



(51) International Patent Classification:

A61K 49/22 (2006.01) A61B 8/08 (2006.01)

(21) International Application Number:

PCT/US2020/012557

(22) International Filing Date:

07 January 2020 (07.01.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/789,295	07 January 2019 (07.01.2019)	US
62/825,612	28 March 2019 (28.03.2019)	US
62/895,553	04 September 2019 (04.09.2019)	US

(71) Applicant: **CALIFORNIA INSTITUTE OF TECHNOLOGY** [US/US]; 1200 East California Boulevard, M/C 6-32, PASADENA, California 91125 (US).

(72) Inventors: **SAWYER, Daniel P.**; c/o California Institute of Technology, 1200 East California Boulevard, M/C 6-32, PASADENA, California 91125 (US). **FARHADI, Arash**; c/o California Institute of Technology, 1200 East

California Boulevard, M/C 6-32, PASADENA, California 91125 (US). **BAR-ZION, Avinoam D.**; c/o California Institute of Technology, 1200 East California Boulevard, M/C 6-32, PASADENA, California 91125 (US). **SHAPIRO, Mikhail**; c/o California Institute of Technology, 1200 East California Boulevard, M/C 6-32, PASADENA, California 91125 (US).

(74) Agent: **CASH, Brian J.** et al.; 155 North Lake Avenue, Suite 700, Pasadena, California 91101 (US).

(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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, 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.

(54) Title: BURST ULTRASOUND RECONSTRUCTION WITH SIGNAL TEMPLATES AND RELATED METHODS AND SYSTEMS

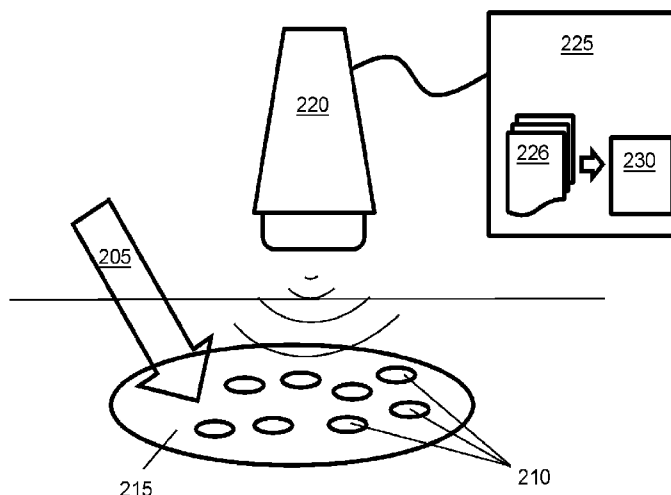


Fig. 2

(57) Abstract: The application of a step function increase in acoustic pressure during ultrasound imaging using gas vesicle contrast, along with capturing successive frames of ultrasound imaging and extracting time-series vectors for pixels of the frames, allows for improved imaging down to even the cell level. Template vectors can be used to implement signal separation of the time-series vectors to improve detection.



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, 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, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, 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).

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

BURST ULTRASOUND RECONSTRUCTION WITH SIGNAL TEMPLATES AND RELATED METHODS AND SYSTEMS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Application No. 62/895,553, entitled "BURST Ultrasound Reconstruction with Signal Templates" filed on September 4, 2019, with docket number CIT-8337-P, and U.S. Provisional Application No. 62/789295, entitled "Mammalian Expression Of Gas Vesicles As Acoustic Reporter Genes" filed on January 7, 2019, with docket number CIT-8165-P, all of which are incorporated herein by reference in their entirety.

[0002] The present application is also related to co-pending U.S. Application No. _____, entitled "Gas Vesicle Gene Clusters, Genetic Circuits, Vectors, Mammalian Cells, Compositions, Methods And Systems For Contrast-Enhanced Imaging", filed on _____, with docket number P2420-US, which is incorporated herein by reference in its entirety.

STATEMENT OF INTEREST

[0003] This invention was made with government support under Grant No. EB018975 awarded by the National Institute of Health. The government has certain rights in the invention."

FIELD

[0004] The present disclosure relates to gas-filled structures for use in imaging technologies, and related compositions methods and systems to image a target site with particular reference to imaging performed by ultrasound.

BACKGROUND

[0005] Ultrasound is among the most widely used biomedical imaging modalities due to its superior spatiotemporal resolution, safety, cost and ease of use compared to other techniques.

[0006] In addition to visualizing anatomy and physiology, ultrasound can take advantage of contrast agents to more specifically image blood flow, discern the location of certain molecular targets, and resolve structures beyond its normal wavelength limit via super-localization.

[0007] Challenges remain for identifying and developing methods and biocompatible nanoscale contrast agents for ultrasound detection of a target site obtained with high sensitivity and resolution.

SUMMARY

[0008] Provided herein are systems and methods to ultrasound image gas vesicles at high sensitivity by creating time-series vectors from successive images during a step function increase in acoustic pressure. The systems and methods allow for high sensitivity imaging even down to imaging a single cell.

[0009] According to a first aspect, a method of ultrasound imaging to be used on a target site contrasted with gas vesicles (GVs) having an acoustic collapse pressure threshold, the method comprising:

- applying ultrasound to the target site at a peak positive pressure less than the acoustic collapse pressure threshold;

- increasing peak positive pressure (PPP) to above the selective acoustic collapse pressure value as a step function;

- imaging the target site in successive frames during the increasing; and extracting a time-series vector for each of at least one pixel of the successive frames. This method requires non-collapsed GV's that may be expressed in native or non-native host cells, isolated from prokaryotes, or produced via cell-free expression.

[0010] According to a second aspect, a system for imaging a target site contrasted with gas vesicles (GVs) having an acoustic collapse pressure threshold, the system comprising: an ultrasound source capable of producing peak positive pressure both below and above the acoustic collapse pressure threshold; an ultrasound imager configured to capture successive frames from the target site; and a processor configured to: calculate a time-series vector for each of at least one pixel of the successive frames.

[0011] The processor can be further configured to perform a signal separation algorithm on the time-series vectors using at least one template vector. The can further comprise a means for introducing the gas vesicles at the target site. Delivering the GV's to the target site can be using

an acoustic reporter gene to express the GVs. The acoustic reporter gene can be in a mammalian cell such as a human embryonic kidney cell or a bacterial cell such as *E. coli* or *S. typhimurium*.

[0012] The primary advantage of BURST (Burst Ultrasound Reconstruction with Signal Templates) is its improvement in sensitivity of up to 1,000,000-fold compared with conventional B-mode ultrasound. BURST also achieves high specificity by cancelling signal from strong linear scatterers such as biological tissue. Unlike contrast mode ultrasound imaging methods such as amplitude modulation and pulse inversion that rely on linear acoustic wave propagation, the specificity of BURST does not deteriorate at higher acoustic pressures where acoustic wave propagation becomes significantly nonlinear.

[0013] The imaging methods and systems herein described can be used in connection with various applications wherein reporting of biological events in a target site is desired. For example, the imaging methods and systems herein described can be used for visualization of biological events, such as a gene expression, proteolysis, biochemical reactions as well as cell location on a target site (e.g. tumor cells inside a host individual, such as mammalian hosts), facilitating for example the study of the mammalian microbiome and the development of diagnostic and therapeutic cellular agents, among other advantages identifiable by a skilled person, in medical applications, as well diagnostics applications. Additional exemplary applications include uses of imaging methods and systems herein described in several fields including basic biology research, neuroscience, applied biology, bio-engineering, bio-energy, medical research, medical diagnostics, therapeutics, and in additional fields identifiable by a skilled person upon reading of the present disclosure.

[0014] The details of one or more embodiments of the disclosure are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0015] The accompanying drawings, which are incorporated into and constitute a part of this specification, illustrate one or more embodiments of the present disclosure and, together with the detailed description and the examples, serve to explain the principles and implementations of the

disclosure.

[0016] **Figure 1** shows an example of a gas vesicle used for BURST imaging.

[0017] **Figure 2** shows a schematic representation of an exemplary system for implementing BURST detection methods herein described.

[0018] **Figure 3** shows an example method of using BURST for imaging.

[0019] **Figure 4** shows an example of the BURST paradigm. **Panel (a)** shows an illustration of the GV collapse. **Panel (b)** shows three consecutive images from the successive images taken during the collapse. **Panel (c)** shows a contrast-to-noise ratio (CNR) vs. frame number. **Panel (d)** shows example output of the template projection algorithm. **Panel (e)** shows example output of the template unmixing algorithm.

[0020] **Figure 5** shows examples of loBURST and hiBURST collapse signal generation. Panels (a) and (c) show the power spectra resulting from BURST acquisitions. Panels (b) and (d) show their corresponding images. **Panels (a) and (b)** show the power spectra and images acquired using standard BURST imaging parameters. **Panels (c) and (d)** show the power spectra and images acquired using a 10-cycle pulse at 5 MHz. **Panel (e)** shows an image time series acquired with an ultrafast version of hiBURST. **Panel (f)** shows the time domain signal used to generate the power spectrum in **panel (a)** and **panel (g)** shows the time domain signal used to generate the power spectrum in **panel (c)**. **Panel (h)** shows BURST images acquired with the 10-cycle sequence at pressures near the 10-cycle loBURST threshold.

[0021] **Figure 6** shows an example of in vitro BURST imaging. **Panels (a)-(d)** show an array of ultrasound images of a cross section of cylindrical wells containing ARG-expressing Nissle *E. coli* embedded in non-scattering agarose. **Panels (e)-(h)** show ultrasound images of the same conditions as **panels (a)-(d)**, but with the cells embedded in tissue-mimicking material (TMM) inside the wells. **Panels (i)-(l)** show CTR vs log cell concentration for loBURST and hiBURST. **Panel (i)** shows loBURST on agar-embedded cells. **Panel (j)** shows hiBURST on agar-embedded cells. **Panel (k)** shows loBURST on TMM-embedded cells. **Panel (l)** shows hiBURST on TMM-embedded cells.

[0022] **Figures 7A and 7B** show an example of in vivo BURST imaging. **Panel (a)** is an illustration of a colon injection experiment. **Panel (b)** shows a collapse frame AM image of the mouse colon filled with probiotic ARG-expressing *E. coli* Nissle. **Panel (c)** shows a BURST image with template projection. **Panel (d)** is an illustration of an oral gavage experiment. **Panels (e)-(f)** show B-mode and BURST images of a coronal cross section of the mouse abdominal cavity. **Panel (e)** shows control gavage of luciferase-expressing Salmonella, with the BURST image displayed below the corresponding B-mode image. **Panel (f)** shows gavage of ARG-expressing Salmonella, with the BURST image displayed below the corresponding B-mode image. **Panel (g)** shows a plot of mean BURST CTR in the abdominal cavity vs distance of the image plane in the caudal direction from the rib cage for mice gavaged with ARG-expressing Salmonella and for mice gavaged with luciferase-expressing Salmonella. **Panels (h)-(i)** show four image planes following those in panels (e) and (f). **Panel (h)** shows spatial sequence frames for a control mouse, with the BURST images displayed below the corresponding B-mode images. **Panel (i)** shows spatial sequence frames for a mouse with ARG-expressing Salmonella, with the BURST images displayed below the corresponding B-mode images.

[0023] **Figure 8** shows an example of single cell detection compared to a control. **Panel (a)** shows a picture of the example experimental setup. **Panel (b)** shows a plot of the average number of single sources counted in images acquired with hiBURST vs cell concentration. **Panel (c)** shows representative images acquired with hiBURST showing single sources in liquid buffer suspension. **Panel (d)** shows representative images of liquid buffer suspension with collapsed ARG-expressing *E. coli* Nissle.

[0024] **Figure 9** shows an example of in vitro ultrasound imaging of gene expression. **Panel (A)** illustrates an ultrasound paradigm used to extract gas vesicle-specific ultrasound image from ARG-expressing cells. **Panel (B)** shows representative non-linear echoes received during this ultrasound imaging paradigm. **Panel (C)** shows cellular viability after being insonated under 8.3 MPa acoustic pressures. **Panel (D)** shows ultrasound imaging of ARG-expressing cells as a function of expression duration. **Panel (E)** shows example ultrasound imaging of ARG-expressing cells as a function of doxycycline induction concentrations. **Panel (F)** shows example ultrasound imaging of ARG-expressing cells mixed with mCherry-only control cells in varying proportions **Panel (G)** illustrates that ARG-expressing cells can re-express gas vesicles after

acoustic collapse.

[0025] Figure 10 shows examples of shortBURST and longBURST signal generation and illustrates how the signal properties change with number of transmit waveform cycles. **Panel (a)** shows representative echoes received following the application of shortBURST at varying pressure levels, indicated by the text in the corresponding rows of panel (c). **Panel (b)** shows representative echoes received following the application of longBURST at varying pressure levels, indicated by the text in the corresponding rows of panel (c). **Panel (c)** shows the power spectra of shortBURST (dark gray) and longBURST (light gray) at each pressure level, obtained by averaging the time-domain signals over the 64 ray lines in each of the 10 replicates. **Panel (d)** shows the peak intensity observed in the shortBURST and longBURST images as a function of peak positive pressure (PPP). **Panel (e)** shows the persistence and gradual disappearance of several bright sources generated by longBURST. **Panel (f)** shows representative images of obtained by applying hiBURST with varying numbers of waveform cycles. **Panel (g)** shows the mean intensity of the hiBURST images (average over 10 replicates) as a function of depth for different numbers of waveform cycles. **Panel (h)** shows the peak mean intensity as a function of number of waveform cycles. **Panel (i)** shows the full-width at half maximum (FWHM) of the mean intensity vs. depth profiles as a function of number of waveform cycles.

DETAILED DESCRIPTION

[0026] Provided herein are gas-filled protein structures, also referred to as “gas vesicles” (GVs), and related compositions methods and systems for use in ultrasound imaging particularly in contrast enhanced ultrasound imaging.

[0027] The term “contrast enhanced imaging” or “imaging”, as herein indicates a visualization of a target site performed with the aid of a contrast agent administered to the target site to improve the visibility of structures or fluids by devices process and techniques suitable to provide a visual representation of a target site. Accordingly contrast agent is a substance that enhances the contrast of structures or fluids within the target site, producing a higher contrast image for evaluation.

[0028] The term “ultrasound imaging” or “ultrasound scanning” or “sonography” as used herein

indicate imaging performed with techniques based on the application of ultrasound. Ultrasound refers to sound with frequencies higher than the audible limits of human beings, typically over 20 kHz. Ultrasound devices typically can range up to the gigahertz range of frequencies, with most medical ultrasound devices operating in the 1 to 18 MHz range. The amplitude of the waves relates to the intensity of the ultrasound, which in turn relates to the pressure created by the ultrasound waves. Applying ultrasound can be accomplished, for example, by sending strong, short electrical pulses to a piezoelectric transducer directed at the target. Ultrasound can be applied as a continuous wave, or as wave pulses as will be understood by a skilled person.

[0029] Accordingly, the wording “ultrasound imaging” as used herein refers in particular to the use of high frequency sound waves, typically broadband waves in the megahertz range, to image structures in the body. The image can be up to 3D with ultrasound. In particular, ultrasound imaging typically involves the use of a small transducer (probe) transmitting high-frequency sound waves to a target site and collecting the sounds that bounce back from the target site to provide the collected sound to a computer using sound waves to create an image of the target site. Ultrasound imaging allows detection of the function of moving structures in real-time. Ultrasound imaging works on the principle that different structures/fluids in the target site will attenuate and return sound differently depending on their composition. Ultrasound imaging can be performed with conventional ultrasound techniques and devices displaying 2D images as well as three-dimensional (3-D) ultrasound that formats the sound wave data into 3-D images. In addition to 3D ultrasound imaging, ultrasound imaging also encompasses Doppler ultrasound imaging, which uses the Doppler Effect or signal decorrelation to measure and visualize movement, such as blood flow rates. Types of Doppler imaging includes continuous wave Doppler, where a continuous sinusoidal wave is used and pulsed wave Doppler, which uses pulsed waves transmitted at a constant repetition frequency. Doppler measurements can be imaged using color flow imaging which uses the phase shift between pulses to determine velocity information which is given a false color (such as red = flow towards viewer and blue = flow away from viewer) superimposed on a grey-scale anatomical image, power Doppler which uses the amplitude of Doppler signal to detect moving matter, or some other method. Ultrasound imaging can use linear or non-linear propagation depending on the signal level. Harmonic and harmonic transient ultrasound response imaging can be used for increased axial resolution, as

harmonic waves are generated from non-linear distortions of the acoustic signal as the ultrasound waves insonate tissues in the body.

[0030] Other ultrasound techniques and devices suitable to image a target site using ultrasound would be understood by a skilled person.

[0031] The term “target site” as used herein indicates an environment comprising one or more targets intended as a combination of structures and fluids to be contrasted, such as cells. In particular, the term “target site” refers to biological environments such as cells, tissues, organs *in vitro*, *in vivo* or *ex vivo* that contain at least one target. A target is a portion of the target site to be contrasted against the background (e.g. surrounding matter) of the target site. Accordingly, a target can include any molecule, cell, tissue, body part, body cavity, organ system, whole organisms, collection of any number of organisms within any suitable environment *in vitro*, *in vivo* or *ex vivo* as will be understood by a skilled person. Exemplary target sites include collections of microorganisms, including, bacteria or archaea in a solution *in vitro*, as well as cells grown in an *in vitro* culture, including, primary mammalian cells, immortalized cell lines, tumor cells, stem cells, and the like. Additional exemplary target sites include tissues and organs in an *ex vivo* culture and tissue, organs, or organ systems in a subject, for example, lungs, brain, kidney, liver, heart, the central nervous system, the peripheral nervous system, the gastrointestinal system, the circulatory system, the immune system, the skeletal system, the sensory system, within a body of an individual and additional environments identifiable by a skilled person. The term “individual” or “subject” or “patient” as used herein in the context of imaging includes a single plant or animal and in particular higher plants or animals and in particular vertebrates such as mammals and more particularly human beings. Types of ultrasound imaging of biological target sites include abdominal ultrasound, vascular ultrasound, obstetrical ultrasound, hysterosonography, pelvic ultrasound, renal ultrasound, thyroid ultrasound, testicular ultrasound, and pediatric ultrasound as well as additional ultrasound imaging as would be understood by a skilled person.

[0032] In embodiments herein described the ultrasound imaging of target site is performed in connection with the administration to the target site of gas vesicle protein structures.

[0033] The wordings “gas vesicles”, “GV”, “gas vesicles protein structure”, or “GVPS”, refer to

a gas-filled protein structure natively intracellularly expressed by certain bacteria or archaea as a mechanism to regulate cellular buoyancy in aqueous environments [4]. In particular, gas vesicles are protein structures natively expressed almost exclusively in microorganisms from aquatic habitats, to provide buoyancy by lowering the density of the cells [4]. GVs have been found in over 150 species of prokaryotes, comprising cyanobacteria and bacteria other than cyanobacteria [5, 6], from at least 5 of the 11 phyla of bacteria and 2 of the phyla of archaea described by Woese (1987) [7]. Exemplary microorganisms expressing or carrying gas vesicle protein structures and/or related genes include cyanobacteria such as *Microcystis aeruginosa*, *Aphanizomenon flos aquae*, *Oscillatoria agardhii*, *Anabaena*, *Microchaete diplosiphon* and *Nostoc*; phototropic bacteria such as *Amoebobacter*, *T. hiodiclyon*, *Pelodiclyon*, and *Ancalochloris*; non phototropic bacteria such as *Microcycilus aquaticus*; Gram-positive bacteria such as *Bacillus megaterium*; Gram-negative bacteria such as *Serratia*, as well as additional microorganisms identifiable by a skilled person.

[0034] In particular, a GV in the sense of the disclosure is an intracellularly expressed structure forming a hollow structure wherein a gas is enclosed by a protein shell, which is a shell substantially made of protein (at least 95% protein). In gas vesicles in the sense of the disclosure, the protein shell is formed by a plurality of proteins herein also indicated as GV proteins or “gvp”s, which form in the cytoplasm a gas permeable and liquid impermeable protein shell configuration encircling gas. Accordingly, a protein shell of a GV is permeable to gas but not to surrounding liquid such as water. In particular, GV protein shells exclude water but permit gas to freely diffuse in and out from the surrounding media [8] making them physically stable despite their usual nanometer size, unlike microbubbles, which trap pre-loaded gas in an unstable configuration.

[0035] GV structures are typically nanostructures with widths and lengths of nanometer dimensions (in particular with widths of 45 - 250 nm and lengths of 100 - 800 nm) but can have lengths up to 2µm in prokaryotes or 8 to 10 µm in mammalian cells as will be understood by a skilled person upon reading of the present disclosure. In certain embodiments, the gas vesicles protein structure have average dimensions of 1000 nm or less, such as 900 nm or less, including 800 nm or less, or 700 nm or less, or 600 nm or less, or 500 nm or less, or 400 nm or less, or 300 nm or less, or 250 nm or less, or 200 nm or less, or 150 nm or less, or 100 nm or less, or 75 nm

or less, or 50nm or less, or 25 nm or less, or 10 nm or less. For example, the average diameter of the gas vesicles may range from 10 nm to 1000 nm, such as 25 nm to 500 nm, including 50 nm to 250 nm, or 100 nm to 250 nm. By “average” is meant the arithmetic mean.

[0036] GVs in the sense of the disclosure have different shapes depending on their genetic origins [8]. For example, GVs in the sense of the disclosure can be substantially spherical, ellipsoid, cylindrical, or have other shapes such as football shape or cylindrical with cone shaped end portions depending on the type of bacteria providing the gas vesicles.

[0037] Representative examples of endogenously expressed GVs native to bacterial or archaeal species are the gas vesicle protein structure produced by the Cyanobacterium *Anabaena flos-aquae* (Ana GVs) [4], and the Halobacterium *Halobacterium salinarum* (Halo GVs) [9]. In particular, Ana GVs are cone-tipped cylindrical structures with a diameter of approximately 140 nm and length of up to 2 μ m and in particular 200-800 nm or longer. Halo GVs are typically spindle-like structures with a maximal diameter of approximately 250 nm and length of 250-600 nm.

[0038] Additional, GVs can be found based on the fact that in bacteria or archaea expressing GVs, the genes (herein also gvp genes) encoding for the proteins forming the GVs (herein also GV proteins), are organized in a gas vesicle gene cluster of 8 to 14 different genes depending on the host bacteria or archaea, as will be understood by a skilled person.

[0039] The term “Gas Vesicle Genes Cluster” or “GVGC” as described herein indicates a gene cluster encoding a set of GV proteins capable of providing a GV upon expression within a bacterial or archaeal cell. The term “gene cluster” as used herein means a group of two or more genes found within an organism's DNA that encode two or more polypeptides or proteins, which collectively share a generalized function or are genetically regulated together to produce a cellular structure and are often located within a few thousand base pairs of each other. The size of gene clusters can vary significantly, from a few genes to several hundred genes [10]. Portions of the DNA sequence of each gene within a gene cluster are sometimes found to be similar or identical; however, the resulting protein of each gene is distinctive from the resulting protein of another gene within the cluster. Genes found in a gene cluster can be observed near one another on the same chromosome or native plasmid DNA, or on different, but homologous

chromosomes. An example of a gene cluster is the Hox gene, which is made up of eight genes and is part of the Homeobox gene family. In the sense of the disclosure, gene clusters as described herein also comprise gas vesicle gene clusters, wherein the expressed proteins thereof together are able to form gas vesicles.

[0040] The term “gene” as used herein indicates a polynucleotide encoding for a protein that in some instances can take the form of a unit of genomic DNA within a bacteria, plant, or other organism.

[0041] The term “polynucleotide” as used herein indicates an organic polymer composed of two or more monomers including nucleotides, nucleosides or analogs thereof. The term “nucleotide” refers to any of several compounds that consist of a ribose or deoxyribose sugar joined to a purine or pyrimidine base and to a phosphate group and that are the basic structural units of nucleic acids. The term “nucleoside” refers to a compound (as guanosine or adenosine) that consists of a purine or pyrimidine base combined with deoxyribose or ribose and is found especially in nucleic acids. The term “nucleotide analog” or “nucleoside analog” refers respectively to a nucleotide or nucleoside in which one or more individual atoms have been replaced with a different atom or a with a different functional group. Accordingly, the term polynucleotide includes nucleic acids of any length, and in particular DNA RNA analogs and fragments thereof.

[0042] The term “protein” as used herein indicates a polypeptide with a particular secondary and tertiary structure that can interact with another molecule and in particular, with other biomolecules including other proteins, DNA, RNA, lipids, metabolites, hormones, chemokines, and/or small molecules. The term “polypeptide” as used herein indicates an organic linear, circular, or branched polymer composed of two or more amino acid monomers and/or analogs thereof. The term “polypeptide” includes amino acid polymers of any length including full-length proteins and peptides, as well as analogs and fragments thereof. A polypeptide of three or more amino acids is also called a protein oligomer, peptide, or oligopeptide. In particular, the terms “peptide” and “oligopeptide” usually indicate a polypeptide with less than 100 amino acid monomers. In particular, in a protein, the polypeptide provides the primary structure of the protein, wherein the term “primary structure” of a protein refers to the sequence of amino acids

in the polypeptide chain covalently linked to form the polypeptide polymer. A protein “sequence” indicates the order of the amino acids that form the primary structure. Covalent bonds between amino acids within the primary structure can include peptide bonds or disulfide bonds, and additional bonds identifiable by a skilled person. Polypeptides in the sense of the present disclosure are usually composed of a linear chain of alpha-amino acid residues covalently linked by peptide bond or a synthetic covalent linkage. The two ends of the linear polypeptide chain encompassing the terminal residues and the adjacent segment are referred to as the carboxyl terminus (C-terminus) and the amino terminus (N-terminus) based on the nature of the free group on each extremity. Unless otherwise indicated, counting of residues in a polypeptide is performed from the N-terminal end (NH₂-group), which is the end where the amino group is not involved in a peptide bond to the C-terminal end (-COOH group) which is the end where a COOH group is not involved in a peptide bond. Proteins and polypeptides can be identified by x-ray crystallography, direct sequencing, immunoprecipitation, and a variety of other methods as understood by a person skilled in the art. Proteins can be provided *in vitro* or *in vivo* by several methods identifiable by a skilled person. In some instances where the proteins are synthetic proteins in at least a portion of the polymer two or more amino acid monomers and/or analogs thereof are joined through chemically-mediated condensation of an organic acid (-COOH) and an amine (-NH₂) to form an amide bond or a “peptide” bond.

[0043] As used herein the term “amino acid”, “amino acid monomer”, or “amino acid residue” refers to organic compounds composed of amine and carboxylic acid functional groups, along with a side-chain specific to each amino acid. In particular, alpha- or α- amino acid refers to organic compounds composed of amine (-NH₂) and carboxylic acid (-COOH), and a side-chain specific to each amino acid connected to an alpha carbon. Different amino acids have different side chains and have distinctive characteristics, such as charge, polarity, aromaticity, reduction potential, hydrophobicity, and pKa. Amino acids can be covalently linked to form a polymer through peptide bonds by reactions between the amine group of a first amino acid and the carboxylic acid group of a second amino acid. Amino acid in the sense of the disclosure refers to any of the twenty naturally occurring amino acids, non-natural amino acids, and includes both D and L optical isomers.

[0044] In embodiments herein described identification of a gene cluster encoding GV proteins

naturally expressed in bacteria or archaea as described herein can be performed for example by isolating the GVs from the bacteria or archaea, isolating the protein for the protein shell of the GV and deriving the related amino acidic sequence with methods and techniques identifiable by a skilled person. The sequence of the genes encoding for the GV proteins can then be identified by methods and techniques identifiable by a skilled person. For example, gas vesicle gene clusters can also be identified by persons skilled in the art by performing gene sequencing or partial- or whole-genome sequencing of organisms using wet lab and *in silico* molecular biology techniques known to those skilled in the art. As understood by those skilled in the art, gas vesicle gene clusters can be located on the chromosomal DNA or native plasmid DNA of microorganisms. After performing DNA or cDNA isolation from a microorganism, the polynucleotide sequences or fragments thereof or PCR-amplified fragments thereof can be sequenced using DNA sequencing methods such as Sanger sequencing, DNaseq, RNASeq, whole genome sequencing, and other methods known in the art using commercially available DNA sequencing reagents and equipment, and then the DNA sequences analyzed using computer programs for DNA sequence analysis known to skilled persons.

[0045] In some embodiments, identification of a gene cluster encoding for GV proteins [9-11] can also be performed by screening DNA sequence databases such as GenBank, EMBL, DNA Data Bank of Japan, and others. Gas vesicle gene cluster gene sequences in databases such as those above can be searched using tools such as NCBI Nucleotide BLAST and the like, for gas vesicle gene sequences and homologs thereof, using gene sequence query methods known to those skilled in the art. For example, genes of the gene cluster for the exemplary haloarchaeal GVs (which have the largest number of different gvp genes) and their predicted function and features are illustrated in Example 26 of related U.S. Application No. 15/613,104, filed on June 2, 2017 which is incorporated herein by reference in its entirety.

[0046] A GV gene cluster encoding for GV proteins typically comprises Gas Vesicle Assembly (GVA) genes and Gas Vesicle Structural (GVS) genes.

[0047] The term Gas Vesicle Structural (GVS) proteins as used herein indicates proteins forming part of a gas-filled protein structure intracellularly expressed by certain bacteria or archaea and can be used as a mechanism to regulate cellular buoyancy in aqueous environments [8]. In particular, GVS shell comprises a GVS identified as gvpA or gvpB (herein also referred to

as gvp A/B) and optionally also a GVS identified as gvpC.

[0048] In particular gvpB gene is a gene encoding for gas vesicle structural protein B. gvpB genes is highly homologous to gvpA gene encoding for gas vesicle structural protein A. A gvp A/B is a protein of the GV shell that has a higher than 70% identity to the following consensus sequence: SSSLAEVLDRILDKGXVIDAWARVSLVGIEILTIEARVVIASVDTYLR (SEQ ID NO: 1) wherein X can be any amino acid. In particular in a gvp A/B of prokaryotes, the consensus sequence of SEQ ID NO: 1 typically forms a conserved secondary structure having an alpha-beta- beta-alpha structural motif formed by portions of the consensus sequence comprising the amino acids LDRILD (SEQ ID NO:2) having an alpha helical structure, RILDKGXVIDAWARVS (SEQ ID NO:3) wherein X can be any amino acid, having a beta strand, beta strand structure, and DTYLR (SEQ ID NO:4) having an alpha helical structure, as will be understood by a skilled person.

[0049] As used herein, "homology", "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to the nucleotide bases or residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity or similarity is used in reference to proteins, it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted with a functionally equivalent residue of the amino acid residues with similar physiochemical properties and therefore do not change the functional properties of the molecule.

[0050] A functionally equivalent residue of an amino acid used herein typically refers to other amino acid residues having physiochemical and stereochemical characteristics substantially similar to the original amino acid. The physiochemical properties include water solubility (hydrophobicity or hydrophilicity), dielectric and electrochemical properties, physiological pH, partial charge of side chains (positive, negative or neutral) and other properties identifiable to a person skilled in the art. The stereochemical characteristics include spatial and conformational arrangement of the amino acids and their chirality. For example, glutamic acid is considered to be a functionally equivalent residue to aspartic acid in the sense of the current disclosure. Tyrosine and tryptophan are considered as functionally equivalent residues to phenylalanine.

Arginine and lysine are considered as functionally equivalent residues to histidine.

[0051] A person skilled in the art would understand that similarity between sequences is typically measured by a process that comprises the steps of aligning the two polypeptide or polynucleotide sequences to form aligned sequences, then detecting the number of matched characters, i.e. characters similar or identical between the two aligned sequences, and calculating the total number of matched characters divided by the total number of aligned characters in each polypeptide or polynucleotide sequence, including gaps. The similarity result is expressed as a percentage of identity.

[0052] As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[0053] As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length protein or protein fragment. A reference sequence can comprise, for example, a sequence identifiable a database such as GenBank™ and UniProt™ and others identifiable to those skilled in the art.

[0054] Thus, a gvp A/B protein in a prokaryote of interest can be identified for example by isolating GVs from a prokaryote of interest, isolating the protein from the protein shell of the GV and obtaining the amino acid sequence of the isolated protein. In addition to, or in the alternative to, isolating the GVs and isolating the protein, the method can include obtaining amino acidic sequences of the shell proteins of the GV of the prokaryote of interest from available database. The method further comprises performing a sequence alignment of the obtained amino acidic sequences against the gvp A/B protein consensus sequence of SEQ ID NO:1.

[0055] In particular the isolating GVs from a prokaryote of interest can be performed following methods to isolate gas vesicles as described in U.S. Application No. 15/613,104, filed on June 2, 2017. Isolating the protein for the protein shell of the GV and obtaining the related amino acidic sequence can be performed with tandem liquid chromatography mass-spectrometry alone or in combination with obtaining amino acid sequences of the isolated protein with wet lab techniques or from available databases comprising the sequences of the prokaryote of interest as well as additional techniques and approaches identifiable by a skilled person. Obtaining amino acid sequences of GV shell proteins of the prokaryote of interest can be performed by screening available databases of gene and protein sequences identifiable by a skilled person. Performing a sequence alignment of the sequences of the isolated GV proteins or proteins encoded in the genome of a prokaryote of interest can be performed (using Protein BLAST as described herein) against the gvp A/B protein consensus sequence of SEQ ID NO:1. In particular, a sequence alignment can be performed using gvp A/B protein sequences from the closest phylogenetic relative to the prokaryote of interest.

[0056] The optional gvpC gene encodes for a gvpC protein which is a hydrophilic protein of a GV shell, including repetitions of one repeat region flanked by an N-terminal region and a C terminal region. The term “repeat region” or “repeat” as used herein with reference to a protein refers to the minimum sequence that is present within the protein in multiple repetitions along the protein sequence without any gaps. Accordingly, in a gvpC multiple repetitions of a same repeat is flanked by an N-terminal region and a C-terminal region. In a same gvpC, repetitions of a same repeat in the gvpC protein can have different lengths and different sequence identity one with respect to another. In performing alignment steps sequence are identified as repeat when the sequence shows at least 3 or more of the characteristics described in US application 15,663,635 published as US 2018/0030501 (incorporated herein by reference in its entirety) which also include additional features of gvpC proteins and the related identification.

[0057] In a GVGC, the GVS genes are comprised with Gas Vesicle Assembly genes. The Gas Vesicle Assembly genes are genes encoding for GVA proteins. GVA proteins comprise proteins with various putative functions such as nucleators and/or chaperons as well as proteins with an unknown specific function related to the assembly of the GV.

[0058] In a prokaryotic cell GVA genes are all the genes within one or more operons comprising

at least one of a gvpN and a gvpF excluding any gvp A/B and gvpC gene possibly present within said one or more operons. Therefore, GVA genes can be identified by identifying an operon in a prokaryote including at least one of a gvpN and a gvpF excluding any gvp A/B and gvpC gene.

[0059] Preferably the one or more operons comprising all the GVA genes of a prokaryote can be identified and detected by detecting a gvpN gene encoding for a GVP protein consensus sequence

RALXYLQAGYXVHXRGPA GTGKTTLAMHLAXXLXRPV MLIXGDDEFXTSDLIGSESGY
XXXKKVVDNYIHSVVKVEDEL RQNWVDNRLTXACREGFTLVYDEFNRSRPEXNNVLLS
VLEEKILXLP (SEQ ID NO: 5) wherein X indicates any amino acid or a sequence of any length having at least 50%, and more preferably 60% or higher, most preferably from 50% to 83% identity.

[0060] GvpN genes of various microorganisms have a sequence encoding for a gvpN protein within the consensus SEQ ID NO: 5. In particular, gvpN gene in the sense of the disclosure is gene encoding for sequence

MTVLTDKRKKGSGAFIQDDETKEVLSRALS YLKSGYSIHFTGPAGGGKTSLARALAKKR
KRPVMLMHGNHELNNKDLIGDFTGYTSKKVIDQYVRSVYKKDEQVSENWQDGRLL EA
VKNGYTLIYDEFTRSKPATNNIFLSILEEGVLPLYGVKMTDPFVRVHPDFRVIFTSNPAEY
AGVYDTQDALLDRLITMFIDYKDIDRETAILTEKTDVEEDEARTIVTLVANVRNRSGDEN
SSGLSLRASLMIA TLATQQDIPIDGSDEDFQTL CIDILHHPLTKCLDEENAKSKAEKIILEE
CKNIDTEEK (SEQ ID NO: 6)

or a sequence of any length having at least 30% sequence identity with respect to SEQ ID NO:6, preferably at least 50%, and more preferably 60% or higher,

and gvpF gene in the sense of the disclosure is gene encoding for sequence

MSETNETGIYIFSAIQTDKDEEFGAVEVEGT KAETFLIRYKDAAMVAAEVPMKIYHPNR
QNLLMHQNAVA AIMDKNDTVIPISFGNVFKSKEDVKV LLENLYPQFEKLFPAIKGKIEVG
LKVIGKKKEWLEKKVNENPELEKVSASVKGKSEAAGYYERIQLGGMAQKMFTSLQKEV
KTDVFSPL EEAEEAAKANEPTGETMLLNASFLINREDEAKFDEKVN EAHENWKDKADF
HYS GPWPAYNFVNIRLKVEEK (SEQ ID NO:7)

or a sequence of any length having at least 20% sequence identity with respect to SEQ ID NO:7, preferably at least 50%, more preferably 60%, and at least 70% or higher.

[0061] The term “operon” as described herein indicates a group of genes arranged in tandem in a prokaryotic genome as will be understood by a skilled person. Operons typically encode proteins participating in a common pathway are organized together as understood by those skilled in the art. Typically, genes of an operon are transcribed together into a single mRNA molecule referred to as polycistronic mRNA. Polycistronic mRNA comprises several open reading frames (ORFs), each of which is translated into a polypeptide. These polypeptides usually have a related function and their coding sequence is grouped and regulated together in a regulatory region, containing a promoter and an operator. Typically, repressor proteins bound to the operator sequence can physically obstruct the RNA polymerase enzyme from binding the promoter, preventing transcription. An example of a prokaryotic operon is the lac operon, which natively regulates transport and metabolism of lactose in *E. coli* and many other enteric bacteria.

[0062] In an operon, each ORF typically has its own ribosome binding site (RBS) so that ribosomes simultaneously translate ORFs on the same mRNA. Some operons also exhibit translational coupling, where the translation rates of multiple ORFs within an operon are linked. This can occur when the ribosome remains attached at the end of an ORF and translocates along to the next ORF without the need for a new RBS. Translational coupling is also observed when translation of an ORF affects the accessibility of the next RBS through changes in RNA secondary structure.

[0063] In some embodiments, a GV cluster comprises one of gvpN or gvpF. In several embodiments GV clusters include both gvpN and gvpF as will be understood by a skilled person. Accordingly, for a certain prokaryote, GVA genes in the sense of the disclosure indicate all the genes that are comprised in the one or more operons having at least one of a gvpN and/or a gvpF herein described and excluding any Gas Vesicle Structural (GVS) genes of the prokaryotes possibly comprised within the one or more operons.

[0064] Thus, GVA genes comprised in a gas vesicle gene cluster in a prokaryote can be identified for example by obtaining genome sequence of the prokaryote of interest and performing a sequence alignment of the protein sequences encoded in the genome of the

prokaryote of interest against a gvpN protein sequence and/or a gvpF protein sequence.

[0065] In particular, obtaining the genome sequence of the prokaryote of interest, can be performed either using wet lab techniques identifiable by a skilled person upon reading of the present disclosure, or obtained from databases of gene and protein sequences also identifiable by a skilled person upon reading of the present disclosure. Performing a sequence alignment of the protein sequences encoded in the genome of the prokaryote of interest can be performed using Protein BLAST or other alignment algorithms identifiable by a skilled person. Exemplary gvpN protein sequence and/or a gvpF protein sequence, that can be used in performing the alignment are sequences SEQ ID NO:6 and/or SEQ ID NO:7. In particular, a sequence alignment can be performed using gvpN and/or gvpF protein sequences from the closest phylogenetic relative to the prokaryote of interest. Accordingly, one or more operons that comprise the gvpN and/or gvpF genes can be identified, and any other gvps within the one or more operons can also be identified, wherein the other gvps are comprised in ORFs within the one or more operons, excluding any ORFs encoding gvp A/B or gvpC genes comprised in the one or more operons of the GV gene cluster.

[0066] Accordingly, GVA genes can also be identified based on the configuration of operon and Gene Clusters identified through homology, phylogenesis also using the gvp A/B, gvpN and/or gvpF consensus of SEQ ID Nos: 1, 6, and 7 herein provided preferably gvp A/B consensus of SEQ ID NO:1 and gvpN consensus of SEQ ID NO: 5.

[0067] GVS genes of a GVGC of the disclosure, identified with methods herein indicated, typically comprise gvpA or gvpB which have similar sequences and are equivalent in their purpose and optionally gvpC. Exemplary sequences for gvpA and gvpB genes of GV gene clusters in the sense of the disclosure, which can also be used to identify additional GVS and GVGC through homology and alignment.

[0068] GVA genes of a GVGC of the disclosure, identified with methods herein indicated, typically comprise proteins identified as gvpN, F, G, L, S, K, J, and U. GVA genes and proteins can also comprise gvpR and gvpT (see e.g. *B.megaterium* GVA) gvpV, gvpW (see *Anabaena flos aque* and *Serratia* GVA) and/or gvpX, gvpY and gvpZ (see e.g. *Serratia* GVA). Exemplary sequences for GVA genes of GV gene clusters in the sense of the disclosure which can also be

used to identify additional GVAs and GVGC through homology and alignment.

[0069] In GVGC herein described co-expression of the GVS genes and the GVA genes in connection with regulatory sequence capable of operating in a host cell are configured to provide a GV type, with a different GVGC typically resulting in a different GV type.

[0070] The wording “GV type” in the sense of the disclosure indicates a gas vesicle having dimensions and shape resulting in distinctive mechanical, acoustic, surface and/or magnetic properties as will be understood by a skilled person upon reading of the present disclosure. In particular, a skilled person will understand that different shapes and dimensions will result in different properties in view of the indications in provided in US Application S/N 15/613,104 and US S/N 15/663,600 and additional indications identifiable by a skilled person. Typically, larger volume results in stronger per-particle scattering, smaller diameter generally results in higher collapse pressure after removal of gvpC, and different dimensions result in different ratios of T2/T2* relaxivity per volume-averaged magnetic susceptibility [12].

[0071] Accordingly, in embodiments herein described, GVGC can be selected based on desired properties of the corresponding GV type. In particular, to this extent, a skilled person can use naturally occurring GVGC or can provide modified GVGC wherein some of the naturally occurring gvp genes are omitted, or can provide hybrid GVGC in which GVAs and GVS genes of naturally occurring GVGCs are mixed to provide GV types having the shape and dimensions resulting in the desired properties. Typically, a gene cluster of gvp genes (GVGC) comprises at least gvpF, gvpG, gvpL, gvpS, gvpK, gvpJ, and gvpU. Preferably a gene cluster of gvp genes (GVGC) comprises a gvpN

[0072] The term “hybrid gene cluster” or “hybrid cluster” as used herein indicates a cluster comprising at least two genes native to different species and resulting in a cluster not natively in any organisms. Typically, a hybrid gene cluster comprises a subset of gas vesicle genes native to a first bacterial species and another subsets of gas vesicle genes native to one or more bacterial species, with at least one of the one or more bacterial species different from the first bacterial specie Accordingly, a hybrid GV gene clusters including a combination of GV genes which is not native in any naturally occurring prokaryotes.

[0073] For example, in one exemplary embodiment, all the gvp genes B, N, F, G, L, S, K, J and

U are from *B. megaterium*. Mega GVs are typically cone-tipped cylindrical structures with a diameter of approximately 73 nm and length of 100-600 nm, encoded by a cluster of eleven or fourteen different genes, including the primary structural protein, gvpB, and several putative minor components and putative chaperones [11] as would be understood by a person skilled in the art.

[0074] **Figure 1** shows a rendition of engineered GVs illustrating gvpA (101) as the main building block of GVs. GvpA is a structural protein that assembles through repeated units to make up the bulk of GVs. GvpC (102) is a scaffold protein with 5 repeat units that assemble on the outer shell of GVs. GvpC can be engineered to tune the mechanical and acoustic properties of GVs as well as act as a handle for appending moieties (103) on to.

[0075] A gvpC protein is a hydrophilic protein of a GV shell, which includes repetitions of one repeat region flanked by an N-terminal region and a C terminal region. The term “repeat region” or “repeat” as used herein with reference to a protein refers to the minimum sequence that is present within the protein in multiple repetitions along the protein sequence without any gaps. Accordingly, in a gvpC multiple repetitions of a same repeat is flanked by an N-terminal region and a C-terminal region. In a same gvpC, repetitions of a same repeat in the gvpC protein can have different lengths and different sequence identity one with respect to another.

[0076] As indicated above GV structures are typically nanostructures with widths and lengths of nanometer dimensions (in particular with widths of 45 - 250 nm and lengths of 100 - 800 nm) but can have lengths up to 2µm or up to 8-10 µm as will be understood by a skilled person. In certain embodiments, the gas vesicles protein structure have average dimensions of 1000 nm or less, such as 900 nm or less, including 800 nm or less, or 700 nm or less, or 600 nm or less, or 500 nm or less, or 400 nm or less, or 300 nm or less, or 250 nm or less, or 200 nm or less, or 150 nm or less, or 100nm or less, or 75 nm or less, or 50nm or less, or 25 nm or less, or 10 nm or less. For example, the average diameter of the gas vesicles may range from 10 nm to 1000 nm, such as 25 nm to 500 nm, including 50 nm to 250 nm, or 100 nm to 250 nm. By “average” is meant the arithmetic mean.

[0077] GVs in the sense of the disclosure have different shapes depending on their genetic origins. For example, GVs in the sense of the disclosure can be substantially spherical, ellipsoid,

cylindrical, or have other shapes such as football shape or cylindrical with cone shaped end portions depending on the type of bacteria or archaea providing the gas vesicles.

[0078] In embodiments herein described, GVs in the sense of the disclosure are capable of withstanding pressures of several kPa, but collapse irreversibly at a pressure at which the GV protein shell is deformed to the point where it flattens or breaks irreversibly, allowing the gas inside the GV to escape and subsequently dissolve in surrounding media, herein also referred to as a critical collapse pressure, or acoustic collapse pressure threshold, as there are various points along a collapse pressure profile.

[0079] A collapse pressure profile as used herein indicates a range of pressures over which collapse of a population of GVs of a certain type occurs. In particular, a collapse pressure profile in the sense of the disclosure comprise increasing acoustic collapse pressure values, starting from an initial collapse pressure value at which the GV signal/optical scattering by GVs starts to be erased to a complete collapse pressure value at which the GV signal/optical scattering by GVs is completely erased. The collapse pressure profile of a set type of GV is thus characterized by a mid-point pressure where 50% of the GVs of the set type have been collapsed (also known as the “midpoint collapse pressure”), an initial collapse pressure where 5% or lower of the GVs of the type have been collapsed, and a complete collapse pressure where at least 95% of the GVs of the type have been collapsed. In embodiments herein described a selectable critical collapse pressure (herein also “collapse threshold”) can be any of these collapse pressures within a collapse pressure profile, as well as any point between them. The critical collapse pressure profile of a GV is functional to the mechanical properties of the protein shell and the diameter of the shell structure.

[0080] The term the “acoustic pressure” as used herein indicates the pressure exerted by a sound wave, such as ultrasound wave, propagating through a medium. In ultrasound imaging, this wave is typically generated by an ultrasound transducer, and the pressure resulting at any time and point in the medium is determined by transducer output and patterns of constructive and destructive interference, attenuation, reflection, refraction and diffraction. Ultrasound images are generated by transmitting one or more pulses into the medium and acquiring backscattered signals from the medium, which depend on medium composition, including the presence of

contrast agents.

[0081] In embodiments herein described, the collapse behavior of GVs under ultrasound exhibits a spectral pattern, as the GVs can collapse over a range or spectra of continuous increasing acoustic collapse pressure values, starting from an initial collapse pressure value at which the GV signal starts to be erased to a complete collapse pressure value at which the GV signal is completely erased. Therefore, for some embodiments of the method, the method begins with applying ultrasound to a target site at a PPP less than the acoustic collapse pressure threshold. The collapse pressure also can vary based on the frequency of the acoustic signal.

[0082] The acoustic collapse pressures of a given GV type can be characterized by an acoustic collapse pressure profile, which is a normalized sigmoid function $f(p)$ defined as follows:

$$f(p) = (1 + e^{(p-p_c)/\Delta p})^{-1} \quad (1)$$

where p is the applied pressure, p_c is the collapse mid-point and Δp is the variance, the latter two being parameters obtained from fitting with a sigmoid function. The acoustic collapse pressure profile shows normalized ultrasound signal intensities as a function of increasing pressures.

[0083] The acoustic collapse pressure profile of a given GV type can be determined by imaging GVs with imaging ultrasound energy after collapsing portions of the given GV type population with a collapsing ultrasound energy (e.g. ultrasound pulses) with increasing peak positive pressure amplitudes to obtain acoustic pressure data point of acoustic pressure values, the data points forming an acoustic collapse curve. The acoustic collapse pressure function $f(p)$ can be derived from the acoustic collapse curve by fitting the data with a sigmoid function such as a Boltzmann sigmoid function.

[0084] Accordingly, acoustic collapse pressure profile in the sense of the disclosure include a set of initial collapse pressure values, a midpoint collapse pressure value and a set of complete collapse pressure values. The initial collapse pressures are the acoustic collapse pressures at which 5% or less of the GV signal is erased. A midpoint collapse pressure is the acoustic collapse pressure at which 50% of the GV signal is erased. Complete collapse pressures are the acoustic collapse pressures at which 95% or more of the GV signal is erased.

[0085] The initial collapse pressures can be obtained by solving the fitted equations for p such that $f(p) \leq 0.05$. The midpoint collapse pressure can be obtained by solving the fitted equations for p such that $f(p) = 0.5$. The complete collapse pressures can be obtained by solving the fitted equations for p such that $f(p) \geq 0.95$. In some embodiments, the acoustic collapse pressure threshold can be set to either the initial collapse pressure, the midpoint collapse pressure, the complete collapse pressure, or some other value in the collapse profile where collapse occurs. For most practical applications, the acoustic collapse pressure threshold would be set at least as high as the midpoint collapse pressure. If the contrast material is composed of multiple types of GVs, where each type has a different collapse pressure threshold, then the effective collapse pressure threshold for the material can be set to the highest collapse pressure threshold of all of the GV types.

[0086] If the imaging is being performed on living tissue, then care must be taken to not have the PPP pressure damage the tissue. This limit on PPP depends on the target site being imaged (and its surrounding tissue).

[0087] Since method ultrasound imaging of the instant disclosure are based on the acoustic collapse pressure of a GV type, GV types can be tested to identify an acoustic collapse pressure before the related use. In some embodiments, a GV type can also be modified by engineering the corresponding GVGC to provide a GVGV detectable in the target cell and having a desired acoustic collapse pressure as will be understood by a skilled person.

[0088] Identification of a GVGC corresponding to a GV type and detection of the related acoustic collapse pressure in a target cell can be performed through a testing method which can be performed in the target cell where detection of the GV type is desired or in testing cells having a cell environment equivalent to the cell environment of the target cell in terms of expression of GV genes and GV formation and thus provide a model to verify ability of the gvp genes to provide a GVGC for the target cells. If the GVGC is known, it might be possible to look up its acoustic collapse pressure profile or threshold in a database of GVGC.

[0089] In the method, the GVGC cluster can be introduced in the target cell or testing cell using

engineered polynucleotide constructs contacted with the target cell or testing cell for a time and under conditions to allow expression of the GVGC and formation of the GV type (e.g. using the methods described in US application 15,663,635 published as US 2018/0030501 incorporated herein by reference). The method further comprises detecting the acoustic collapse pressure of the GV type in the target cell or testing cell. Preferably the testing can be performed in a target cell or testing cell, that have been modified, either chemically or genetically, to have the same cellular turgor pressure as mammalian cells according to methods identifiable by a skilled person.

[0090] Additionally, or in the alternative, the GVs can be introduced to the target site pre-formed (e.g. formed *in vitro* from a bacteria culture) before the detecting.

[0091] Several detectable GVGC with one or more detection method of interests have been identified and can be used for production of GV types in various cells through various genetically engineered constructs as will be understood by a skilled person upon reading of the present disclosure and US application S/N 15/663,635 herein incorporated by reference in its entirety.

[0092] In some embodiments those GVGC can comprise gvp genes A/B, C and N (gvpB, gvpC and gvpN genes) from a same or different prokaryote. Preferably the GVGC comprises gvpN gene as presence of gvpN protein is known or expected to result in an increased detectability of the related GV type (better signal under ultrasound collapse).

[0093] Exemplary gene clusters which have provided to be detectable in mammalian cells and *E. Coli* comprise gvp genes from *B. megaterium* (herein also mega-gvp) and/or *Anabaena flos-aquae* (herein also Ana-gvp), and in particular those summarized in **Table 1**. The acoustic collapse pressures for the clusters are listed in **Table 1** for frequencies between 5 MHz and 20 MHz.

Table 1 Exemplary GVGCs			
Type of cluster	gvp genes of the GVGC	Acoustic Pressure	Collapse

Table 1 Exemplary GVGCs			
Type of cluster	gvp genes of the GVGC	Acoustic Pressure	Collapse
Naturally occurring in <i>B. megaterium</i>	Mega-gvpB, Mega-gvpN Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpS, Mega-gvpK, Mega-gvpJ, Mega-gvp-R, Mega-gvp-T and Mega-gvpU	1.9 MPa	
Engineered	Mega-gvpB, Mega-gvpN Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpS, Mega-gvpK, Mega-gvpJ, and Mega-gvpU	1.9 MPa	
Naturally occurring in <i>Anabaena flos-aquae</i>	Ana-gvpA, Ana-gvpC, Ana-gvpN, Ana-gvpJ, Ana-gvpK, Ana-gvpF, Ana-gvpG, Ana-gvpV, Ana-gvpW	0.9 MPa	
Engineered	Ana-gvpA, Ana-gvpN, Ana-gvpJ, Ana-gvpK, Ana-gvpF, Ana-gvpG, Ana-gvpV, Ana-gvpW	0.6 MPa	
Hybrid engineered	Ana-gvpA gen, Mega-gvpR, Mega-gvpN, Mega-gvpF, Mega-gvpG, Mega-gvpL, Mega-gvpS, Mega-gvpK, Mega-gvpJ, gvpT and gvpU	0.6 MPa	
Hybrid engineered	Ana-gvpA, Ana-gvpC, Mega-gvpN Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpS, Mega-gvpK, Mega-gvpJ, and Mega-gvpU	2.2 MPa	
Hybrid engineered	Ana-gvpA, Ana-gvpC, Mega-gvpN Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpR Mega-gvpS, Mega-gvpT Mega-gvpK, Mega-gvpJ, and Mega-gvpU	2.2 MPa	
Hybrid engineered	Ana-gvpA, Ana-gvpC Ana-gvpN; Mega- Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpS, Mega-gvpK, Mega-gvpJ, and Mega-gvpU	2.2 MPa	
Hybrid engineered	Ana-gvpA, Ana-gvpC Ana-gvpN; Mega- Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpR, Mega-gvpS, Mega-gvpT, Mega-gvpK, Mega-gvpJ, and Mega-gvpU	2.2 MPa	

[0094] Additional GVGCs can be identified based on the genes and exemplary sequences reported in **Example 1** herein described and the related mechanical and acoustic properties such as acoustic collapse pressure of each GV type is also identifiable by a skilled person upon reading of the present disclosure.

[0095] Based on the above acoustic collapse pressure values, a standard collapse pressure of 4.3 MPa has been established which will result in the collapse of the GV types reported in **Table 1** and is still below 4.6 MPa, a pressure that, according to limits on ultrasound imaging pressure set by the U.S. Food and Drug Administration (USFDA), could be considered damaging to a target site comprising living cells for a longBURST pulse sequence at 6 MHz, assuming peak negative pressure is equal in magnitude to peak positive pressure. In view of known values of acoustic collapse pressure for GVs this standard collapse pressure is expected to work for most GV types and can be used in the testing method to identify acoustic properties of GVs herein described.

[0096] Accordingly different GV types can be provided to be used in a method of ultrasound imaging to be used on a target site contrasted with gas vesicles (GVs) having an acoustic collapse pressure threshold, which comprises: applying ultrasound to the target site at a peak positive pressure less than the acoustic collapse pressure threshold; increasing peak positive pressure (PPP) to above the selective acoustic collapse pressure value as a step function; and imaging the target site in successive frames during the increasing; and extracting a time-series vector for each of at least one pixel of the successive frames.

[0097] In particular, in methods of the instant disclosure, applying ultrasound refers to sending ultrasound-range acoustic energy to a target. The sound energy produced by the piezoelectric transducer can be focused by beamforming, through transducer shape, lensing, or use of control pulses. The soundwave formed is transmitted to the body, then partially reflected or scattered by structures within a body; larger and smoother structures typically reflecting, and smaller or rougher structures typically scattering. The return sound energy reflected/scattered to the transducer vibrates the transducer and turns the return sound energy into electrical signals to be analyzed for imaging. The frequency and pressure of the input sound energy can be controlled and are selected based on the needs of the particular imaging task and, in some methods described herein, collapsing GVs.

[0098] The increasing peak positive pressure (PPP) to above the selective acoustic collapse pressure value as a step function can be performed by implementing an automated pulse sequence on a programmable ultrasound system and transducer in which the voltage applied to the transducer, and thus the PPP, increases during certain successive pulses.

[0099] To create images, particularly 2D and 3D imaging, scanning techniques can be used where the ultrasound energy is applied in lines or slices which are composited into an image. The images can be captured in successive frames, showing images at successively different times typically ranging from 100 microseconds to 100 milliseconds between image frames, depending on the amount of motion in the target.

[00100] In some embodiments, imaging the target site can be performed by scanning an ultrasound image of the target site in successive frames. In some cases, imaging the target site includes transmitting an imaging ultrasound signal from an ultrasound transmitter to the target site, and receiving a set of ultrasound data at a receiver. The visible image is formed by ultrasound signals backscattered from the target site. The ultrasound data can be analyzed using a processor, such as a processor configured to analyze the ultrasound data and produce an ultrasound image from the ultrasound data. In certain embodiments, the ultrasound data detected by the receiver includes an ultrasound signal reflected by the target site of the subject. The imaging can be any type of ultrasound imaging, including the standard B-mode or a contrast mode sequence such as amplitude modulation (AM) or pulse inversion (PI).

[00101] Methods for performing ultrasound imaging are known in the art and can be employed in methods of the current disclosure. In certain aspects, an ultrasound transducer, which comprises piezoelectric elements, transmits an ultrasound imaging signal (or pulse) in the direction of the target site. Variations in the acoustic impedance (or echogenicity) along the path of the ultrasound imaging signal causes backscatter (or echo) of the imaging signal, which is received by the piezoelectric elements. The received echo signal is digitized into ultrasound data and displayed as an ultrasound image. Conventional ultrasound imaging systems comprise an array of ultrasonic transducer elements that are used to transmit an ultrasound beam, or a composite of ultrasonic imaging signals that form a scan line. The ultrasound beam is focused onto a target site by adjusting the relative phase and amplitudes of the imaging signals. The imaging signals are reflected back from the target site and received at the transducer elements. The voltages produced at the receiving transducer elements are summed so that the net signal is indicative of the ultrasound energy reflected from a single focal point in the subject. An ultrasound image is then composed of multiple image scan lines.

[00102] In certain embodiments, the ultrasound signal has a transmit frequency of at least 1 MHz, 5 MHz, 10 MHz, 20 MHz, 30MHz, 40 MHz or 50 MHz. For example, an ultrasound data is obtained by applying to the target site an ultrasound signal at a transmit frequency from 4 to 11 MHz, or at a transmit frequency from 14 to 22 MHz.

[00103] In the embodiments herein described, the collapsing ultrasound and imaging ultrasound are selected to have a collapsing pressure and an imaging pressure amplitude based on the acoustic collapse pressure profile of the GV structure type used in the contrast agent. In some instances, the ultrasound pressure, including the collapsing ultrasound pressure and the imaging ultrasound pressure can be referred to as the “peak positive pressure” of the ultrasound pulses. The term “peak positive pressure” refers to the maximum pressure amplitude of the positive pulse of a pressure wave, typically in terms of the difference between the peak pressure and the ambient pressure at the location in the person or specimen that is being imaged.

[00104] In some embodiments, the GV contrast agent is detected by burst ultrasound reconstruction with signal templates (BURST), which involves applying an ultrasound step function pressure differential to the location of the GV contrast agent and capturing successive frames of the ultrasound image during the increase of pressure. In some embodiments, the ultrasound step function pressure differential increases the acoustic pressure from a pressure below the collapse threshold of the GVs to an acoustic pressure above the collapse threshold of the GVs. Example step function pressure differentials can include increasing the ultrasound peak positive pressure (PPP) from a value under 1 MPa to a value over 1 MPa, such as 3 MPa or higher, 3.7 MPa or higher, 4 MPa or higher, 4.3 MPa or higher, or other values. BURST allows for an ability to detect smaller number of cells than conventional imaging, and even allows sensitivity down to imaging individual cells in the imaging plane. See e.g. **Example 2**.

[00105] The term “peak positive pressure” (or PPP) as used herein refers to the pressure difference from zero to the highest positive pressure (the peak of the positive part of a pressure wave) of the signal. As used herein, the PPP is measured or calculated at the target site, not at the transducer/source. Some attenuation is expected as the ultrasound permeates matter to reach the target site.

[00106] The term “step function” as used herein refers to a strong increase or decrease in value

over a short period of time. The BURST step function is an increase of PPP. The strength of the increase does not need to be particularly strong, so long as there is a clear transition from a PPP below the collapse threshold to a PPP above the collapse threshold, such that the collapse rate prior to the step function increase is very low (ex. <5%) and the collapse rate after the step function increase is high (ex. >80%). Typically, an increase of at least 3MPa is required for most GVs, but the actual value will depend on individual GV collapse sensitivity. Typically, larger pressure increases lead to larger gains in sensitivity. Detection of single cells typically requires a pressure increase of 4 MPa. Because the step function consists of several discrete ultrasound pulses, the speed of the step function transition is equal to the time between ultrasound pulses, which typically matches the time between images frames of 100 microseconds to 100 milliseconds. A step function can include an impulse (a step function increase followed shortly by a step function decrease).

[00107] In some embodiments, the detection includes detecting a transient signal from the GV contrast in the time domain of the ultrasound image. An example of a transient signal is an increase in contrast in the image over less than a second. For example, a transient signal might be present over a few hundred microseconds. The transient signal appears as a strong increase in contrast signal during and after the collapse of the GVs.

[00108] In some embodiments, the detection of the transient signal can be accomplished by imaging the target site in successive frames during the step function increase of pressure (for example, including frames from before collapse, during collapse, and after collapse) and extracting a time-series vector for each pixel from the successive frames.

[00109] The term “time-series vector” as used herein refers to a vector of data taken from multiple points in time for a common pixel location in an image.

[00110] In some embodiments, the method can also comprise performing a signal separation algorithm on the time-series vectors using at least one template vector. Signal separation allows for greater sensitivity of imaging against background noise. Signal separation algorithms include template projection and template unmixing. The at least one template vector can include linear scatterers, noise, gas vesicles, or a combination thereof. The successive frames can comprise a frame prior to GV collapse, a frame during GVs collapse, and a frame after GVs collapse.

[00111] Signal separation algorithms include template projection and template unmixing. In an embodiment, the method of imaging can include template projection and/or template unmixing of template vectors with the pixel vectors. The signal separation algorithm can be implemented in software or firmware/hardware.

[00112] The term “signal separation” as used herein refers to a method of separating the signal from the noise for an image (set of data).

[00113] The term “template vector” as used herein refers to a vector obtained from a previously known signal to allow signal separation for a possibly noisy signal under consideration.

[00114] The unique temporal responses of GVs, linear scatterers, and non-scattering material to this stimulus allows us to use known signal templates to separate the signal due to GVs from signal due to noise or linear scatterers such as biological tissue. Signal templates can be estimated empirically by averaging pixel time series from regions of interest (ROIs) containing known samples, as exemplified in **Example 2** (see **Figure 4**), and, in general, BURST can be used with any number of unique signal templates. However, in one problem setting, the physical properties of GVs and biological tissue can be used to specify exact signal templates *a priori*: a flat function for noise, a step function for linear scatterers, and an impulse function for GVs. Hence, as few as three frames (pre-collapse, collapse, and post-collapse) can be used to distinguish these templates, which would correspond to the following template vectors for:

$$\text{linear scatterers } (u_s = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}), \quad (2)$$

$$\text{noise } (u_n = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}), \text{ and} \quad (3)$$

$$\text{GVs } (u_g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}). \quad (4)$$

[00115] In template projection, the final BURST intensity I for each pixel can be a normalized similarity score computed as the projection of the template vector of interest (in our case u_g) onto

the pixel vector:

$$I = \frac{(u_g^T p)}{\|p\|} \quad (5)$$

[00116] Because the template vector can be projected onto the pixel vector, rather than vice versa, template projection is scale-invariant: pixel locations with clear impulse time traces will have the highest intensity in the final BURST image even if the peak intensities of the time traces are orders of magnitude lower in intensity than those corresponding to surrounding linear scatterers, as is the case exemplified in **Example 2** (see in particular **Figure 4**). In fact, for a given GV signal intensity, stronger scatterer signals will be more efficiently canceled. This contrasts with conventional techniques for improving specificity in ultrasound imaging of contrast agents, which typically improve CTR by an approximately fixed amount that often leaves a visible residual signal *in vivo* [1]. Moreover, as AM and PI rely on linearity of acoustic scattering, their specificity deteriorates rapidly with increasing acoustic pressures, as observed in the exemplary detection of **Example 2** (see in particular **Figure 4**).

[00117] Despite these advantages, template projection has its limitations. Firstly, its scale invariance means that pixel values in the final template projection image do not always directly correspond to physical quantities, making quantification difficult. Second, the performance of template projection might be compromised in scenarios where GV signal is colocalized with strong linear scatterer signal.

[00118] In template unmixing, the colocalization problem can be addressed by modeling each pixel vector as a linear combination of the template vectors. This model can be represented by the linear equation

$$Vw = p, \quad (6)$$

where the template vectors are concatenated into the template matrix

$$V = [u_s \ u_n \ u_g], \quad (7)$$

and w contains the weights for each template. For each pixel vector p , obtain the least squares solution for the template weights by the pseudoinverse:

$$w = (V^T V)^{-1} V^T p \quad (8)$$

[00119] Technically, because negative weights have no meaning in this model, a proper estimation of the template weights would require the appropriate constrained linear least squares solution, which is typically two orders of magnitude slower to compute. However, empirically, setting all negative values of the unconstrained solution to zero results in a final image that is not appreciably different from that obtained using the constrained solution.

[00120] Template unmixing tends to cancel linear scatterers less efficiently than template projection due to the lack of scale invariance (see **Example 2** and in particular **Figure 4, panel (e)**), showing that template projection is preferable in scenarios where GVs are known to not be co-localized with linear scatterers. However, because co-localization occurs in many interesting *in vivo* contexts, and because it is often desired to quantify BURST contrast, template unmixing can be considered the more robust and versatile algorithm generally and it is used for BURST images herein, unless otherwise specified. This method is also applicable in scenarios where a nonlinear signal is produced at the target site. In fact, at high PPP (ex. above 3 MPa), intrinsic nonlinearities in most media result in production of strong nonlinear signal even in the absence of contrast agents. Because the signal separation algorithms employed rely only on detectable changes in the signals generated by GVs, BURST is equally applicable to targets that produce linear signals and those that produce nonlinear signals.

[00121] The PPP used to collapse the GV contrast can be divided into two or more regimes. For example, lower pressure can be considered “loBURST” and higher pressure can be considered “hiBURST”, separated by what the predominant mechanism is for the signal. See e.g. **Example 3**. For PPPs in between loBURST and hiBURST, which may be used when a tradeoff between the benefits and drawbacks of each is desired, the generated signal will consist of a mixture of the mechanisms characterizing each regime.

[00122] In an embodiment, the PPP is in a loBURST regime (relatively low PPP), where the dominate mechanism of the signal is due to an acoustic wave generated by the collapse of the GV shell and the resulting rapid displacement of fluid volume. The loBURST regime is characterized by a signal composed predominantly of dim sources, dominated by the fundamental and second harmonic peaks. An example of loBURST is a PPP of around 3.7 MPa

for a half-cycle duration. The minimum loBURST PPP will depend on the type of GV used, in particular the collapse threshold. The hiBURST PPP can be lower in setups with lower frequencies, larger number of waveform cycles, or less attenuating tissue types, since these factors all contribute to enhancing cavitation.

[00123] In an embodiment, the PPP is in a hiBURST regime (relatively high PPP), where the dominant mechanisms include stable cavitation of nanobubbles liberated from the GVs following collapse, and a limited amount of inertial cavitation in some cases. The hiBURST regime is characterized by a signal composed predominately of bright sources, and the emergence of higher (>2) harmonic peaks. An example of hiBURST is a PPP of around 4.3 MPa for 1.5 half-cycles.

[00124] Operating in a loBURST or hiBURST regime can depend on what is optimal for a particular use case. In cases where it is desirable to maximize sensitivity or detect single cells, such as with highly scattering tissue or low cell and/or GV concentrations, hiBURST is often optimal. However, because hiBURST results in a greater amount of cavitation, it results in a reduction in viability of GV-expressing bacteria. Thus, in cases where it is desirable to minimize effects on host cells and/or surrounding tissue and where cell and/or GV concentrations are sufficient for detection by loBURST, loBURST is often optimal. The loBURST PPP can be lower in setups with GVs that have a lower collapse threshold. The maximum loBURST PPP will increase with the frequency and decrease with the number of waveform half-cycles. For example, a PPP of 4.3 MPa that would normally define a hiBURST regime with a waveform using the standard 3 half-cycles will instead correspond to loBURST when using a short waveform with only 1 half-cycle. This will also depend on the specific transducer model used and its ability to realize the specified number of half-cycles with minimal ringdown.

[00125] Accordingly, in some embodiments, the increasing PPP can be increasing the PPP to a hiBURST regime or increasing the PPP to a loBURST regime. The hiBURST regime can be 4.3 MPa or higher and the loBURST regime can be 3.7 MPa or lower. Other values of hiBURST and loBURST can be used, so long as loBURST is less than hiBURST. The distinction of hiBURST and loBURST is mainly characterized by the differences in the mechanisms behind the signals produced.

[00126] Additionally, the duration of the increased PPP can affect the sensitivity of the imaging. For example, the number of half-cycles in the transmit waveform can be divided into two or more regimes. For example, smaller numbers of half-cycles can be considered “shortBURST” and larger numbers of half-cycles can be considered “longBURST”, separated by what the predominant mechanism is for the signal. See e.g. **Example 10**.

[00127] In an embodiment, the waveform is in a shortBURST regime (relatively small number of half-cycles), where the dominate mechanism of the signal is due to an acoustic wave generated by the collapse of the GV shell and the resulting rapid displacement of fluid volume. The loBURST regime is characterized by a signal composed predominantly of dim sources, dominated by the fundamental and second harmonic peaks. An example of shortBURST is a waveform with 1 half-cycle with a PPP of 4.3. The loBURST regime coincides with the shortBURST regime since both are defined by the dominant signal generation mechanism.

[00128] In an embodiment, the waveform is in a longBURST regime (relatively large number of half-cycles), where the dominant mechanisms include stable cavitation of nanobubbles liberated from the GVs following collapse, and a limited amount of inertial cavitation in some cases. The longBURST regime is characterized by a signal composed predominately of bright sources. An example of longBURST is a waveform with 5 half-cycles with a PPP of 4.3 MPa.

[00129] The BURST technique can be implemented by a combination of hardware, software, and biotechnology. In an embodiment, an example of which is shown in **Figure 2**, a system for imaging a target site (215) can include: a means (205) for introducing GVs (210) to the target site (215), the GVs (210) having a collapse threshold; an ultrasound source (220) capable of producing PPP both below and above the collapse threshold; and an ultrasound imager (225) configured to capture successive frames (226) from the target site (215). A processor (230) can be configured to calculate the pixel vectors from the successive frames (226) and then perform a signal separation algorithm (such as threshold unmixing) on the pixel vectors from a set of template vectors. The processor can be part of the imager (as shown in **Figure 2**), or it can be a part of a separate device. The means for introducing GVs to the target site can include techniques such as injecting GVs to the site, introducing cells containing GVs to the site, modifying cells at the site to produce the GVs, modifying cells to produce GVs then introducing those cells to the

site, combinations thereof, or any similar technique. The ultrasound source typically needs to be capable of producing the PPP both below and above the collapse threshold in order to perform the step function for the BURST. The ultrasound imager needs to be capable of capturing the successive frames of the image.

[00130] Figure 3 shows an example method of using BURST for imaging. In an embodiment, the process begins with the introduction of GVs to the target site (305). The means for introducing the GVs can vary, including injection of the GVs in solution, injection of GVs in host cells, injection of host cells with acoustic reporter genes (either naturally occurring or engineered cells), or engineering cells at the target site to have acoustic reporter genes. The target site can be *in vitro* or *in vivo*.

[00131] An example of introducing GVs to a target site is injecting isolated GVs into the tail vein of a mouse. Another example is mixing engineered GV-expressing bacteria with molten agarose and injecting the solution into the colon of an animal model. Another example is gavaging a solution of GV-expressing bacteria into an animal model and waiting for the cells to propagate through the gastrointestinal tract. Another example is growing a tumor on a mouse model where the tumor is grown from mammalian cells with acoustic reporter genes.

[00132] An ultrasound PPP below the collapse threshold of the GVs is applied to the target site (310), which can be started before, during, or after the GVs are introduced. Image frames are captured in sequence from the ultrasound image (315). This can be performed before, during, or after the introduction of GVs, but the frames taken prior to the introduction of the GVs might not have any value to the BURST process (but may have other use). Once the GVs are present and images frames are being captured, the ultrasound PPP can be rapidly increased to value over the collapse threshold of the GVs (320), which can be described as a step function change in PPP.

[00133] As the PPP is increased, the image frames continue to be captured. Any number of frames can be captured, but at a minimum three frames should be captured – one before the GVs collapse, one during the GVs collapse, and one after the GVs collapse. After the GVs collapse, the capturing of image frames can end (325). For each pixel of the image frames, a time-series vector can be extracted (330). Either all pixels of the frame can have time-series vectors extracted, or only those pixels within a region of interest within the frames can be represented by

time-series vectors. When the time-series vectors are found, signal separation can be performed on them using template vectors (335). Signal separation can be performed by any method, such as template projection or template unmixing.

[00134] Four mechanisms can contribute to the transient acoustic signal observed with loBURST and/or the much stronger transient signal observed with hiBURST: 1) the same linear scattering that creates contrast when imaging below the collapse threshold of the GV, 2) an acoustic wave generated by the rapid volume change that occurs during GV collapse, 3) stable cavitation of nanobubbles liberated from the GVs following collapse, and 4) inertial cavitation of liberated nanobubbles. In the case of (1), the signal strength is due to an increase in scattering amplitude in proportion to the higher pressures applied, while the signal transience is explained by the collapse of the GVs after the initial scattering event. For (3) and (4), signal transience would result from the sub-millisecond dissolution times of the nanobubbles. While these mechanisms are not mutually exclusive, their fundamental physical differences suggest the resulting signal amplitudes are likely to differ by orders of magnitude. Thus, the transient collapse signal from hiBURST or loBURST can be considered to be due predominately to a single mechanism, and for the dominant mechanism to differ between hiBURST and loBURST.

[00135] By imaging ARG-expressing *E. coli* in liquid buffer suspension at 10^5 cells/ml and recording the frequency spectra and temporal properties of the resulting BURST signal at various pressure levels, this difference can be shown. In order to achieve sufficient frequency resolution to discern higher harmonics from broadband enhancement, acquire data with a pulse sequence using 10 cycles at 5 MHz (see **Example 3** and **Figure 5, panels (a), (b), (e), (g)**) in addition to the standard BURST sequence using $\frac{1}{2}$ cycle at 6 MHz (see **Example 3, Figure 5 panels (c), (d), (f), (h)**), chosen to match the center frequency of our transducer and maximize axial resolution. In a small window of transitional pressure levels just below the hiBURST threshold, the signal tends to be a combination of elongated “bright” sources and a thinner band of lower intensity (see **Example 3** and **Figure 5, (d)-(e), top**). At lower pressures or concentration, it is apparent that the low-intensity band is composed of smaller, point-like “dim” sources (see **Example 3** and **Figure 5, panel (h)**). This indicates that the loBURST regime is characterized by a signal composed predominantly of dim sources, while the hiBURST regime signal is composed predominantly of bright sources.

[00136] It is observed that there are markedly different temporal properties for these two types of sources. Though both appear transient in the standard BURST pulse sequence with an inter-frame delay on the order of 10 msec, an ultrafast implementation of BURST with an inter-frame delay of 100 μ sec shows that many bright sources persist after several high-pressure transmits (see **Example 3** and **Figure 5, panel (e)**). In contrast, the band of dim sources always vanishes after the first high-pressure frame. Because mechanism (2) depends on an irreversible collapse of the GV shell, it can be ruled out as a cause of the bright sources. Though it is conceivable that the dim sources could result from cavitation of much smaller nanobubbles, this is unlikely because a sample preparation protocol ensures both that there are no free GVs present and that all ARG-expressing cells have similar numbers of GVs, so mechanisms (3) and (4) can be ruled out for the dim sources.

[00137] The loBURST mechanism can be narrowed down with the observation that the density of dim sources increases with pressure while their intensity remains relatively constant (see **Example 3** and **Figure 5, panel (h)**). If the sources were generated by mechanism (1), the opposite would be observed: there should be scattering from all cells in the field of view at an intensity that increases proportionally with incident pressure. Instead, the observations are consistent with a stochastic collapse model in which a given GV collapses with a probability proportional to the peak positive acoustic pressure. Therefore, the loBURST signal is the result of mechanism (2): an acoustic wave generated by the collapse of the GV shell and the resulting rapid displacement of fluid volume.

[00138] The hiBURST mechanism can be determined with the signal spectra. Below the hiBURST threshold with the 10-cycle pulse sequence, the spectrum is dominated by the fundamental and second harmonic peaks, which are both also observed in the post-collapse spectra (while all scattering occurs at the fundamental frequency in a linear medium, the intrinsic nonlinearity of water causes significant scattering at the second harmonic at elevated pressure levels). Above the hiBURST threshold, appearance of the bright sources is accompanied by both the emergence of higher harmonic peaks, a characteristic of stable cavitation, and a broadband enhancement in the power spectrum (see **Example 3** and **Figure 5, panel (c)**), a characteristic of inertial cavitation [2]. Based on the relative amplitude of the broadband and harmonic enhancements, mechanisms (3) and (4) both play a significant role in signal generated by

hiBURST with the 10-cycle pulse sequence. It is more difficult to assess the contribution of inertial cavitation to hiBURST signal with the standard $\frac{1}{2}$ cycle pulse sequence because, while a slight enhancement is observed across higher frequencies at pressures above the hiBURST threshold (see **Example 3** and **Figure 5, panel (a)**), there is not sufficient frequency resolution to distinguish harmonic enhancement from broadband enhancement. However, because the observed enhancement is weak relative to the 10-cycle case, and because inertial cavitation is typically generated by pulses with a large number of cycles, it can be concluded that the hiBURST signal involving few-cycle pulse sequences is predominantly generated by mechanism (3), with the possible presence of a limited amount of inertial cavitation.

[00139] In some embodiments, the imaging method herein described can further comprise delivering the GVs to the target site. Delivering the GVs to the target site can include using an acoustic reporter gene to express the GVs. The target site can comprise a mammalian cell with the acoustic reporter gene or a bacterial cell with the acoustic reporter gene.

[00140] In methods herein described, administering the contrast agent can be performed in any way suitable to deliver a GV to the target site to be imaged. In some embodiments, the contrast agent can be administered to the target site locally or systemically. The GVs can be delivered by the use of acoustic reporter gene (ARG) engineering.

[00141] The term “acoustic reporter gene” (or ARG) as used herein indicates genes used to express GVs in bacterial cells. The term “mammalian acoustic reporter gene” (or mARG) as used herein indicates genes used to express GVs in mammalian cells.

[00142] The wording “local administration” or “topic administration” as used herein indicates any route of administration by which a GV is brought in contact with the body of the individual, so that the resulting GV location in the body is topic (limited to a specific tissue, organ or other body part where the imaging is desired). Exemplary local administration routes include injection into a particular tissue by a needle, gavage into the gastrointestinal tract, and spreading a solution containing GVs on a skin surface.

[00143] The wording “systemic administration” as used herein indicates any route of administration by which a GV is brought in contact with the body of the individual, so that the

resulting GV location in the body is systemic (i.e. non limited to a specific tissue, organ or other body part where the imaging is desired). Systemic administration includes enteral and parenteral administration. Enteral administration is a systemic route of administration where the substance is given via the digestive tract, and includes but is not limited to oral administration, administration by gastric feeding tube, administration by duodenal feeding tube, gastrostomy, enteral nutrition, and rectal administration. Parenteral administration is a systemic route of administration where the substance is given by route other than the digestive tract and includes but is not limited to intravenous administration, intra-arterial administration, intramuscular administration, subcutaneous administration, intradermal, administration, intraperitoneal administration, and intravesical infusion.

[00144] Accordingly, in some embodiments of methods herein described, administering a contrast agent can be performed topically or systemically by intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, rectal, vaginal, and oral routes. In particular, a contrast agent can be administered by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e. g., oral mucosa, vaginal, rectal and intestinal mucosa, etc.) and can optionally be administered together with other biologically active agents. In some embodiments of methods herein described, administering a contrast agent can be performed by injecting the contrast agent into a subject at the target site of interest, such as in a body cavity or lumen. In some embodiments, it can be performed by spreading a solution containing the contrast agent on a region of the skin.

[00145] In some embodiments, the GV are provided by transforming cells within a target site with polynucleotide construct directed to deliver genes encoding for the GP proteins forming one or more gas vesicles type.

[00146] GV production in prokaryotes can be natural or engineered. An initial inquiry to determine if a given prokaryote will produce GVs is to determine if there is a gene cluster containing *gvpF* and *gvpN*. *gvpN* is not strictly needed for GV production, but GVs produced with *gvpN* typically have better acoustic properties (in the case of BURST, a stronger collapse signal). If there is such a gene cluster (determined, for example, by sequencing) and if the prokaryote contains *gvp A/B*, then the prokaryote will likely produce useful GVs (for BURST) if

those genes are expressed.

[00147] GVs can also be produced in mammalian cells through engineering (e.g. inserting gvps by means of a plasmid). The gvps for GV production in mammalian cells match those used for prokaryotes. For both prokaryotic and mammalian production, there are a number of permutations of gvps that can produce different GV types (GVs with different structural properties, such as shape, size, collapse threshold, etc.) with gvpF and gvp A/B being the conserved genes (and gvpN being an optional, but useful, gene).

[00148] In addition or in the alternative to detecting an acoustic collapse pressure for corresponding GV types, in exemplary embodiments where a GV type is to be used in the BURST (burst ultrasound reconstruction with signal templates) imaging described herein, the method of detection can be performed to further identify the a peak positive pressure (PPP) to be applied in connection with the specific GV type and can comprise imaging with ultrasound a target site comprising the cell following the introduction of the GVGC, over successive frames, at a peak positive pressure (PPP) well below the known or expected collapse threshold pressure for the GVs. While the frames are being taken, increasing the PPP step-wise to a value well over the expected collapse threshold pressure for at least 9 half-cycles. Frames from before, during, and after the application of the increased pressure undergo template unmixing to detect a BURST signal from the collapsing GVs, if present.

[00149] Further details concerning the BURST detection, and related methods and systems in accordance with the present disclosure will become more apparent hereinafter from the following detailed disclosure of examples by way of illustration only with reference to an experimental section.

EXAMPLES

[00150] The BURST imaging methods and systems herein disclosed are further illustrated in the following examples, which are provided by way of illustration and are not intended to be limiting.

[00151] In particular, the following examples illustrate exemplary methods and protocols for methods and systems to perform BURST imaging in accordance with the present disclosure. A

person skilled in the art will appreciate the applicability and the necessary modifications to adapt the features described in detail in the present section, to detection of additional gas vesicle structures and related genetic circuits, vectors, genetically engineered mammalian cells, compositions, methods and systems according to embodiments of the present disclosure.

Example 1: Amino acid sequences of exemplary GV proteins including GVS and GVA proteins

[00152] Several gvp genes and related proteins have been identified and are available in accessible databases.

[00153] In particular, **Table 2** shows amino acid sequences of exemplary GVS (gvp A/B or gvpC) and GVA proteins from several exemplary prokaryotic species. In particular, these exemplary amino acid sequences can be used as reference amino acid sequences in some embodiments for homology-based searches for related GVS and GVA proteins.

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
gvpA/B		
Ana-family-consensus_gvpA	MAVEKTNSSSSLAEVIDRILDKGIVIDAWVRVSLVGIELL AIEARXVIASVETYLKYAEAVGLTXSAAVPAX	8
Aphanizomenon-flos-aquae_gvpA	MAVEKTNSSSSLAEVIDRILDKGIVIDAWVRVSLVGIELL AIEARIVIASVETYLKYAEAVGLTQSAAVPA*	9
Aphanothece-halophytica_gvpA	MAVEKTNSSSSLGEVVDRLDKGVVVLDWVRVSLVGIE LLAVEARVVVASVETYLKYAEAVGLTSSAAVPAE*	10
Anabaena-flos-aquae_gvpA	MAVEKTNSSSSLAEVIDRILDKGIVIDAWVRVSLVGIELL AIEARIVIASVETYLKYAEAVGLTQSAAVPA*	11
Ancylobacter-aquaticus_gvpA	MAVEKINASSSLAEVVDRLDKGVVVDAWVRVSLVGIE LLAVEARVVVAGVDITYLKYEAVGLTASAQAA*	12
Aquabacter-spiritensis_gvpA	MAVEKINASSSLAEVVDRLDKGVVVDAWVRVSLVGIE	13

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	LLAVEARVVVAGVDITYLKYEAVGLTAGAQAA*	
Arthrospira-sp-PCC-8005_gvpA	MAVEKVNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LSVEARVVIASVETYLKYEAVGLTAQAAVPSV*	14
Calothrix-sp-strain-PCC-7601_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LAIEARIVIASVETYLKYEAVGLTQSAAVPA*	15
Dactylococcopsis-salina-PCC-8305_gvpA1	MAVEKTNSSSSLGEVDRILDKGVVVDLWVRVSLVGIE LLAVEARVVIASVETYLKYEAVGLTSSAAVPAE*	16
Dolichospermum-circinale-AWQC131C_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL AIEARIVIASVETYLKYEAVGLTQSAAVPA*	17
Dolichospermum-lemmermannii_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL AIEARIVIASVETYLKYEAVGLTQSAAVPA	18
Enhydrobacter-aerosaccus_gvpA1	MAVEKMNASSSLAEVDRILDKGIVDAWVRVSLVGIE LLAVEARVVVAGVDITYLKYEAVGLTAGAEAA*	19
Lyngbya-confervoides-BDU141951_gvpA	MAVEKVNSSSSLAEVDRILDKGIVDAWVRVSLVGIE LLAIEARVVIASVETYLKYEAVGLTAQAAVPAS*	20
Nostoc-punctiforme-PCC-73102_gvpA	MAVEKVNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LSIEARIVIASVETYLRYAEAVGLTSQAAVPSAA*	21
Nostoc-sp-PCC-7120_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LAIEARIVIASVETYLKYEAVGLTQSAAMPA*	22
Microchaete-diplosiphon_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LAIEARIVIASVETYLKYEAVGLTQSAAVPA*	23
Microcystis-aeruginosa-NIES-843_gvpA1	MAVEKTNSSSSLAVIDRILDKGIVDAWARVSLVGIEL AIEARVVIASVETYLKYEAVGLTQSAAVPA*	24
Microcystis-aeruginosa-NIES-843_gvpA2	MAVEKTNSSSSLAVIDRILDKGIVDAWARVSLVGIEL AIEARVVIASVETYLKYEAVGLTQSAAVPA*	25

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Microcystis-aeruginosa-NIES-843_gvpA3	MAVEKTNSSSSLAVIDRILDKGIVDAWARVSLVGIELL AIEARVVIASVETYLKYAEAVGLTQSAAVPA*	26
Microcystis-flos-aquae-TF09_gvpA	MAVEKTNSSSSLAVIDRILDKGIVDAWARVSLVGIELL AIEARVVIASVETYLKYAEAVGLTQSAAVPA*	27
Phormidium-tenue-NIES-30_gvpA	MAVEKVNSSSSLAEVVDRILDKGIVDAWVRVSLVGIEL LAIEARVVIASVDTYLKYAEAVGLTAQAAVPA*	28
Planktothrix-agardhii_gvpA	MAVEKVNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LSIEARIVIASVETYLKYAEAVGLTAQAAVPSV	29
Planktothrix-rubescens_gvpA	MAVEKVNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LSIEARIVIASVETYLKYAEAVGLTAQAAVPSV*	30
Pseudanabaena-galeata-PCC-6901_gvpA	MAVEKVNSSSSLAVIDRILDKGIVDAWVRVSLVGIEL LSIEARVVIASVETYLKYAEAVGLTASAAVPA	31
Stella-vacuolata_gvpA	MAVEKINASSSLAEVVDRLDKGVVVDWVRVSLVGIE LLAVEARVVVAGVDTYLKYAEAVGLTAGAQTAA*	32
Trichodesmium-erythraeum-IMS101_gvpA3	MAVEKVNSSSSLAVIDRILDKGVVVDWIRLSLVGIEL LTIEARIVVASVETYLKYAEAVGLTTLAAAPGEAAA*	33
Trichodesmium-erythraeum-IMS101_gvpA4	MAVEKVNSSSSLAVIDRILDKGVVVDWVRLSLVGIE LLTIEARIVIASVETYLKYAEAVGLTTLAAEPAA*	34
Tolypothrix-sp.-PCC-7601_gvpA1	MAVEKTNSSSSLAVIDRILDKGIVVDAWVRVSLVGIEL LAIEARIVIASVETYLKYAEAVGLTQSAAVPA*	35
Tolypothrix-sp.-PCC-7601_gvpA2	MAVEKTNSSSSLAVIDRILDKGIVVDAWVRVSLVGIEL LAIEARIVIASVETYLKYAEAVGLTQSAAVPA*	36
Halo-family-consensus_gvpA	MAQPDSSSLAEVLDRVLDKGVVVDVWARXSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEA- XPAPEA	37

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Halobacterium-salinarum_gvpA1	MAQPDSSGLAEVLDRVLDKGVVVDVWARVSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAAPEA	38
Halobacterium-salinarum_gvpA2	MAQPDSSSLAEVLDRVLDKGVVVDVWARISLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAPAP EA	39
Halobacterium-salinarum-NRC-1_gvpA1	MAQPDSSGLAEVLDRVLDKGVVVDVWARVSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAAPEA*	40
Halobacterium-salinarum-NRC-1_gvpA2	MAQPDSSSLAEVLDRVLDKGVVVDVWARISLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAPAP EA*	41
Haloferax-mediterranei-ATCC-33500_gvpA	MVQPDSSSLAEVLDRVLDKGVVVDVWARISLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAAPTPE A*	42
Halogeometricum-borinquense-DSM-11551_gvpA	MAQPDSSSLAEVLDRVLDKGVVVDVWARVSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTATAEAAPTPE A*	43
Halopenitus-persicus-strain-DC30_gvpA	MAQPDSSGLAEVLDRVLDKGVVVDVWARVSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAAPEA	44
Haloquadratum-walsbyi-C23_gvpA	MAQPDSSSLAEVLDRVLDKGIVVDTFARISLVGIEILTVE ARVVVASVDTFLHYAEEIAKIEQAELTAGAEA*	45
Halorubrum-vacuolatum-strain-DSM-8800_gvpA	MAQPDSSSLAEVLDRVLDKGVVVDVYARLSLVGIEILT VEARVVAASVDTFLHYAEEIAKIEQAELTAGAEAAPTPE A*	46
Halopiger-xanaduensis_gvpA1	MAQPQRRPDSSSLAEVLDRILDKGVVIDVWARISVVGIE LLTIEARVVVASVDTFLHYAEEIAKIEQATAEGDLEEELE LEVEPRPESSPQSAAE*	47
Natrialba-magadii-ATCC-43099_gvpA	MAQPQRRPDSSSLAEVLDRVLDKGVVIDIWARVSVVGI ELLTVEARVVVASVDTFLHYAEEIAKIEQATAEGDLEDL	48

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	EELEVEPRPESSPKSATE*	
Natrinema-pellirubrum-DSM-15624_gvpA1	MAQPQRRPDSSSLAEVLDRVLDKGVVIDVWARISVVGIELLTIEARVVVASVDITFLHYAEEIAKIEQATAEGDLDEL EELEVEPRPESSPKSAE*	49
Natronobacterium-gregoryi-SP2_gvpA1	MAQPQRRPDSSSLAEVLDRILDKGVVIDVWARVSVVGIELLTIEARVVVASVDITFLHYAEEIAKIEQATAEGDLEDL EELEVEPRPESSPQSATE*	50
Methanosaeta-thermophila_gvpA1	MVTSTPDSSSLAEVLDRILDKGIVVDVWARVSLVGIEILTVEARVVVASVDITFLHYSEEMAKIEQAIAAAPSAA*	51
Methanosaeta-thermophila_gvpA2	MVTSTPDSSSLAEVLDRILDKGIVVDVWARVSLVGIEILTVEARVVVASVDITFLHYSEEMAKIEQAIAAAPGVPA*	52
Methanosarcina-barkeri-3_gvpA1	MVSQSPDSSSLAEVLDRILDKGIVVDVWARVSLVGIEILAI EARVVVASVDITFLHYAEEITKIEIAAKEEKPAAIAA*	53
Methanosarcina-vacuolata_gvpA1	MVSQSPDSSSLAEVLDRILDKGIVVDTWARVSLVGIEILAI EARVVVASVDITFLHYAEEITKIEIAAREEKPVIAA*	54
Methanosarcina-vacuolata_gvpA2	MVSQSPDCSSSLAEVLDRILDKGIVVDTWARVSLVGIEILAI EARVVVASVDITFLHYAEEITKIEIAAREEKPVIAA*	55
Haladaptatus-paucihalophilus-DX253_gvpA	MVQAEPNSSSLADVLDRILDKGVVIDVWARISVVGIEVL TVEARVVVASVDITFLHYAKEMAKLERASSEDEIDFEQV EVASPEASTS*	56
Mega-family-consensus_gvpA	MSIQKSTXSSSLAEVIDRILDKGIVIDAFARVSXVGIEILTI EARVVIASVDITWLRYAEAVGLL-D-VEE-GLP-RX-	57
Bacillus-megaterium_gvpA	MSIQKSTDSSSLAEVIDRILDKGIVIDAFARVSLVGIEILTI EARVVIASVDITWLRYAEAVGLLTDKVEEEGLPGRTEER GAGLSF*	58
Bacillus-megaterium_gvpB	MSIQKSTNSSSLAEVIDRILDKGIVIDAFARVSVVGIEILTI EARVVIASVDITWLRYAEAVGLLRDDVEENGLPERSNSS EGQPRFSI*	59

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Serratia-family-consensus	MAKVQKSTDSSSLAEVVDRILDKGIVIDAWXKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTAXAAAPA*	60
Burkholderia-sp- Bp5365_gvpA1	MAKVQKSTDSSSLAEVVDRILDKGIVIDVWAKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTATAAAPT*	61
Desulfobacterium- vacuolatum-DSM-3385_gvpA	MAKVQKTTDSSSLAEVVDRILDKGIVVDAWAKISLVGI ELISIEARVVIASVETYLKYAEAIGLTAAAAAPA*	62
Desulfomonile-tiedjei-DSM- 6799_gvpA1	MAKIAKSTDSSSLAEVVDRILDKGIVIDAWAKVSLVGIE LLSVEARVVIASVETYLKYAEAIGLTASAAAPA*	63
Isosphaera-pallida-ATCC- 43644_gvpA1	MAKVTKSTDSSSLAEVVDRILDKGIVIDAFKVS LVGIE LLSVEARVVIASVETYLKYAEAIGLTASAATPA*	64
Lamprocystis-purpurea-DSM- 4197_gvpA1	MAKVANSTDSSSLAEVVDRILDKGIVIDAWIKVSLVGIE LLAIEARIVIASVETYLKYAEAIGLTAPAAAPA*	65
Lamprocystis-purpurea-DSM- 4197_gvpA2	MAKVANSTDSSSLAEVVDRILDKGIVIDAWLKVSLVGIE LLAVEARVVIASVETYLKYAEAIGLTAPAAAPA*	66
Legionella-drancourtii- LLAP12_gvpA1	MAKVQKSTDSSSLAEVIDRILDKGIVIDVWAKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTATASHPA*	67
Psychromonas- Ingrahamii_gvpA1	MANVQKTTDSSGLAEVIDRILDKGIVIDAFVKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTASAATPA*	68
Psychromonas- Ingrahamii_gvpA4	MANVQKSTDSSGLAEVVDRILEKGIVIDAFVKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTASAATPA*	69
Serratia-39006_gvpA1	MAKVQKSTDSSSLAEVVDRILDKGIVIDAWVKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTASAATPA*	70
Thiocapsa-rosea-strain-DSM- 235-Ga0242571-11_gvpA1	MAKVANSTDSSSLAEVVDRILDKGIVIDAWVKVSLVGIE LLAIEARVVIASVETYLKYAEAIGLTAPAAAPA*	71

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Other gvpAs		
Bradyrhizobium-oligotrophicum-S58_gvpA1	MAIEKATASSSLAEVIDRILDKGVVIDAFVRVSLVGIELLSIELRAVVASVETWLKYAEAIGLVAQMPA*	72
Desulfotomaculum-acetoxidans-DSM-771_gvpA1	MAVKHSVASSSLVEVIDRILEKGIVIDAWARVSLVGIELLAIEARVVVASVDTFLKYAEAIGLTKFAAVPA*	73
Octadecabacter-antarcticus-307_gvpA1	MAVNKMNSSSSSLAEVVDRLDKGVVIDAWVRVSLVGIELIAVEARVVIAGVD TYLKYAEAVGLTAEA*	74
Octadecabacter-arcticus-238_gvpA1	MAVSKMNSSSSSLAEVVDRLDKGVVIDAWVRVSLVGIELIAVEARVVIAGVD TYLKYAEAVGLTAEA*	75
Pelodictyon-luteolum-DSM-273_gvpA1	MAVEKTIGSSSLVEVIDRILDKGVVVDWVRMSLVGIELLAIEARVVVASVETYLKYAEAIGLTAKAA*	76
Pelodictyon-luteolum-DSM-273_gvpA2	MAVEKTIGSSSLVEVIDRILDKGVVVDWVRVSLVGIELLAIEARVVVASVETYLKYAEAIGLTAKAA*	77
Pelodictyon-phaeoclathratiforme_gvpA1	MSVEKTIGSSSLVEVIDRILDKGVVVDWVRVSLVGIELLAIEARVVVASVETYLKYAEAIGLTAKAA*	78
Rhodobacter-capsulatus-SB-1003_gvpA1	MAIEKSLASASIAEVIDRVLDKGIVVDAFVRISLVGIELLAIELRAVVASVETWLKYAEAIGLTVDPQTP*	79
Rhodobacter-sphaeroides_gvpA1	MAIEKSVASASIAEVIDRILDKGVVIDAFVRVSLVGIELIAIEVRAVVASIETWLKYAEAVGLTVDPATT*	80
gvpF		
Anabaena-flos-aquae_gvpF	MSIPLYLYGIFPNTIPETLELEGLDKQPVHSQVVDEF CFLYSEARQEKYLA SRRLLT HEKVLEQTMHAGFRVLLPLRFGLVVKDWETIMSQLINPHKDQLNQLFQKLAGKREVS IKIFWDAKAELQTMESHQDLKQQRDNMEGKKLSMEEVIQIGQLIEINLLARKQAVIEVFSQELNPFAQEIVVSDPMT EEMIYNAAFLIPWESESEFSERVEVIDQKFGDRLRIRYNN	81

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	FTAPYTFAQLDS*	
Ancylobacter aquaticus strain UV5_gvpF	MSATLSAPGTANVAVEATAAADGKYLYGIIIEAPAPATF DVPAIGGRGDVVHTIALGRLAAVVSNSPRIDYDNSRRN MLAHTKVLEAVMARHTLLPVCFGTVGSDAEVIIIEKILRE RRDELAGLLGQMHRMELGLKASWREEIIFEEVLAENP AIRKLRLDALVGRSPDQSHYERIQGERIGQALQRKRQDD EERILERVPRPFVHKTRLNKLIGDRMVINA AFLVDAAVES RLDASIRAMDEEWGGRLAFKYVGPVPPYNFVTITIH*	82
Aphanizomenon flos-aquae NIES-81_gvpF	MNTGLYLYGIFPDPIPETVDLQGLDKQSVHSQVVDGFSF LYSDACQEKYLASRRNLLTHEKVLEQAMHEGFHVLLPL RFGLVVKDWETIQKQLIEPYKEQLNELFQKL AGQREVSI KILWDSKSELQAMMESNQDLKQQRDNMEGKKLKMEEI IQIGQLIESNLAARKQTVIQEFFNNLHPLAKEIIESEPMT EMIYNAAFIPWETESVFSERVEAIDRKFGDRLRIRYNNF TAPYTFAQLAS*	83
Aphanothece halophytica (strain PCC 7418)_gvpF	MAEGFYLYGIFPPPGPQTIAVQGLDKQPIFSHTVEGFTFL YSEAQQSRYLASRRNLITHTKVLEEAMEQGFRLLPLQF GLVVPDWESVSQDLLQHQS ETLQLLFQRLEGKREVSLKI YWETDAELNALLEENPDLKARRDNLEGKNLSMDEV IQI GQALEQAMERRKQEVITRFEDALIPFAVETQENDVLTET MIYNTAFLIPWESEPEFGEAVETVDAEFAPRLKIRYNNFT PPYNFVELRE*	84
Aquabacter spiritensis strain DSM 9035_gvpF	MMQTDTLAPAETVAEGKYLYCLIDAPAPDTFASPGIGG RGDVVHTITVGRLLAAVVS DSPRIEYENSRRNMMAHTKV LEEVMARHTMLPVCFGTVATGPDPISGKILEGRRDEL VG LLEQMRGRLELGLKATWREDVIFAEILQENPAIAKL RDS LVGRSPEKSHFERIRLGEMIGQAMERKRRDDEERILERV RPFVHKTKLNKPIGDRMILNAAVLVEAAREAGLDQAVR QMDAEW GARLSFKYVGPVPPYNFVTITIH*	85
Bacillus-megaterium_gvpF	MSETNETGIYIFSAIQTDKDEEFGAVEVEGTKAETFLIRY KDAAMVAAEVP MKIYHPNRQNLLMHQNAVA AIMDKN DTVIPISFGNVFKSKEDVKV LLENLYPQFEKLFPAIKGKI EVGLKVIGKKEWLEKKVNENPELEKVSASVKGKSEAA GYIERIQLGGMAQKMFTSLQKEVKTDVFSPL EEAEEAA KAN EPTGETMLLNASFLINREDEAKFDEKVNEAHENWK DKADFHYS GPWPAYNFVNIRLKVEEK*	86

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Bradyrhizobium oligotrophicum S58_gvpF	MSNQPIYVYGLIRAEDHQPLAVRAVGDSEQPVNIIGSGN VAALVSTIDLPEIMPTRRHMLAHTKVLEAAMANGPVL MRFGIIVPNPATLLRVIGFRHQELRARLDEIDGRIEVALK ASWDEQFMWRQLASEHPDLAVSGRTMMGRGEQQSY DRIELGRAIGAALAEERRTAARLQLLQTVTPFAVQVKELT PVDDAMFAHLALLVEKGAEPSLYQTVEALERSNDSGLK FRYVAPIPPYNFVAVTLDWEQHEQAPRR*	87
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpF	MNSRNGARYLYAVQHARDVPASLPAGIGGA AVRALTD GDVAAIVSDTGLAKVRPERRHLLAHHTVIQSLAAAGTV LPVAFGTIATSEVALRRMLRKHRNALAGELARLVDHVE MSVRLNWDVTDLFRHLIDVRPDLKAARDAMLALGSAV TRDDKIELGSRFERVLNEERARHAALVDEALDACCKEIR RDPPRHETEILHLTCLVRHAELGRFESGVAAASRELDDS LVLKYSGPCPPHHFVNLMNSL*	88
Chlorobium luteolum DSM 273_gvpF1	MERDGKYYICIIGADCECDFGPIGIGGRGDLVSTIGFEGIS MVVSDHPLNRFVVDPDGILAHQRVIEAVMKEHESVIPV RFGTVAATPDEIRNLLDRRYGELSELLRLRNKVEFNVT GRWHDMAAIYKEVERTHPEIKEQRARIESMRDGDGEAL KQSLILDGTGHQIEAALEVMKEEKFDASLFRKTAMAS KMNRTTSPDMFMNA AFLIDRGREVEFDGIMEILGQKDA DRCDYRYSGPLAIFNFVDLRILPEKWEL*	89
Chlorobium luteolum DSM 273_gvpF2	MAHEAAEQDGLYIYGIINNSGELDFGPIGIGGREERVYA VIHNDIAAVVSRTVVKEFEP RRANMIAHQKVLEAVMVS HAVLPVRFSTVSPGHDDMKVEKILEEDYLRLKLLVKM EGKKEMGLKVMANEEKVYESIITGYDNIRYL RDKLINLP PEKTHYQRVKIGELVAAALEKEVGTYKDAVLDALSPIA EEVKVND SYGSMMVLNAAFLIRTAREEEFDRAVNALD DRYHDMMTFKYVGTLPPYNFVNISINIKGR*	90
Chlorobium luteolum DSM 273_gvpF3	MNQSIYIYGIVNEPALAASFVETDPDIYAVASMGCSAIV ENRPAIDL GELDRESLARMLLQHQQTLERLMESGMQLIP LKLGT FVSSAADAACIIEDGYNLIERIFRETEDAHELEV VKWSSFADLLQEVVSEGDVQELKREVEARQSSSTEDAI AVGRLIKEKIDRRNAALSASVLRQLGERASQSKRHETM DDEMVLNAAFLVNRGDVDAFVATVEALDSQYLNALHF RIVGPLPCYSFYTLEVTALFEFIAEKRAVLGLDARSCEA DVKKAYHAKAKVAHPDVHVPAGANN GADFTVLNEAY	91

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	MTLHDYYSSALRNSASSRHGHEGQDSSSVVFSVKILN*	
Dactylococcopsis salina PCC 8305_gvpF	MTEGFYLYGIFPPPGPKTIETQGLDKQPIFSHTVEGFTFL YSEAQQSRYLASRRNLITHTKVLEEAMENGSRITLLPLQF GLIVPDWETVVQDLLQHQAESLHFFLEKLEKREVSLEKI YWETNAELNALLEENPALKARRDNLEGGKQLSMDEVIQI GQALEQEMEGRKQDIISRFEEVLIPFAFEIKENDVLTETM IYNDAFLINWDAESDFGEQLEAIDAIEFSPRLKIRYNNFTP PYNFVELRE*	92
Desulfobacterium vacuolatum_DSM 3385_gvpF	MSKKNLKRNGRYLYAIIIEASEEKTFGSIGMDGSDVYLIV EDKTAADVSDVPNKKIRPQRKNIAAHHAHLNKIMEEITP LPMAFGIIADGEQAIRKILADNRDVFREQFATVSGKVEM GMRISYDVPNIFEYFISTDSEIRAARDQYFGGNREPSQEA KLELGRMFNRQLNANREEYTNQVIEILDDYCDIENK CRNEQEVTSACLINRSDQKRFEFGVFESARHFDNNFSF EYNGPWSPHNFVNILIEL*	93
Desulfomonile tiedjei DSM 6799_gvpF	MEKATIKTTGSGNGRYLYAVVPGSQERVYGCINGGNGV YTIAAKDVAAVSDVPHQKIRPERRHFAAHQAVLKRV MLDGDLLPMSFGIISQGPKA VRAILSRNNKSVQQQLKRI SGKAEMGIKVTWDVPNIFEYFIDVNRELREARNKL VQP NYLPTQQEKIEIGRMFEEILNLERERHTKQVERVMSKRC SEIKRSKCRTEIEVMNLSCLVDRITLLSDFEAGVLEAASH FDDSFADFNGPWAPHNFVDLEIDV*	94
Desulfotomaculum acetoxidans_DSM 771_gvpF1	MSTGRYVYCVINSIEPLTFMSGPVGNEPEGVFTVHYKEL AAVVSQSSEKYNVCRENTIAHQKVLEEVLVSHPLL PV RFGTVAQNEEIVKKFLLQERYAELRSMLHNVTGKVQM GLKVLWTD MKTVYQEIVEENPQIKNLKKKLESKPAETI HYEMIDLGMVNQALLRKKEKQKEMVLKPLQKIALET KESFLYGDQMFVNADFLISRSSLDDFNAKVNELGEFFNE QALFKYIGPLPPYNFVTLTYVNF*	95
Desulfotomaculum acetoxidans_DSM 771_gvpF2	MVKNHNTDHLKELIYGLIGGTPFKDELEKISVIQENTPI YGVWHKNIGFAVSAAPDYPLKDL SKESIIQLFVDHQQV LECLRQKFS LIPVKLGTVLESVTEAAAVLANNEEKFN DL LNYLKDKVELNLSVSWNDLNEVVAKIGEEDVKKLKQ SLLAQEQVSQEDLIKIGKIISFQMQQKKQAAREYIISLR NLWEDYFINEVVDENSILNLTLLAITGKVDDVNKKIEYL NQIYRDSLDFSLTKSLLPQGFSTVSIKKITMDQLLLAKDI	96

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	LKL PDTASLQDINAARRALLHCYHPDKNDHAAVNKVQ EINAAAYKLLECYQENSSDFNVDLITDYYIMKVIKADKS NVNSMNME*	
Dolichospermum circinale_gvpF	MNTDLAHKNFGLYLYGIFPDITPETLEIKGLDGKSVHSQ VVDGFTFLYSQACQEKYLASRRNLLAHERVLEQTMHE GFHVLLPLRFGLVVKDWETIMSQLINPHKEQLHKLFEKL AGQREVSIKILWDAKAEQAMMESNHDLRQQRDNMEG KKLSMEEVIQIGQLIESNLQARKQAVIEVFTRELNPLAQE IVVSEPMTEEMIYNAAFLIPWDSEPLFSERVESIDQKFGN RLRIRYNNFTAPYTFALLDS*	97
Enhydrobacter aerosaccus strain ATCC 27094_gvpF	MNPPEAYIAGRATAAKSVEDRKARPQDLAEGKYVYAIIA CDEPREFKNRGIGERGDKVHTINHRQMAAVVSDSPTID YERSRRNMMAHTVVLEEVMEKFDLLPLRFGTVASSAES VERQLLVPRYGEISAMLEKMRGRSEFGLKAFWHEGVA FGEIVRENARVRKLRDALQGRSLEESYYQRIQLGEEVEK ALTAIRARDEELILSRLRPFMRDIRTNKIISDRMVLNAAF LVERGDVPALDEAIRQLDQEFSERLMFKYVGPVPPYNF VNIAINWER*	98
Isosphaera pallida_ATCC- 43644_gvpF	MRNAPPTRPGSVTPASPGKPVIDGPARYLYAFTHDLPEG PLADLEGLPGARVVVVADGRVAAVVSPCPLGKVRPER QRVAGHHHVLKHLQDTLGKAILPASFGMVADSEEDLR ALLRHHSAAIAEGLVRVQGKVEMTVKLRWAPDNVAQA VLGRDPELRQLRDQLYSNGQTPTRDQSLDLGRRFHHL ERQRDHYAAYLRAALSPLLSELVEEDLRDERDLVHWA CLIENQRRAGFEAALDRLAEELEDDLVELETPWPVPHHF VDLDLDDDDHDDDEEE*	99
Legionella drancourtii LLAP12_gvpF	MDSTSKKPAASNLYLYAIASVNENQEPISFHGIEEQPIDL VPYKDIMLVVSNLSKKKVRPERKNVAVHHAVLNHLMK HNTSMLPIRFGMIADNRKEVQRLLTINYDMLHTKLKMM AGR VEMGVSLSWDVPNIFEYLLNRHSQRETRDKLLAN PAHEPSRDEKIEIGALFSQILDEEREVYTDITLSLLSPVCC DVVKSTYRNDTEIMNIFCLISAARRDEFEEKIIEASTILDD NFVIKYTGWPVPHNFSKLNLSLE*	100
Lyngbya confervoides BDU141951_gvpF	MPQLLYLYGIFPAPGPQDLEVQGLDQQPIHTHIIDEFVFL YSVAQQERYLASRKNLLGHERVLEAAMKVGYRTLLPL QFGLIETWDRVIKELITPRGDALKRLFAKLEGRREVSVK	101

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	LLWGPDAELNQLMEEDAGLRAERDRLEGQQLSMDQIV DIGQAIETAMTERKDDVINAFRQRLNALAIEVLENDPLT DAMIYNTAYLIPWEDEVKFSQAIEELDEQFEDRLRIRYN NFTAPYNFAQLDQLS*	
Microcystis aeruginosa NIES-843_gvpF	MTVGLYLYGIFPEPVPDGLVLQIDNEPVHSEMIEGFSFL YSAAHKEKYLASRRYLICHEKVLETVMAGFTTLLPLRF GLVIKTWESVTEQLISPYKTQLKELFAKLSGQREVSIF WDNQWELQAALSNPKLKQERDAMMGKLNLMEEIIHI GQLIEATVLQRKQDIIQVFRDQLNHRAQEVIESDPMTDD MIYNAAYLIPWEQEPEFSQNVEAIDQQFGDRLRIRYNNL TAPYTFAQLV*	102
Nostoc punctiforme ATCC 29133_gvpF	MSFYIYGILTLPAPQNLNLEGLDRQPQIKILDDFAVIYS EAQQERYLASRRNLLSHEKVLEEIMQAGDRYLLPVQFG LLVSSWETVSQQLIRPHQEELTQLLAKLSGCREVSVKVF WDTEAEIQGLLAEHPNLKTERDKLVGQPLSMERVIQIGQ VIEQGMSDRKQGIIDVFKGTLNSIAIEVVENTPQVDTMIY NSAYLIPWEAESQFSEHVESLDRQFENRLRIRYNNFTAP YNFARLRLTTSN*	103
Nostoc sp. PCC 7120_gvpF	MSSGLYLYGIFDPPIPETVTLQGLDSQLVYSQIIDGFTFLY SEAKQEKYLASRRNLISHEKVLEQAMHAGFRTLLPLRF GLVVKNWETVVTQLLQPYKAQLRELQKLGRREVS KIFWDSKAELQAMMDSHQDLKQKRDQMEGKALSME VIHIGQLIESNLLSRKESIIQVFFDELKPLADEVIESDPMT DMIYNAFLIPWENESIFSQQVESIDHKFDERLRIRYNNF TAPYTFAQIS*	104
Octadecabacter antarcticus 307_gvpF1	MKREVVVRMTDENTINSKYLIAIKCREQREFIARGIGER GDAVHTIAYKGLAAVVS DSPVMEYDQSRRNMMAHTA VLEELMEEFTLLPVRFN TVAPEAGAIEERLLVPRHEFT QLLGQIDKRVELGIKAFWHDGMIFEEVLREND SIRKMR DALEGKSVDGSYYERIQLGEKIEQAMIKRVEDEEILSR IRQHVHKSRSNKTIGDRMVLNGAFLVDANKESDFDKAV QLLDQDLGNRLMFKYVGPVPPYNFVNIVVNWGVV*	105
Octadecabacter antarcticus 307_gvpF2	MTVVAEENMTGSGVGLYVCAIVA EWESNSALIKCANE AQGEIQLIGQG GITAVVMVPPEQPVSRDRQELVRQLL VHQQQLVERFTEIAPVLPVKFGTLAPDRESVELGLERG REKFTAFGGLSGKTQFEITVTWADVFAKIAKLPAVVKLK	106

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	VDLVATSESDRPINLDRVGRLVKETLDHQRAQTGKVLL DALLPLGVDSIVNPILNDSIVLNLALLVDTDQADALDRC LDELSTFHGALSFRFCVGPMPPHSFATVEINYIEPTQVSH ACCVLELDAAHNFEEIRSAYHRLARQTQQDIAPDVVVD NKSSSVGIAVLNDAYKTLLSFVDAGGPVVVSVQRQEDA YATDIPSSGG*	
Octadecabacter arcticus 238_gvpF1	MTDEKKVNSKYLYAIIQCREPRELKARGIGERGDVVHT VVHKGLAAVVS DSPVMEYDQSRRNMMMAHTAVLEELM EEFTLLPVRFNTVAPEAVAIEERLLVPRHDEFTQLLGQID KRVELGLKAFWHDGMIFGEVLREND SIRKMRDSLKGQS VDGSYYERIQLGKIEKALTEKRLEDEEMILSRIRPHVHK SRSNKTIGDRMVLNGAFLVDAEKESKFDEAVQSLDQDL SDRLMFKYVGPVPPYNFVNIVVNWGES*	107
Octadecabacter arcticus 238_gvpF2	MRAQKVIPAAEENISGNVGLYVCAIVAERVSCSALIQCA NDAPGEIQLIGHGDFTA VVMVPEKDQLVSPDRKELMQQ LLVHQQLEKFMEIAPVLPVKFATLAPNRESVELGLEVG SEKFSAAFNSLSGKVQFEVIVTWDVAEVFAEIAKEPAVA KLKVDLAAMPESYGSVSLEQLGKLVKETLELRRRAETGK VLLDALVQVGVDNVNSILDDSIILNLALLVEAKRADAF DRCLDEL DSTYHGALTFRFCVGPLPPHSFATVEITYLEPA KVTEACDILELDVARSTEEVRSAYHRLARKSHPDIVPDV AVGETASVSM AVLTDAYKTLLSFVGAGGSVVVSVQRQ EASYAADISSAG*	108
Pelodictyon phaeoclathratiforme_gvpF1	MDIETTKEGRYIYGIIRNSEFIDFGQIGIGKRNDRVYGVY KDICAVVSSTPIIQYEARRANMIAHQKVLEEV MKRFNVL PVRFTISPHDNDDAIKILITDYSRFD ELLIKMKGKKELG LKVMADETRIYENIIQKYDNIRSLRDKLLNQPADKIH YQ RVKIGEMVADALKKEIESYKQQLDILSPIAEDIKITDNY GNLMILNAAFLIKEVKESEFDDSVNKLDEKYGNIMTFK YVGTLPPYNFVNLSINTKGV*	109
Pelodictyon phaeoclathratiforme_gvpF2	MEKDGYVYCIIASTYECNFGAIGIGGRGDLVNTIGFQG LSMVVSDHPLNHFVLNPDNILAHQRVIEVVMSQFNSVIP VRFGTVAATPDEIRNLLDRRYGELSELLERFENKVEYNL KASWRCMIDIYKEIDKEHVELKQLRREIEGLKDEEKRL LIVEAGHIIENELQKKKEVEAYEIVTYLRKTVVAHKHNK TTGEAMFMNTAFLNKGREVEFDNIMNDLGEQYKDRS DYYYTGPLPIFNFDLRILPEKWEL*	110

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Pelodictyon phaeoclathratiforme_gvpF3	MDRQGIYTYGFIPNHYLTDIKTILIESGIYSIEYGSIAALVS DTMVDDIEYLNREDLAYLLVDHQKKIELIMSTGCSTIIP MQLGTIVNSGNDVIKIVKNGLRIINKTFDDIADIQEFDLV VMWNNFPDLIKKISDTPQIRIMKEEIANKGSYDQADSINI GKIIKKKIDEKNSKVNLDIMNSLSSLCICVKKHESMNDE MPLNSAFLIKKDKENSFIEMVNQLDIKYENLLRYKIVGP LPCYSFYTLESKLLNKKEIEKAEKILGIDAYKSESDIKKA YRAKAAHAHPDKNNTISAIDNDDFIEINKAYQILLEYSS VFKDSPDHKPDEPFYL VKIKK*	111
Phormidium tenue NIES-30_gvpF	MADRYLYYGIFPAPGPAELPLMGLDEQVVQAQQLGDFT FLYSLACQKRYLSSRNLLGHEKVLEAAMEQGHRTLLP LQFGLIVESWNQVQEDLVTPYAEDLTQLFGRLNGCREV SIKVQWEPSTELEMMMAENADLRAQRDQLEGTLGME QVIFIGQQIESALEERKQGIVDQFRQALSPLAKDVLENAP QTDVMIYNAAFLIPWESEAEFSQAVDAIDSTFGDRLRIR YNNFTAPYNFAQLN*	112
Planktothrix agardhii str. 7805_gvpF	MGNGLYLYGILPTNRVRPLALHGLDKQPIQTHPVDEFSS LYSETQQERYLASRRNLLGHEDVLEKVMQHGYSVLPL QFGLIVKDWDHVKAQLIIPYQDRLKELFKLEGKREVG VKIFWEETEELDLLMTENQELREKRDSEKRLSMDEII GIGQEIERAMQDRQQGIIDKFQQILNPLAQEIVENDNLT SAMIYNAAYLIPWDIEPQFGDKIEELDHHFNNRLRIRYNN FTAPFNFAQLNP*	113
Psychromonas ingrahamii 37_gvpF	MAENKKKVRKSSSKVIAKPKVIYAITAGGLQDLGNLVG INKSDIYTIEKESISFVSDLSPPSPRPRPDRRNIMAHNEIL KQLMSKTSVLPVRFGTATGERAVNRFCSQYNAQLLEQ LDRVQDRVEMGIKVTWNPNIYDYFVDNHSELREERD RVYDGNKNPRRDDRLNLGHMYDALVTEARLSHQTDLE EILPGCDEIHSIPPKDEKVVVNLACL VQRADLEVFEERV VEAGKTLDNITYDIELNGPWAPHNFVELDLKTMGTGR*	114
Serratia sp. ATCC 39006_gvpF	MMSIDKSRNHRKVLVYALCVSDDSTPNYKIRGLEAAPV YSIDQDGLRAVVSDDLSTRLRPERRNITAHQAVLHKLTE EGTVLPMRFGVIARNAEAVKNLLVANQDTIREHFERLD GCVEMGLRVSWDVTNIYEYFVATYPVLSETRDEIWNGN SNANNHREEKIRLGNLYESLRSGDRKESTEKVKVLLD YCEEIENPVKKEKDVMLNACLVARERMDEFKGVFEA	115

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	SKLFDNVYLFDYTGWPAPHNFVTLDLHAPTAKKKTLTRAGT LSD*	
Stella vacuolata_ATCC-43931_gvpF	MQTEALAPAAVAAEGKYLYCIIDAPAPATFASPGIGGRGDVVHTLAVGRLAAVVS DTPRIEYENSRRNMMAHTKVL EEVMAHHTLLPVCFGTVGSGDDVIAEKILEGRREELSRL LEEMRGRVELGLKATWREEVIFAEVLDEDP AVRKL RDS LVGRSPEKSHFERIRLGELIGQALLRKR RDEEERILDRVR PFVRKTKL NKPIGDRMILNAAFLVETAREAAALDQSVRE MDADWGARLSFKYVGPVPPYNFVTITIHW*	116
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpF	MQQAKRQDVAAGRYIYAIIPDRGDHSLGRIGLDESEVYTIGDGRVAAVVS DLSGGRIRPQRRNMAAHQEVLKQVLR EVSPLPAAFGLMADDEAAIIRILKDNQDAFLNQLERVDG SLEMGLRMSWDVPNIFEYFVG AHP ELQELRDDFFRDGS NLTDQDQMITLGRSFERLLEQDREEYTEQVESVMRSCCR EIKRNC RTEKEVLHLACLVD RDAAGRFEQVVLQAARP FDN NYAFDFNGPWAPHNFVEMDIHV*	117
Tolypothrix sp. PCC 7601_gvpF	MDAGLYLYGIFSDPIPTVSLKGLDSQPVSQVIEGFTFLYSDAKQEKYLASRRNLISHEKVLEQAMQEGFRTLLPLR FGLVVKNWETVISQLIQPCERQLRDLFQKLAGKREVS V KILWDTKAELQAMMQSNPDLKQKRDQMEGKNLSMEE VIEIGQLIESNLQQRKEAVIKTFFDELKPLAEEVVESEPM MEEMIYNAAFLIPWDQEALFSQRVEAIDKKFGDRLRIRY NNFTAPYTFAQIS*	118
Trichodesmium erythraeum IMS101_gvpF	MEFGFYVYGLIQEKGKMDESKDESKNGLKGSNESKDELKGLDKEDVKIQDVDEF AVL YSIKKERYLASRRNLITHE KVLESAMEAGYRNLLPMQFGLVVSEWEKFSQDFTKPCE QQIHDLFTKLKNNREVG I KIY WEPDAELEKLENDKDL KEERDSLKDKKL TMDQVIDIGQKIEQGMNERKQNIIEIF QETLNKMAIEVIENEVQTEKMIYNAAYLIPWDQEEDFG EK VETIDSKLCERGNFTIRYNSFTAPYNFARIRQQD*	119
gvpF/L		
Ancylobacter aquaticus strain UV5_gvpFL1	MTDLLVFAVVPADRFDPAILAEGDGLPPGLRAIAAGPLA AVVGAAPEGGLKGRERSALLPWLLASQKVMERLLANA PVL PVALGT VVEDEGRVRHMLDAGAAILGEGFQAVGD GIEMNLSVLWHLDTVVARLLPGVAPELRQAAAGGDAIE	120

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	RQALGVVL AGLVSAERRRARARVIEALQAVTRDFAIGE PTEPGGVN LALLVDRAAEEALGAALDAEFDGALT FRLVGPLPPYSFASVQVHLSPAAAVCGARAALGVEPDA SPETVKAAYRRAARETHPDLVPMGGEDEEAEATADET SRFVVLSDAYRVLEGEHAPVSLRRLDSVLTE*	
Ancylobacter aquaticus strain UV5_gvpFL2	MLYVYAITADYAAGANHLLPAKGIVPGVPVQRFGTGAL GAVASPVPTVFGKEALHALLDDADWTRARILAHQRV VSSLLPLATVLPKFGTLVAGEASLAAALTSQHDALDAT VARLRGAREWGVKLFEEAPTRTIRAEPPVGAGAGLAFF RRKKEEQETRAAAEAALDRCVAASHRRLASHARAABA NPLQPPELHGHPGTMGLNGAYLVAAENEAARVCFSE LEQAYAALGARYVRTGPWAAYNFTGGGLV*	121
Aquabacter spiritensis strain DSM 9035_gvpFL1	MSGLLVFAIVPADRIEPLLAPAEGLPPGLETVVAAGFA AIVGTAPEGGLKGRDRGSLLPWLLASQKVIERLMARGP VLPAAALGSVLEDESRRHMLVCGQAALAAAFETLNGC WQTDLSVRWDL SRTVAHLMTELPPGLRAAAETGDETA RRSLGAALAGLVAGERRRIQSRIGAVLGAVARDLIVSDP VEPEGVVGVALLDAPASAQVDAALDRLDGEFEGRLTF RLVGPLAPYSFATVQIHLGPAAGLAGAHAELGLEAGAP LEAVKAAHYRLIVGLHPDLVPHGSPGDDADDAASGKG GRAARFAAVTAAYRTLQAEHAPVSLRRQDGLSPG*	122
Aquabacter spiritensis strain DSM 9035_gvpFL2	MLYVYAITADHPGPHDAGSLPGEIVPGAPVRLLPFGDL AAAVSPVSAVDFGPEALPARLQDQVDWTGQRLAHQRV VDSLVDVATVLPKMFCTLFSGAAALRAALADNRAALE ATVVRLRGAREWGVKLFWEAPPAEPAPVERGPGAGAA FFQRKRDAQRLRAEAEALAHGVAESHRRLAARARAA VANPVQPAAVHRRRGEMALNGAYLVPRADEAAWRES LAELERTYAGAGIRYELTGPWGPYNFTGGGLAGS*	123
Bradyrhizobium oligotrophicum S58_gvpFL1	MTMNLVGITTPDVAGAI AAAGGRLADVETRAVEAGGL VALLALSKAPFWHLRRSRTALRSMLTAQRILEAAAVY GPLLPARPGTLIRNDAEACMLLR SQCRHLAEGRLRHGTS RQYQITISWDPVAALAARRDHQDLVEAAAASADGAAD KAASMIQRFMSDQQARFEAEAMRALAAVAEDVITLPV NQPDMLMNAVLLAPGAEP ELERVLEALDRGLRGKNLI RLIGPLPPVSFAAVSIERPGRQRIAAARRLLGIGEA TRTC DLRRAYLDKAHAHHPDTG GHAADASIVGAAAEAFRL ARVAEARASAGQDDVILVDIRRDQQRSLST*	124

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Bradyrhizobium oligotrophicum S58_gvpFL2	MSKANLGIGLVHGVVTAQSAALLPQIVDAFDATETIIVVN TEQQALLISDIPQYLRGHVEADTLFSDPARISTLAMKHH RILQAAAVVTDVVPVRLGTLVRGPGSGARDLLNREAVRF AGHLVTIHNALEFSVRILPTEQPSRRVARVPSSGRDYLR IRRDERCGQRPVVDITLQELASRAVAIRERQSASRSGG RTPALAEAAFLVDRHALAAFDDCAGRIERQIAENGLAL DIFGPWPAYSFVDGARENLG*	125
Bradyrhizobium oligotrophicum S58_gvpFL3	MSSPRLIGLLAADDVPADLADQIMSCGPVAAAIRFAPAA ASSSESLDHHAADVAVWCRAAFLPSRAGIPISPELLQSI RSAWYHRSTIEHIEGRVEISVELERRDGV RDGGIDGGGR AYLRATAHDLRACEVGVATAANLLAMYSERADADLIA RTAPLPAIRLRASVLVRRVAPRLARQFDSMLS AISDRL VCRVTGPWPYPYSFSTIREPS*	126
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpFL	MVWLTYAVLTPKRSITLPPGVAGARLEIVDGAHLRTIVS EHPRAPSATIPSALDFGQTVAALFRHGAI VPMRFPTCLDS KQAVRDWLDDSDMYRDLLQRIDGCVEMGLRFRLEA PRAQPRQAGGPGHAYLAARGAPNSVARSHGERIAAVL RNL YRDWRFDGLVEGFVSLSFLVRQTTLDDFVDRCRQA ARETAFPLYMSGPWPPYSFATDERSSAPEPHRALRLMR RPSTAVSISANVA APEKKDSAR*	127
Desulfobacterium vacuolatum-DSM 3385_gvpFL	MTLHLLYCVFSSGEMEKTRKLVPPGIDGEPVHEICSNKIS GVVSTL GKPPDTHVKSLLAYHGVIDSYHQNRTVIPMRF AAVFRTYAHMITALNNNEKSYLLQLKRLHDCTEMCVR FISNSPCCVKKKEPAISP KKISGTTFLQQRKAMYEQQNR LPPEIHEKTRDILQHFRGLYMEFKQESQPLEKDCPSLSLQ GAEKTDGNALLISLFFLISKKNISLFRSRFQNICGSSSGRH MMNGPWPPFNFINTESNLTDPS*	128
Desulfomonile tiedjei DSM 6799_gvpFL	MLGSLAAIQFLSISSYGADMKFLMYCIFTENSIEPPHSL VGVNRSPVRIISCDGLAAAVSVITQKEIPRDPATGLDYH KVIQWFHERIGVIPLRLGTCLGHESDVVQLLHSHGARYK SLLKELDGCVEMGIRVIHDRPGPQELASKSPFISRFNGTE SGTDYLMRRKVLFD ADEFAISR NREIVERYHSPFTGLYV SFKAQTSKFSPLGTDNRNSVL TSLYFLIPRQSADSFRAIYG DLRSGLHERIMLSGPWPYPYNFVLPEDCL*	129
Enhydrobacter aerosaccus	MEGHRIYIYGIVRDAADGGPAPVPPVAGLDGGALRAIA	130

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
strain ATCC 27094_gvpFL	GYGLAAIASAVDLSKAGIPFEEQLKDPDRATALVLEHHR VLQQAIDAQTVLPMRFGALFQDDRGVTDALKNRCGL MDALGRIDGAREWGVKIFCDRAVAARQLSATSASVVQA AEKELSGLAEGRAFFLRRRLERLRTEETDRAVAHEVDV SRQALCELARASAPLKLQPAAVHGRGEDMVWNGAFLV PRSGEERFLSRLEV VVQSRSDLGLHYEVTGPWPPFSFVD GQLEGGGDACPDGA*	
Octadecabacter antarcticus 307_gvpFL	MRSATSIVYAYGVL TNCS DIALDMPRSDLAGLVKNGPL RILPFGNIAAVVCD FVLPNGSDLETLL EDSRSAERLILNH HQVLSYIVSQHTILPLRFGAAFTEDAGVIAALGGRCSEL QKALGRIDGALEWGVKTFCDRKLLKQVRGTGSEISDL ESEIAKQGEKGKAFFLRRRKERLILEEVEEILEQCVVGTQE QLEPSVIEEALVKLQPPTVHGHEHDMLSNISYLIARGTE DAFMQSLEDLRLAHAPYGLE YQMNGPWPAYSFSDQQL EGGVNDQ*	131
Octadecabacter arcticus 238_gvpFL	MSSATSIVYVYGVLTNCS DLVLD FPPGDLA GIVESGPLRI LPFGDIGALVCD FILPDGSDLK TILED SRSAERMILNHH VLADMVSRYTILPLRFGAVFAEDAGVIAALGGRYSTLQ KELDRIDGAIEWGVKSFCNRKMFSECVAETVSEISVLEK EIA DQGEKGKAFFLRRRIQRLILDEVEKTLEQCLVGAQDQ LKSRAIEETLVKLQPPTVHG HKHEMVSNRSYLIARGAED AFMQSLDDLRV VYAPFGFDYQINGPWPAYSFSDQQLGG GVNDK*	132
Rhodobacter capsulatus SB 1003_gvpFL1	MGHYLYG LLAPPARGTLAQMQAAAAGVTS LGGPVALS AVEGMLLVHCPCDLAEISQTRRNMLAHTRMLEALMPL ATCLPVRFGVIAQDLAEVARMIHERRAELVGHAQRLLD PVEIGLRVRFPRDRAL AQLMAETPDFVAERDRLMGQGA GAHFARADFGRR LAEALDARRTRDQKRLLAALRPHVR DHVLRAP EEDVEVLRAEFLIPAAGVDAFSRIAHD LAAAL GFAGAAEP ELQVIGPAPPYHFLSLSLAFDNTSEAA*	133
Rhodobacter capsulatus SB 1003_gvpFL2	MAHEIIAILPCEAAQLPSGLTG VVGRGATAVLAPAPGW AERLTGGPKQTAVRHHSRLEALMAMGSVLPFAAGIAC PEEAALLRLDAPLIARLAAEIGPRRH FQLALDWDESRV LAAFRDSELA PLFSGAAVTPEALRQAITALADRLSATA LRLLD PVAEDPVEQPRAPGCLLNLVLLRPEDEPRLDAA LQAIDALWSEGLRLRLIGPSAPISHALVDIDRADVAALA AAADLLKVAPEAGPEAVTEAAKAALRSPDLAANAAEQI	134

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	RAAARLLL RAGDIAALGLSGAATLPHLVHLRPGGRKSG LTSSGEAA*	
Rhodobacter capsulatus SB 1003_gvpFL3	MTGLALHGFVSPDGWSAAAAPPARCAVVLGGVAALVS EAGDALDTPETAQAAALAHHALISAWHRRGPVLPVRLG TVFSSQAALQTALAPKAAQLRAALDALADKEEMVLTIV PAARPPDLPPPAATGADWLRARKAVRDRGQARQTDRQ QTLAGLQDALRAQGVASLAAPAPREGGSRWHLLIARD DGAGLDRWLAAQADRFDAAGLDLTLDGPWPPYRFAAE ILEALDG*	135
Rhodobacter capsulatus SB 1003_gvpFL4	MSEPRISGLAPWRADLPDVIGCHGGWVLMGAAADETP EARLRRQVGWCRAAVDVLPLSPRLAPTRAEALVATR GPDLERAHHRHGRQLQVIVQLEMCRDGLVLRREISGGR SWLQDRAERATREARANADFEAQVRRVVRALFPREGQ VVTLAPSGTAGQLRLRAVLVPRAGLQAFAAALSADLD RDGRGGLWDVIAPLPPLAFAAALEAGPGGAVT*	136
Rhodobacter sphaeroides 2.4.1_gvpFL1	MIYLYGLLEEPASGHEVLAGMAGVTGPIALARLPGGILI YSSATEADILPRRRLLLAHTRVLEAAAWFGNLLPMRFG MMASTLAEVAAMLASRLTELCAAFDRVRGRVELGLRL SFPREPALAATLATAPDLAAERARLLALRRPDPMAQAE FGRRLAERLDARRGETQRLLFQSLRPLWVDHRLRVPDS DVQVIAVDVLVEDGAQDRLAAALVKAAADCSFAPTAE PSVRVIGPVPLFNFVDLVLSPRREEVA*	137
Rhodobacter sphaeroides 2.4.1_gvpFL2	MRLREVAVVLEGHPPSVLPEGTEAICEAGLTAILGMPPG LLSGRRALLEHAACRQAVLERLMAFGTVLPVLTGNCLT PAEAAAALAANSPLRQELRRLAGRVQFQVLVQWHAA LVPKRTDPDETAEDLRLRFTHRIADALARVAERHVNPL REDMLANQALLLLQTRTDDLDRSLEQIDALWTEGLRIR RIGSPPPVSFASLNFRRVSSAAIRRARHRFDLEGPVDPIRL RALRRDLLLLRASEAERAELAAAALVLDLLTRCAASGGD LHLVRIWSEGQAVPSDLEDA*	138
Rhodobacter sphaeroides 2.4.1_gvpFL3	MSGLLLLGVVSGLGISPAITSPHLRLDGDGYAAILLSLDR LPPDPASPDWAVQAALAQNAILSAYAATEDVLPVALGA AFTGIAAVKRHLDAERATLDAGMERLAGRAEYVAQLIA EQVADGAAPAPASGS AFLKARSARHEQRRHLARERTGF ARATAEELASLSCSASARPLKPDGPLLDLSLLVARDRV GLLEAAEASSRAGSRLALSRLIGPCAPFSFLPETRGHD*	139

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Rhodobacter sphaeroides 2.4.1_gvpFL4	MAGDARSVRHLHLAAMRDCETFLPFPPAATIAVDEAIA WCGRRTNALAEIDRFSRQRQLTVSARLIAPLLPDAAAS GAGWLRARRDASAHQARLRTVLMQIMSLGVCIPG RLQDEVQVNLLVPAAETHPVLHELRLRVGDALWSA CTVTGPWPPYAFISWETA*	140
Rhodococcus hoagii 103S_gvpFL1	MSEQESAPDGGGPVVYVYGLVPADVEVKEDATGIGSPP RPLKIVHHEDVAALVSEIDPDTPLGSSDDLRAHAAVLDS TATVAPVPLPLRFGAVLTDTDVVAELLEPYRDEFHEAL EQLEGKVEFVVKGYVEDAILREILADDPEAARLRDVV REQPEDTTRDERLALGERISQALTAKREQDTGRIVEALQ PAATAVAPREPTDDEEAGSVAVLISADGVDELDKAVAR LIDDWQGRVEVTVTGPLAAYDFVKTRAPGT*	141
Rhodococcus hoagii 103S_gvpFL2	MTPDDGVWVYAVTGDGSPGGISGIRGVAGEELRTVTD SGFTAVVGTVRLDTFGEEALRRNLEDLDWLADTARRH DAVVAICAGGATVPLRLATVYFDDDRVRTMLRDNAE QLGEALQQIADRSEWGVAYLERPRSEPRDAREKTGRP SGTAYLMQRRQAQAAREQAESAAGRRADEIFAELARW AVAGVRQPPSPDLAGRRSQEILNTSFLVDNGRHREFVT AVEELDARLSDVDLVLTPGPWPYPYSFTSVEASAR*	142
Serratia sp. ATCC 39006_gvpFL	MSLLLYGIVAEDTQLALEPDGSPHAGEEPMQLVKAATL AALVKPCEADVSRPAAALAFGQQIMHVHQQTTIPIRY GCVLADEDAVTQHLLNHEAHYQTQLVELENCDEMGR LSLASAEDNAVTPQASGLDYLRSLRAYAVPEHAERQ AALLNNAFTGLYRRHCAEISMFGQRTYLLSYLVPRTG LQAFRDQFNTLANNMTDIGVISGPWPYPYNFAS*	143
Stella vacuolata-ATCC- 43931_gvpFL1	MSGLLVFAIVPADGIEPGILAPREELPANLRAVAADGFA AVVGAAPEGGLKGRDRSVLLPRLASQKVIERLMARGP VLPVTLGTVLEDEARVRHMLAAGAPMLEAAFGTLGDC WQMDLSVRWDLNQQVARLMGEVPGDVRAAAGSGDE AARRALGEALAGLAAGERRRVQSRLAAALRDVARDLI VSEPVEPESVVDIAILVERPALAEVEAALDRLDAEFEGR LKFRLVGPLAPHSFATVQVHLAPEAALAGACAELGVER GAGLQDVKVAYHRALVRFPDLAPHGDDGGPEDEHDG GEGRASRLTAVTAAYRALQAEHAPISLRRQDGIQAVNQE QDASAAMGQQRGIVPGRELQALRM*	144

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
<i>Stella vacuolata</i> -ATCC-43931_gvpFL2	MLYVYAIAADHPDPDNAMFGGEGIVPDAPVRLLQLGDL AVAASLVSAADFAADALRAHLEDARWTALRVLAHQRV VDSLLPHATVLPKMFCTLFSGEALKQALAHNRAALQA TVERLRGAREWGVKLYWEAPRNPAPPSAGQGEAGAGA AFFQKRKRDQQRQRAEAEAAVARCVAASHRRLADAARA AVANPVQPPAVHRQPGEMALNGAYLVARAAEPAWRE VLAELERTHADGGIRYELTGPWGPYNFTGSGLVGS*	145
<i>Thiocapsa rosea</i> strain DSM 235 Ga0242571-11_gvpFL	MSDRPRPMLHCILRSPPGSIARAEAGLRWIERDGLAALV ADREPSEIAGASSVGLQRYADIVAEIHACAAVIPVRFGC LLAGDEAVGKLLHRSRDLHGLLDQVGDCLEFGIRLLL PADAPAATDDDAAPRLHANAPSDPRADPDMGPGLSHLL AIRHRLDVEASLAARAREAREVIKGRVAGRFREVEEL GQIDGRSLLSLYFLVPREQGEHFVECLRQDASSLRGTGL LTGPWPPYNFVGAIIDDIRSLD*	146
gvpG		
<i>Anabaena-flos-aquae</i> _gvpG	MLTKLLLLPIMGPLNGVWVIAEQIQERTNTEFDAQENLH KQLLSLQSFIDGEIGEEFEIQEEILLKIQALEEEARLEL EAEQEEARLELEAEQEDFEYPPQFTAENVKQHLVLLP*	147
<i>Bacillus-megaterium</i> _gvpG	VLHKLVTAPINLVKIGEKVQEEADKQLYDLPTIQKLI QLQMMFELGEIPEEAFQEKEDELLMRYEIAKRREIEQWE ELTQKRNEES*	148
<i>Ancylobacter aquaticus</i> strain UV5_gvpG	MGMLTDVVFAPAVGPKGVWLARIIEQAERTLYDE GVIRAALLDLEQQLEAGEIDEDAYETQETVLLERLKIAR ERMRSGL*	149
<i>Aphanizomenon flos-aquae</i> NIES-81_gvpG	MLTKLLLLPIMGPLNGLVWIGEIQIQERTNTEFDAQENLH KQLLNLSLQSFIDGEISEEDFEIQEEELLLKIQALEEEARLE LELAEEEEARLELELEQEEEDFVVKPQLTTEIDRDKDLV LLP*	150
<i>Aphanothece halophytica</i> (strain PCC 7418)_gvpG	MVFKLLLLPITGPIEGVTWLGEQILERANQELDEKENLN KRLLSLQSLDLGEISEEYDEQEEILLAMQAMEDEEN NQAEETD*	151
<i>Aquabacter spiritensis</i> strain	MSLVTDVLFAPAVGPKGVWLARLIAEQAERTLYDED	152

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
DSM 9035_gvpG	VLRAALLDLEQRFEAGEISEADYETEEDILLARLKIARER MRSGL*	
Bradyrhizobium oligotrophicum S58_gvpG	MLFQILTSPVSGPFRMVSWIGGAIRDAVDTKMNDPAEIK RALAALEQQLEAGSLSEQDYERMELIERLQSSLRHG SGNGG*	153
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpG	MFILDNLLAAPIKGMFWIFEEIAQAAEEETIADIEMIKAA LVELYRELESGQIDETEFETRERALLDRLDSLETS*	154
Chlorobium luteolum DSM 273_gvpG	MFILDDILLAPLSGMVFLGRKINEIVQNEMSDEGAVKEQ LMKLQFRFEMDELSEEEYDRLEDELLSTLAEIRAQKENR *	155
Dactylococcopsis salina PCC 8305_gvpG	MVFKLLLLPITGPIEGITWLGEQILERADQELDSKENLNK RLLSLQLSLDLGEISEEEYDEQEEIILLAMQAMEDEENE EEES*	156
Desulfobacterium vacuolatum_DSM 3385_gvpG	MFLVDDILFFPAKSLVWVFRELHNAVQQEKTNESDAL TELSELYMMLETGKITEEEDEREQILDRLEIQRDQ*	157
Desulfomonile tiedjei DSM 6799_gvpG	MERYTMFLDDILFLPMNGVLWICNEIHDAAEQELHNE SDAITAQLQKLYTLLEAGDIGESEFDVLEAELLDRDLAI QERGALLEA*	158
Desulfotomaculum acetoxidans_DSM 771_gvpG	MLGKLLLSPI LGPVMGVKFIAEKIKQQADQELYDKSKIK QDLMELQIKLELEEITEEYYLQREEELLVRLDELASMET EEEEV*	159
Dolichospermum circinale_gvpG	MLTQLLLLPIMGPLNGVWVIAEQIERTNTEFDAQENLH KQLLSLQLSFDIGEISEEEFEIQEEIILLKIQALEEEARLEL EAEQEEARLELEAEQEQRARLELEAEQEELNQPQLTPKI DTYRHLVKL*	160
Enhydrobacter aerosaccus strain ATCC 27094_gvpG	MGMLARLLTLPVSAPVGGVLWIARKIEEEANAERWDR NKITGALSELELDELGAIDVEEYDAREAVLLQKLKELQ EVEND*	161

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Isosphaera pallida_ATCC-43644_gvpG	MFLVDDILLAPAHSLMFLLRREIHQAAL EELRRDAQKVR EELAECYRALETGALTDEEFASLETDLRLDALEELAR FNSDEDDDPEDWDVEDDDPAEAVW*	162
Legionella drancourtii LLAP12_gvpG	MLLLGSILMAPVHGLMAIFEKIKEAVDEEKQHDIERIKS ELMALYTKLESGELSEADFEKQEKILLDKLDSLEDEDD*	163
Microcystis aeruginosa NIES-843_gvpG	MFLDLLFLPVTGPIGGLIWIGEKIQUERADIEYDEAENLHK LLLSLQLSYDMGNISEEEFEIQEEELLLKIQALEEEEEAE ESESSL*	164
Nostoc punctiforme ATCC 29133_gvpG	MVLRFLLLPTGPLMGVTWLGEKILEQASTEIDDKENLS KQLLALQLAFDMGEIPEEEFEIQEEALLAILEAEQEERD QTQEY*	165
Nostoc sp. PCC 7120_gvpG	MLGKILLLPVMGPINGLMWIGEQIQUERTNTEFDAQENLH KQLLSLQLKFDMGEISEEEFDIQEEILLKIQALEAEERL NAESEDDDLVDQPIFILASEENPVYQDQSRFSEEYEDK EDLVLS*	166
Octadecabacter antarcticus 307_gvpG	MGHILNTLMSPLIGPMKGVFWVAEQIKDQTD AEIYDDSK ILVELSELELLLDLEKIELKDFEAKEDVLLKRLQEIRKAK KNDSV*	167
Octadecabacter arcticus 238_gvpG	MSHILNTLMGPLIGPMKGLLWVAEQIKDQADAELYDDSD KILVALSELELSFDLEQIELKEFEAQEDVLLQRLQAIRKA KQNDTD*	168
Pelodictyon phaeoclathratiforme_gvpG	MFILDDILFAPLNGLIFIACKINDVVEKETSDEGVVKERL MALQLRFELDEIDEVEYDREDELLQKLERIRLNKQNQ*	169
Phormidium tenue NIES-30_gvpG	MLFKLLFAPVLGPIEGISWVANKLLEQADVPTNDLESQ KQLLALQLAFDMGEVAEADFEIQEEILLAIQAI EDEEDE DE*	170
Planktothrix agardhii str. 7805_gvpG	MILRLLSPITAPFEGVIWIGEQLLERAEAEELDDKENLGK RLLALQLAFDMGDIPEEDFEVQEEELLLKIQALEDEANQ ENDEID*	171
Psychromonas ingrahamii	MFILDDILLAPYSGIKWLFKEIQRQAQEELDGEADRITTD	172

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
37_gvpG	LTNL YRQFESNEITEQEFEERETVLLDRLDELQEE SNLLD EEYDEEYEDDDDEEYEDDDDEEYEDDDDEEYEDDDDEEYED DDKNDKDKND DHDNDDDDENKDENDKYNDEER*	
Rhodobacter capsulatus SB 1003_gvpG	MGLLRKLLLAPVELPITGALWIVEKIAETAES ELDTPGT VRLLRGLEQQLEAGEITEEEYEFAEEILLDRLKRGQAA EARS GGP*	173
Rhodobacter sphaeroides 2.4.1_gvpG	MGLLTSLTLTPFRGPF DGT LWI AARIGEAAEQSWNDPA ALRAALVEAERQLLAGELSEETYDAIELDLLERLKG TAR *	174
Rhodococcus hoagii 103S_gvpG	MGLFSAIFGLPLAPVRGVVWIGEVVRRQVEEETTSPAA MRRDLEAIEEGRRSGEISEDEAAQA EDEILHRVTRRRDA GASGEE*	175
Serratia sp. ATCC 39006_gvpG	MLLIDDILFSPVKGVMWIFRQIHELAEDELAGEADRIRES LTDLYMLETGQITEDEFEQQEAVLLDRLDALDEEDDM LGDEPGDDEDD EYEEDDDDEEDDDDEEDDDDEDDDD DDEEDDDDDDEDDDDDEDEPEGTTK*	176
Stella vacuolata_ATCC- 43931_gvpG	MGLVTNVAFAPVVGPLKGVLWLARLIADQAERTLYDE DLVRAALLDLEQRLDAGQISEADYDAEEEEILLARLKIAR ERMRSGL*	177
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpG	MLIVDDLLAAPFKGIWVFEEIHKSAEQRARRDEIMA ALSALYRALEQGEITDDTFDTREQALLDELDA LDARED ANELGSEDEDDLDGAGEDAS*	178
Tolypothrix sp. PCC 7601_gvpG	MEVMIMLGKILLFPVMGPISGLMWIGEQIQERTDTEFDA QENLHKQLLSLQSF DIGEISEEDFEEQEEELLLKIQALEE EKARLEAESIEDEEDEV EPTYFIAEVEEDKVLAEAFRGN KKYEDNENLVLSP*	179
Trichodesmium erythraeum IMS101_gvpG	MLLRLLTLPISGPLEGVTWL GKKLQEQVDTEIDETENLS KKLLTLQLAFDMGEISEEDFEDQEEELLLAIQALEEQKL KEEEEDA*	180
gvpJ		
Anabaena-flos-aquae_gvpJ	MLPTRPQTNSSRTINTSTQG STLADILERVLDKGIVIAGDI	181

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	SISIASTELVHIRIRLLISSVDKAKEMGINWWESDPYLSTK AQRLVEENQQQLQHRLESLEAKLNSLTSSSVKEEIPLAAD VKDDL YQTS AKIPSPVDTPIEVLDFQAQSSGGTPPYVNT SMEILDFQAQTS AESSSPVGSTVEILDFQAQTSEESSSPV VSTVEILDFQAQTSEESSSPVGSTVEILDFQAQTSEEIPSS VDP AIDV*	
Bacillus-megaterium_gvpJ	MAVEHNMQSSTIVDVLEKILDKGVVIAGDITVGIADVEL LTIKIRLIVASVDKAKEIGMDWWENDPYLSSKGANNKA LEEENKMLHERLKTLEEKIETKR*	182
Ancylobacter aquaticus strain UV5_gvpJ1	MNEQRMEHSLQAVGLADILERVLDKGIVIAGDITISLVE VELLNIRLRLVVASVDRAMSMGINWWQSDPHLNHAR ELAEENKLLRERLDRLEAAVPSALPADAALEPSLAGE DARHGG*	183
Ancylobacter aquaticus strain UV5_gvpJ2	MPSRHSGEIAVADLLDRALHKGLVWGEATISVAGVDL VYLGLKLLLTSTDTVNRMR E AANAPPDERHLHAD*	184
Aphanizomenon flos-aquae NIES-81_gvpJ	VTSTPILPTRPQTNSSRAINTSTQGSTLADILERVLDKGIV IAGDISISIASTELIHIRIRLLIASVDKAKEMGINWWETDP YLSTKAQRLVEENQQQLQNRLENLESQINLLTSAKVQEIQI SLVETTEDNTHQT TEDNTHQT HEESIPLPIDSQLDV*	185
Aphanothece halophytica (strain PCC 7418)_gvpJ	MVNPNTNPKSYQSKGITNSTQSSSLADILERVLDKGIVI AGDITVSVGSTELL SIRIRLLVSSVDKARELGINWWEGD PYLSSQANLLKEENQALQNRLENMEAE LRRLKGETNPE PSFLSESEDNS*	186
Aquabacter spiritensis strain DSM 9035_gvpJ1	MSEQRMEHSLQAVGLADILERVLDKGIVIAGDISISLVE VDLLNIRLRLVVASVDRAMSMGINWWQSDPHLNHAR QLEEEENRLLRERLDRLEAALAPPEGGMLRAEVEVAHGG *	187
Aquabacter spiritensis strain DSM 9035_gvpJ2	MPDPEPIIPRTSGDVALADLLDRALHKGLVLWGEATISV AGVDLVYLGLKVLLASTDTANRMRDAAAASAAGSHLP GG*	188
Arthrospira platensis NIES- 39_gvpJ	MTLQSRSSSPQRGVPMSTSGSSLADILERVLDKGIVIAGD ISVSVGSTELL SIRIRLLIASVDKAKEIGINWWESDPYLSS QAQQLSQSNQQLLEEVKRLQEEVRSLKALTSQSSQPVTP	189

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	PNSEDD*	
Bradyrhizobium oligotrophicum S58_gvpJ1	MTFTVHQPTGGDRLADILERVLDKGIVVAGDVTISLVGIELLNKIRLIVATVDRALELGINWWEADPRLTTRASELS VENEELKKRLALLEADAGRNRQRPRKRRVRSIAATSGASHER*	190
Bradyrhizobium oligotrophicum S58_gvpJ2	MTYRADLDYLEPAASSEGSLELLDHLDRGVLLWGELRISVADVELIEVGLKLMLASARTADRWRTTQTRASIAPGDCP*	191
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpJ1	MRSADGEPVSAELAQRSLCESLDRILNKGAVISAQVVVSVADVDDLVLHLRLLTSVETALVGRAMPREEASR*	192
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpJ2	MADLLERVLDKGVVITGDIRINLVDVELLTIRIRLLVCSVDKAKELGIDWWNADTFFLGPDRGQSALPGRASAVDVAAGSAVHADAAHR*	193
Chlorobium luteolum DSM 273_gvpJ1	MPELKHAVNATGLADILERVLDKGIVIAGDIKIADIDL LTIKIRLMVASVDKAIEMGINWWQEDPYLSTGAKTSEQ TRLLGEINQRIEKLESINR*	194
Chlorobium luteolum DSM 273_gvpJ2	MQEDLYTANRQVTLLDILDRVLNKGVVISGDIIISVAGIDL VYVGLRVLLSSVETMERLDAARAEGQLQ*	195
Chlorobium luteolum DSM 273_gvpJ3	MAVEKTIGSSSLVEVIDRILDKGVVVDWVRVSLVGIEL LAIEARVVVASVETYLKYAEAIGLTAKAA*	196
Chlorobium luteolum DSM 273_gvpJ4	MAVEKTIGSSSLVEVIDRILDKGVVVDWVRVSLVGIEL LAIEARVVVASVETYLKYAEAIGLTAKAA*	197
Dactylococcopsis salina PCC 8305_gvpJ	MVNSNTNQPKSYQSKGITNSTQSSSLADILERVLDKGIVI AGDISVSVGSTELLTIRIRLLISSVDRAREIGINWWESDPY LSSQAHLMKREENQALQSRLENMEAELRRLKGETNLDQSLGESDQRSLQ*	198
Desulfobacterium	MAYIDIDNDASKQISICEALDRVLNKGAVITGELTISVAD IDLIYLSLQAVLTSVETARHMFDSQINDAVKEVK*	199

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
vacuolatum_DSM 3385_gvpJ1		
Desulfobacterium vacuolatum_DSM 3385_gvpJ2	MPIQRTAQHSIESTNIADLLERVLVDKGIVIAGDIKISLVDI ELLSIQLRLVICSVDKAKEMGMDWWVNNPVFMPNKG QNDEIADTLTKINSRLEHLEKATISGS*	200
Desulfomonile tiedjei DSM 6799_gvpJ1	MMDEEEHVSLCEALDRVLNKGAVIAGEVTISVANVDLI YLGLQVVLASVDITIRGKRNELLRHDVGLHLTADNA*	201
Desulfomonile tiedjei DSM 6799_gvpJ2	MSIQASTRHSIQSTNLADLLERVLVDKGVVIAGDIKIKLVD VELLTIQIRLVVCSVDKAKEMGMDWWTNNPAFQPALA QISE*	202
Desulfotomaculum acetoxidans_DSM 771_gvpJ1	MGPQMGPSTGNLSLLDVIDRILDKGLVINADISVSIVG VELLGIKIKAAVASFETAAYGLQFPTGTEINEKVSEAA KQLKEICPECCKSGRDELLHEGCPWCGWISARALRE TEHSQR*	203
Desulfotomaculum acetoxidans_DSM 771_gvpJ2	MLPIREERATLTDLLDRVLVDKGLLLNADILISVAGVPLIG ITLKAAIAGMETMKKYGLLIDWDQESRLAERRLRSSRH*	204
Enhydrobacter aerosaccus strain ATCC 27094_gvpJ1	MAVTNGRMEHSIQGSSLADILDRILDKGIVIAGDVTISLV GVLLNIRLRLVASVDKAIEMGINWWEADPYLTSQTK ASSEQTELLQQRLERIEGLLAGQATKEQPL*	205
Enhydrobacter aerosaccus strain ATCC 27094_gvpJ2	MPVQTAHDGELALADLLDRALNKGVVWLGDATISLAG VELVYVGLRVLVASCSTMEKYRSSPRKGSMPARGES*	206
Isosphaera pallida_ATCC-43644_gvpJ1	MIVCSSSTPERIGPPMNLPPPHHAPWCYDSPDLETPLDP AERIALCEVLDRVLNKGVIHGEITISVAGVDLVYLGLN LLTTSVETAQSWKFRGMIE*	207
Isosphaera pallida_ATCC-43644_gvpJ2	MAITRSSRPDVTHTSGATLADVLERVLVDKGLVIAGDIK IKLVDVELLTIQIRLVVASVDKAREMGLDWWTRSPSS LAATTCPALTPPKQEATPPATRIQAPTESAQTTPDQSHPS DPSASNIDEVAELRRHIELMQLRDEARQRAHREELAALR AQLTRLTELLDSPR*	208

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Legionella drancourtii LLAP12_gvpJ1	MIIEDKPVSLCETLDRVLNKGVVVAGTVTISVADVLLY LDLHCLLSSMKGMNLIGSERER*	209
Legionella drancourtii LLAP12_gvpJ2	MELQKSPTHSIGSTTIADLLERILDKGIVIAGDIKVNLVQ VELLTIQIRLLICSVDKAKEIGMDWWTHQNDVQSKNGS MPIQEYVTQMEERLKNLENTLASSKNAI*	210
Lyngbya confervoides BDU141951_gvpJ	MTGQSLSRSSSANRQMATATQGSTLVDVLERVLDKGIV IAGDISVSVGSTELLTIRIRLLVASVDKAREMGINWWEN DPYLSARSQELLTANEQLQSRIESLEQELKSLRSQED*	211
Microcystis aeruginosa NIES- 843_gvpJ	MTSSTFAGSLRNQSNNSLKTATQGSSLADILERVLDKGI VIAGDISVSIASTELINIRIRLLIASVDKAREMGINWWEG DPYLHSQSQALLAENRELSRLQTLETELETLKSLTQLS AMESHDTSPNDEAHSSDA*	212
Nostoc punctiforme ATCC 29133_gvpJ	MSTNTNRGAITTTSTQGSTLADILERVLDKGIVIAGDISISV GSTELLNIRIRLLISSVDKAKEIGINWWESDPYLSNQTRT LLATNQQLQERLASLETQLSLKALNPINHQNAGD*	213
Nostoc sp. PCC 7120_gvpJ	MTTTPHPTRPQTNSNRVIPTSTQGSTLADILERVLDKGI VIAGDISISIASTELIHIRIRLLISSVDKAREMGINWWEND PYLSSKSQRLVEENQQLQQRLESLETQLRLLTSAAKEET TLTANNPEDLQPMYEVNSQEGDNSQLEA*	214
Octadecabacter antarcticus 307_gvpJ1	MNDGKMEHSLNATNLADILERVLDKGIVIAGDVTISLV GVLLNIKLRLLIASVDKAMEMGINWWAHDPLTAGA QAPAVADPAMLERMDRLEAALATALASNQTTPMKGHK *	215
Octadecabacter antarcticus 307_gvpJ2	MTNKAQGGQDLALADLLDRALSTGVVIWGEATISLAG VDLVYVGLKVLVASVDAAERMKAASLVDRPTDRGQQI *	216
Octadecabacter arcticus 238_gvpJ1	MNNGKMEHSLDATNLADILERVLDKGIVIAGDVTISLV GVLLNIKLRLLIASVDKAMEMGINWWAHDPLYTAGA QAPVGVDPAMLERMDRLEAALAKALASNQTTPAEGQS S*	217
Octadecabacter arcticus	MTNETQGGQDLALADLLDRALSTGVVIWGEATISLAGV	218

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
238_gvpJ2	DLVYVGLKVLVASVDAAQRMKDASLVDRPTDGGQ*	
Pelodictyon phaeoclathratiforme_gvpJ1	MPELKHAVNATGLADILERVLDKGIVIAGDIKIADIIDL LTIKIRLLIASVDKAMEMGINWWQEDTYLSTKAKDKEQ QLLRDDLQQRIEKLEALTKIT*	219
Pelodictyon phaeoclathratiforme_gvpJ2	MQDEFYSKNKEITILDVLDRLVTKGVVITGDIVISVADID LVYVGLRLLLSSVETMEKNKQNSIKM*	220
Phormidium tenue NIES-30_gvpJ	MATATQGSSLVDVIERVLDKGIVIAGDISVSVGSTELLSI RIRLISSVDKAREIGINWWESDPYLSSTNELLEANQQL QSRLETLEAELKALRSAEPVS*	221
Planktothrix agardhii str. 7805_gvpJ	MNSQQLPNSNIQRGVPTSTQGSSLADILERVLDKGIVIAGD ISVSVGSTELLNIRIRLLIASVDKAREIGINWWESDPYLS QTKVLTESNQQLLEQVKFLQEEVKALKALLPQENQPNPI SDPHK*	222
Planktothrix rubescens_gvpJ	MNSQQRPSNIQRGVPTSTQGSSLADILERVLDKGIVIAG DISVSVGSTELLNIRIRLLIASVDKAREIGINWWESDPYLS SQTQVLTESNQELLEQVKLLQEEVKALKALLPQENQPK EME*	223
Psychromonas ingrahamii 37_gvpJ1	MANVQKSTDSSGLAEVVDRILEKGIVIDAFVKVSLVGIE LLSIEARVVIASVETYLKYAEAIGLTASAATPA*	224
Psychromonas ingrahamii 37_gvpJ2	MPMANVSINPELTAQECEKISLCDALDRIINKGVVIHGEI TISVANVDLISLGVRLILSNVETREQSNTPKEEV*	225
Psychromonas ingrahamii 37_gvpJ3	MATGKPQSMTHSVKSTTVADLLERILDKGIVVTGDIKIK LVDVELLTVELRLVICSVDKAVEMGMDWWNNNPAFAP QAPAQEGELSSIEKRLEKIEKALVK*	226
Rhodobacter capsulatus SB 1003_gvpJ1	MGYRSASQPEGLADVLERILDKGIVIAGDVSVSLVGIEL LTIRLRLLIATVDKAREMGIDWWSHDPYLNGLRPGEP APETETETAALRDRLAQLEAQLSALGAQVGAAPALAE ALRGLAAAGSSALCAAPEASSADVVPVFRRYKEAP*	227
Rhodobacter capsulatus SB 1003_gvpJ2	MDDRFSRLRFGPEEVFDAPSGGLADLLDGLLGHGIVLH GDLWLTVADVELVYVGLSAVLASPEALRSHE*	228

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Rhodobacter sphaeroides 2.4.1_gvpJ1	MSFQMQSPLQQDSLADVLERILDKGIVIAGDISISLVGIE LLTIRLRLLVATVDKAREMGINWWESDPRLCITQAPASD GSAALLDRLERIETQIGQLAAAREG*	229
Rhodobacter sphaeroides 2.4.1_gvpJ2	MTDSAPTLQFATAEEALQSSETRLVDVVDALLSQGIAIR GELWLTIAADVLDVFLGLDLLLLANPDRLQCRVPDAA*	230
Rhodococcus hoagii 103S_gvpJ	MTRSGSGANYPQQYSQGLGGAGHEPANLGDILERVLD KGIVIAGDIRVNLLDIELLTIKLRLVIASLETAREVGIDW WEHDPWLSGNNRDLEENERLRARIEALESGERRVADV TDPHRAVQPAESPAAEVRDDDA*	231
Serratia sp. ATCC 39006_gvpJ1	MPVNKQYQDEQQQVSLCEALDRVLNKGVVIVADITISV ANIDLIYLSLQALVSSVEAKNRLPGRE*	232
Serratia sp. ATCC 39006_gvpJ2	MSGNKKLTHSTDSTTVADLLERLLDKGVVISGDIRIRLV EVELLTLEIRLLICSVDKAVEMGLDWWSGNPAFDSRAR VSSAPAPELEERLQRLEARLEAAPSVEETHL	233
Stella vacuolata_ATCC- 43931_gvpJ1	MSGQRMESVQAVGLADILERVLDKGIVIAGDISISLVE VELLTIRLRLVVASVDRAMSMGINWWQSDPNLNSHAR QLEEDNRLLRERLDRLEAALALPEMAGERLADAGQGG GAEQGVTHGR*	234
Stella vacuolata_ATCC- 43931_gvpJ2	MSDPEPIIPTSGDIALADLLDRALHKGLVLWGEATISV AGVDLVYLGLKVLVASTETADRMRAAAASQSADPKVR AG*	235
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpJ1	MMLAIGEHPDCPEEIQRVSLCEALDRILNKGAVVSGELT IAVANVDLLYLSLQLVITSVETAKREMLYVRH*	236
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpJ2	MSVQRSTLTHSTNSTSVADLLERVLDKGIVIAGDIRIKLV DIELLTIQLRLVICSVDKAREMGIDWSDNAMFKGLSS QASAASLPGTAAASGIEDRLARLESLLVKQSAAAETVL*	237
Tolypothrix sp. PCC 7601_gvpJ	MADILERVLDKGIVIAGDISVSIASSELLHIRIRLLISSVDK AKELGINWWENDPYLSSKSQRLVEENQQLQQRLESLEA QLRSLTAAKINNPELFPVNAEDNGQSDEENVPLPMNYQ PND*	238

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Trichodesmium erythraeum IMS101_gvpJ1	MFIRVDFLLDKGVVDWVRLSLVVIELLTIEAKIVIASV EAYLKYSEAFCFNY*	239
Trichodesmium erythraeum IMS101_gvpJ2	MAVEKVNSSSLAEVIDRILDKGVVVDWVRLSLVGIEL LTIEARIVVASVETYLKYAEAVGLTTLAAAPGEAAA*	240
Trichodesmium erythraeum IMS101_gvpJ3	MAVEKVNSSSLAEVIDRILDKGVVVDWVRLSLVGIE LLTIEARIVIASVETYLKYAEAVGLTTLAAEPAA*	241
Trichodesmium erythraeum IMS101_gvpJ4	MKTSANIATSASGNGLADVLERVLDKGVVIAGDISVSIA STELLNIKIRLLISSVERAKEIGINWWESDPYFSSQNNSLV QANEKLLERVASLESEIKALRSN*	242
Trichodesmium erythraeum IMS101_gvpJ5	MKTSANIAKSAGGDSLADVLERVLDKGVIAGDISVSIA STELLNIKIRLLISSVERAKEIGINWWESDPSLSSQNNSLV QVNQKLLERVASLESEIEALKYSQ*	243
gvpK		
Anabaena-flos-aquae_gvpK	MVCTPAENFNNSLTIASKPKNEAGLAPLLLTVLELVRQL MEAQVIRRMEEDLLSEPDLERAADSLQKLEEQLHLCE MFEVDPADLNINLGEIGTLLPSSGSYYPGQPSSRPSVLEL LDRLNTGIVVDGEIDLGAQIDLIHAKLRLVLTSPKI*	244
Bacillus-megaterium_gvpK	MQPVSQANGRIHLDPDQAEQGLAQLVMTVIELLRQIVE RHAMRRVEGGTLTDEQIENLGIALMNLEEKMDDELKEVF GLDAEDLNIDLGPLGSL*	245
Ancylobacter aquaticus strain UV5_gvpK	MTAPCTAETLENALRGRIDIDPEKVEQGLVKLVMLVE TVRQVVERQAIRRVEGGTLTEEETERLGLALMRLEEK AELRLHFGLEDGDLDLKLQPLGEL*	246
Aphanizomenon flos-aquae NIES-81_gvpK	MVYSPVENSNDFLNVIPVENSNEFLNTSPKKKSNSETGL APLLLTVLELIRQLMEAQIIRRMEEDLLSEDLERTAESL QKLEEQLNLCQIFDIDPADLNINLGDFGSLLPASGSYY GETGNRPSILELLDRLNTGIVVDGEIDIGVAQLDIHAK LRLVLTSPKI*	247
Aphanothece halophytica	MSADESNLSQVNLNPATSNSDAGLAPLLLTVTELIRQLM	248

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
(strain PCC 7418)_gvpK	EAQVIRRM DGGL NEEEL DRAGDSLQRLEAEIIRLCEIFE IDPKDLNVDL GELGTLMPKNGGYYPGESSDDPSILELLD RILHKG VVIDGNLDL GIAQLSLIQLARLHLVLTSQPINGK*	
Aquabacter spiritensis strain DSM 9035_gvpK	MTGFAGGPAVTETLESVLQGRVDIDPERVEQGLVKLV MVVETLRQVIERQAIRRVEAGALTDEEIERLGLTLLRLE EKMAELRVQFNLSEADLSLKLRLPLGEL*	249
Bradyrhizobium oligotrophicum S58_gvpK	MSASSHSEAPGLRLQLGDLDTALAAVFTDAAPNGSINL DPDKIEHDLARLVLTLIEFLRRLLELQAIRRMEANELSE EEERVGLALMRAAAQVSRLARELGVDPRENLQLGPLG RLL*	250
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpK	MNAPHAAAVSDAAALAAALEQALAQQAAPPPRATQRF DVATASAGNGLAKLVLALMKLLHELLERQALRRIEAGS LNDDEIERLGLALMRQAEEIERLAAQFGFTDADLNLDL GPLGRLF*	251
Chlorobium luteolum DSM 273_gvpK	MHEDKVQFQASSVEEALRQLEGMKQGKESRIEANPDN VESGLARLVLTLIELLRKLMEKQAMRRIDGGSLEAQID ELGETLMKLEMKMDLKKTFNLTDSDLNLDLGPLGDL M*	252
Dactylococcopsis salina PCC 8305_gvpK	MSEESNLSRVDLNPASSNSDAGLAPLLTVTELIRQLM EAQVIRRM DAELLTEAELDRAGESLQRLEEEILRLCEIFD VDPADLNVHLGELGTLLPKEGGYYPGETSDQPSILELLD RVLHTGVVIDGNLDL GIAQLNLIQAKLHLVLTSQPINN*	253
Desulfobacterium vacuolatum_DSM 3385_gvpK	MIKDPEAKDFKIESDSIDAFARVMHADTSSCSSSVTAG QRQQLKIDEENIKNGLAQLVMTLIKLLHELLERQAIRRI ESGSLDDDQIERLGLTLMQQCEEIDRLRKLFDLEEDLN LDLGPLGKLL*	254
Desulfomonile tiedjei DSM 6799_gvpK	MNPMNIAKVESDSLGDFAEIMQTDWISSLHSDKEEKRL NLNQDSVKNGLGQLVLT VKLLHDLLERQAIRRMEAGT LTDTEIDRLGTTLMMQAQEIERLRSEFGLEEDLNLDLG PLGKLL*	255
Desulfotomaculum acetoxidans_DSM 771_gvpK	MYIDISEGSLKQGVGLLLALVEIHKDALKIQALKRIEGD SLTEDEIERLGNALHELEEALVEIEMEHNLNQNVVQNIRE GLDNVVNEVVDTFNPERWIAENEFN*	256

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Dolichospermum circinale_gvpK	MLSTPADNFDESLTTVSKSKNEAGLAPLLLTVLELLRQL MEAQVIRRMEDNLLSESELERAAADSIQKLEEQILHLCET FEVDPaelNINLGDFTLLPQSGSYYPGETGSRPSVLELL DRLLNTGVVLDGEIDLGLAQLDLIHAKLRLVLTSKPI*	257
Enhydrobacter aerosaccus strain ATCC 27094_gvpK	MTKLLEAKTVDPDKAGDDL VKL VLALVETLRQLVERQ AIRRVDSGVLNDDEVERLGLALLRLEEKMSSELKAHFGF GDEELTLKLGLSLGELARDV*	258
Isosphaera pallida_ATCC-43644_gvpK	MSDSLFEVRSPSAAPPSPVNPVGVADEWTAVLKDWDTLT AQLRQATAPPNAENSARSHATTGRIDLDPEQVGDGLAK LVLTLELIRQLLERQAIRRLDAGSLDHEQTERLGLTLM RLAQRMEELKTHFGLQGEDLNLDLGPLGKLL*	259
Legionella drancourtii LLAP12_gvpK	MNDKREEDNALPQRINLQPDVKNGLGKLVLIQLIHE LLERQAIGRIEAGDLSDEQIDRLGITLMKQAEEDKLREV FGLTQEDLNLDLGPLGKLL*	260
Microcystis aeruginosa NIES-843_gvpK	MTLACTPYDSDNQALLTRPESNSQAGLAPLLLTVVELV RQLLEAQIIRMEKGVLSSEDLDRAAESIQKLQEQLYLC EIFEVEPEELNVHLGEFGTLLPEAGSYYPGEEGIKPSVLE LVDRLLNTGVVVEGNVDLGLAQLDLIHLKRLVLTSQP V*	261
Nostoc punctiforme ATCC 29133_gvpK	MQAISKSKGSDSGLAPLLLTVVELIRQLMEAQVIRRM AGTLNDSELDRAAESLQKLEQQVQVLCIFDIDPADLNI NLGEMGNLLPQSGGYYPGETSSQPSILELLDRLLNTGVV VEGDLDLGLAQLSLVHAKLRLVLTSKPL*	262
Nostoc sp. PCC 7120_gvpK	MVCTPVEKSPNLLPTTSKANSKAGLAPLLLTVVELIRQL MEAQVIRRMEDCLSESELEQASESLQKLEEQVLNLCHI FEIEPADLNINLGDVGTLLPSPGSYYPGEIGNKPSVLELL DRLLNTGIVVDGEIDLGLAQLNLIHAKLRLVLTSRPL*	263
Octadecabacter antarcticus 307_gvpK	MKTTSDSQFDSMKKILTSSKEDSASCDPTDLLPNKSLP PSLSTSPETAADDLVKL VLAVIDTVRQVMEKQAIRRVES GALAEAEIERLGLTLMRLLEARMVELKSHFGLSNEDLN HFGTVQDLKDILNDEE*	264
Octadecabacter arcticus	MKTQNDTQFDSMKKILTDSGGDPNPNNGSPDQTQHASL	265

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
238_gvpK	PSNLSTDPETAADDLVKLVLAVIDTVRQVMERQAIRRV DSGALADEEIERLGLTLMRLEERMADLKSHFGLSNEDL NLNFGTVQDLKDILNDEE*	
Pelodictyon phaeoclathratiforme_gvpK	MDSKILYYAGSADEIIEELEKLKPGIQGRINATPDNVES GLAKLVLTIELIRKLEKQAMRRIDGNSLSESQIEELGET LMKLEKKMEELKGIFNLTDKDLNLDLGPLGDLN*	266
Phormidium tenue NIES- 30_gvpK	MTSENAEPDLSTTLALQPPAKTDAGLAPLLTVIELVRQ LMEAQVIRRMESGDLDDNDLERAADSLRKLQVVS CEIFDVDPADLNIDLGEIGTLLPKEGNYYPGQKNQNP ELLDRLLDTGVVVEGDVDLGMAQLNLIHAKLRLVLT SKPI*	267
Planktothrix agardhii str. 7805_gvpK	MSSSEPSIETIITPKSSRKDAGLAPLVLTVELIRQLME AQVIRRMENGLSEEELDRAAQSLQQLQVLEIKLCEI FEIDPTDLNIELSEFGTLLPKSGSYYPGENTQNPSIL ELLDRLMTGIVVEGSVDLGLAQLNLIHAKLRLVLT SKPL*	268
Psychromonas ingrahamii 37_gvpK	MPFEHFKSNNQADVNSDTKPAASVGGLNLESDDLKNG LGRLVLTVLKLLHELLERQALRRMDAGSLQDDEIER LGLAFMKQAEEIDRLRKEFGLEVEDLNLDLGPLGRLL*	269
Rhodobacter capsulatus SB 1003_gvpK	MSAAMHLELGDVDAVLSQAARSLAAGGRLTDPERVE QDLARLVLGIVELLRKLMEQAIIRMEAGSLTPEQEET LGLTLMRAEAALHEVAAKFGLQPADLILDGLPLGRSV*	270
Rhodobacter sphaeroides 2.4.1_gvpK	MTYFPFPLLLRDDRPPTEAPVTAPRIALDPDRLEHDLA RILLGLMEMLRQIMELQAIIRMEAGSLSESQQEQLGT TLMRAEAAIHEMAARFGLTPADLSLDLGPLGRIT*	271
Rhodococcus hoagii 103S_gvpK	MRRRIDSDPESVERGLVALVLTVELLRQLMERQALRR VDAGDLSDDQIERIGTTLMLLEEKMEELREHFGLEPED LNLDLGPLGPLLAED*	272
Serratia sp. ATCC 39006_gvpK	MTTNQLSHHSPVFGPTSPAQRPIANRHKIDIDGERVR DGLAQLVLTVLKLLHELLERQAIIRMDSGSLSDEEVE RLGLALMRQAEELTHLCDVFGFKDDDLNLDLGPLGRLL*	273
Stella vacuolata_ATCC- 43931_gvpK	MTGFLNGPADVETLETALRGRVDIDPERVEQGLVKLV LMVVETLRQVIERQAIIRVESGSLTDDEVERLGLTLMR LE	274

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	EKMDQLRRQFDLGEEDLSMRLRLPLQEL*	
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpK	MSDTRTGTAPSSAASAAPDTSTLQRANLLADLLETQVA AAGRRIIDPERVQRGLGQLVLTVVKLLHVLLERQAIIR VDGGDLDEDEIEQLGLALMRQSEEIERLRRLGLLEEQDL NLDLGPLGKLF*	275
Tolypothrix sp. PCC 7601_gvpK	MAMVCTPSENSNDLLATNSKANNQAGLVPLLLTVVELI RQLMEAQVIRRMEEEECLSEDLERAAESLQKLEEQVLN LCQIFEIDPADLNIHLGELGSLPAAGSYYPGETGNTPSV LELLDRLLNTGVVVDGELDLGVAQLNIHAKLRLVLTS KPLNTK*	276
Trichodesmium erythraeum IMS101_gvpK	MSLENSPEESLIVPIDKSKSNPEAGLAPLLTVIELLRELM QAQVIRRM DAGILSDEQLERAAEGLRQLEEQVIKLCVKF DIPTEDLNLDLGEIGTLLPKSGEYYPGEKSENPSVLELLD RILNTGVVLDGTVDLGLAELDLIHARLRLVLTA*	277
gvpL		
Ancylobacter aquaticus strain UV5_gvpL	MLYLYAILESPPQKPLPPGIGGAAPLFVESHALVCAASE AADAIAAREPSQIWRHQEVVAALMEGRPVLPLRFGTVV EDSAACLRLRLARHHAELSAQLDRVRHCVEFALRVAGLS ELADPGLDPNATPAGLPGGASHLRTLVRREGWPVSSA AFPHDTLTAHAASRLWARSPPQDLRASFLVQRRSAS AFLDDVNALQRLRPDLGITVTGPWPPYSFSDPDLSSGRE *	278
Aphanothece halophytica (strain PCC 7418)_gvpL	MLYTYCFLFSPEKTLSPQGFKGDLQMIKGAIAAVVEP NLPKAELEDDQKL VQAVVHHDWVICELFRGLTVLPLR FGTYFRGEADLRSHLAAYEESYQQKL TALTGKVEVTLK LTPIPFSEEGSSSTAKGKAYLQAKKQRYQQSNTYTQQ QEALEKLQEEIKKTYPQLIHDEPKENTERFYLLIDSHSFS VFGEKMEQWKQFLSSWSIEISDPLPPYHFL*	279
Aquabacter spiritensis strain DSM 9035_gvpL	MLYLYAVLEAPPPARSLPPGIGGGAPHFIEAFELVCAAS ETPNRSVAPEPAEVWRHQVVEALIDRAPALPLRFGTL VEDASACRRLLTRHRDALGAQLGRVRHCVEFALRVSG PEEVAPDPGIGGGPGTSYLRTLARREAGWPPSTAVFPHD GLAAHAAERLLWARSTSQPDLRASFLVRKPNVAAFLAD VSALQVRPDLGITCTGPWPPYSFSDPDLSGVSP*	280

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Bacillus-megaterium_gvpL	MGELLYLYGLIPTKEAAAIEPFPSYKGFEDGEHSLYPIAFD QVTAVVSKLDADTYSEKVIQEKMEQDMSWLQEKAFHH HETVAALYEEFTIIPKFKCTIYKGEESLQAAIEINKEKIEN SLTLLQGNEEWNVKIYCDDTELKKGISETNESVKAKKQ EISHLSPGRQFFEKKKIDQLIEKELELHKNKVCEEIHDKL KELSLYDSVKKNWSKDVGTGAEEQMAWNSVFLPSLQI TKFVNEIEELQQRLENKGWKFEVTGPWPYPYHFSSFA*	281
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpL	MNDALYLFCFARAEPLAPAWAKRAPGEPRLQLLHEGN LAAVLCDVSRSEFAGADAERRLADPAWIAGRVAHVAA AIEWTMRYSPVIPAQFGTLFSGAGRVIALLMESCHAHIGR VLDHVEGKTEWAVKGWLDQRQAADSQAALLRADEPE SAARTAGARYLRERQLQARAGQNLRDWLEQSVPPISAR LQRHAVEMCSRPCRASDSEHEIVANWAFLVRNRDVPAF RRQAEIDADEFATWGLHFDGSGPWPYPYSFCAPLTEETT WSG*	282
Chlorobium luteolum DSM 273_gvpL	MPCRLTVTWKSLRTAGLLPTAKGIQGRTERMAQNILYV YCIVRQLPGADIVARYPDLVFIEAGSAYVAACYVSPLEY SDASMKLKLADDEWLDRNAREHLSVNVMMMAQQTIIPF NFGTIFKSRESLSGFLGDYGRKLDESFDALLEGREEWAVK AYCNESFLLKNLHLESPAIAAIEQEIQAAASPGKAYLLKK KKEAMSASALEGVHQGHAKAVWGELAALSKEHVLNR LIPEDVSGVDGRMIVNGVFLIANTDVGAFIRTTEDLGER YRDAGVFLDVTGPWPYPYDFVDIPY*	283
Dactylococcopsis salina PCC 8305_gvpL	MLYTYCLIASSPSALSLPSGFRGELQLIKQGAIAAIVEAE LPLEELEENDQKLIQAVIHDAVICEIFQQIPLPLRFGTY FPTEKDLLEHLDFKAEKYQKKLQEIQDKVELTLKLTPLP FSTENASPMKQKGNLYLAKKQRYQEQTNYQSQQQAE LNQLQTQINQDYPQFIHGEPKENIERFYLLIKERDRSVFS EQLEQWKKDFPTWTIEVSDPLPPYHFIE*	284
Desulfobacterium vacuolatum-DSM 3385_gvpL	MEKKKAVYLYCVTRANKFNAPGITGIDANTPVCFEHLE NFVAVYNIIPLNTFVGTSAEENMKNIDWIGPRAMRHN VIERMMQESSVYPARFATLFSSMENLRETLHLKSGLISR FLNQTQHKCEYSLKGFINKQLLEFLIKTKFKQEKKQLD GLSPGKKYFAQHQNKKVETGINQWIKRRCGIFLDHLT KRNPEVSPRELFTKTEKNNLEMMFNLAFLIHNSKSAF LQEISQAEKEFSQTGISLVVSGPWAPYSFCKTTRGEGL*	285

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Desulfomonile tiedjei DSM 6799_gvpL	MSNVLYLFCLARTGLVDHIEGTGITGTEDLILKNFSGVT AVTCEVPEDDFSGESAEIKLQDLAWVGPRAVRHRIIEE IMQYSPVFPAPFGSLFSSEKRLGTLIESNIDAIREFLDHTA DKQEWSVKGLVCKSKAVDEIFTGKLKILSETLSSSPAGM RYFKERQMRSEAEKELSGKVKAAC TVVGEKLLACSNN FRQRKNISFGKAEGDKQLVVNWAF LVDHSRISYFLDQV EHANSNYQAGGLAFEC SGWPWPYSFCPSLHMEPTR*	286
Desulfotomaculum acetoxidans-DSM 771_gvpL	MNLIDDCKAKYIYCIGENPGNWPSEVMGVEGSLVYHV VYRDIAAVVHDCAEQPYNSDDNNKVIDWVLGHQLVVD KACSCYSSVLPFTFNSIVKGKEDLSSHEILVNWLEDNYD NFKLKLKGKIKGKEYSVQLFLDKQVSLSLQSESDILEL QVELLGSAGKAYFVQEKINKKIGELMANRADSYCRQF YHEISSVVSECKLCKLKQAGRNEIMIINLVCLAGDNEVE VLGDVLEKIKSNDIAIKIKFSGPWPAYSFV*	287
Enhydrobacter aerosaccus strain ATCC 27094_gvpL	MLYVYGIADNAFEVLRGAGLLNSDVFAVPAGCLAAAA SKLAQGGIETTPQGVWRHEQVLRQLMQDHAVLPLRFG TICRDRETLTDRLMEASDDLVRGLGRVRGKVEIALRIVD EREHEAHPVPSETPTVDAIGGGRTAYLRARRRHAAE MGREARAERVGKMLSAYIDVGAEDLVCSVAPEGDHAV SVSCLLRDQLATLQAALERFQSDHPAIGLSWTGPWTP YSFVAPSLFGVGLP*	288
Legionella drancourtii LLAP12_gvpL	MNKALYLFCLTPASDLPMMEGELLPNFSPLFIHPFQTFN AILSWVPAKEYQEQTSDSNLINTEEFMQRVFFHEL VVEK IMRDEAVFPIGFGTLFSSIASLEEQILTHQTLISSCLANLN QKDEYAVRVYLNQDKALESLLSVMLQERESSWASSSPG VQYLKKQQLHNEIQRNLNQHLGGMLDEVLSMFQRHAT DFKSRENTAQSSDIHGTSILHWAFLIPRVVSSIFKEQVDL MNAKYNPFGLHFVLTGPWPAYSFCTLQSVEAP*	289
Lyngbya confervoides BDU141951_gvpL	MRWHRSEAVISYCDLSMIYLYALCPNSTETNNLPEGIGT AQVEVLTVGTLGAVIERDVIDIAQIQKDDAQLMAAVLA HDRILSHLFTYSPLLPLRFGTQFSNSEAVTTFLKTQGETY RQKLSHLQDRAEYLVKLIPQLDLPAIASDLKGREYFLA KKQRLQDHTAALNQQADELQTFLTDLATQDIPLVRSAP QDHEERLHVLLSRD TDTTEQVIMTWQEQLPNWQVVC EPLPPYHFAA*	290

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Octadecabacter antarcticus 307_gvpL	MKRLYVYGIVGATSFDDPLPNGHDEASVFALVSGDIAV AVSFVERSAVEASAANVWLHDNVLSALMTRYAVLPMR FGTIAVGATQLLEGIVKRQKQLMKDLMRLNENVEIALHI SGKNWEKVNQKVTCKNTDQAITQGTAYLLGRQQSLYG SDKTQLLVQNVRAIRSGLDPLMKDVIWPIDKPQALPFK ASCLINRNDVASFVQIVNDIAAQNLDAVVTCTGPWAPY SFVGKSGVEGET*	291
Octadecabacter arcticus 238_gvpL	MTKLYVYGIVGATHFDVKLPNGHDEAPVFAIVSGDLAV AVSSLSAVEASAANVWLHENVLSALMEGHAVLPMR FGTIATGAAQLLGDIVKRRGQLMKDLTRLDGKVEIALRI SGKNREKVEQRIAGQIVDTNVTQGVAYLQEKQQNLYGS FYTQSSVQCARRAIRSQLDPFIVEAIWPTDEPQMLPFRAS CLIKKGDIARFVQTVDDVVVKVSDIRVTCTGPWAPYSF VGQSGSEAET*	292
Pelodictyon phaeoclathratiforme_gvpL1	MVAIQERLIYIFCVTSEPPLLQQYQLQKGICVVDVDGLF VTTMDVTDNDFAEENQLQSNLSDVWLDTKVREHLVDI TSIMQHVKSLIPNFGTLYKSESSLMQFIKYAEFFKKNL VYLEEKEEWAVKLYCNKNKIVENITHLSKKVSDINALIQ NSSIGKAYILGKKKNEIIEIINIYNTYSKKIFTKFSILSEE FRFNPIPNNETLEKEDDMILNVVLLLNKANVESFIETSDQ LIIHQHQNIGLNIEITGPWPCYSFINISH*	293
Pelodictyon phaeoclathratiforme_gvpL2	MPLIYAFDSINYIDSFSSYVDAISLKS KIKLEIISTSTLSAI VSRTTDEKKQACQNDVMYATIIGDIAAKYSILPMRYGS IVSSPFDVTELLKNHNETFVTIHKITDKEEYSLRILYSHQ DKEKNNIEDLFDLPQNVDPDILHGNTDSKKYLLNKYIKHL SEEKRLQYIDKIQSIVACNLQKITDLIVYNKQTTTGFI VDVFMIERSKKSELLDLVIQMQLTFSEHNVVLSGPWPPYN FSNINIG*	294
Psychromonas ingrahamii 37_gvpL1	MKNSNHSGLDPNQALYL YCFVHADSIQSVTSQAIEKDSP VFIYQWQDIAAVLSHVPTS YFTGYDDEEPEQTARILPRT QLHEQVIEEVMRQSPVFPAQFGTLFSSQESLEQEISQQYL AITHTLKEVSGSVEWAVKGVLDGRGVAEKALYSQQLTE QQNSLSSSPGMRHLQEQLRRETQSKLNSWLHQLYTDI ATPLSELSGDFFQRKIPSSIEEGKEVILNWAFLVPESAGD DFHAQIDKLNQRLNSFGLVIQCSGPWPPYSFCNQSS*	295

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
<i>Psychromonas ingrahamii</i> 37_gvpL2	MKNSNHSGLDPNQALYL YCFVHADSIQSVTSQAIEKDSP VFIYQWQDIAAVLSHVPTS YFTGYDDEEPEQTARILPRT QLHEQVIEEVMRQSPVFP AQFGLTFSSQESLEQEISQQYL AITHTLKEVSGSVEWAVKGVLD RGVAEKALYSQQLTE QQNSLSSSPGMRHLQEQLRRETQSKLNSWLHQLYTDI ATPLSELSGDFFQRKIPSSIEEGKEVILNWAFLVPESAGD DFHAQIDKLNQRLNSFGLVIQCSGPWPPYSFCNQSS*	296
<i>Serratia</i> sp. ATCC 39006_gvpL	MTMNTEAQTEQAIYLYGLTLPDLAAPPILGVDNQHPINT HQCAGLNAVISPVALSDFTGEKGEDNVQNVTLTPRIC RHAQIIDLMAQGPVYPLPFGTLFSSQNALEQEMKSRAT DVFVSLRRITGCQEWALEATLDRKQAVDVLFTGLDSDG RFCLPEAIGRRHLEEQLRRRLTTELSDWLAHALTAMQ NELHPLVRDFRSRRLDDKILHWAYLLPVEDVA AFYQQQ VADIVERYEAYGFSFRVTGPWAAYSFCQPDES*	297
<i>Stella vacuolata</i> -ATCC- 43931_gvpL	MLYLYAVLEALPAARTLPAGIGGGELLFVEAFELVCAA SETPERAIAPEPTQVWRHQVVEALIDCAAALPLRFGTL VEDAVACRLLTRHREALCAQLDRVRHCVEFALRVSG REEVGSDHVIGGGPGVSYMRALARREASWPPSTGT DGLAAHAADRLLSRSASQPD LRASFLVLKPNVA AFLA DVSALQRM RPD LGITCTGPWPPYSFSDPDL SGMSP*	298
<i>Thiocapsa rosea</i> strain DSM 235 Ga0242571-11_gvpL	MDAFYCFCFAPACLASDLRFDDCGWEDPIEIRRLAGLD VILSRVPLGRFAGAEAEQRLADLEWLVPRAQAHDRVIT RTMERSTVFPLTFATLFSSLPALALEVAARRRALLDFFE RMAGREEWAVKVSMDRERV IATRMQSLYPEGGDVPAG GRGYLLKQRRRGEAEQAIGPW LKGQIGCLDEALRPSCE TLLIRPLRDEMVASRACLVARDLGPSLSEAIERSREAF DQGLDLHCSGPWPLY SFCGTP*	299
<i>Trichodesmium erythraeum</i> IMS101_gvpL	MSYYVYGFLYLPESCLALPKGMEKEVELVPYQNIAAVV EANVSIEAIQETEEKLLEAILAHDRVVREIFQQVSMLPLR FGNAFALRENIINDLQNNQQQYLNILTKLQQQAEYTITF TPVSYPSTLEVSKVRGKAYLLAKKQFEQQQAFQTKQR QQWENIRQLIFKNYPKAVFRDSTESKIKQVHLLANRDA RVITTEELSTWQTECSYWQITLSEQLPPYHFV*	300
gvpN		

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Anabaena-flos-aquae_gvpN	MTTCKVNHKRAVLRRLRPGQFVVTPAIERVAIRALRYLK SGFPVHLRGPAGTGKTTLAMHLANCLDRPVMMLFGDD QFKSSDLIGSESGYTHKKVLDNYIHSVVKLEDEFKQNW VDSRLTLACREGFTLVYDEFNRSRPEVNNVLLSALEEKI LSLPPSSNQPEYLSVNPQFRVIFTSNPEEYAGVHSTQDAL MDRLVTISMPEPDEITQTEILIQKTNIDRESANFIVRLVKS FRLATGAEKTSGLRSCLMIKVCADNNIPVTTESLDFPDI AIDILFNRSLSMSESTNIFLELLDKFSAAEELEILNNRVTG DNDFLIDNSQFVSQQLAGQPN*	301
Ancylobacter aquaticus strain UV5_gvpN	MTSEAASKDPISLLSGFGAGAASSGPKAGGRSTPSALTP RPRTGFVEAEQVRDLTRRGLGFLNAGYPLHFRGPAGTG KTTALHVAQAQLGRPVIITGDNELGTADLVGSQRGYHY RKVVDQFIHNVTKLEETANQHWTDHRLTTACREGFTLV YDEFTRSRPETHNVLLGVFEERMLFLPAQAREECYIKVH PEFRAIFTSNPQEYAGVHASQDALADRLATIDVDYPDRA MELAVASARTGMPEASAARIIDL VRAFRASGDYQQTPT MRAGLMIARVAAQEGFEVSVDPRFVQLCSDALESRIFS GQRAEEVAREQRRALHALIDTHCPSAAKPRARRAGGA VRASIEGAQS*	302
Aphanizomenon flos-aquae NIES-81_gvpN	MTKTNHKRAVLRVRPGQFVVTPAIEQVAIRALLYLKSG FPIHLRGPAGTGKTTALHLAHCLDRPVMMLFGDDEFKS SDLIGSESGYTHKKLLDNYIHSVVKVEDEFKQNWVDSR LTLACREGFTLVYDEFNRSRPEVNNVLLSALEEKILSLPP SSNQPEYLSVSPQFRAIFTSNPEEYCGVHSTQDALMDRL VTINMPEPDEITQTEILIQKTNIQKESAHLIVRLVKSFR TGAEKTSGLRSCLMIKVCADNNLVAEPENSFFQEIAME ILSNRTHLSVNESTDIFLDVISQFSNKEIEILNDAELGSLPT MDTLANTDLGNDVPLEKEASDYVIQQKNNEFKGFQKPS TKVLN*	303
Aphanothece halophytica (strain PCC 7418)_gvpN	MTTVLHARPKGFVSTPTIDRISRRAWRYLQSGFSIHLRG PAGTGKTTLAMHLADLLNRPIMLLYGDDEFKSTDLIGS NTGYTRKKVVDNYIHSVVKKEEDELQQWVDSRLTMAC REGFTLVYDEFNRSRPEVNNVLLSALEEKLLVLPPDSHR SEYVRVSPNFRAIFTSNPEEYWGVTGTQDALLDRVVTIN VPEPDLETQREIIVQKVGINADDGDMIVNFVRNFRDRAE MENS SGLRSCLMIAQVCHQHEIPVQTSNEDFQDICYDIL TSRCPLSTQESISLLEQLFREYELELVVEDEDEDVPSVIVE	304

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	GETEDLSSDEKPHLRLSHPFNGTEND*	
Aquabacter spiritensis strain DSM 9035_gvpN	MSTEPAPLVSPSQDVETTPQRPAPAEALAVGYRLSARPASATLTPRPRADFVETDQVKDLTRRGLGFLRAGYPLHFRGPAGTGKTTALHVAQAQLGRPVIITGDNELGTADLVGSQRGYHYRKVVVDQFIHNVTKLEETANQRWTDHRLTTACREGYTLVYDEFTRSRPETHNVLLGVFEELKILFLPAQAREECYIRVHPDFRAIFTSNPQEYAGVHASQDALADRLATIDVDYPDRGMELAVASARTGLGETEAARIIDLVRFRASGDYQQTPTMRASLMARVAAQEGLRVSIDDPGFVQLCMDALESRMFSGARLEAATRETSRAALLALLAVHCPSEAPIVRVTAARRAKKADAS*	305
Arthrospira platensis NIES-39_gvpN	MTTVLRAVPKGFVNTPAIERITVRALRYLQSGFSVHLRGPAGTGKTTALHLADLLNRPIMLIFGDDELKSSDMIGNQTGYTRKKVVDNFIHSVVKLEDSLKQNWIDSRLLACREGFTLVYDEFNRSRPEVNNVLLSALEEKLLVLPPNNSRSEYIRVNPHFRAIFTSNPPEYCGVYSTQDALLDRLITMNPPEDEATQQEILIKVAVTPPEAQITIVTLVQQFREATHAIAPSKIQTVARQQTNADKASGLRPSMLARICQEHNIPIVPIPDFQEVCRDILLSRAIGDITELESRLHQIFDHLSGLENDQIIALPPREELTTSSVPNNLSDETEQKIYTYIKSDGARVSEIEIALGLNRVQTTDALRSLLRKSYLTQQDNRLFVVYEGD*	306
Bacillus-megaterium_gvpN	MTVLTDKRKKGSGAFIQDDETEKVLRSALSYLKSGYSIHFTGPAGGGKTSALARALAKKRKRPVMLMHGNHELNNKDLIGDFTGYTSKKVIDQYVRSVYKKDEQVSENWQDGRLL EAVKNGYTLIYDEFTRSKPATNNIFLSILEEGVLPLYGVKMTDPFVRVHPDFRVIFTSNPAEYAGVYDTQDALLDRLITMFIDYKDIDRETAILTEKTDVEEDEARTIVTLVANVRNRSKDENSGLSLRASLMATLATQQDIPIDGSDDEFQTLCDILHHPLTKCLDEENAKSKAEKIILEECKNIDTEEK*	307
Bradyrhizobium oligotrophicum S58_gvpN	MLRSDRAAIAGGQGRGSRAQGDVARNDAAAGSRAAIAQISRPDADNAALSPAPRTDLFENPQLASMAARALTYLNAGIPVHLRGPAGTGKTTMAMQLAARLGRPVLTLTGDDGLTAAHLVGREIGTKSRQVVDYVHVSRRVETETSSMWCDAVLAQAVVEGLTFVYDEFTRSPQANNPLLSVVEERILIFPAGSRKERLVHAHPEFRILTSNPPEYAGVSRPQDALLDRLITFDLDDYDRETEIGIVSNRTGLAYAEAGVIVD	308

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	LVRGVRRWPKAHHPPSMRSAIMIARIVARELITPSVDDP RFVRLCLDVLA AKAKPTDRDDRDRFAATLLRLMNNHC PAGAI DGG*	
Burkholderia thailandensis sp. Bp5365 strain MSMB43_gvpN	MEASAEFVQTPAVRNLTALTYLGAGYGVHLAGPSGT GKTTLAFHIAAQLGRQVVL MHGDELGSADLVGRGAG YRRSRVVDNFIHSVVKTEEEMTTTWIDNRLTTACQHGL TLIYDEFNRSRPEANNALLPVLSEGILNLPNRMTGAGYL TVHPGFRAIFTSNPEEYVG VHKTNALMGRITIQVGHY DRETEVEIVRARS GIARADAERIVDLTRRLRDADDNGH HPSIRAAIALARALSYCGGEATPDNAGYVWACRDILGV DLEQDARTRSQAGRRTKARR*	309
Chlorobium luteolum DSM 273_gvpN	MRAAVNDNEMNTVLAPRPMANFVETEYIRDITERGLTY LKAGFPVHFRGPSGTGKTTVAMHLAGKIGRPVVVIHGD SEYKTS DLIGSEQGYKFRRLNDNFIHSVHKYEEDMSKQ WVNNRLSIAIKKGFTLVYDEFTRSRPEANNILLPILQEKM LSTSASNEEDYYMKVHPEFRAIFTSNPEEYAGVNRTQD ALRDRMVTMDLDYFDYETELRVTHAKSEL TLEDSEKIV QVVRGLRESGKTEFDPTVRGSIMARTLHIMQVRPEKTN DAVRKVFQDILTSETSRVGSKTNQEKVRAIVNDLIEAYL *	310
Dactylococcopsis salina PCC 8305_gvpN	MTTVLHARPKGFVSTPTIDRISGRWRYLQSGFSIHLRG PAGTGKTTLAMHLADLLNRPIMLLYGDDEFKSTDLIGS NTGYTRKKVVDNYIHSVVKEEDELQQWVDSRLTMAC REGFTLVYDEFNRSPPEVNNVLLSALEEKLLVLPPDSNR SEYVRVSPNFRAIFTSNPEEYWG VHGTDALLDRVVTIN VPEPDLETQQEITQKVGINANDGEKIVNFVRQFRDRAA VKNSSGLRSLMIAQVCHQHEIPVQTSDEGFRDICYDILS SR	311
Desulfobacterium vacuolatum_DSM 3385_gvpN	MSASMSSMKETRQRMSAPEQDNVVPEAGSDFVETPYV KDITDRALAYLHVGYPVHFGPAGTGKTTLAFHVA AKL KRTVMLIHGDDEFGSSDLIGKDSGYRKAKVVDNYIHSV VKTEESMNTVWADNRLTIACQQGCTLVYDEFTRSRPEA NNAFLSVLEEKILNIPSLRDIDQGYLQVHPEFRAIFTSNPE EYAGVHKTQDAMMDRLITITLDHFDRDTEVQVTMSKS DLPQKDAEKIVDIVRKL RKTGVNNHRPTIRACIAIGKILK HMGGGASKDNFVFKQICRDVLNVDTTKVTRDGEPLLP KIDELINSL*	312

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Desulfomonile tiedjei DSM 6799_gvpN	MNGAELRIASIEETE VITANNENIVPEAGDRFVNTPHVEEL TARAMAYLEVGYSVHFSGVAGTGKTTLAFHAAAKLGR PVILVHGDHEFGSSDLIGRDAGYKKSRLVDNFIHSSVVK EEEMRSLWVDNRLTTACRDGYTLIYDEFTRSRPEANNV LLSILEEKILNPSLRRTGEGYLEVHPSFRAIFTSNPEEYA GVHKTQDALMDRIITINVDHYDRETEIEITRAKSGVCKQ DATVIVDIIRELRLLGVNNHRPTIRAAIAIARVLAHTGEH ADQHNSVFQWLCKDVLSTDTVKVSRGGSPLMAKKVEE VIRKVCGRGTGGRSGKPVGSKEETSE*	313
Desulfotomaculum acetoxidans_DSM 771_gvpN	MQLNGLDKNSIINPVVLSDFVVTDYISNVVDRALAYIKA GFAIHLRGRSGTGKTSIAMYISSKLNRPVLVIHGDEEFRT SDLIGGRYGYRIRKTIDNFVQSVVKVEEDLVERWVDSR LTTACKNGYTLVYDEFTRSRPEANNILLSVLQERLLDIS VARGAEEGYVKVHPDFTAIFTSNPEDYAGVYGSQDALR DRMVTLDLDNYDKETEISIIKSKSKLSREDSESVVNILRD LRELGDCEYGPTIRGGIMIAKTLQVLGAPVDKNEMFR QICEEVLASETSRAGNLQALRKVRKVINELFNKYA*	314
Dolichospermum circinale_gvpN	MSITKVNHKRAVLRLRPGQFVVTPIAERVVIRALRYLRS GFPIHLRGPAGTGKTTLMHLANCLDRPVMMLFGDDQF KSSDLIGSESGYTHKKLLDNYIHSVVKVEDEFKQNWVD SRLTLACREGFTLVYDEFNRSRPEVNNVLLSALEEKILSL PPSSNQPEYLSVNPQFRVIFTSNPEEYCGVHSTQDALMD RLVTINMPEPDEITQTEILIQKTNIGRESANLIVRLVKSFR LATGAEKTSGLRSCLMIAKICADHDIPASTEDLDFREIAI DILFNRAQLSISESTDIFMGLLEQFSAAEIKVLNDTHFPTD ELLINNSQFITQELVTQPNTTELATDIPQELRKTEQN*	315
Enhydrobacter aerosaccus strain ATCC 27094_gvpN	MSMDQAAEEIGVVTTIEPRPRADFVRTQSVEATARRALG YLNAGFSVHFRGPAGTGKTTLALHLAALLGRPMVMITG DEEMLTSTLVGTQHGYHFRRVVDRIHTVTKTEETADK RWADHRLTTACREGYTLIYDEFTRSRPEANNVLLSVLEE GLLVLPANQNEPYIKVHPNFRVIFTSNPQEYAGVHDA QDALGDRIVTIDMGHADRELELAIAAARSGLPPTQVAPI VDMVREFRETGEYDQTPTLRTSIMICRMMSQERLAPTIE DQQFVQICMDILGGKSLPGGKGDKNKRAQQQKMLLSLIE HHCPARSFTSVGEV*	316
Isosphaera pallida_ATCC-	MDYESTALQLKPRPDFVATPWVRELADRALGYLTAGY	317

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
43644_gvpN	PVHFSGPAGTGKTTLAMHLAALVNRPVVLLHGDDEFGS SDLVGDHLGFRSTKVVDNFIHSVVKTEQSVSKTWVDHR LTTACRHGFTLIYDEFNRSRPEANNILLTILEERLLELPPI AGGRDGSGLRVHPEFRAIFTSNPEEYAGVHKTQDALL DRMITISMGGHDEATEITEITAAKSGLSRDEAARIVELAR AVRALKPLRHPPTIRSCLMIKVAALRKVPIDPNDALFL AICRDVLRIDALPVDDPEATFAELIRRVFAPTPAVAPPRV PTTGFAANRVVPIPRRPLAASASPPPGANGHAHLR*	
Legionella drancourtii LLAP12_gvpN	MMTQENNGSLTDSKNNDKLIRFVNNRSDNILLEASEEFT ETPHIRGISERALAYLDIGYPIHLLGPAGTGKTTVALHIA AQLGRPVLIIHGDDFTGADLVGRGTGYHHSKLVDNFIH SVLKTEEEMTTMWTDNRLTTACEQGYTLIYDEFNRSRA EANNALLSVLSEGILNLPGRRERDGIGYVDVHSNFRAIF TSNSEEYVGIHKTQNALADRLIAIKMDYPDQQSEIQIEK KSTLPRKDIEIIVNLARELRKSEKRPSIRGCIAIARVLAY HNRHAHADDPFQAVCQDIFGISKEFLKQLLHPMDSGLQ KRSEKNQESIKKYKTKNQKL*	318
Lyngbya confervoides BDU141951_gvpN	MSTVLQARPRNFVSTPAVERIARRALRYLQSGYSVHLR GPAGTGKTTALHLADLLSRPIMLVFGDDEFKTSDLIGN QSGYTRKKVVDNYIHSVVKVEDEL RHNWVDSRLTLAC REGFTLVYDEFNRSRPEVNNVLLSALEEKLLVLPPSGHR PEYLRVNPHFRAIFTSNPEEYAGVHGTQDALLDRLITIH MPEPDEL TQQILIQKVGIEPADALMIVRLVKAFKSQMG NHSATSLRPSLMIANICHEHGVAMMTEDADFRDVCSDV LLSRVTNELSPATHLWDLFNELTASADVLGPESNSTDV SPQPEADKPVETKSGKSKSTTKSKAKESAKASEEADAE GDDASAPELDEIESSILTFLTAREASLSEIESELSTRFK AVDALRSLVEAGYLQKQNGAGKPAIYGLVPEES*	319
Microcystis aeruginosa NIES- 843_gvpN	MTVTETQTRRAVLSLRPGQFVVTSPSIDQIATRALRYLNS GFSIHLGCGPAGTGKTTLAMHLANCLARPVMLIFGDDDDF TSSDLIGSQSGYTHKKLMDNYIHSVLKVEDELKHNWVD SRLTMACREGFTLVYDEFNRSRPEVNNVLLSALEEKILT LPPTSHQPDYLVNSQFRAIFTSNPEEYCGVHATQDALM DRLVTINMPEPDQLTQTEILAQKTGIGREDALFIVNLVKT FRVKTATEKTSGLRSCLMIKVCASHDIAANSADSDFRD ICADVLLSRTNLSVDKSRAILWEILEDNPLESLSFLEEEEP SDAQVSTSEPSTGNQSLKAIQSLLRGNLPQRKD*	320

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Nostoc punctiforme ATCC 29133_gvpN	MTTVLNASPQRFVNTPAVQRIAQRALRYLQSGFSIHLRG AAGVGKTTLAMHLADLLNQPIILLFGDDEFKTSDLIGNQ LGYTRKKVVDNFIHSVIKVEDEVQRHWVDARLTACK GFTLVYDEFNRSHPEVNNVLLSVLEERLLVLPTNQHRAE YIRVHPQFRAILTSNPQEYCGVHATQDALMDRVITIDMP TPDELSQQEIVVHKTGIDSEKAEVIVRIVRTFWSRSGSGQ GGGLRSLMIAKICHEHEISVNPDPSPFQDICADILLSRT NQPLIEATRLLEEVLSEFYHRINTQSQPSEIIPNNQNQIVL EQRVPYEHEVYNYLCNSPGRRFSELA VELGIDRSQIVAA LKSLREQGVLVQMKGNAESPSISQTVAFDSGHLINK*	321
Nostoc sp. PCC 7120_gvpN	MTLTANNKKRAVLRVRPGQFVVTPEAQVAIRALRYLT SGFAIHLRGPAGTGKTTLAMHLANCLDRPIMLIFGDDEF KSSDLIGSESGYTHKKLLDNYIHSVLKVEDEFKQNWVD SRLTLACREGFTLVYDEFNRSRPEVNNVLLSALEEKILTL PPSSNQPEYLHVNPQFRAIFTSNPPEYCGVHSTQDALMD RLVTINMPEPDELQTTEILAQKTALNRADALLIVRLVKA FRSRTGGKEKTSGLRSLMIAKVCAEHNILVSPQSSDFREI CADVLFNRTNWSASEAATIFLELLNHLDLQQIEEFKNSII PEDTDAIAEGGFPTIIDSHFGTLDSEVLEQPGVEDSIPFEQ EIYLYLQQYKSAALALVQQEFELSRVATNALNSLEQK GLVSKNNHVTIEEPNQS*	322
Octadecabacter antarcticus 307_gvpN	MNSNLRATNSGGPDISKTMPEAREDFVQTESVKSISRR ALAYINAGYSVHFRGPAGTGKTTMAMHTAALLGRP VVLITGDEEMITSNLVGAESGYNYRKVTDNYIHTVSKIEESS DRSWNDHRLTTACREGYTLIYDEFTRSRAEANNVLLSV LEEGILVLPQNRGEPFIKVHPNFRVIFTSNPQEYAGVHE AQDALSDRIVTIDIGEADRELEVSIASSRSGLEVAKTEPIV DMVRAFRDTEYDQTPTLRACIVICRMVANEKLNNTID DPFFVQICLDVLGSKSTFGGKEHDKRTQQRKLLLDNLK HYCPSKVSTKPSAKDDESKSTLIQVSSRGS*	323
Octadecabacter arcticus 238_gvpN	MMPEARKDFVQTDVSVKSRRALAYINAGYSVHFRGP AGTGKTTMAMHTAALLGRPVMITGDEEMVTSNLVGA ESGYNYRKVTDNYIHTVSKVEESSDRSWNDHRLTTACR EGYTLIYDEFTRSRAEANNVLLSVLEEGILVLPQNRGE PFIKVHPDFRIVFTSNPQEYAGVHDAQDALSDRIVTIDIG AADRELEVSIASSRSGLEVAKTAPIVDMVRAFRDTEYD QTPTLRACIMICRMVANEKLNPTIDDSYFVQICLDVLGS	324

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	KSMFGAKEQGKRTQQEKLLEDNLSHHCSPPPSKPSAK EAEAKPRSIQATSRGPA*	
Pelodictyon phaeoclathratiforme_gvpN	MRRQGCDSMNNTVLEPKMPNPFVETDYIRDITSRGLTY MKAGFPVHFRGPSGTGKTTVALHLASKIGRPVVIHGD EYKTSDLIGSEQGYKYRRLDDNFIHSHVHKYEEDMTKQW VNNRLTIAIKKGFTLVYDEFTRSRPEANNILLPILQEKMM STSSSNEEEYYMKVHPEFRAIFTSNPEEYAGVNRTQDAL RDRMVTMDLDYFDYETELMITHAKSGMSLDDAEKIVKI VRGLRESGKTEFDPTIRGSIMIAKTLNVLNARPDKTNELF KKVCQDILTSETSRVSGSKTNQERVRGIVNELIDLHS*	325
Phormidium tenue NIES- 30_gvpN	MNTVLQARPRNFVSTPTLERTSIRALRYLQSGYSIHLKG PAGTGKTTALHLADLLARPIMLLFGDDEFKTSDLIGNQ SGYTRKKVVDNYIHSVVKVEDEL RHNWTD SRLTLACR EGFTMVYDEFNRSRPEVNNVLLSALEEKLLVLPSPNNR AEYIRVSPHFRAILTSNPEEYCGVHGTQDALQDRLITINM PEPDELAQQQILVQKVGIDSSAALQIVQLVKAFQSAVAP DMVSSLRPSLMIATICHDDHILPLAENADFRDVCSDILLA RSKEPAPDATRHLWNLFNRFVVSQAALVNDLSLKPEAH PTARFHGEEEDDAPLQPLEALVESDIDDVAVEDQPVIGP QDLQGETLPEAVIPEPQGETVVETPAEAEALPEEIARVQ VSPDDIETRIFDYLDATGTASLVNIEAALDLNRFQAVNA VKSMLDQGLIEKQETDQGLQGYQLSSN*	326
Planktothrix agardhii str. 7805_gvpN	MTTVLQARPKGFVNTPTIEQLTIRALRYLQSGFSLHLRG PAGTGKTTLAMHLADLLNRPIVLIFGDDELKSSDLIGNQ LGYTRKKVVDNFIHSHVVKLEDEL RQNWIDSRLTLACKE GFTLVYDEFNRSRPEVNNVLLSALEEKLLVLPNNRSRSE YIRVNPHFRAIFTSNPEEYCGVYGTQDALLDRLITIDMPE PDDETQQEILIQIGISPEDAKNIEIVKIYLEITTQKKEIKP VQNGKAARPHIDKASGLRPGLIIAKICHEHDISIQENNQD FIKVCADILLSRTNLSL TEAQNKLEKVIKTVLTDGDTSN NSFLPPSETQLTENNSLEIEEQVYQYLQKTT SARVSEIEV ALGLNRVQTTNVLRSLLKQGHLLKQDNRFFAVNKQGE LIQP*	327
Planktothrix rubescens_gvpN	MTTVLQARPKGFVNTPTIEQLTIRALRYLQSGFSLHLRG PAGTGKTTLAMHLADLLNRPIVLIFGDDELKSSDLIGNQ LGYTRKKVIDNFIHSHVVKLEDEL RQNWIDSRLTLACKEG FTLVYDEFNRSRPEVNNVLLSALEEKLLVLPNNRSRSEYI	328

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	RVNPHFRAIFTSNPEEYCGVYGTQDALLDRLITIDMPEP DDETQQEILIQKIGISPEDAKNIIIVKIYLEITTQKKEIKPV QNGKAARPHIDKASGLRPLIIAKICHEHDSIQENNQDFI KVCADILLSRTNLSL TEAQNKLEKVIKTVLTDGDTSTNS FLPLSETQLTENNSLEIEEQVYQYLQKTTSARVSEIEVAL GLNRVQTTNVLRSLKQGHLLKQQDNRFFAVNKQGELIQ P*	
<i>Psychromonas ingrahamii</i> 37_gvpN1	MSIENLNNVSEIKIEQSDDDHIYPEASEDFVETPYIKEVTE RAMLYLDAGYPVHFAGPAGTGKTTLAFHIAALRQRPVT LIHGNHEFGTSDLIGKESGYRRHRVVDNYVHSVVKEEE ELQSLWSDNRLTTCCRN GDTL VYDEFNRSTPEANNVLL SILEEGILNLPSSRSDGYLEVHPQFRAIFTSNPQEYAGTH ATQDALVDRMITIMLHYPDRHTEVRVAIAKSGINSDEA GSIVDIVNEFRELCSKIVSSGPKTMPTVRASIAIARVLV QKGEHAFRDNTFFHRICRDVLCMYTQQVVSFSNRSVLDK QLEDLIMKFCPATYKSSGSKIRA*	329
<i>Psychromonas ingrahamii</i> 37_gvpN2	MSINNLNISTIKIEQPENDNIYPEASAEFVQTPYIQEVTER ALLYLDA GYPVHFAGPAGTGKTTLAFHIAALRKRPTLI HGNHEFGSSDLIGKESGYRRHRLVDNYVHSVVKEEEEEL KSLWVDNRLTTCCRN GDTL VYDEFNRSTPEANNVLLSI LEEGILNLP SLRSMGDGYLEVHPSFRAIFTSNPQEYAGTH ATQDALVDRMITIMLNYPDRDTEVRVAVAKSGISNEEA GFIVDIVNEFREL SNHKSLSGQKSMPTVRASIAISRVLQ KGEHAFRDNTFFHRVCHDVLCMYIQKISPSNRSFLDKQ LEV LIGKFC PAAKSALVPKVVK*	330
<i>Rhodobacter capsulatus</i> SB 1003_gvpN	MTIPRDLPWGDARTPLFEDEELRSLLDRAEIYLRGIAIH FRGPAGVGKTTALHLAQRFPVTFVGVNDWLGRADI FGRDLGETVSTVQDHYISSVRRRAERKSRIDWQEAPLAR AMRDGHVLVYDEFSSRPEANAALLSVIEEGVLPLSDPA AGRSHIV AHPDFR VILTSNPRDYVGQAVPDALLDRMIT FSLDGMSFETEVGIVATAARTDPADARAICALIHLLRAE KPGTMEISMRS GIMARLARAAGVAPDPADPVFVQICAD VLGTRMRGSDIDDVMALLLRPDPAPAACAGGAR*	331
<i>Rhodobacter sphaeroides</i> 2.4.1_gvpN	MTVLSPSLPHAAGIDAALVENPWLGLRRSGRYFQNAET EALFARALGYARAGVCVHLAGPAGLGKTTLALRIAQAL GRPVAFMTGNEWLGSRDFIGGEIGQTVTSVVDRIYQSV RRTEQSARIDWKESILGQAMRCGQTFIYDEFTRASPEAN	332

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	AALLSVLEEGVLVSTDGASRHQYIEAHPDFRVLLTSNPH EYQGVKAAPDALIDRMVTLRLEEPSAPTLAGIVALRSGL DPATARRIVDLILSVQRSGEMQAPPSMRTAILVARLAAP LRLAGRLSDAALAEIAADVLRGRGLEADAAAFEAKLAA PTPGETAR*	
Serratia sp. ATCC 39006_gvpN	MIKQNTVSQYTVDDDLVVPEASEHFVATSYVNDIHERAL VYLRAGYPVHFAGPSGIGKTTLAFHLAALWGRPVMTMLQ GNEEFVSSDLTGKDIGYRKSSLVDNYIHSVLKTEEQMNR MWVDNRLTTACRNGDMLIYDEFNRSKAETNNVLLSVL SEGILNLPGLRGVGEGLDVHPEFRAIFTSNPEEYAGTH KTQDALMDRMITINIGLVDRDTELQILHARSELELKEAA YIVDIIRELRGNEHETKHGLRAGIAIAHILHQQGIKPRYG DKLFHAICYDVLSDAAKIQHAGRSIYREMGVIRKIC PPIGSDTVKASTQKIKAVE*	333
Stella vacuolata_ATCC- 43931_gvpN	MSTEPAPVMPPSTDIEFGSQRPARPKPAEALAVGYRLSA RPAAPSTLTLRPRADFVETDQVKDLTRRGLGFLRAGYPL HFRGPAGTGKTTLALHVAQAQLGRPVIITGDNELGTAD LVGSQRGYHYRKVVDQFIHNVTKLEETANQRWTDHRL TTACREGYTLVYDEFTRSRPETHNVLLGVFEELKILFLPA EAREECYIRVHPDFRAIFTSNPQEYAGVHASQDALADRL ATIDVDYPNRAVELAVASARTGLAEAEAAARIIDLVRFA RASGDYQQTPTMRASLMIAVAAQEGLRISVDDPGFVQ LCMDALESRIFSGARQEADARARHRVALLGLLATHCPS EAPVARVATVARAKRKSAS*	334
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpN	MSAKPLQDASEVSALNNDNVQPEASDTFVCTPSVEALA ERASAYLQAGYPVHLAGPAGTGKTTLAFHAAAKRGRP VKLIHGNDLGLADMVGQDNGYRRNTLVNDYIHSVVK TQEEVRTFWIDNRVTTACLNGETLIYDEFNRSRPEVNNI FLSILGEGILNLPNRRHQGAGYLEVHPEFRVIFTSNPEEY AGTHKTQDALMDRMITMKIGHYDRETEIRVTRAKSGLP PSEVAIVVDIVRELRGQSVNHRPTLRACIAIARIMADRR ISARSNNSFFRDICRDILDMDSAKVRRDGNALGESPVDD VVASISARARRPKIVEPKGLHKEI*	335
Tolypothrix sp. PCC 7601_gvpN	MTNTENHKKRAVLRVRPGQFVVTPEIAEKVAIRALRYLT SGFAIHLRGPAGTGKTTLAMHLANCLDRPIMLIFGDDEF KSSDLIGSESGYTHKKLLDNYIHNVLKVEDELKQNWVD SRLTLACREGLTLVYDEFNRSRPEVNNVLLSALEEKILT	336

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	LPPSSNQPEYLHVHPKFRAIFTSNPEEYCGVHSTQDALM DRLVTINMPEPDEQTQIEILTHKTGIHHEYAQLIARLVKA FRSATGAEKTSGLRSCLMVAKVCAEHDILVTPENTDFRE ICADVLFNRTNLSASDATTFLLELLNHVQVKPVEPVDDS DPYDVAEAEIVGAAEPQTDIAEPVTLDESLLSDQPN*	
Trichodesmium erythraeum IMS101_gvpN1	MTTVLNVSPDRFVSTPGVERVTQRASRYLESYGYSVHLR GPAGVGKTTALHLAHLRQQPIFLMIGDDEFKTSDLIGN KSGYTRKKLVDNYIHTVLKVEDEL RDNWIDSRLTLACK EGFTLIYDEFNRSRPEVNNVLLSVLEEKMLVLPPSQNQ EYIQVHPQFRVILTSNSEEWTGVHATQDALLDRVVTIG MEQPDISTEQNIVIKTGINPLKAEVIKLVRSVRQRVVK EDLGLRSALMISKVCHDHDIPLDGKDSFSDLCADILIS RPNLPRQEALQQLDEVLEEFFPADQPSSSDVGLEKEGSL *	337
Trichodesmium erythraeum IMS101_gvpN2	MTTVLNVSPDRFVSTPSVERVTQRASRYLESYGYSVHLR GPAGVGKTTALHLAHLRQQPIFLMIGDDEFKTSDLIGN KSGYTRKKLVDNYIHTVLKVEDELKHNWIDSRLTLACK EGFTLIYDEFNRSRPEVNNVLLSVLEEKMLVLPPSQNQ EYIQVHPQFRVILTSNSEEWTGVHATQDALLDRVVTIG MGQPDISTEQNIIKKTGINPLKAEVIKLVRSVRERLETE DLGSLRSALMISKVCHDHDIPLDGKDSNFSDLCADILISR ANLPRQEALKQLDEVLEELFPADQLSISDIGLKKEGSL*	338
gvpV		
Anabaena-flos-aquae_gvpV	MIKNIQVFFMKTISNRSISRAKISTMPRPKSDASSQLDLY KMTVEKQRIQRDMYSIKERMGLLQQLDILNQIEATE KTIHKLRQPHSNTAQNIVRSNIFVESNNYQTFEVEY*	339
Aphanizomenon flos-aquae NIES-81_gvpV	MKSFRHRSIIRAKISTMPRHISEASSQLELYKMVAEKQRI SRELSSIKERMATLQKRLDSLNNIDNTEKTIHKLRQPHS STAQNIVRSKNVSVESNNYQTFEIEY*	400
Arthrospira platensis NIES- 39_gvpV	MRYKYHRQIQPKLSAIPRQKSQANLYRNSYLLAVEKKR LTEELEVLQSRSHIEQRLALIEDQLGELEKDVTQLSVPPS PKPQNNLPVNNPEPPQSNPTNSSHINTFMVDY*	
Burkholderia thailandensis sp. Bp5365 strain	MPIPKKGLHDIRFRHAPGATPLPVHSMYMRISCIEMEKS RRTIERRAAQRRIAAVDSRVADLEREKARLYAAIDNEAP	401

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
MSMB43_gvpV	QAGDIRGSFRIRY*	
Desulfobacterium vacuolatum_DSM 3385_gvpV	MLKNRNRSGVQNIKTHAGKVDHVSHPHMAYMRISC LEMEKARKNKEKSGAQKRIDMINQRLMEIEKEKAHIQRI LGDTSLALESSNVHDHSEIKGGFKIKY*	402
Desulfomonile tiedjei DSM 6799_gvpV	MNIRMKGNSRGLRDIRTHSGKVDRVGLPYMAYMSISCL EMEKARREKERLSALTRIKNIEQRIEIEAEKDLLKGVG ERTRDLQKASTPRDQSAQCKGGFKIRY*	403
Legionella drancourtii LLAP12_gvpV	MMPALVKGLRNIKTMSNRLDKVQSPHEAFISAAALHRE KQRHLQELAILRNRLDEINLRLEQINEQQNQVAEAFDISP PRAVKSALRTGIQSKTGSTSHGFKIKY*	404
Microcystis aeruginosa NIES- 843_gvpV	MTTTRPPRPIRSKISTMPRKQSEADHQLELYKLITEKQRI QEKLEMMERQIQQLKNRLTFVTEQIETTEQSIQNLRTAN PPSVAKKPDSPKTVAHSSNNSSNFQTFYLEY*	405
Nostoc punctiforme ATCC 29133_gvpV	MHRTPNRRQIQAKLSTMPQRSQATVYLNAYKMMLEK ERLEEELEKLEARRHQIQQLAILNSQTIPEENMTHQQA NTDLENNTPKFNTLTLEY*	406
Nostoc sp. PCC 7120_gvpV	MLSIIQVFPMKVRNRGIIRPKITTMPRNKSEASSQLELY KLVTQQRIKQELAFIEQRTVLLKQRLSTLKTQIEGTERS INHLRHSELKYSRIALPKIFSETNNYQAFDIEY*	407
Planktothrix agardhii str. 7805_gvpV	MRPFRSQPPILPKISTMPRQKTEATLYRSLYQLAVEKKR LQEELESLGQRFETVTQRLQQIETQIQGLETDVKQIAPPK PPETKPNQPSTPTPTKAEPGSVSTFTLDY*	408
Psychromonas ingrahamii 37_gvpV1	MTAAKRKTLRGLADIRTISSCGTSGQEAYQMYLKRGVL EMEKLRRQKEKNSALERVTNINRRLMAIDTDIDFLCQSL KVIEKRTNQENSIVEKSVSRGFKLRY*	409
Psychromonas ingrahamii 37_gvpV2	MIFSKKKNALRGLADIRTLSCGTSGQEAYQMYLKRGV LEMEKLRRQKEKNSALERVRNINYRLMAIDADIDFLCQ SLKVIEERTNKENSISNESVTYKKGFKLRY*	410
Serratia sp. ATCC 39006_gvpV	MAISTRPLRTLSDIKTHSGRVSGEHQTYRDFQIGALELE RWRRTREERAASSRIASIDERIADIDKEKAALLADATAA SAVAENNDKSEAAEKKKKSSGLRIKY*	411

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
Thiocapsa rosea strain DSM 235 Ga0242571_11_gvpV	MSKFTQPSRSVRDIKTLAGMADDVRAPHKMYMRLFAL ETERHRRLQERASAMLRVDNIDARCAEIAEEMEQLLQIL GVEAVAPGGPPANARPGSGRVPTQPHRGRGKGTGAGR QTTSGETSVGEAVKIRY*	412
gvpW		
Anabaena-flos-aquae_gvpW	MELENLYTYAFLEIPSSPLILPQGAANQVVLLINGTELA AIVPEGFLESFQNNDEKIIQMALSHDRVICELFQQITVPLR FGTYFTSTNNLLNHLKSHEKEYQNKLEKINGKNEFTLKL IPRMIEEIVPSEGGGKDYFLAKKQRYQNQNNFSIAQAAE KQNLIDLITKVNQLPVVVQEQQEQIYLLVSCQDKTLL LEQFLTWQKACPRWDLLLGDCLPPYHFI*	413
Aphanizomenon flos-aquae NIES-81_gvpW	MELENLYTYAFLKTPSFLHLPQGSTTSVIQIDGNLSAI VEPGISLDSFQDDDEKIVQMAIEHDRVICDIFRQITVPLR FGTYFANTDNLLTHLESYGQEYLDKLEKINCKTEFILKLI PRMITEESPVLESGRHYFLAKKQHYQRQKNFILAQASEK EILINFISKINQIPVIIQEQQEEVRIYLLVNYQDKTLLLEQF LTWQQTCPRWDLFLGEGIPPYHFI*	414
Arthrospira platensis NIES-39_gvpW	MYVYAFIKSQSISWKSQGIYEPVVLLEAGALAAVVEP NLQAENLSADNEEELMRAVLTHDRIVCQIFEETTVPVR FGTCFDSEARLCEHLTTEGDRYFRQLEKLTGRAEYLLEA IPQPFNQEKPSDDTAPPTKGRDYFLQKKRLHQQLNFE QQQEQQWQDFINAIASKYPIVQGKATEDAERIYLLIPRS QEVALVEWVAQQQQNIDLWEFSLGNAVPAHYHFL*	415
Dolichospermum circinale_gvpW	MKLENFYTYAFLEIPRFPLVLPQGAASQVILINGSGMSAI VEPGISLESFQNNDEKIIQMALSHDRVICELFQQVTVLPL RFGTCFTSTNNLLNYLELHRQEYQEKLEKINGKIEFTLK LIPQTMEEPAPLERGGRDYFLAKKQRYQDQNNFRIAQA AEKQNLIDSISKVNQLPFVQKEKEEEVNIYLLVKSEDKTL LLEQFLNWQKACPRWDLLLGEPLPPYHFI	416
Microcystis aeruginosa NIES-843_gvpW	MKLYNLYTYAFLKTPIESLKLPGVGMANPLLLITGGELSA VVEPEVGLDTLQNDDERLIQSVLCHDRVICQLFQQTTIL PLRFGTSFLEAENLLTHLCSHGQEYQEKIEELEGGKEYL LKCIPIRKPEEPVLFSESKGRQYFLAKKQLYEAQQDFYTL QGSEWQNLVNLITQSYPTRIITAPGTESRIYLLVNLQEE	417

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	PLLIEQVLHWQKACPRWELQLGQVSPPYHFT*	
Nostoc punctiforme ATCC 29133_gvpW	MSIYAYALLVPTASPLVLPLGMERNTEL VYSSGLAALVE PEISLEAIQATDERLLQAVLNHDHVIRELFQQTPLLPLRF GRGFTSVEKLLNHLENHQEQYLETLTQLADKVEYSVKV TACSLDDSDTIDARGKAYLLAKKQRYQTQQAFAQQ CEQWELLNELILKTYTNVICETRQSDVRQIHFLAQRNDS TLSTQLFSLWQVQC SHWQLALSEPLPPYHFLKNTLI*	418
Nostoc sp. PCC 7120_gvpW	MRSNPFYTYAFLNTPDIPLRLPSGNLGQLLLIHGHKLSA VVEPGISLESSQNNDDEEVIKMLAHDRVICELSQQTTVL PLRFGTYFNSEETLLNHIESHAQEYQKKLDHIQKTEYT LKLIPRKFEELAKVSGNGRDYFLAKKLHYEHQKNFIG DQNREKNHNLINLIMDVYRSSAIQDYVEEVRLHLLVDRH DKTLLFKQVLTLEKCPHWNLILGEPLPPYHFV*	419
gvpR		
Bacillus-megaterium_gvpR	MEIKKIMQAVNDFFGEHVAPPHKITSVEATEDEGWRVI VEVIEEREYMKKYAKDEMLGTYESCFVNKEKEVISFKRL DVRYSIAIGIEA*	420
gvpS		
Bacillus-megaterium_gvpS	MSLKQSMENKDIALIDILDVILDKGVAIKGDLIISIAGVD LVYLDLRLVISSVETLVQAKEGNHKPITSEQFDKQKEEL MDATGQPSKWTNPLGS*	421
Rhodococcus hoagii 103S_gvpS	MSATPDRRIALVDLLDRVLGGGVVVAGEITLSIADVDM VHISLRTLVSVSALTRPPDEKPENDG*	422
gvpT		
Bacillus-megaterium_gvpT	MATETKLDNTQAENKENKNAENGSKENKNGSKASKTTS SGPIKRAVAGGIIGATIGYVSTPENRKSLLDRIDTDELKS KASDLGTKVKEKSKSSVASLKTSAGSLFKKDKDKSKDD EENVNSSSSETEDDNVQEYDELKEENQTLQDRLSQLEE KMNMLVELSLNKNQDEEAEDTDSDEEENDENDENDEN EQDDENEEETSKPRKKDKKEAEEEESEDEDESEEEEEEDS RSNKKNKVKTEEEDEDESEEEKKEAKPKKSTAKKSKN	423

Table 2: Amino acid sequences of exemplary gvpA/B, gvpF, gvpF/L, gvpG, gvpJ, gvpK, gvpL, gvpN, gvpV, gvpW, gvpR, gvpS, gvpT, and gvpU proteins

Species, protein	Amino acid sequence	SEQ ID NO.:
	TKAKKNTDEEDDEATSLSEDDTTA*	
gvpU		
Bacillus-megaterium_gvpU	MSTGPSFSTKDNLTLEYFVKASNKHGFSLDISLNVNGAVI SGTMISAKKEYFDYLSSETFEEGSEVAQALSEQFSLASEASE SNGEAEAHFIHLKNTKIYCGDSKSTPSKGKIFWRGKIAE VDGFFLGKISDAKSTSKKSS*	424

[00154] The exemplary GVGC cluster formed by Ana-gvpA, Ana-gvpC, Mega-gvpN Mega-gvpF, Mega-gvpG, Mega-gvpL Mega-gvpR Mega-gvpS, Mega-gvpT Mega-gvpK, Mega-gvpJ, and Mega-gvpU was used as ARG in the experiments summarized in the following Examples.

Example 2: BURST signals

[00155] **Figure 4** shows an example of the BURST paradigm. **Panel (a)** shows an illustration of the GV collapse (401) in response to a step increase in acoustic pressure (402), along with the transient acoustic signal created (403). Shortly after the collapse (401), the signal is diminished with the GV in a collapsed state (404). **Panel (b)** shows three consecutive images from the successive images taken during the collapse. In this example, these are images ten (410), eleven (411), and twelve (412) of a 50-frame sequence. The frames were taken for a BURST sequence applied to a tissue-mimicking phantom with wells containing plain 1% agarose (414) or 10^8 cells/ml ARG-expressing *E. coli* embedded in 1% agarose (415). The acoustic pressure is ramped from 0.27 MPa in the first 10 frames, including frame ten (410) to 3.2 MPa for the remaining 40 frames, including frames eleven (411) and twelve (412). Scale bar: 1 mm. **Panel (c)** shows a contrast-to-noise ratio (CNR) vs. frame number, showing the qualitative differences in the temporal dynamics of mean pixel intensity for different materials, corresponding to the regions of interest (420, 421, 422) identified in **panel (b)**. **Panel (d)** shows example output of the template projection algorithm, showing selective enhancement of tissue signal (420), GV signal (421), and noise (422). **Panel (e)** shows example output of the template unmixing algorithm,

showing estimated contribution of tissue signal (420), GV signal (421), and noise (422) to every pixel. In this example, the noise (422) and tissue (420) signal levels are fairly constant over time, but BURST can also be used where the signals change over time.

[00156] The following protocol was used to obtain the results illustrated in **Figure 4**. Plasmids encoding ARGs were transformed into chemically competent *E. coli* BL21(A1) cells (Thermo Fisher Scientific™) and grown in 5 ml starter cultures in LB medium with 50 µg ml⁻¹ kanamycin, 1% glucose for 16 h at 37 °C. Large-scale cultures in LB medium containing 50 µg ml⁻¹ kanamycin and 0.2% glucose were inoculated at a ratio of 1:100 with the starter culture. Cells were grown at 37 °C to OD_{600nm} = 0.5, then induced with 0.5% l-arabinose and 0.4 mM IPTG for 22 h at 30 °C.

[00157] Ultrasound imaging was performed using a Verasonics Vantage™ programmable ultrasound scanning system and an L22-14v 128-element linear array transducer (Verasonics™). Image acquisition was performed using a custom imaging script with a 64-ray-lines protocol with a synthetic aperture to form a focused excitation beam. An aperture of 65 elements was used. The transmit waveform was set to a frequency of 15.625 MHz for the L22 transducer, 67% intra-pulse duty cycle, and a 3/2-cycle pulse.

[00158] Phantoms for imaging were prepared by melting 1% (w/v) agarose in PBS and casting wells using a custom 3D-printed template that included a pair of 2mm diameter wells. *E. coli* cells at 2× the final concentration at 25 °C were mixed in a 1:1 ratio with molten agarose or molten TMM (at 56 °C) and immediately loaded into the phantom. The concentration of cells was determined before loading by measuring their OD_{600nm}. An arbitrary number of additional signal categories and corresponding templates can be used in the signal unmixing algorithms, including templates for different types of GVs, though the quality of the signal unmixing will tend to degrade as the number of signal categories increases. However, the three original signal templates were included for any version of BURST since they each model signal components that will be present to some degree in all setups.

[00159] Because the results found in **Figure 4** are intended primarily to illustrate the BURST method, the protocol described above need not be followed exactly to obtain similar results. For instance, purified GVs or GVs expressed in a different type of cell could be used, different

pressure levels could be applied as long as they are above the collapse threshold, and higher or lower concentrations of GVs could be used as long as they allow for detectable signal.

[00160] The GV template represents transient signal produced by GV collapse, the tissue template represents persistent signal that varies in proportion to the pressure applied, and the noise template represents persistent signal that does not vary in response to pressure applied. There are no limitations on the linearity of the signals, as mentioned earlier. The unmixing results will remain valid for all relative signal amplitudes, though GV signal may become undetectable in practice if the relative amplitudes of the noise and tissue signals are sufficiently large. Thermal noise, electronic noise, and many other mechanisms can contribute to the overall noise levels.

[00161] An arbitrary number of additional signal categories and corresponding templates can be used in the signal unmixing algorithms, including templates for different types of GVs, though the quality of the signal unmixing will tend to degrade as the number of signal categories increases. However, the three original signal templates can be included in any version of BURST since they each model signal components that will be present to some degree in all setups. The GV template represents transient signal produced by GV collapse, the tissue template represents persistent signal that varies in proportion to the pressure applied, and the noise template represents persistent signal that does not vary in response to pressure applied. There are no limitations on the linearity of the signals, as mentioned earlier. The unmixing results will remain valid for all relative signal amplitudes, though GV signal may become undetectable in practice if the relative amplitudes of the noise and tissue signals are sufficiently large. Thermal noise, electronic noise, and many other mechanisms can contribute to the overall noise levels.

Example 3: BURST at different pressure values

[00162] Figure 5 shows examples of loBURST and hiBURST collapse signal generation (0-80°dB, 2mm scalebars). **Panels (a) and (c)** show the power spectra resulting from BURST acquisitions with liquid buffer suspension of intact acoustic reporter gene (ARG) *E. coli* Nissle at 10^4 cells/ml, with PPP ranging from 3.2 MPa to 4.3 MPa. **Panels (b) and (d)** show their corresponding images, with the brighter pixels indicating higher dB. **Panels (a) and (b)** show the power spectra and images acquired using standard BURST imaging parameters: $\frac{1}{2}$ -cycle pulse at

6 MHz. **Panels (c) and (d)** show the power spectra and images acquired using a 10-cycle pulse at 5 MHz to increase frequency resolution and ensure the second harmonic peak is inside the bandwidth of the transducer. **Panel (e)** shows an image time series acquired with an ultrafast version of hiBURST, showing that many of the single sources observed in liquid buffer ARG-expressing cell suspension persist for several hundred microseconds (arrows). **Panel (f)** shows the time domain signal used to generate the power spectrum in **panel (a)** and **panel (g)** shows the time domain signal used to generate the power spectrum in **panel (c)**, both at 4.3 MPa (hiBURST). **Panel (h)** shows BURST images acquired with the 10-cycle sequence at pressures near the 10-cycle loBURST threshold, showing the emergence of single dim sources.

[00163] The following protocol was used to obtain the results illustrated in **Figure 5**. Plasmids encoding ARGs were transformed into electro-competent *E. coli* Nissle 1917 (Ardeypharm GmbH) and grown in 5 ml starter cultures in LB medium with 50 µg ml⁻¹ kanamycin, 1% glucose for 16 h at 37 °C. Large-scale cultures in LB medium containing 50 µg ml⁻¹ kanamycin and 0.2% glucose were inoculated at a ratio of 1:100 with the starter culture. Cells were grown at 37 °C to OD_{600nm} = 0.3, then induced with 3 µM IPTG for 22 h at 30 °C.

[00164] An L11-4v transducer (Verasonics™) was mounted on a computer-controlled 3D translatable stage (Velmex™) above a 4 L bucket containing 3.8 L water that had been circulated through a water conditioner for 1 hour to remove air bubbles. 200 ml of 20x PBS was then gently added to the water, with the mouth of the PBS-containing bottle at the level of the surface of the water to avoid creating bubbles. A piece of acoustic absorber material was placed at the bottom of the bucket to reduce reflections. A MATLAB™ script was written to control the Verasonics system in tandem with the Velmex stage, which was programmed to move 1 cm after each of 5 replicate BURST pulse sequences. Intact Nissle cells were added to the bucket for a final concentration of 10⁴ cells/ml. After each set of replicate acquisitions, the bucket was stirred gently with a glass rod and another set of acquisitions were made at the next pressure level.

Example 4: comparison of BURST with previous techniques

[00165] To compare the performance of BURST with existing techniques under a range of well-controlled conditions, several concentrations of ARG-expressing Nissle *E. coli* in an agarose phantom were imaged using various imaging techniques (see e.g. **Example 3** and **Figure 5**). The

phantom consisted of a rectangular block of agarose gel with several pairs of cylindrical wells that were filled with ARG-expressing Nissle *E. coli* embedded in phantom material. In each pair, the well on the left contained cells whose GVs had been hydrostatically collapsed to serve as a control, while the well on the right contained cells with intact GVs. In half the well pairs, the cells were embedded in plain 1% agarose. In the other half, the cells were embedded in tissue-mimicking material (TMM)[3] to emulate the challenges of in vivo imaging. Cell concentrations ranged from 10^9 cells/ml to 10^3 cells/ml. For each phantom material and concentration, images were acquired using four different ultrasound imaging techniques: 1) standard B-mode, 2) pre-collapse/post-collapse difference, 3) loBURST, and 4) hiBURST. For consistency and quantifiability, template unmixing was used to process all BURST images.

[00166] The following protocol was used to obtain the results illustrated in **Figure 6**. Plasmids encoding ARGs were transformed into electro-competent *E. coli* Nissle 1917 (Ardeypharm GmbH) and grown in 5 ml starter cultures in LB medium with 50 $\mu\text{g ml}^{-1}$ kanamycin, 1% glucose for 16 h at 37 °C. Large-scale cultures in LB medium containing 50 $\mu\text{g ml}^{-1}$ kanamycin and 0.2% glucose were inoculated at a ratio of 1:100 with the starter culture. Cells were grown at 37 °C to $\text{OD}_{600\text{nm}} = 0.3$, then induced with 3 μM IPTG for 22 h at 30 °C.

[00167] **Figure 6** shows an example of in vitro BURST imaging. **Panels (a)-(d)** show an array of ultrasound images of a cross section of cylindrical wells containing ARG-expressing Nissle *E. coli* embedded in non-scattering agarose, within an agarose phantom. Each image contains a pair of wells, the left well containing cells with collapsed GVs and the right well containing cells with intact GVs. Rows correspond to cell concentrations, which range over seven orders of magnitude. Columns correspond to different image processing techniques, as indicated by the bottom labels. The top edge of each image corresponds to a depth of 16 mm, the bottom to a depth of 23 mm. The left edge of each image corresponds to a lateral coordinate of -7 mm, the right to +7 mm. Scalebars: 2 mm. **Panels (e)-(h)** show ultrasound images of the same conditions as **panels (a)-(d)**, but with the cells embedded in tissue-mimicking material (TMM) inside the wells. **Panels (i)-(l)** show CTR vs log cell concentration for loBURST and hiBURST. **Panel (i)** shows loBURST on agar-embedded cells. **Panel (j)** shows hiBURST on agar-embedded cells. **Panel (k)** shows loBURST on TMM-embedded cells. **Panel (l)** shows hiBURST on TMM-embedded cells.

[00168] In line with previously reported results, ARG contrast in B-mode images was clearly detectable at 10^9 cells/ml in non-scattering agarose and only marginally detectable at 10^8 cells/ml (**Figure 6, panel (a)**). Clutter is reduced in difference images relative to B-mode, but this technique did not improve upon the B-mode detection limit (**Figure 6, panel (b)**). In the TMM conditions, ARG contrast was not detectable in either B-mode or difference images for any cell concentration (**Figure 6, panels (e)-(f)**). The residual signal observed in TMM control wells in **Figure 6, panels (f)-(h)** supports the conclusion that this condition was made more challenging by the presence of microscopic air bubbles inadvertently introduced during mixing of cell samples with molten TMM, which has a significantly higher viscosity than plain 1% agarose.

[00169] Both hiBURST and loBURST improved these detection limits to 10^4 cells/ml in plain agarose (**Figure 6, panels (i)-(j)**) and 10^6 cells/ml in TMM (**Figure 6, panels (k)-(l)**). Moreover, loBURST CTR appears to be linearly proportional to cell concentration in all conditions except 10^9 cells/ml in plain agarose, demonstrating its potential for quantifying ARG-expressing cell concentration across at least five orders of magnitude.

[00170] At 10^9 and 10^8 cells/ml in plain agarose, hiBURST had suboptimal CTR relative to both the same concentrations with loBURST and even some lower concentrations with hiBURST (**Figure 6, panels (c)-(d), (i)-(j)**). This counterintuitive result is mostly due to the generation of cavitation events in the control wells from collapsed GVs, which are known to act as weak cavitation nuclei [2]. There also appears to be amplitude-dependent acoustic shielding, in which acoustic energy is absorbed by cavitation events caused by the higher-pressure pulse in the top portion of the well, shielding the interior. hiBURST also did not significantly improve the detection limit relative to loBURST, mostly due to the presence of microbubbles in the wells that cause confounding transient signal at the higher pressures as well as amplitude-dependent attenuation of the transmitted pulse. Because such microbubbles are not present in biological tissue, hiBURST will likely still offer advantages in certain in vivo imaging contexts.

[00171] These results demonstrate the potential of BURST to image ARG-expressing cells co-localized with strong scatterers at 10^6 cells/ml, which are relevant conditions for imaging rare gut microbial species.

Example 5: BURST imaging of *in vivo* gut microbe distribution

[00172] To test the *in vivo* specificity and robustness of BURST under a protocol used in previous work on GV imaging *in vivo*, probiotic ARG-expressing *E. coli* Nissle cells in agarose gel were imaged within the colon of an anesthetized mouse at 10^7 cells/ml, an order of magnitude lower than the previous *in vitro* detection limit (see **Figure 6, panel (a)**). To demonstrate the maximum contrast to tissue ratio (CTR) achievable with conventional imaging in this setting, an AM image of the gel-filled colon at the moment of collapse is shown in Figure 5, panel (b). Template projection of this image together with the other frames in the time series resulted in a BURST image with 40 dB higher CTR (**Figure 7B, panel (c)**). This result demonstrates the *in vivo* robustness of BURST and its orders-of-magnitude improvement in CTR relative to conventional imaging methods.

[00173] All *in vivo* experiments were performed on mice, under a protocol approved by the Institutional Animal Care and Use Committee of the California Institute of Technology. No randomization or blinding were necessary in this study. Mice were anesthetized with 1–2% isoflurane, maintained at 37 °C on a heating pad, depilated over the imaged region, and imaged using an L11-4v transducer attached to a manipulator. For colon imaging, an L22-14v transducer was used. For imaging of gavaged *Salmonella typhimurium* in the gastrointestinal tract, mice were placed in a supine position, with the ultrasound transducer positioned over the upper abdomen such that the transmit focus of 12 mm was close to the top of the abdominal wall. Two hours prior to imaging, mice were gavaged with 200 μ l of buoyancy-enriched *Salmonella typhimurium* at a concentration of 10^9 cells/ml.

[00174] Because BURST amplifies changes in pixels across frames, any tissue motion in the timeseries may confound the final image. To mitigate this during *in vivo* imaging, we implemented a custom BURST script that transmits and acquires three 32-aperture focused beams at a time, improving the frame rate by a factor of 3. The smaller aperture meant that hiBURST pressures could not be achieved, so all *in vivo* images were acquired using loBURST.

[00175] After each acquisition, the manipulator was used to translate the transducer 1 mm forward to the next image plane. An attempt was made to time each acquisition to coincide with the part of the mouse's breathing cycle with the least motion.

[00176] Prior to processing with template unmixing, a 2×2 median filter followed by a gaussian blur filter with $\sigma = 1$ was applied to each 2D image frame of each image plane of each mouse. Template unmixing was applied using 1 low-pressure frame (frame 5) and 2 high-pressure frames (frames 6-7). The images output from template unmixing were then concatenated into a 3D array to which a $1 \times 1 \times 2$ 3D median filter was applied to remove isolated motion artifacts. The resulting 2D BURST images were then dB scaled and overlaid on the square-root-scaled B-mode image representing frame 1 in the corresponding timeseries. The BURST images were overlaid in locations where the BURST image pixel values exceeded a threshold of 105 dB, which was chosen as the minimum threshold at which no residual motion artifacts were visible in the lower abdomen, where no BURST signal was expected. BURST images were pseudo-colored with the hot colormap and B-mode images with the gray colormap. Quantification was performed by manually drawing ROIs conservatively covering the upper half of the abdominal cavity in each image plane for each mouse.

[00177] **Figures 7A and 7B** show an example of in vivo BURST imaging. **Panel (a)** is an illustration of a colon injection experiment. **Panel (b)** shows a collapse frame AM image of the mouse colon filled with probiotic ARG-expressing *E. coli* Nissle at 10^7 cells/ml. **Panel (c)** shows a BURST image with template projection, generated from an image time series. Scalebars = 1 mm. **Panel (d)** is an illustration of an oral gavage experiment. **Panels (e)-(f)** show B-mode images (PPP = 0.93 MPa) of a coronal cross section of the mouse abdominal cavity 17 mm caudal to the rib cage, acquired 2 hours post-gavage. A heatmap of the corresponding BURST image is overlaid in locations where the BURST CNR exceeds 105 dB. **Panel (e)** shows control gavage of luciferase-expressing *Salmonella*. **Panel (f)** shows gavage of ARG-expressing *Salmonella*. **Panel (g)** shows a plot of mean BURST CTR in the abdominal cavity vs distance of the image plane in the caudal direction from the rib cage for mice gavaged with ARG-expressing *Salmonella* and luciferase-expressing *Salmonella*. Error bars: SEM, $n = 4$. **Panels (h)-(i)** show four image planes following those in **panels (e) and (f)** (18 mm to 21 mm) from the same representative mice with the same display settings. **Panel (h)** shows spatial sequence frames (1 mm spacing) for a control mouse (luciferase-expressing *Salmonella*), displaying no signal. **Panel (i)** shows spatial sequence frames (1 mm spacing) for a mouse with ARG-expressing *Salmonella*, the frames displaying a BURST signal.

[00178] BURST was used to noninvasively image the spatial distribution of a pathogenic bacteria propagating naturally through the GI tract of a mammalian host, a procedure that could not be performed using previous techniques. An attenuated strain of Salmonella was used as a model pathogen for the mouse GI tract. Two groups of four mice were gavaged with 10^9 cells in 200 μ l 2 hours prior to anesthetization and imaging. The experimental group was gavaged with buoyancy-enriched ARG-expressing Salmonella and the control group with luciferase-expressing Salmonella. No fasting, bicarbonate administration, or other pretreatments were used. Because the 3D spatial distribution of cells was not known a priori, loBURST data was acquired for the entire abdominal cavity of each mouse in 20-30 transverse image planes with 1 mm spacing (see **Figure 7A, panel (d)**). Display images were generated by overlaying grayscale low-pressure B-mode images with heatmaps of all BURST image pixels exceeding a CNR threshold of 105 dB (**Figures 7A and 7B, panels (e)-(f)**).

[00179] In all but one experimental mouse, contiguous patches of supra-threshold BURST signal, approximately 2 mm $\times$ 1 mm, were observed spanning several contiguous frames in the middle of the abdomen 1 mm below the abdominal wall (**Figure 5, panel (i)**), the expected location of the small intestine. No supra-threshold BURST signal was observed in the abdominal cavities of control mice (**Figure 5, panel (h)**). Aggregating mean BURST CTR in the upper half of the abdominal cavity in each image plane for all mice, there is a statistically significant enhancement of BURST CTR in the experimental group for all image planes between 16 mm and 22 mm, inclusive. These results demonstrate the capability of BURST to noninvasively image gene expression of cells co-localized with strong scatterers in a live animal host with no prior knowledge of their spatial distribution.

Example 6: single cell imaging

[00180] An advantage of BURST is the ability to resolve imaging to detect contrast at the individual cell level. An example of this is imaging in degassed liquid buffer a linear range of concentrations of ARG-expressing Nissle, on the order of 10^2 - 10^3 cells/ml, as well as pre-collapsed controls. Based on hydrophone measurements of the pressure profile of the $\frac{1}{2}$ cycle BURST pulse sequence and the observed loBURST pressure threshold, it is estimated that all ARG-expressing cells in a 1 mm $\times$ 19.5 mm $\times$ 1 mm field of view (FOV) experience sufficient

pressure to generate collapse signal by either the loBURST or hiBURST mechanism. This value can be used to estimate the expected number of sources in each BURST image for each cell concentration. Both bright and dim sources can be counted as a single source.

[00181] The following protocol was used to obtain the results illustrated in **Figure 5**. Plasmids encoding ARGs were transformed into electro-competent *E. coli* Nissle 1917 (Ardeypharm GmbH) and grown in 5 ml starter cultures in LB medium with 50 µg ml⁻¹ kanamycin, 1% glucose for 16 h at 37 °C. Large-scale cultures in LB medium containing 50 µg ml⁻¹ kanamycin and 0.2% glucose were inoculated at a ratio of 1:100 with the starter culture. Cells were grown at 37 °C to OD_{600nm} = 0.3, then induced with 3 µM IPTG for 22 h at 30 °C.

[00182] For validation of single-cell detection, an L11-4v transducer (Verasonics) was mounted on a computer-controlled 3D translatable stage (Velmex) above a 4 L bucket containing 3.8 L water that had been circulated through a water conditioner for 1 hour to remove air bubbles. 200 ml of 20x PBS was then gently added to the water, with the mouth of the PBS-containing bottle at the level of the surface of the water to avoid creating bubbles. A piece of acoustic absorber material was placed at the bottom of the bucket to reduce reflections. A MATLAB script was written to control the Verasonics system in tandem with the Velmex stage, which was programmed to move 1 cm after each of 10 replicate BURST pulse sequences. After each set of BURST acquisitions (starting with plain PBS), 30 µl of 10⁶ cells/ml intact Nissle cells were added to the bucket, which was gently stirred with a glass rod. A separate bucket with freshly conditioned water and buffer was used for the collapsed control cells. A MATLAB script was used to display a 1 mm × 19.5 mm segment, centered at the point of highest average intensity, of all BURST images (all replicates, all concentrations, and collapsed vs. intact cells) in a random order, blinding the experimenter to the condition when performing source counting.

[00183] Comments for replicating results: One should use the following guidelines for accurate counting:

- If a bright signal that spans 2 or 3 columns is observed, with the intensity decreasing from left to right, then that counts as one source.
- However, if the intensity increases from left to right, usually that should count as two,

with the logic being that it is more likely two separate sources or one source that coalesced with another one. Also, if the source spans more than three columns, that is counted as two as well.

- Err on the side of false positives when looking at very weak sources that only slightly stand out from the noise. Sources from bubbles tend to be very bright, and weak sources also don't have the problem of spanning more than one column, so they are what can most confidently be classified as single cells.
- If a bright source with chunks of black pixels in it is observed, that corresponds to a persistent bubble that moved slightly between collapse frames and was partially cancelled, and so should not be counted as a source.
- If the source is partially outside of the frame, it still counts as long as its brightest point seems to be inside the frame.

[00184] Figure 8 shows an example of single cell detection compared to a control. **Panel (a)** shows a picture of the example experimental setup. **Figure 8 Panel (b)** shows a plot of the average number of single sources counted in images acquired with hiBURST vs cell concentration for both intact and collapsed ARG-expressing *E. coli* Nissle for the example. The expected number of cells in the transducer's field of view, based on cell counting by fluorescence microscopy and hydrophone measurements of the transducer's peak pressure profile, is also plotted for comparison. **Panel (c)** shows, for the example, representative images acquired with hiBURST showing single sources in liquid buffer suspension of intact ARG-expressing *E. coli* Nissle, and the number of sources increasing with cell concentration. **Panel (d)** shows, for the example, representative images of liquid buffer suspension with collapsed ARG-expressing *E. coli* Nissle.

[00185] In images of buffer containing cells with intact GVs, the number of sources was found to increase linearly with cell concentration (**Figure 8, panel (c)**). In images of buffer containing cells with collapsed GVs, the average number of sources had no significant dependence on cell concentration and the global average for the collapsed condition was 0.65 ± 0.95 (**Figure 8, panel (d)**). This establishes intact GVs as the causal agent for the observed sources,

demonstrating that few or no observed sources are generated by causes other than the collapse of intact GVs expressed in single cells.

[00186] Most significantly, the number of sources observed in images of cells with intact GVs closely tracks the expected number (**Figure 8, panel (b)**), with SEM error bars overlapping at all concentrations. The number of sources begins to level off at concentrations above 720 cells/ml, but this is an expected consequence of increased probability of overlapping sources at higher concentrations, demonstrating that most or all single cells expressing intact GVs generate observable sources when exposed to BURST, and each ARG-expressing cell generates one and only one source.

[00187] These results demonstrate the ability of BURST to reliably image gene expression in single cells with high sensitivity and specificity.

Example 7: In vitro ultrasound imaging in mammalian cell

[00188] To create phantoms for in vitro ultrasound imaging, wells were casted with molten 1% w/v agarose in PBS using a custom 3D-printed template. ARG-expressing and mCherry-only control cells were allowed to express gas vesicles using the specified inducer concentrations and expression duration. They were then trypsinized and counted via disposable hemocytometers in bright field microscopy. Next, cells were mixed at a 1:1 ratio with 50 °C agarose and loaded into the wells before solidification. The volume of each well is 60 μ l and contain 6×10^6 cells. The phantoms were submerged in PBS, and ultrasound images were acquired using a Verasonics Vantage programmable ultrasound scanning system and L22-14v 128-element linear array transducer with a 0.10-mm pitch, an 8-mm elevation focus, a 1.5-mm elevation aperture, and a center frequency of 18.5 MHz with 67% –6 dB bandwidth (Verasonics). Each frame was formed from 89 focused beam ray lines, each with a 40-element aperture and 8 mm focus. A 3-half-cycle transmit waveform at 17.9 MHz was applied to each active array element. For each ray line, the AM code is implemented using one transmit with all elements in the aperture active followed by 2 transmits in which the odd- and then even-numbered elements are silenced. Each image contains a circular cross-section of a well with a 4 mm diameter and center positioned at a depth of 8 mm. In AM mode, signal was acquired at 0.9 MPa (2V) for 10 frames and the acoustic pressure was increased to 4.3 MPa (12V) to collect 46 frames. There after the acoustic pressure

was increased to 8.3 MPa (25V) to ensure complete collapse of gas vesicles. Gas vesicle-specific signal was determined by subtracting the area under the curve of the first sequence by the post-collapse imaging sequence.

[00189] Figure 9 shows an example of in vitro ultrasound imaging of gene expression. **Panel (A)** illustrates an ultrasound paradigm used to extract gas vesicle-specific ultrasound image from ARG-expressing cells. **Panel (B)** shows representative non-linear echoes received during this ultrasound imaging paradigm. Insonated acoustic pressures in the white region are 0.9 MPa and within the grey box are 4.3 MPa. **Panel (C)** shows cellular viability after being insonated under 8.3 MPa acoustic pressures. **Panel (D)** shows ultrasound imaging of ARG-expressing cells as a function of expression duration. Cells were induced with 1 μ g/mL of doxycycline and 5 mM sodium butyrate. **Panel (E)** shows example ultrasound imaging of ARG-expressing cells as a function of doxycycline induction concentrations. Cells were allowed to express gas vesicles for 72 hours in the presence of 5 mM sodium butyrate. **Panel (F)** shows example ultrasound imaging of ARG-expressing cells mixed with mCherry-only control cells in varying proportions. Cells were induced with 1 μ g/mL of doxycycline and 5 mM sodium butyrate for 72 hours prior to imaging. **Panels (D)-(F)** show representative ultrasound images of cells embedded in agarose phantoms. To generate each image, a set of nonlinear ultrasound images are acquired (55 frames totaling 1.65 seconds), the cells are insonated with 8.3 MPa ultrasound and a set of nonlinear ultrasound images are re-acquired for the background. The total ultrasound signal from each series is calculated and the square-root of the difference is displayed (top). Region of interest quantification for each replicate is shown as a shaded circle with the mean shown as a dark circle (bottom). **Panel (G)** illustrates that ARG-expressing cells can re-express gas vesicles after acoustic collapse. Representative ultrasound image of ARG-expressing cells mixed in Matrigel that were induced with 1 μ g/mL of doxycycline and 5 mM sodium butyrate for 72 hours (before and after 8.3 MPa acoustic insonation). The ARG-expressing cells laden in Matrigel are induced for an additional 72 hours and imaged using ultrasound (bottom left). Images are generated from the square-root of the different between the nonlinear ultrasound signal at the moment of gas vesicle collapse (frame 11) from the nonlinear ultrasound signal at frame 15. Region of interest quantification for each replicate is shown as a shaded circle with the mean shown as a dark circle.

Example 8: longBURST and shortBURST Characterization

[00190] The results illustrated in **Figure 10** were obtained by the same protocol used in Example 3, except that 10 replicates were used instead of 5.

[00191] **Figure 10** shows examples of shortBURST and longBURST signal generation and illustrates how the signal properties change with varying pressure levels and number of transmit waveform cycles. **Panel (a)** shows representative echoes received following the application of shortBURST at varying pressure levels, indicated by the text in the corresponding rows of **panel (c)**. The number of sources increases with the PPP, and all of the sources are small and dim. **Panel (b)** shows representative echoes received following the application of longBURST at varying pressure levels, indicated by the text in the corresponding rows of **panel (c)**. The number of sources increases with the PPP and start out with only dim sources being observed at lower PPP. However, unlike with shortBURST, elongated bright sources begin to appear as the PPP is increased. **Panel (c)** shows the power spectra of shortBURST (dark gray, lower curve) and longBURST (light gray, upper curve) at each pressure level, obtained by averaging the time-domain signals over the 64 ray lines in each of the 10 replicates. The spectral resolution is not sufficient to identify harmonic peaks, but there appears to be a slight broadband enhancement observed in higher frequencies in the longBURST spectra at higher pressure levels. **Panel (d)** shows the peak intensity observed in the shortBURST and longBURST images as a function of peak positive pressure (PPP). While the shortBURST peak intensity does not increase with pressure, the peak intensity for longBURST, which is dominated by the bright sources, increases significantly with pressure. This is further evidence that the dim sources are produced by the GV collapse event, whose intensity is independent of PPP after the PPP exceeds the collapse threshold. **Panel (e)** shows the persistence and gradual disappearance of several bright sources generated by longBURST with an ultrafast acquisition script. Dim sources are evident in the first high-pressure frame but disappear completely after 100 μ s. 10 bright sources are observed in the first high-pressure frame, 5 are observed after 100 μ s, and 3 are observed after 200 μ s. This provides evidence that the bright sources are produced by nanobubbles liberated from collapsed GVs. **Panel (f)** shows representative images obtained by applying hiBURST with varying numbers of waveform cycles. **Panel (g)** shows the mean intensity of the hiBURST images (average over 10 replicates) as a function of depth for different numbers of waveform cycles (0.5

cycles being the lowest curve, 8.5-10.5 cycles being the highest curves). **Panel (h)** shows the peak mean intensity as a function of number of waveform cycles, demonstrating that an increase in the number of cycles increases signal intensity, which would be expected if the signal were partially produced by inertial cavitation. **Panel (i)** shows the full-width at half maximum (FWHM) of the mean intensity vs. depth profiles as a function of number of waveform cycles, providing further evidence for the cavitating bubble mechanism of signal generation.

Example 9: Hypothetical

[00192] For a hypothetical example, suppose that you have a new bacteria strain, we will call *A. Hypothetica*, and you suspect that it can produce GVs. In an initial step, the proteins from *A. Hypothetica* are sequenced and it is determined that they have a sequence in a gene cluster that is a close match to gvpF. To verify, the GVs are expressed and isolated, via lysing, as a contrast agent. As a control, a portion of these isolated GVs are collapsed using a hydrostatic pressure well above the hydrostatic collapse threshold of all known GVs – in this example, 12 MPa. The contrast agent is injected into a target site of a known signal attenuation for ultrasound at a selected frequency – in this example, approximately 3 dB/cm at 3.5 MHz. The target site is imaged at a starting PPP of 0.5 MPa, calculated using the known attenuation and depth of the target site. While frames are captured, the PPP is suddenly increased to a hiBURST level (e.g. 4.3 MPa) for a longBURST duration (e.g. 8 half-cycles). The frames from before, during, and after the step increase of PPP undergo template unmixing to discern a BURST signal against the background signals. The injection and imaging procedure is repeated with the collapsed control sample. If the signal observed in the target site containing contrast agent is significantly greater than the signal observed in the target site containing the control sample then GVs were present. Additional tests at different increased PPP levels can be performed on new batches of GVs to determine an acoustic collapse profile of the GVs, with the point where approximately 50% of GVs collapsing (profile midpoint) being selected as the acoustic collapse threshold of the GVs.

[00193] In summary provided herein is, the application of a step function increase in acoustic pressure during ultrasound imaging using gas vesicle contrast, along with capturing successive frames of ultrasound imaging and extracting time-series vectors for pixels of the frames, allows for improved imaging down to even the cell level. Template vectors can be used to implement

signal separation of the time-series vectors to improve detection as will be understood by a skilled person.

[00194] The examples set forth above are provided to give those of ordinary skill in the art a complete disclosure and description of how to make and use the embodiments of the materials, compositions, systems and methods of the disclosure, and are not intended to limit the scope of what the inventors regard as their disclosure. Those skilled in the art will recognize how to adapt the features of the exemplified methods and arrangements to additional gas vesicles, related components, genetic or chemical variants, as well as in compositions, methods and systems herein described, in according to various embodiments and scope of the claims.

[00195] All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the disclosure pertains.

[00196] The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background, Summary, Detailed Description, and Examples is hereby incorporated herein by reference. All references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually. However, if any inconsistency arises between a cited reference and the present disclosure, the present disclosure takes precedence.

[00197] The terms and expressions which have been employed herein are used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the disclosure claimed. Thus, it should be understood that although the disclosure has been specifically disclosed by embodiments, exemplary embodiments and optional features, modification and variation of the concepts herein disclosed can be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this disclosure as defined by the appended claims.

[00198] It is also to be understood that the terminology used herein is for the purpose of

describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. The term "plurality" includes two or more referents unless the content clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the disclosure pertains.

[00199] When a Markush group or other grouping is used herein, all individual members of the group and all combinations and possible sub-combinations of the group are intended to be individually included in the disclosure. Every combination of components or materials described or exemplified herein can be used to practice the disclosure, unless otherwise stated. One of ordinary skill in the art will appreciate that methods, device elements, and materials other than those specifically exemplified may be employed in the practice of the disclosure without resort to undue experimentation. All art-known functional equivalents, of any such methods, device elements, and materials are intended to be included in this disclosure. Whenever a range is given in the specification, for example, a temperature range, a frequency range, a time range, or a composition range, all intermediate ranges and all subranges, as well as, all individual values included in the ranges given are intended to be included in the disclosure. Any one or more individual members of a range or group disclosed herein may be excluded from a claim of this disclosure. The disclosure illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein.

[00200] A number of embodiments of the disclosure have been described. The specific embodiments provided herein are examples of useful embodiments of the invention and it will be apparent to one skilled in the art that the disclosure can be carried out using a large number of variations of the devices, device components, methods steps set forth in the present description. As will be obvious to one of skill in the art, methods and devices useful for the present methods may include a large number of optional composition and processing elements and steps.

[00201] In particular, it will be understood that various modifications may be made without departing from the spirit and scope of the present disclosure. Accordingly, other embodiments are within the scope of the following claims.

REFERENCES

1. Maresca, D. *et al.* Nonlinear ultrasound imaging of nanoscale acoustic biomolecules. *Appl. Phys. Lett.* 110, 073704 (2017).
2. Bar-Zion, A. *et al.* Acoustically Detonated Biomolecules for Genetically Encodable Inertial Cavitation. *bioRxiv* 620567 (2019).
3. Ramnarine, K. V., Anderson, T. & Hoskins, P. R. Construction and geometric stability of physiological flow rate wall-less stenosis phantoms. *Ultrasound in medicine & biology* 27, 245–250 (2001).
4. Walsby, A.E., *Gas vesicles*. Microbiol. Rev., 1994. 58(1): p. 94-144.
5. Walsby, A.E., *Gas-vacuolate bacteria (apart from cyanobacteria)*, in *The Prokaryotes*. 1981, Springer. p. 441-447.
6. Walsby, A.E., *Cyanobacteria: planktonic gas-vacuolate forms*. The Prokaryotes, a Handbook on Habitats, Isolation, and Identification of Bacteria, 2013. 1: p. 224-235.
7. Woese, C.R., Bacterial evolution. Microbiological reviews, 1987. 51(2): p. 221.
8. Walsby, A.E., Gas vesicles. Microbiol Rev, 1994. 58(1): p. 94-144.
9. Pfeifer, F., Distribution, formation and regulation of gas vesicles. Nat Rev Microbiol, 2012. 10(10): p. 705-15.
10. Smith TF, W.M., Identification of common molecular subsequences. J Mol Biol, 1981. 147(1): p. 3.
11. Altschul SF, M.T., Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 1997. 25(17): p. 14.
12. Lu et al. “Acoustically modulated magnetic resonance imaging of gas-filled protein nanostructures” Nature Materials, 2018

CLAIMS

1. A method of ultrasound imaging to be used on a target site contrasted with gas vesicles (GVs) having an acoustic collapse pressure threshold, the method comprising:
 - applying ultrasound to the target site at a peak positive pressure less than the acoustic collapse pressure threshold;
 - increasing peak positive pressure (PPP) to above the selective acoustic collapse pressure value as a step function;
 - imaging the target site in successive frames during the increasing; and
 - extracting a time-series vector for each of at least one pixel of the successive frames.
2. The method of claim 1, further comprising performing a signal separation algorithm on the time-series vectors using at least one template vector.
3. The method of claim 2, wherein the signal separation algorithm includes template projection.
4. The method of claim 2, wherein the signal separation algorithm includes template unmixing.
5. The method of any of claims 2-4, wherein the at least one template vector includes linear scatterers, noise, gas vesicles, or a combination thereof.
6. The method of any of claims 1-5, wherein the successive frames comprise a frame prior to GV collapse, a frame during GV collapse, and a frame after GV collapse.
7. The method of any of claims 1-6, further comprising delivering the GV to the target site.
8. The method of claim 7, wherein the delivering the GV to the target site comprises using an acoustic reporter gene to express the GV.
9. The method of claim 8, wherein the target site comprises a mammalian cell with the acoustic reporter gene.
10. The method of any of claims 1-9, wherein the increasing includes increasing the PPP to a

hiBURST regime.

11. The method of claim 10, wherein the PPP in hiBURST regime is 4.3 MPa or higher.

12. The method of any of claims 1-11, wherein the increasing includes increasing the PPP to a loBURST regime.

13. The method of claim 12, wherein the PPP in loBURST regime is no higher than 3.7 MPa.

14. A system for imaging a target site contrasted with gas vesicles (GVs) having an acoustic collapse pressure threshold, the system comprising:

an ultrasound source capable of producing peak positive pressure both below and above the acoustic collapse pressure threshold;

an ultrasound imager configured to capture successive frames from the target site; and

a processor configured to:

calculate a time-series vector for each of at least one pixel of the successive frames.

15. The system of claim 14, wherein the processor is further configured to perform a signal separation algorithm on the time-series vectors using at least one template vector.

16. The system of claim 15, wherein the signal separation algorithm includes template projection.

17. The system of claim 15, wherein the signal separation algorithm includes template unmixing.

18. The system of any of claims 15-17, wherein the at least one template vector includes linear scatterers, noise, gas vesicles, or a combination thereof.

19. The system of any of claims 14-18, wherein the successive frames comprise a frame prior to GV collapse, a frame during GV collapse, and a frame after GV collapse.

20. The system of any of claims 14-19, further comprising a means for introducing the gas

vesicles at the target site.

21. The system of claim 20, wherein the delivering the GVs to the target site comprises using an acoustic reporter gene to express the GVs.

22. The system of claim 21, wherein the acoustic reporter gene is in a mammalian cell.

1/16

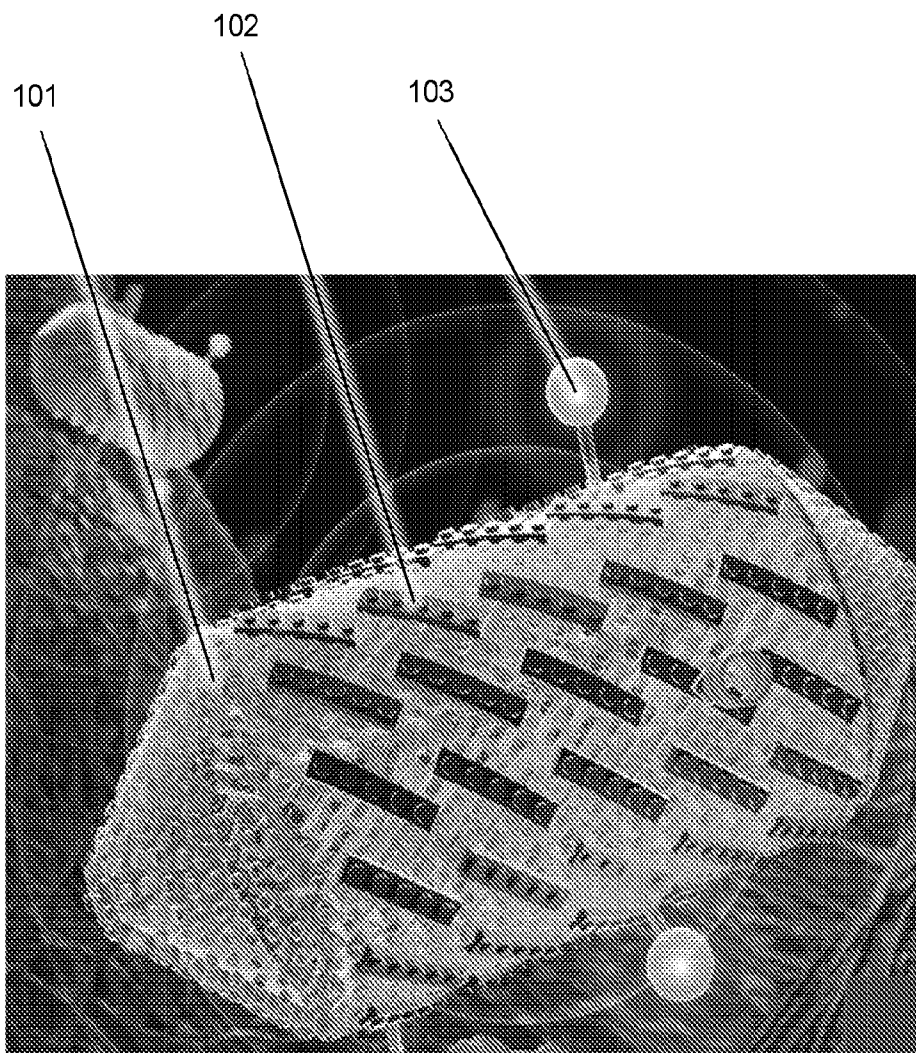
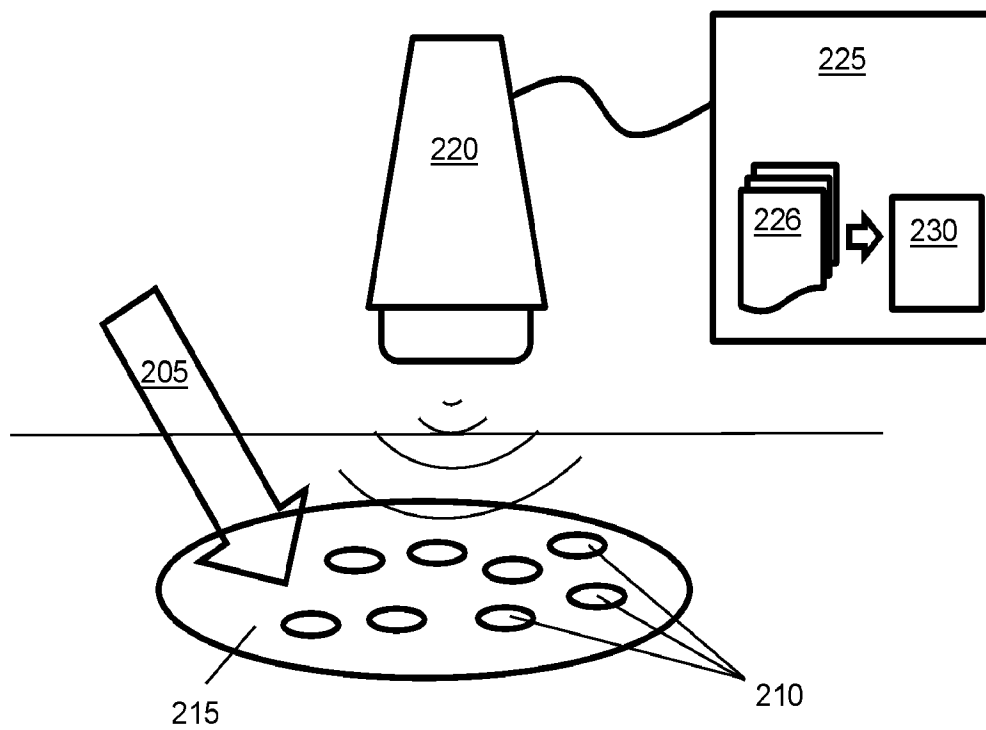
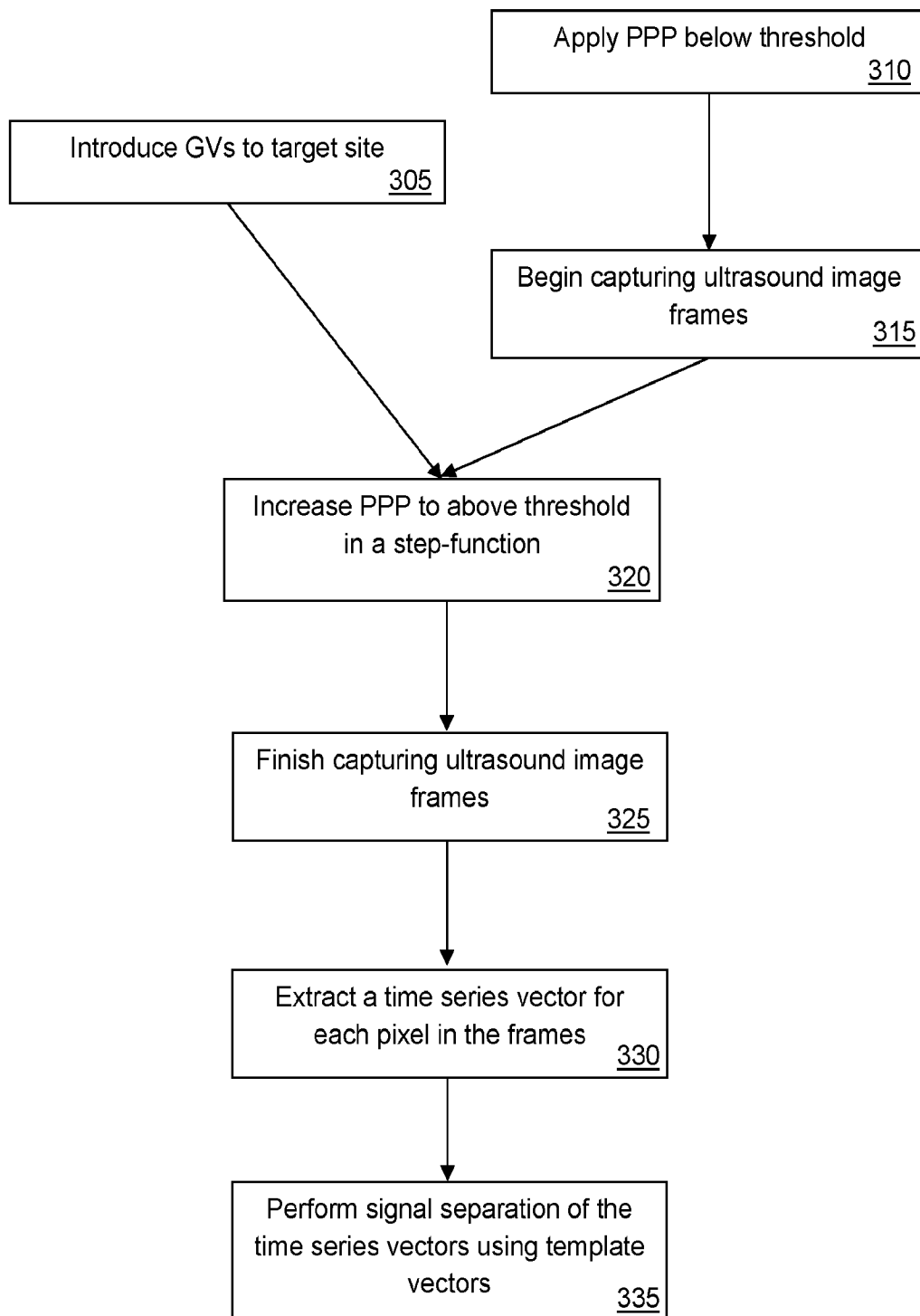


Fig. 1

2/16

*Fig. 2*

3/16

*Fig. 3*

4/16

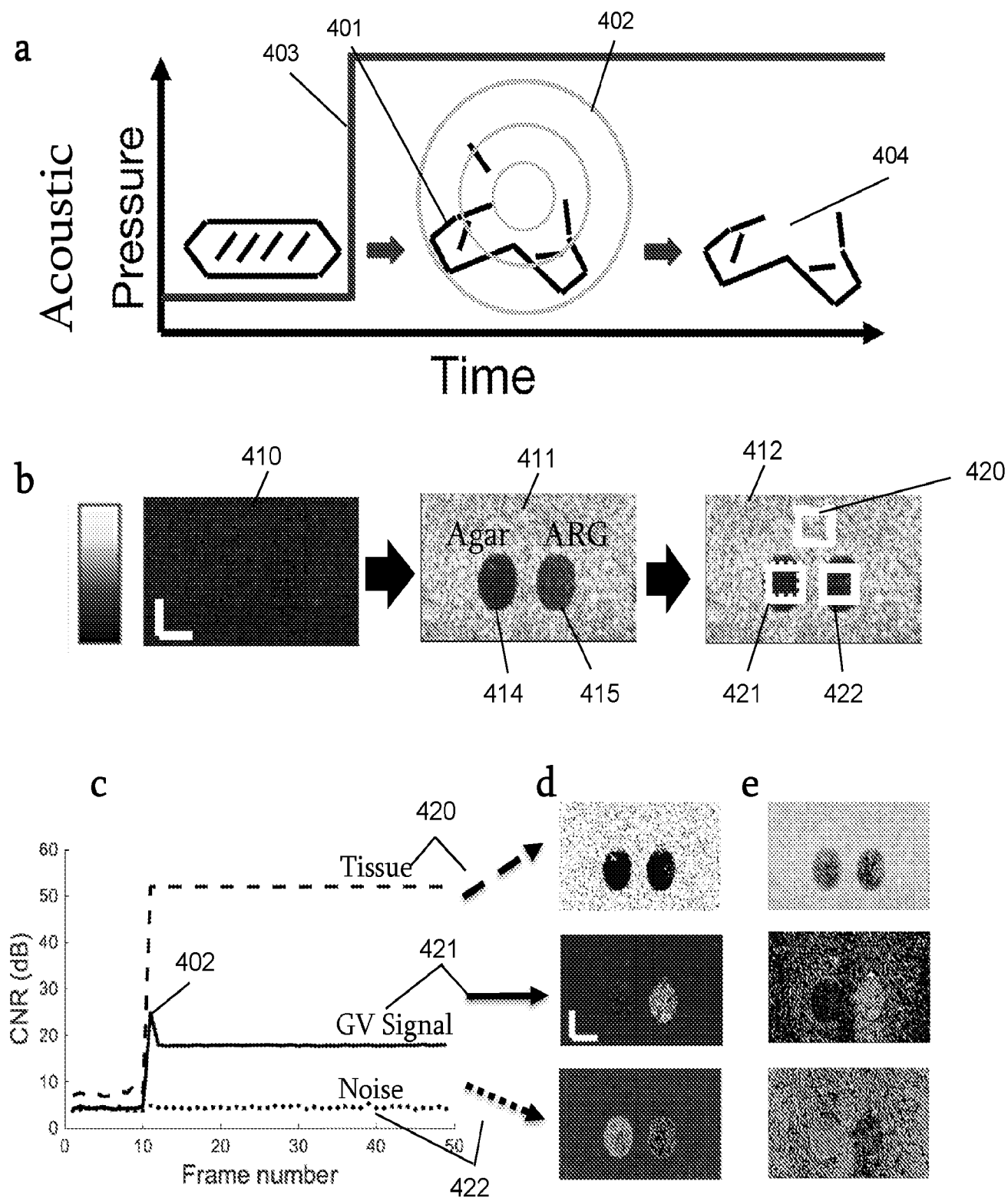


Fig. 4

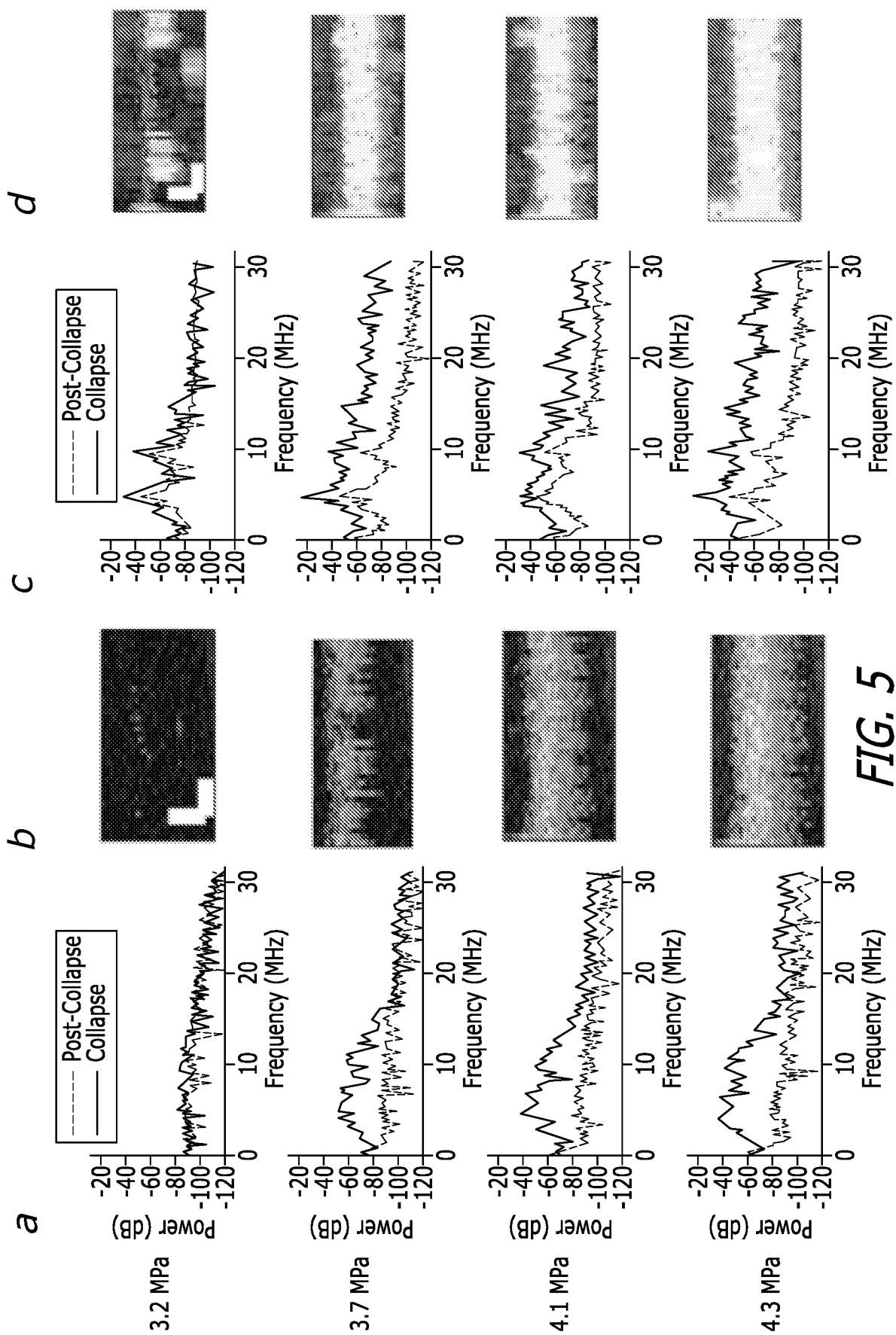


FIG. 5

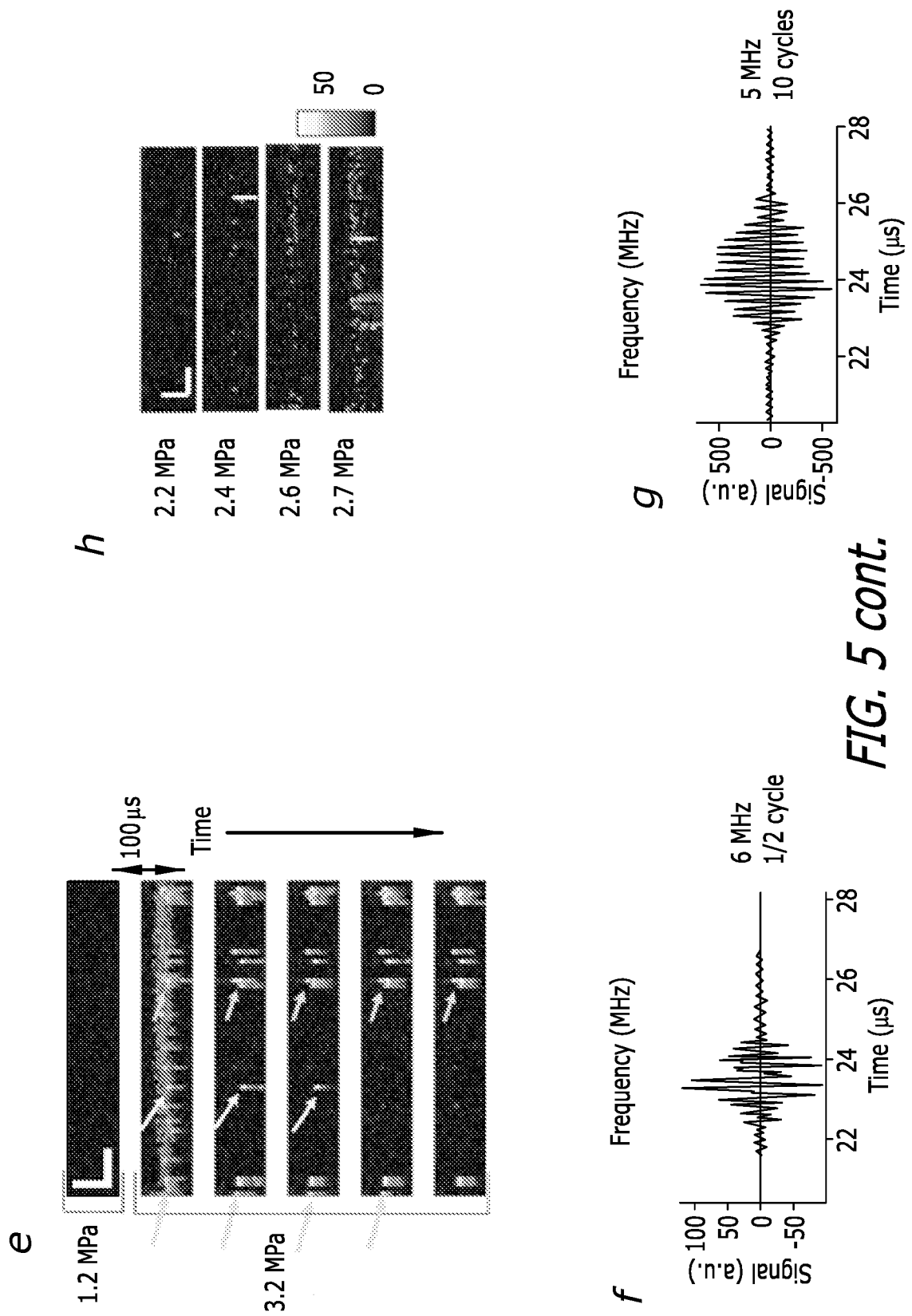


FIG. 5 cont.

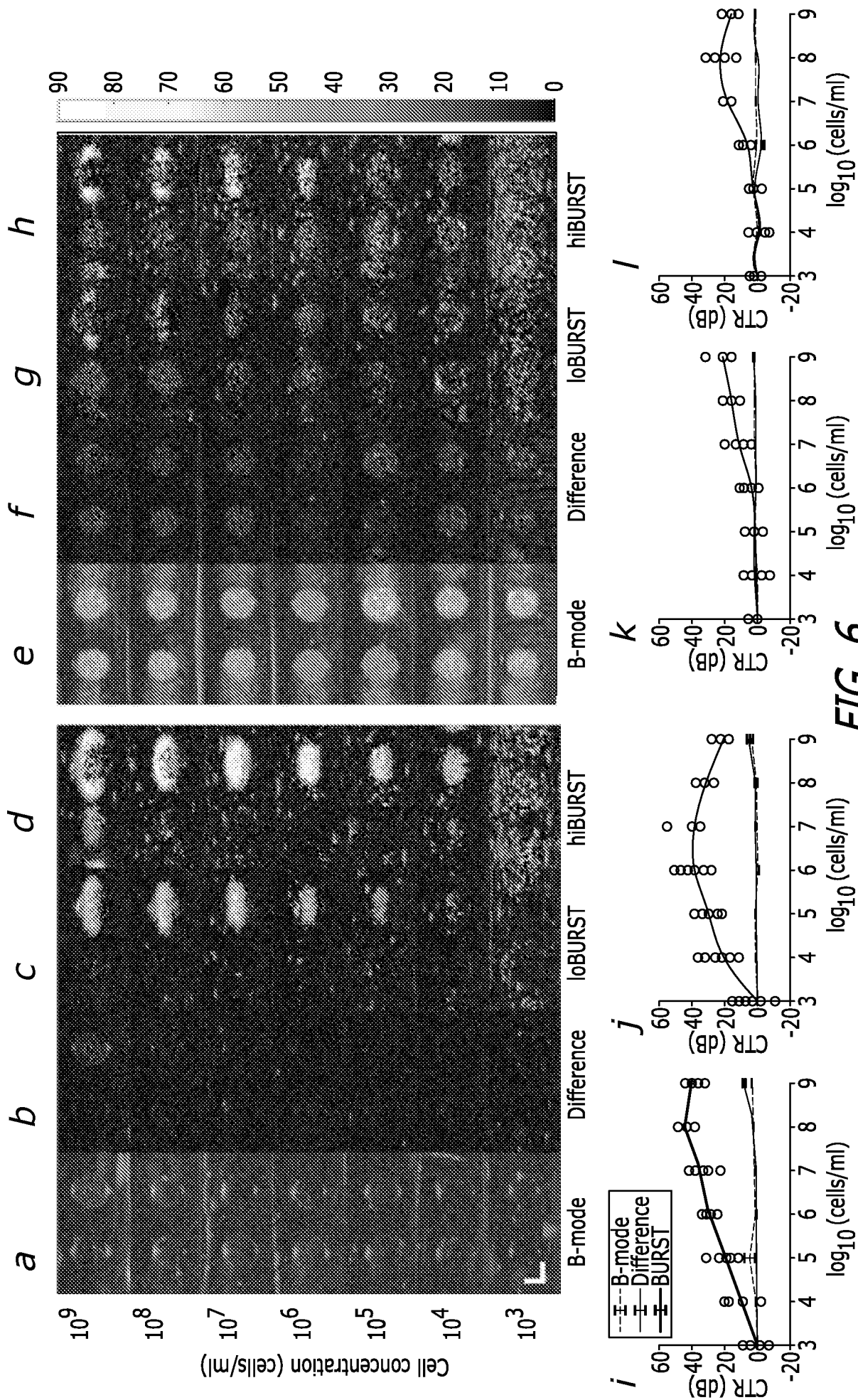


FIG. 6

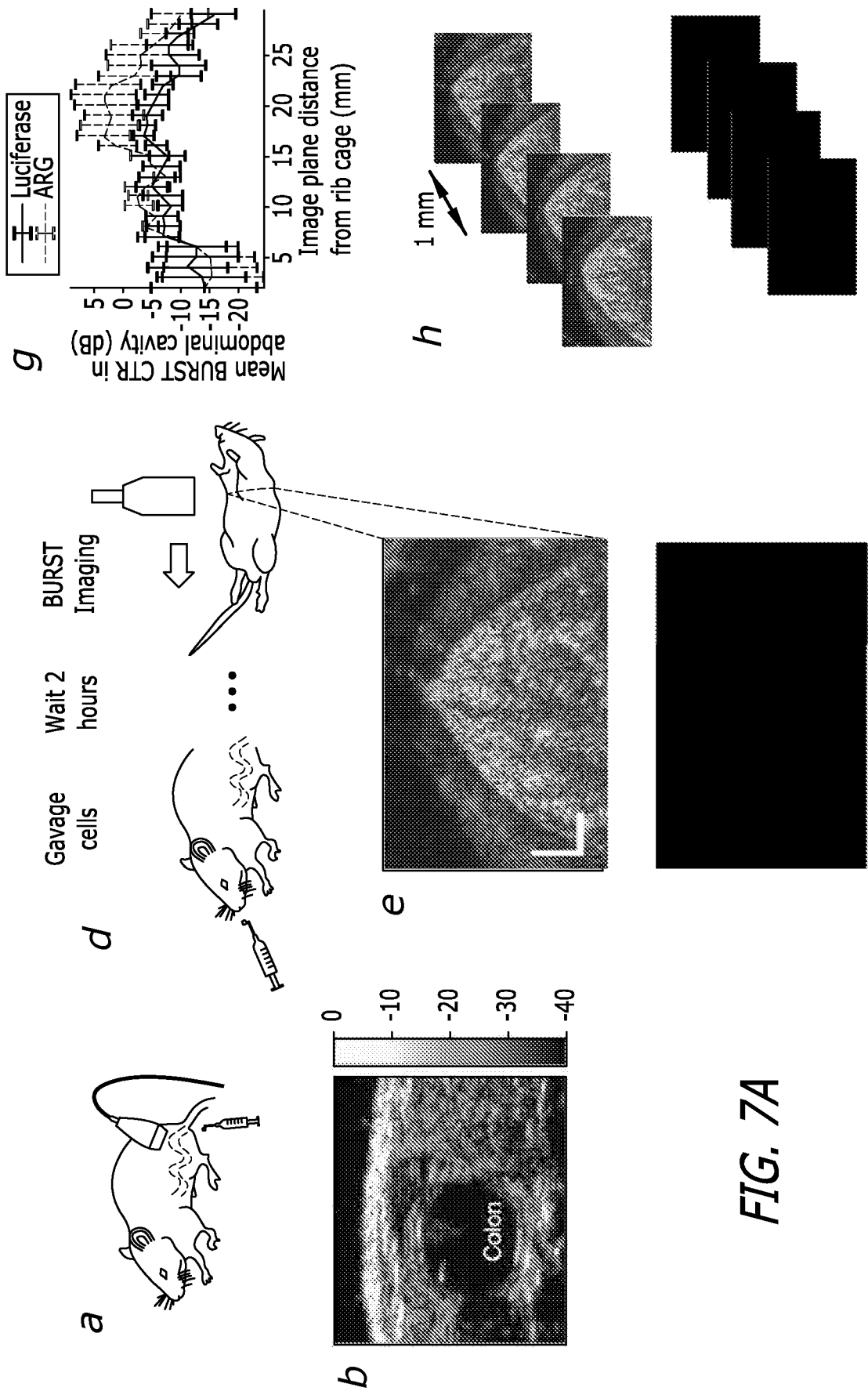
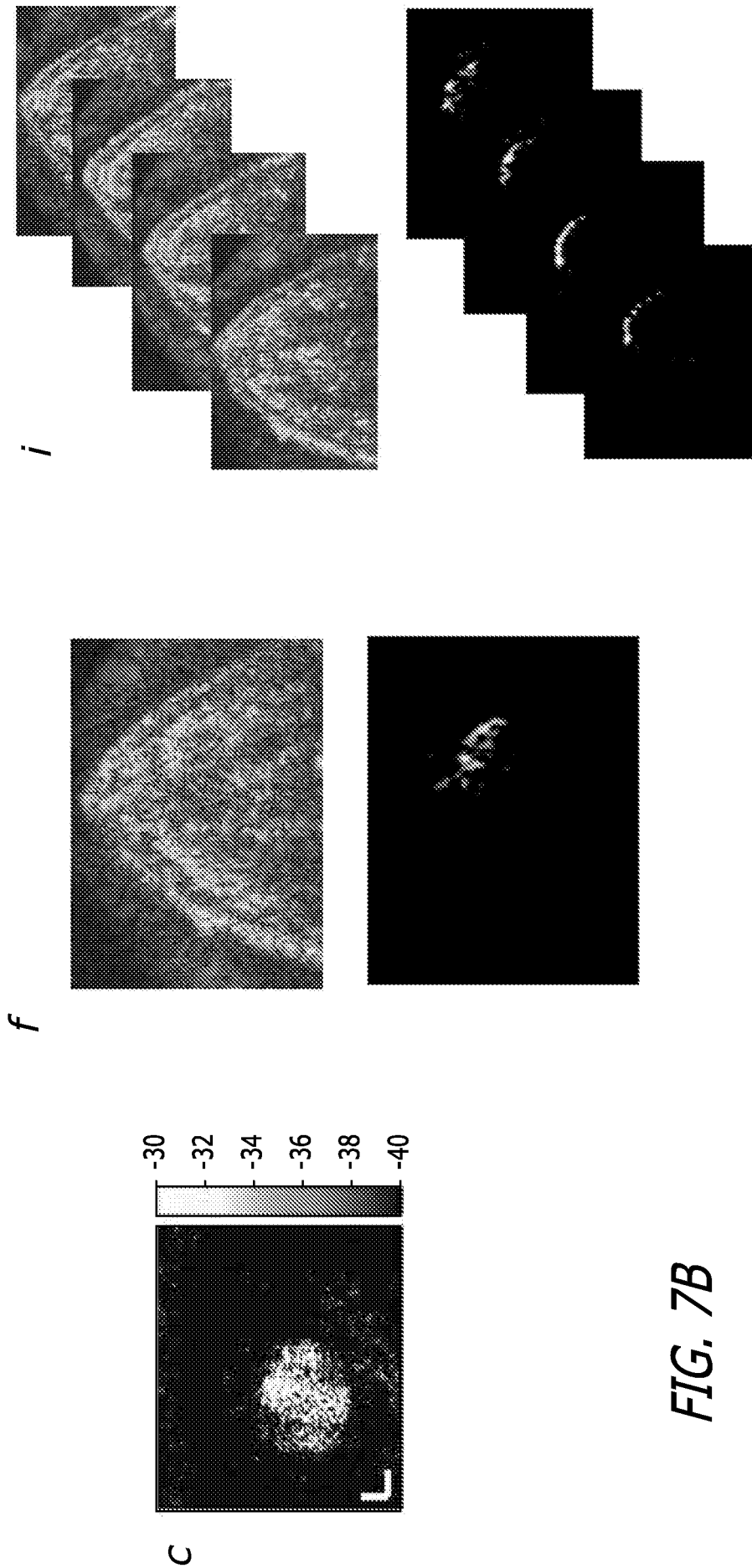


FIG. 7A



10/16

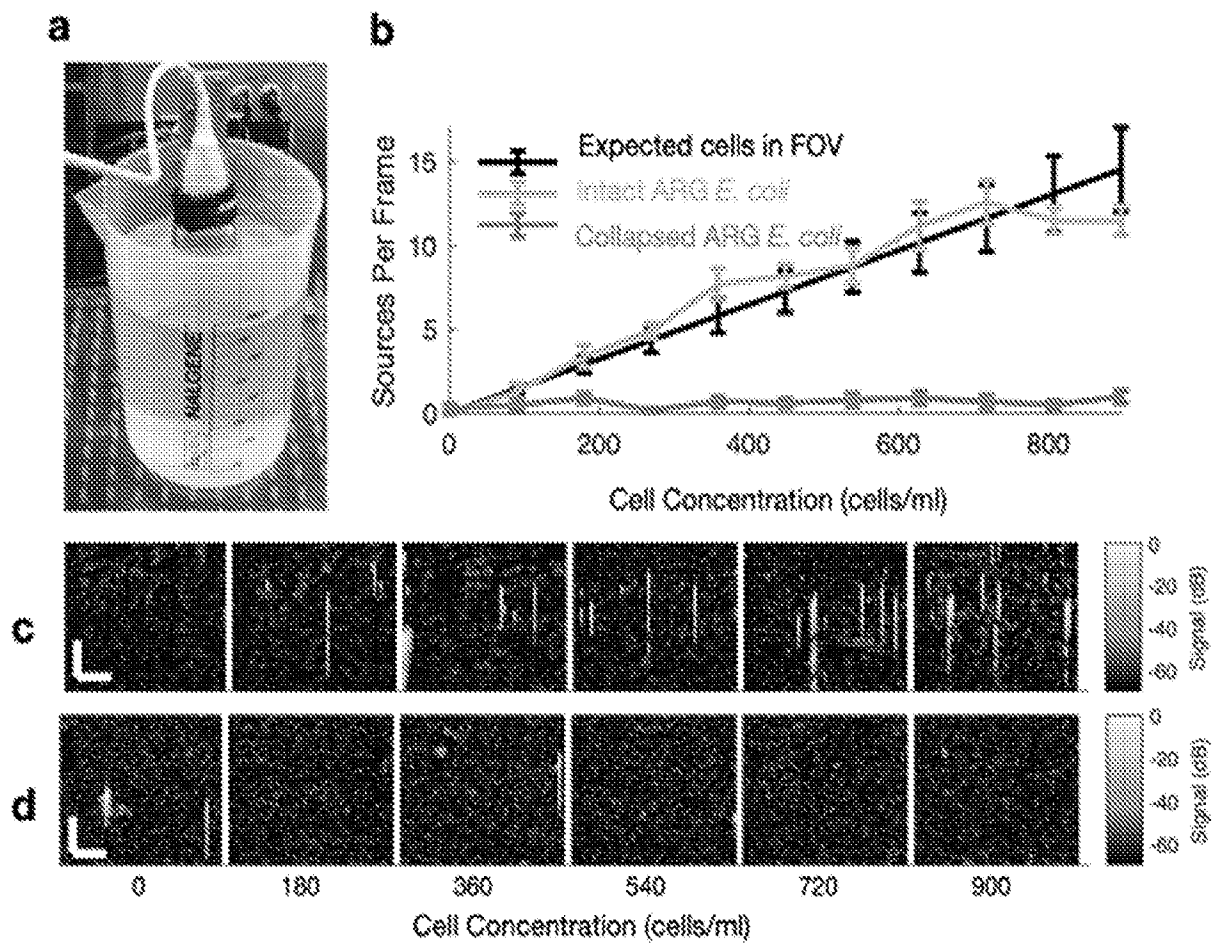


Fig. 8

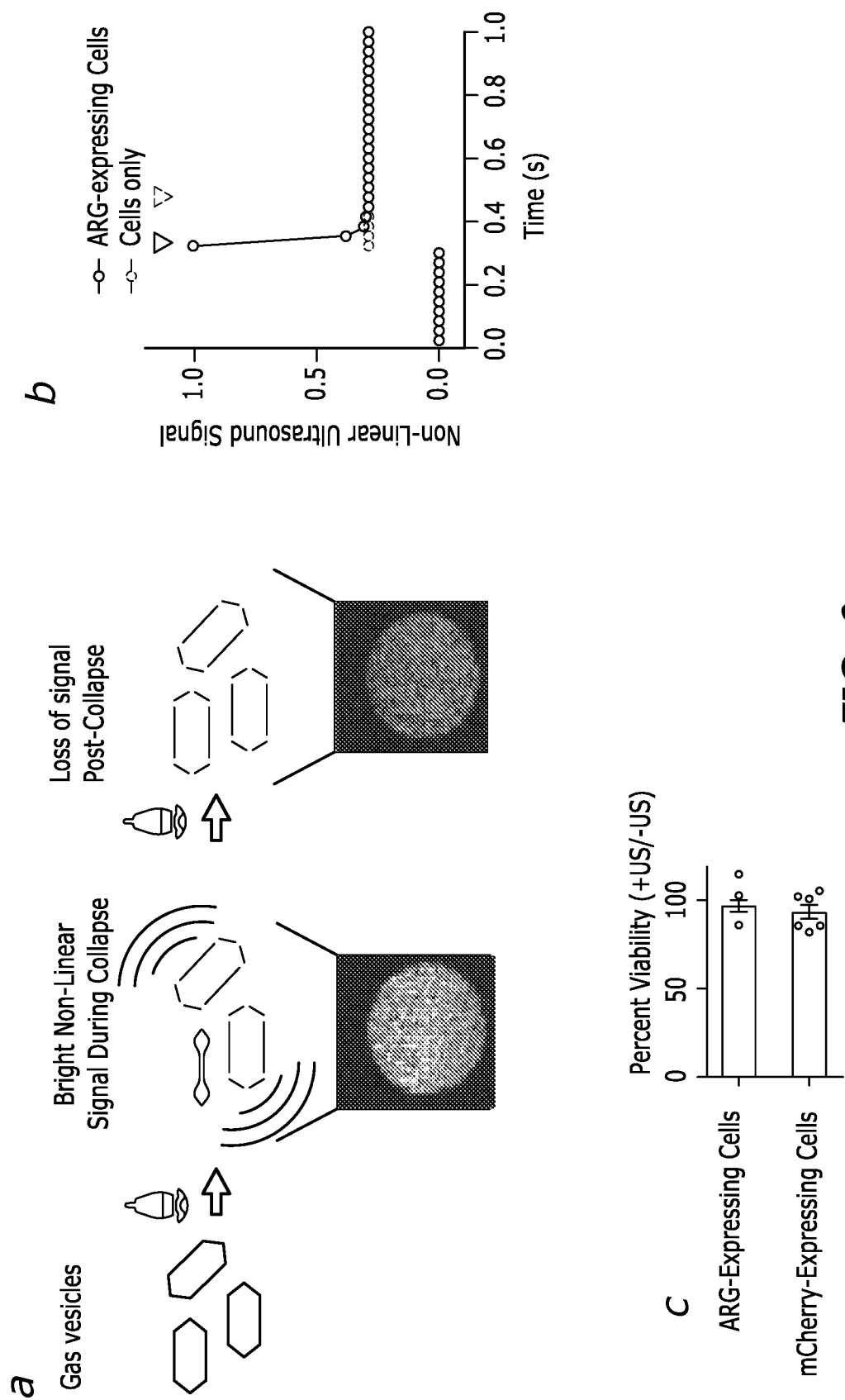


FIG. 9

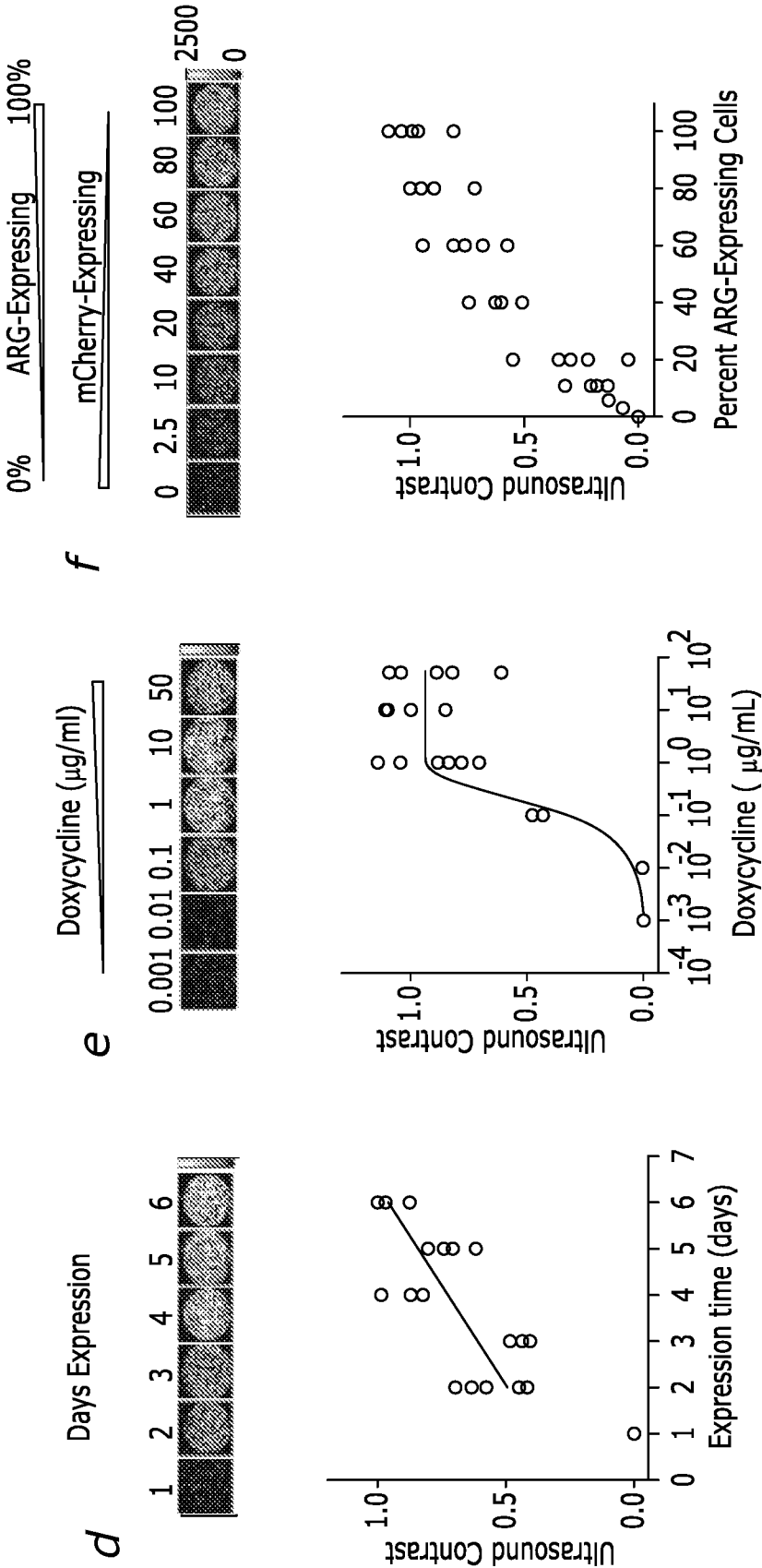


FIG. 9 cont.

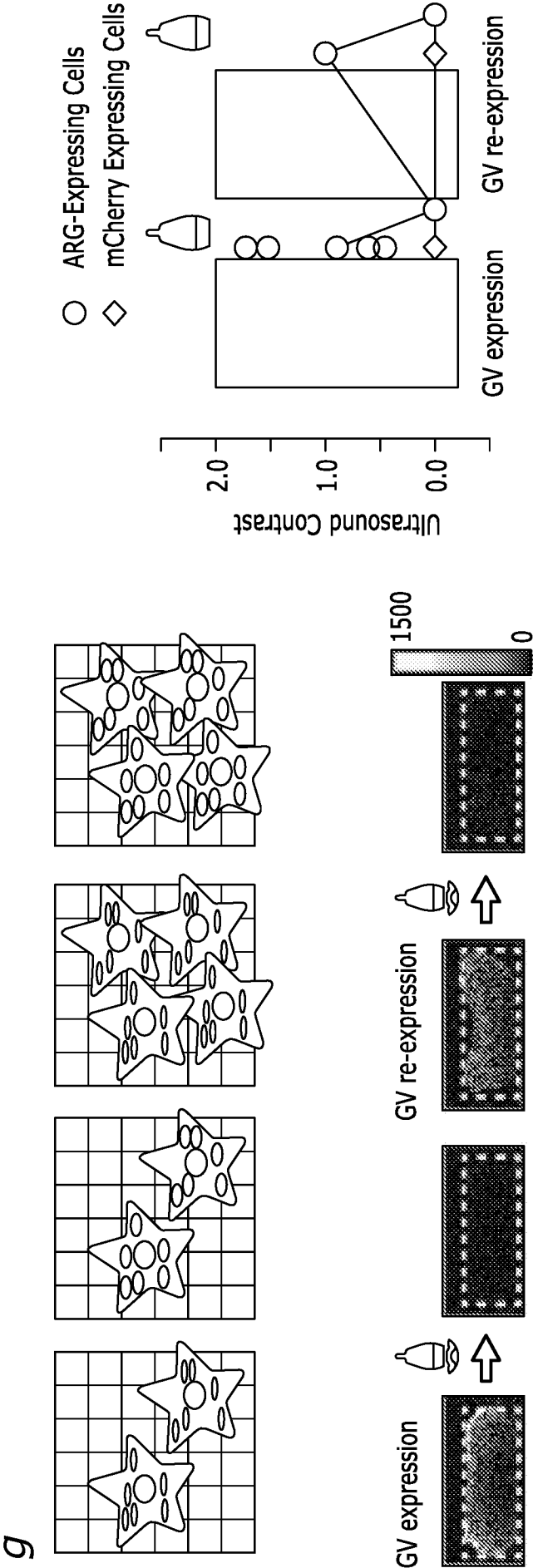
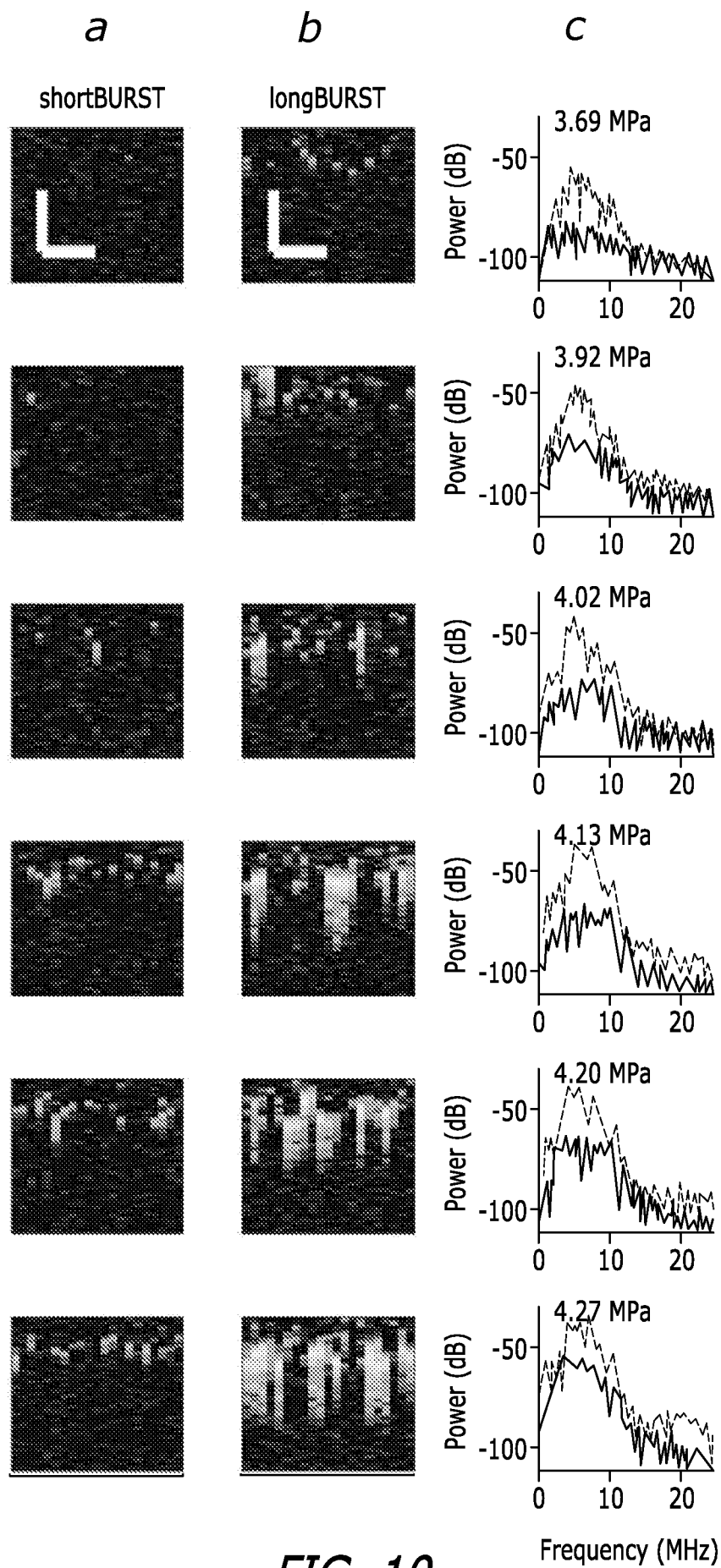


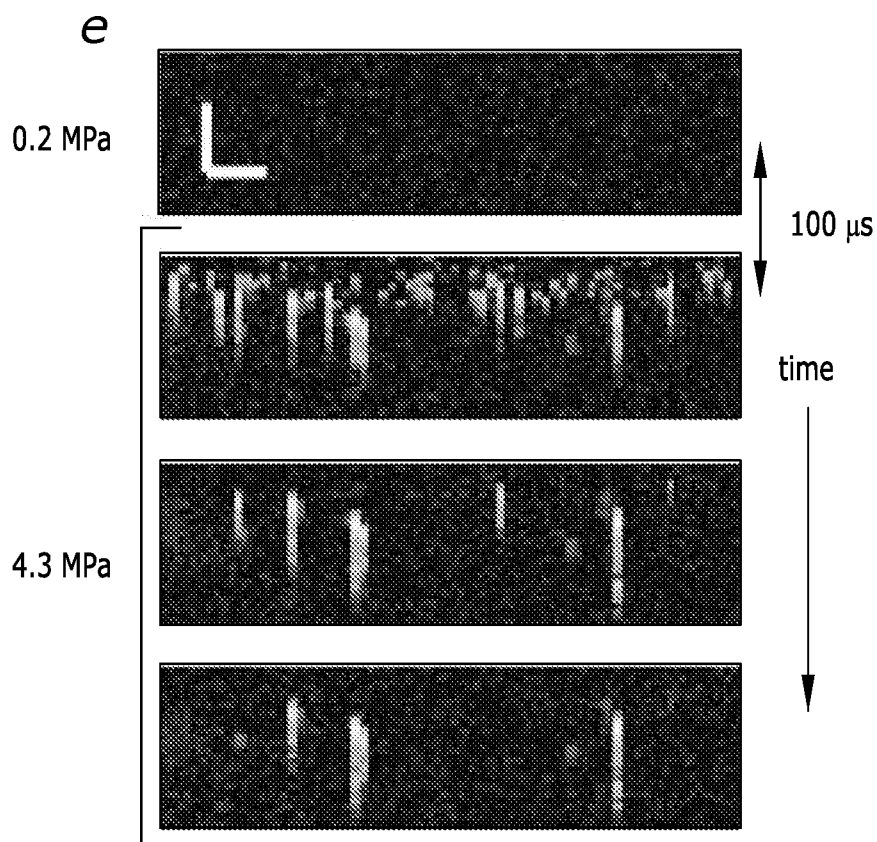
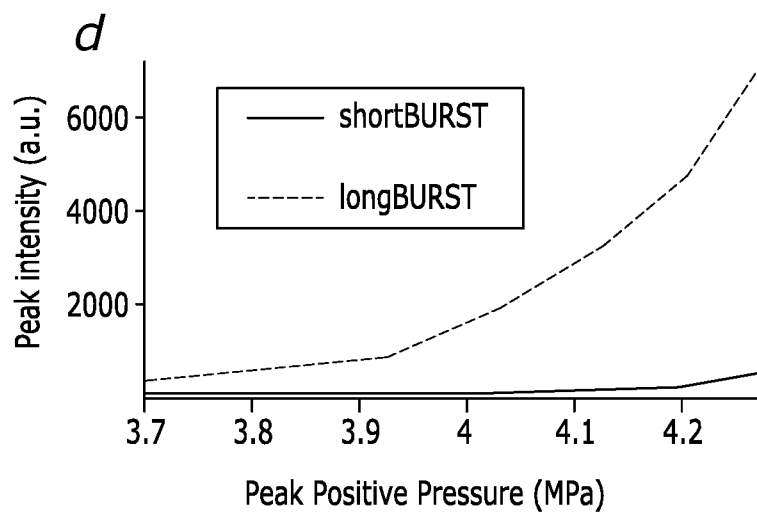
FIG. 9 cont.

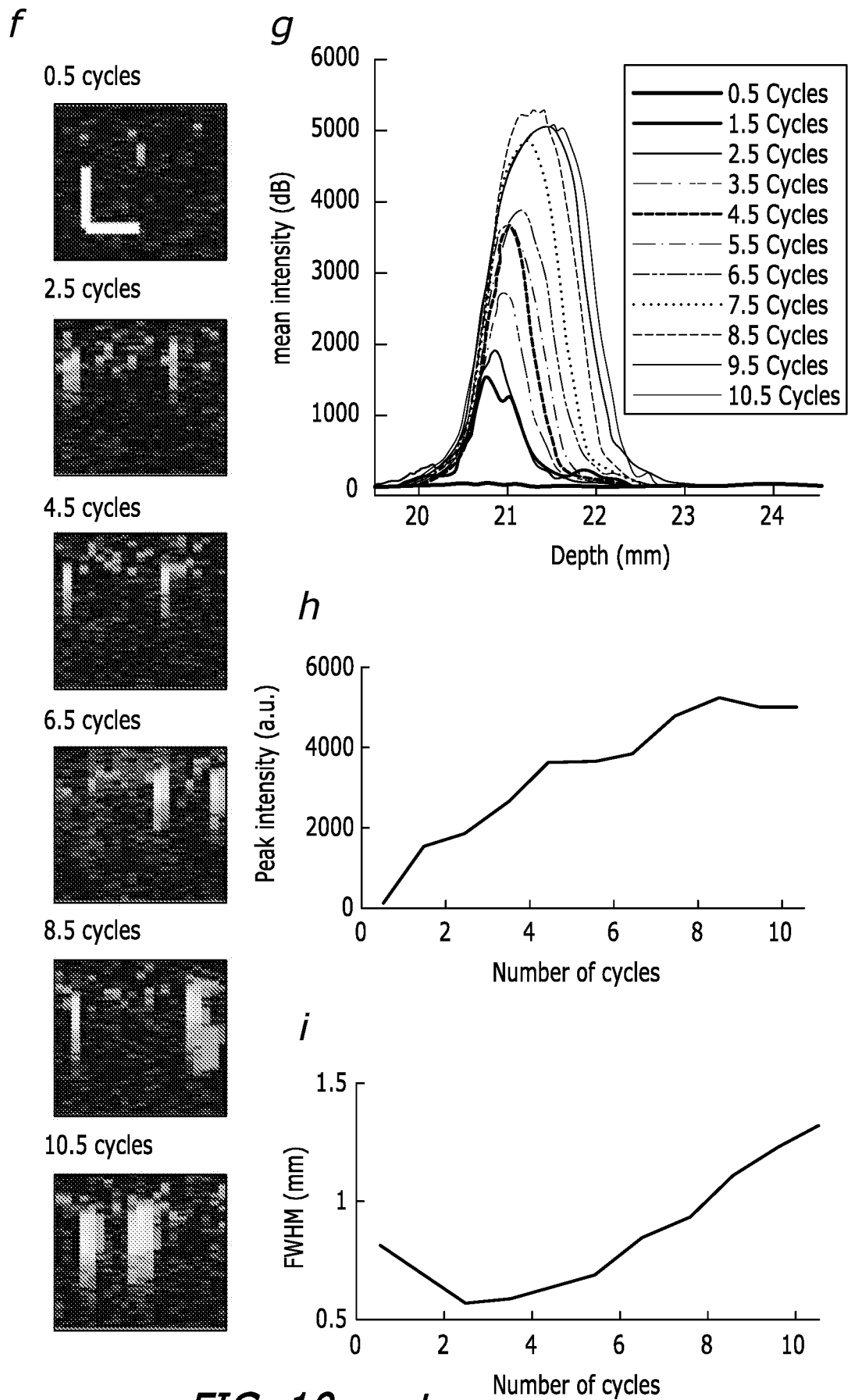
14/16

**FIG. 10**

SUBSTITUTE SHEET (RULE 26)

15/16

*FIG. 10 cont.*

**FIG. 10 cont.**

A. CLASSIFICATION OF SUBJECT MATTER**A61K 49/22(2006.01)i, A61B 8/08(2006.01)i**

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)

A61K 49/22; A61B 5/055; C12N 1/00; C12Q 1/10; A61B 8/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords:ultrasound imaging, gas vesicle, acoustic collapse pressure threshold, peak positive pressure, time-series vector, successive frame

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOURDEAU, R. W. et al., 'Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts', Nature, 04 January 2018, Vol. 553, pp. 86-90, METHODS abstract; page 89, light column, 2nd paragraph; figure 5; "In vitro ultrasound imaging", "In vivo ultrasound and bioluminescence imaging", "Imaging processing" sections in METHODS	1-5,14-18
X	LU, G. J. et al., 'Acoustically modulated magnetic resonance imaging of gas-filled protein nanostructures', Nature Materials, May 2018, Vol. 17, pp. 456-463, Methods abstract; page 456, right column, 2nd paragraph-page 458; figures 2, 3, 5; Methods	1-5,14-18
X	US 2018-0030501 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 01 February 2018 claims 1, 24, 28; paragraphs [0042]-[0044], [0411], [0412], [0419], [0420]; figures 7, 9	1-5,14-18
A	US 2006-0216810 A1 (JU, LU-KWANG) 28 September 2006 the whole document	1-5,14-18
A	CN 105232045 A (XIAMEN UNIVERSITY) 13 January 2016 the whole document	1-5,14-18

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"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

01 May 2020 (01.05.2020)

Date of mailing of the international search report

01 May 2020 (01.05.2020)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

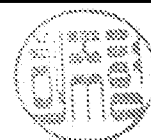
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

HEO, Joo Hyung

Telephone No. +82-42-481-8150



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2020/012557

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	BAR-ZION, A. et al., 'Acoustically detonated biomolecules for genetically encodable inertial cavitation' , BioRxiv, Epub. 29 April 2019, 620567, pp. 1-11 abstract; page 1, right column-page 2, right column; page 5, right column-page 7, left column; page 9, right column; figures 1, 2, 5, 6	1-5,14-18

International application No.
PCT/US2020/012557

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 8,9,11,13,21,22
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claims 8, 9, 11, 13, 21, 22 are regarded to be unclear because they refer to claims which do not comply with PCT Rule 6.4(a).
3. ☒ Claims Nos.: 6,7,10,12,19,20
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2020/012557

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018-0030501 A1	01/02/2018	US 2018-038922 A1	08/02/2018
US 2006-0216810 A1	28/09/2006	AU 5359198 A	03/06/1998
		EP 0956337 A1	17/11/1999
		EP 0956337 A4	22/12/2004
		US 2002-0068356 A1	06/06/2002
		US 2002-0132312 A1	19/09/2002
		US 6036940 A	14/03/2000
		US 6413763 B1	02/07/2002
		US 7022509 B2	04/04/2006
		US 9107949 B2	18/08/2015
		WO 98-21311 A1	22/05/1998
CN 105232045 A	13/01/2016	CN 105232045 B	29/12/2017