



Temperature-inducible expression system for thermophilic organisms

Jensen, Sheila Ingemann; Mol, Viviënne Henrieke; Nielsen, Alex Toftgaard

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Jensen, S. I., Mol, V. H., & Nielsen, A. T. (2023). Temperature-inducible expression system for thermophilic organisms. (Patent No. WO2023111304).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



(51) International Patent Classification:

C12N 15/10 (2006.01)

C12N 15/75 (2006.01)

C12N 15/63 (2006.01)

c/o Danmarks Tekniske Universitet, Anker Engelunds Vej 101, 2800 Kongens Lyngby (DK).

(21) International Application Number:

PCT/EP2022/086436

(74) Agent: INSPICOS P/S; Agern Allé 24, 2970 Horsholm (DK).

(22) International Filing Date:

16 December 2022 (16.12.2022)

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

21215326.6

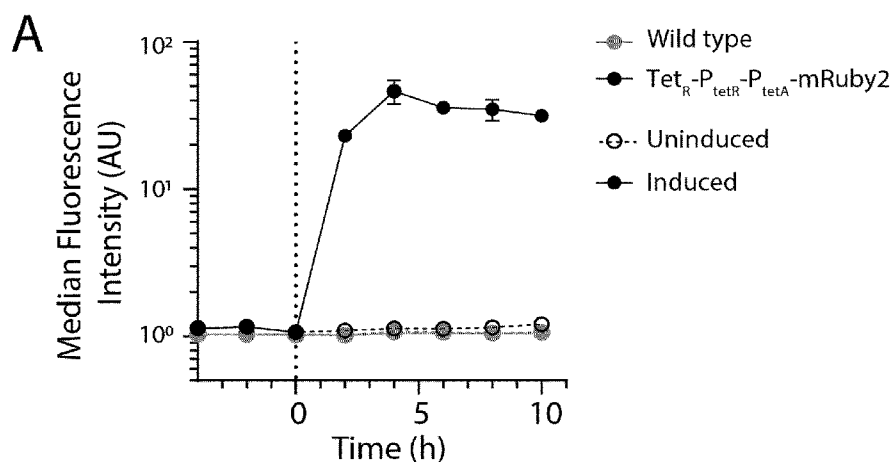
16 December 2021 (16.12.2021) EP

(71) Applicant: DANMARKS TEKNISKE UNIVERSITET [DK/DK]; Anker Engelunds Vej 101, 2800 Kongens Lyngby (DK).

(72) Inventors: JENSEN, Sheila Ingemann; c/o Danmarks Tekniske Universitet, Anker Engelunds Vej 101, 2800 Kongens Lyngby (DK). MOL, Viviënne Henrieke; c/o Danmarks Tekniske Universitet, Anker Engelunds Vej 101, 2800 Kongens Lyngby (DK). NIELSEN, Alex Toftgaard;

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

(54) Title: TEMPERATURE-INDUCIBLE EXPRESSION SYSTEM FOR THERMOPHILIC ORGANISMS



(57) Abstract: Provided are thermophilic bacteria comprising an expression control system comprising a Tet repressor protein, use of such an expression system to control the expression of a gene of interest in thermophilic bacteria by temperature regulation and methods for producing a gene product of interest in thermophilic bacteria by using this system by temperature regulation.



DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI,
SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed
contained color or greyscale and is available for download
from PATENTSCOPE*

TEMPERATURE-INDUCIBLE EXPRESSION SYSTEM FOR THERMOPHILIC ORGANISMS

FIELD OF THE INVENTION

The present invention relates to thermophilic bacteria comprising an expression control system comprising a Tet repressor protein and use of such a system to control the expression of a gene of interest in thermophilic bacteria by temperature regulation.

BACKGROUND OF THE INVENTION

In moving towards a more sustainable bioeconomy, ever-increasing interest is shown in fermentative bioproduction as an alternative to petrochemical synthesis. Thermophilic fermentation ($>45^{\circ}\text{C}$) has been shown to possess inherent advantages over traditional, mesophilic production ($20\text{--}45^{\circ}\text{C}$), including, but not limited to, lower risk of contamination, less cooling energy requirement, efficient pretreated-biomass consumption and higher conversion rates. This altogether provides economic incentive to pursue thermophilic production for (bulk) production of chemicals, which is reflected by the increasing research interest in these production hosts. One factor that delays large-scale application of thermophiles is the present toolbox to optimize productivity in target strains.

One commonly used approach to optimize product formation in fermentations is through the induction of the production-related proteins and pathways by the addition of specific inducer compounds. By doing so, one can ensure that there is a separate growth and production phase of the process. In this manner, no carbon is lost to the products in upstream stages of production, where the target is maximal production of biomass. Conversely, induction of a separate production phase leads to increased control over carbon flux towards the products of interest, in turn resulting in higher overall productivities.

The Tet repressor and promoter (TetR- P_{tet}) system is a well-known example of a system which can be used to control the expression of a protein of interest by the addition of an inducer molecule. The Tet repressor (TetR) binds to the tet operon (tetO) sequence present in the tetA promoter and inhibits protein expression from this promoter. Upon addition of the inducer molecule tetracycline or the chemical analog anhydrotetracycline, TetR loses its ability to bind to the tetO sequence, and expression from the tetA promoter can occur. Use of TetR- P_{tet} systems at mesophilic temperatures has been described (Corrigan et al., *Plasmid*;61:126-129 2008; Kamionka et al., *Appl. Environ. Microbiol.*;71:728-733 2005; Wickstrum et al., *PloS One*;8:e76743 2013; Bertram and Hillen, *Microb. Biotechnol.*;1:2-16 2008).

A disadvantage of using externally added inducers is that it leads to an added material cost, as well as introducing an additional risk of contamination upon addition. Finally, specifically for thermophilic fermentations, high temperatures can result in degradation of inducer molecules, which can create heterogeneous populations in the fermentation, being disadvantageous to overall productivity. Therefore, novel systems are needed to allow induction of expression without the need to add expensive, exogenous compounds to the fermentation.

SUMMARY OF THE INVENTION

It has been found by the present inventors that an expression control system comprising a Tet repressor (TetR) protein and a promoter comprising a tet operator (tetO) sequence can be used to control protein expression in a thermophilic host cell by temperature-regulation at thermophilic temperatures, in contrast to the classical induction of such a system by addition of the inducer molecule tetracycline or derivatives thereof.

So, in a first aspect, the present invention relates to a thermophilic host cell comprising a heterologous expression control system comprising a Tet repressor (TetR) protein and a first promoter comprising a tet operator (tetO) sequence to which the TetR protein can bind, wherein the first promoter is operatively linked to at least one gene of interest.

In a second aspect, the present invention relates to use of an expression control system comprising a TetR protein and a first promoter comprising a tetO sequence to which the TetR protein can bind to control the expression of at least one gene of interest in a thermophilic host cell by temperature regulation.

In some embodiments, the thermophilic host cell further comprises a gene encoding the TetR protein, optionally wherein the gene encoding the TetR protein is operatively linked to a second promoter.

In some embodiments,

(a) when the thermophilic host cell is incubated in a culture medium having a first temperature, the expression of the gene of interest is not induced; (b) when the thermophilic host cell is incubated in a culture medium having a second temperature, the expression of the gene of interest is induced; and (c) the second temperature is at least about 4°C higher than the first temperature.

In a third aspect, the present invention relates to a method of controlling the expression of at least one gene of interest in a thermophilic host cell comprising an expression control system

comprising a TetR protein and a first promoter comprising a tetO sequence, wherein the first promoter is operatively linked to the at least one gene of interest, the method comprising:

(i) incubating the thermophilic host cell in a culture medium having a first temperature; and

(ii) incubating the thermophilic host cell in a culture medium having a second temperature which is at least about 4°C higher than the first temperature, wherein steps i) and ii) can be conducted in any order,

wherein the expression of the at least one gene of interest is induced at the second temperature but not at the first temperature, thereby controlling the expression of the at least one gene of interest.

In a fourth aspect, the present invention relates to a method of producing a gene product of interest in a thermophilic host cell comprising an expression control system comprising a TetR protein and a first promoter comprising a tetO sequence to which the TetR protein can bind, wherein the first promoter is operatively linked to a gene of interest encoding the gene product of interest, the method comprising:

(i) incubating the thermophilic host cell in a culture medium having a first temperature; and

(ii) incubating the thermophilic host cell in a culture medium having a second temperature which is at least about 4°C higher than the first temperature; and

(iii) optionally, isolating the gene product,

wherein the expression of the gene encoding the gene product of interest is induced at the second temperature but not at the first temperature, thereby producing the gene product of interest.

In some embodiments, the method according to the third or fourth aspect comprises a prior step of transforming or transfecting the thermophilic host cell with one or more nucleic acid constructs comprising the first promoter, the at least one gene of interest and a gene encoding the TetR protein, optionally wherein the gene encoding the TetR protein is operatively linked to a second promoter.

In some embodiments, the medium has not been supplemented with tetracycline or an analog thereof which binds to the TetR protein.

In some embodiments, the first temperature is at most about 60°C, the second temperature is at least about 65°C, or both.

In some embodiments, (a) when the TetR protein is bound to the tetO sequence of the first promoter, the promoter is repressed; and (b) when the TetR protein is not bound to the tetO sequence of the first promoter, the promoter is active.

In some embodiments, the TetR protein

- 5 a) comprises an amino acid sequence at least 60% identical to a sequence selected from SEQ ID NOs:1, 32-56, and 79-82,
- b) has a melting temperature between 40°C and 90°C, preferably between 55°C and 90°C, more preferably between 60°C and 90°C,
- 10 c) comprises a helix-turn-helix (HTH) region, which comprises a DNA-binding domain and a sub-region of at least 4 consecutive amino acid residues, which sub-region destabilizes upon temperature increase, optionally wherein the sub-region has a change in root mean square fluctuations (Δ RMSF) score above 1.5 times the average Δ RMSF score for the whole protein in dimeric form when measured by molecular dynamic analysis using the software program CABS-flex 2.0, or
- 15 (d) a combination of (a) and (b), (a) and (c) or all of (a) to (c).

In some embodiments, (a) the TetR protein comprises an amino acid sequence at least 60% identical to SEQ ID NO:1; and/or (b) the tetO sequence comprises a nucleotide sequence at least 65% identical to SEQ ID NO:5.

- 20 In some embodiments, the expression control system comprises a TetR protein comprising an amino acid sequence at least 90% identical to SEQ ID NO:1, a tetO nucleotide sequence at least 90% identical to SEQ ID NO:5, a first promoter comprising a nucleotide sequence at least 90% identical to SEQ ID NO:14, and, optionally, a second promoter comprising a nucleic acid sequence at least 90% identical to SEQ ID NO:13.

- 25 In some embodiments, the thermophilic host cell has an optimum growth temperature in the range of about 55°C to about 65°C, such as at about 60°C.

In some embodiments, the thermophilic host cell is a bacterium.

- 30 In further embodiments, the bacterium is selected from the group consisting of: *Parageobacillus thermoglucosidasius*, *Geobacillus toebii*, *Geobacillus stearothermophilus*, *Geobacillus thermodenitrificans*, *Geobacillus kaustophilus*, *Geobacillus thermoleovorans*, *Geobacillus thermocatenulatus*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharotyticum*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter mathranii*, *Thermoanaerobacter pseudoethanolicus*, *Thermoanaerobacter brockii*, *Thermoanaerobacter kivui*, *Thermoanaerobacter brockii*,

Caldanaerobacter subterraneus, *Clostridium thermocellum*, *Clostridium thermosuccinogenes*, *Thermoclostridium stercorarium*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus smithii*, *Bacillus methanolicus*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor bescii*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Moorella thermoacetica*, *Moorella thermoautotrophica*, *Thermus thermophilus*, *Thermus aquaticus*, *Thermotoga maritima*, *Pseudothermotoga lettingae*, *Pseudothermotoga thermarum*, *Chloroflexus aurantiacus*, *Anaerocellum thermophilum*, *Rhodothermus marinus*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Thermococcus barophilus*, *Thermococcus kodakarensis*, *Pyrococcus abyssi*, and *Pyrococcus furiosus*.

In a sixth aspect, the present invention relates to a thermophilic host cell comprising a heterologous expression control system comprising a DNA-binding protein, which is at least 80% identical to an amino acid sequence selected from: SEQ ID NOs:1, 32-56, and 79-82, and a DNA sequence to which the DNA-binding protein can bind.

All embodiments of aspects 1-5 apply mutatis mutandis to the sixth aspect of the invention.

These and other aspects and embodiments of the invention are described in further detail below.

LEGENDS TO THE FIGURES

Fig. 1: A) and B) mRuby2 expression over time at 52°C for a wild type strain and a strain with the Tet_R-P_{tetR}-P_{tetA}-mRuby2 construct integrated, as determined by flow cytometry. Expression was induced by addition of anhydrotetracycline. AU = arbitrary units.

Fig. 2: A) β-galactosidase (BgaB) expression over time at 60°C for a wild type strain and a strain with the Tet_R-P_{tetR}-P_{tetA}-BgaB construct integrated, as determined by absorbance measurements. Expression was induced by addition of anhydrotetracycline. The BgaB readout is expressed in Miller units. B) Growth curves of the cultures represented in (A), measured through OD₆₀₀.

Fig. 3: BgaB expression over time for a wild type strain and a strain with the Tet_R-P_{tetR}-P_{tetA}-BgaB construct integrated, as determined by absorbance measurements. Expression was induced by a shift in temperature. Both strains were measured in uninduced (60°C constant) and induced (60°C to 65°C shift) states.

Fig. 4: Schematic illustration of the temperature regulation-principle of the present invention. At temperatures below 60°C, the tetR repressor represses the expression of the gene of interest (GOI), which is under the control of the tetA promoter. At temperatures above 65°C, the tetR protein is destabilized and can no longer repress expression from the tetA promoter, and thereby the GOI is expressed.

Fig. 5: Molecular dynamic (MD) simulation of the DNA-bound TetR homodimer, at low (dotted line) and high (solid line) temperature, shown as average root mean square fluctuation (RMSF) value in Ångström (Å) per residue of the crystal structure. Light grey shading indicates the conserved helix-turn-helix (HTH) region of the TetR family, where dark grey indicates the DNA-binding domain of the protein.

Fig. 6: Percentage identity matrix as a result of Multiple Alignment using Fast Fourier Transform (MAFFT) for all TetR family proteins used in this work. Shading indicates percentage identity score. SEQ ID NOs referring to the sequence listing of the amino acid sequences used for generating this data can be found in Table 9.

Fig. 7: Percentage identity matrix as a result of Multiple Alignment using Fast Fourier Transform (MAFFT) for the helix-turn-helix (HTH) regions of the TetR family proteins used in this work. Shading indicates percentage identity score. SEQ ID NOs referring to the sequence listing of the amino acid sequences used for generating this data can be found in Table 9.

Fig. 8: Difference in root mean square fluctuation ($\Delta RMSF$ in Ångström, Å) values for 8 representative proteins in the TetR family, when comparing low versus high temperature. The dotted line shows the average $\Delta RMSF$ values across the whole protein ($\bar{x}_{\Delta RMSF}$). Light grey shading indicates the helix-turn-helix (HTH) region, where dark grey depicts the DNA-binding domain of each protein. Destabilized regions are defined where $\Delta RMSF > 1.5\bar{x}_{\Delta RMSF}$ for at least four consecutive residues, and their amino acid positions indicated in each plot. PDB ID numbers of crystal structures used for each protein can be found in Table 8.

DETAILED DISCLOSURE OF THE INVENTION

Definitions

As used herein, the term “thermophilic cell” refers to any cell, such as a microbial cell, *e.g.*, a bacterial, eukaryotic or archaeal cell, which has an optimum growth temperature of at least about 40°C, such as at least about 45°C, such as at least about 50°C, such as at least about

55°C. Typically, a thermophilic cell as referred to herein has an optimum growth temperature in the range of about 40°C to about 125°C or about 45°C to about 125°C or about 55°C to about 125°C, such as in the range of about 40°C to about 80°C or about 45°C to about 80°C or about 55°C to about 80°C.

- 5 As used herein, the term "mesophilic cell" refers to any cell, such as a microbial cell, e.g., a bacterial cell, with has an optimum growth temperature of at most about 45°C, such as an optimum growth range in the range of about 20 to about 45 °C. Non-limiting examples of mesophilic cells are bacteria of the *Escherichia* genera, such as *Escherichia coli*.

- 10 As used herein, a "host cell" is a cell, such as a thermophilic cell, which comprises at least one transgene, e.g., a heterologous gene.

- As used herein, the term "Tet repressor" or "TetR" refers to a protein which is a member of the TetR family of transcriptional repressors found in bacterial cells, as well as functionally active fragments and variants thereof (see, e.g., Ramos et al., Microbiology and Molecular Biology Reviews 2005;69(2):326–356 and Cuthbertson and Nodwell, Microbiology and Molecular Biology Reviews 2013;77(3) doi:10.1128/MMBR.00018-13). At least some TetR proteins contains a helix-turn-helix (HTH) structural motif, which reportedly provides for their binding to the target DNA sequence (Ramos et al., 2005). Preferred TetR proteins are from mesophilic gram-negative bacterial cells. An example of a TetR protein is the *Escherichia coli* TetR protein (SEQ ID NO:1; UniProt ID: P0ACT4). A TetR protein according to the present invention may
15 for example be a protein with at least about 80% sequence identity to SEQ ID NO:1. Another example of a TetR protein is PfmR from the thermophilic bacterium *Thermus thermophilus* HB8 (SEQ ID NO:74; UniProt ID: Q53WD9). A TetR protein typically binds to DNA in dimerized form, and, unless otherwise indicated or contradicted by context, the term "TetR protein" may refer to a protein in either monomeric or dimeric form.

- 25 As used herein, the term "tet operator" or "tetO" refers to a target DNA sequence of a TetR protein as described herein. The tetO is typically part of or associated with a promoter, whose ability to allow the expression of a gene to which the promoter is operably linked is regulated by the binding of the TetR protein to the tetO. Non-limiting tetO sequences include native tetO sequences from bacterial cells, such as mesophilic bacterial cells, as well as variants thereof
30 where one or more mutations have been introduced. An example of a tetO sequence is the *E. coli* tetO1 sequence (SEQ ID NO:4).

As used herein, the term "gene" or "coding sequence" refers to a nucleic acid sequence that encodes a biological molecule, such as a protein, optionally including or operably linked to regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding

sequences) the coding sequence. A "transgene" is a gene that has been introduced into a cell, by a genetic engineering technique, such as by transformation. A transgene may be native or heterologous to the cell into which it is introduced. Gene names are herein set forth in italicised text with a lower-case first letter whereas protein names are set forth in normal text with a capital first letter.

Unless otherwise stated, "sequence identity", as used for amino acid sequences or nucleotide sequences herein, is determined by comparing two optimally aligned sequences of equal length according to the following formula: $(N_{\text{ref}} - N_{\text{dif}}) \cdot 100 / N_{\text{ref}}$, wherein N_{ref} is the number of residues or nucleotides in one of the two sequences and N_{dif} is the number of residues or nucleotides which are non-identical in the two sequences when they are aligned over their entire lengths and in the same direction. Hence, the nucleotide sequence AGGTCCTA will have a sequence identity of 87.5% with the sequence AGTTCCTA ($n_{\text{dif}}=1$ and $n_{\text{ref}}=8$).

The sequence identity can be determined by conventional methods, e.g., Smith and Waterman (Adv. Appl. Math.;2:482 1981), by the 'search for similarity' method of Pearson and Lipman (Proc. Natl. Acad. Sci. USA;85:2444 1988), using the CLUSTAL W algorithm of Thompson et al. (Nucleic Acids Res.;22:467380 1994), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group), or the Needleman-Wunsch algorithm (Needleman and Wunsch, J. Mol. Biol.;48:443-453 1970) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., Trends Genet.;16:276-277 2000), e.g., as provided at the European Bioinformatics Institute website (www.ebi.ac.uk). The BLAST algorithm (Altschul et al., Mol. Biol.;215:403-410 1990), for which software may be obtained through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/), may also be used. When using any of the mentioned algorithms, the default parameters for "Window" length, gap penalty, etc., may be used.

The term "expression", as used herein, refers to the process in which a gene is transcribed into mRNA, and may optionally include the subsequent translation of the mRNA into an amino acid sequence, *i.e.*, a protein or polypeptide.

As used herein, the term "promoter" refers to a region of DNA that may be upstream from the start of transcription, and that may be involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A promoter may be operably linked to a coding sequence for expression in a cell, or a promoter may be operably linked to a nucleotide sequence encoding a signal sequence which may be operably linked to a coding sequence for expression in a cell. An "active" promoter is a promoter from which a gene can be actively transcribed, e.g., because there are no transcription factors negatively regulating (*i.e.*,

repressing) gene expression from the promoter and/or because there are transcription factors present which positively regulate (i.e., promote) gene expression from the promoter. A "repressed" promoter is a promoter from which a gene cannot be actively transcribed or from which only a very low level of transcription occurs (so-called "leaky" expression). As used
5 herein, "induced" expression means expression occurring from an active promoter, whereas when expression is "not induced", it means that there is no expression or only leaky expression from a repressed promoter.

The term "operably linked," when used in reference to a regulatory sequence and a coding sequence, means that the regulatory sequence can affect the expression of the linked coding
10 sequence. "Regulatory sequences" or "control elements" refer to nucleotide sequences that influence the timing and level/amount of transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters; translation leader sequences; introns; enhancers; stem-loop structures; repressor binding sequences; termination sequences; polyadenylation recognition sequences, etc. Particular
15 regulatory sequences may be located upstream and/or downstream of a coding sequence operably linked thereto. Also, particular regulatory sequences operably linked to a coding sequence may be located on the associated complementary strand of a double-stranded nucleic acid molecule.

As used herein, the term "expression control system" refers to any system comprising at least
20 two components and capable of controlling the expression of one or more genes of interest. Components of an expression control system may include, but are not limited to, one or more promoters, one or more genes to be expressed, one or more transcription factors and/or genes encoding said transcription factors, wherein the various components act together to control the expression of the one or more genes.

A "transcription factor" is a protein involved in the process of transcribing DNA into RNA. A
25 transcription factor may, for example, initiate and/or regulate the transcription of genes, and usually comprises a DNA-binding domain that gives it the ability to bind to regulatory sequence such as, e.g., enhancer, promoter and operator sequences. One example of a transcription factor is a repressor, which is a DNA- or RNA-binding protein that inhibits the expression of
30 one or more genes by binding to an operator sequence or associated silencer sequence.

The term "temperature regulation", as used herein in the context of regulation of gene expression, refers to any means and mechanism of regulation that results in induction and/or
abolishment of gene expression caused by a shift in temperature of the cell being regulated.

The term "optimum growth temperature", as used herein about thermophilic cells, refers to a temperature or temperature interval where the growth-rate of the thermophilic cells is the highest.

By "tetracycline analog", as used herein, is meant an analog of tetracycline (i.e.,
5 (4S,4aS,5aS,6S,12aR)-4-(dimethylamino)-1,6,10,11,12a-pentahydroxy-6-methyl-3,12-dioxo-4,4a,5,5a-tetrahydrotetracene-2-carboxamide; sold under the name of Sumycin among others), and capable of controlling a TetR-P_{tetA} expression control system in a similar manner as tetracycline. A tetracycline analog preferably belongs to the tetracyclines family, containing the typical structure of at least four fused rings and triggering a conformational change in the
10 TetR protein upon binding which causes it to lose its binding to its target DNA sequence. Examples of tetracycline analogs include, but are not limited to, anhydrotetracycline, doxycycline, demeclocycline, minocycline, chloro-tetracycline, sancycline, metacycline, and tigecycline (see, e.g., Krueger et al., *Biotechniques*;37(4) doi:10.2144/04374BM04 2018).

Specific embodiments of the invention

15 As described in Example 1, it was found by the present inventors that a TetR-P_{tetA} expression control system retained its function, i.e. the expression of a gene under the control of the P_{tetA} promoter was repressed by TetR, when the system was used in the thermophilic bacterium *Parageobacillus thermoglucosidasius* at a temperature of 60°C. Surprisingly, when the temperature was raised to 65°C, expression of the reporter gene was induced.

20 Thus, the present inventors have unexpectedly discovered that expression control systems previously used in mesophilic host cells can be sufficiently stable also for use in thermophilic host cells at thermophilic temperatures. Furthermore, it was found that a TetR-P_{tetA} expression control system can, at thermophilic temperatures, be regulated by another mechanism than chemical induction by tetracycline, namely by temperature. Without being limited to theory,
25 and as supported by the results in Example 2, a temperature increase may destabilize a TetR protein in a thermophilic cell, causing it to lose its ability to bind to the P_{tetA} promoter and thereby its ability to repress the expression of the gene whose expression it controls.

This discovery enables the use of this well-known expression control system also in thermophilic organisms, which are advantageous production hosts for fermentative
30 bioproduction (as described in the background section). The toolbox available for controlling expression in these organisms is thereby expanded. Moreover, temperature-regulation of this expression control system enables its use without the external addition of inducer molecules, which can save expenses as well as decrease the likelihood of adding contaminants.

The expression control system can be used in any suitable thermophilic cell. Thermophilic cells may, for example, be isolated from natural sources (e.g., isolated from hot springs) or obtained from commercial sources. They may also potentially be generated in a laboratory, e.g., by adapting mesophilic cells to grow at higher temperatures by adaptive laboratory evolution.

- 5 Preferably, in the context of the present invention, the thermophilic cell is a thermophilic bacterial cell, such as a gram-positive thermophilic bacterial cell.

Particularly contemplated are thermophilic host cells with an optimum growth temperature which is higher than about 40°C, such as at least about 45°C, such as at least about 50°C, such as at least about 55°C, such as at least about 56°C, at least about 57°C, such at least
10 about 58°C, such as at least about 59°C, such as at least about 60°C. For example, the thermophilic host cell may have an optimum growth temperature in the range of about 40°C to about 125°C, such as about 45°C to about 125°C, such as about 50°C to about 125°C, such as about 50°C to about 100°C, such as about 50°C to about 80°C, such as about 50°C to about 75°C, such as about 50°C to about 70°C, such as about 50°C to about 65°C, such as about
15 50°C to about 60°C, such as about 55°C to about 125°C, such as about 55°C to about 100°C, such as about 55°C to about 80°C, such as about 55°C to about 75°C, such as about 55°C to about 70°C, such as about 60°C to about 125°C, such as about 60°C to about 100°C, such as about 60°C to about 80°C, such as about 60°C to about 75°C, such as about 60°C to about 70°C, such as about 60°C to about 65°C, such as about 50°C to about 60°C, such as about
20 55°C to about 65°C. For example, the thermophilic host cell may have an optimum growth temperature in the range of about 50°C to about 65°C, such as about 55°C to about 60°C, such as at about 60°C.

The optimum growth temperature of a thermophilic cell can be determined by methods known to a person of skill in the art. Typically, the growth rate of the thermophilic cells is measured
25 over a range of about 20°C to about 40°C at temperature intervals of about 2°C to 10°C. For the thermophilic cell in question, a growth medium is chosen which contains appropriate media components that facilitate growth. The effect of temperature on growth rate can then be determined from the exponential increase in optical density measured at 440 nm. If a more detailed assessment is needed, the assay can be repeated using e.g. different media
30 components, media adjusted to a different pH, and/or a smaller temperature range which includes the initially determined optimum growth temperature but with a smaller interval between the growth temperatures, e.g., in the range of 1 to 5 °C.

Non-limiting examples of thermophilic bacterial cells are those belonging to a genus selected from: *Geobacillus*, *Parageobacillus*, *Thermoanaerobacterium*, *Thermoanaerobacter*,
35 *Caldanaerobacter*, *Bacillus*, *Thermoclostridium*, *Anoxybacillus*, *Caldicellulosiruptor*, *Moorella*,

Thermus, *Thermotoga*, *Pseudothermotoga*, *Chloroflexus*, *Anaerocellum*, *Rhodothermus*, *Sulfolobus*, *Thermococcus*, *Pyrococcus*, and *Clostridium*.

For example, the host cell may be a bacterium belonging to a species selected from: *Parageobacillus thermoglucosidasius*, *Geobacillus toebii*, *Geobacillus stearothermophilus*,
 5 *Geobacillus thermodenitrificans*, *Geobacillus kaustophilus*, *Geobacillus thermoleovorans*,
Geobacillus thermocatenulatus, *Thermoanaerobacterium xylanolyticum*,
Thermoanaerobacterium saccharolyticum, *Thermoanaerobacterium thermosaccharolyticum*,
Thermoanaerobacter mathranii, *Thermoanaerobacter pseudoethanolicus*, *Thermoanaerobacter*
brockii, *Thermoanaerobacter kivui*, *Thermoanaerobacter brockii*,
 10 *Caldanaerobacter subterraneus*, *Clostridium thermocellum*, *Clostridium thermosuccinogenes*,
Thermoclostridium stercorarium, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus smithii*,
Bacillus methanolicus, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus*
gonensis, *Caldicellulosiruptor bescii*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor*
kristjanssonii, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*, *Moorella*
 15 *thermoacetica*, *Moorella thermoautotrophica*, *Thermus thermophilus*, *Thermus aquaticus*,
Thermotoga maritima, *Pseudothermotoga lettingae*, *Pseudothermotoga thermarum*,
Chloroflexus aurantiacus, *Anaerocellum thermophilum*, *Rhodothermus marinus*, *Sulfolobus*
acidocaldarius, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Thermococcus barophilus*,
Thermococcus kodakarensis, *Pyrococcus abyssi*, and *Pyrococcus furiosus*.

20 Non-limiting examples of thermophilic cells are bacteria of the *Parageobacillus* genus. Preferably, the thermophilic host cell is a *Parageobacillus thermoglucosidasius*. Other examples of thermophilic cells are *Caldicellulosiruptor bescii*, *Clostridium thermocellum*, and *Moorella thermoacetica*.

The expression control system introduced into the thermophilic cell may comprise a TetR
 25 protein and a first promoter comprising a tetO sequence to which the TetR protein can bind. The expression control system or the thermophilic host cell may also further comprise a gene encoding the TetR protein, optionally wherein the gene encoding the TetR protein is operatively linked to a second promoter. Alternatively, the expression control system introduced into the thermophilic cell may comprise a gene encoding a TetR protein, a first promoter comprising a
 30 tetO sequence to which the TetR protein can bind, wherein the gene encoding the TetR protein is operatively linked to a second promoter. Typically, in the expression control system or the thermophilic host cell, the first promoter is operatively linked to a gene of interest. The expression control system may further comprise one or more additional regulatory sequences as described elsewhere herein, e.g., additional tetO sequences.

In one non-limiting embodiment, the expression control system introduced into the thermophilic host cell comprises the following components, e.g., as a part of one or more nucleic acid constructs, such as a plasmid:

- (i) a P_{tetA} promoter comprising a first tetO (tetO1) sequence to which the TetR protein can bind and operatively linked to a gene of interest;
- (ii) a gene encoding a TetR protein, operatively linked to a P_{tetR} promoter which comprises a second tetO (tetO2) sequence to which the TetR protein can bind;
- (iii) An origin of replication, suited for Gram-negative, mesophilic hosts. E.g., BBR1;
- (iv) An origin of replication suited for the thermophilic host cell, e.g., RepB, and/or two homology arms, allowing genome integration of the construct;
- (v) A selection marker suitable for selection in mesophiles and/or thermophiles. E.g., a kanamycin resistance gene;
- (vi) Terminators.

Although it is contemplated that one or more components of the expression control system may exist naturally in (i.e., be endogenous to) the thermophilic cell, at least one component of the expression control system is typically heterologous to the thermophilic cell.

In some cases, the thermophilic host cell may take up a synthesized linear DNA strand, either single- or double-stranded, in the absence of a plasmid vector. Thus, in some embodiments, a linear piece of DNA encoding the expression control system according to the invention and sequences for homologous recombination may be taken up by the cell and become integrated into its genome. The host cell may alternatively be transduced with a viral vector encoding the expression control system.

As described herein, the temperature-regulated expression control system can provide for thermophilic host cells, uses and methods characterized by the following: When the thermophilic host cell is incubated in a culture medium having a first temperature, the expression of the gene of interest is not induced, whereas when the thermophilic host cell is incubated in a culture medium having a second temperature, the expression of the gene of interest is induced. The second temperature may be at least about 3°C, such as at least about 4°C, such as at least about 5°C, such as at least about 6°C, such as at least about 7°C, such as at least about 8°C, such as at least about 9°C, such as at least about 10°C higher than the first temperature, e.g., at least 15°C, 20°C, or 25°C higher, such as at most 30°C higher than the first temperature. The second temperature may, for example, be at least about 4°C higher than the first temperature. The second temperature may alternatively be at least about 5°C higher, such as about 5°C higher, than the first temperature.

The first and second temperatures can be selected depending on the characteristics of the expression control system in question as well as on the growth characteristics of the thermophilic cell. It may be advantageous if the expression control system provides for temperature-based regulation at temperatures close to the optimum growth temperature of the thermophilic host cell in question. Accordingly, the first temperature, the second temperature, or both may be within about 5°C, such as within about 10°C, of the optimum growth temperature. However, in other cases it may be more beneficial if the expression control system provides regulation at temperatures distant from the optimum growth temperature, since this may provide for separate phases of growth and production.

Many TetR proteins are known in the art (see, e.g., Ramos et al., 2005). TetR proteins suitable for the expression control systems, thermophilic host cells, uses and methods described herein are not limited to any specific TetR proteins, originating from or expressed in any particular species. Particularly contemplated are, however, TetR proteins from mesophilic bacteria and functionally active fragments and variants thereof. Non-limiting examples of TetR proteins include those shown in Table 1, those shown in Table 8, as well as functionally active variants and fragments thereof. Particularly preferred is a TetR protein comprising the amino acid sequence of SEQ ID NO:1, or a functionally active fragment or variant thereof. In some embodiments, the TetR protein has the sequence of SEQ ID NO:1. Functionally active fragments and variants of, e.g., *E. coli* TetR (SEQ ID NO:1; 207 amino acids) can be identified by a person of skill in the art.

Table 1 – Examples of TetR proteins

Organism	% identity to <i>E. coli</i> TetR (SEQ ID NO:1)	Database Reference Sequence
<i>Escherichia coli</i>	100	P0ACT4 (UniProt ID) (SEQ ID NO:1)
<i>Providencia stuartii</i>	99.52	WP_183124523.1 (NCBI) (SEQ ID NO:79)
<i>Enterobacter roggenkampii</i>	99.52	WP_165473055.1 (NCBI) (SEQ ID NO:80)
<i>Shigella flexneri</i>	99.50	WP_072195351.1 (NCBI) (SEQ ID NO:81)
<i>Acinetobacter baumannii</i>	99.33	WP_047939682.1 (NCBI) (SEQ ID NO:82)

For example, the full-length sequence can be truncated from the N- and/or C-terminus, typically to remove residues or portions not necessary for its DNA-binding activity and/or expression control activity. For example, such a fragment may comprise at least about 60%, such as at least about 70%, such as at least about 80%, such as at least about 90%, such as at least about 95%, of the full length of the protein, counting by its amino acid residues in the native, mature sequence. Alternatively, such a fragment can be obtained by removing, e.g., 1, 2, 3, 4, 5, 10, 20 or more amino acid residues from the C-terminus, N-terminus, or both. In TetR proteins where the HTH motif is located in the N-terminal portion, the fragment is preferably obtained by removing one or more amino acids only from the C-terminal portion.

Variants comprising one or more amino acid substitutions, deletions or insertions can also be prepared and identified by screening or testing them for their DNA-binding activity and/or expression control activity. For example, a TetR protein as described herein may comprise an amino acid sequence at least about 60%, such as at least about 65%, such as at least about 70%, such as at least about 75%, such as at least about 80%, such as at least about 85%, such as at least about 90%, such as at least about 95%, such as at least about 96%, such as at least about 97%, such as at least about 98%, such as about 99%, identical to any one of the sequences with SEQ ID NOs:1, 32-56, and 79-82, more preferably to that of *E. coli* TetR (SEQ ID NO:1). Alternatively, such a variant may comprise up to 10 mutations, such as 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 mutations. In some particularly contemplated TetR variants, the variant differs from the native sequence only by conservative amino acid substitutions.

Variants of *E. coli* TetR are known in the art. For example, Ettner et al. (1996) generated 50 single amino acid TetR variants. Muller et al. (1995) generated tetR variants which did not respond to tetracycline inducer but most retained their DNA binding. The mutations were introduced as single amino acid substitutions distributed in different secondary structures throughout the protein sequence; predominantly in the interior of the protein. Only 6 of the 93 variants tested showed decreased binding to tetO. Kamionka et al. (2004) generated a reverse tetR (i.e. which binds to tetO in the presence of tetracycline but does not bind to tetO in the absence of tetracycline) by introducing 2 point mutations; G96E and L205S. Scholz et al. (2004) introduced mutations to change TetR from responding to tetracycline to another inducer; 4-de(dimethylamino)-6-deoxy-6-demethyl-tetracycline (cmt3). The variants still bound to tetO in absence of inducer. Additionally, since the crystal structure of *E. coli* TetR in complex with tetO has been reported (Orth et al., 2001), amino acids suitable for substitution or deletion can be identified and tested. Generally, protein segments outside of the secondary structures (α -helices and β -strands) are more flexible and would allow for mutagenesis without causing changes in functionality. Moreover, conservative amino acid substitutions (e.g., R to K mutations or E to D mutations) can be introduced.

A "conservative" amino acid substitution in a protein is one that does not negatively influence protein activity. Typically, a conservative substitution can be made within groups of amino acids sharing physicochemical properties, such as, *e.g.*, basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagines), hydrophobic amino acids (leucine, isoleucine, valine and methionine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, and threonine). Most commonly, substitutions can be made between Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, Asp/Gly.

In some embodiments, the C- or N-terminus of the TetR protein, or functionally active variant or fragment thereof, is fused to a second polypeptide, such as a protein tag or linker. Suitable protein tags for facilitating detection of a protein or for improving stability, solubility or other properties of a protein are well-known in the art. Non-limiting proteins tags for facilitating protein detection include Green Fluorescent protein (GFP) and Red Fluorescent Protein (RFP). Linkers which may, for example, be used as spacers in fusion proteins of a TetR protein with another protein include (G₄S)₄-His₆.

Suitable assays for testing the expression control activity (and, without being limited to theory, thereby also the DNA binding activity) of such fragments or variants or fusion proteins in an expression control system are provided in Example 1. For example, a *P. thermoglucosidasius* DSM2542Z Δ ldh::TetR-PtetR-PtetA-BgaB test strain can be prepared where the gene encoding the *E. coli* TetR protein is replaced by a gene encoding the fragment or variant to be tested. Preferably, when the thermophilic host cell is incubated in a culture medium having a first temperature, the expression of the gene of interest is not induced; when the thermophilic host cell is incubated in a culture medium having a second temperature, the expression of the gene of interest is induced; and the second temperature is at least about 4°C higher than the first temperature, *e.g.*, at least about 5°C higher than the first temperature. For example, the first temperature may be about 60°C and the second temperature about 65°C.

As described in Example 1, the coding sequence for a TetR protein may be modified to remove predicted restriction sites and for the purpose of codon optimization. For example, the coding sequence of SEQ ID NO:2 may be modified to SEQ ID NO:3 (see Table 4).

In some embodiments, the TetR protein may have a melting temperature between 40°C and 90°C, such as between 45°C and 55°C, between 50°C and 60°C, between 55°C and 65°C, between 60°C and 70°C, between 65°C and 75°C, between 70°C and 80°C, between 75°C and 85°C, or between 80°C and 90°C.

In some embodiments, the TetR protein comprises a helix-turn-helix (HTH) region, which comprises a DNA-binding domain and a sub-region of at least 4 consecutive amino acid residues, which sub-region destabilizes upon temperature increase.

5 In further embodiments, the sub-region has a change in root mean square fluctuations (Δ RMSF) score above 1.5 times the average Δ RMSF score for the whole protein in dimeric form when measured by molecular dynamic analysis using the software program CABS-flex 2.0.

The calculation of Δ RMSF scores is described in Example 2.

10 The sub-region may comprise between 4 and 100 consecutive amino acid residues, such as at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 40, and such as at least 50, consecutive amino acid residues.

15 The sub-region may overlap with the DNA-binding domain either partially or fully, or it may not overlap with the DNA-binding domain at all.

In some embodiments, the sub-region overlaps with the DNA-binding domain.

In some embodiments, the sub-region overlaps fully with the DNA-binding domain.

In some embodiments, the sub-region overlaps partially with the DNA-binding domain.

In some embodiments, the sub-region does not overlap with the DNA-binding domain.

20 By "overlaps fully" is meant that the whole sub-region is comprised within the boundaries of the DNA-binding domain. By "overlaps partially" is meant that only a part of the sub-region, such as at least 1, at least 2, at least 3, or at least 4 consecutive amino acid residues of the sub-region, is within the boundaries of the DNA-binding domain.

In some embodiments, a TetR protein may comprise a sub-region comprising or consisting of:

25 (a) amino acid residues 26-47 of tetR (SEQ ID NO:1);

(b) amino acid residues 36-62 of amtR (SEQ ID NO:33);

(c) amino acid residues 31-42 of icaR (SEQ ID NO:37);

(d) amino acid residues 26-44 of lmrA (SEQ ID NO:32);

(e) amino acid residues 36-46 of qacR (SEQ ID NO:36);

(f) amino acid residues 39-58 of smcR (SEQ ID NO:41);

5 (g) amino acid residues 18-46 of pfmR (SEQ ID NO:51);

(h) amino acid residues 23-53 of fadR (SEQ ID NO:53); or

(i) a variant of any one of (a) to (h), which has 1, 2 or 3 amino acid substitutions, deletions or insertions.

Preferably, the sub-region domain overlaps at least partially with the DNA binding domain.

- 10 The expression control system introduced into the thermophilic host cell may comprise a second promoter which is operatively linked to a gene encoding the TetR protein. Suitable second promoters which can induce the expression of a TetR protein to which it is linked in a thermophilic bacterial cell can be chosen by a person of skill in the art depending on, e.g., the thermophilic host cell and/or the TetR protein. In theory any constitutive promoter could be
- 15 used, as long as it works in the organism of choice at a high enough level to have sufficient TetR protein to allow for controlling the promoter(s) containing the tetO binding sequence. It could also be in a circuit (e.g., be controlled by another input) as well as be designed to be self-regulated. As an example, the gene encoding the TetR protein may be operatively linked to the PXyIR promoter. As described in Example 1, the second promoter may be a modified
- 20 PXyIR promoter, such as the one with SEQ ID NO:13, which is originally derived from *Bacillus subtilis*. This promoter is based on the pRMC2 plasmid from Corrigan et al., 2009, wherefrom the modified *B. subtilis* PXyIR promoter was copied. The -10 box was then replaced with the *B. subtilis* consensus sequence (Corrigan et al., 2009; Fagan and Fairweather, 2011). Suitable promoters also include any native promoter which regulates the expression of a native TetR
- 25 protein in a bacterial cell, e.g., an *E. coli* cell. Preferred promoters are at least about 85%, such as at least about 90% or 100% identical to the promoter with SEQ ID NO:13.

The expression control system may comprise a first promoter which comprises a tetO sequence which can be bound by the TetR protein, e.g., the *E. coli* TetR protein having the amino acid sequence of SEQ ID NO:1, and/or fragments or variants thereof. Many tetO sequences are

known in the art. The tetO sequence suitable for the expression control systems, thermophilic host cells, uses and methods described herein are thereby not limited to any specific tetO sequence, originating from or present in any particular species. Moreover, variants of native tetO sequences can be prepared and tested for their binding activity to a TetR protein as described herein. Preferred tetO sequences include those that comprise a nucleotide sequence having at least about 65%, such as at least about 70%, such as at least about 75%, such as at least about 80%, such as at least about 85%, such as at least about 90%, and such as at least about 95%, such as 100% sequence identity to the tetO1 sequence tetO1-pRMC2 (SEQ ID NO:5). For example, it has previously been shown that certain mutated tetO sequences having a sequence identity to the tetO1 sequence from *E. coli* (SEQ ID NO:4, 94.7% identical to SEQ ID NO:5) of 68.4% still retain the ability to be bound by TetR proteins (Reichheld et al., 2009, Sizemore et al., 1990). Some specifically contemplated tetO sequences and variants thereof are shown in Table 2. Preferably, the tetO sequence comprises SEQ ID NO:5.

Table 2 – Examples of tetO sequences

Name	5'→3' Sequence	% identity to tetO1	SEQ ID NO
tetO1 ¹	ACTCTATCAATGATAGAGT	100	4
tetO1-pRMC2 ²	ACTCTATCATTGATAGAGT	94.7	5
tetO2 ³	TCCCTATCAGTGATAGAGA	78.9	6
tetO2-M3 ⁴	TCCCTAACAGTGTTAGAGA	68.4	7
tetO2-M4 ⁴	TCCCTTTCAGTGAAAGAGA	68.4	8
tetO2-M5 ⁴	TCCCAATCAGTGATTGAGA	68.4	9
tetO2-M6 ⁴	TCCGTATCAGTGATACAGA	68.4	10
tetO2-pWH1012-5T ⁵	TCCCCATCAGTGATGGAGA	68.4	11

¹ Fagan et al., 2011; ² Corrigan et al., 2009; ³ Bertram et al., 2007; ⁴ Reichheld et al., 2009; ⁵ Sizemore et al., 1990

Any promoter or promoter region which can include a tetO sequence to which the TetR protein can bind and which can be operatively linked to at least one gene of interest can be used as the first promoter in the expression control systems, thermophilic host cells, uses and methods described herein. Suitable first promoters which can induce the expression of a gene of interest to which it is linked in a thermophilic bacterial cell can be chosen by a person of skill in the art depending on, e.g., the thermophilic host cell and/or the gene of interest. Preferably, a first promoter should have a suitable -10 and -35 site and a suitable tetO binding site. These sites

could possibly be adjusted to give the system a preferred induction temperature and/or a preferred level of leakiness. Furthermore, the sequence prior to the gene of interest should preferably have an appropriate ribosomal binding site. Suitable promoters also include any native promoter which regulates the expression of a native TetA protein in a bacterial cell, e.g., an *E. coli* cell, as well as any modified promoter derived from such native promoter. As described in Example 1, a suitable promoter may be a modified P_{tetA} promoter originally derived from *E. coli*, such as the promoter with SEQ ID NO:12. This promoter is based on the pRPF185 plasmid from Fagan and Fairweather, 2011, wherefrom the modified *E. coli* Tn10 P_{tetA} sequence, adapted for use in Gram positive bacteria, was copied. As is also described in Example 1, the promoter with SEQ ID NO:12 may be further optimized by removing possible restriction sites, resulting in the promoter with SEQ ID NO:14. Preferred promoters are at least about 85%, such as at least about 90% or 100% identical to the promoter with SEQ ID NO:12 or at least about 80%, such as at least about 85%, such as at least about 90% or 100% identical to the promoter with SEQ ID NO:14.

In particular thermophilic host cells, uses and methods, the TetR protein comprises an amino acid sequence at least 60% identical to SEQ ID NO:1 and/or the tetO sequence comprises a nucleotide sequence at least 65% identical to SEQ ID NO:5. For example, in the thermophilic host cell, use or the method, the expression control system may comprise a TetR protein comprising an amino acid sequence at least 90% identical to SEQ ID NO:1, a tetO nucleotide sequence at least 90% identical to SEQ ID NO:5, a first promoter comprising a nucleotide sequence at least 90% identical to SEQ ID NO:14, and, optionally, a second promoter comprising a nucleic acid sequence at least 90% identical to SEQ ID NO:13.

Constructs, e.g., a vector such as a plasmid, comprising (or encoding for) the components of the expression control system described herein can be introduced into the thermophilic host cell using a standard protocol, e.g., a protocol similar to the one used in Example 1. Any suitable procedure for introducing heterologous polynucleotide sequences into thermophilic host cells can be used. For example, only a single copy of each component of the expression control system and transgene may be introduced. Alternatively, more than a single copy, such as two copies, three copies or more than three copies of the expression control system components and transgene may be introduced. As is understood by the skilled artisan, multiple copies of heterologous polynucleotides may be on a single construct or on more than one construct.

Contemplated are thermophilic host cells where the components of the expression control system (or genes encoding them) are transiently introduced into the thermophilic host cell through use of a plasmid or shuttle vector. Preferably, however, the components of the expression control system (or genes encoding them) are permanently introduced into a

chromosome of the thermophilic host cell. Chromosomal integration techniques are known to the skilled artisan and have been described in, for example, Zhou and Wolk, J Bacterid., 184(9):2529-2532, 2002. Briefly, polynucleotide sequences comprising (or encoding) the components of the expression control system and gene of interest may be subcloned into one or more integration vectors. This construct may then be introduced into the thermophilic host cell for double homologous recombination at specific loci on the host cell chromosome, e.g., the lactate dehydrogenase (*ldh*) locus.

In the thermophilic host cell as described herein, the first promoter is operatively linked to a gene of interest. A wide variety of genes/gene products may be relevant to express/produce from the gene of interest according to the uses and methods described herein. The gene product could for example be a protein which is already produced industrially in a thermophilic microorganism but using a different expression control system than the Tet_R-P_{tetA} system or using a Tet_R-P_{tetA} system but with tetracycline induction. The gene product could also be a protein endogenously produced in a mesophilic bacterium, but which could more advantageously be produced in a thermophilic organism using the expression control system according to the present invention. The gene product could also be an RNA sequence transcribed from the gene of interest. Table 3 gives some examples of suitable genes and gene products for expression/production in thermophilic host cells. Typically, the gene product of interest is not a TetR protein.

Table 3: Examples of suitable genes/gene products of interest

Gene name	Protein name	Comments/references
<i>alsS</i> and <i>alsD</i>	Acetolactate synthase and acetolactate decarboxylase, respectively.	Can be from various sources, e.g. <i>B. subtilis</i> , for the production of 2,3-butanediol or acetoin. (Zhou et al., Appl. Microbiol. Biotechnol. 2020)
<i>ribC</i>	Riboflavin synthase	This gene can be used to get the thermophilic cell to produce riboflavin. (Yang et al., Microb. Biotechnol. 2020)
<i>kivD</i> , <i>ilvC</i> , <i>alsS</i> , <i>ilvD</i>	2-ketoisovalerate decarboxylase, ketol-acid reductoisomerase, acetolactate synthase, dihydroxy-acid dehydratase	These genes form a pathway that can be used for the production of isobutanol in thermophiles. (Lin et al., Metabol. Engineer. 2014)
<i>roseRS_3509</i> , <i>hmgr</i> , <i>aca</i> , <i>hmgs</i>	Terpene synthase, hydroxymethylglutaryl-CoA reductase, acetyl-CoA acetyltransferase,	These genes form a pathway for the production of terpenes. (Styles et al., Metabol. Engineer. 2021)

<i>mvk, dmd, pmk and idi, fpps</i>	hydroxymethylglutaryl-CoA synthase, mevalonate kinase, diphosphomevalonate decarboxylase, isopentenyl pyrophosphate isomerase, FPP syn- thase	
<i>xynA1</i>	Xylanase	A xylanase from a <i>Parageobacillus</i> strain was produced in a different <i>Parageobacillus</i> strain, for production of enzymes required for biomass degradation. (Holland et al., BMC Biotechnol. 2019)

Provided are also methods and uses of the expression control system and thermophilic host cells as described herein. In some expression systems, when the thermophilic host cell is incubated in a culture medium having a first temperature, the expression of the gene of interest is not induced; when the thermophilic host cell is incubated in a culture medium having a second temperature, the expression of the gene of interest is induced; and the second temperature is at least about 4°C higher than the first temperature. In some expression systems, methods and uses described herein, the temperature is increased from a temperature of up to 60°C to a temperature of at least 65°C. For example, the first temperature may be ≤59°C, such as 55°C, such as 50°C, and such as 45°C, and the second temperature may be ≥66°C, such as 70°C, such as 75°C, and such as 80°C. In preferred embodiments, the first temperature is around 60°C and the second temperature is around 65°C.

In some expression systems described herein, an increase in temperature results in induction of expression of the gene of interest. Accordingly, these expression systems are useful for uses and methods where, at the first (lower) temperature, the thermophilic host cell culture is multiplied, i.e., grown, and, once a desired number or concentration of thermophilic host cells has been achieved, the temperature is increased and the transcription and expression of the gene or protein of interest induced. The resulting biological molecule can then be isolated, e.g., retrieved from the cell culture and purified as needed.

Table 4: Sequences of specific expression control system components

Sequence name	Sequence	SEQ ID NO
<i>E. coli</i> TetR amino acid sequence	MSRLDKSKVINSALELLNEVGIEGLTTRKLAQ KLGVEQPTLYWHVKNKRALLDALAIEMLDHRH HTHFCPLEGESWQDFLRNNAKSFRCALLSHR DGAHVHLGTRPTEKQYETLENQLAFLCQQGF SLENALYALSAVGHFTLGCVLEDQEHQVAKE ERETPTTDSMPPLLRQAIELFDHQGAEPALF GLELIICGLEKQLKCESGS	1
<i>E. coli</i> TetR nucleotide sequence	ATGATGTCTAGATTAGATAAAAAGTAAAGTGA TTAACAGCGCATTAGAGCTGCTTAATGAGG TCGGAATCGAAGGTTTAAACAACCCGTAAAC TCGCCCAGAAGCTAGGTGTAGAGCAGCCTA CATTGTATTGGCATGTAAAAAATAAGCGGG CTTTGCTCGACGCCTTAGCCATTGAGATGTT AGATAGGCACCATACTCACTTTTGCCCTTTA GAAGGGGAAAGCTGGCAAGATTTTTACGT AATAACGCTAAAAGTTTTAGATGTGCTTTAC TAAGTCATCGCGATGGAGCAAAAGTACATT TAGGTACACGGCCTACAGAAAAACAGTATG AAACTCTCGAAAATCAATTAGCCTTTTTATG CCAACAAGGTTTTTCACTAGAGAATGCATTA TATGCACTCAGCGCTGTGGGGCATTTTACT TTAGGTTGCGTATTGGAAGATCAAGAGCAT CAAGTCGCTAAAGAAGAAAGGGAAACACCT ACTACTGATAGTATGCCGCCATTATTACGAC AAGCTATCGAATTATTTGATCACCAAGGTGC AGAGCCAGCCTTCTTATTCGGCCTTGAATT GATCATATGCGGATTAGAAAAACAACCTTAA TGTGAAAGTGGGTCTTAA	2
Modified <i>E. coli</i> TetR nucleotide sequence (used in Ex. 1)	ATGATGTCTAGATTAGATAAAAAGTAAAGTGA TTAACAGCGCATTAGAGCTGCTTAATGAGG TCGGAATCGAAGGTTTAAACAACCCGTAAAC TCGCCCAGAAGCTAGGTGTAGAGCAGCCTA CATTGTATTGGCATGTAAAAAATAAGCGGG CTTTGCTCGACGCCTTAGCCATTGAGATGTT AGATAGGCACCATACTCACTTTTGCCCTTTA GAAGGGGAAAGCTGGCAAGATTTTTGAGA AATAACGCTAAAAGTTTTAGATGTGCTTTAC TAAGTCATCGCGATGGGGCAAAAGTACATT TAGGTACACGGCCTACAGAAAAACAGTATG AAACTCTCGAAAATCAATTAGCCTTTTTATG CCAACAAGGTTTTTCACTAGAGAATGCATTA TATGCACTCAGCGCTGTGGTCATTTTACTT TAGGTTGCGTATTGGAAGATCAAGAGCATC AAGTCGCTAAAGAAGAAAGGGAAACACCTA CTACTGATAGTATGCCGCCATTATTACGACA AGCTATCGAATTATTTGATCATCAAGGGGC AGAGCCAGCCTTCTTATTCGGCCTTGAATT GATCATATGCGGATTAGAAAAACAACCTTAA TGTGAAAGTGGGTCTTAA	3

<i>E. coli</i> tetO1 sequence	ACTCTATCAATGATAGAGT	4
tetO1-pRMC2 (used in Ex. 1)	ACTCTATCATTGATAGAGT	5
tetO2	TCCCTATCAGTGATAGAGA	6
tetO2-M3	TCCCTAACAGTGTTAGAGA	7
tetO2-M4	TCCCTTTCAGTGAAAGAGA	8
tetO2-M5	TCCCAATCAGTGATTGAGA	9
tetO2-M6	TCCGTATCAGTGATACAGA	10
tetO2-pWH1012-5T	TCCCCATCAGTGATGGAGA	11
P _{tetA} sequence (modified, originally from <i>E. coli</i> ¹)	TTGACACTCTATCATTGATAGAGTATAATTA AAATAAGCTTGATCGTAGCGTTAACAGATCT GAGCTCCTGCAGTAAGCTG	12
P _{tetR} sequence (based on a modified <i>B. subtilis</i> PXylR promoter ^{1,2} – used in Ex. 1)	TGACAAATAACTCTATCAATGATATAATGTC AACAAAAAGGAGGAATTA	13
P _{tetA} sequence (modified, originally from <i>E. coli</i> ¹ – used in Ex. 1)	TTGACACTCTATCATTGATAGAGTATAATTA AAATAAGCTTGATCGTAGCGATTACAGATCT GTCCTCCTGCAGTAAGCTG	14
mRuby2 amino acid sequence (used in Ex. 1)	MVSKGEELIKENMRMKVVMESVNGHQKFC TGE GEGNPYMGQTMRKIVIEGGPLPFAFDIL ATSFMYGSRTFIKYPKGIPDFFKQSFPEGFTW ERVTRYEDGGVVTVMQDTSLEDGCLVYHVQ VRGVNFPNSNGPVMQKKTGWEPNTEMMYPAD GGGLRGYTHMALKVDGGGHLSCSFVTT	15
BgaB amino acid sequence (used in Ex. 1)	MNVLSSICYGGDYNPEQWPEEIWYEDAKLM QKAGVNLVSLGIFSWSKIEPSDGVDFEWD KVIDILYDHGVYINLGTATATTPAWFVKKYPD SLPIDESGVIFSFGSRQHYCPNHPQLITHIKR LVRAIAERYKNHPALKMWHVNNEYACHVSK CFCENCAVAFRKWLKERYKTIDELNERWGTN FWGQRYNHWDEINPPRKAPTFINPSQELDYY RFMND SILKLFLTEKEILRGVTPDIPVSTNFM GSFKPLNYFQWAQHIDIVTWDSYPDPREGLP IQHAMMNDLMRSLRKGQPFILMEQVTSHVN WRDINVPKPPGVMRLWSYATIARGADGIMFF QWRQSRAGAEKFHGAMVPHFLNENNRIYRE VTQLGQELKKLDCLVGSRIKAEVAIIFDWEN WWAVELSSKPHNKLRYIPIVEAYYRELYKRNI AVDFVRPSDDLTKYKVVIAPMLYMKEGEDE NLRQFVANGGTLIVSFFSGIVDENDRVHLGG YGPPLRDILGIFVEEFVPYPETKVNKIYSNDGE YDCTTWADIIRLEGAELATFKGDWYAGLPA VTRNCYKGEGEYVGTYPDSNYLGRLLQVF AKHHINPILEVAENVEVQQRETDEWKYLIIN HNDYEVTLSPEDKIYQNMIDGKCFRGGELRI QGV DVAVLREHDEAGKV	16

Additional TetR proteins		SEQ ID NOs: 32-56 and 79-82
--------------------------	--	-----------------------------------

¹ Fagan et al., 2011; ² Corrigan et al., 2009

EXAMPLE 1

Construction of inducible expression system

This example demonstrates the construction of an inducible promoter system, where a shift in temperatures within the thermophilic range is used to activate expression from the promoter system. The functionality of the system is shown with a thermostable β -galactosidase.

For stable and homogeneous expression of constructs, integrative expression is preferred over plasmid-based expression. Previous work showed the suitability of the lactate dehydrogenase (*ldh*) locus (GenBank: ALF09597.1) for the integration of target constructs (Sheng et al., Biotechnol. Biofuels;10:1-18 2017). First, a plasmid backbone that allows the integration of target constructs into the *ldh* locus was created. Homologous regions of 1000 base pairs, flanking the genomic region that was to be modified, were amplified using polymerase chain reaction (PCR). The primers used are shown in Table 5 – primer sequences are shown in Table 7. The flanks were inserted into a pMTL61110 (Sheng et al., Biotechnol. Biofuels;10:1-18 2017) backbone through standard USER cloning (Bitinaite et al., Nucleic Acids Res.;35:1992-2002 2007) approaches, to create pMTL61110- Δ *ldh*. To insert the final temperature sensitive inducible system into the integrative backbone vector, an initial iteration using the endogenous nucleotide sequence of the gene encoding the *Escherichia coli* TetR protein (SEQ ID NO:2 (gene) and SEQ ID NO:3 (protein)) and the P_{tetR}-P_{tetA} promoter sequence (SEQ ID NO:13 (PtetR) and SEQ ID NO:12 (PtetA)) from the plasmid pRPF185 (Fagan and Fairweather, 2011) was performed. As described above, the P_{tetR} promoter sequence is a modified PXylR promoter originally derived from *B. subtilis*. It was originally copied from the pRMC2 plasmid (Corrigan et al., 2009), and the -10 box was replaced with the *B. subtilis* consensus sequence (Fagan and Fairweather, 2011). The P_{tetA} promoter sequence is originally derived from the *E. coli* Tn10 P_{tetA} but has been adapted for use in Gram positive bacteria (Fagan and Fairweather, 2011). The Tet_R-P_{tetR}-P_{tetA} region was amplified from pRPF185 by PCR, with the primer pair as indicated in Table 5. To evaluate the performance of the expression system, two different reporter proteins were used. To evaluate functioning at 52°C, mRuby2 (SEQ ID NO:15) was used as a fluorescent marker. The marker was amplified from pcDNA3-mRuby2 using primers indicated in Table 5.

As mRuby2 does not function above 53°C, to evaluate the systems functioning above 53°C, a thermostable β -galactosidase from *Geobacillus stearothermophilus* (BgaB, SEQ ID NO:16) was selected as reporter protein and PCR amplified from pMTLP13B using primers indicated in Table 5 (Ølshøj Jensen et al., AMB Express;7 2017). The PCR products were cloned into pMTL61110- Δ/dh using USER cloning, resulting in pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}-mRuby2 and pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}-BgaB. The resulting construct was verified by Sanger sequencing.

To test the construct in the model thermophile *Parageobacillus thermoglucosidasius* DSM2542², transformation of the pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}-mRuby2 and pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}-BgaB plasmids was performed through a high osmolarity electroporation method according to protocols as described earlier (Taylor et al., Plasmid;60:45-52 2008). Transformants were selected on BBL™ Trypticase™ Soy Agar (TSA) (BD Bioscience) plates supplemented with 12.5 µg/mL kanamycin. However, upon multiple successive attempts, few to no colonies were obtained. Restriction-modification (R-M) systems have been known to be wide-spread across the prokaryotic realm (Loenen et al., Nucleic Acids Res.;42:3-19 2014), yet their presence is poorly predicted in novel genomes and strains. Thus, unique restriction sites present within the Tet_R-P_{tetR}-P_{tetA} nucleotide sequence were identified (Table 6). These sites were found in both the Tet_R (5 times) and P_{tetR}-P_{tetA} (2 times) regions of the nucleotide sequence. The recognition sites present in the Tet_R coding sequence were changed through the insertion of silent mutations, whereas the recognition sites in the P_{tetR}-P_{tetA} region were changed by making mutations through sticking to the same nucleotide class (purines and pyrimidines). To insert these designed changes into the promoter system, a Gblock was ordered from IDT Technologies and PCR amplified (Table 5), and inserted into pMTL61110 Δ/dh , resulting in pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}(mod)-mRuby2 and pMTL61110- Δ/dh -Tet_R-P_{tetR}-P_{tetA}(mod)-BgaB. In these final constructs, the Tet_R nucleotide sequence was SEQ ID NO:3, the P_{tetR} sequence was SEQ ID NO:13, the P_{tetA} sequence was SEQ ID NO:14, the mRuby2 (amino acid) sequence was SEQ ID NO:15, and the BgaB (amino acid) sequence was SEQ ID NO:16 (Table 4).

Table 5: Overview of primer pairs and templates used in Example 1

Target name	Primer pair	Template
Tet _R -P _{tetR} -P _{tetA}	1395_TetR-U-fw 1980_tetR(BS)-U-rv	pRPF185 (Fagan and Fairweather, 2011)
BgaB	P124_bgaB_U_Fw P125_bgaB_U_Rev	pMTLP13B (Ølshøj Jensen et al., AMB Express;7 2017)

Tet _R -P _{tetR} -P _{tetA} (mod)	P353_TetRnew_UFW P354_TetRnew_URRev	Ordered Gblock (TetR: SEQ ID NO:3; P _{tetR} : SEQ ID NO:13; P _{tetA} : SEQ ID NO:14)
LDH upstream homology region	PNJ851 PNJ852	<i>P. thermoglucosidasius</i> DSM2542 ^Z gDNA
LDH downstream homology region	PNJ853 PNJ854	<i>P. thermoglucosidasius</i> DSM2542 ^Z gDNA
mRuby2	P64 P340	pcDNA3-mRuby2 (Lam et al., Nat. Methods 2012)

Upon removing the seven restriction enzyme recognition sites, transformation of *P. thermoglucosidasius* DSM2542^Z was possible and resulted in an increased number of colonies, indicating the importance of removing the recognition sites from the nucleotide sequence. To

5 integrate the Tet_R-P_{tetR}-P_{tetA}-mRuby2 or Tet_R-P_{tetR}-P_{tetA}-BgaB sequence of the plasmid into the *ldh* locus of the *P. thermoglucosidasius* DSM2542^Z genome, homologous recombination at the desired locus was forced by growing transformants in liquid cultures (2SPY with 12.5 µg/mL kanamycin, 65°C, 200 rotations per minute) and plating on selective solid medium at 65°C. The vector is unable to replicate at this temperature, and so survival only ensues from

10 integration of the complete construct. Colonies were screened for correct integration with colony PCR. To remove the selection marker from the genome, correct integrants were cultured in 2SPY medium at 65°C for 2-5 days through serial passaging. Double cross-overs were screened through plating on TSA and replica plating on TSA with 12.5 µg/mL kanamycin. Sensitive clones were screened through colony PCR for correct, stable double cross-over and

15 removal of the resistance cassette. The resulting strains created were *P. thermoglucosidasius* DSM2542^Z $\Delta ldh::Tet_R-P_{tetR}-P_{tetA}-mRuby2$ and *P. thermoglucosidasius* DSM2542^Z $\Delta ldh::Tet_R-P_{tetR}-P_{tetA}-BgaB$.

Table 6 gives an overview of restriction enzyme recognition sites identified in the Tet_R-P_{tetR}-P_{tetA} nucleotide sequence, which were not found in other DNA regions previously known to show

20 efficient transformation into *P. thermoglucosidasius* strains. The restriction enzymes and known isochimers are shown, as well as the organisms these enzymes have previously been identified in.

Table 6: Restriction enzyme recognition sites in the Tet_R-P_{tetR}-P_{tetA} nucleotide sequence

Restriction enzyme (and isochimers)	Recognition sequence	Previously identified in	SEQ ID NO
BsaXI	ACNNNNCTCC	<i>Bacillus stearothermophilus</i> 25B	17
AleI, OIil	CACNNNNGTG	<i>Aureobacterium liquefaciens</i> , <i>Oceanospirillum linum</i> 4-5D1	18
RleAI	CCCACA	<i>Rhizobium leguminosarum</i> DSM 5629	-
BstENI, EcoNI, XagI	CCTNNNNNAGG	<i>Geobacillus stearothermophilus</i> EN, <i>Escherichia coli</i> CDC A-193 (ATCC 12041), <i>Xanthobacter agilis</i> Vs18-132	19
SacI, Eco53kI, BpuAmI, Ecl136II, EcoICRI, MxaI, Psp124BI, SstI	GAGCTC	<i>Streptomyces achromogenes</i> (ATCC 12767), <i>Escherichia coli</i> 53k, <i>Bacillus pumilus</i> , <i>Enterobacter cloacae</i> RFL136, <i>Myxococcus xanthus</i> F18E, <i>Pseudomonas</i> sp. 124B, <i>Streptomyces Stanford</i>	-
BstEZ359I, BstHPI, HpaI, KspAI, SsrI	GTTAAC	<i>Bacillus stearothermophilus</i> EZ359, <i>Bacillus stearothermophilus</i> HP, <i>Haemophilus parainfluenzae</i> , <i>Kurthia</i> sp. N88, <i>Staphylococcus saprophyticus</i> B6	-
BstSNI, Eco105I, SnaBI	TACGTA	<i>Bacillus stearothermophilus</i> SN, <i>Escherichia coli</i> RFL105, <i>Sphaerotilus natans</i>	-

Initially, the classical, inducer-dependent functioning of the promoter system at 52°C and 60°C was investigated. For the evaluation at 52°C, the *P. thermoglucosidasius* DSM2542² $\Delta dh::Tet_R$ -P_{tetR}-P_{tetA}-mRuby2 strain was used. Single colonies were used to inoculate 5 ml of 2SPY medium, and the cultures were grown for 8h at 52°C or 60°C, 200 rotations per minute (RPM). 10 µL of the culture was used to inoculate 50 mL thermophile minimal medium (TMM, as used previously; Pogrebnyakov et al., PLoS One;12 2017) supplemented with 10 g/L D-Glucose and 0.2% yeast extract (YE). These cultures were left to grow at 52°C or 60°C, 200 RPM overnight. The following morning, preheated baffled shake flasks with 50 ml TMM, 10 g/l D-Glucose and 0.2% YE were inoculated to a starting optical density (OD) of 0.05 with the overnight culture. The cultures were left to grow at 52°C or 60°C, 200 RPM, whilst monitoring the OD of the culture. After one duplication of the culture, flasks were induced by addition of

anhydrotetracycline (stock solution at 2 g/L) to a final concentration of 0.5 µg/mL. No anhydrotetracycline was added to uninduced control cultures. For the following 10 hours post induction, the OD was monitored, and 1 mL samples were taken. Cells from the sample were fixed by incubation for 30-60 min in 2% paraformaldehyde in phosphate-buffered saline (PBS).
 5 After incubation, cells were washed with and stored in PBS at 4°C for up to 2 weeks. The fluorescence of the fixed cells was determined with a MACSQuant VYB cytometer (Miltenyi Biotec, Germany). mRuby2 fluorescence was measured by excitation with a 561 nm laser, using a 615/20 nm band-pass filter. Per 200 µL sample, up to 50,000 events were recorded. Gain settings used were; forward scatter: 580 V, side scatter: 516 V, Y2 channel: 492 V. Data
 10 was analyzed using FlowJo (Tree Star Inc., Asland, OR).

As shown in Figure 1A, the mRuby2 expression from Tet_R-P_{tetR}-P_{tetA} when uninduced at 52°C is indistinguishable from the wild type (WT) control, whereas induction drastically increases mRuby2 expression and does so in a homogenous manner as seen through the flow histograms (Figure 1B). The signal is also stable for a period of at least 8 hours post induction (Figure 1A).
 15 This shows that, unexpectedly, the expression system has a similar function to the way it behaves in mesophilic hosts, even at the 15°C higher temperature.

Next, we characterized the function of the Tet_R-P_{tetR}-P_{tetA} expression system at 60°C, through classical induction, as described above, with the BgaB activity as output. BgaB activity was assayed based on previously described protocols (Zhang et al., J. Biol. Chem.;270:11181-
 20 11189 1995), but the concentration of dibasic sodium phosphate in the permeabilization solution was lowered to 100 mM. Activity was measured at 60°C, and the reaction stopped after 20 minutes. The absorbance (A_{420nm} and A_{600nm}) of 200 µL of the reaction supernatant was measured in a 96-well microtiter plate. The Miller units gives a readout for BgaB activity and was calculated according to Formula I:

$$Miller\ units = 1000 * \frac{A_{420nm}}{(A_{600nm} * V * t)} \quad (I)$$

where the volume of the sample used (V, 20 µL) and exact assay reaction time (t) are also used. Directly upon taking the sample, 20 µL was added to the permeabilization solution and stored at 4°C overnight, until the assay was run with all samples simultaneously. Previous experiments indicated this did not influence the BgaB activity when compared to freshly taken
 30 samples.

It is seen that at 60°C, the Tet_R-P_{tetR}-P_{tetA} expression system still functions as it does at lower temperatures, being highly inducible yet extremely silent when uninduced (Figure 2A). However, the time stability of the system appears different, though this may be an artifact

resulting from the sporulation observed to occur after 4 hours post induction, due to the test being performed in batch culture (Figure 2B).

Finally, to investigate the temperature inducibility of the expression system, single colonies were used to inoculate 5 mL of 2SPY medium and grown for 8h at 60°C, 200 rotations per minute (RPM). 10 µL of the culture was used to inoculate 50 mL thermophile minimal medium (TMM, as used previously; Pogrebnyakov et al., PLoS One;12 2017) supplemented with 10 g/L D-Glucose and 0.2% yeast extract. These cultures were left to grow at 60°C, 200 RPM overnight. The following morning, preheated shake flasks with 50 mL TMM, 10 g/L D-Glucose and 0.2% YE and a magnetic stirrer rod were inoculated to a starting OD of 0.05 with the overnight culture. The cultures were left to grow at 60°C, 200 RPM, whilst monitoring the OD of the culture. After one duplication of the culture, flasks were induced by transfer to a preheated, temperature-controlled water bath set to 65°C placed on a magnetic stirrer plate. Uninduced control flasks were left at 60°C. For the following 8 hours post induction, the OD was monitored, and samples of the culture were taken every two hours to quantify BgaB expression.

As shown in Figure 3, at 60°C the BgaB expression from Tet_R-P_{tetR}-P_{tetA} is very silent and indistinguishable from the wild type background signal. This indicates the Tet_R-P_{tetR}-P_{tetA} functions similarly at 60°C as it does at lower, mesophilic temperatures. Surprisingly, upon shifting culture temperature to 65°C, induction of the BgaB construct from the Tet_R-P_{tetR}-P_{tetA} promoter was observed within two hours, independent of the presence or absence of inducer. Maximal induction was reached after 6 hours post induction and was maintained until the maximum sample point of 8 hours post induction. These data demonstrate that the Tet_R-P_{tetR}-P_{tetA} system can be used by simply shifting culture temperature and thereby induce expression of target proteins (see schematic representation in Figure 4).

Table 7: Primer sequences

Name	Sequence	SEQ ID NO
P353_TetRnew_UFW	ACAATUTCACACAGGAGGCCGA	20
P354_TetRnew_URv	ATCGCCUCAGCTTACTGCAGGAGGAC	21
1395_TetR-U-fw	atcgcgU ttaagaccactttcacatttaagttgt	22
1980_tetR(BS)-U-rv	atcgccUcagc ttactgcaggagctcagatctgt	23
P124_bgaB_U_Fw	GGCGAUaggaggaatataccATGAACGTTTTATCCTCAATTTG	24

P125_bgaB_U_Rev	AGAGTUAGAAAAAGAAACAGAGGCTACTCTCAA	25
PNJ851	ACGAATUCCCTCGGCAAACAGAGCTTTAAACC	26
PNJ852	AACAAGGUGAACATCGTGTGGATAACAAC	27
PNJ853	AACTCUGCCATTATCATTTCGTAATGCC	28
PNJ854	ATCCCCGGGUACGGACTTTTATTTCAATTCGTATGGTGCG	29
P64	GGCGAUaggaggaatatacatggtgtctaagggcgaagag	30
P340	GGTGCGAUTTACTTGTACAGCTCGTC	31

EXAMPLE 2

Simulation of temperature-dependent stability of TetR family members

This example demonstrates an *in silico* simulation of the temperature-dependent (in)stability of members of the TetR family. The results of this simulation indicate that TetR proteins in general, irrespective of whether they are derived from mesophilic, thermophilic or psychrophilic hosts, are amenable to be used for control of gene expression by temperature regulation in the thermophilic temperature range.

As shown in Example 1, upon temperature increase, the TetR protein loses its capacity to bind to DNA and repress expression from the P_{tet} promoter. To identify possible structural and sequence components of TetR that are causative for the structural change upon temperature increase, we deployed molecular dynamic (MD) simulations. MD simulations predict the physical movements of atoms and molecules and is a benchmark for analyzing protein stability under changing conditions (Karplus and McCammon, Nat. Struct. Biol.;9:646-652 2002). From this, flexible regions in the protein can be identified, which are often unstable at the tested conditions.

The flexibility of TetR was modeled at low and high temperatures by rapid, large scale molecular dynamics (MD) simulation and analysis using the online version of the coarse-grained modeling tool CABS-flex 2.0 (Kuriata et al., Nucleic Acids Res.;46:W338-W343 2018).

The atomic coordinates for TetR was obtained from the PDB ID 2XPW. The PDB file was manually inspected using PyMOL version 2.5.4 (Schrödinger and LLC, 2015) to select the chains corresponding to the dimer whose solved structure covered most of the protein sequence. Small missing N- or C-terminal segments were not modelled, as they were expected to have

negligible influence on the simulation results. PyMOL was also used to fix missing chain IDs in the PDB file when needed.

To simulate an entire dimer, two chains corresponding to a monomer each were selected for the PDB file. The standard distance restraint values suggested by CABS-flex 2.0 were used, i.e., the mode was SS2, with a restrain gap along the chain of 3 residues, and the minimal and maximal length of restraint were 3.8 and 8.0 Å, respectively. No additional distance restraints were added. Independent simulations were performed at low (1.4) and high (2.0) temperatures.

The main results generated by CABS-flex 2.0 (publicly available at the internet webpage <http://biocomp.chem.uw.edu.pl/CABSflex2/index>, accessed on 6 December 2022), are conformational models, contact maps, and fluctuation plots, and our analysis focused on the latter. Since the monomers forming a dimer in a simulation are identical, they were treated as replicates of each other and hence their root mean square fluctuations (RMSFs) were averaged for each amino acid residue.

For TetR, the MD simulations reveal regions showing a higher level of increased flexibility upon increased temperature (Figure 5). Of these three regions, two are found at the (generally flexible) termini of the protein. The final region identified falls in the conserved helix-turn-helix (HTH) region of the protein, even more specifically in the known DNA-binding domain. Therefore, at higher temperature, the MD simulations predict a loss of structure in the DNA-binding domain, which can result in a loss of affinity for the tetO sequence, and so trigger expression from P_{tet} .

Conservation of functionality across the TetR family:

The TetR family is classified based on the presence of the highly conserved HTH region, often found near the N-terminus of the protein. The HTH region forms the basis for the DNA-binding domain of the protein, and the residual 70-75% of the protein can dictate alternative functional parameters, such as inducer specificity. The highly conserved nature of the HTH-region results in proteins of the TetR family being found in nearly all sequenced bacteria to date, even though they have very different response elements and inducer specificities (Stanton et al., Nat. Chem. Biol.;10:99–105 2014).

To investigate the level of conservation across the TetR family, in relation to potential for temperature-based induction, a Multiple Alignment using Fast Fourier Transform (MAFFT) was run on a subset of TetR family proteins (Madeira et al., Nucleic Acids Res.;47:W636–W641 2019). The variants with identified inducers and operator sites from Stanton et al. were

selected, as well as three variants from thermophilic and psychophilic (optimum growth <20 °C) hosts (Table 8; Agari et al., Microbiology;157:1589–1601 2011). MAFFT analysis on the whole protein structure as well as on the conserved HTH-domain was performed (Figures 6 and 7, respectively). The sequences used for performing the MAFFT can be found in Tables 9 and 10, respectively. This analysis shows how diverse the TetR family proteins are, both when looking at whole protein sequence and the HTH sequence. Even though the HTH-domain has higher similarity compared to the whole protein analysis, there is still variation in amino acid sequence. Though this is observed on amino acid level, the 3D structure and hence function of the HTH motif is still highly conserved within the family.

Using the crystal structure of the identified variants, the melting temperature can be estimated using predictive algorithms (Table 8; Pucci et al., Bioinformatics;33:3415–3422 2017). For TetR, the crystal structure reveals a predicted melting temperature of 60.5 °C, close to what was observed for the *in vitro* data presented in Example 1. Interestingly, for all the variants tested, regardless of their organism source (i.e. psychophilic, mesophilic or thermophilic), their melting temperature falls in the thermophilic operating range.

As the conserved DNA-binding domain was shown to be potentially causative for the TetR temperature induction (Figure 5), its role in regulating potential temperature inducible systems in the other TetR family proteins was investigated in the same manner. The PDB IDs for the analyzed proteins are shown in Table 8. In the case of IcaR, where all PDBs presented unresolved loops in the internal part of the sequence, the online tool ModLoop (Fiser et al., Protein Sci.;9:1753–1773 2000; Fiser and Sali, Bioinformatics;19:2500–2501 2003).

By performing the same MD simulations, the difference in RMSF values (ΔRMSF) for each residue between low and high temperature can be calculated. From there, destabilized regions are defined as regions where the ΔRMSF for a residue is higher than 1.5 times the overall average RMSF value for the whole protein ($\Delta\text{RMSF} > 1.5\bar{x}_{\Delta\text{RMSF}}$) for at least four consecutive residues. This analysis was performed for a representative subset of the TetR family proteins (Figure 8). Overlaying the destabilized regions with the HTH region and DNA-binding domains shows that every TetR-family protein analyzed has a predicted destabilization region within the HTH- and DNA binding regions. Some contain additional destabilization regions in other parts of the protein. This points to the possibility that temperature-based induction of the TetR proteins in the thermophilic range may be feasible across the TetR family.

Table 8: TetR family proteins used to analyze sequence similarity and for further MD simulations. The growth preference of the host from which the protein was found is indicated. Percentage identity to the whole TetR protein is indicated, when analyzed by MAFFT. The PDB

ID for the crystal structure of the protein or Alpha Fold ID for the predicted structure is provided. Melting temperature predicted based on protein structure as calculated by SCOP is shown (Pucci et al., Bioinformatics;33:3415–3422 2017). NA = Not available.

Protein	Origin	% AA identity to whole tetR	PDB or AlphaFold ID	Predicted T _m (°C)
TetR	Mesophilic	100.00	1QPI	60.5
ImrA	Mesophilic	14.77	1SGM	59.8
AmtR	Mesophilic	15.70	5MQQ	62.2
SrpR	Mesophilic	16.67	AF-Q9R9T9-F1	57.4
TarA	Mesophilic	14.62	NA	
QacR	Mesophilic	15.12	1JT0	68
IcaR	Mesophilic	19.38	2ZCN	67.9
ScbR	Mesophilic	14.88	NA	
ButR	Mesophilic	15.05	NA	
PhIF	Mesophilic	17.86	AF-Q9RF02-F1	60.8
SmcR	Mesophilic	15.08	3KZ9	58.7
McbR	Mesophilic	14.61	4P9F	59.5
PsrA	Mesophilic	13.19	2FBQ	61.4
BetI	Mesophilic	19.88	AF-P17446-F1	62.6
LitR	Mesophilic	14.86	AF-Q8KX64-F1	54.6
Orf2	Mesophilic	11.24	NA	
HapR	Mesophilic	14.12	2PBX	58.1
HlyIR	Mesophilic	12.29	2FX0	65.9
AmeR	Mesophilic	16.28	AF-Q7CWT2-F1	56.3
BM3R1	Mesophilic	18.54	AF-P43506-F1	63.3
Q488L7	Psychrophilic	15.66	AF-Q488L7-F1	46.9
A0A127VTP2	Psychrophilic	12.50	AF-A0A127VTP2-F1	54.7
A0A5J6WF87	Psychrophilic	15.56	AF-A0A5J6WF87-F1	50.0
FadR	Thermophilic	14.04	3ANP	71.6

PaaR	Thermophilic	14.81	AF-Q5SJN5-F1	81.9
PfmR	Thermophilic	17.14	3VPR	87.5

Table 9: SEQ ID NOs of amino acid sequences used in the analyses shown in Figures 6 and 7

Protein	Origin	SEQ ID NO of amino acid sequence used in Fig. 6	SEQ ID NO of amino acid sequence used in Fig. 7
TetR	Mesophilic	1	57
ImrA	Mesophilic	32	58
AmtR	Mesophilic	33	59
SrpR	Mesophilic	34	60
TarA	Mesophilic	35	-
QacR	Mesophilic	36	61
IcaR	Mesophilic	37	62
ScbR	Mesophilic	38	-
ButR	Mesophilic	39	-
PhIF	Mesophilic	40	63
SmcR	Mesophilic	41	64
McbR	Mesophilic	42	65
PsrA	Mesophilic	43	66
BetI	Mesophilic	44	67
LitR	Mesophilic	45	68
Orf2	Mesophilic	46	-
HapR	Mesophilic	47	69
HlyIIR	Mesophilic	48	70
AmeR	Mesophilic	49	71
BM3R1	Mesophilic	50	72
PfmR	Thermophilic	51	73

PaaR	Thermophilic	52	74
FadR	Thermophilic	53	75
Q488L7	Psychrophilic	54	76
A0A127VTP2	Psychrophilic	55	77
A0A5J6WF87	Psychrophilic	56	78

LIST OF REFERENCES

Each reference listed below or described elsewhere herein is hereby incorporated by reference in its entirety.

- 5 Agari, Y., Agari, K., Sakamoto, K., Kuramitsu, S., and Shinkai, A. (2011). TetR-family transcriptional repressor *Thermus thermophilus* FadR controls fatty acid degradation. *Microbiology*. 157:1589–1601.
- 10 Bertram, R., and Hillen, W. (2008). The application of Tet repressor in prokaryotic gene regulation and expression. *Microb. Biotechnol.* 1(1):2–16. doi:10.1111/j.1751-7915.2007.00001.x
- Corrigan, R. M., and Foster, T. J. (2009). An improved tetracycline-inducible expression vector for *Staphylococcus aureus*. *Plasmid*. 61(2):126–129. doi: 10.1016/j.plasmid.2008.10.001.
- 15 Cuthbertson, L., and Nodwell, J. R. (2013). The TetR family of regulators. *Microbiology and Molecular Biology Reviews*. 77(3). doi:10.1128/MMBR.00018-13.
- 20 He, P., et al. (2017). High-level production of α -amylase by manipulating the expression of alanine racamase in *Bacillus licheniformis*. *Biotechnol. Lett.* 39:1389–1394. doi: 10.1007/s10529-017-2359-5
- 25 Holland, A. T. N., et al. (2019). Inhibition of extracellular proteases improves the production of a xylanase in *Parageobacillus thermoglucosidasius*. *BMC Biotechnol.* 19. doi: 10.1186/s12896-019-0511-0
- Ettner, N., et al. (1996). Fast large-scale purification of tetracycline repressor variants from overproducing *Escherichia coli* strains. *Journal of Chromatography A*, 742: 95–105.
- 30 Fagan and Fairweather (2011). *Clostridium difficile* has two parallel and essential Sec secretion systems. *J. Biol. Chem.* 286:27483–27493.
- Fiser, A., Do, R. K. G., and Sali, A. (2000). ModLoop: Automated modeling of loops in protein structures. *Protein Sci.* 9:1753–1773.
- 35 Fiser, A., and Sali, A. (2003). ModLoop: Automated modeling of loops in protein structures. *Bioinformatics*. 19:2500–2501.
- 40 Ito, A., et al. (2012). Heat-inducible transgene expression with transcriptional amplification mediated by a transactivator. *Int. J. Hypertherm.* 28:788–798. doi: 10.3109/02656736.2012.738847
- Kamionka, A., et al. (2005). Tetracycline-dependent conditional gene knockout in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 71(2):728–733. doi: 10.1128/AEM.71.2.728-733.2005.
- 45 Kamionka, A., et al. (2004). Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res.* 32:842–847.
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
- 50 Krueger, C., et al. (2018). Tetracycline derivatives: alternative effectors for Tet transregulators. *Biotechniques*. 37(4). doi:10.2144/04374BM04

Kuriata, A., et al. (2018). CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res.* 46:W338–W343.

5 Lam, A.J., et al. (2012). Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat. Methods.* 9:1005–1012. doi: 10.1038/nmeth.2171. 10.1038/nmeth.2171

Lin, P., et al. (2014). Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*. *Metabol. Engineer.* 24:1-8. doi: 10.1016/j.ymben.2014.03.006.

10 Madeira, F., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47:W636–W641.

15 Menart, V., et al. (2003). Constitutive versus thermoinducible expression of heterologous proteins in *Escherichia coli* based on strong PR,PL promoters from phage lambda. *Biotechnol. Bioeng.* 83:181–190. doi: 10.1002/bit.10660.

20 Müller, G., et al. (1995). Characterization of non-inducible Tet repressor mutants suggests conformational changes necessary for induction. *Nature Structural Biology* 2:693-703.

Orth, P., et al. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nature Structural Biology.* 7(3):215-219.

25 Pogrebnyakov, I., et al. (2017). Genetic toolbox for controlled expression of functional proteins in *Geobacillus* spp. *PLoS ONE.* 12(2):e0171313. doi: 10.1371/journal.pone.0171313

Pucci, F., Kwasigroch, J. M., and Rooman, M. (2017). SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics.* 33:3415–3422.

30 Ramos, J. L., et al. (2005). The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews.* 69(2):326–356.

35 Reichheld, S. E., et al. (2009). The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. *PNAS.* 106(52):22263-22268. doi: 10.1073/pnas.0911566106

Scholz, O., et al. (2004). Activity reversal of Tet repressor caused by single amino acid exchanges. *Mol. Microbiol.* 53:777–789.

40 Schrödinger and LLC. (2015). PyMOL molecular graphics system.

Sheng, L., et al. (2017). Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnol. Biofuels.* 10:1-18.

45 Sizemore, C., et al. (1990). Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants. *Nucleic Acids Res.* 18(10):2875-2880.

50 Stanton, B. C., et al. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* 10:99–105.

Styles, M. Q., et al. (2021). The heterologous production of terpenes by the thermophile *Parageobacillus thermoglucosidasius* in a consolidated bioprocess using waste bread. *Metabol. Engineer.* 65:146-155. doi: 10.1016/j.ymben.2020.11.005

55 Valdez-Cruz, N.A., et al. (2010). Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microb. Cell Fact.* 9:18. doi: 10.1186/1475-2859-9-18

Vasina, J. A., et al. (1998). Scale-up and optimization of the low-temperature inducible cspA promoter system. *Biotechnol. Prog.* 14:714–721. doi: 10.1021/bp980061p.

- 5 Wickstrum, J., et al. (2013). Conditional gene expression in *Chlamydia trachomatis* using the Tet system. *PLoS ONE*. 8(10):e76743. doi: 10.1371/journal.pone.0076743

Yang, Z., et al. (2020). Engineering thermophilic *Geobacillus thermoglucosidasius* for riboflavin production. *Microb. Biotechnol.* doi: 10.1111/1751-7915.13543

10

Zhang, X., et al. (1995). Control of the *Escherichia coli* rrnB P1 promoter strength by ppGpp. *J. Biol. Chem.* 270(19):11181–11189.

15

Zhou, J., Lian, et al. (2020). Metabolic engineering of *Parageobacillus thermoglucosidasius* for the efficient production of (2R, 3R)-butanediol. *Appl. Microbiol. Biotechnol.* 104:4303–4311. doi: 10.1007/s00253-020-10553-8

WO2007074991 A1 (CJ Corp.)

WO17147555 A1 (Lanzatech New Zealand Ltd.)

- 20 WO16060714 A1 (American Sterilizer Co.)

WO13144647 A1 (Univ. Nottingham)

WO08079469 A2 (American Sterilizer Company)

US2006121564 A1 (National Institute of Advanced Industrial Science and Technology, Tokyo)

US2009042301 A1 (Univ. Alberta)

- 25 US10745693 B2 (Wisconsin Alumni Res. Found.)

US4874703 A (Eli Lilly and Company)

EP0685560 A2 (Suntory Ltd.)

CN105838716 A (Univ. Southwest)

CN110066758 A (Wuhan Junan Biological Tech Co Ltd.)

30

CLAIMS

1. A thermophilic host cell comprising a heterologous expression control system comprising a Tet repressor (TetR) protein and a first promoter comprising a tet operator (tetO) sequence to which the TetR protein can bind, wherein the first promoter is operatively linked to at least one gene of interest.
2. Use of an expression control system comprising a TetR protein and a first promoter comprising a tetO sequence to which the TetR protein can bind to control the expression of at least one gene of interest in a thermophilic host cell by temperature regulation.
3. The thermophilic host cell according to claim 1 or the use according to claim 2, wherein the thermophilic host cell further comprises a gene encoding the TetR protein, optionally wherein the gene encoding the TetR protein is operatively linked to a second promoter.
4. The thermophilic host cell or use according to any one of the preceding claims, wherein,
 - (a) when the thermophilic host cell is incubated in a culture medium having a first temperature, the expression of the gene of interest is not induced;
 - (b) when the thermophilic host cell is incubated in a culture medium having a second temperature, the expression of the gene of interest is induced; and
 - (c) the second temperature is at least about 4°C higher than the first temperature.
5. A method of controlling the expression of at least one gene of interest in a thermophilic host cell comprising an expression control system comprising a TetR protein and a first promoter comprising a tetO sequence, wherein the first promoter is operatively linked to the at least one gene of interest, the method comprising:
 - (i) incubating the thermophilic host cell in a culture medium having a first temperature; and
 - (ii) incubating the thermophilic host cell in a culture medium having a second temperature which is at least about 4°C higher than the first temperature, wherein steps i) and ii) can be conducted in any order,wherein the expression of the at least one gene of interest is induced at the second temperature but not at the first temperature, thereby controlling the expression of the at least one gene of interest.
6. A method of producing a gene product of interest in a thermophilic host cell comprising an expression control system comprising a TetR protein and a first promoter comprising a tetO sequence to which the TetR protein can bind, wherein the first promoter is

operatively linked to a gene of interest encoding the gene product of interest, the method comprising:

(i) incubating the thermophilic host cell in a culture medium having a first temperature; and

(ii) incubating the thermophilic host cell in a culture medium having a second temperature which is at least about 4°C higher than the first temperature; and

(iii) optionally, isolating the gene product,

wherein the expression of the gene encoding the gene product of interest is induced at the second temperature but not at the first temperature, thereby producing the gene product of interest.

7. The method according to any one of claims 5 and 6, comprising a prior step of transforming or transfecting the thermophilic host cell with one or more nucleic acid constructs comprising the first promoter, the at least one gene of interest and a gene encoding the TetR protein, optionally wherein the gene encoding the TetR protein is operatively linked to a second promoter.

8. The thermophilic host cell, the use or the method according to any one of claims 4 to 7, wherein the medium has not been supplemented with tetracycline or an analog thereof which binds to the TetR protein.

9. The thermophilic host cell, the use or the method according to any one of claims 4 to 8, wherein the first temperature is at most about 60°C, the second temperature is at least about 65°C, or both.

10. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein,

(a) when the TetR protein is bound to the tetO sequence of the first promoter, the promoter is repressed; and

(b) when the TetR protein is not bound to the tetO sequence of the first promoter, the promoter is active.

11. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein the TetR protein

a) comprises an amino acid sequence at least 60% identical to a sequence selected from SEQ ID NOs:1, 32-56, and 79-82,

b) has a melting temperature between 40°C and 90°C, preferably between 55°C and 90°C, more preferably between 60°C and 90°C,

c) comprises a helix-turn-helix (HTH) region, which comprises a DNA-binding domain and a sub-region of at least 4 consecutive amino acid residues, which sub-region destabilizes upon temperature increase, or

(d) a combination of (a) and (b), (a) and (c) or all of (a) to (c).

5

12. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein

(a) the TetR protein comprises an amino acid sequence at least 60% identical to SEQ ID NO:1; and/or

10

(b) the tetO sequence comprises a nucleotide sequence at least 65% identical to SEQ ID NO:5.

13. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein the expression control system comprises a TetR protein comprising an amino acid sequence at least 90% identical to SEQ ID NO:1, a tetO nucleotide sequence at least 90% identical to SEQ ID NO:5, a first promoter comprising a nucleotide sequence at least 90% identical to SEQ ID NO:14, and, optionally, a second promoter comprising a nucleic acid sequence at least 90% identical to SEQ ID NO:13.

15

20

14. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein the thermophilic host cell has an optimum growth temperature in the range of about 55°C to about 65°C, such as at about 60°C.

25

15. The thermophilic host cell, the use or the method according to any one of the preceding claims, wherein the thermophilic host cell is a bacterium.

16. The thermophilic host cell, the use or the method according to claim 15, wherein the bacterium is selected from the group consisting of: *Parageobacillus thermoglucosidasius*, *Geobacillus toebii*, *Geobacillus stearothermophilus*, *Geobacillus thermodenitrificans*, *Geobacillus kaustophilus*, *Geobacillus thermoleovorans*, *Geobacillus thermocatenulatus*, *Thermoanaerobacterium xylanolyticum*, *Thermoanaerobacterium saccharolyticum*, *Thermoanaerobacterium thermosaccharolyticum*, *Thermoanaerobacter mathranii*, *Thermoanaerobacter pseudoethanolicus*, *Thermoanaerobacter Brockii*, *Thermoanaerobacter kivui*, *Thermoanaerobacter Brockii*, *Caldanaerobacter subterraneus*, *Clostridium thermocellum*, *Clostridium thermosuccinogenes*, *Thermoclostridium stercorarium*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus smithii*, *Bacillus methanolicus*, *Bacillus flavothermus*, *Anoxybacillus kamchatkensis*, *Anoxybacillus gonensis*, *Caldicellulosiruptor bescii*, *Caldicellulosiruptor saccharolyticus*, *Caldicellulosiruptor*

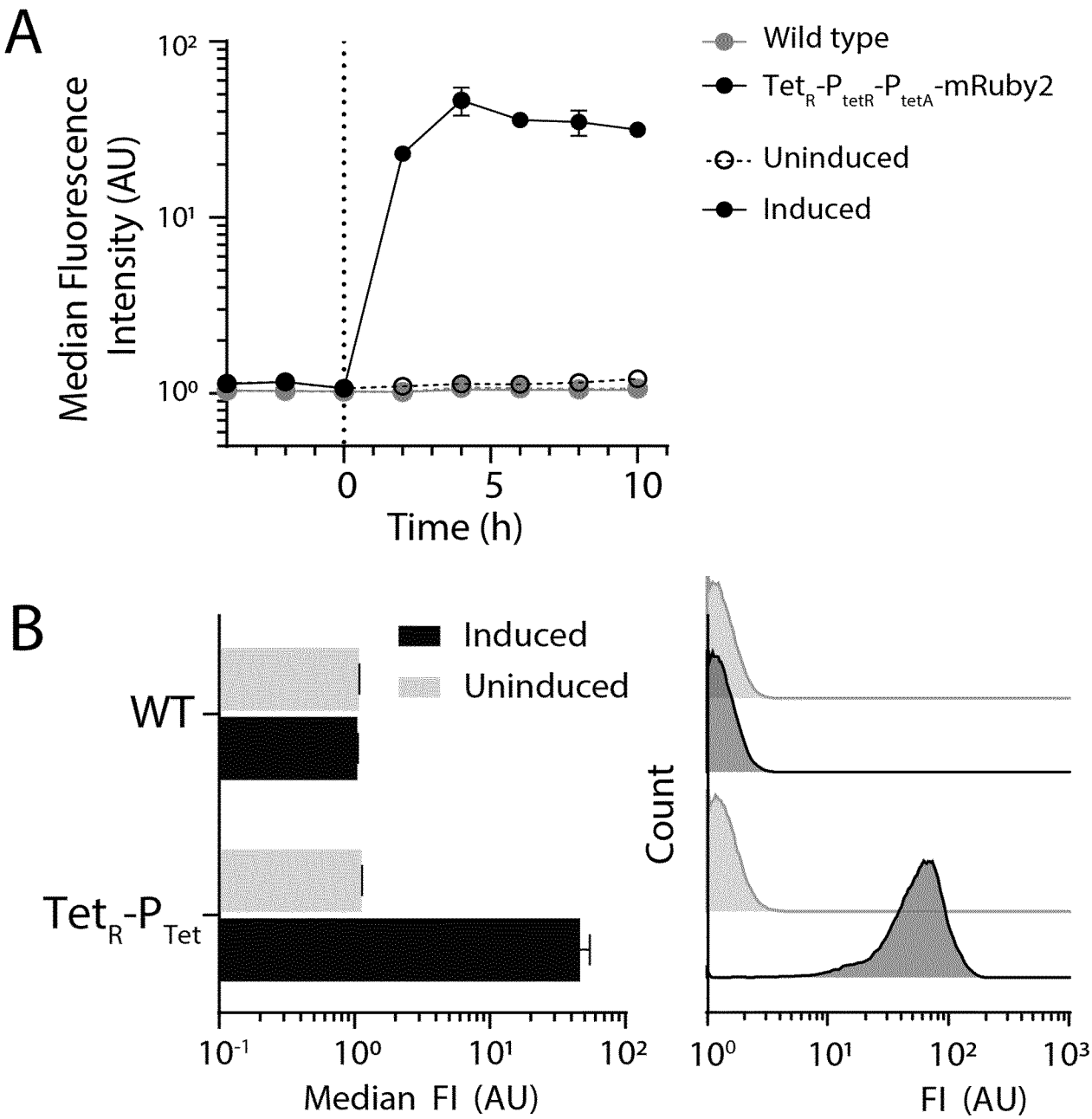
30

35

5 *kristjanssonii*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor lactoaceticus*,
Moorella thermoacetica, *Moorella thermoautotrophica*, *Thermus thermophilus*, *Thermus*
aquaticus, *Thermotoga maritima*, *Pseudothermotoga lettingae*, *Pseudothermotoga*
thermarum, *Chloroflexus aurantiacus*, *Anaerocellum thermophilum*, *Rhodothermus*
marinus, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*,
Thermococcus barophilus, *Thermococcus kodakarensis*, *Pyrococcus abyssi*, and
Pyrococcus furiosus.

Fig. 1

1/8



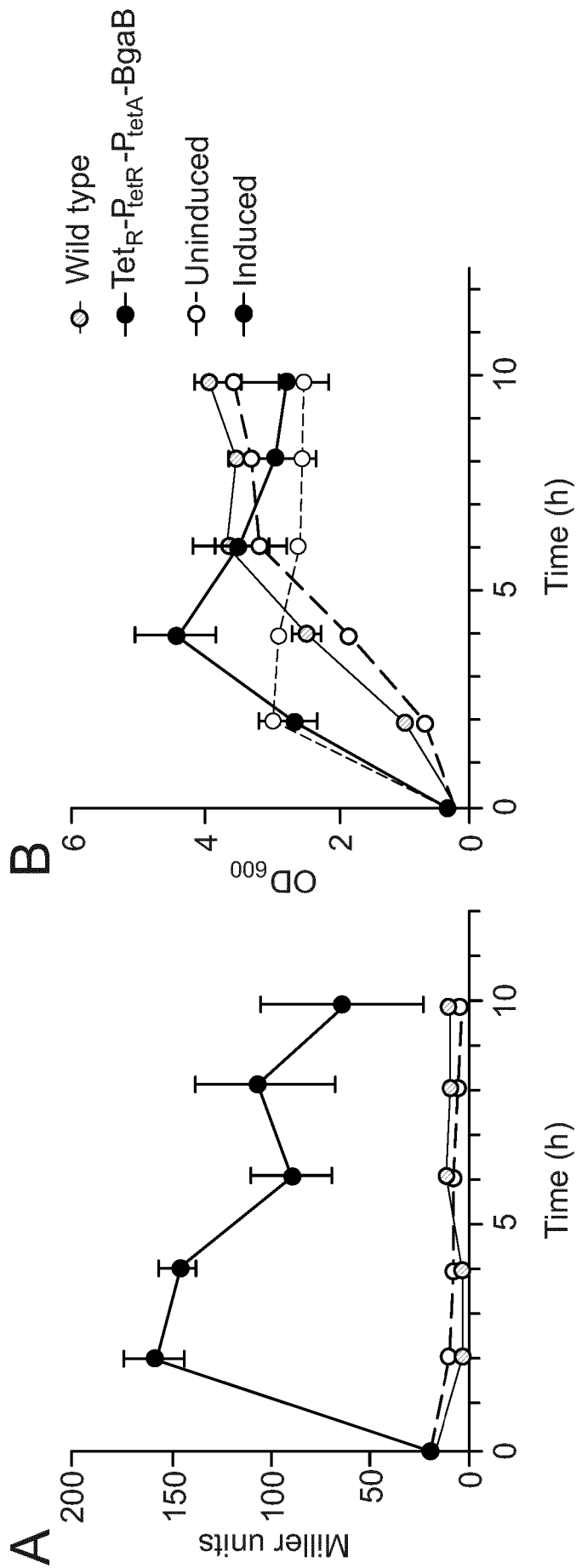


Fig. 2

Fig. 3

3/8

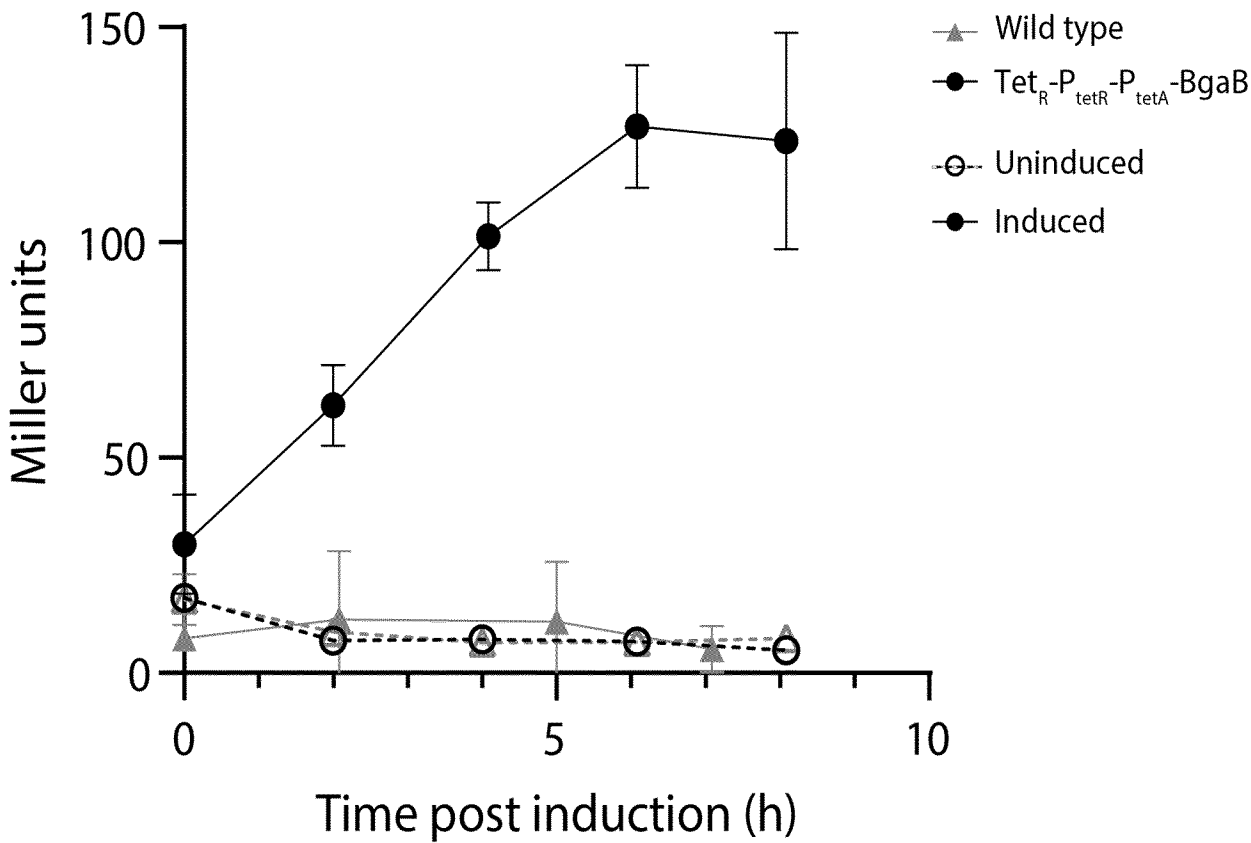


Fig. 4

4/8

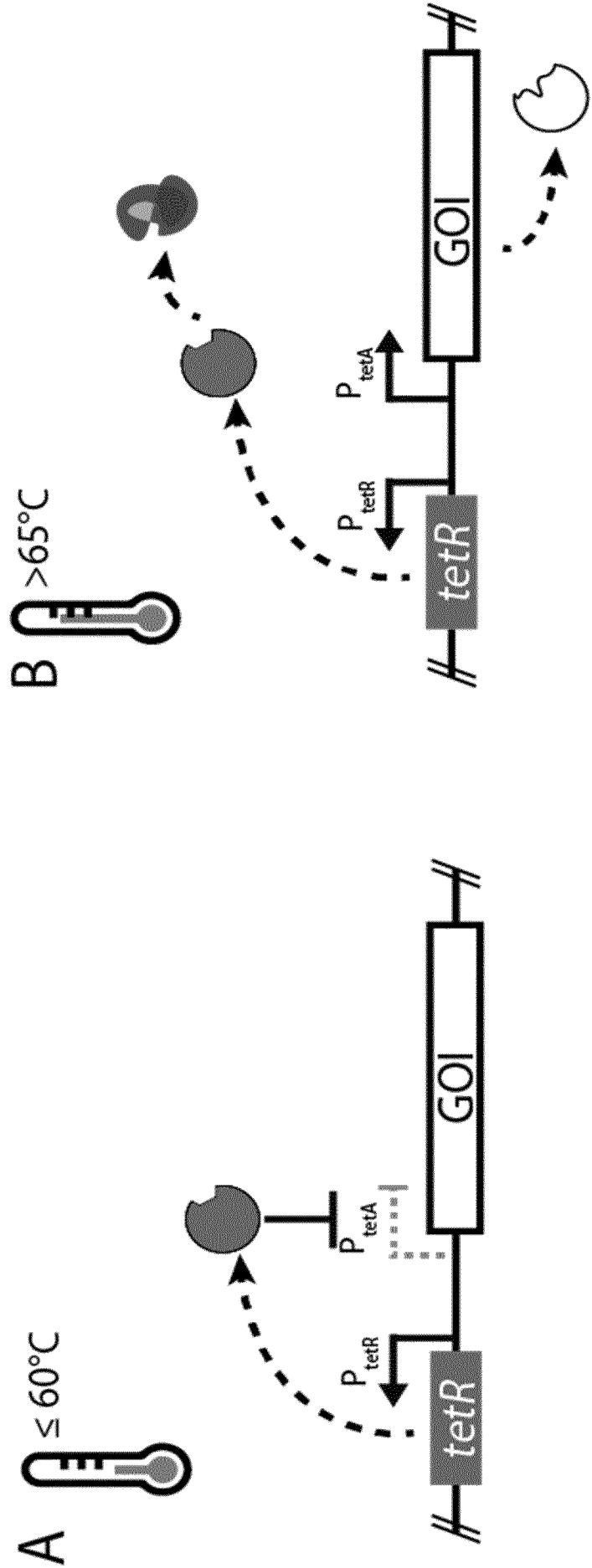


Fig. 5

5/8

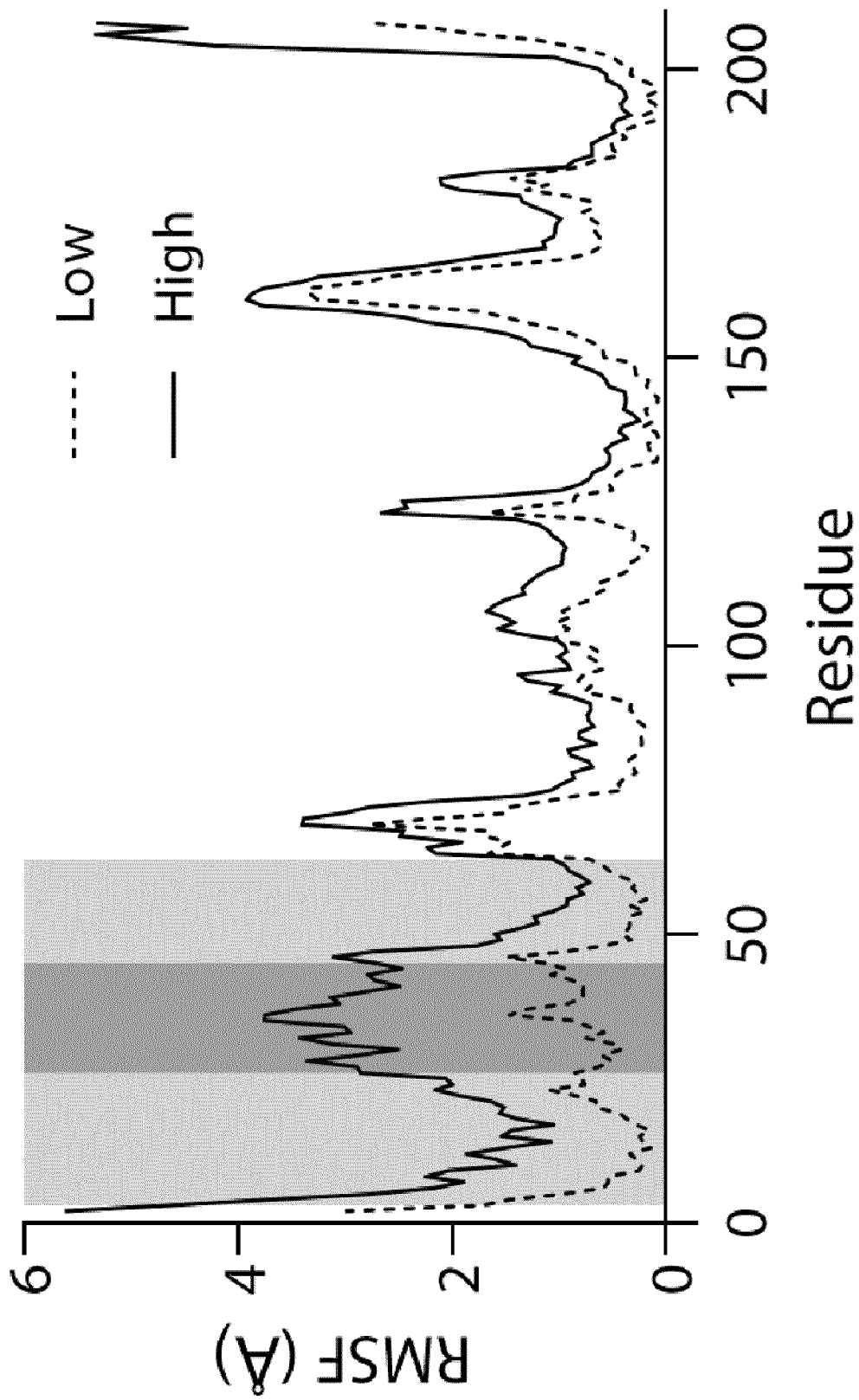


Fig. 6

6/8

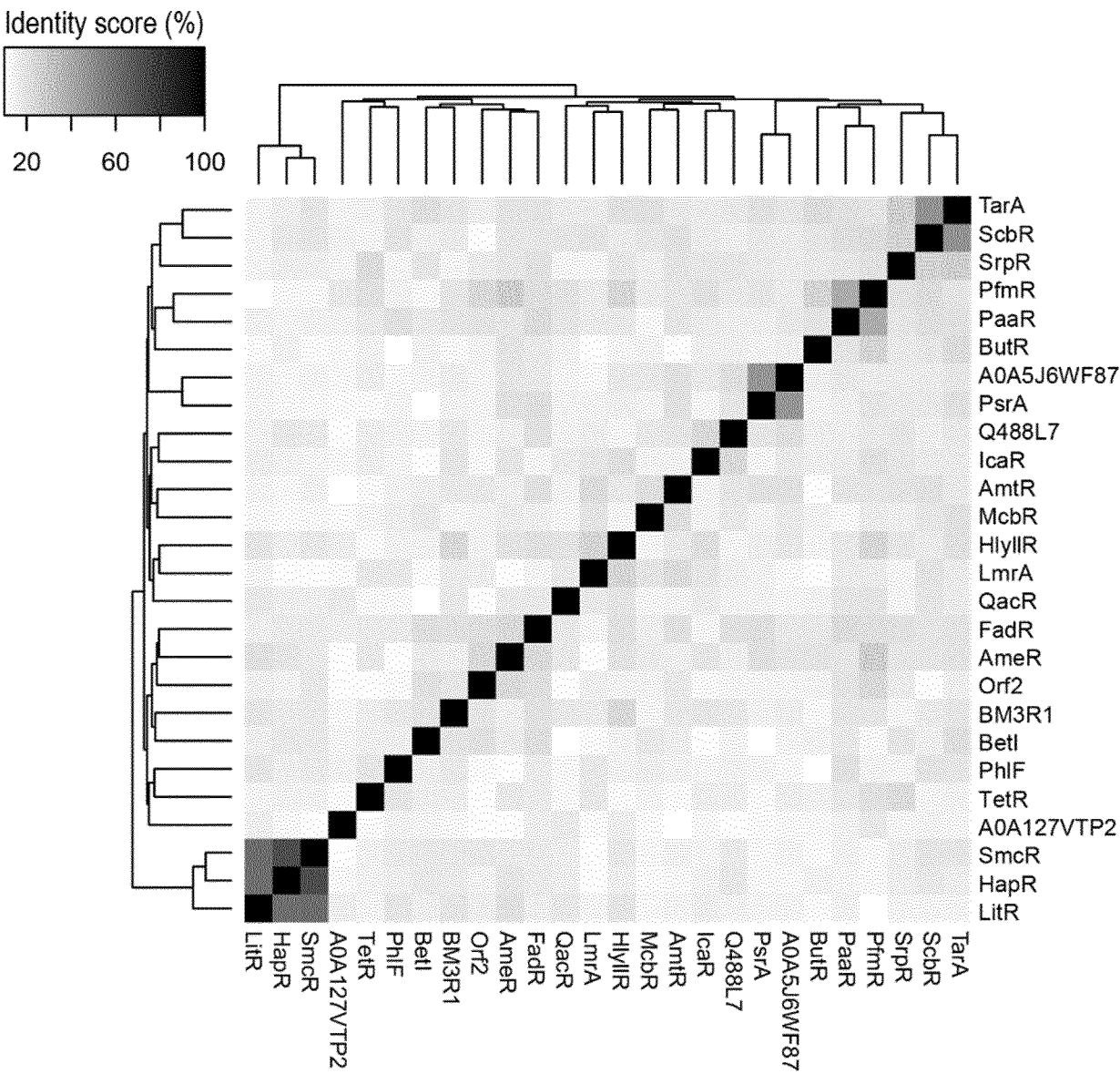


Fig. 7

7/8

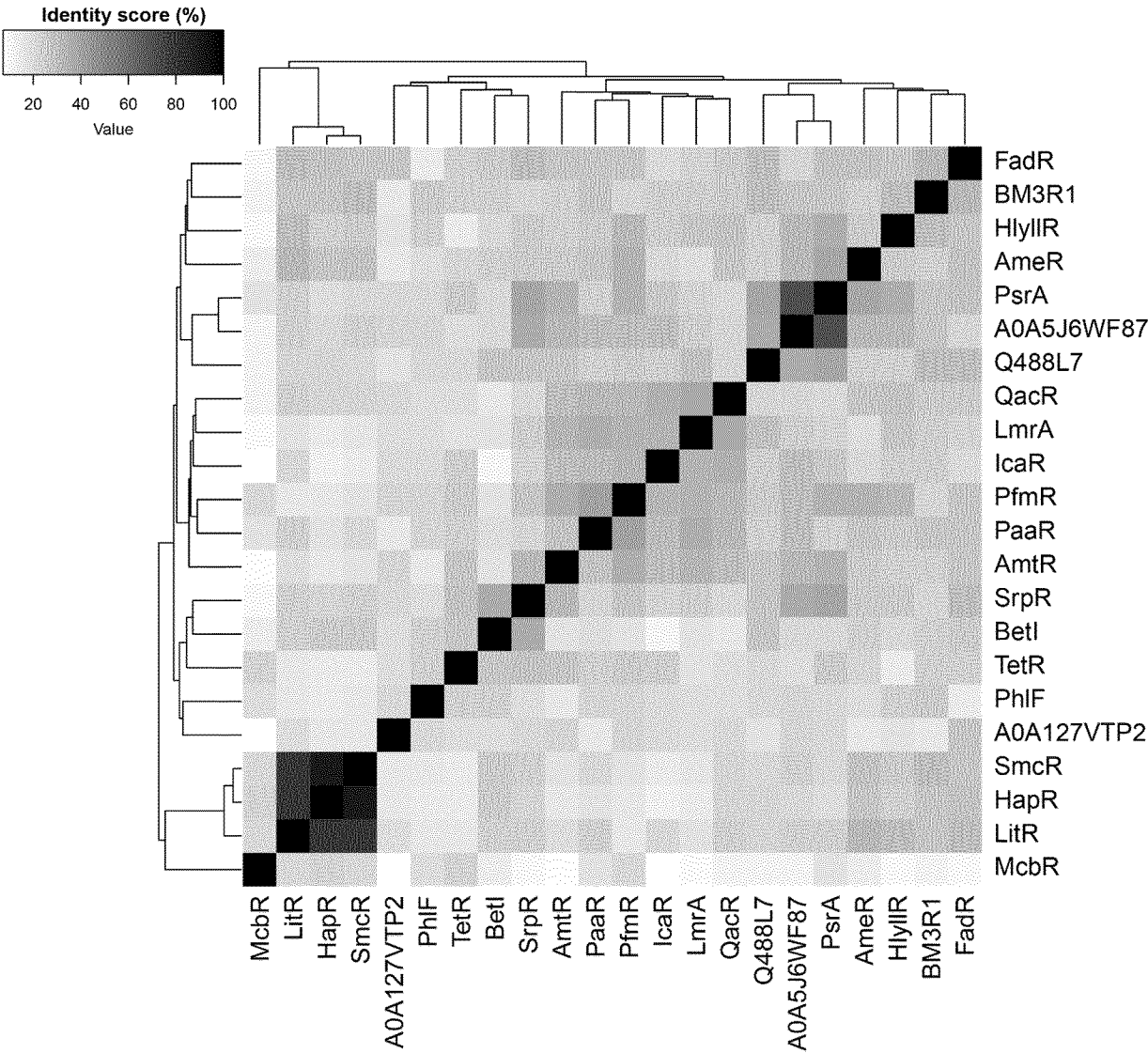
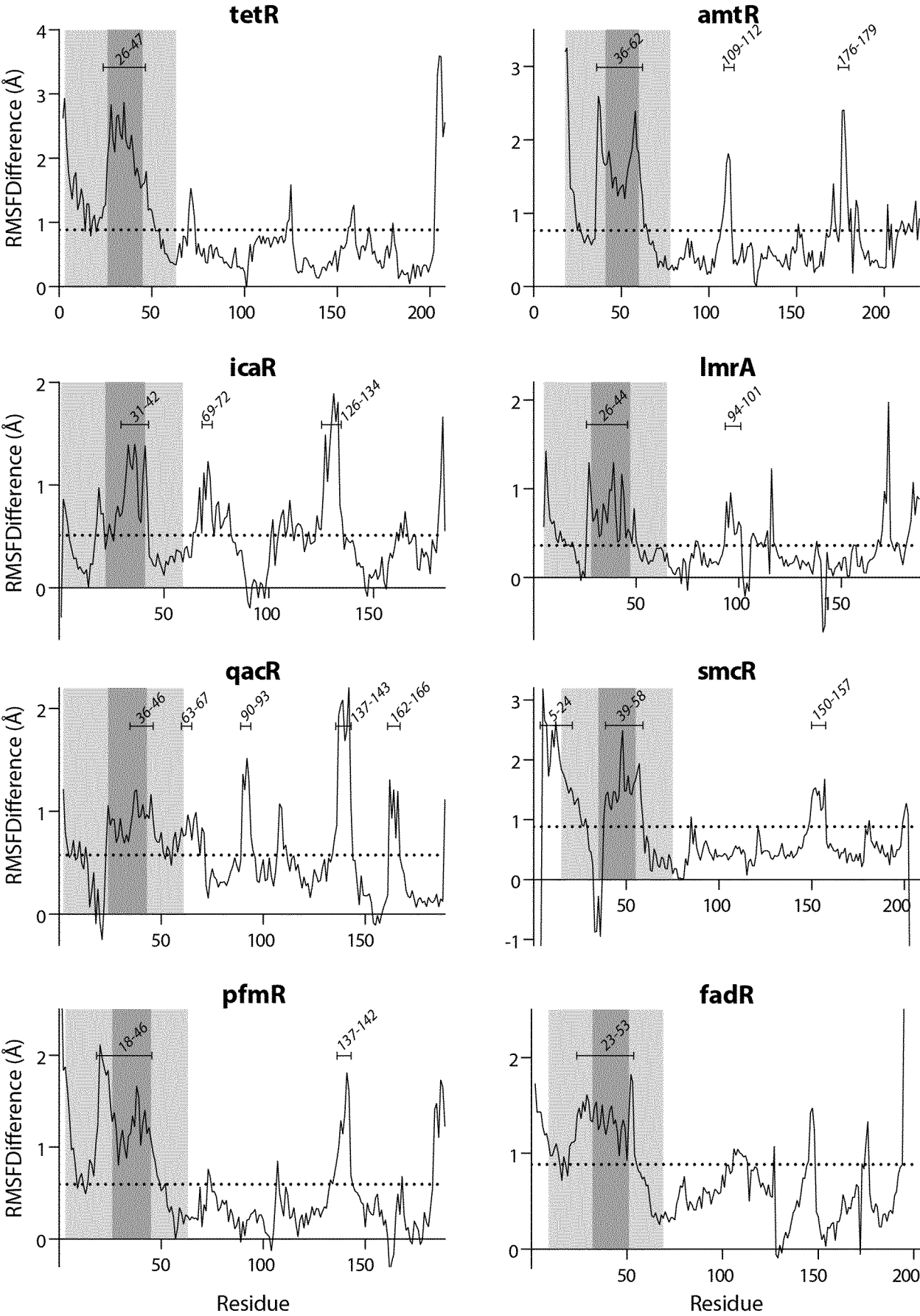


Fig. 8

8/8



INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2022/086436

A. CLASSIFICATION OF SUBJECT MATTER INV. C12N15/10 C12N15/63 C12N15/75 ADD.								
According to International Patent Classification (IPC) or to both national classification and IPC								
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12N								
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched								
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, Sequence Search, WPI Data								
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td> POGREBNIYAKOV IVAN ET AL: "Genetic toolbox for controlled expression of functional proteins in Geobacillus spp.", PLOS ONE, vol. 12, no. 2, 2 February 2017 (2017-02-02), page e0171313, XP055919020, DOI: 10.1371/journal.pone.0171313 see abstract and pages 2-7 <div style="text-align: center;">----- -/--</div> </td> <td>1-16</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	Y	POGREBNIYAKOV IVAN ET AL: "Genetic toolbox for controlled expression of functional proteins in Geobacillus spp.", PLOS ONE, vol. 12, no. 2, 2 February 2017 (2017-02-02), page e0171313, XP055919020, DOI: 10.1371/journal.pone.0171313 see abstract and pages 2-7 <div style="text-align: center;">----- -/--</div>	1-16
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.						
Y	POGREBNIYAKOV IVAN ET AL: "Genetic toolbox for controlled expression of functional proteins in Geobacillus spp.", PLOS ONE, vol. 12, no. 2, 2 February 2017 (2017-02-02), page e0171313, XP055919020, DOI: 10.1371/journal.pone.0171313 see abstract and pages 2-7 <div style="text-align: center;">----- -/--</div>	1-16						
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.								
* Special categories of cited documents : <table border="0"> <tr> <td style="vertical-align: top;"> "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="vertical-align: top;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family				
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family							
Date of the actual completion of the international search		Date of mailing of the international search report						
7 March 2023		15/03/2023						
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Vix, Olivier						

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2022/086436

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>US 2018/195085 A1 (SAEKI KOICHI [JP]) 12 July 2018 (2018-07-12) The query sequence SEQ ID NO:1 has 98.07 % identity (99.03 % similarity) over 207 positions in a common overlap (range (q:s): 1-207:1-207) with subject GSP:BDD28272 (length: 207) from WO2016125364-A1 published on 2016-08-11, same family member as US2018195085.</p> <p>-----</p>	1-16
Y	<p>WANG KUN ET AL: "A TetR-family transcription factor regulates fatty acid metabolism in the archaeal model organism <i>Sulfolobus acidocaldarius</i>", NATURE COMMUNICATIONS</p> <p>,</p> <p>vol. 10, no. 1 4 April 2019 (2019-04-04), XP055918914, DOI: 10.1038/s41467-019-09479-1 Retrieved from the Internet: URL: http://www.nature.com/articles/s41467-019-09479-1 the whole document</p> <p>-----</p>	1-16
Y	<p>YANG ZHIHENG ET AL: "Engineering thermophilic <i>Geobacillus thermoglucosidasius</i> for riboflavin production", MICROBIAL BIOTECHNOLOGY</p> <p>,</p> <p>vol. 14, no. 2 25 February 2020 (2020-02-25), pages 363-373, XP055919027, GB ISSN: 1751-7915, DOI: 10.1111/1751-7915.13543 Retrieved from the Internet: URL: https://onlinelibrary.wiley.com/doi/full/10.1111/1751-7915.13543 see abstract and pages 363-366</p> <p>-----</p>	1-16
Y	<p>KELLNER NIKOLA ET AL: "Developing genetic tools to exploit <i>Chaetomium thermophilum</i> for biochemical analyses of eukaryotic macromolecular assemblies", SCIENTIFIC REPORTS</p> <p>,</p> <p>vol. 6, no. 1 11 February 2016 (2016-02-11), XP055918803, DOI: 10.1038/srep20937 Retrieved from the Internet: URL: https://www.nature.com/articles/srep20937.pdf see abstract and pages 2-7</p> <p>-----</p>	1-16

-/--

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2022/086436

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>AGARI Y. ET AL: "Transcriptional Repression Mediated by a TetR Family Protein, PfmR, from <i>Thermus thermophilus</i> HB8", JOURNAL OF BACTERIOLOGY (PRINT) , vol. 194, no. 17 1 September 2012 (2012-09-01), pages 4630-4641, XP055918805, US ISSN: 0021-9193, DOI: 10.1128/JB.00668-12 Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415517/pdf/zjb4630.pdf see abstract and pages 4631-4634 -----</p>	1-16
Y	<p>KLUGE JANINA ET AL: "Inducible promoters and functional genomic approaches for the genetic engineering of filamentous fungi", APPLIED MICROBIOLOGY AND BIOTECHNOLOGY , vol. 102, no. 15 2 June 2018 (2018-06-02), pages 6357-6372, XP055918797, Berlin/Heidelberg ISSN: 0175-7598, DOI: 10.1007/s00253-018-9115-1 Retrieved from the Internet: URL:https://link.springer.com/content/pdf/10.1007/s00253-018-9115-1.pdf see abstract and pages 6358-6361, Fig. 1-2, pg 6367-6368 -----</p>	1-16
Y	<p>BERTRAM RALPH ET AL: "The application of Tet repressor in prokaryotic gene regulation and expression", JOURNAL OF MICROBIAL BIOTECHNOLOGY, BIOTECH PUBLICATIONS, KAMPUR, IN , vol. 1, no. 1 1 January 2008 (2008-01-01), pages 2-16, XP002703937, ISSN: 0256-8551, DOI: 10.1111/J.1751-7915.2007.00001.X Retrieved from the Internet: URL:http://onlinelibrary.wiley.com/doi/10.1111/j.1751-7915.2007.00001.x/abstract [retrieved on 2007-08-24] cited in the application the whole document ----- -/--</p>	1-16

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2022/086436

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>TAKAI KEN ET AL: "Acquired Thermotolerance and Temperature-Induced Protein Accumulation in the Extremely Thermophilic Bacterium <i>Rhodothermus obamensis</i>", JOURNAL OF BACTERIOLOGY (PRINT) / vol. 180, no. 10 15 May 1998 (1998-05-15), pages 2770-2774, XP055921060, US ISSN: 0021-9193, DOI: 10.1128/JB.180.10.2770-2774.1998 Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107234/pdf/jb002770.pdf see abstract and pages 2771-2773 -----</p>	1-16
Y	<p>US 2009/215168 A1 (LEE BRADY D [US] ET AL) 27 August 2009 (2009-08-27) The query sequence SEQ ID NO:52 has 100.00 % identity (100.00 % similarity) over 182 positions in a common overlap (range (q:s): 3-184:1-182) with subject GSP:BFG11776 (length: 182) from US2017058281-A1 published on 2017-03-02. The query sequence SEQ ID NO:52 has 100.00 % identity (100.00 % similarity) over 182 positions in a common overlap (range (q:s): 3-184:1-182) with subject GSP:BFG11776 (length: 182) from US2017058281-A1 published on 2017-03-02. -----</p>	11
Y	<p>JP 2002 325574 A (RIKAGAKU KENKYUSHO) 12 November 2002 (2002-11-12) The query sequence SEQ ID NO:52 has 100.00 % identity (100.00 % similarity) over 203 positions in a common overlap (range (q:s): 1-203:1-203) with subject CAS:2002_856298_475531523_1 (length: 203) from JP2002325574-A published on 2002-11-12. -----</p>	11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2022/086436

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2022/086436

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018195085 A1	12-07-2018	JP 6769880 B2	14-10-2020
		JP 2020195381 A	10-12-2020
		JP WO2016125364 A1	30-11-2017
		US 2018195085 A1	12-07-2018
		WO 2016125364 A1	11-08-2016

US 2009215168 A1	27-08-2009	AU 2009244775 A1	12-11-2009
		BR PI0907407 A2	24-09-2019
		CA 2712103 A1	12-11-2009
		CN 102083988 A	01-06-2011
		EP 2245040 A2	03-11-2010
		JP 2011512799 A	28-04-2011
		US 2009215168 A1	27-08-2009
		US 2014273238 A1	18-09-2014
		US 2016040174 A1	11-02-2016
		US 2017058281 A1	02-03-2017
		WO 2009137145 A2	12-11-2009

JP 2002325574 A	12-11-2002	NONE	
