

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 February 2003 (27.02.2003)

PCT

(10) International Publication Number
WO 03/016536 A2

- (51) International Patent Classification⁷: **C12N 15/63**, 1/21, C07K 14/195, C12N 9/12
- (74) Agent: **HOFMAN-BANG ZACCO A/S**; Hans Bekkevolds Allé 7, DK-2900 Hellerup (DK).
- (21) International Application Number: PCT/DK02/00535
- (22) International Filing Date: 13 August 2002 (13.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/311,871 13 August 2001 (13.08.2001) US
- (71) Applicant (for all designated States except US): **DTU, TECHNICAL UNIVERSITY OF DENMARK** [DK/DK]; Anker Engelunds Vej 1, Building 101A, DK-2800 Kgs. Lyngby (DK).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **AHRING, Birgitte, Kiær** [DK/DK]; Stakkeledet 27, DK-2970 Hørsholm (DK). **CLAUSEN, Anders** [DK/DK]; Slotsbakken 20, DK-2970 Hørsholm (DK). **MIKKELSEN, Marie, Just** [DK/DK]; Stenløsevej 52, DK-2700 Brønshøj (DK).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PLASMIDS FROM AN EXTREMELY THERMOPHILIC MICROORGANISM AND DERIVED EXPRESSION VECTORS

(57) Abstract: The present invention concerns the isolation of plasmids from extremely thermophilic anaerobic microorganisms and their use in genetic transformation of thermophilic and mesophilic microorganisms. More particular the invention concerns the use of thermostable plasmid vectors as tools for creating shuttle vectors for genetic transformation of extremely thermophilic anaerobic microorganisms.



WO 03/016536 A2

Plasmids from an extremely thermophilic microorganism and derived expression vectors

TECHNICAL FIELD OF THE INVENTION

- 5 The present invention concerns genetic transformation of extremely thermophilic anaerobic microorganisms. More particular the invention concerns thermostable plasmid vectors as tools for genetic transformation of extremely thermophilic anaerobic microorganisms.

10 BACKGROUND OF THE INVENTION

- Extremely thermophilic anaerobic microorganisms, growing at or above 70 °C have long attracted considerable industrial attention. The high temperature optimum of these organisms, and their capability of converting a variety of simple and complex carbohydrates, make them especially interesting as bio-
- 15 catalysts in a number of industrial processes. Polysaccharide degrading enzymes like xylanases, cellulases and amylases with high temperature optimum and high efficiencies have been cloned from thermophilic anaerobic microorganisms. Bio-transformation of agricultural products and waste from agriculture, forestry, industry and municipal sources with thermophilic an-
- 20 aerobic microorganisms under simultaneous production of industrial important compounds such as fuel ethanol, are becoming increasingly interesting areas for industrial production.

- Thermophilic conditions will ensure high conversion rates and high temperatures will increase the product formation rate and reduce the risk of contamination. In the case of ethanol, as an example, thermophilic anaerobic micro-
- 25 organisms are potential candidates for bio-ethanol production from agricultural products or from waste. Ethanol can substitute for the octane booster Methyl-Tertiary-Butyl-Ether (MTBE) in car fuel or be used directly as a liquid fuel. In contrast to yeasts and the majority of mesophilic bacteria, the ther-
- 30 mophilic anaerobic microorganisms are able to metabolise a wider range of substrates, including complex carbohydrate polymers, to yield ethanol. Since

the fermentation catalysed by these thermophiles takes place at temperatures just below the boiling point of ethanol, which is about 90°C, it is easy to strip the ethanol from the bioreactor during production. These advantageous properties of thermophilic anaerobes could be exploited for biocatalysis processes using high efficiency reactors and facilitated product recovery (Lynd, 1989; Weimar et al., 1984). Improvement of product yield and minimisation of by-product formation is, however, still essential in order to make these biocatalytic processes economically viable on an industrial scale.

The genetic or metabolic engineering approaches needed to optimise product yield and minimise un-wanted by-product formation of anaerobic thermophilic microbial strains, are hampered by the lack of suitable thermostable plasmid-based genetic systems.

Plasmids from thermophilic anaerobic microorganisms

Despite the large number of thermophilic anaerobic microorganisms isolated and studied, only a few plasmids have been identified. The plasmids are listed in Table 1.

All pTA plasmids were isolated from partially described thermophilic clostridia capable of converting a number of carbohydrates into ethanol (Kurose *et al.*, 1989b). From this study only 3 out of 320 strains contained one or two plasmids. The four plasmids which were found ranged in size from 1.5 kb to 7 kb. Restriction maps were constructed from the plasmids. No special phenotypes were found to be associated with these plasmids, and they remain cryptic (Kurose *et al.*, 1989a). No information as to copy number is available. In addition, sequence information for these plasmids is lacking, making these plasmids less attractive as the basis for a genetic system for anaerobic thermophilic microorganisms. The fact that all pTA plasmids have been derived from poorly described

isolates means that the characterisation of these plasmids in their natural hosts is hampered by the lack of information concerning the host strains.

Tsoi et al. (1987) reported that plasmids were present in four out of seven *Clostridium thermocellum* strains from their culture collection. They ranged in size from 25 kb to 50 kb, which initially makes them unattractive for vector construction. In addition no sequence information is available.

5 In thermophilic anaerobic bacteria fermenting hemicellulose, Weimer et al. (1984) found that four out of seven isolated strains contained one plasmid each. All plasmids were small, 1.9-2.4 kb. The plasmid-containing strains were not classified, but were believed to belong to either *Thermoanaerobacterium*, *Thermoanaerobacter* or *Thermobacteroides*. Determination of plas-
10 mid size was done by comparing covalently closed circular plasmid migration patterns to *E. coli* plasmids. Comparing migration patterns of different covalently closed circular plasmids does, however, not necessarily give any information of the actual plasmid size. Only comparison of linearized forms will give the desired information. The presence of plasmids in these strains did
15 not correlate with any gross morphological or physiological characteristics of the organisms (Weimer et al., 1984). The fact that all pDP plasmids have been derived from scarcely described isolates means that the characterisation of these plasmids in their natural hosts is hampered by the lack of information concerning the host strains. In addition no sequence information is
20 available.

Thermoanaerobacterium saccharolyticum DSM571 was shown to harbour two different plasmids, pNB1 4.9 kb and pNB2 2.0 kb (Belogurova et al., 1991). The copy number of pNB1 is 1-2 per cell. The cells were resistant to streptomycin and kanamycin up to 20 µg/ml. By addition of streptomycin or
25 kanamycin the copy number of pNB1 increased to 5-10 plasmid copies per cell without affecting the growth rate. In addition, hydrogen production was strongly affected by the presence of antibiotics. Kanamycin and streptomycin increased the hydrogen production by almost 100%, whereas other antibiotics (ampicillin, tetracycline, chloramphenicol and cycloserine) inhibited hy-
30 drogen production and cell growth completely (Belogurova et al., 1991). pNB2 has been fully sequenced (Delver et al., 1996) whereas only limited sequence information from pNB1 is available. A 400 bp fragment within a PstI-HindIII region, found on both plasmids, show 100% homology

Plasmid name	Size	Functions associated	Sequence information available	Isolated from (reference)	Temperature Optimum (strain)
pTA106	1.5 kb	Cryptic	No	Thermophilic <i>Clostridia</i> (Kurose <i>et al.</i> , 1989a)	60 °C
pTA729	1.5 kb	Cryptic	No	Thermophilic <i>Clostridia</i> (Kurose <i>et al.</i> , 1989a)	60 °C
pTA688S	2.3 kb	Cryptic	No	Thermophilic <i>Clostridia</i> (Kurose <i>et al.</i> , 1989a)	60 °C
pTA688L	7.0 kb	Cryptic	No	Thermophilic <i>Clostridia</i> (Kurose <i>et al.</i> , 1989a)	60 °C
pNB1	4.9 kb	Possibly Km ^R , Sm ^R	No	<i>Thermoanaerobacterium thermosaccharolyticum</i> DSM 571 (Belogurova <i>et al.</i> , 1991)	55-60 °C
pNB2	2.0 kb	Possibly Km ^R , Sm ^R	Yes	<i>Thermoanaerobacterium thermosaccharolyticum</i> DSM 571 (Belogurova <i>et al.</i> , 1991; Delver <i>et al.</i> , 1996)	55-60 °C
pFA1	25 kb	Cryptic	No	<i>Clostridium thermocellum</i> F1 (Tsoi <i>et al.</i> , 1987)	60 °C
pFA7	25 kb	Cryptic	No	<i>Clostridium thermocellum</i> F7 (Tsoi <i>et al.</i> , 1987)	60 °C
pFB1	45-50 kb	Cryptic	No	<i>Clostridium thermocellum</i> F1 (Tsoi <i>et al.</i> , 1987)	60 °C
PSH49	8.0 kb	Cryptic	No	Thermophilic anaerobe FERM 7495 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985a; AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985f)	70 °C
PSH29	6.1 kb	Cryptic	No	Thermophilic anaerobe FERM 7494 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985d; AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985c)	69 °C
PSH62	5.3 kb	Cryptic	No	Thermophilic anaerobe FERM 7493 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985e; AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985b)	66 °C
pFB7	45-50 kb	Cryptic	No	<i>Clostridium thermocellum</i> F7 (Tsoi <i>et al.</i> , 1987)	60 °C
pDP5009	3.8 kb	Cryptic	No	Not classified bacterium. Possibly <i>Thermoanaerobacterium</i> , <i>Thermoanaerobacter</i> , <i>Thermobacteroides</i> (Weimer <i>et al.</i> , 1984)	60 °C
pDP5010	2.4 kb	Cryptic	No	Not classified bacterium. Possibly <i>Thermoanaerobacterium</i> , <i>Thermoanaerobacter</i> , <i>Thermobacteroides</i> (Weimer <i>et al.</i> , 1984)	60 °C
pDP5012	2.4 kb	Cryptic	No	Not classified bacterium. Possibly <i>Thermoanaerobacterium</i> , <i>Thermoanaerobacter</i> , <i>Thermobacteroides</i> (Weimer <i>et al.</i> , 1984)	60 °C
pDP5016	1.9 kb	Cryptic	No	Not classified bacterium. Possibly <i>Thermoanaerobacterium</i> , <i>Thermoanaerobacter</i> , <i>Thermobacteroides</i> (Weimer <i>et al.</i> , 1984)	60 °C

Table 1. Plasmids found in thermophilic anaerobic bacteria. Km^R, kanamycin resistance Sm^R, streptomycin resistance.

(Kalyuzhnyi *et al.*, 1992) so it was suggested that the smaller plasmid pNB2 originally has derived from pNB1. More sequence information is needed to elucidate this hypothesis. It is, however, very likely that some kind of recombination event has occurred. The sequence of pNB2 reveals three open reading frames (ORF's). One of the ORF's, ORF289, shows homology to replication proteins encoded by rolling circle (RC) plasmids of the pC194/pUB110 family (Delver *et al.*, 1996). Strongest similarity was found to RC plasmids of the pC194 family, in particular three plasmids showed significant similarity, pTB913, pBC1 and pST1. These plasmids were originally identified in the moderate thermophilic bacteria *Bacillus sp.*, *Bacillus coagulans* and *Streptococcus thermophilus*, respectively (Ammendola *et al.*, 1992; Muller *et al.*, 1986; Janzen *et al.*, 1992). Rolling circle plasmids replicate via single stranded DNA (ssDNA) intermediates. ssDNA intermediates of pNB2 were indeed found in *T. saccharolyticum* (Delver *et al.*, 1996), suggesting that pNB2 actually replicates via the rolling circle mechanism. The majority of rolling circle plasmids containing single stranded origins (ss-origins) are known to be host specific (Gruss *et al.*, 1989), and not surprisingly pNB2 failed to replicate in both *E. coli* and *B. subtilis*. In addition, no significant similarity to any known ss-origins from other RC plasmids was found. pNB2 derives from a bacterial strain with a temperature optimum at 60 °C, indicating that this organism belongs to the thermophilic microorganisms but not to the extremely thermophilic microorganisms. This means that the plasmid stability will be decreased at temperatures above 60 °C and it is consequently less attractive to use pNB2 at temperatures above 60 °C.

Three plasmids, pSH29, pSH49 and pSH62, have been isolated from undescribed thermophilic anaerobic isolates. pSH29 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985d) from isolate FERM 7494 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985c), which has a growth optimum at 69°C (growth range 50-78°C). pSH49 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985a) from isolate FERM 7495 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985f), which has a growth optimum at 70°C (growth range 50-80°C). pSH62 (AGENCY OF IND SCI & TECHNOLOGY (AGEN), 1985e) from isolate FERM 7493, which has a growth optimum at 66°C (growth range 50-74°C). The plasmids range in size from 5.3 kb to 8.0 kb. No

information as to the function or DNA sequence of the plasmids is available, and the bacterial isolates are poorly characterised.

Plasmid constructs used in transformation of thermophilic anaerobic bacteria

- 5 The first attempts to transform thermophilic anaerobic bacteria were done using the mesophilic plasmids pUB110 and pGS13 as vectors and *T. thermohydrosulfuricus* as recipient (Table 2) (Soutschek-Bauer *et al.*, 1985).

Transformants were selected on the basis of increased kanamycin and/or chloramphenicol resistance. Plasmid bands were barely visible when attempts were made to recover the plasmids from transformed *T. thermohydrosulfuricus*. These plasmid preparations could however be used to re-transform *B. subtilis*. One third of the re-transformed *B. subtilis* clones contained plasmids with increased size. Both plasmids could be stably maintained at 55 °C. Initial attempts to transform *T. thermohydrosulfuricus* protoplasts failed due to the sensitivity of this strain to polyethylene glycol (PEG). Therefore, an alternative protocol for transformation of intact cells was developed. The cells are made competent by treatment with Tris-buffer at pH 8.3 followed by DNA uptake in the presence of PEG. Between 2 and 5 µg plasmid DNA was used, resulting in $(1-2) \times 10^2$ transformants. The entire procedure was performed under anaerobic conditions using an anaerobic chamber.

Clostridium thermocellum has been used as recipient in genetic studies by Tsoi *et al.*, 1987. The mesophilic plasmid vectors pHV33 and pMK419 (Table 2) were introduced into *C. thermocellum* by protoplast transformation. pHV33 and pMK419 are *B. subtilis*-*E. coli* shuttle vectors based on pUB110, pC194 and pUC9/pUC19. The protoplasts were stabilized by addition of sorbitol, but the number of regenerated protoplasts remained extremely low. Catalase treatment of the surface of agar plates improved the regeneration process dramatically, eventually resulting in 0.01-10% regeneration. Approximately 7 µg plasmid DNA produced 70 transformants.

Plasmid name	Size	Genetic marker	Gram positive origin	Hybrid description	Replicates in (reference)	Temperature Maximum
pUB110	4.5 kb	Km ^r	<i>S. aureus</i>		<i>T. thermohydrosulfuricus</i> DSM 568 (Soutschek-Bauer et al., 1985)	55 °C
pGS13	2.9 kb	Km ^r , Cm ^r	<i>S. aureus</i>	PUB110 + pC194	<i>T. thermohydrosulfuricus</i> DSM 568 (Soutschek-Bauer et al., 1985)	55 °C
pHV33		Km ^r , Tc ^r , Ap ^r	<i>S. aureus</i>	PUB110 + pBR322	<i>C. thermocellum</i> F7 (Tsoi et al., 1987)	60 °C
pMK419	5.6 kb	Km ^r , Ap ^r	<i>S. aureus</i>	PMK4 (pC194+pUC9) + pUC19	<i>C. thermocellum</i> F7 (Tsoi et al., 1987)	60 °C
PCTC1	7.2 kb	Em ^r	<i>S. lactis</i>	PAMβ1 and Gram negative repli-con R2	<i>Ta. thermosaccharolyticum</i> (Klapatch et al., 1996)	60 °C
pIKM1	6.3 kb	Km ^r	<i>B. subtilis</i>	PIMP1 (pIM13+pUC9) + kan ^r from <i>S. faecalis</i> .	<i>Thermoanaerobacterium</i> sp. Strain JW/SL-YS485 (Mai et al., 1997)	60 °C
pRUKM1	7.2 kb	Km ^r	<i>B. stearothermophilus</i>	PRP9 + pUC18 + kan ^r from <i>S. faecalis</i> .	(Mai et al., 2000)	60 °C
pUXK	7.8 kb	Km ^r	Integration vector	PUC19 + kan ^r from <i>S. faecalis</i> + xynA	(Mai et al., 2000)	60 °C
pCL1	11.9 kb	Cm ^r	Thermophilic <i>Clostridium</i>	PTA688L + pBR328 (Cm ^r)	Thermophilic <i>Clostridium</i> (Kurose et al., 1989a)	No Transformation
pCS1	7.2 kb	Cm ^r	Thermophilic <i>Clostridium</i>	pTA688S + pBR328(Cm ^r)	Thermophilic <i>Clostridium</i> (Kurose et al., 1989a)	60 °C

Table 2. Candidate vector systems for extremely thermophilic anaerobic bacteria. Km^r kanamycin resistance, Cm^r chloramphenicol resistance, Tc^r tetracycline resistance, Sm^r streptomycin resistance, Em^r erythromycin resistance. * Plasmids originally isolated from *S. aureus* and shown here replicate in *B. subtilis* and are in most cases maintained in *B. subtilis*.

The transformation experiments done by Soutschek-Bauer and Tsoi were all based on plasmids from Gram-positive mesophilic microorganisms. This type of plasmids is often referred to as Gram positive replicons. Stability of the plasmids decreased with increasing temperature, and the plasmids became difficult to recover.

The first report using a replicon originally isolated from an extremely thermophilic organism came from Kurose et al. (Kurose *et al.*, 1989a). Thermophilic *Clostridia* were screened for plasmids and these plasmids were eventually used for the development of thermophilic *Clostridia-E. coli* shuttle vectors.

The plasmids pCL1 11.2 kb and pCS1 7.2 kb (table 2) were introduced into a poorly described thermophilic *Clostridium* by PEG-mediated transformation of intact cells. 1 µg of plasmid pCS1 produced approximately 100 chloramphenicol resistant colonies. pCS1 could be recovered from 60 °C cultures and visualized on agarose gels. Based on restriction enzyme digestion profiles, these recovered plasmids were apparently identical to the original plasmid. The transformants were, however, unstable, and the resistance towards chloramphenicol was easily lost upon repeated cultivation. Stability of the transformants was highly dependent on the choice of recipient. The plasmid pCL1 failed to produce any chloramphenicol resistant colonies and consequently no transformation had taken place.

The mesophilic plasmid pCTC1 (Table 2), based on pAMβ1 and the Gram-negative R2, replicon was used to electrotransform *Thermoanaerobacterium thermosaccharolyticum* (Klapatch *et al.*, 1996). Prior to electroporation the cells were washed in ice-cold water and re-suspended in 20% glycerol. The transformation efficiency was 5 transformants/10 µg DNA. When pCTC1 was prepared from transformed *Ta. thermosaccharolyticum* the transformation efficiency increased to 52 transformants/µg DNA. Two liters of cell culture were used to recover plasmids from transformed cells. The extracted plasmid-preps were used in Southern blotting experiments to verify transformation. In addition, *E. coli* was re-transformed with *Ta. thermosaccharolyticum* extracted pCTC1. Plasmids from *E. coli* were characterized by restriction enzyme digestion and found to be identical to the original plasmid. Plasmid sta-

bility decreased with increasing temperature and no plasmid bands were detectable at 60 °C.

pIKM1, pRUKM1 and pUXK (integration vector) (Table 2), containing the thermostable kanamycin cassette from *S. aureus*, was used to electro-transform *Thermoanaerobacterium* sp. strain JW/SL-YS485 (Mai *et al.*, 1997; Mai *et al.*, 2000). The plasmids could be maintained under selective pressure at 60 °C. Transformation frequencies varied between 10 and 100 transformants/μg DNA, and apparently no consistent or significant change in transformation efficiency was observed when electro-poration parameters were varied over a wide range. One litre of cell culture was used to recover the plasmids. Recovered plasmids could be visualised on agarose gels and used to re-transform *E. coli*. Recovered and re-transformed plasmids were identical to the original plasmids, as determined by restriction enzyme digestion. However, plasmid recovery was difficult and large volumes were essential in order to visualise recovered plasmids, suggesting that the plasmids were present in low numbers at elevated temperatures.

Transformation of thermophilic anaerobic microorganisms has been conducted with mesophilic plasmids only, with the exception of two plasmids (pCL1 and pCS1). They are all difficult to recover and they are quite unstable, and even the thermostable plasmids were only present in low numbers when the host cells were cultivated at elevated temperatures. The highest temperature reported for recovery of plasmids from transformants is 60 °C. Since no sequence information is available for the thermostable plasmids used in these reported transformation experiments, their use in the construction of shuttle vectors must be based solely on their restriction maps. This means that shuttle vector design can not take account of regions of the plasmid with potential importance for replication and copy number control, and their usefulness is thus largely a question of chance.

30 SUMMARY OF THE INVENTION

The invention provides plasmids and shuttle vectors, capable of self-replication in extremely thermophilic anaerobic micro-organisms. The inven-

tion thereby facilitates the establishment of genetic systems for these industrially important micro-organisms. The new plasmids were isolated following a plasmid screen of a number of extremely thermophilic anaerobic microbial strains. These new plasmids, designated pBAS2, 3653 bp (Fig. 1, pBAS 5 1863 bp (Fig. 2) and pBAL 8294 bp (Fig. 3) were all isolated from the extremely thermophilic anaerobic microorganism, *Anaerocellum thermophilum* DSM6725, which has a temperature optimum of 72-75°C. All three isolated plasmids, namely pBAS2, pBAS and pBAL have been characterized by DNA sequencing, that has shown that each of these plasmids is new isolates. The 10 DNA sequence of pBAS showed it to be a sub-species of pBAS2, whereby nucleotide sequence 1 – 1863 of pBAS corresponds to nucleotide sequence 638-2501 of pBAS2. Both pBAS2 and pBAL contain ORFs encoding proteins which show homology to known proteins of potential industrial interest. Thus proteins encoded by ORF 11 and ORF 61 of pBAS2 (Fig. 1) show homology 15 to a recombinase protein and a replication protein, respectively. In pBAL the ORFs 31, 33 and 34 encode proteins showing homology to a DNA polymerase/DNA repair protein, a sigma factor, and a replication protein, respectively.

The present invention provides isolated and sequenced plasmids, pBAS2, 20 pBAS and pBAL, originating from extremely thermophilic microorganisms which are stable at growth temperatures of at least up to 75 °C and are thus ideally suited for genetic expression systems intended to operate at high temperatures.

A thermostable shuttle vector, pEAKS (Fig. 4) has been constructed from a 25 fusion of pBAS and pBluescript SK+/KanR, that can be transformed into thermophilic anaerobic kanamycin-sensitive strains, conferring kanamycin resistance. The pEAKS vector includes replication origins, to facilitate propagation in both extreme thermophilic microbial hosts as well as *E. coli*, two drug resistance markers for selection of transformants and a multiple cloning 30 site, and thereby provides a new and useful genetic system for enhancing the industrial utility of extreme thermophilic anaerobic microorganisms.

Further, the invention provides an extremely thermophilic host cell transformed with the pEAKS vector, which is stably maintained during growth at 70 °C.

5 Since pBAS2, pBAS and pBAL are derived from an extremely thermophilic anaerobic microorganism with a temperature optimum at 72-75 °C, they are the most thermostable plasmids ever isolated from extremely thermophilic anaerobic microorganisms.

10 Knowledge of the DNA sequence of the plasmids pBAS2, pBAS and pBAL greatly facilitates their use in the construction of various vectors. In addition pBAS2, pBAS and pBAL are small plasmids, and thus well suited for the construction of genetic systems for extremely thermophilic anaerobic microorganisms.

15 Expression vectors can be constructed from the pBAS2, pBAS or pBAL plasmids provided by the invention by inserting heterogenous protein expression cassettes into the vector sequence. The expression vectors can be transformed into appropriate host cells for the production of heterogenous proteins, which are recovered and subsequently applied for industrial processes at temperatures of 60-75 °C. Host cells, which have acquired new catalytic properties following transformation with the expression vector, may
20 also be used directly in biological processes at temperatures of 60-75 °C. Furthermore the yield and/or stability of thermophilic proteins may be enhanced by expression at high temperatures. In cases where it is desired to screen variants thermophilic proteins for improved catalytic properties, the expression vectors and host cells provided by the invention have particular
25 utility, since the expressed proteins can be screened *in vivo*.

DEFINITIONS

PCR

Polymerase Chain Reaction. *In vitro* amplification of DNA fragments.

Cloning

Construction of a recombinant DNA molecule and amplification of this molecule in *Escherichia coli*.

Restriction enzyme

- 5 Restriction enzymes recognise specific sequences on DNA molecules and cuts/digests the DNA molecule into smaller fragments.

Restriction map

Physical map of a plasmid based on the location of restriction sites using different restriction enzymes.

10 Transformation

Introduction of foreign DNA, encoding a new phenotype, into a cell. Selection of viable cells expressing the new phenotype.

PEG transformation

The use of the chemical polyethylene glycol to media transformation.

15 Electro-transformation

The use of electric fields to mediate transformation.

Protoplast-transformation

Partial removal of the outer cell wall to facilitate transformation.

Open Reading Frame (ORF)

- 20 Putative protein encoding sequence on DNA molecules.

Plasmid

Self-replicating extrachromosomal genetic element.

Vector

A vector is identical to a plasmid. The term vector is used to emphasise that the plasmid is used for genetic engineering.

Plasmid replication

- 5 Multiplication of plasmids inside the cell.

Rolling circle plasmid

- 10 Replication mode of some circular plasmids. A nick is introduced in one of the DNA strands. The free end serves as template for DNA replication, while the plasmid continues to unwind. Replication proceeds until the entire plasmid has been replicated.

Theta plasmid

Replication mode, where no nicks are introduced. The two DNA strands separate and the free end(s) is used as template for DNA replication, while the two strands continue to separate.

- 15 Mesophilic microorganisms

Microorganisms with temperature optimum for growth from 20-37 °C.

Moderate thermophilic microorganisms

Microorganisms with temperature optimum for growth from 37-50 °C.

Thermophilic microorganisms

- 20 Microorganisms with temperature optimum for growth from 50-65 °C.

Extremely thermophilic microorganisms

Microorganisms with temperature optimum for growth from 65-80 °C.

Single stranded DNA (ssDNA)

DNA molecule where the double stranded DNA has separated into single strands.

Origin

- 5 Section of plasmid that is mandatory for replication.

ss-Origin

Replication initiates at a single stranded DNA sequence within the replication origin, after the two DNA strands have separated.

Cryptic plasmid

- 10 Plasmid with no known function.

Plasmid-preps

Extraction of plasmid DNA from transformed cells.

Similarity

- 15 Relationship between DNA or protein molecules not necessarily derived from a common background.

Replication initiation

Physical location on the plasmid where replication initiates.

Replication terminus

Physical location on the plasmid where replication terminates.

- 20 Genetic engineering

Introduction of foreign DNA into a cell.

Metabolic engineering

Introduction of foreign DNA into a cell with the purpose of altering the metabolism of said cell.

Genetic system

- 5 Tools to introduce and stably maintain foreign DNA into a cell.

Transformation efficiency

Number of transformed cells per μg of DNA.

Shuttle vector

- 10 Plasmid, that can be replicated and maintained in two different microorganisms.

Covalently closed circular plasmid

Circular compact topological form of a plasmid, migrating fastest in agarose gels.

Southern blotting

- 15 Sensitive technique to identify small amounts of known DNA in a sample. The DNA sample is applied to a membrane, which is incubated in a defined hybridization solution containing a labelled DNA-probe, complementary to a known DNA sequence.. If the known DNA is present in the sample, then the labelled DNA-probe will recognise it and bind to it. Because the DNA-probe is
20 labelled, for example with a radioactive isotope, it can be visualised by an X-ray autoradiography.

Agarose gels

- 25 Gel of agarose used to separate DNA molecules of different sizes. When an electrical field is applied to DNA embedded in an agarose gel, the DNA migrates according to the length of the DNA molecule. Long DNA molecules

migrate slower than small DNA molecules. The migration of DNA molecules in the agarose gel can be visualised and the migration patterns analysed.

DNA sequencing

5 Technique to determine the specific sequence of deoxy-ribonucleic acids (DNA) of a DNA molecule.

Heterogenous protein

10 Heterogenous protein is a protein encoded by a gene originating from a host cell genome which is heterogenous to the new host in which the protein is expressed. Such proteins include, but are not restricted to, the enzymes alcohol dehydrogenase, carbohydrase, amylase, cellulase, beta-glucanase, beta-glucosidase, alpha-glucosidase, xylanase, oxidoreductase, protease and lipase.

BRIEF DESCRIPTION OF FIGURES

15 **Figure 1**

Physical map of plasmid pBAS2. Broad arrows indicate open reading frames, where some have been assigned a putative function. Restriction sites for restriction enzymes cutting the plasmid at 3 or less positions are shown on the map. The sub-species plasmid pBAS, nucleotides 1-1863, corresponds to
20 nucleotides 638 to 2501 of pBAS2.

Figure 2

Physical map of plasmid pBAS. Base pairs are counted clockwise from the *EcoRI* site. Black arrows indicate open reading frames with no homology to
25 known genes. Grey arrows are open reading frames, which have been assigned a putative function.

Figure 3

Physical map of plasmid pBAL. Base pairs are counted clockwise from the *EcoRI* site. Black arrows indicate open reading frames with no homology to known genes. Grey arrows are open reading frames, which have been assigned a putative function.

Figure 4

Kanamycin resistance gene was obtained from plasmid pUB110, originally found in a *Staphylococcus aureus* strain. The resulting thermostable shuttle vector pEAKS contains replication origin from the *E. coli* vector pBluescript SK⁺ and replication origin from the thermostable plasmid pBAS. As selective markers ampicillin resistance expressed in *E. coli* and kanamycin resistance expressed in thermophilic anaerobic microorganisms are included.

DETAILED DESCRIPTION OF THE INVENTION

- 15 *Anaerocellum thermophilum* DSM6725 is a strict anaerobic microorganism with a temperature optimum at 72-75 °C which is freely available from a public culture collection at DSM - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, D-3300 Braunschweig, Germany, under the accession number DSM6725.
- 20 *Anaerocellum thermophilum* DSM6725 was found to harbour three plasmids, pBAS2, pBAS and pBAL, none of which have previously been reported. pBAS2 is 3653 bp (Fig. 1) with SEQ. ID NO: 15, pBAS is 1863 bp (Fig. 2) with SEQ ID NO: 22 and pBAL is 8294 bp (Fig. 3) with SEQ. ID NO: 1, and the GC content of the plasmids is 43%, 43% and 39%, respectively.
- 25 Sequencing of pBAS2 revealed 8 open reading frames, where ORFs 11 and 61 showed homology to genes encoding a recombinase protein and a replicase protein, respectively. The following ORFs of pBAS2 have been designated the sequence ID numbers: SEQ ID NO: 16 = recombinase, SEQ ID NO: 17 = replicase (ORF 11), SEQ ID NO: 18 = ORF41, SEQ ID NO: 19 =
- 30 ORF42, SEQ ID NO: 20 = ORF43, SEQ ID NO: 21 = ORF44. Furthermore,

a detailed analysis of the DNA structure of pBAS2 revealed a putative replication initiation site and replication terminus.

Sequencing of pBAS revealed that it was a sub-species plasmid (Fig. 2), showing complete sequence identity to a portion of pBAS2, and hence containing only some of the ORFs predicted in pBAS2. Nucleotides 1-1863 in pBAS correspond to nucleotides 638-2501 in pBAS2.

Plasmid pBAL contains 14 open reading frames, of which three showed similarity to known proteins. One ORF showed homology to replication proteins found on various *Staphylococcus* plasmids. These plasmids, pNVH97, pSK1, pI9789 and pIP680, all replicate via the theta mechanism, indicating that pBAL also employs this type of replication mechanism. Significant homology was also found to sigma factor K from *Bacillus thuringiensis* and to DNA repair protein from *Campylobacter jejuni*. The ORFs of pBAL have been designated the following sequence ID numbers: SEQ ID NO: 2 = ORF 31 (DNA polymerase); SEQ ID NO: 3 = ORF32; SEQ ID NO: 4 = ORF 33 (sigma factor); SEQ ID NO: 5 = ORF 34 (replicase); SEQ ID NO: 6 = ORF41; SEQ ID NO: 7 = ORF42; SEQ ID NO: 8 = ORF43; SEQ ID NO: 9 = ORF44; SEQ ID NO: 10 = ORF51; SEQ ID NO: 11 = ORF52; SEQ ID NO: 12 = ORF53; SEQ ID NO: 13 = ORF61; SEQ ID NO: 14 = ORF62.

The pBAS2 and pBAL plasmids found in *Anaerocellum thermophilum* DSM6725, showed no nucleotide sequence similarity to each other. Similarly the proteins encoded by the ORF's in the pBAS2 and pBAL plasmids did not reveal any similarity.

The described plasmids, pBAS2, pBAS and pBAL, are derived from an extremely thermophilic anaerobic microorganism with a temperature optimum at 72-75 °C. This makes them the most thermostable plasmids isolated from extremely thermophilic anaerobic microorganisms. PBAS2, pBAS and pBAL are therefore expected to be considerably more thermostable than the plasmids reported by others, since most of the latter plasmids are isolated from cultures with temperature optimum at 60 °C.

As opposed to all other plasmids reported, with the exception of pNB2, the full DNA sequence of pBAS and pBAL has been determined. Knowledge of the DNA sequence allows for identification of replication regions and proteins with potential importance for plasmid copy number control, thus providing a far more solid basis for the construction of vectors. In addition both plasmids are small and easy to handle, making them especially useful as the basis for genetic systems for extremely thermophilic anaerobic microorganisms.

The thermostable expression plasmid (pEAKS) is provided according to the invention as an example of several possible expression plasmids. It is to be understood that pEAKS is intended only to be illustrative and in no way limitative. A person skilled in the art will readily understand that once the extremely thermophilic plasmids pBAS2, pBAS and pBAL have been isolated and sequenced, one may derive extremely thermophilic expression plasmids based on these genetic elements.

Marker genes are included in plasmid and shuttle vectors to facilitate the identification and selection of host cells transformed with the vector/plasmid. Marker genes may encode proteins conferring resistance to antibiotics, such as an ampicillin resistance gene, a kanamycin resistance gene, a tetracycline resistance gene, or an erythromycin resistance gene. Alternatively they may encode visual markers, such as those based on beta-galactosidase marker gene expression. Particularly useful marker genes are those which encode proteins which can be actively expressed in the host cell of interest. In the case of an extreme thermophilic anaerobic host cell, it is important to select a marker gene encoding a heat-stable protein. The kanamycin resistance gene is particularly useful in this context, since both the kanamycin resistance protein and the kanamycin antibiotic are rather heat-stable. This is particularly important for plasmid stability, where the maintenance of the plasmid in the extreme thermophilic anaerobic host may depend on antibiotic selection pressure. The inclusion of more than one selection marker is particularly useful in shuttle vectors, which should be maintained in two host cell types.

Multiple cloning sites are included in the shuttle vector provided by the invention to facilitate the introduction of additional expression cassettes, which may comprise marker genes or heterogenous protein expression cassettes.

Expression vectors of the invention are used for transformation of host cells, thus constructing microorganisms suitable for industrial production processes at temperatures of 60-75 °C for the manufacture of recombinant proteins which are subsequently recovered. Said proteins can be an alcohol dehydrogenase, a carbohydrase, an amylase, a cellulase, a beta-glucanase, a beta-glucosidase, an oxidoreductase, a protease, an oxidoreductase, or a lipase and they may be applied in a wide variety of industrial processes characterised by high temperatures. Examples of recombinant proteins used in industrial processes are cellulases, amylases, xylanases, β -galactosidases, β -glucosidases, α -glucosidases, and glucanases.

Expression vectors of the invention are also used for transformation of host cells to modify the catalytic properties of the cells, thus improving their performance as biocatalysts in industrial processes at temperatures of 60-75 °C. This is known as metabolic engineering and the modified cells may be used for e.g. degradation of biopolymers such as lignocellulosic materials.

Of course the expression vectors of the invention may also be effective at lower temperatures between room temperature and 60 °C.

EXAMPLES

20

Example 1. Screening of thermophilic anaerobic micro-organisms for plasmids.

Bacterial strains and plasmids.

Screening for plasmids was done on thermophilic anaerobic isolates from our strain collection and on DSM6725, which was purchased from DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany.

Media and Culture conditions.

Anaerobic cultures were grown in a BA medium as previously described (Angelidaki *et al.*, 1990), but the medium was amended with 1 g/l yeast extract (Difco) and cysteine was not added. The medium was reduced with 0.25 g/l sodium sulfide. Appropriate carbon sources, cellulose, glucose, xylose, mannose or galactose was added at 5 g/l and incubation was at 70 °C and pH 6.8.

Plasmid Extraction procedures.

60 ml overnight cultures were used for extraction of plasmids. Cells were harvested by centrifugation and washed once in 10 ml TE pH 8.0, pelleted and resuspended in TE containing 10 mg/ml lysozyme. The cells were lysed according to the method described by (O'Sullivan *et al.*, 1993), starting from step 2, and the plasmid DNA was subsequently purified according to standard molecular methods (Ausubel, 1987).

Example 2. Isolation and characterisation of pBAS2, pBAS and pBAL.

The plasmids pBAS2, pBAS and pBAL were detected and isolated from *Anaerocellum thermophilum* DSM6725 as described in example 1.

The presence of the plasmids, designated pBAS2, pBAS and pBAL, in *Anaerocellum thermophilum* DSM6725 was verified by agarose gel electrophoresis using 1.2 % agarose, stained with ethidium bromide and illuminated by UV light.

Cloning and sequencing.

For cloning purposes the cloning vector pBluescript SK⁺ and cloning host *E. coli* DH5- α was used.

Initially all three plasmids were characterized by digestion with 23 different restriction endonucleases. Restriction endonuclease digests were carried out according to the suppliers recommendation (New England Biolabs). Plasmids

pBAS and pBAL showed a unique *EcoRI* site and were cloned into *EcoRI*/SAP (shrimp alkaline phosphatase, Boeringer Mannheim) treated pBluescript SK⁺ vector. pBAS was sequenced by "plasmid walking". In the first round of sequencing the pBluescript KS⁺ primers T3 and T7 were used and the sequence information obtained was then used to design sequencing primers for the next round of sequencing. pBAL revealed an additional unique *SaI* site, which was used for subcloning into *SaI*/*EcoRI* treated pBluescript KS⁺, eventually yielding two plasmids pBAAN 4.2 kb and pBAB 4.0 kb. These two plasmids were then sequenced by "plasmid walking". Restriction mapping and sequencing of pBAS and pBAS2 revealed that pBAS was a plasmid sub-species of pBAS2. Since part of the nucleotide sequence of pBAS2 was homologous to pBAS, the sequence of the unknown part of the plasmid was determined according to the following methods. The unknown part of pBAS2 was amplified using the following primers:

pBAScw1: cccctgcaggctattctgtgggtagtgag (SEQ ID NO:23) pBASccw1: cccctgcagccttgaaacgcaaataagggtcc (SEQ ID NO:24)

This fragment was then sequenced by plasmid walking starting with primers:

pBASeco_ccw1:ccttcacacgagcaggattcc (SEQ ID NO:25)

pBASeco_cw1:gcttttgattgcaagatggtattgc (SEQ ID NO:26)

The known part of the plasmid, common to pBAS, was verified by restriction enzyme digestion. The nucleotide residues 1 – 1863 of pBAS correspond to nucleotide residues 638 to 2501 of pBAS2.

Figures 1, 2 and 3 show the three plasmids and the complete sequences of pBAS2, pBAS and pBAL are given in SEQ ID NOs 15, 22 and 1, respectively.

25

Example 3. Derivation of pEAKS expression vector from pBAS.

pBAS was used as the thermostable backbone for construction of a vector replicating in thermophilic anaerobic microorganisms. A 1841 bp fragment of

plasmid pBAS, including potential replication region and replication protein was amplified by PCR using the following primer sequences:

pBAS upper1-20:gcgggatccgctcgacgaattcgaaaagcagataag (SEQ.ID NO:27)

pBAS lower184:gggctgcagtcctcgtttaacaactaca (SEQ. ID NO:28)

5 thereby introducing unique *Bam*H1 and *Pst*I restriction sites in the PCR product. Subsequently the *Bam*H1, *Pst*I digested pBAS PCR product was cloned into pBluescript SK⁺/Kan^R which was restriction enzyme digested with *Bam*H1 and *Pst*I. pBluescript SK⁺/Kan^R was constructed by introducing the kanamycin resistance gene from pUB110 into the pBluescript vector. The
10 Kan^R gene in pUB110 was excised by digestion with the restriction enzymes, *Mun*I and *Nci*I, treated with T4 DNA polymerase, and then ligated into *Eco*RV digested pBluescript. The resulting thermophilic expression plasmid pEAKS is shown in Fig. 3.

The kanamycin-sensitive, thermophilic anaerobic (*Clostridium*-like) isolate
15 (Sonne *et al.* 1993), X7B and the kanamycin sensitive strain *Thermoanaerobacter mathranii* A3M4 were transformed to kanamycin resistance by the thermostable shuttle vector pEAKS (including the kanamycin resistance gene), by using electroporation and PEG mediated transformation.

Initially transformations were conducted at 55 °C and kanamycin resistant
20 colonies were selected. The selected colonies were cultivated at both 55 °C and at 70 °C in selective liquid medium (kanamycin added), without any apparent differences in cell growth rates. Following growth of the transformants in selective liquid medium for 20 generations, they were still stable.

25 **Example 4.**

Since both plasmids pBAS2, pBAS and pBAL are small and easy to handle, they are ideal as thermophilic backbone for constructing plasmid vectors for genetic engineering and metabolic engineering of thermophilic microorganisms. For example for improving ethanol yield from thermophilic anaerobic

microorganisms, by introduction of additional copies of genes responsible for ethanol formation (e.g. *Adh*, Alcohol dehydrogenase).

In addition the plasmids pBAS2, pBAS and pBAL can be used as basis for constructing expression vectors for over-expressing industrially important thermostable enzymes e.g. cellulases, amylases, xylanases, β -galactosidases, β -glucosidases etc. The shuttle vector, pEAKS, which incorporates pBAS, can be used for the expression of thermostable enzyme. The inducible promoter of the *xylA* gene from *Thermoanaerobacter ethanolicus* (Erbeznik *et al.* 1998), is cloned into the *Pst*I or *Eco*R1 site of pEAKS, together with a downstream MCR sequence, and *xyl A* terminator. A gene of interest, encoding for example an ethanol dehydrogenase, can be cloned into the MCS of the plasmid, and subsequently transformed into the extreme thermophilic anaerobic microbial host. Expression of the heterogenous protein from the gene inserted into pEAKS can be regulated by the level of xylose or glucose added to the medium.

Example 5.

Sequence analysis revealed that pBAL encodes at least three proteins: 1) a DNA polymerase/DNA repair protein 2) a sigma K factor and 3) a replication protein. Their gene products, especially DNA polymerase/DNA repair protein encoded by pBAL can be used in DNA manipulating processes used in molecular biology techniques. For example in the PCR reaction.

Example 6.

This example disclosed the construction of a high copy number shuttle vector based on pBAS2 comprising a kanamycin resistance gene.

The entire pBAS2 plasmid is amplified using pBAScw1 and pBASccw1 primers. The resulting molecule is digested with *Pst* I restriction enzyme and inserted into the *Pst* I digested *E. coli* vector pBluescript SK+. The resulting plasmid is called pBlueBAS2.

pUB110 is digested with Mun I and Nci I and the fragment containing the kanamycin resistance gene is treated with T4 DNA polymerase. The kanamycin resistance cassette is then inserted into the EcoRV site of pBlueBAS2.

5

Example 7.

This example disclosed the construction of a low copy number shuttle vector based on pBAS2 comprising a kanamycin resistance gene.

10

The entire pBAS2 plasmid is amplified using pBAScw3 and pBASccw3 primers.

pBAScw3: ccccgatcccctattcttgtgggtagtgcag

15

pBASccw3: ccccgatccccttgaaacgcaaataaggtcc

The product is digested with BamH I and inserted into the low copy number plasmid pOU71 digested with BamH I. The resulting plasmid is called p71BAS2.

20

pUB110 is digested with Mun I and Nci I and the fragment containing the kanamycin resistance gene is treated with T4 DNA polymerase. The kanamycin resistance cassette is then inserted into the EcoR I and T4 DNA polymerase treated p71BAS2. The resulting plasmid is called p71BAS2K.

25

REFERENCES

- Japanese unexamined published patent application no. JP 60-186287 A, Agency of Industrial Science and Technology (AGEN). New plasmid derived from anaerobic thermophilic bacteria - is used as vector in breeding anaerobic thermophilic bacterial. September 21, 1985a.
- Japanese unexamined published patent application no. JP 60-186277 A, Agency of Industrial Science and Technology (AGEN). New microorganism Thermoanaerobacter No. 62 - contains new plasmid useful as vector in recombinant DNA experiment. September 21, 1985b.
- Japanese unexamined published patent application no. JP 60-186278 A, Agency of Industrial Science and Technology (AGEN). New microorganism Thermoanaerobacter No.29 - containing new plasmid characterised by specified restriction end nuclease cleavage map. September 21, 1985c.
- Japanese unexamined published patent application no. JP 60-186286 A, Agency of Industrial Science and Technology (AGEN). New plasmid - derived from anaerobic thermophilic bacteria. September 21, 1985d.
- Japanese unexamined published patent application no. JP 60-186285 A, Agency of Industrial Science and Technology (AGEN). New plasmid derived from anaerobic thermophilic bacteria - is used as vector in DNA recombination using thermophilic bacteria as most. September 21, 1985e.
- Japanese unexamined published patent application no. JP 60-186279 A, Agency of Industrial Science and Technology (AGEN). New Thermoanaerobacter No. 49 - containing new plasmid (pSH 49), useful as vector in recombinant DNA experiment. September 21, 1985f.
- Ammendola, S., Raia, C.A., Caruso, C., Camardella, L., D'Auria, S., De Rosa, M. & Rossi, M. Thermostable NAD⁺ dependent alcohol dehydrogenase from *Sulfolobus solfataricus*: gene and protein sequence determination and relationship to other alcohol dehydrogenases (1992), *Biochemistry* **31**, 12514-12523.

- Angelidaki, I., Petersen, S.P. & Ahring, B.K. Effects of lipids on the anaerobic digestion and reduction of lipid inhibition upon addition of bentonite (1990), *Applied Microbiology and Biotechnology* **33**, 469-472.
- 5 Ausubel, F.M. Current protocols in molecular biology (1987) Ed: F.M. Ausubel, R. Brent, R.E. Kingston, D.D. More, J.G. Seidman, J.A. Smith, K. Stuhl, John Wiley & Sons, Inc.
- Belogurova, N.G., Mosolova, T.P., Kalyuzhnyy, S.V. & Varfolomeyev, S.D. (1991), *Appl. Biochem. Biotechnol.* **27**(1), 1-8.
- 10 Delver, E.P., Belogurova, N.G., Tupikova, E.E., Varfolomeyev, S.D. & Belogurov, A.A. (1996), *Molecular & General Genetics* **253**, 166-172.
- Erbeznik M, Dawson KA, Strobel HJ. Cloning and characterization of transcription of the xylAB operon in *Thermoanaerobacter ethanolicus*. (1998), *J Bacteriol.* **180**(5), 1103-9
- Gruss, A. & Ehrlich, S.D. (1989), *Microbiol. Rev.* **53**(2), 231-241.
- 15 Janzen, T., Kleinschmidt, J., Neve, H. & Geis, A. (1992), *FEMS Microbiology Letters* **74**(2-3), 175-180.
- Kalyuzhnyi, S.V., Mosolova, T.P., Belogurova, N.G. & Varfolomeev, S.D. Plasmids from thermophilic clostridia (1992), *Appl. Biochem. Biotechnol.* **34-35**, 383-389.
- 20 Klapatch, T.R., Guerinot, M.L. & Lynd, L.R. (1996), *Journal of Industrial Microbiology* **16**(6), 342-347.
- Kurose, N., Miyazaki, T., Kakimoto, T., Yagyu, J., Uchida, M., Obayashi, A. & Murooka, Y. (1989a), *Journal of Fermentation and Bioengineering* **68**, 371-374.
- 25 Kurose, N., Yagyu, J., Ohmori, M., Matsumoto, M., Ohsato, K., Yamada, T., Uchida, M. & Obayashi, A. (1989b), *Agricultural and Biological Chemistry* **53**, 3179-3186.

- Lynd, L.R. Production of ethanol from lignocellulosic materials using thermophilic bacteria (1989), in A Fiechter (Editor) *Advances in Biochemical Engineering/Biotechnology* **38**, 1-52
- 5 Mai, V., Lorenz, W.W. & Wiegel, J. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance (1997), *FEMS Micro-biology Letters* **148**(2), 163-167.
- Mai, V. & Wiegel, J. (2000), *Appl. Environ. Microbiol.* **66**(11), 4817-4821.
- Muller, R.E., Ano, T., Imanaka, T. & Aiba, S. (1986), *Mol. Gen. Genet.* **202**(1), 169-171.
- 10 O'Sullivan, D.J. & Klaenhammer, T.R. Rapid mini-prep isolation of high-quality plasmid DNA from *Lactococcus* and *Lactococcus* spp. (1993), *Applied and Environmental Microbiology* **59**(8), 2730-2733.
- Sonne-Hansen, J., Mathrani, I.M., & Ahring, B.K. Xylanolytic anaerobic thermophiles from Icelandic hot-springs (1993) *App. Microbio. Biotech.* **38**, 537-
15 541.
- Soutschek-Bauer, E., Hartl, L. & Staudenbauer, W.L. Transformation of *Clostridium thermohydrosulfuricum* DSM 568 with plasmid DNA (1985), *Biotechnology Letters* **7**(10), 705-710.
- 20 Tsoi, T.V., Chuvil'skaya, N.A., Atakishieva, Ya.Yu., Dzhavakhishvili Ts. D., Akimenko, V.K., and Boronin, A.M. (1987), *Molekulyarnaya Genetika Mikrobiologiya i Virusologiya* No. 11, 18-23.
- Weimer, P.J., Wagner, W., Knowlton, S. & Ng, T.K. Thermophilic anaerobic bacteria which ferment hemicellulose: characterization of organisms and identification of plasmids (1984), *Archives of Microbiology*, **138**, 31-36.

CLAIMS

1. 1. A plasmid comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:15 and SEQ ID NO:22, a nucleotide sequence which has more than 80% identity to a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:15 and SEQ ID NO:22, and nucleotide fragments of SEQ ID NO:1, SEQ ID NO:15 and SEQ ID NO:22 wherein said fragments comprise at least 20 nucleotides.
2. A plasmid or DNA sequence according to claim 1, wherein said plasmid or nucleotide sequence is isolated from an extremely thermophilic microorganism or a thermophilic microorganism.
3. A plasmid or DNA sequence according to any one of claims 1-2, wherein said plasmid or nucleotide sequence is isolated from *Anaerocellum thermophilum*.
4. A plasmid according to any one of claims 1-3, which comprises a marker gene.
5. A plasmid according to claim 4, wherein said marker gene is a beta-galactosidase gene.
6. A plasmid according to claim 4, wherein said marker gene is a drug resistance gene.
7. A plasmid according to claim 6, wherein said marker gene is selected from the group consisting of an ampicillin resistance gene, a kanamycin resistance gene, a tetracycline resistance gene, an erythromycin resistance gene, and a chloramphenicol resistance gene.
8. A plasmid according to any one of claims 1-7, which comprises a multiple cloning site.

9. A plasmid according to claim 8, wherein said multiple cloning site comprises a restriction site for at least one restriction enzyme selected from the group consisting of EcoRI, HindIII, SacI, BamHI, XbaI, SalI and PstI.
- 5 10. A plasmid according to any of claim 1-9 for use as a shuttle vector, comprising a replication region for use in a secondary host different from the primary thermophilic host from which the plasmids with SEQ ID NOS 1, 15 and 22 have been isolated.
- 10 11. A plasmid according to claim 10 wherein the secondary host is a mesophilic microorganism, preferably E. coli.
12. A plasmid according to any of claims 1-11, comprising one or more expression control sequences.
13. Method of obtaining a shuttle vector according to any of claims 10 or 11, comprising:
- 15 a) digesting a plasmid according to any of claims 1-9 with one or more restriction enzymes to obtain a fragment of said plasmid comprising a replication region,
- b) digesting a plasmid suited for the secondary host with one or more restriction enzymes to obtain a fragment of said plasmid comprising a replication region, and
- 20 c) ligating said fragments to obtain a plasmid autonomously replicable in both the primary and secondary host.
14. A plasmid according to claim 11, which has more than 60%, preferably more than 70%, more preferably more than 80% nucleic acid identity to pEAKS.
- 25 15. A plasmid according to any one of claims 1-12 or 14, wherein a gene has been inserted which codes for a heterogenous protein.
16. A plasmid according to claim 15, wherein said heterogenous protein is an enzyme.
- 30

- 5 17. A plasmid according to claim 16, wherein said enzyme is selected from the group consisting of an alcohol dehydrogenase, a carbohydrase, an amylase, a cellulase, a beta-glucanase, a beta-glucosidase, an alpha-glucosidase, a xylanase, an oxidoreductase, a protease and a lipase.
18. A host cell transformed with a plasmid according to any one of claims 1-12 or 14-17.
19. A host cell according to claim 18, which is a thermophilic microorganism.
- 10 20. A host cell according to claim 19, which is an extremely thermophilic microorganism.
21. A host cell according to any one of claims 18-20, wherein said host cell is an anaerobic microorganism.
- 15 22. A host cell according to claim 18-21, wherein said host cell is a bacterium.
- 20 23. A host cell according to claim 22, which is a microorganism selected from the group consisting of *Thermoanaerobacter*, *Thermoanaerobacterium*, *Thermoanaerobium*, *Thermoanaerobacteroides*, *Anaerocellum*, *Caldicellusiruptor*, *Clostridium*, *Bacillus*, *Thermobacillus*, *Thermus* and *Thermotoga*.
24. A host cell according to claim 18, which is a mesophilic microorganism, preferably *E. coli*.
25. Use of a host cell according to any one of claims 18-24, for producing a protein and recovering said protein.
- 25 26. Use of a host cell according to any one of claims 18-25, wherein the expressed protein confers said host cell with a changed phenotype.

27. Use of a host cell according to any of claims 18-24 in a process of degradation or fermentation of biomaterial.
28. A protein encoded by any one of the open reading frames of SEQ ID NO:1 or open reading frames of SEQ ID NO:15.
- 5 29. A protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs:2-14 and 16-20.
30. A protein, characterised by being a variant of the DNA polymerase/DNA repair protein encoded by SEQ ID NO:2, wherein the amino acid identity between said protein and said DNA polymerase is at least 60%, preferably at least 80%, more preferably at least 90%, said protein exhibiting DNA polymerase activity.
- 10 31. Use of a DNA polymerase/DNA repair protein of any one of claims 29-30 for the synthesis or repair of DNA.

Figure 1

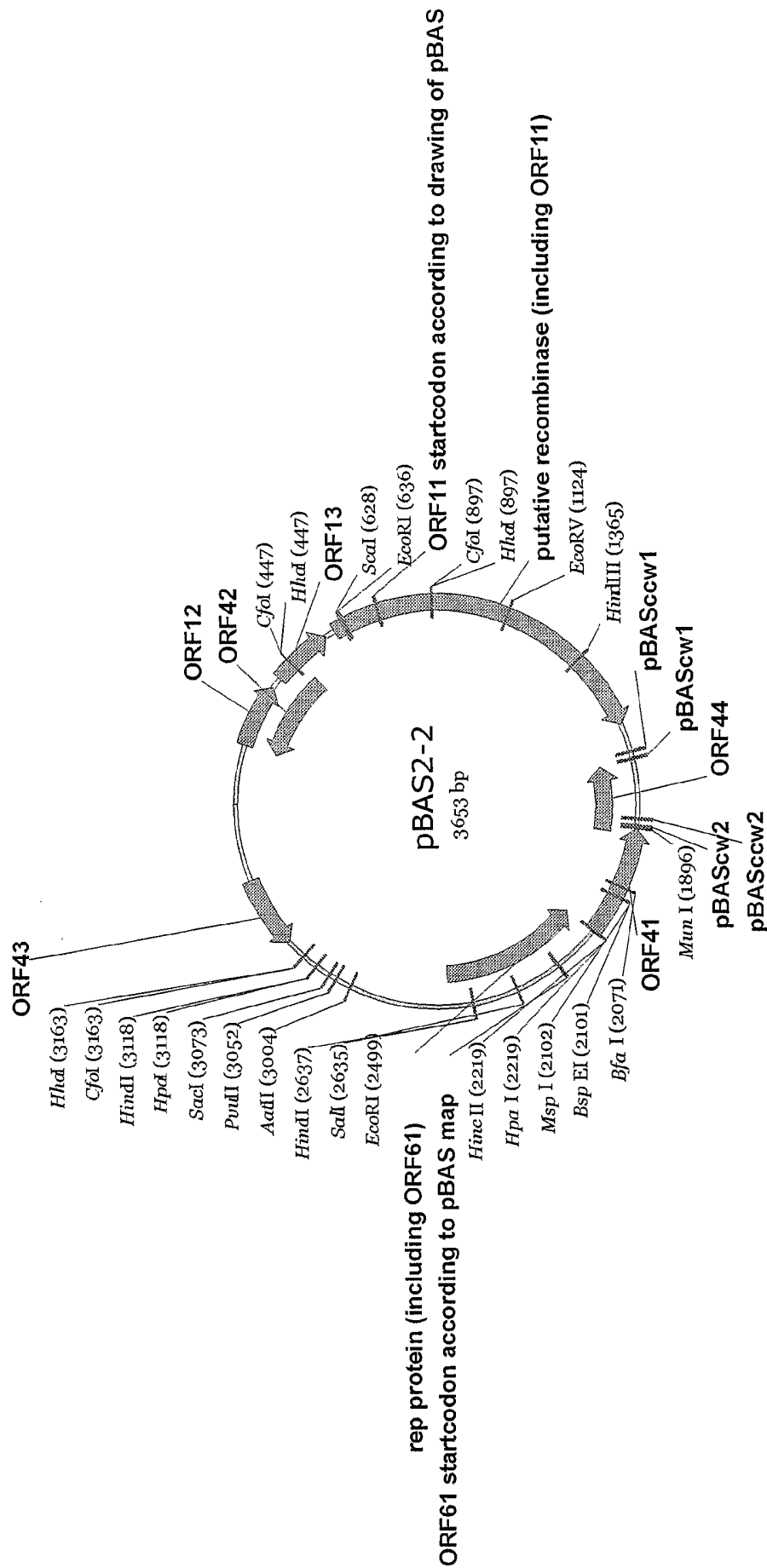


Figure 2

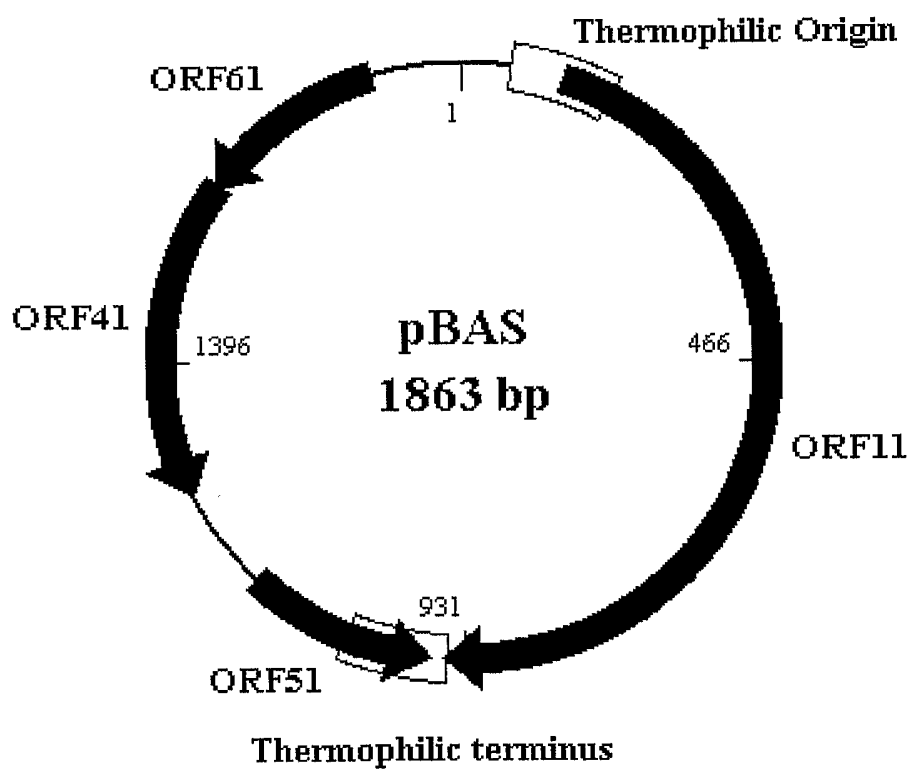
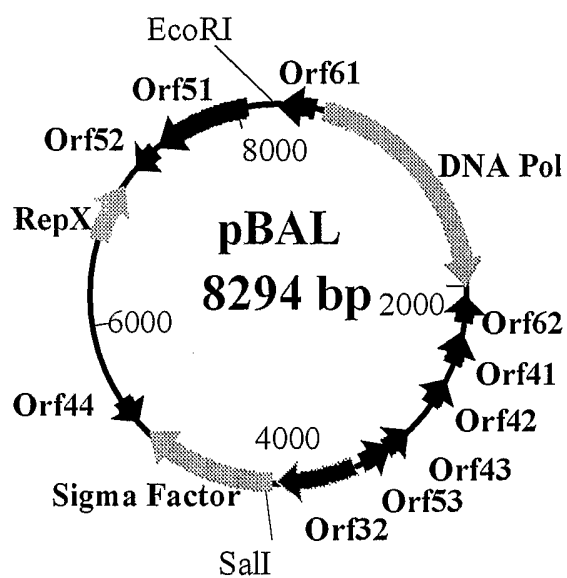


Figure 3



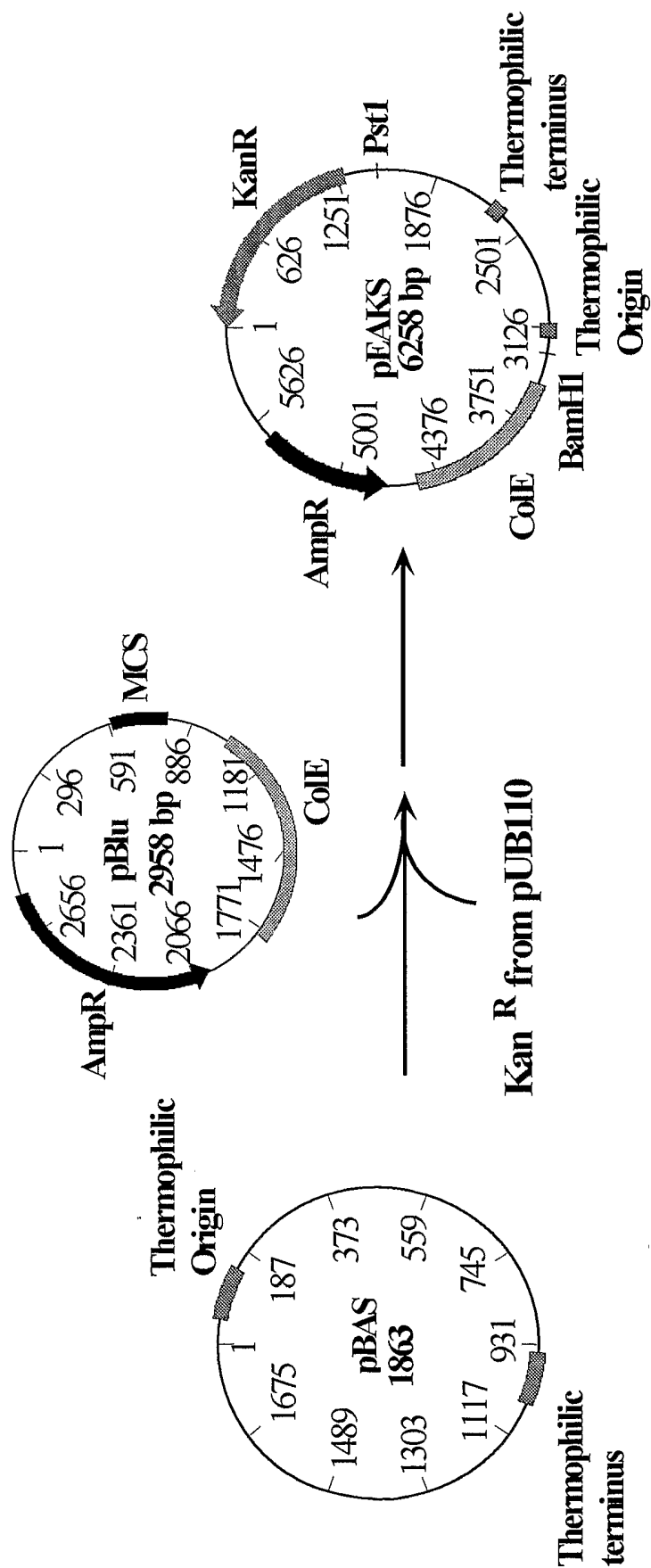


Figure 4

P200100395 WO SEQ.ST25
SEQUENCE LISTING

<110> DTU, Technical University of Denmark

<120> Plasmids from an extremely thermophilic microorganism and derived expression vectors

<130> P200100395 WO

<150> US 60/311,871

<151> 2001-08-13

<160> 28

<170> PatentIn version 3.1

<210> 1

<211> 8294

<212> DNA

<213> Anaerocellum thermophilum

<220>

<221> CDS

<222> (339)..(2003)

<223> Orf31, DNA polymerase

<220>

<221> CDS

<222> (3588)..(4166)

<223> Orf32

<220>

<221> CDS

<222> (4197)..(5129)

<223> Orf33, Sigma factor

<220>

<221> CDS

<222> (6617)..(7024)

<223> Orf34, Replication protein

<400> 1

ttctacaaag gcataggggtt aagactaact gcacttaatg gcaaaaagcc attccttgcc	60
gactggtact ataaagaact aagcgacgaa gagcttttag agtactgggc tcagggctat	120
aacatcgggtg ggaagatggg agaaaaaagc tgctggatgg tggatatcga catagaccac	180
ccagcagcaa acaaaatagt tgatttgtac cttccaatag atacgttgaa attcgggcgg	240
gcaacgaaac ctctatcaca tctgctgtat ttttcaaaaa attgtaagtc gcttaagaga	300
gattttcaaat tcagtcaatc agactgaaaa aactacca ttg tcg aga tcc gca gta	356
Leu Ser Arg Ser Ala Val	
1 5	
ctg gtc aag cag acc atg att cca cct tca atc cat cca gac act aaa	404
Leu Val Lys Gln Thr Met Ile Pro Pro Ser Ile His Pro Asp Thr Lys	
10 15 20	

P200100395 WO SEQ.ST25

gaa gaa tta aga ttt gtg ggt gaa ctt aaa gaa ccc tta aaa atc gaa Glu Glu Leu Arg Phe Val Gly Glu Leu Lys Glu Pro Leu Lys Ile Glu 25 30 35	452
gct gat ttg cta act gcc gct gtt aat aaa atc gca gct gcg act cta Ala Asp Leu Leu Thr Ala Ala Val Asn Lys Ile Ala Ala Ala Thr Leu 40 45 50	500
atc ggc ttg cat tgg tat gaa ggg cag cgt caa gat tgt gct tta gcg Ile Gly Leu His Trp Tyr Glu Gly Gln Arg Gln Asp Cys Ala Leu Ala 55 60 65 70	548
ctg gct ggt ggt ctt ttg cgt gct ggg tgg tct gaa gaa gag gca gaa Leu Ala Gly Gly Leu Leu Arg Ala Gly Trp Ser Glu Glu Glu Ala Glu 75 80 85	596
aaa ttt ata cag gct gtt tgc tta gtt gca gat gat gaa gag tca aaa Lys Phe Ile Gln Ala Val Cys Leu Val Ala Asp Asp Glu Glu Ser Lys 90 95 100	644
aag cgg ata aaa act gta att caa act gct aag aaa caa gag gag aac Lys Arg Ile Lys Thr Val Ile Gln Thr Ala Lys Lys Gln Glu Glu Asn 105 110 115	692
aag gcg acc aca ggc tgg cgt aga ttg gca caa ttg ata gga aat gag Lys Ala Thr Thr Gly Trp Arg Arg Leu Ala Gln Leu Ile Gly Asn Glu 120 125 130	740
gtt gtt gca aaa gtt att gaa tgg ctg ggg ata aaa caa acc aaa gca Val Val Ala Lys Val Ile Glu Trp Leu Gly Ile Lys Gln Thr Lys Ala 135 140 145 150	788
gag ctg cta tcc atg gca gaa ctt ttg agt ttg gag ctg cca cct gtt Glu Leu Leu Ser Met Ala Glu Leu Leu Ser Leu Glu Leu Pro Pro Val 155 160 165	836
gaa ttt cta ctt gaa ggt ctt ttg cca aat gaa ggc ctt act ata tta Glu Phe Leu Leu Glu Gly Leu Leu Pro Asn Glu Gly Leu Thr Ile Leu 170 175 180	884
gca ggc agg gaa aag ata ggg aaa agc tgg ctt tgt ctt gaa ctt gcc Ala Gly Arg Glu Lys Ile Gly Lys Ser Trp Leu Cys Leu Glu Leu Ala 185 190 195	932
ttg agt tta gca gca gga aca gga ttt ctg ggc agg cag act aaa cga Leu Ser Leu Ala Ala Gly Thr Gly Phe Leu Gly Arg Gln Thr Lys Arg 200 205 210	980
cct tac aaa gtg ctc tat ctt gca ctt gaa gac agc aaa aga aga ttg Pro Tyr Lys Val Leu Tyr Leu Ala Leu Glu Asp Ser Lys Arg Arg Leu 215 220 225 230	1028
cag caa cga att aaa caa tta gga act ggc att tca caa aac tgt ttt Gln Gln Arg Ile Lys Gln Leu Gly Thr Gly Ile Ser Gln Asn Cys Phe 235 240 245	1076
gtt ctt ttc aac att gaa agc tta caa gaa tta gaa aag ata atc ata Val Leu Phe Asn Ile Glu Ser Leu Gln Glu Leu Glu Lys Ile Ile Ile 250 255 260	1124
gac cac aac ttt gaa gta gtg ata gta gac acc atg caa gcc ttt aaa Asp His Asn Phe Glu Val Val Ile Val Asp Thr Met Gln Ala Phe Lys 265 270 275	1172

P200100395 WO SEQ.ST25

aga gct gta aaa ctt ttt	aca aac aca aag ctt aac acc tat gag cag	1220
Arg Ala Val Lys Leu Phe	Thr Asn Thr Lys Leu Asn Thr Tyr Glu Gln	
280	285 290	
gat tat aac att agc cta atg ttg aac gac att gca aaa agg ctt aat	1268	
Asp Tyr Asn Ile Ser Leu Met Leu Asn Asp Ile Ala Lys Arg Leu Asn		
295	300 305 310	
tgc acc ttt ata att gtc cat cac aca aag aag tca ggc gaa ctg gat	1316	
Cys Thr Phe Ile Ile Val His His Thr Lys Lys Ser Gly Glu Leu Asp		
	315 320 325	
ttt gta aac gaa gtg ctt ggg aga gta act gca ggt gct gat acc atc	1364	
Phe Val Asn Glu Val Leu Gly Arg Val Thr Ala Gly Ala Asp Thr Ile		
	330 335 340	
ctg cat ttg acc aga gga aga acc gaa gct gat ggt ttt cta aaa gtc	1412	
Leu His Leu Thr Arg Gly Arg Thr Glu Ala Asp Gly Phe Leu Lys Val		
	345 350 355	
aca ggc aga gat gtt gaa gaa cag gag ctt gcg ctg cgg ttt gaa aat	1460	
Thr Gly Arg Asp Val Glu Glu Gln Glu Leu Ala Leu Arg Phe Glu Asn		
	360 365 370	
ggg cac tgg tct ttg tta gga act gcg aaa gaa tat gca att tcg aaa	1508	
Gly His Trp Ser Leu Leu Gly Thr Ala Lys Glu Tyr Ala Ile Ser Lys		
375	380 385 390	
gaa agg cgg gaa att tta gaa gtc ata aaa caa ttg ggt ggg tct gct	1556	
Glu Arg Arg Glu Ile Leu Glu Val Ile Lys Gln Leu Gly Gly Ser Ala		
	395 400 405	
aca cca aaa caa att gcc gaa gcc tta gga aaa aca aga agc aac gtt	1604	
Thr Pro Lys Gln Ile Ala Glu Ala Leu Gly Lys Thr Arg Ser Asn Val		
	410 415 420	
aac aag cta ttg cta aaa ctg gct ggg gat gga tat gtg aag gca gta	1652	
Asn Lys Leu Leu Leu Lys Leu Ala Gly Asp Gly Tyr Val Lys Ala Val		
	425 430 435	
aat gga ata tac agc ctt gcc aca gag aca gaa aaa agc att cac tct	1700	
Asn Gly Ile Tyr Ser Leu Ala Thr Glu Thr Glu Lys Ser Ile His Ser		
	440 445 450	
att cac tct att cac tct att cac tct att cac tct att cac tct gtg	1748	
Ile His Ser Ile His Ser Ile His Ser Ile His Ser Ile His Ser Val		
455	460 465 470	
gac aga gtg aat aga gaa gtg aat gag gga acg atg gaa aat ttt agc	1796	
Asp Arg Val Asn Arg Glu Val Asn Glu Gly Thr Met Glu Asn Phe Ser		
	475 480 485	
gtt cac tct gta gag ccc ttg agc caa cta aat ttt tcg ggg gtg agt	1844	
Val His Ser Val Glu Pro Leu Ser Gln Leu Asn Phe Ser Gly Val Ser		
	490 495 500	
gaa cag agt gaa cgg agt gaa cgc act tta ttg aaa ttc gaa aaa gat	1892	
Glu Gln Ser Glu Arg Ser Glu Arg Thr Leu Leu Lys Phe Glu Lys Asp		
	505 510 515	
aaa gtt tta cat tcg acg ttt gac ttg aaa aag gaa gag tgg cag gaa	1940	
Lys Val Leu His Ser Thr Phe Asp Leu Lys Lys Glu Glu Trp Gln Glu		
	520 525 530	

Side 4

P200100395 WO SEQ.ST25

560	565	570	
aaa gtt gcg ggc ttt agg tgt ccg cag aat gta aaa tgc ggg caa tat			3695
Lys Val Ala Gly Phe Arg Cys Pro Gln Asn Val Lys Cys Gly Gln Tyr			
575	580	585	590
gac tgg gaa agt tta aaa cgt gag ttt atg ctt gga gat ttc aag act			3743
Asp Trp Glu Ser Leu Lys Arg Glu Phe Met Leu Gly Asp Phe Lys Thr			
	595	600	605
ctc aaa gag ttt gca cag gct aaa aat ata agc tat gga ttt ttg aga			3791
Leu Lys Glu Phe Ala Gln Ala Lys Asn Ile Ser Tyr Gly Phe Leu Arg			
	610	615	620
aac aaa gca aaa ggc tgg act cag gag aaa aga caa ctt agc aaa acc			3839
Asn Lys Ala Lys Gly Trp Thr Gln Glu Lys Arg Gln Leu Ser Lys Thr			
	625	630	635
aag agt cag ctt gtg gtt gaa aag aca ttg caa aaa caa ata gaa aaa			3887
Lys Ser Gln Leu Val Val Glu Lys Thr Leu Gln Lys Gln Ile Glu Lys			
	640	645	650
gca agc gac tac aat act ttg cat gtt gag ttt tgg gac agg ctt ctt			3935
Ala Ser Asp Tyr Asn Thr Leu His Val Glu Phe Trp Asp Arg Leu Leu			
	655	660	670
gac ttg gtg tgg cag gct ttg cat gat gag aaa acc atc aaa acc aag			3983
Asp Leu Val Trp Gln Ala Leu His Asp Glu Lys Thr Ile Lys Thr Lys			
	675	680	685
gat ggc aga ata aac att tat gca ctt gaa aag ctg gct ttg gtg gtt			4031
Asp Gly Arg Ile Asn Ile Tyr Ala Leu Glu Lys Leu Ala Leu Val Val			
	690	695	700
gag agg gtg caa aaa gga cag aga ctt gct ctt gga ctt gat gag aaa			4079
Glu Arg Val Gln Lys Gly Gln Arg Leu Ala Leu Gly Leu Asp Glu Lys			
	705	710	715
aac tca aat caa ccg cag att gaa gac agt att aga gaa ctt atc aaa			4127
Asn Ser Asn Gln Pro Gln Ile Glu Asp Ser Ile Arg Glu Leu Ile Lys			
	720	725	730
gca ctt cat gaa cca gat gag cct tct gtt tta aac taa gcttgcaata			4176
Ala Leu His Glu Pro Asp Glu Pro Ser Val Leu Asn			
	735	740	745
taaggagtgg caataaacia atg cag cag tca gaa aaa aga cag ttt gcc aag			4229
	Met Gln Gln Ser Glu Lys Arg Gln Phe Ala Lys		
	750	755	
gta caa agc cag ctt tta att gtc gac aac ctg aag ctt gtg tat cat			4277
Val Gln Ser Gln Leu Leu Ile Val Asp Asn Leu Lys Leu Val Tyr His			
	760	765	770
gtt gca aat aaa ttt atg cca tgt cca aag ggc tat tac tat gaa gtt			4325
Val Ala Asn Lys Phe Met Pro Cys Pro Lys Gly Tyr Tyr Tyr Glu Val			
	775	780	785
gat gac ctt gtc agt gaa ggg tac att gga ctt gtt gtt gca gca aag			4373
Asp Asp Leu Val Ser Glu Gly Tyr Ile Gly Leu Val Val Ala Ala Lys			
	790	795	800
aac tat gac cct gcc aaa ggc agt ttt tca acc tat gct tgc aag gtg			4421
Asn Tyr Asp Pro Ala Lys Gly Ser Phe Ser Thr Tyr Ala Cys Lys Val			

P200100395 WO SEQ.ST25

810	815	820	
att gag agc aag ata aaa aga agc ttg cca aaa tac gag tta tct tgt			4469
Ile Glu Ser Lys Ile Lys Arg Ser Leu Pro Lys Tyr Glu Leu Ser Cys			
825	830	835	
gcc atc tct ttg gac agt cca gta tca aaa gaa ccc gaa gaa gat aca			4517
Ala Ile Ser Leu Asp Ser Pro Val Ser Lys Glu Pro Glu Glu Asp Thr			
840	845	850	
tct aca gtt gca gat ata atc tca gat ggt tat tca gtc gaa agc gaa			4565
Ser Thr Val Ala Asp Ile Ile Ser Asp Gly Tyr Ser Val Glu Ser Glu			
855	860	865	
gtg att agg aaa gac aca gtc caa aag cta cat agg tat ttt gcc atc			4613
Val Ile Arg Lys Asp Thr Val Gln Lys Leu His Arg Tyr Phe Ala Ile			
870	875	880	885
tta acc gaa gag gaa aaa gaa aga gtc tta tat tac atc tcc aca gga			4661
Leu Thr Glu Glu Glu Lys Glu Arg Val Leu Tyr Tyr Ile Ser Thr Gly			
890	895	900	
aag aat tta cct gcc agt gat aat gtg ttc aaa aaa gca aga aga aag			4709
Lys Asn Leu Pro Ala Ser Asp Asn Val Phe Lys Lys Ala Arg Arg Lys			
905	910	915	
ata ata gcc aag atg aga caa gaa cag ttt gag gca gac ctt gac gac			4757
Ile Ile Ala Lys Met Arg Gln Glu Gln Phe Glu Ala Asp Leu Asp Asp			
920	925	930	
ttg aca gtg ttt ata tcc tgc cca gct gtg cat ata gca agt ggt agt			4805
Leu Thr Val Phe Ile Ser Cys Pro Ala Val His Ile Ala Ser Gly Ser			
935	940	945	
aat ttt ggc tat tca cca gtt gag gca act gtt att gat aga gaa gag			4853
Asn Phe Gly Tyr Ser Pro Val Glu Ala Thr Val Ile Asp Arg Glu Glu			
950	955	960	965
aaa aga aaa cag ctt gag gtt ttg tct gca ata cca caa ttg gac aag			4901
Lys Arg Lys Gln Leu Glu Val Leu Ser Ala Ile Pro Gln Leu Asp Lys			
970	975	980	
ctg agg gtt tta aaa caa caa ggc gaa cag gac aga atg aaa ctt ctt			4949
Leu Arg Val Leu Lys Gln Gln Gly Glu Gln Asp Arg Met Lys Leu Leu			
985	990	995	
gtt gcc aag tgg gaa gta gaa gag ttt ata gaa cgc aac tgg gat			4994
Val Ala Lys Trp Glu Val Glu Glu Phe Ile Glu Arg Asn Trp Asp			
1000	1005	1010	
aat ctc aca atc cag aat gca agg ctt ttg cag gac tat ttt gtg			5039
Asn Leu Thr Ile Gln Asn Ala Arg Leu Leu Gln Asp Tyr Phe Val			
1015	1020	1025	
tat tgt ctt tcc cat ctt tca atg gta caa aag tat ggc agc aag			5084
Tyr Cys Leu Ser His Leu Ser Met Val Gln Lys Tyr Gly Ser Lys			
1030	1035	1040	
tac aga gag caa ata aaa aga gtg gtc aaa aag tta gct ttg taa			5129
Tyr Arg Glu Gln Ile Lys Arg Val Val Lys Lys Leu Ala Leu			
1045	1050	1055	
aaatttttttc ttgcgaaaaa agcgaaaaaa gcgaaaaaag tcttgagaat aggaatatca			5189

P200100395 WO SEQ.ST25

```

aggtttttag tatactact tttttcgag ctattgcgaa aaaagttcag aaaaagccta 5249
aaaaccttgt ggcgactgaa ttacaacctt gcgaaaaaag tgtgaaaaaa gtccatatac 5309
cctatagggg gtataaaact ttggtatatc cccccacctt gacaagtagg caagccatct 5369
tgagaccggc gggcccagct ccaaaaacct gcgattgaca cgcacacgag gggggttaaa 5429
caaaaaaggt taaaaaacia aaaaaatttc gcatttttagc ttaaaaagtc tgaaataatt 5489
ttaaaaattgc ttaaaacatt gcaattttta agaggctaaa aagctttcac aggacaaaaa 5549
ggaaagacag agaacagata caaaatttca aaagatcatt ttgaaccttt gaaacattca 5609
cagcatagcc agtaatttaa atgttcaa atgacattat aattcctttc cagtaattcc 5669
attccttacca tcgtttataa gcacatcaaa atggcaagat ggtattttat attcccctat 5729
ctcattctgc acgttgtaga gactttaaaa gcgctttgag aaggttgtca aacgtgtcaa 5789
atgcaaagtt taacgtgaca ggttgacaat ttaacagagc tgattttgac cgcagcggtc 5849
aaagtggaac aagcttaaca accatcgaga aaagattaac aggttaacag ttttaaattg 5909
tcaagaagtg tcaaattgtt ttataaacgc tgcctgtgtt aagaaatggg aaggtaagct 5969
acagctttgc ctctctttta agcttatcaa ccttggcaca tctcaaaaat gggcagggtat 6029
tcaagtttaa ccagccgatg aaattcgtgg ctggtttcaa gccaaagccg accaaattcg 6089
atgcttggac atgctaacia atgctaacct ttggagaaga cttacaaagg gtattctaac 6149
ttggaacatc ctataatacag ggtgtcgaaa ttcacgacac ccataaaact attcggaat 6209
ttcggagtc ctaaaaaaag agcaacataa aacttacaat tttgcagttc ctaaaagggtg 6269
acttcctaac ttggaaagtc attttaaagg ggaatggacc tataaatcct tgtagttttt 6329
caatcaggac cagcgcgctg gggtttgc atgtgaccgc aagttttcaa tgtcttttgg 6389
agcttgtctg ttctttgcag gctttagttt aaactccaga attttgagca aaaagctact 6449
tagtaaaact taaggttcgt taaggttttg aatgcgtaga aaatcagcaa attaaatota 6509
aataccccg ccaaagaaaa gccattgacc aagttaaagc gtacatataa actggtttaa 6569
agcagtttgc aaacagttta caagcacgtt tacaagagg tgtaaac atg aaa aac 6625
                               Met Lys Asn

att  caa  cag  tta  tgc  gat  gag  ctg  ggt  ata  tcg  aga  aca  aca  gtt  6670
Ile  Gln  Gln  Leu  Cys  Asp  Glu  Leu  Gly  Ile  Ser  Arg  Thr  Thr  Val
1060                               1065                               1070

tat  aag  tac  atc  aga  agg  ctt  gga  ata  gag  gta  aag  aag  gaa  ggc  6715
Tyr  Lys  Tyr  Ile  Arg  Arg  Leu  Gly  Ile  Glu  Val  Lys  Lys  Glu  Gly
1075                               1080                               1085

aat  att  gca  ttt  ata  aac  gac  gat  gac  att  gag  aaa  ata  aaa  gaa  6760
Asn  Ile  Ala  Phe  Ile  Asn  Asp  Asp  Asp  Ile  Glu  Lys  Ile  Lys  Glu
1090                               1095                               1100

gca  cta  tca  aca  gcg  agt  aca  aac  agt  ttg  cac  aca  gat  tac  aaa  6805
Ala  Leu  Ser  Thr  Ala  Ser  Thr  Asn  Ser  Leu  His  Thr  Asp  Tyr  Lys

```

P200100395 WO SEQ.ST25

1105	1110	1115	
ctt gag tat ata cag tct tta cag cag caa atc gaa aca ttg caa			6850
Leu Glu Tyr Ile Gln Ser Leu Gln Gln Gln Ile Glu Thr Leu Gln			
1120	1125	1130	
aaa caa gta gac ttt ttg aag gaa cag ctc aga gca aaa gac gaa			6895
Lys Gln Val Asp Phe Leu Lys Glu Gln Leu Arg Ala Lys Asp Glu			
1135	1140	1145	
cag ata gcc aag ctg att caa acc aat caa aac ttt cag gtg ctt			6940
Gln Ile Ala Lys Leu Ile Gln Thr Asn Gln Asn Phe Gln Val Leu			
1150	1155	1160	
tta aaa gaa aaa gaa gaa caa att tac cag ctt gaa gga caa aaa			6985
Leu Lys Glu Lys Glu Glu Gln Ile Tyr Gln Leu Glu Gly Gln Lys			
1165	1170	1175	
caa aaa ggc agt ttt tta aaa aag ctg ttt ggc cga taa gtttttgttt			7034
Gln Lys Gly Ser Phe Leu Lys Lys Leu Phe Gly Arg			
1180	1185	1190	
gtcgtgtgcc agtataatac agttttgcct ttatataccgt cgttttataag ccacgtctaa			7094
aaacagggta gtattttttat atgtatcatc aattagatgc atttctaaaa gcctttatgc			7154
gcttctgaag ggcatttttag aaacaaccag caggggaaca tatcaagggt gcgggtagaa			7214
tagcaccttt ccagaaggga ttaaccgctt ccagcagcat tgaaaaactg aactggcaaa			7274
taggtaaaaa atagacaagg cagggttaata aatacaaaat ggtatccaaa aacatacgaa			7334
aagcaccaaa aagaaacaaa taaacaaaat cctaattgca caaaattata tttttgcgca			7394
atcaaaaaat gtttgggtaa cacttaaaag tgggtgtgat tgaaaacaat ttttgtgcgt			7454
ataacaataa caggggggtga aagctatgaa aatgcgcaca gtgaacttga gactgccaga			7514
agacattttt aatgaaatcg gcgagattgc aaagaaaaaa gggcttacta aaagcagtct			7574
gataagggtg attgttgtgg agtatattgga aaagcaaaag aaagaagggtg gcaagaatgg			7634
ataatctgcc gaggggtgaa ataccaaagg acaaagagaa aatcaagcaa ttaatagctg			7694
ctcttgaaca acaattgcaa aaggacacaa acgaaaagga taaagaaata cacagaaaag			7754
ctcttgaaga tttaaaagcc agtttaaaaca aaaaagacca cctgtaaaaa cttttgttcc			7814
ctataggggt gtgatataat aacttacaaa gaagataaag acaagtaaag gaggtgcaaa			7874
gtatgacaca gcaaataaca ataagagtcc ctgctaaaga acttgaagaa tggcgactca			7934
agattaaaga gctttcagac agactggggc atgggaattg ctttaagcga attagtaaga			7994
caaggcggtc caaaagcaat tgagggaactt gagtcaagaa taaaatgagc acgctgtaaa			8054
ttttgtgggc gggacggaga caagcgtgct cattctcttt ttttaactt tgcttaactt			8114
ttttcaaaac ttttcttagc ttttttcaaa gaatttgata cattttgtaa ctctaatttt			8174
accacaatgg tgcaaaaatg agacaccaaa gtttgaacaa aaaacaaaag tagacgttgt			8234
aaagattcat gaaaggaagg aggctaaaaa atgcgaaagg acatcagaaa gtaccttgaa			8294

P200100395 WO SEQ.ST25

```

<210> 2
<211> 554
<212> PRT
<213> Anaerocellum thermophilum

<400> 2

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

Ile His Pro Asp Thr Lys Glu Glu Leu Arg Phe Val Gly Glu Leu Lys
          20          25          30

Glu Pro Leu Lys Ile Glu Ala Asp Leu Leu Thr Ala Ala Val Asn Lys
          35          40          45

Ile Ala Ala Ala Thr Leu Ile Gly Leu His Trp Tyr Glu Gly Gln Arg
          50          55          60

Gln Asp Cys Ala Leu Ala Leu Ala Gly Gly Leu Leu Arg Ala Gly Trp
65          70          75          80

Ser Glu Glu Glu Ala Glu Lys Phe Ile Gln Ala Val Cys Leu Val Ala
          85          90          95

Asp Asp Glu Glu Ser Lys Lys Arg Ile Lys Thr Val Ile Gln Thr Ala
          100          105          110

Lys Lys Gln Glu Glu Asn Lys Ala Thr Thr Gly Trp Arg Arg Leu Ala
          115          120          125

Gln Leu Ile Gly Asn Glu Val Val Ala Lys Val Ile Glu Trp Leu Gly
          130          135          140

Ile Lys Gln Thr Lys Ala Glu Leu Leu Ser Met Ala Glu Leu Leu Ser
145          150          155          160

Leu Glu Leu Pro Pro Val Glu Phe Leu Leu Glu Gly Leu Leu Pro Asn
          165          170          175

Glu Gly Leu Thr Ile Leu Ala Gly Arg Glu Lys Ile Gly Lys Ser Trp
          180          185          190

Leu Cys Leu Glu Leu Ala Leu Ser Leu Ala Ala Gly Thr Gly Phe Leu
          195          200          205

Gly Arg Gln Thr Lys Arg Pro Tyr Lys Val Leu Tyr Leu Ala Leu Glu
          210          215          220

```

P200100395 WO SEQ.ST25

Asp	Ser	Lys	Arg	Arg	Leu	Gln	Gln	Arg	Ile	Lys	Gln	Leu	Gly	Thr	Gly	225	230	235	240
Ile	Ser	Gln	Asn	Cys	Phe	Val	Leu	Phe	Asn	Ile	Glu	Ser	Leu	Gln	Glu	245	250	255	
Leu	Glu	Lys	Ile	Ile	Ile	Asp	His	Asn	Phe	Glu	Val	Val	Ile	Val	Asp	260	265	270	
Thr	Met	Gln	Ala	Phe	Lys	Arg	Ala	Val	Lys	Leu	Phe	Thr	Asn	Thr	Lys	275	280	285	
Leu	Asn	Thr	Tyr	Glu	Gln	Asp	Tyr	Asn	Ile	Ser	Leu	Met	Leu	Asn	Asp	290	295	300	
Ile	Ala	Lys	Arg	Leu	Asn	Cys	Thr	Phe	Ile	Ile	Val	His	His	Thr	Lys	305	310	315	320
Lys	Ser	Gly	Glu	Leu	Asp	Phe	Val	Asn	Glu	Val	Leu	Gly	Arg	Val	Thr	325	330	335	
Ala	Gly	Ala	Asp	Thr	Ile	Leu	His	Leu	Thr	Arg	Gly	Arg	Thr	Glu	Ala	340	345	350	
Asp	Gly	Phe	Leu	Lys	Val	Thr	Gly	Arg	Asp	Val	Glu	Glu	Gln	Glu	Leu	355	360	365	
Ala	Leu	Arg	Phe	Glu	Asn	Gly	His	Trp	Ser	Leu	Leu	Gly	Thr	Ala	Lys	370	375	380	
Glu	Tyr	Ala	Ile	Ser	Lys	Glu	Arg	Arg	Glu	Ile	Leu	Glu	Val	Ile	Lys	385	390	395	400
Gln	Leu	Gly	Gly	Ser	Ala	Thr	Pro	Lys	Gln	Ile	Ala	Glu	Ala	Leu	Gly	405	410	415	
Lys	Thr	Arg	Ser	Asn	Val	Asn	Lys	Leu	Leu	Leu	Lys	Leu	Ala	Gly	Asp	420	425	430	
Gly	Tyr	Val	Lys	Ala	Val	Asn	Gly	Ile	Tyr	Ser	Leu	Ala	Thr	Glu	Thr	435	440	445	
Glu	Lys	Ser	Ile	His	Ser	Ile	His	Ser	Ile	His	Ser	Ile	His	Ser	Ile	450	455	460	
His	Ser	Ile	His	Ser	Val	Asp	Arg	Val	Asn	Arg	Glu	Val	Asn	Glu	Gly	465	470	475	480

P200100395 WO SEQ.ST25

Thr Met Glu Asn Phe Ser Val His Ser Val Glu Pro Leu Ser Gln Leu
 485 490 495

Asn Phe Ser Gly Val Ser Glu Gln Ser Glu Arg Ser Glu Arg Thr Leu
 500 505 510

Leu Lys Phe Glu Lys Asp Lys Val Leu His Ser Thr Phe Asp Leu Lys
 515 520 525

Lys Glu Glu Trp Gln Glu Ala Ala Glu Gln Phe Lys Cys Glu Cys Gly
 530 535 540

His Asp Phe Ser Ile Cys Arg Ser Phe Ser
 545 550

<210> 3
 <211> 192
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 3

Met Phe Thr Phe Phe Asp Leu Leu Ser Ser Leu Lys Ser Leu Ser Ile
 1 5 10 15

Val Glu Lys Pro Lys Val Ala Gly Phe Arg Cys Pro Gln Asn Val Lys
 20 25 30

Cys Gly Gln Tyr Asp Trp Glu Ser Leu Lys Arg Glu Phe Met Leu Gly
 35 40 45

Asp Phe Lys Thr Leu Lys Glu Phe Ala Gln Ala Lys Asn Ile Ser Tyr
 50 55 60

Gly Phe Leu Arg Asn Lys Ala Lys Gly Trp Thr Gln Glu Lys Arg Gln
 65 70 75 80

Leu Ser Lys Thr Lys Ser Gln Leu Val Val Glu Lys Thr Leu Gln Lys
 85 90 95

Gln Ile Glu Lys Ala Ser Asp Tyr Asn Thr Leu His Val Glu Phe Trp
 100 105 110

Asp Arg Leu Leu Asp Leu Val Trp Gln Ala Leu His Asp Glu Lys Thr
 115 120 125

Ile Lys Thr Lys Asp Gly Arg Ile Asn Ile Tyr Ala Leu Glu Lys Leu
 130 135 140

Ala Leu Val Val Glu Arg Val Gln Lys Gly Gln Arg Leu Ala Leu Gly
 Side 11

side 12

P200100395 WO SEQ.ST25

Ser Cys Pro Ala Val His Ile Ala Ser Gly Ser Asn Phe Gly Tyr Ser
 195 200 205

Pro Val Glu Ala Thr Val Ile Asp Arg Glu Glu Lys Arg Lys Gln Leu
 210 215 220

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

Gln Gln Gly Glu Gln Asp Arg Met Lys Leu Leu Val Ala Lys Trp Glu
 245 250 255

Val Glu Glu Phe Ile Glu Arg Asn Trp Asp Asn Leu Thr Ile Gln Asn
 260 265 270

Ala Arg Leu Leu Gln Asp Tyr Phe Val Tyr Cys Leu Ser His Leu Ser
 275 280 285

Met Val Gln Lys Tyr Gly Ser Lys Tyr Arg Glu Gln Ile Lys Arg Val
 290 295 300

Val Lys Lys Leu Ala Leu
 305 310

<210> 5
 <211> 135
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 5

Met Lys Asn Ile Gln Gln Leu Cys Asp Glu Leu Gly Ile Ser Arg Thr
 1 5 10 15

Thr Val Tyr Lys Tyr Ile Arg Arg Leu Gly Ile Glu Val Lys Lys Glu
 20 25 30

Gly Asn Ile Ala Phe Ile Asn Asp Asp Asp Ile Glu Lys Ile Lys Glu
 35 40 45

Ala Leu Ser Thr Ala Ser Thr Asn Ser Leu His Thr Asp Tyr Lys Leu
 50 55 60

Glu Tyr Ile Gln Ser Leu Gln Gln Gln Ile Glu Thr Leu Gln Lys Gln
 65 70 75 80

Val Asp Phe Leu Lys Glu Gln Leu Arg Ala Lys Asp Glu Gln Ile Ala
 85 90 95

P200100395 WO SEQ.ST25

Lys Leu Ile Gln Thr Asn Gln Asn Phe Gln Val Leu Leu Lys Glu Lys
 100 105 110

Glu Glu Gln Ile Tyr Gln Leu Glu Gly Gln Lys Gln Lys Gly Ser Phe
 115 120 125

Leu Lys Lys Leu Phe Gly Arg
 130 135

<210> 6
 <211> 78
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 6

Leu Ser Thr Leu Leu Val Leu Glu Thr Phe Lys Gln Leu His Cys Thr
 1 5 10 15

Phe Lys Lys Cys Tyr Arg Phe Ala Tyr Arg Met Pro Val Tyr Ser Asn
 20 25 30

Cys Ser Leu Ser Leu Leu Phe Ser Thr Asp Leu Ser Gly Ile Glu Ser
 35 40 45

Thr Lys Ser Thr Leu Phe Cys Pro Cys Phe Ala Leu Thr Leu Leu Ser
 50 55 60

Cys Pro Gly Leu Tyr Phe Met Arg Phe Cys Val Leu Val Ser
 65 70 75

<210> 7
 <211> 67
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 7

Met Ala Val Phe Gly Leu Leu Gly Arg Cys Phe Val Thr His Ser Pro
 1 5 10 15

Phe Gln Ser Leu Phe Phe Lys Arg His Lys Ser Asn Asn Ala Lys Ser
 20 25 30

Cys Val Ser Ala His Pro Ile Thr Ile Leu Cys His Phe Phe Leu Pro
 35 40 45

Asn Phe Asn Leu Thr Phe Ala Leu Tyr Pro Cys Leu Cys Ala Leu Cys
 50 55 60

Tyr Tyr Leu

P200100395 WO SEQ.ST25

65

<210> 8
 <211> 65
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 8

Met Thr Ala Thr Lys Phe Pro Gln Ser Leu Thr Phe Phe Asp Ala Gln
 1 5 10 15

Lys Ala Phe Gln Asn Ser Ser Tyr Lys His Leu Leu Asp Val Asn Trp
 20 25 30

Met Gly Ser Asn Ile Pro Cys Phe Gly Leu Ser Ile Thr Glu Ser Ile
 35 40 45

Pro Ser Ala Leu Phe Phe Cys Phe Ser Phe Thr Ser Leu Leu Pro His
 50 55 60

Val
 65

<210> 9
 <211> 61
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 9

Leu Phe Asn Pro Pro Arg Val Arg Val Asn Arg Arg Phe Leu Glu Leu
 1 5 10 15

Gly Pro Pro Val Ser Arg Trp Leu Ala Tyr Leu Ser Arg Trp Gly Asp
 20 25 30

Ile Pro Lys Phe Tyr Thr Pro Tyr Arg Val Tyr Gly Leu Phe Ser His
 35 40 45

Phe Phe Arg Lys Val Val Ile Gln Ser Pro Gln Gly Phe
 50 55 60

<210> 10
 <211> 226
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 10

Cys Leu Arg Pro Ala His Lys Ile Tyr Ser Val Leu Ile Leu Phe Leu
 1 5 10 15

P200100395 WO SEQ.ST25

Thr Gln Val Pro Gln Leu Leu Leu Glu Arg Leu Val Leu Leu Ile Arg
 20 25 30

Leu Lys Gln Phe Pro Cys Pro Ser Leu Ser Glu Ser Ser Leu Ile Leu
 35 40 45

Ser Arg His Ser Ser Ser Ser Leu Ala Gly Thr Leu Ile Val Ile Cys
 50 55 60

Cys Val Ile Leu Cys Thr Ser Phe Thr Cys Leu Tyr Leu Leu Cys Lys
 65 70 75 80

Leu Leu Tyr His Asn Pro Ile Gly Asn Lys Ser Phe Tyr Arg Trp Ser
 85 90 95

Phe Leu Phe Lys Leu Ala Phe Lys Ser Ser Arg Ala Phe Leu Cys Ile
 100 105 110

Ser Leu Ser Phe Ser Phe Val Ser Phe Cys Asn Cys Cys Ser Arg Ala
 115 120 125

Ala Ile Asn Cys Leu Ile Phe Ser Leu Ser Phe Gly Ile Ser Thr Leu
 130 135 140

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

Leu His Asn Asn Tyr Pro Tyr Gln Thr Ala Phe Ser Lys Pro Phe Phe
 165 170 175

Leu Cys Asn Leu Ala Asp Phe Ile Lys Asn Val Phe Trp Gln Ser Gln
 180 185 190

Val His Cys Ala His Phe His Ser Phe His Pro Leu Leu Leu Tyr
 195 200 205

Ala Gln Lys Leu Phe Ser Ile Asn Thr Thr Phe Lys Cys Tyr Pro Asn
 210 215 220

Ile Phe
 225

<210> 11
 <211> 60
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 11

Leu Phe Leu Phe Gly Ala Phe Arg Met Phe Leu Asp Thr Ile Leu Tyr
 Side 16

P200100395 WO SEQ.ST25

1 5 10 15

Leu Leu Pro Cys Leu Val Tyr Phe Leu Pro Ile Cys Gln Phe Ser Phe
20 25 30

Ser Met Leu Leu Glu Ala Val Asn Pro Phe Trp Lys Gly Ala Ile Leu
35 40 45

Pro Ala Pro Leu Ile Cys Ser Pro Ala Gly Cys Phe
50 55 60

<210> 12
 <211> 79
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 12

Met Leu Lys Gly Phe Leu Asn Ile Ala Lys Cys Cys Ile Pro Asn Pro
1 5 10 15

Arg Leu Val Gly Cys Ser Lys Leu Glu Cys Pro Ser Asn Cys Leu Leu
20 25 30

Ser Ser Phe Trp Leu Ile Gln Asn Gln Lys Ala Pro Pro Arg Leu Phe
35 40 45

Cys Ala Phe Phe Ser Leu Ile Ile Cys Ser Leu Gln Pro Arg Ile Leu
50 55 60

Leu Gly His Asp Ser Tyr Lys Ile Pro Ala Lys Leu Asp Ile Phe
65 70 75

<210> 13
 <211> 80
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 13

Leu Lys Asn Thr Ala Asp Val Ile Glu Val Ser Leu Pro Ala Arg Ile
1 5 10 15

Ser Thr Tyr Leu Leu Glu Gly Thr Asn Gln Leu Phe Cys Leu Leu Leu
20 25 30

Gly Gly Leu Cys Arg Tyr Pro Pro Ser Ser Ser Phe Phe Leu Pro Ser
35 40 45

Ser His Arg Cys Tyr Ser Pro Glu Pro Ser Thr Leu Lys Ala Leu Arg
50 55 60

P200100395 WO SEQ.ST25

Arg Leu Val Leu Tyr Ser Thr Ser Arg Gln Gly Met Ala Phe Cys His
 65 70 75 80

<210> 14
 <211> 66
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 14

Leu Tyr Ile Val Pro Cys Pro Ser Val Phe Phe Phe Ser Val Glu Asn
 1 5 10 15

Val Ala Met Leu Cys Cys Leu Leu Asp Lys Asn Lys Arg Ala Pro Ser
 20 25 30

Gln Val Pro Cys His Gln Val Phe Ala Leu Phe Leu Leu Tyr Cys Ser
 35 40 45

Lys Asn Leu His Cys Leu Leu Ala Ile Lys Leu Pro Leu Pro Leu Leu
 50 55 60

Leu Pro
 65

<210> 15
 <211> 3653
 <212> DNA
 <213> Anaerocellum thermophilum

<220>
 <221> CDS
 <222> (599)..(1588)
 <223> Putative recombinase

<400> 15
 tattgtagtgt ttatgatcta tttcagaagc gaaaagtaaa ccgcggtagc gggttacttt 60
 ttgcgtaata tttggaattg tcagtttggc acgggaaatg gaatcgggca ggtataggga 120
 aaggcaatag cagtggacgg cgttccgatt ccgctttggt tgtgcgtag cacaatgagtc 180
 ttgctgttat cacgctctgc tctgtggtag tacaccgcaa ggggtacacc atttgagatt 240
 ttgccgcata tgagagttta gcagtttgca aaacaaagtt cgagttaga aaagttaggc 300
 aaggggcgat gacgtgcacg tctgacacgc ttgacgtgca ccgagccgag agtagagcgt 360
 gatgacatag aaaaagccca gtagcagggt gtgcgtaccg tgacatgcta ctgtggaaga 420
 caagtaaata gtcaaaaaac ggtgcgctta ctgttcgtgg gctcatggga ataccgtgag 480
 cattctggac aggttcaaca gctgaaagg gcaaaacccc ccttgaaagg gtttaaatac 540
 acacgtcctt tttgtcgttt tttgtcgggt taaatattta tgcaggggat gataggga 598

P200100395 WO SEQ.ST25

atg gga aag ccg tcc ata atc aag caa gta ctc aat gaa ttc gaa aag	646
Met Gly Lys Pro Ser Ile Ile Lys Gln Val Leu Asn Glu Phe Glu Lys	
1 5 10 15	
cag ata agg ttc ggg gaa agt aag cat gaa gca aag aga gaa gag cgg	694
Gln Ile Arg Phe Gly Glu Ser Lys His Glu Ala Lys Arg Glu Glu Arg	
20 25 30	
gag aga tgt gaa gtt act gga gag acg tgg aat cct gct cgt gtg gaa	742
Glu Arg Cys Glu Val Thr Gly Glu Thr Trp Asn Pro Ala Arg Val Glu	
35 40 45	
ggc att ttt agt ttc tca acc tac agg gag tac gtt aag gag gcg tta	790
Gly Ile Phe Ser Phe Ser Thr Tyr Arg Glu Tyr Val Lys Glu Ala Leu	
50 55 60	
gaa ttt gcg aat tgg gct cgt act gaa aag ggg tgt aag gat tta gaa	838
Glu Phe Ala Asn Trp Ala Arg Thr Glu Lys Gly Cys Lys Asp Leu Glu	
65 70 75 80	
caa gca ccg gcc tat gtg tct gaa tat ttg cag tcg cat ata gac aaa	886
Gln Ala Arg Ala Tyr Val Ser Glu Tyr Leu Gln Ser His Ile Asp Lys	
85 90 95	
ggg tat agc gcg tgg act gtt aag aaa gaa gca gca gcc ctg gcg aaa	934
Gly Tyr Ser Ala Trp Thr Val Lys Lys Glu Ala Ala Ala Leu Ala Lys	
100 105 110	
ctg tat cat tgt cgt aca act gac ttt aaa gta gag ctt ccc gca aga	982
Leu Tyr His Cys Arg Thr Thr Asp Phe Lys Val Glu Leu Pro Ala Arg	
115 120 125	
cac agg gaa gag att gag aga agc agg gga tac aaa gac cac gat agg	1030
His Arg Glu Glu Ile Glu Arg Ser Arg Gly Tyr Lys Asp His Asp Arg	
130 135 140	
gag ttt agc aaa gag agg aat aga gac att atc atc ttt tca aaa gct	1078
Glu Phe Ser Lys Glu Arg Asn Arg Asp Ile Ile Ile Phe Ser Lys Ala	
145 150 155 160	
act ggg ctg aga aga agg gaa ttg gaa aga gtg agt tct cgg gat atc	1126
Thr Gly Leu Arg Arg Arg Glu Leu Glu Arg Val Ser Ser Arg Asp Ile	
165 170 175	
ttt cgt ggg cct gac gga aga tta tat gtg cac gtg agc aac ggc aag	1174
Phe Arg Gly Pro Asp Gly Arg Leu Tyr Val His Val Ser Asn Gly Lys	
180 185 190	
ggc ggt aga gaa agg gat gtt cat gtt ttg cag aaa tac gag aga gag	1222
Gly Gly Arg Glu Arg Asp Val His Val Leu Gln Lys Tyr Glu Arg Glu	
195 200 205	
gtt gag agg ata gtc aga gag cgg gaa gga aga gac agg ctg ttt gac	1270
Val Glu Arg Ile Val Arg Glu Arg Glu Gly Arg Asp Arg Leu Phe Asp	
210 215 220	
agg gtc ccc ata agg atg gac gta cac agc tat agg agg gaa tat gca	1318
Arg Val Pro Ile Arg Met Asp Val His Ser Tyr Arg Arg Glu Tyr Ala	
225 230 235 240	
aga gag cgt tac aga gaa gtt gag cgt gag ata agt cgt gag aga aag	1366
Arg Glu Arg Tyr Arg Glu Val Glu Arg Glu Ile Ser Arg Glu Arg Lys	
245 250 255	

P200100395 WO SEQ.ST25

ctt ttc gac aga gtt gag gat ctc gtt cgt agt agg ctt aca agg ctc	1414
Leu Phe Asp Arg Val Glu Asp Leu Val Arg Ser Arg Leu Thr Arg Leu	
260 265 270	
tat cct gac agg ttt aga gaa att ggc gaa aga caa ctt act cgt gaa	1462
Tyr Pro Asp Arg Phe Arg Glu Ile Gly Glu Arg Gln Leu Thr Arg Glu	
275 280 285	
ctc aca aga gct gat ggg ctt tat cat cgc agt gat ggt agg gag ttt	1510
Leu Thr Arg Ala Asp Gly Leu Tyr His Arg Ser Asp Gly Arg Glu Phe	
290 295 300	
gac cgc ctg gca ttg tgg gaa gtt tca aac gac ttg gga cat aat cga	1558
Asp Arg Leu Ala Leu Trp Glu Val Ser Asn Asp Leu Gly His Asn Arg	
305 310 315 320	
att gac gtt gtt gca aga cac tat ctg gat taagcgaata aaggctcaag	1608
Ile Asp Val Val Ala Arg His Tyr Leu Asp	
325 330	
aaagtggata aaaaaacagg gggtatattg atatatcccc cctgtttttt gtgcgtctac	1668
aggaccttat ttgcgtttca aggctattct tgtggggtag tgcagtaaaa aaggttgcct	1728
ttgtatcttg cctgtagaat cttgccgttt tcaaaaacaa attcaacggt gcacttatca	1788
gcaagtttct tgattattcc ctgctcaaat tcgtcgaacc tgtacccttc cattgctctt	1848
gctgtgaaat aaccagaag gtacattgct tcttcacgtg taaactcaat tgttagcttt	1908
ttcactctct tttacccct ttacgagact gagatattct ttgtacaatc tgtcaatacg	1968
ttctgtatcg tcaggttgta cgtttttgag tttttcgtag agattatgaa acaacagggt	2028
ctctgctgtt ccaataggggt atttctttct ttccgcgtgg actagccttt tcagttgggt	2088
tagcttttacg ttccggatct tgctcagtag gtcttcgacg atattttcaa tctgttttgc	2148
ttcgtcttct tcgatgtaga agcatagttg tttgatttca agtagcctct gaataagttc	2208
ctgctttgtt aacattcctt gtcacccttc cccttttcag tcttttttat tttgtcgtat	2268
tctgattcga actttctcag ttggtcagct ctgccgtaac gtatgcgtcg aggatttagc	2328
atgaatcgtc cacggcctat tttaatcaac aaatctttct cgattagctt tttgattgca	2388
agatgggtatt gcttttttga catgtcgatg tttcctttga tgtcgtattc gtatgaaagt	2448
gcaaacgtgt agttgttaaa cgagattttg tgttgtacaa tgtatgagag aattcgcatt	2508
gctgattttg caaggctttt atcttccatt atgagtttga tagaaatgaa tggaagtttg	2568
acaaagtctg tgtcttcaaa atcatgtccg atcagtatga aagtttgtac ttcgcctggt	2628
tcagggtcga ccattttttt gtaccctatc ttcttcattt gaatccccct ttttttacct	2688
ttttccaatt ctaaagccat tataatacac ttaaagtaac ttgtcaagta acttcagggtg	2748
taaaaaatta cacttcagggt gtaataaatt acactttcca ttcagtattt tcaaggcttt	2808
gtgggtaact ttattcttat ctatgtatat atcgctgctg ttagcaggct tgaaaatttc	2868
cagttaggat aagcaggaac aacggtcgct gacgctgaac actgacgaaa tagctgacgc	2928

P200100395 WO SEQ.ST25

```

cccaaagtcc acaacagtgc caaaccgata aaaaaaacat gctaacgcaa acatagacta 2988
acgcacgact gacgtcgtga tgtgtgtgtg ggcctaccta cacacaaaaa gaactaacia 3048
cagctgacta acgtctgaag agctctaaca acacttttgc aacgctgagc taacggacag 3108
ctcaacgtta acacccgcta aaaaaattag agcaggattt gagcaagagg ggcgctgcaa 3168
ccccctttga aaccgctttc ataaccgaa aagctcttca cccaccgcat caatattcat 3228
acccgcttgc ataattattc atccctgatt gtaccacaaa acccagaaaa gtcctgggaa 3288
aatcatgaag attcttccat tttccaattt ccaggaagcg aattgaaatt ccaggggatt 3348
ttttgccttt ttccaggga tgaactggaa tttccagcgg tttttccagg gacttaacca 3408
cacgaaaggt actttccgcc atcaaaggct tgcaaaccga taggaaacta ctggcaggca 3468
aaaccttcca ccagtagca aatctcttct accttgtttt gcgagcgta gcaggagct 3528
gtccgtagc tttgtgttag catccttctg ttagcctttg gcacgttagt agacttttgt 3588
tagttctctc ttcttgtttt tatcttcatt gtgattgttt tccagaagac tgcttgacat 3648
tttgt 3653

```

<210> 16
 <211> 330
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 16

```

Met Gly Lys Pro Ser Ile Ile Lys Gln Val Leu Asn Glu Phe Glu Lys
1           5           10           15

```

```

Gln Ile Arg Phe Gly Glu Ser Lys His Glu Ala Lys Arg Glu Glu Arg
20           25           30

```

```

Glu Arg Cys Glu Val Thr Gly Glu Thr Trp Asn Pro Ala Arg Val Glu
35           40           45

```

```

Gly Ile Phe Ser Phe Ser Thr Tyr Arg Glu Tyr Val Lys Glu Ala Leu
50           55           60

```

```

Glu Phe Ala Asn Trp Ala Arg Thr Glu Lys Gly Cys Lys Asp Leu Glu
65           70           75           80

```

```

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

```

```

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

```

```

Leu Tyr His Cys Arg Thr Thr Asp Phe Lys Val Glu Leu Pro Ala Arg
115          120          125

```

P200100395 WO SEQ.ST25

His Arg Glu Glu Ile Glu Arg Ser Arg Gly Tyr Lys Asp His Asp Arg
 130 135 140

Glu Phe Ser Lys Glu Arg Asn Arg Asp Ile Ile Ile Phe Ser Lys Ala
 145 150 155 160

Thr Gly Leu Arg Arg Arg Glu Leu Glu Arg Val Ser Ser Arg Asp Ile
 165 170 175

Phe Arg Gly Pro Asp Gly Arg Leu Tyr Val His Val Ser Asn Gly Lys
 180 185 190

Gly Gly Arg Glu Arg Asp Val His Val Leu Gln Lys Tyr Glu Arg Glu
 195 200 205

Val Glu Arg Ile Val Arg Glu Arg Glu Gly Arg Asp Arg Leu Phe Asp
 210 215 220

Arg Val Pro Ile Arg Met Asp Val His Ser Tyr Arg Arg Glu Tyr Ala
 225 230 235 240

Arg Glu Arg Tyr Arg Glu Val Glu Arg Glu Ile Ser Arg Glu Arg Lys
 245 250 255

Leu Phe Asp Arg Val Glu Asp Leu Val Arg Ser Arg Leu Thr Arg Leu
 260 265 270

Tyr Pro Asp Arg Phe Arg Glu Ile Gly Glu Arg Gln Leu Thr Arg Glu
 275 280 285

Leu Thr Arg Ala Asp Gly Leu Tyr His Arg Ser Asp Gly Arg Glu Phe
 290 295 300

Asp Arg Leu Ala Leu Trp Glu Val Ser Asn Asp Leu Gly His Asn Arg
 305 310 315 320

Ile Asp Val Val Ala Arg His Tyr Leu Asp
 325 330

<210> 17

<211> 163

<212> PRT

<213> Anaerocellum thermophilum

<400> 17

Met Ala Leu Glu Leu Glu Lys Gly Lys Lys Arg Gly Ile Gln Met Lys
 1 5 10 15

P200100395 WO SEQ.ST25

Lys Ile Gly Tyr Lys Lys Met Val Asp Pro Glu Thr Gly Glu Val Gln
 20 25 30

Thr Phe Ile Leu Ile Gly His Asp Phe Glu Asp Thr Asp Phe Val Lys
 35 40 45

Leu Pro Phe Ile Ser Ile Lys Leu Ile Met Glu Asp Lys Asp Leu Ala
 50 55 60

Lys Ser Ala Met Arg Ile Leu Ser Tyr Ile Val Gln His Lys Ile Ser
 65 70 75 80

Phe Asn Asn Tyr Thr Phe Ala Leu Ser Tyr Glu Tyr Asp Ile Lys Gly
 85 90 95

Asn Ile Asp Met Ser Lys Lys Gln Tyr His Leu Ala Ile Lys Lys Leu
 100 105 110

Ile Glu Lys Asp Leu Leu Ile Lys Ile Gly Arg Gly Arg Phe Met Leu
 115 120 125

Asn Pro Arg Arg Ile Arg Tyr Gly Arg Ala Asp Gln Leu Arg Lys Phe
 130 135 140

Glu Ser Glu Tyr Asp Lys Ile Lys Lys Thr Glu Lys Gly Lys Gly Asp
 145 150 155 160

Lys Glu Cys

<210> 18

<211> 109

<212> PRT

<213> Anaerocellum thermophilum

<400> 18

Met Leu Thr Lys Gln Glu Leu Ile Gln Arg Leu Leu Glu Ile Lys Gln
 1 5 10 15

Leu Cys Phe Tyr Ile Glu Glu Asp Glu Ala Lys Gln Ile Glu Asn Ile
 20 25 30

Val Glu Asp Leu Leu Ser Lys Ile Arg Asn Val Lys Leu Asn Gln Leu
 35 40 45

Lys Arg Leu Val His Ala Glu Arg Lys Lys Tyr Pro Ile Gly Thr Ala
 50 55 60

P200100395 WO SEQ.ST25

Glu Asn Leu Leu Phe His Asn Leu Tyr Glu Lys Leu Lys Asn Val Gln
65 70 75 80

Pro Asp Asp Thr Glu Arg Ile Asp Arg Leu Tyr Lys Glu Tyr Leu Ser
85 90 95

Leu Val Lys Gly Val Lys Glu Ser Glu Lys Ala Asn Asn
100 105

<210> 19
<211> 99
<212> PRT
<213> Anaerocellum thermophilum

<400> 19

Met Ser Pro Arg Thr Val Ser Ala Pro Phe Phe Asp Tyr Leu Leu Val
1 5 10 15

Phe His Ser Ser Met Ser Arg Tyr Ala Gln Pro Ala Thr Gly Leu Phe
20 25 30

Leu Cys His His Ala Leu Leu Ser Ala Arg Cys Thr Ser Ser Val Ser
35 40 45

Asp Val His Val Ile Ala Pro Cys Leu Thr Phe Leu Thr Ala Asn Phe
50 55 60

Val Leu Gln Thr Ala Lys Leu Ser Asp Ala Ala Lys Leu Gln Met Val
65 70 75 80

Tyr Pro Leu Arg Cys Thr Thr Thr Glu Gln Ser Val Ile Thr Ala Arg
85 90 95

Leu Met Cys

<210> 20
<211> 70
<212> PRT
<213> Anaerocellum thermophilum

<400> 20

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

Glu Ile Pro Val His Ser Leu Glu Lys Gly Lys Lys Ser Leu Gly Ile
20 25 30

Ser Ile Arg Phe Leu Glu Ile Gly Lys Trp Lys Asn Leu His Asp Phe
35 40 45

P200100395 WO SEQ.ST25

Pro Arg Thr Phe Leu Gly Phe Val Val Gln Ser Gly Met Asn Asn Tyr
 50 55 60

Ala Ser Gly Tyr Glu Tyr
 65 70

<210> 21
 <211> 73
 <212> PRT
 <213> Anaerocellum thermophilum

<400> 21

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

Leu Gly Tyr Phe Thr Ala Arg Ala Met Glu Gly Tyr Arg Phe Asp Glu
 20 25 30

Phe Glu Gln Gly Ile Ile Lys Lys Leu Ala Asp Lys Cys Asn Val Glu
 35 40 45

Phe Val Phe Glu Asn Gly Lys Ile Leu Gln Ala Arg Tyr Lys Gly Asn
 50 55 60

Leu Phe Tyr Cys Thr Thr Pro Gln Glu
 65 70

<210> 22
 <211> 1863
 <212> DNA
 <213> Anaerocellum thermophilum

<400> 22

ttcgaaaagc agataagggt cggggaaagt aagcatgaag caaagagaga agagcgggag 60
 agatgtgaag ttactggaga gacgtggaat cctgctcgtg tggaaggcat ttttagtttc 120
 tcaacctaca gggagtacgt taaggaggcg ttagaatttg cgaattgggc tcgtactgaa 180
 aaggggtgta aggatttaga acaagcacgg gcctatgtgt ctgaatattt gcagtcgcat 240
 atagacaaag ggtatagcgc gtggactgtt aagaaagaag cagcagccct ggcgaaactg 300
 tatcattgtc gtacaactga ctttaaagta gagcttcccg caagacacag ggaagagatt 360
 gagagaagca ggggatacaa agaccacgat agggagttta gcaaagagag gaatagagac 420
 attatcatct tttcaaaagc tactgggctg agaagaaggg aattggaaag agtgagttct 480
 cgggatatct ttcgtgggccc tgacggaaga ttatatgtgc acgtgagcaa cggcaagggc 540
 ggtagagaaa gggatgttca tgttttgcag aaatacgaga gagagggtga gaggatagtc 600
 agagagcggg aaggaagaga caggctgttt gacaggggtcc ccataaggat ggacgtacac 660

Side 25

P200100395 WO SEQ.ST25

```

agctatagga gggaatatgc aagagagcgt tacagagaag ttgagcgtga gataagtcgt 720
gagagaaaagc ttttcgacag agttgaggat ctcgttcgta gtaggcttac aaggctctat 780
cctgacaggt ttagagaaat tggcgaaaga caacttactc gtgaactcac aagagctgat 840
gggctttatc atcgcagtga tggtagggag tttgaccgcc tggcattgtg ggaagtttca 900
aacgacttgg gacataatcg aattgacggt gttgcaagac actatctgga ttaagcgaat 960
aaaggctcaa gaaagtggat aaaaaaacag ggggtatatt gatatatccc ccctgttttt 1020
tgtgcgtcta caggacctta tttgcgtttc aaggctattc ttgtggggta gtgcagtaaa 1080
aaaggttgcc tttgtatctt gcctgtagaa tcttgccggt ttcaaaaaca aattcaacgt 1140
tgcacttatc agcaagtttc ttgattatc cctgctcaaa ttcgtcgaac ctgtaccctt 1200
ccattgctct tgctgtgaaa taaccagaa ggtacattgc ttcttcacgt gtaaactcaa 1260
ttgttagctt tttcactctc ttttaccccc tttacgagac tgagatattc tttgtacaat 1320
ctgtcaatac gttctgtatc gtcagggtgt acgtttttga gtttttcgta gagattatga 1380
aacaacaggt tctctgctgt tccaataggg tatttctttc tttccgctg gactagcctt 1440
ttcagttggt ttagctttac gttccggatc ttgctcagta ggtcttcgac gatattttca 1500
atctgttttg cttcgtcttc ttcgatgtag aagcatagtt gtttgatttc aagtagcctc 1560
tgaataagtt cctgctttgt taacattcct tgtcaccctt ccccttttca gtctttttta 1620
ttttgtcgta ttctgattcg aactttctca gttggtcagc tctgccgtaa cgtatgcgtc 1680
gaggatttag catgaatcgt ccacggccta ttttaatcaa caaatctttc tcgattagct 1740
ttttgattgc aagatgggat tgcttttttg acatgtcgat gtttcctttg atgtcgtatt 1800
cgtatgaaag tgcaaacgtg tagttgttaa acgagatttt gtgttgtaaa atgtatgaga 1860
gaa 1863

```

<210> 23
 <211> 33
 <212> DNA
 <213> Anaerocellum thermophilum

<400> 23
 cccctgcag gctattcttg tggggtagtg cag 33

<210> 24
 <211> 32
 <212> DNA
 <213> Anaerocellum thermophilum

<400> 24
 cccctgcag ccttgaaacg caaataaggt cc 32

<210> 25
 <211> 22

P200100395 WO SEQ.ST25

<212> DNA
<213> Anaerocellum thermophilum

<400> 25
ccttccacac gagcaggatt cc 22

<210> 26
<211> 26
<212> DNA
<213> Anaerocellum thermophilum

<400> 26
gctttttgat tgcaagatgg tattgc 26

<210> 27
<211> 35
<212> DNA
<213> Anaerocellum thermophilum

<400> 27
gcgggatccg tcgacgaatt cgaaaagcag ataag 35

<210> 28
<211> 27
<212> DNA
<213> Anaerocellum thermophilum

<400> 28
gggctgcagt ctcgtttaac aactaca 27