Val
                165                 170                 175

Ile Ser Ser Ala Tyr Arg Thr Ser Pro Thr Arg Tyr Val His His Pro
            180                 185                 190

Thr Gly Ala Lys Asp Thr Thr Phe Ser Leu Cys Asp Tyr Leu Asp Tyr
        195                 200                 205

Glu Ala Val Asn Leu Phe Ile Ser Glu Val Tyr Glu Lys Tyr Lys Ala
    210                 215                 220

Tyr Met Gly Asn Glu Phe Gly Lys Thr Ile Ile Gly Phe Phe Ala Asp
225                 230                 235                 240

Glu Pro Asp Tyr Ser Ile Ser Gly Leu Pro Tyr Thr Asp Asn Ile Phe
                245                 250                 255

Asp Ile Phe Tyr Asn Glu Gln Gly Tyr Asp Val Lys Lys Tyr Ile Pro
            260                 265                 270

Tyr Phe Phe Lys Glu Gln Leu Asp Glu Lys Ile Lys Arg Val Lys Ala
        275                 280                 285

Asp Tyr Trp Asp Val Trp Ser Asn Ile Phe Thr Asn Thr Phe Phe Lys
    290                 295                 300

Gln Ile Tyr Lys Trp Cys Glu Ala Asn Gly Leu Lys Phe Val Val His
305                 310                 315                 320

Leu Asn His Glu Asp Met Ile Glu His Leu Thr Lys Ser Glu Gly Gln
                325                 330                 335

Phe Phe Ser His Met Lys Tyr Val His Ile Pro Ala Ile Asp Val Ile
            340                 345                 350

Trp Arg Gln Ile Trp Tyr Asp Lys Lys Ala Ile Phe Pro Lys Tyr Ala
        355                 360                 365

Ser Ser Val Ser His Ile Lys Asn Ile Ala Gln Thr Phe Ser Glu Ser
    370                 375                 380

Phe Ala Val Tyr Gly Gln Gly Ile Ser Val Glu Gln Ile Lys Trp Val
385                 390                 395                 400

Val Asp Tyr Gln Phe Ala Met Asp Ile Asn Leu Phe Leu Thr Ser Ile
                405                 410                 415

Phe Lys Tyr Leu Tyr Asp His Pro Gln Asn Tyr Phe Phe Pro Glu Val
            420                 425                 430

Ile Lys Tyr Ile Asn Thr Ile Ser Tyr Leu Leu Tyr Val Ser Thr Pro
        435                 440                 445

Cys Thr Lys Val Leu Val Tyr Phe Pro Thr Pro Asp Leu Trp Ala Gly
    450                 455                 460

Glu Asn Met Ser Ala Ser Lys Ala Met Glu Ile Gly Asn Ala Leu Leu
465                 470                 475                 480

Glu Asn Gln Ile Asp Phe Asp Phe Phe Asp His Ser Leu Leu Glu Tyr
                485                 490                 495
```

-continued

```
Leu Glu Ile Lys Asn His Arg Ile Tyr Ala Asn Asn Arg Lys Glu Tyr
            500                 505                 510

Asp Ile Val Ile Leu Pro Pro Ile Lys Tyr Leu Pro Gln Asp Leu Phe
        515                 520                 525

Arg Phe Leu Lys Leu Phe Ser Ser Lys Gly Gly Lys Ile Ile Phe Phe
    530                 535                 540

Glu Asn Ser Pro Leu Phe Val Tyr Asn Lys Thr Phe Thr Ser Phe Phe
545                 550                 555                 560

His Phe Val Asp Arg Glu Ile Gly Val Val Val Glu Ser Ile Glu Gln
                565                 570                 575

Leu Ser Lys Met Val Glu Lys Asp Val Thr Val Val Asp Ser Lys Asp
            580                 585                 590

Val Arg Val Leu His Lys Arg Ile Glu Gly Asn Asn Leu Ile Phe Leu
        595                 600                 605

Phe Asn Val Ser Gly Thr Ser Phe Leu Gly Lys Ile Ile Leu Lys Phe
    610                 615                 620

Ser Lys Lys Asn Val Tyr Ile Trp Asp His Ile Gln Asn Lys Phe Leu
625                 630                 635                 640

Met Val Ser Asn Ile Lys Ser Asn Lys Lys Asn Ile Gln Leu Glu Leu
                645                 650                 655

Tyr Ile His Pro Tyr Gln Thr Leu Val Leu Ile Ala Ser Asp Glu Tyr
            660                 665                 670

Val Asp Gly Ile Gln Lys Thr Thr Leu Leu Gly Ser Leu Pro Arg Thr
        675                 680                 685

Val Leu Glu Leu Asn Asp Asn Trp Glu Ile His Phe Asp Lys Asp Phe
    690                 695                 700

Val Leu Phe Ser Asp Leu Lys Asp Trp Gln Ser Leu Gly Phe Gly Asp
705                 710                 715                 720

Tyr Ser Gly Ser Val Val Tyr Arg Lys Ile Phe Ser Phe Ser His Asp
                725                 730                 735

Asp Phe Ile Lys Asn Lys His Leu Phe Leu Asn Cys Pro Asn Val Lys
            740                 745                 750

Tyr Ser Ala Lys Val Trp Leu Asn Lys Arg Tyr Leu Gly Val Arg Ala
        755                 760                 765

Phe Ser Pro Phe Met Trp Asp Ile Thr Glu Ala Leu Lys Ile Gly Glu
    770                 775                 780

Asn Glu Leu Val Ile Glu Val Gln Asn Thr Pro Ala Ala Ala Leu Leu
785                 790                 795                 800

Gly Thr Gln Glu Lys Leu Glu Lys Leu Arg Lys Glu Ala Glu Lys Asn
                805                 810                 815

Phe Tyr Leu Ser Ile Ser Leu Lys Phe Asp Leu Glu Met Val Gln Ser
            820                 825                 830

Gly Leu Leu Pro Pro Val Ala Ile Val Ser Leu Glu
        835                 840


<210> SEQ ID NO 146
<211> LENGTH: 882
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 146

atgctaaaaa gacaggatat ttatatccgt gaccctttta ttatgcctgt aaaagaaaaa     60
```

-continued

```
agggtgtact atatgtttgg tacaactgat aaaaattgct ggggaaatga aaaagcaaca    120

ggctttgatt attatgtaac ctatgacctt gaaaattttg atgggccata tccagcattc    180

aggcctcctc agaatttttg ggcagacaga aatttctggg cgccagaggt gcacttttat    240

aacggcaaat attatatgtt tgcaacattt tgttcaccag aaggaaaaag gggaactgcg    300

attcttgttt cagatactgt agaaggacca tatgttgaac atagtcttgg acctgttaca    360

ccaaaggatt ggatgtgcct tgatggaaca ctttacatag atgaagattc aaatccctgg    420

cttgtatttt gccgtgaatg ggttgaggtt tatgatggtc agattttggc agcaagactt    480

tcaaatgatt tgaaaagagt tattgatgag cccgttttac tctttacagc ctccagcgca    540

ccttggacag caattataag agttaaagat ggcagagaat gttttgtgac agatggtccg    600

tttttattca gaacccaaaa cggtactctt ttgatgctat ggtcaagttt tgatagagaa    660

agaagatatt gtgttggagt tgcaaaatca ctgtcaggca atattttagg tccatggcag    720

cacaaaacag agccaatttt ttctaatgac ggtggtcatg gcatgctttt tagaaccttt    780

gaaggtgacc ttatgctttc aattcatact ccaaattcaa ggggcaatga gaggcctttg    840

tttataaaaa ttgatgataa aaatttagaa gacgagtttt ga                       882
```

```
<210> SEQ ID NO 147
<211> LENGTH: 293
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 147

Met Leu Lys Arg Gln Asp Ile Tyr Ile Arg Asp Pro Phe Ile Met Pro
1               5                   10                  15

Val Lys Glu Lys Arg Val Tyr Tyr Met Phe Gly Thr Thr Asp Lys Asn
            20                  25                  30

Cys Trp Gly Asn Glu Lys Ala Thr Gly Phe Asp Tyr Tyr Val Thr Tyr
        35                  40                  45

Asp Leu Glu Asn Phe Asp Gly Pro Tyr Pro Ala Phe Arg Pro Pro Gln
    50                  55                  60

Asn Phe Trp Ala Asp Arg Asn Phe Trp Ala Pro Glu Val His Phe Tyr
65                  70                  75                  80

Asn Gly Lys Tyr Tyr Met Phe Ala Thr Phe Cys Ser Pro Glu Gly Lys
                85                  90                  95

Arg Gly Thr Ala Ile Leu Val Ser Asp Thr Val Glu Gly Pro Tyr Val
            100                 105                 110

Glu His Ser Leu Gly Pro Val Thr Pro Lys Asp Trp Met Cys Leu Asp
        115                 120                 125

Gly Thr Leu Tyr Ile Asp Glu Asp Ser Asn Pro Trp Leu Val Phe Cys
    130                 135                 140

Arg Glu Trp Val Glu Val Tyr Asp Gly Gln Ile Leu Ala Ala Arg Leu
145                 150                 155                 160

Ser Asn Asp Leu Lys Arg Val Ile Asp Glu Pro Val Leu Leu Phe Thr
                165                 170                 175

Ala Ser Ser Ala Pro Trp Thr Ala Ile Ile Arg Val Lys Asp Gly Arg
            180                 185                 190

Glu Cys Phe Val Thr Asp Gly Pro Phe Leu Phe Arg Thr Gln Asn Gly
        195                 200                 205
```

-continued

```
Thr Leu Leu Met Leu Trp Ser Ser Phe Asp Arg Glu Arg Arg Tyr Cys
    210                 215                 220

Val Gly Val Ala Lys Ser Leu Ser Gly Asn Ile Leu Gly Pro Trp Gln
225                 230                 235                 240

His Lys Thr Glu Pro Ile Phe Ser Asn Asp Gly Gly His Gly Met Leu
                245                 250                 255

Phe Arg Thr Phe Glu Gly Asp Leu Met Leu Ser Ile His Thr Pro Asn
            260                 265                 270

Ser Arg Gly Asn Glu Arg Pro Leu Phe Ile Lys Ile Asp Asp Lys Asn
        275                 280                 285

Leu Glu Asp Glu Phe
    290


<210> SEQ ID NO 148
<211> LENGTH: 1788
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 148

atgataagaa aaaatggagt ttataaaagc aagattttca aaaagaattt tgtgcagata     60

gttattgttc ccattgtgat aataactatt cttgggcttt tctcatgcat cattatagaa    120

caatacgtta aaaatgaaat aaacaaaaat ttagagacaa tgctaataca aagcaaaaac    180

aatgtcgagc ttatgctcgg tgagatagac tatctttata tggtatttgg gataaacaaa    240

gatgtgaccc ttcagataaa gaggattttg aactcaatgt atttttcttt agaagatatc    300

tggcagatta acatggtcaa aaatgtttta aattcaatct catattcaaa gccgttcata    360

cattccatct atgtttattt cgaaaatcct gaagggaatt ttatagttac cccagatgga    420

atgactaatt ttcagtattt ttatgacaaa tggtggtttg atcagtataa agaaaataaa    480

gcattaatgt gggtagagag aagaaaaatt caaccttaca attttactgg agaatcaatt    540

gatgttttga ccatctacaa aaggataaaa tctgcatatt ctgatgtgaa tgagggtgtt    600

attgttctta atctgtatta cgaccaggta aaaaagctct taagccttaa aagttcgctc    660

cctcagcatg caatgtacat attagatcaa aatggaaatg ttttggtatc aaatgaatca    720

gataactcta atacctcaag tatggccctc ctaaaaaaag aaacagacaa ctatctcaca    780

aaaagattag agtcaaaaaa atacaactta acctttgttt cagtaattcc caaaaattat    840

ctttacagca tccctatcag acttttcaag gtgacactgg tgctactttt aatttttata    900

gttatcgctt ttgctgcctc atactacatt gccaaagtaa attacaggaa tattaaaaag    960

attatagata caataaattc agcaacagaa ggaaaaccac caaaagaaat taaaattact   1020

tcaaatgatg aatatgggta tatcatgtac aatgtaatca agaactttat tgaaaaacat   1080

tatctgacaa cacgccttca agctttggag cttttagcat tgcaggctca gattaaccct   1140

cattttcttt tcaatacttt agaacacata tatcttaaaa ctttagcact tacaggcacc   1200

ccaaacgaga ttacaaaaat gatagaaaac ctttcggcta tactcaaata ttctctgagc   1260

aatccaaaaa ttactatctt cctaagggaa gaaattaaag ctacacaggc atatattgag   1320

cttgtaaaag caagatataa agataagttt gatgtgtttt gggactatag tgaagatgtg   1380

cttgagataa aagtgatgaa gcttttattc cagccgctca tagaaaattc aatctatcat   1440

gggataaaac cttgcgaaaa gagatgtgga ataaaaatca ggataagaaa attaaaagat   1500
```

-continued

```
accagtgatt ggctttgtat atgggtaatt gacaatggaa ttgggatgag caaagaaaag   1560

ttagaggagg tacaaggcag gctttcacag gattttgact tttcagatca tattgggctt   1620

ttaaacacca atgaaaggtt aaagctcaac tatgggggta actttaaact caaggtttgg   1680

agcaagctgg gtttggggac aattgtaaaa ataattcttc ctgtgaattt tgaggaccga   1740

aaggagaatg aaatagatgc taaaaagaca ggatatttat atccgtga                1788


<210> SEQ ID NO 149
<211> LENGTH: 595
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 149

Met Ile Arg Lys Asn Gly Val Tyr Lys Ser Lys Ile Phe Lys Lys Asn
1               5                   10                  15

Phe Val Gln Ile Val Ile Val Pro Ile Val Ile Ile Thr Ile Leu Gly
            20                  25                  30

Leu Phe Ser Cys Ile Ile Ile Glu Gln Tyr Val Lys Asn Glu Ile Asn
        35                  40                  45

Lys Asn Leu Glu Thr Met Leu Ile Gln Ser Lys Asn Asn Val Glu Leu
    50                  55                  60

Met Leu Gly Glu Ile Asp Tyr Leu Tyr Met Val Phe Gly Ile Asn Lys
65                  70                  75                  80

Asp Val Thr Leu Gln Ile Lys Arg Ile Leu Asn Ser Met Tyr Phe Ser
                85                  90                  95

Leu Glu Asp Ile Trp Gln Ile Asn Met Val Lys Asn Val Leu Asn Ser
            100                 105                 110

Ile Ser Tyr Ser Lys Pro Phe Ile His Ser Ile Tyr Val Tyr Phe Glu
        115                 120                 125

Asn Pro Glu Gly Asn Phe Ile Val Thr Pro Asp Gly Met Thr Asn Phe
    130                 135                 140

Gln Tyr Phe Tyr Asp Lys Trp Trp Phe Asp Gln Tyr Lys Glu Asn Lys
145                 150                 155                 160

Ala Leu Met Trp Val Glu Arg Arg Lys Ile Gln Pro Tyr Asn Phe Thr
                165                 170                 175

Gly Glu Ser Ile Asp Val Leu Thr Ile Tyr Lys Arg Ile Lys Ser Ala
            180                 185                 190

Tyr Ser Asp Val Asn Glu Gly Val Ile Val Leu Asn Leu Tyr Tyr Asp
        195                 200                 205

Gln Val Lys Lys Leu Leu Ser Leu Lys Ser Ser Leu Pro Gln His Ala
    210                 215                 220

Met Tyr Ile Leu Asp Gln Asn Gly Asn Val Leu Val Ser Asn Glu Ser
225                 230                 235                 240

Asp Asn Ser Asn Thr Ser Ser Met Ala Leu Leu Lys Lys Glu Thr Asp
                245                 250                 255

Asn Tyr Leu Thr Lys Arg Leu Glu Ser Lys Lys Tyr Asn Leu Thr Phe
            260                 265                 270

Val Ser Val Ile Pro Lys Asn Tyr Leu Tyr Ser Ile Pro Ile Arg Leu
        275                 280                 285

Phe Lys Val Thr Leu Val Leu Leu Leu Ile Phe Ile Val Ile Ala Phe
    290                 295                 300

Ala Ala Ser Tyr Tyr Ile Ala Lys Val Asn Tyr Arg Asn Ile Lys Lys
```

-continued

```
305                 310                 315                 320

Ile Ile Asp Thr Ile Asn Ser Ala Thr Glu Gly Lys Pro Pro Lys Glu
                325                 330                 335

Ile Lys Ile Thr Ser Asn Asp Glu Tyr Gly Tyr Ile Met Tyr Asn Val
            340                 345                 350

Ile Lys Asn Phe Ile Glu Lys His Tyr Leu Thr Thr Arg Leu Gln Ala
        355                 360                 365

Leu Glu Leu Leu Ala Leu Gln Ala Gln Ile Asn Pro His Phe Leu Phe
    370                 375                 380

Asn Thr Leu Glu His Ile Tyr Leu Lys Thr Leu Ala Leu Thr Gly Thr
385                 390                 395                 400

Pro Asn Glu Ile Thr Lys Met Ile Glu Asn Leu Ser Ala Ile Leu Lys
                405                 410                 415

Tyr Ser Leu Ser Asn Pro Lys Ile Thr Ile Phe Leu Arg Glu Glu Ile
            420                 425                 430

Lys Ala Thr Gln Ala Tyr Ile Glu Leu Val Lys Ala Arg Tyr Lys Asp
        435                 440                 445

Lys Phe Asp Val Phe Trp Asp Tyr Ser Glu Asp Val Leu Glu Ile Lys
    450                 455                 460

Val Met Lys Leu Leu Phe Gln Pro Leu Ile Glu Asn Ser Ile Tyr His
465                 470                 475                 480

Gly Ile Lys Pro Cys Glu Lys Arg Cys Gly Ile Lys Ile Arg Ile Arg
                485                 490                 495

Lys Leu Lys Asp Thr Ser Asp Trp Leu Cys Ile Trp Val Ile Asp Asn
            500                 505                 510

Gly Ile Gly Met Ser Lys Glu Lys Leu Glu Glu Val Gln Gly Arg Leu
        515                 520                 525

Ser Gln Asp Phe Asp Phe Ser Asp His Ile Gly Leu Leu Asn Thr Asn
    530                 535                 540

Glu Arg Leu Lys Leu Asn Tyr Gly Gly Asn Phe Lys Leu Lys Val Trp
545                 550                 555                 560

Ser Lys Leu Gly Leu Gly Thr Ile Val Lys Ile Ile Leu Pro Val Asn
                565                 570                 575

Phe Glu Asp Arg Lys Glu Asn Glu Ile Asp Ala Lys Lys Thr Gly Tyr
            580                 585                 590

Leu Tyr Pro
        595


<210> SEQ ID NO 150
<211> LENGTH: 759
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 150

atgtcttaca agatgattat cgtggaggat gaaagtgaga taagacaggg gctttttacc      60

tgctttccat ggaataaact tggttttgaa gttgttggcc tgtttgaaaa tggaaagcag     120

gctttagatt atattttgca aaataaggtt gatgttgttt tttgcgatat taaaatgcct     180

gttatgagtg gaattgacct tgcaagggtt ttgtttttaa ataacatccc tgtaaaagtt     240

gtttttataa gtggttatca ggactttgaa tttgctcaaa aagcaatgaa atacggggta     300

aggtactaca taacaaaacc agcaacgtat gatgaagtaa tagagatatt tgggctaata     360
```

-continued

```
aaaaaagagc ttgacctaca gtcaaaaaat ccgcaaactc acagtgaaga ggtaaaaaaa    420

caagatttaa tttctattga caatcctgaa aatgtgattg aaactgtgaa agagtatata    480

aaaagaaact ataaagaggc cacgcttgag gatgctgcaa agcttgtcta tatgaatcct    540

tattatctta gccggctttt caaatgcaaa actggtaaaa acttttctga ttatctgatg    600

gaagtcagga tgagaaaagc tctggaactt atgaaaattc ctatatataa acttcatgag    660

ataggtgaaa tggttggtta caagaaccct aagaacttta caagagcttt taaaaaatat    720

ttcggcaagt ctccttcaga gtttgttcag gagagataa                           759
```

<210> SEQ ID NO 151
<211> LENGTH: 252
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 151

```
Met Ser Tyr Lys Met Ile Ile Val Glu Asp Glu Ser Glu Ile Arg Gln
1               5                   10                  15

Gly Leu Phe Thr Cys Phe Pro Trp Asn Lys Leu Gly Phe Glu Val Val
            20                  25                  30

Gly Leu Phe Glu Asn Gly Lys Gln Ala Leu Asp Tyr Ile Leu Gln Asn
        35                  40                  45

Lys Val Asp Val Val Phe Cys Asp Ile Lys Met Pro Val Met Ser Gly
    50                  55                  60

Ile Asp Leu Ala Arg Val Leu Phe Leu Asn Asn Ile Pro Val Lys Val
65                  70                  75                  80

Val Phe Ile Ser Gly Tyr Gln Asp Phe Glu Phe Ala Gln Lys Ala Met
                85                  90                  95

Lys Tyr Gly Val Arg Tyr Tyr Ile Thr Lys Pro Ala Thr Tyr Asp Glu
            100                 105                 110

Val Ile Glu Ile Phe Gly Leu Ile Lys Lys Glu Leu Asp Leu Gln Ser
        115                 120                 125

Lys Asn Pro Gln Thr His Ser Glu Glu Val Lys Lys Gln Asp Leu Ile
    130                 135                 140

Ser Ile Asp Asn Pro Glu Asn Val Ile Glu Thr Val Lys Glu Tyr Ile
145                 150                 155                 160

Lys Arg Asn Tyr Lys Glu Ala Thr Leu Glu Asp Ala Ala Lys Leu Val
                165                 170                 175

Tyr Met Asn Pro Tyr Tyr Leu Ser Arg Leu Phe Lys Cys Lys Thr Gly
            180                 185                 190

Lys Asn Phe Ser Asp Tyr Leu Met Glu Val Arg Met Arg Lys Ala Leu
        195                 200                 205

Glu Leu Met Lys Ile Pro Ile Tyr Lys Leu His Glu Ile Gly Glu Met
    210                 215                 220

Val Gly Tyr Lys Asn Pro Lys Asn Phe Thr Arg Ala Phe Lys Lys Tyr
225                 230                 235                 240

Phe Gly Lys Ser Pro Ser Glu Phe Val Gln Glu Arg
                245                 250
```

<210> SEQ ID NO 152
<211> LENGTH: 858
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

-continued

<400> SEQUENCE: 152

atgtactccc aaaacaaaat agtaaaaaat actatttctc tttttataac ttatgcattt 60

ttgattctat ttggactgtt tatgatttat ccacttttgt gggttgtatc tgcggcattt 120

aaatcaaatg atgagatatt caaatcgctg tctttatttc ctcaaaaaat tgttacagac 180

tcatttataa aaggatggca ggggacaggg caatatacat ttggtagatt ttttgcaaat 240

aaccttatac ttgttattcc agttgtattt tttacaataa tatcctctac acttgtggca 300

tatgggtttg caaggtttaa ttttccactg aaaagattat tttttgtgat actaatttca 360

acattaatgc tgccagactc tgtaaacctt attccaagat atattctttt caatgcattt 420

ggatgggtag atagttacaa accatttata attccatcaa tgtttgcatc aactccattt 480

tttgtgttta tgatgataca gtttatgaga gggcttccaa gagaaataga agaagctgct 540

ataattgatg gctgcaactc tttccagata cttttgcgta taacgggtcc tctttgtaaa 600

acagcaatga tttcaatggg aattttccag tttatctgga cttggaatga ctttttagga 660

ccccttatat atatcaacag cgttgaaaaa tacacaattg cccttggact tagaatgtgt 720

gttgacagtg cagctgcaat tgcttggaac cagatcatgg ctatgactgt tatagcaatg 780

ctaccgtgta ttatcatctt ctttgctgct caaaaatact tcgttgaagg aattgcaaca 840

agcggaataa aaggttaa 858

<210> SEQ ID NO 153
<211> LENGTH: 285
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 153

Met Tyr Ser Gln Asn Lys Ile Val Lys Asn Thr Ile Ser Leu Phe Ile
1 5 10 15

Thr Tyr Ala Phe Leu Ile Leu Phe Gly Leu Phe Met Ile Tyr Pro Leu
20 25 30

Leu Trp Val Val Ser Ala Ala Phe Lys Ser Asn Asp Glu Ile Phe Lys
35 40 45

Ser Leu Ser Leu Phe Pro Gln Lys Ile Val Thr Asp Ser Phe Ile Lys
50 55 60

Gly Trp Gln Gly Thr Gly Gln Tyr Thr Phe Gly Arg Phe Phe Ala Asn
65 70 75 80

Asn Leu Ile Leu Val Ile Pro Val Val Phe Phe Thr Ile Ile Ser Ser
85 90 95

Thr Leu Val Ala Tyr Gly Phe Ala Arg Phe Asn Phe Pro Leu Lys Arg
100 105 110

Leu Phe Phe Val Ile Leu Ile Ser Thr Leu Met Leu Pro Asp Ser Val
115 120 125

Asn Leu Ile Pro Arg Tyr Ile Leu Phe Asn Ala Phe Gly Trp Val Asp
130 135 140

Ser Tyr Lys Pro Phe Ile Ile Pro Ser Met Phe Ala Ser Thr Pro Phe
145 150 155 160

Phe Val Phe Met Met Ile Gln Phe Met Arg Gly Leu Pro Arg Glu Ile
165 170 175

Glu Glu Ala Ala Ile Ile Asp Gly Cys Asn Ser Phe Gln Ile Leu Leu
180 185 190

-continued

```
Arg Ile Thr Gly Pro Leu Cys Lys Thr Ala Met Ile Ser Met Gly Ile
        195                 200                 205

Phe Gln Phe Ile Trp Thr Trp Asn Asp Phe Leu Gly Pro Leu Ile Tyr
    210                 215                 220

Ile Asn Ser Val Glu Lys Tyr Thr Ile Ala Leu Gly Leu Arg Met Cys
225                 230                 235                 240

Val Asp Ser Ala Ala Ala Ile Ala Trp Asn Gln Ile Met Ala Met Thr
                245                 250                 255

Val Ile Ala Met Leu Pro Cys Ile Ile Ile Phe Phe Ala Ala Gln Lys
            260                 265                 270

Tyr Phe Val Glu Gly Ile Ala Thr Ser Gly Ile Lys Gly
        275                 280                 285


<210> SEQ ID NO 154
<211> LENGTH: 897
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 154

atggtagttc gctatagaaa aaaagacttt gttggttttt tatatatact accttggctt     60

attggatttt tgatttttag gctatatcca tttattatgt ctttctacta ctctttcagt    120

gattatacaa tgttgaagcc gcctcgctat gttggattat ataacttcat ttatatgttc    180

acaaaggatg aactgtttcc aaaagcactt ttaaacacta taaagtatgt tataataact    240

gttccgctta aaatctcgtt tgcacttttt gttgcaataa tattgaacat gaaactgaaa    300

ggaataaatc ttttcagaac agtatattat cttccttcta tcttcggtgg gtctgttgca    360

atctcgattt tgtggagatt tttgtttatg aaagaaggta tagtaaacaa gttcttaagt    420

ctttttagaa tagaaggtat aaactggctt ggagacccaa gaatagctat gttttcagta    480

agccttcttg cggtgtggca gtttgggtca tctatggtac tatttttagc aagacttaaa    540

gagataccat cagaacttta cgaagcagca ctggttgatg gagcatcaag actaaaaatg    600

tttacaaaga taactcttcc tatgatttca cctataatgt ttttcaacct tgtgatgcag    660

accataaatg ctttccaaga atttactgga ccgtacatca tcacaggcgg tggacctgtc    720

aattccacct atcttttgag tatgctcata tatgacaatg catttaagta ttttagaatg    780

ggttatgcgg cagcactttc ctgggttcag tttgtgataa tattaatctt tactgcattt    840

atatttaggt cttctactta ttggacatat tacgagtatg atgaagggag gttctaa       897


<210> SEQ ID NO 155
<211> LENGTH: 298
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 155

Met Val Val Arg Tyr Arg Lys Lys Asp Phe Val Gly Phe Leu Tyr Ile
1               5                   10                  15

Leu Pro Trp Leu Ile Gly Phe Leu Ile Phe Arg Leu Tyr Pro Phe Ile
            20                  25                  30

Met Ser Phe Tyr Tyr Ser Phe Ser Asp Tyr Thr Met Leu Lys Pro Pro
        35                  40                  45

Arg Tyr Val Gly Leu Tyr Asn Phe Ile Tyr Met Phe Thr Lys Asp Glu
    50                  55                  60
```

-continued

```
Leu Phe Pro Lys Ala Leu Leu Asn Thr Ile Lys Tyr Val Ile Ile Thr
65                  70                  75                  80

Val Pro Leu Lys Ile Ser Phe Ala Leu Phe Val Ala Ile Ile Leu Asn
                85                  90                  95

Met Lys Leu Lys Gly Ile Asn Leu Phe Arg Thr Val Tyr Tyr Leu Pro
            100                 105                 110

Ser Ile Phe Gly Gly Ser Val Ala Ile Ser Ile Leu Trp Arg Phe Leu
        115                 120                 125

Phe Met Lys Glu Gly Ile Val Asn Lys Phe Leu Ser Leu Phe Arg Ile
    130                 135                 140

Glu Gly Ile Asn Trp Leu Gly Asp Pro Arg Ile Ala Met Phe Ser Val
145                 150                 155                 160

Ser Leu Leu Ala Val Trp Gln Phe Gly Ser Ser Met Val Leu Phe Leu
                165                 170                 175

Ala Arg Leu Lys Glu Ile Pro Ser Glu Leu Tyr Glu Ala Ala Leu Val
            180                 185                 190

Asp Gly Ala Ser Arg Leu Lys Met Phe Thr Lys Ile Thr Leu Pro Met
        195                 200                 205

Ile Ser Pro Ile Met Phe Phe Asn Leu Val Met Gln Thr Ile Asn Ala
    210                 215                 220

Phe Gln Glu Phe Thr Gly Pro Tyr Ile Ile Thr Gly Gly Gly Pro Val
225                 230                 235                 240

Asn Ser Thr Tyr Leu Leu Ser Met Leu Ile Tyr Asp Asn Ala Phe Lys
                245                 250                 255

Tyr Phe Arg Met Gly Tyr Ala Ala Ala Leu Ser Trp Val Gln Phe Val
            260                 265                 270

Ile Ile Leu Ile Phe Thr Ala Phe Ile Phe Arg Ser Ser Thr Tyr Trp
        275                 280                 285

Thr Tyr Tyr Glu Tyr Asp Glu Gly Arg Phe
    290                 295


<210> SEQ ID NO 156
<211> LENGTH: 1317
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 156

gtgataaaaa gtaaaaggtt aattgcaact cttgtgttag tagtttttac tatgtcgatc      60

ttctttgctt tttcgaccgc tgggtctgag aaagctaagg cagcatcgaa aaaggttaca     120

ctcaggttta tgtggtgggg cggagaggca agacacaaag ccactttggc agcaattcag     180

gcgtatatga agaaataccc taatgtaaga attaatgcag agtatggcgg tattgaaggt     240

tacatgcaga agctcattac ccagcttgtg ggaagaaccg ctccagatat aatccagatt     300

gacgttacat ggattggtga gctgagcagc cagggagatt tctttgcaga ccttaaaact     360

ttcaaagagg tcaacttaaa gccatttgaa gagaagtttt taaaagactg gtgctattca     420

aacggaaaac ttattggact tccaacaggt gttaatgctt cggtacttca atataataaa     480

gagtttttta agaagtttaa tatcgacgaa aatacagttt ggacgtggga taacttacta     540

tcaatagctg aaaaagtaca caaaaaagat aaaaatagtt atttgcttaa ttttgatcaa     600

atcctctgtt actatgtttt gacatcgtat attggtcaaa aaacaggaaa ggattggatt     660

ttagatgatt atacattagg atttaatagg aatcagttga tagaagcttt tacttatttg     720
```

-continued

```
aaaaaattat ttgatgtagg agctattcaa ccttttgcag aaagtgcgcc atttcaaggt    780

aagccagaac aaaatccaaa atggctaaaa ggagaattag ggattttatg gaattggact    840

tcaacttatg ctgcaaataa agctatgatt ccaagtttgg cgatgacatt accacccagg    900

ggcaacaact tgaaaaacta tgcagtaact gtcagaccgt cgcaattgtt atctgtaaat    960

aaactttcga agaatgctaa agaagctgca aaatttatta actggttttt aaacgataaa   1020

caagctgctc taatactcac tgatgtaaga ggagttcctg caagttcaag tgccagagat   1080

gcattgttaa aggcaaataa attagatcca gaaatattga gggttacaaa cgaagcggta   1140

aagtatgcag caaaaccaca gaatgcacta tcacagaatc aagaaatagc aaatatagca   1200

tatgatatca ttcagcagct tgcatacaaa cagctaacac cgacacaggc tgcagataaa   1260

ttgatagcat tatataaaca aaaactttct gagcttaaaa gaatgcagtc tcgatag      1317
```

```
<210> SEQ ID NO 157
<211> LENGTH: 438
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 157

Val Ile Lys Ser Lys Arg Leu Ile Ala Thr Leu Val Leu Val Val Phe
1               5                   10                  15

Thr Met Ser Ile Phe Phe Ala Phe Ser Thr Ala Gly Ser Glu Lys Ala
            20                  25                  30

Lys Ala Ala Ser Lys Lys Val Thr Leu Arg Phe Met Trp Trp Gly Gly
        35                  40                  45

Glu Ala Arg His Lys Ala Thr Leu Ala Ala Ile Gln Ala Tyr Met Lys
    50                  55                  60

Lys Tyr Pro Asn Val Arg Ile Asn Ala Glu Tyr Gly Gly Ile Glu Gly
65                  70                  75                  80

Tyr Met Gln Lys Leu Ile Thr Gln Leu Val Gly Arg Thr Ala Pro Asp
                85                  90                  95

Ile Ile Gln Ile Asp Val Thr Trp Ile Gly Glu Leu Ser Ser Gln Gly
            100                 105                 110

Asp Phe Phe Ala Asp Leu Lys Thr Phe Lys Glu Val Asn Leu Lys Pro
        115                 120                 125

Phe Glu Glu Lys Phe Leu Lys Asp Trp Cys Tyr Ser Asn Gly Lys Leu
    130                 135                 140

Ile Gly Leu Pro Thr Gly Val Asn Ala Ser Val Leu Gln Tyr Asn Lys
145                 150                 155                 160

Glu Phe Phe Lys Lys Phe Asn Ile Asp Glu Asn Thr Val Trp Thr Trp
                165                 170                 175

Asp Asn Leu Leu Ser Ile Ala Glu Lys Val His Lys Lys Asp Lys Asn
            180                 185                 190

Ser Tyr Leu Leu Asn Phe Asp Gln Ile Leu Cys Tyr Tyr Val Leu Thr
        195                 200                 205

Ser Tyr Ile Gly Gln Lys Thr Gly Lys Asp Trp Ile Leu Asp Asp Tyr
    210                 215                 220

Thr Leu Gly Phe Asn Arg Asn Gln Leu Ile Glu Ala Phe Thr Tyr Leu
225                 230                 235                 240

Lys Lys Leu Phe Asp Val Gly Ala Ile Gln Pro Phe Ala Glu Ser Ala
                245                 250                 255
```

-continued

```
Pro Phe Gln Gly Lys Pro Glu Gln Asn Pro Lys Trp Leu Lys Gly Glu
            260                 265                 270

Leu Gly Ile Leu Trp Asn Trp Thr Ser Thr Tyr Ala Ala Asn Lys Ala
        275                 280                 285

Met Ile Pro Ser Leu Ala Met Thr Leu Pro Pro Arg Gly Asn Asn Leu
    290                 295                 300

Lys Asn Tyr Ala Val Thr Val Arg Pro Ser Gln Leu Leu Ser Val Asn
305                 310                 315                 320

Lys Leu Ser Lys Asn Ala Lys Glu Ala Ala Lys Phe Ile Asn Trp Phe
                325                 330                 335

Leu Asn Asp Lys Gln Ala Ala Leu Ile Leu Thr Asp Val Arg Gly Val
            340                 345                 350

Pro Ala Ser Ser Ser Ala Arg Asp Ala Leu Leu Lys Ala Asn Lys Leu
        355                 360                 365

Asp Pro Glu Ile Leu Arg Val Thr Asn Glu Ala Val Lys Tyr Ala Ala
    370                 375                 380

Lys Pro Gln Asn Ala Leu Ser Gln Asn Gln Glu Ile Ala Asn Ile Ala
385                 390                 395                 400

Tyr Asp Ile Ile Gln Gln Leu Ala Tyr Lys Gln Leu Thr Pro Thr Gln
                405                 410                 415

Ala Ala Asp Lys Leu Ile Ala Leu Tyr Lys Gln Lys Leu Ser Glu Leu
            420                 425                 430

Lys Arg Met Gln Ser Arg
        435


<210> SEQ ID NO 158
<211> LENGTH: 1227
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 158

atgaaaaaag tattcttaca atttggttct gggttaggtc ccatgtctcg atcccttcca      60

atcgcagaag ggttacaaag agaaggatat atcgtaaagt attttggatt tgaaaatgca     120

aaaccatata tgaataaaat aggaattgaa gagttatcag aaaattttaa tatcaaggat     180

attaaaaaag gagtgcaaac tcctaattgg tattgtgcag agcaattttg ggaaataata     240

ggatatggta atatggaatg ggtagaaaaa aaagttgaag aattaataga atatttaaaa     300

gatttttctc ctgattttat aatatcagac cttggtatat taagttgtat tgctgcaaga     360

ataatggaca tacctttgat agctataact caaagttgtt atcatcctaa cattgctttt     420

ggaagaataa gatggtggga agaagaacaa aatttaaagt ttacattaac tgagaaatta     480

aataattatt ttaagaaaaa aggtgtttca caattaaatt cttttgaaga aatttttact     540

ggtagtttaa ccataattcc cagttttcct gaatttgatc caataaataa tccttcagaa     600

tttaacacat attatgttgg tcccatatta tgggatccat tagacatggc taaagaagag     660

tatataaaat tgtttaacag agataaaaat aagcctacaa ttttttgcta tacagcaaga     720

ttttacgaca atgtgggtga aagtggaatt attattttta aaacattact ttcagcatta     780

aaaaaatttg atgctaacat tatttttct acagggagtg attcggacag gaaaatagca     840

aaagagattt taaactctta cggaattgat gaagagaaat ttagcattat tgattgggtt     900

ccaatgggaa ttgcttatgg aaactctgat gttgttatcc atcatggagg ccatggaagt     960
```

-continued

```
tgtttaggtc aatttttgta tgaggtacct tcattaatta tacctactca tactgaacga   1020

gagtataatg caagaatttg caccaatatg ggagtttcta aatttataaa aagagaagac   1080

attgaaaaag cagatatatt agctgaaatt gttgagattt taactaactc aagttttaaa   1140

gaaagattac acttttggca tactaaacta aatgaatata attttacagg tgtaaataaa   1200

gttttagaat tgattcaaaa attataa                                       1227
```

<210> SEQ ID NO 159
<211> LENGTH: 408
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 159

```
Met Lys Lys Val Phe Leu Gln Phe Gly Ser Gly Leu Gly Pro Met Ser
1               5                   10                  15

Arg Ser Leu Pro Ile Ala Glu Gly Leu Gln Arg Glu Gly Tyr Ile Val
            20                  25                  30

Lys Tyr Phe Gly Phe Glu Asn Ala Lys Pro Tyr Met Asn Lys Ile Gly
        35                  40                  45

Ile Glu Glu Leu Ser Glu Asn Phe Asn Ile Lys Asp Ile Lys Lys Gly
    50                  55                  60

Val Gln Thr Pro Asn Trp Tyr Cys Ala Glu Gln Phe Trp Glu Ile Ile
65                  70                  75                  80

Gly Tyr Gly Asn Met Glu Trp Val Glu Lys Lys Val Glu Glu Leu Ile
                85                  90                  95

Glu Tyr Leu Lys Asp Phe Ser Pro Asp Phe Ile Ile Ser Asp Leu Gly
            100                 105                 110

Ile Leu Ser Cys Ile Ala Ala Arg Ile Met Asp Ile Pro Leu Ile Ala
        115                 120                 125

Ile Thr Gln Ser Cys Tyr His Pro Asn Ile Ala Phe Gly Arg Ile Arg
    130                 135                 140

Trp Trp Glu Glu Glu Gln Asn Leu Lys Phe Thr Leu Thr Glu Lys Leu
145                 150                 155                 160

Asn Asn Tyr Phe Lys Lys Lys Gly Val Ser Gln Leu Asn Ser Phe Glu
                165                 170                 175

Glu Ile Phe Thr Gly Ser Leu Thr Ile Ile Pro Ser Phe Pro Glu Phe
            180                 185                 190

Asp Pro Ile Asn Asn Pro Ser Glu Phe Asn Thr Tyr Tyr Val Gly Pro
        195                 200                 205

Ile Leu Trp Asp Pro Leu Asp Met Ala Lys Glu Glu Tyr Ile Lys Leu
    210                 215                 220

Phe Asn Arg Asp Lys Asn Lys Pro Thr Ile Phe Cys Tyr Thr Ala Arg
225                 230                 235                 240

Phe Tyr Asp Asn Val Gly Glu Ser Gly Ile Ile Ile Phe Lys Thr Leu
                245                 250                 255

Leu Ser Ala Leu Lys Lys Phe Asp Ala Asn Ile Ile Phe Ser Thr Gly
            260                 265                 270

Ser Asp Ser Asp Arg Lys Ile Ala Lys Glu Ile Leu Asn Ser Tyr Gly
        275                 280                 285

Ile Asp Glu Glu Lys Phe Ser Ile Ile Asp Trp Val Pro Met Gly Ile
    290                 295                 300
```

-continued

```
Ala Tyr Gly Asn Ser Asp Val Val Ile His His Gly Gly His Gly Ser
305                 310                 315                 320

Cys Leu Gly Gln Phe Leu Tyr Glu Val Pro Ser Leu Ile Ile Pro Thr
                325                 330                 335

His Thr Glu Arg Glu Tyr Asn Ala Arg Ile Cys Thr Asn Met Gly Val
            340                 345                 350

Ser Lys Phe Ile Lys Arg Glu Asp Ile Glu Lys Ala Asp Ile Leu Ala
        355                 360                 365

Glu Ile Val Glu Ile Leu Thr Asn Ser Ser Phe Lys Glu Arg Leu His
    370                 375                 380

Phe Trp His Thr Lys Leu Asn Glu Tyr Asn Phe Thr Gly Val Asn Lys
385                 390                 395                 400

Val Leu Glu Leu Ile Gln Lys Leu
                405


<210> SEQ ID NO 160
<211> LENGTH: 1317
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 160

atgaagtact tcaaagacat tccagaagta aaatatgaag gaccacagtc ggacaaccca     60

tttgctttca agtactacaa tcctgacgaa atcattgacg gcaagccttt gaaagaccac    120

cttcgttttg ctattgctta ctggcacaca ttctgtgcaa caggaagcga tccgtttgga    180

caacctacaa ttgttcgtcc ttgggataag ttttcaaacc gaatggacaa cgcaaaagca    240

agggttgagg cagcatttga atttttgaa ctgttagatg taccattttt ctgcttccat    300

gacagagata ttgcacctga aggggaaaat ttaaaagagt caaataagaa tttggatgag    360

attgtttctt taataaaaga gtatttgaaa accagcaaga caaaagtatt atggggaaca    420

gcaaacctat tttcacatcc gcgatatgtt catggtgctg caacatcctg caatgccgat    480

gtttttgcat atgcagcagc gcaagtgaaa aaggcgttag aggttacaaa agagcttggc    540

ggcgaaaact atgtgttctg gggcggaagg gaaggttatg agacacttct aaatacagat    600

atgggattgg aacttgataa ccttgcaaga tttttgcata tggcggttga gtatgcaaag    660

gaaataggtt ttgacggaca gtttttaata gaaccaaaac caaaagagcc aactaagcat    720

cagtacgatt ttgattcggc tcatgtttat ggatttttga aaaagtatga tcttgacaaa    780

tacttcaagc tcaacataga ggtaaaccat gcaaccttag caggacatga tttccaccat    840

gagttgagat ttgcgcgaat aaacaacatg cttggttcaa ttgacgctaa catgggcgat    900

ttgcttttgg gctgggatac agatcagttc ccaacagatg taagacttac tacacttgct    960

atgtatgagg ttattaaagc tggtggtttt gacaaaggtg gacttaactt tgacgcaaag   1020

gtaagaagag gttcttttga gcttgaagac ttggtcattg gtcacattgc tggcatggat   1080

gcttttgcta aaggcttcaa gattgcgtat aagcttgtta aagatggcgt atttgataaa   1140

tttatagatg agagatacaa gagctacaaa gaaggaatcg gtgctaagat tgtaagcggt   1200

gaagcaaact tcaagatgtt agaggaatat gctctgtctc ttgacaagat agaaaataaa   1260

tctggcaagc aagagcttct tgagatgatt ttgaacaaat atatgttcag cgaataa      1317


<210> SEQ ID NO 161
<211> LENGTH: 438
```

-continued

<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 161

Met Lys Tyr Phe Lys Asp Ile Pro Glu Val Lys Tyr Glu Gly Pro Gln
1 5 10 15

Ser Asp Asn Pro Phe Ala Phe Lys Tyr Tyr Asn Pro Asp Glu Ile Ile
20 25 30

Asp Gly Lys Pro Leu Lys Asp His Leu Arg Phe Ala Ile Ala Tyr Trp
35 40 45

His Thr Phe Cys Ala Thr Gly Ser Asp Pro Phe Gly Gln Pro Thr Ile
50 55 60

Val Arg Pro Trp Asp Lys Phe Ser Asn Arg Met Asp Asn Ala Lys Ala
65 70 75 80

Arg Val Glu Ala Ala Phe Glu Phe Phe Glu Leu Leu Asp Val Pro Phe
85 90 95

Phe Cys Phe His Asp Arg Asp Ile Ala Pro Glu Gly Glu Asn Leu Lys
100 105 110

Glu Ser Asn Lys Asn Leu Asp Glu Ile Val Ser Leu Ile Lys Glu Tyr
115 120 125

Leu Lys Thr Ser Lys Thr Lys Val Leu Trp Gly Thr Ala Asn Leu Phe
130 135 140

Ser His Pro Arg Tyr Val His Gly Ala Ala Thr Ser Cys Asn Ala Asp
145 150 155 160

Val Phe Ala Tyr Ala Ala Ala Gln Val Lys Lys Ala Leu Glu Val Thr
165 170 175

Lys Glu Leu Gly Gly Glu Asn Tyr Val Phe Trp Gly Gly Arg Glu Gly
180 185 190

Tyr Glu Thr Leu Leu Asn Thr Asp Met Gly Leu Glu Leu Asp Asn Leu
195 200 205

Ala Arg Phe Leu His Met Ala Val Glu Tyr Ala Lys Glu Ile Gly Phe
210 215 220

Asp Gly Gln Phe Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys His
225 230 235 240

Gln Tyr Asp Phe Asp Ser Ala His Val Tyr Gly Phe Leu Lys Lys Tyr
245 250 255

Asp Leu Asp Lys Tyr Phe Lys Leu Asn Ile Glu Val Asn His Ala Thr
260 265 270

Leu Ala Gly His Asp Phe His His Glu Leu Arg Phe Ala Arg Ile Asn
275 280 285

Asn Met Leu Gly Ser Ile Asp Ala Asn Met Gly Asp Leu Leu Leu Gly
290 295 300

Trp Asp Thr Asp Gln Phe Pro Thr Asp Val Arg Leu Thr Thr Leu Ala
305 310 315 320

Met Tyr Glu Val Ile Lys Ala Gly Gly Phe Asp Lys Gly Gly Leu Asn
325 330 335

Phe Asp Ala Lys Val Arg Arg Gly Ser Phe Glu Leu Glu Asp Leu Val
340 345 350

Ile Gly His Ile Ala Gly Met Asp Ala Phe Ala Lys Gly Phe Lys Ile
355 360 365

Ala Tyr Lys Leu Val Lys Asp Gly Val Phe Asp Lys Phe Ile Asp Glu
370 375 380

-continued

```
Arg Tyr Lys Ser Tyr Lys Glu Gly Ile Gly Ala Lys Ile Val Ser Gly
385                 390                 395                 400

Glu Ala Asn Phe Lys Met Leu Glu Glu Tyr Ala Leu Ser Leu Asp Lys
                405                 410                 415

Ile Glu Asn Lys Ser Gly Lys Gln Glu Leu Leu Glu Met Ile Leu Asn
            420                 425                 430

Lys Tyr Met Phe Ser Glu
        435


<210> SEQ ID NO 162
<211> LENGTH: 1914
<212> TYPE: DNA
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 162

atgagaatgc tattaaaaag gtctcttgct ctgctggtaa gtattgtcct tgtattttcg     60

ttgttcttaa gtgtttttcc acagcaagca agagcacaag ataccataaa aattgtaggt    120

aactggcagg atgctggtaa ctggaatttt gacagctcta atattgtttt gagtgaaacc    180

tcaactccgg gactttatta tggtgagtat acttttaaaa ctggtggttc ttatgagttc    240

aaagcagtta taaatggtag tatatggtgt acaggcgctc caaaagtagc agataataat    300

actaacatac ctcttaatgt tacagatggt cagacagtta agttttggtt cttcaaaaat    360

tctcaacttg taattgacag cactcatttt ccaaatgggc ctgatagttt agtagggcaa    420

aattctttta aatttgtagg agtaaataat gaatggaacc ctaatgatgg taagtatcaa    480

tttataagag tcccagatgc aacatatacc tatttatatg ataatagtta caatgatttt    540

tcgataccat atggatttaa gattataata ggtggttttg ggaacctatc atgggcatgg    600

aatggtgcaa aagaaggttc agttgttaaa tttaaagagg gtggagataa tattgacctg    660

aaagaattta aagaatctaa tgataatctt ctaaaaacaa aatttttcct tgatatccta    720

aacggctggc tctttacaga aaaagatttg acaaatatcc aaccagtaga ttttgcaaat    780

aacggctctg taattggtgg aacttcagtt catcttacat ggacaccata tacaccaaat    840

gcaaatccac tttcagccaa gctttactac aaaataattg atgtgagcaa tgatagcgag    900

ttggttagtc tcaccgatta cagcacaatc cacaaaatag atatcccaag agagtggata    960

ggaaagacaa ttaaaattat tgcaaatgcc aagataggtg aaataacagg acctaatgtt   1020

gagtttacat taaatattgt tgacttgcca caaaacttaa tagttgaagc agctaactat   1080

ataacataca actctataga tgagacattt ttaaatttgg ctcaaggtga tacagttcaa   1140

aacataactt ctaacttcgc agttgttaca tcatatgttt acaatgtggt gtacgaggga   1200

aatagttatc cactgacatt taatattgac tgggagagca gtaattcatc tgtattaaca   1260

ataaatggta gtacagttgt ggttacaagg ccaacacaag gtgaccttca ggttgagttg   1320

agagcaagag caagatttgg tgctatttca gcagatggac aaaaaatatt tactcttaca   1380

gtaaagaaat tcctattagg agtagatggt ggaatacctg ttacatttaa cgtaacagtt   1440

ccagattata ctcctgataa tgacaatatt tacattgctg gagattttaa gactgataag   1500

cttccaaaat gggaccctgt aggtatcaaa ttgataaagg tcggagataa aaaatatagt   1560

ataacaatgt atcttccacc gaatgtaaca attgaatata aatatacccg cggaagttgg   1620

tcaaaagtag aaaaggatgc atttggtaat gaaatatcaa atagagttct gcagataaat   1680
```

-continued

```
aatcaggctg tgacaaaaaa tgatgtagtt gaagcatttg cggaccttgg agctgtcaag   1740

caaggtcttc caacggtgaa tttggtcatc aatgttcctc aacaaactgt aattgataca   1800

ggtgatggag gaagggtaat aaaagcagaa ggtggtgcta tttcagttcc aaatactaag   1860

aacactgttt tgatagatat tcggagtgca ttaaatgcta ttatattaaa gtaa         1914


<210> SEQ ID NO 163
<211> LENGTH: 637
<212> TYPE: PRT
<213> ORGANISM: Anaerocellum thermophilum DSM 6725

<400> SEQUENCE: 163

Met Arg Met Leu Leu Lys Arg Ser Leu Ala Leu Leu Val Ser Ile Val
1               5                   10                  15

Leu Val Phe Ser Leu Phe Leu Ser Val Phe Pro Gln Gln Ala Arg Ala
            20                  25                  30

Gln Asp Thr Ile Lys Ile Val Gly Asn Trp Gln Asp Ala Gly Asn Trp
        35                  40                  45

Asn Phe Asp Ser Ser Asn Ile Val Leu Ser Glu Thr Ser Thr Pro Gly
    50                  55                  60

Leu Tyr Tyr Gly Glu Tyr Thr Phe Lys Thr Gly Gly Ser Tyr Glu Phe
65                  70                  75                  80

Lys Ala Val Ile Asn Gly Ser Ile Trp Cys Thr Gly Ala Pro Lys Val
                85                  90                  95

Ala Asp Asn Asn Thr Asn Ile Pro Leu Asn Val Thr Asp Gly Gln Thr
            100                 105                 110

Val Lys Phe Trp Phe Phe Lys Asn Ser Gln Leu Val Ile Asp Ser Thr
        115                 120                 125

His Phe Pro Asn Gly Pro Asp Ser Leu Val Gly Gln Asn Ser Phe Lys
    130                 135                 140

Phe Val Gly Val Asn Asn Glu Trp Asn Pro Asn Asp Gly Lys Tyr Gln
145                 150                 155                 160

Phe Ile Arg Val Pro Asp Ala Thr Tyr Thr Tyr Leu Tyr Asp Asn Ser
                165                 170                 175

Tyr Asn Asp Phe Ser Ile Pro Tyr Gly Phe Lys Ile Ile Ile Gly Gly
            180                 185                 190

Phe Gly Asn Leu Ser Trp Ala Trp Asn Gly Ala Lys Glu Gly Ser Val
        195                 200                 205

Val Lys Phe Lys Glu Gly Gly Asp Asn Ile Asp Leu Lys Glu Phe Lys
    210                 215                 220

Glu Ser Asn Asp Asn Leu Leu Lys Thr Lys Phe Phe Leu Asp Ile Leu
225                 230                 235                 240

Asn Gly Trp Leu Phe Thr Glu Lys Asp Leu Thr Asn Ile Gln Pro Val
                245                 250                 255

Asp Phe Ala Asn Asn Gly Ser Val Ile Gly Gly Thr Ser Val His Leu
            260                 265                 270

Thr Trp Thr Pro Tyr Thr Pro Asn Ala Asn Pro Leu Ser Ala Lys Leu
        275                 280                 285

Tyr Tyr Lys Ile Ile Asp Val Ser Asn Asp Ser Glu Leu Val Ser Leu
    290                 295                 300

Thr Asp Tyr Ser Thr Ile His Lys Ile Asp Ile Pro Arg Glu Trp Ile
305                 310                 315                 320
```

-continued

```
Gly Lys Thr Ile Lys Ile Ile Ala Asn Ala Lys Ile Gly Glu Ile Thr
                325                 330                 335

Gly Pro Asn Val Glu Phe Thr Leu Asn Ile Val Asp Leu Pro Gln Asn
            340                 345                 350

Leu Ile Val Glu Ala Ala Asn Tyr Ile Thr Tyr Asn Ser Ile Asp Glu
        355                 360                 365

Thr Phe Leu Asn Leu Ala Gln Gly Asp Thr Val Gln Asn Ile Thr Ser
    370                 375                 380

Asn Phe Ala Val Val Thr Ser Tyr Val Tyr Asn Val Val Tyr Glu Gly
385                 390                 395                 400

Asn Ser Tyr Pro Leu Thr Phe Asn Ile Asp Trp Glu Ser Ser Asn Ser
                405                 410                 415

Ser Val Leu Thr Ile Asn Gly Ser Thr Val Val Val Thr Arg Pro Thr
            420                 425                 430

Gln Gly Asp Leu Gln Val Glu Leu Arg Ala Arg Ala Arg Phe Gly Ala
        435                 440                 445

Ile Ser Ala Asp Gly Gln Lys Ile Phe Thr Leu Thr Val Lys Lys Phe
    450                 455                 460

Leu Leu Gly Val Asp Gly Gly Ile Pro Val Thr Phe Asn Val Thr Val
465                 470                 475                 480

Pro Asp Tyr Thr Pro Asp Asn Asp Asn Ile Tyr Ile Ala Gly Asp Phe
                485                 490                 495

Lys Thr Asp Lys Leu Pro Lys Trp Asp Pro Val Gly Ile Lys Leu Ile
            500                 505                 510

Lys Val Gly Asp Lys Lys Tyr Ser Ile Thr Met Tyr Leu Pro Pro Asn
        515                 520                 525

Val Thr Ile Glu Tyr Lys Tyr Thr Arg Gly Ser Trp Ser Lys Val Glu
    530                 535                 540

Lys Asp Ala Phe Gly Asn Glu Ile Ser Asn Arg Val Leu Gln Ile Asn
545                 550                 555                 560

Asn Gln Ala Val Thr Lys Asn Asp Val Val Glu Ala Phe Ala Asp Leu
                565                 570                 575

Gly Ala Val Lys Gln Gly Leu Pro Thr Val Asn Leu Val Ile Asn Val
            580                 585                 590

Pro Gln Gln Thr Val Ile Asp Thr Gly Asp Gly Gly Arg Val Ile Lys
        595                 600                 605

Ala Glu Gly Gly Ala Ile Ser Val Pro Asn Thr Lys Asn Thr Val Leu
    610                 615                 620

Ile Asp Ile Arg Ser Ala Leu Asn Ala Ile Ile Leu Lys
625                 630                 635
```

1. A method of processing plant biomass, the method comprising:

growing *Anaerocellum thermophilum* on a substrate that comprises plant biomass under conditions effective for the *A. thermophilum* to convert at least a portion of the plant biomass to a water soluble product or a water insoluble product; and

isolating at least a portion of the water soluble product or water insoluble product.

2.-**3**. (canceled)

4. The method of claim **1** wherein the conditions comprise a temperature of at least 70° C.

5. (canceled)

6. The method of claim **1** wherein the plant biomass comprises spent biomass.

7.-**11**. (canceled)

12. The method of claim **1** wherein the water soluble product comprises methanol, ethanol, butanol, fatty acids, hydrogen gas, succinic acid, citric acid, oxaloacetic acid, malic acid, adipic acid, fumaric acid, pyruvic acid, a monosaccharide, or a disaccharide.

13. (canceled)

14. The method of claim **1** wherein the water soluble product or water insoluble product comprises a biofuel.

15-17. (canceled)

18. The method of claim **1** wherein the *A. thermophilum* produces a water insoluble product that comprises alkyl fatty acids.

19.-**21**. (canceled)

22. A method of transferring one or more polynucleotides of *A. thermophilum* to a recipient microorganism, the method comprising:

providing an expression vector appropriate for the recipient microorganism comprising an *A. thermophilum* PBU polynucleotide; and

introducing the expression vector into the recipient microorganism.

23. The method of claim **22** wherein the recipient microorganism comprises *Saccharomyces cerevisiae.*

24.-**26**. (canceled)

27. The method of claim **22** wherein the recipient microorganism comprises an extremophile.

28.-**34**. (canceled)

35. The method of claim **22** wherein the recipient microorganism comprises a thermophilic microbe.

36.-**39**. (canceled)

40. The method of claim **22** wherein the *A. thermophilum* polynucleotide comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence of a plant biomass utilization (PBU) polynucleotide.

41.-**43**. (canceled)

44. The method of claim **40** wherein the PBU polynucleotide comprises a polysaccharide hydrolases and related enzymes (PHR) polynucleotide.

45.-**76**. (canceled)

77. A genetically-modified microorganism comprising one or more *A. thermophilum* plant biomass utilization (PBU) polynucleotides.

78. The genetically-modified microorganism of claim **77** wherein the PBU polynucleotide comprises a nucleotide sequence having at least 80% identity to the nucleotide sequence of a PBU polynucleotide.

79. (canceled)

80. The genetically-modified microorganism of claim **78** wherein the PBU polynucleotide comprises one or more coding regions from a gene cluster chosen from: SYb001 and SYb037.

81. (canceled)

82. The genetically-modified microorganism of claim **78** wherein the PBU polynucleotide comprises a polysaccharide hydrolases and related enzymes (PHR) polynucleotide.

83-**85**. (canceled)

86. The genetically-modified microorganism of claim **77** wherein the microorganism comprises a eukaryote.

87. (canceled)

88. The genetically-modified microorganism of claim **77** wherein the microorganism comprises an extremophile.

89. The genetically-modified microorganism of claim **77** wherein the microorganism comprises a thermophilic bacterium.

90. The genetically-modified microorganism of claim **77** wherein the microorganism comprises a mesophilic microbe.

91. An isolated polypeptide comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a PBU polypeptide.

92. (canceled)

93. The isolated polypeptide of claim **91** wherein the PBU polypeptide comprises a PHR polypeptide.

94.-**114**. (canceled)

115. A method of processing plant biomass, the method comprising:

growing *Anaerocellum thermophilum* on a substrate that comprises plant biomass under conditions effective for the *A. thermophilum* to convert at least a portion of the plant biomass to a water soluble product or a water insoluble product; and

converting at least a portion of the water soluble product or water insoluble product to a biofuel or commodity chemical.

116. The method of claim **115** wherein the conditions comprise a temperature of at least 70° C.

117. The method of claim **115** wherein the plant biomass comprises spent biomass.

118. The method of claim **115** wherein the biofuel or commodity chemical comprises methanol, ethanol, butanol, fatty acids, hydrogen gas, succinic acid, citric acid, oxaloacetic acid, malic acid, adipic acid, fumaric acid, pyruvic acid, a monosaccharide, or a disaccharide.

* * * * *