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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 November 2015 (05.11.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/164978 A1

(51) International Patent Classification:
G01N 33/18 (2006.01) *C07C 51/42* (2006.01)
B01D 15/08 (2006.01) *C07K 14/195* (2006.01)
C02F 1/28 (2006.01) *C12N 1/19* (2006.01)
C07C 17/38 (2006.01) *C12N 1/21* (2006.01)

(21) International Application Number:
PCT/CA2015/050373

(22) International Filing Date:
1 May 2015 (01.05.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/987,032 1 May 2014 (01.05.2014) US

(71) Applicant: **UNIVERSITY OF MANITOBA** [CA/CA]; Technology Transfer Office, 631 Drake, Centre, 181 Freedman Crescent, Winnipeg, Manitoba R3T 5V4 (CA).

(72) Inventor: **STETEFELD, Joerg**; c/o University of Manitoba, 631 Drake Centre, 181 Freedman Crescent, Winnipeg, Manitoba R3T 5V4 (CA).

(74) Agents: **POLONENKO, Daniel R.** et al.; Gowling Lafleur Henderson LLP, Suite 1600, 421 - 7 Avenue SW, Calgary, Alberta T2P 4K9 (CA).

(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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(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).

[Continued on next page]

(54) Title: BIOREMEDIATION OF POLYCYCLIC AROMATIC HYDROCARBON AND/OR POLYCHLORINATED BIPHENYL CONTAMINATION

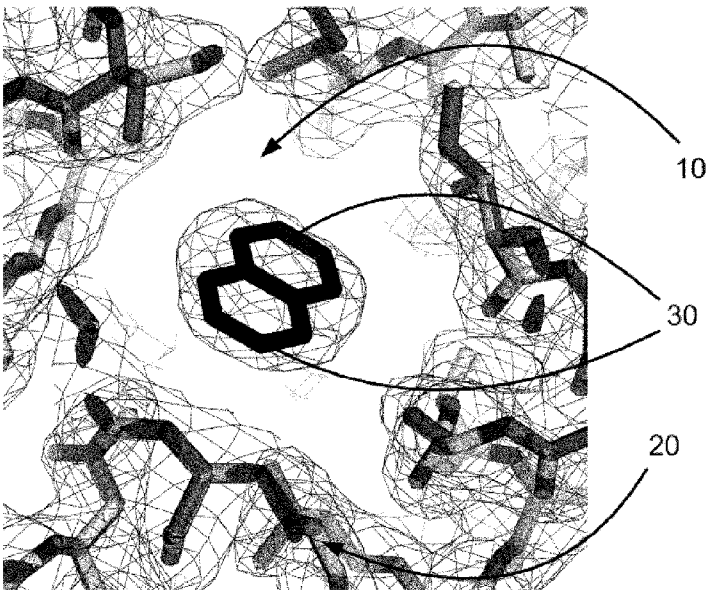

Fig. 9

(57) Abstract: The present disclosure provides uses and methods for using compositions comprising a tetrabrachion protein from Staphylothermus marinus or a fragment thereof for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from natural waterways, from municipal water supply systems, from industrial supply systems, from wastewater storage facilities, and from water purification and recycling facilities.

WO 2015/164978 A1

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))*

— *with sequence listing part of description (Rule 5.2(a))*

TITLE: **BIOREMEDIATION OF POLYCYCLIC AROMATIC HYDROCARBON AND/OR POLYCHLORINATED BIPHENYL CONTAMINATION**

FIELD OF THE INVENTION

The present disclosure relates to methods and compositions for detection and/or recovery of polycyclic aromatic hydrocarbons and/or polychlorinated biphenyls and their associated constituents in fluid samples. More specifically, the present disclosure pertains to compositions comprising a tetrabrachion protein from *Staphylothermus marinus* and/or fragments thereof, and to uses of the compositions for detecting and/or recovering chemical elements from solutions and/or suspensions.

BACKGROUND OF THE INVENTION

Polycyclic aromatic hydrocarbons such as naphthalene and naphthenic acids are naturally occurring constituents of crude oil and are generally categorized as saturated acyclic and cyclic carboxylic acids in which the carboxylic acid group is attached to an aliphatic sidechain or alternatively, to a cycloaliphatic ring (a single benzene ring or multiple-fused benzene rings). More precisely, polycyclic aromatic hydrocarbons can be described as a complex mixture of saturated alkyl-substituted acyclic and cyclic aliphatic carboxylic acids with a general formula $C_nH_{2n+Z}O_2$, where n indicates the carbon number, and Z specifies the number of rings in the molecule. Polycyclic aromatic hydrocarbons are quite corrosive and are removed from crude oil (e.g., bitumen) during its extraction from oil fields and oil sands, and during subsequent refining of the recovered crude oil. The consequence is that water used for recovery and processing of crude oil becomes increasingly enriched with polycyclic aromatic hydrocarbons. Significant amounts of dissolved polycyclic aromatic hydrocarbons remain in waste process water and consequently, are stored in large tailings ponds. It is also known that coal deposits comprise naphthenic acid constituents that are assimilated into water used to recover and process coal ores. Additionally, polycyclic aromatic hydrocarbons, commonly recovered during crude oil refining and isolated

from conventional petroleum distillates, are subsequently used as source materials in the manufacture of corrosion inhibitors, wood preservatives, lubricant and fuel additives, driers for paints and inks, and in the production of metal soaps, and as a consequence, may end up as contaminants in industrial waste water streams.

Polychlorinated biphenyls (PCBs) are another class of compounds characterized by a core of two benzene rings, and are closely related to polycyclic aromatic hydrocarbons. PCBs are of particular concern because of their environmental toxicity and serious health effects on human, animal, and aquatic species. PCBs are very stable compounds and do not readily decompose because of their resistance to chemical oxidation and reduction in natural environments. Furthermore, PCBs have long half-lives ranging from eight to fifteen years, and are generally insoluble in water. Furthermore, the destruction of PCBs by chemical, thermal, and biochemical processes is extremely difficult, and consequently, most countries have banned the commercial production and use of PCBs. Because of the difficulties in safely destroying PCBs and their long persistence in the environment, considerable quantities of PCBs have been stored in evacuated buildings and municipal landfill sites. However, many of these storage sites were not design to safely contain PCBs and as result PCBs are able to escape into the atmosphere and groundwaters.

The problem with both classes of compounds, i.e., polycyclic aromatic hydrocarbons and PCBs, is that their elevated levels natural waterways are extremely toxic to aquatic life including snails, bacteria, fish, plants, mammals, zooplankton, and algae. Furthermore, it is possible for these pollutants to move from the natural waterways into municipal water supply systems.

SUMMARY OF THE INVENTION

The present disclosure provides use of tetrabrachion protein or a fragment thereof from *Staphylothermus marinus* or a composition comprising tetrabrachion protein or a fragment thereof for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from fluid systems exemplified by waterways, municipal water supply systems, industrial water supply systems, waste water storage facilities,

water purification and recycling facilities and from related sludges, sediments, and suspensions. The present disclosure further discloses methods for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from solutions and/or suspensions using a tetrabrachion protein or a fragment thereof.

According to an aspect of the present disclosure, use of a tetrabrachion protein or a fragment thereof from *S. marinus* for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from natural waterways, from municipal water supply systems, from industrial supply systems, from wastewater storage facilities, and from water purification and recycling facilities is disclosed.

According to another aspect of the present disclosure, a use of a composition comprising a tetrabrachion protein or a fragment thereof from *S. marinus* and a carrier therefor for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment is disclosed.

According to a further aspect of the present disclosure, a method for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment is disclosed, wherein the method comprises: (a) introducing *S. marinus* into the solution to produce tetrabrachion protein or a fragment thereof; and (b) providing conditions that permit the tetrabrachion protein or the fragment thereof to recover a naphthenic acid and/or a PCB from a water system and/or a suspension and/or a sludge and/or a sediment.

According to another aspect of the present disclosure, a method for detecting and/or recovering a naphthenic acid and/or a PCB from a water system and/or a suspension and/or a sludge and/or a sediment is disclosed, wherein the method comprises: (a) introducing tetrabrachion protein or a fragment thereof from *S. marinus* into the solution; and (b) providing conditions that permit the tetrabrachion protein or the fragment thereof to recover the naphthenic acid and/or the PCB from a solution and/or a suspension.

According to another aspect of the present disclosure, a method for detecting and/or recovering a naphthenic acid and/or a PCB from a water system and/or a

suspension and/or a sludge and/or a sediment is disclosed, wherein the method comprises: (a) a composition comprising tetrabrachion protein or a fragment thereof from *S. marinus* into the a water system and/or a suspension and/or a sludge and/or a sediment; and (b) providing conditions that permit the tetrabrachion protein or the fragment thereof to recover the naphthenic acid and/or the PCB from the water system and/or suspension and/or sludge and/or sediment.

BRIEF DESCRIPTION OF THE DRAWINGS

The features of the present disclosure will become more apparent from the following description in which reference is made to the appended drawings wherein:

Fig. 1 shows an exemplary diagrammatic representation of the canopy-like arrangement of the tetrabrachion stalk anchored to the *Staphylothermus marinus* cell membrane;

Fig. 2 shows an exemplary diagrammatic representation of the tetrabrachion stalk illustrating the dimensions of the tetrabrachion stalk. The spherical balls represent the STABLE protease that binds to the RHCC polypeptide chain fragment of tetrabrachion;

Fig. 3 shows the RHCC domain of tetrabrachion. Fig. 5(a) shows a side view of the four helices of the RHCC domain of tetrabrachion at 1.8 Å resolution. The N-terminus is at the bottom, and the C-terminus is at the top of the figure. Fig. 5(b) shows an axial view from the N-terminus of the RHCC domain;

Fig. 4 shows an amino acid sequence of a tetrabrachion protein from *S. marinus* (SEQ ID NO: 1);

Fig. 5 shows a 52-amino-acid sequence of the RHCC polypeptide chain fragment of tetrabrachion (SEQ ID NO: 2);

Fig. 6 shows a listing of the amino acid sequence (SEQ ID NO: 3) of a RHCC polypeptide chain fragment of tetrabrachion and the codon optimized nucleotide

sequence (SEQ ID NO: 2) encoding the RHCC polypeptide chain fragment used in the exemplary embodiments of the present disclosure;

Fig. 7 is a chart showing the fluorescence anisotropy of various concentrations of $RHCC_{tet}$ in the presence of 20 μM naphthalene measured at 340 nm with an excitation wavelength of 286 nm;

Fig. 8 shows a RHCC polypeptide chain fragment of tetrabrachion with two naphthalene molecules within the hydrophobic channel; and

Fig. 9 is an electron density presentation from a cross-section of a co-crystal structure of a RHCC polypeptide chain fragment with two naphthalene molecules within the hydrophobic channel.

DETAILED DESCRIPTION

The present disclosure pertains to use of compositions comprising a right-handed coiled coil (RHCC) structure or domain of the tetrabrachion protein from a *Staphylothermus marinus* microorganism for detection and/or recovery of polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment.

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 invention belongs. Certain terms are discussed in the specification to provide additional guidance to the practitioner in describing the methods, uses and the like of embodiments of the invention, and how to make or use them. It will be appreciated that the same thing may be said in more than one way. Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein. No significance is to be placed upon whether or not a term is elaborated or discussed herein. Recital of one or a few synonyms or equivalents does not exclude use of other synonyms or equivalents, unless it is explicitly stated. Use of examples in the specification, including examples of terms, is for illustrative purposes only and does not limit the scope and meaning of the embodiments of the invention herein. Although

any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described.

To facilitate understanding of the disclosure, the following definitions are provided.

As used herein, the term "waterway" means any naturally occurring or man-made body of water exemplified by lakes, ponds, pools, reservoirs, oceans, seas, and/or conduits wherein water moves along and/or into and/or out of a body of water such as exemplified by rivers, streams, channels, canals and the like.

As used herein, a "water system" means a man-made infrastructure for conveying water between a source and a destination, and includes pumping devices, filtering apparatus, treatment apparatus for settling, and/or clarifying and/or purifying and/or chemically treating, and/or disinfecting and/or sterilizing the water. Examples of water systems are municipal water systems for receiving water from waterways, treating the received water to make it potable water, conveying the water for consumption, receiving wastewater and sewage and other water-borne materials and conveying the same to sewage treatment plants from which the treated water is discharged into a water and/or alternatively, further treated to produce potable water for municipal use. Other examples of water systems include industrial water systems for receiving water from waterways, optionally treating the water, and then using the water for or in industrial processes exemplified by processing natural ores for separation and recovery of minerals, refining of petroleum products, extraction of hydrocarbons from oil sands, manufacturing of chemicals, food processing and the like.

As used herein, "solution" refers to any chemical element-containing liquid including, without limitation, any organic liquid, waste water, mine tailings, oils, coal and effluents produced by various processes.

As used herein, "suspension" refers to any suspension or slurry that is a heterogeneous mixture containing solid particles of chemical elements.

As used herein, "sludge" refers to any residual, semi-solid material left behind from an industrial wastewater treatment process or of from a sewage treatment process, or it may be used herein as a generic term for solids separated from suspension in a liquid.

As used herein, "sediment" refers to naturally occurring materials that have broken down by processes of weathering and/or erosion and are transported by water or have settled out of water.

As used herein, "recover", "recovered", "recovering" or "recovery" refers to the obtaining of and/or extraction of and/or separation of one or more chemical elements from a solution and/or a suspension.

As used herein, "STABLE" refers to the stalk-associated archaeabacteria endo-protease proteins that bind specifically to the right-handed coiled coil polypeptide chain fragment of the tetrabrachion protein from *Staphylothermus marinus*

As used herein, "RHCC" refers to the right-handed coiled coil polypeptide chain fragment of tetrabrachion from *S. marinus* comprising 52 amino acid residues and to which the STABLE protease binds.

As used herein, "host cell" refers to a cell of any microorganism into which a nucleic acid construct encoding tetrabrachion or a fragment thereof can be transformed. The host cell is not to be considered limiting in any manner, and can be, but is not limited to, a mammalian cell, an insect cell, a bacterial cell or a yeast cell, exemplified by and including, without limitation, *Escherichia* sp., *Pseudomonas* sp., *Bacillus* sp., *Saccharomyces* sp., *Schizosaccharomyces* sp. and *Candida* sp.

As used herein, the term "synthetic DNA" means DNA sequences that have been prepared entirely or at least partially by chemical means. Synthetic DNA sequences may be used, for example, for modifying native DNA sequences in terms of codon usage and expression efficiency.

The word "comprise" or variations such as "comprises" or "comprising" will be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers.

As used herein, the word "complexed" means attached together by one or more linkages.

The term "a cell" includes a single cell as well as a plurality or population of cells.

The term "about" or "approximately" means within 20%, preferably within 10%, and more preferably within 5% of a given value or range.

The term "nucleic acid" refers to a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acid includes polyribonucleic acid (RNA) and polydeoxyribonucleic acid (DNA), both of which may be single-stranded or double-stranded. DNA includes cDNA, genomic DNA, synthetic DNA, and semisynthetic DNA.

The term "gene" refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids.

The term "recombinant DNA molecule" refers to a DNA molecule that has undergone a molecular biological manipulation.

The term "vector" refers to any means for the transfer of a nucleic acid into a host cell. A vector may be a replicon to which another DNA segment may be attached so as to bring about the replication of the attached segment. A "replicon" is any genetic element (e.g., plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*, i.e., capable of replication under its own control. The term "vector" includes plasmids, DNA-protein complexes, and biopolymers. In addition to a nucleic acid, a vector may also contain one or more regulatory regions, and/or selectable markers useful in selecting, measuring, and monitoring nucleic acid transfer results (transfer to which tissues, duration of expression, etc.).

The term "cloning vector" refers to a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment. Cloning vectors may be capable of replication in one cell type, and expression in another ("shuttle vector").

A cell has been "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. A cell has been "transformed" by exogenous or heterologous DNA when the transfected DNA effects a phenotypic change. The transforming DNA can be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

The term "nucleic acid molecule" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester anologs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA—DNA, DNA-RNA and RNA—RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms.

Modification of a genetic and/or chemical nature is understood to mean any mutation, substitution, deletion, addition and/or modification of one or more residues. Such derivatives may be generated for various purposes, such as in particular that of enhancing its production levels, that of increasing and/or modifying its activity, or that of conferring new pharmacokinetic and/or biological properties on it. Among the derivatives resulting from an addition, there may be mentioned, for example, the chimeric nucleic acid sequences comprising an additional heterologous part linked to one end, for example of the hybrid construct type consisting of a cDNA with which one or more introns would be associated.

Likewise, for the purposes of the invention, the claimed nucleic acids may comprise promoter, activating or regulatory sequences, and the like.

The term "promoter sequence" refers to a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3′ direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3′ terminus by the transcription initiation site and extends upstream (5′ direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background.

The term "homologous" in all its grammatical forms and spelling variations refers to the relationship between proteins that possess a "common evolutionary origin," including homologous proteins from different species. Such proteins (and their encoding genes) have sequence homology, as reflected by their high degree of sequence similarity. This homology is greater than about 75%, greater than about 80%, greater than about 85%. In some cases the homology will be greater than about 90% to 95% or 98%.

"Amino acid sequence homology" is understood to include both amino acid sequence identity and similarity. Homologous sequences share identical and/or similar amino acid residues, where similar residues are conservative substitutions for, or "allowed point mutations" of, corresponding amino acid residues in an aligned reference sequence. Thus, a candidate polypeptide sequence that shares 70% amino acid homology with a reference sequence is one in which any 70% of the aligned residues are either identical to, or are conservative substitutions of, the corresponding residues in a reference sequence.

The term "polypeptide" refers to a polymeric compound comprised of covalently linked amino acid residues. Amino acids are classified into seven groups on the basis of the side chain R: (1) aliphatic side chains, (2) side chains containing a hydroxylic (OH) group, (3) side chains containing sulfur atoms, (4) side chains containing an acidic or amide group, (5) side chains containing a basic group, (6) side chains containing an aromatic ring, and (7) proline, an imino acid in which the side chain is fused to the amino group. A polypeptide of the invention preferably comprises at least about 14 amino acids.

The term "protein" refers to a polypeptide which plays a structural or functional role in a living cell.

A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced (if the coding sequence contains introns) and translated into the protein encoded by the coding sequence.

The terms "codon optimization" and "codon optimized" mean the selection of appropriate DNA nucleotides for the synthesis of oligonucleotides of a sequence encoding a tetrabrachion protein or a fragment thereof using codons that are typically utilized within the target host.

The term "corresponding to" is used herein to refer to similar or homologous sequences, whether the exact position is identical or different from the molecule to which the similarity or homology is measured. A nucleic acid or amino acid sequence alignment may include spaces. Thus, the term "corresponding to" refers to the sequence similarity, and not the numbering of the amino acid residues or nucleotide bases.

The term "derivative" refers to a product comprising, for example, modifications at the level of the primary structure, such as deletions of one or more residues, substitutions of one or more residues, and/or modifications at the level of one or more residues. The number of residues affected by the modifications may be, for example, from 1, 2 or 3 to 10, 20, or 30 residues. The term derivative also comprises the molecules comprising additional internal or terminal parts, of a peptide nature or otherwise. They may be in particular active parts, markers, amino acids, such as methionine at position −1. The term derivative also comprises the molecules comprising modifications at the level of the tertiary structure (N-terminal end, and the like). The term derivative also comprises sequences homologous to the sequence considered, derived from other cellular sources, and in particular from cells of human origin, or from other organisms, and possessing activity of the same type or of substantially similar type. Such homologous sequences may be obtained by hybridization experiments. The hybridizations may be performed based on nucleic acid

libraries, using, as probe, the native sequence or a fragment thereof, under conventional stringency conditions or preferably under high stringency conditions.

The term "operatively linked" means that the particular sequences, for example a regulatory element and a coding region of interest, interact either directly or indirectly to carry out an intended function, such as mediation or modulation of gene expression. The interaction of operatively linked sequences may, for example, be mediated by proteins that interact with the operatively linked sequences. A coding region of interest, such as the nucleotide sequence encoding tetrabrachion protein, may also be introduced within a vector along with other sequences, typically heterologous, to produce a chimeric construct. A transcriptional regulatory region and a sequence of interest are "operably linked" when the sequences are functionally connected so as to permit transcription of the sequence of interest to be mediated or modulated by the transcriptional regulatory region.

The terms "regulatory region" and "regulatory element" mean a nucleic acid sequence that has the property of controlling the expression of a sequence that is operatively linked with the regulatory region. Such regulatory regions may include promoter or enhancer regions, and other regulatory elements recognized by one of skill in the art. By "promoter" it is meant the nucleotide sequences at the 5' end of a coding region, or fragment thereof, that contain all the signals essential for the initiation of transcription and for the regulation of the rate of transcription. There are several types of regulatory elements, including those that are inducible, constitutive or the like. A regulatory element may be derived from any suitable source provided that the regulatory element is active in the host cell. A regulatory element may be an animal nucleic acid sequence, a bacterial nucleic acid sequence, a viral nucleic acid sequence, a protozoan nucleic acid sequence, or a yeast nucleic acid sequence, provided that the regulatory element functions within the host cell in which it is used. A regulatory element may comprise, in whole or in part, synthetic nucleic acid sequences not found in nature (a synthetic regulatory element).

Staphylothermus marinus is a marine hyperthermophilic Archaea microorganism, isolated from geothermally heated sediments and from a "black

smoker" on the ocean floor. *S. marinus* requires elemental sulfur for growth. The optimum temperature for growth of *S. marinus* is 85°C in minimal medium and 92°C in rich medium. *S. marinus* is capable of tolerating wide ranges of pH (2-10), high redox potential, pressure and salinity. *S. marinus* possesses a filiform glycoprotein complex tetrabrachion that forms the surface (S-) layer of the organism (Fig. 1). The S-layer forms an assembly of protein molecules that coat the outside of the cell, thus providing the cell membrane of *S. marinus* protection against potentially damaging solutes and macromolecules.

Tetrabrachion protein comprises an α-helical stalk of 70 nm in length that is anchored to the cell membrane at its C-terminal end, and an N-terminal domain consisting of four arms each approximately 24 nm in length formed of β-strands. The arms of the N-terminal domain form end to end contacts with the canopy-like meshwork of the *S. marinus* S-layer and give rise to a "quasi-periplasmic space" (Fig. 2).

The α-helical stalk of tetrabrachion comprises four copies of a "heavy" chain polypeptide that together form a parallel four-stranded α-helical coiled coil that is membrane-anchored (Fig 3(a)). At the top of the coiled coil portion, there is a hinge domain at which point the four heavy chains diverge from each other and the four N-terminal arms are formed of "light" chain polypeptides (Fig. 3(b)).

The S-layer stalk of tetrabrachion is not uniform throughout its length. For example, the first 130 amino acid residues after the hinge show a classical heptad repeat motif that is characteristic of "left handed coiled coils". However, after Pro1160, the heptad repeat is replaced by an undecad repeat (an 11 amino acid residue repeat) that results in the formation of a right handed coiled coil (RHCC) structure or domain. A protease known as STABLE (stalk-associated archaeabacteria endo-protease) binds to the RHCC domain of tetrabrachion. A major feature of the tetrabrachion stalk domain is its extreme thermostability even in the presence of 1% (w/v) sodium dodecyl sulfate (SDS), 6M guanidine, or 70% (w/v) sulfuric acid.

The RHCC structure of the stalk is a 52-amino-acid residue tetrameric bundled protein. (Fig. 4). The RHCC polypeptide chain fragment forms a parallel right-handed

coiled coil with an average length of about 72 Å and width of about 25 Å. It has a 11/3 residue repeat, with the N-terminal part more supercoiled than the C-terminal part due to the presence of a stutter between Ile11 and Thr16 (Fig. 5). There is also a unique 7,4 residue repeat.

The RHCC molecule contains four large globular cavities along the inside of the tetramer (Fig. 6), which have sizes in the range 150 $Å^3$ to 340 $Å^3$. The original X-ray structure of RHCC in the native state revealed that the cavities are occupied by water molecules. Because of the lack of buried polar groups and the resulting weak water-protein interactions, these water molecules are clustered into groups. Clusters of nine water molecules and five water molecules are found in cavities two and three, respectively. Cavity one at the N-terminus and cavity four at the C-terminus are occupied by two water molecules and one water molecule, respectively.

It is to be appreciated by those skilled in the art that a "fragment" of the tetrabrachion protein can be any portion or domain of the full-length tetrabrachion protein and may be, but is not limited to, the 52 amino acid residue of the RHCC structure of the tetrabrachion protein. The amino acid sequence of full-length tetrabrachion protein is known (Genbank Accession No. AAC44118), and is described as SEQ ID NO: 1 in the present disclosure. Therefore, a person skilled in the art would understand that the present disclosure contemplates the full-length amino acid sequence of SEQ ID NO: 1 (see Fig. 4) or any portion or domain of the full-length amino acid sequence.

In an aspect of the present disclosure, a use of tetrabrachion protein or a fragment thereof from *Staphylothermus marinus* for recovering a chemical element from a solution and/or suspension is provided.

It is to be understood that tetrabrachion protein and fragments thereof can be made using a variety of cell production systems, such as, but not limited to, mammalian cells, insect cells, yeast, for example, *Saccharomyces cerevisiae*, and bacteria such as but not limited to *Escherichia coli*, *Bacillus subtilis* or *Pseudomonas aeruginosa*.

If desired, the codons of the nucleotide sequence encoding tetrabrachion or a fragment thereof may be optimized for the host cell expressing the construct.

Transformation of a suitable host cell are well known in the art (for example, Maniatis et al, 1982, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, or Ausubel, *et al.* (eds), 1989, *Current Protocols in Molecular Biology*, Vol. 1, Green Publishing Associates, Inc., and John Wiley & Sons, Inc., New York) and can be accomplished by a variety of means well described in the art such as transfection, calcium chloride or other chemically induced transformation or electroporation. Regardless of the transformation method, once a modified host cell is generated it can be cultured and the tetrabrachion protein or fragment thereof can be expressed and isolated. In an alternative embodiment, the host cell expressing the tetrabrachion protein or fragment thereof can be used directly for detection and/or recovery of a chemical element from a solution and/or suspension as will be described in more detail below.

In another aspect of the present disclosure, a use of a composition comprising tetrabrachion protein or a fragment thereof from *Staphylothermus marinus* and a carrier therefor for recovering detection and/or recovery polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment, is provided. As used herein, the "carrier" may be, without limitation, a host cell, such as *Saccharomyces cerevisiae* or *Escherichia coli*, or a gel, matrix or other type of attachment surface. Furthermore, the composition of the present disclosure may comprise nanotubes composed of tetrabrachion protein or fragments thereof for detection and/or recovery of chemical elements from solutions and/or suspensions.

In another aspect of the present disclosure, a method for polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment, using tetrabrachion protein or a fragment thereof is provided. The method comprises: (a) introducing *S. marinus* or tetrabrachion protein or a fragment thereof from *S. marinus* or a composition comprising tetrabrachion protein or a fragment thereof from *S. marinus* into a solution; and (b) providing conditions that permit the tetrabrachion protein or the fragment thereof to detect and/or recover of the

chemical element from a solution and/or suspension. In the step of introducing (a), where tetrabrachion protein or a fragment thereof from *S. marinus* is introduced into the solution or suspension, the tetrabrachion protein or fragment thereof that is introduced can be made as described above. It is optional to express the tetrabrachion protein or a fragment thereof more than once, for example two times or three times, before introducing the expressed tetrabrachion protein or a fragment thereof into the solution or suspension.

In accordance with a further aspect of the present disclosure, there is provided a method for detecting and/or recovering polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment, wherein, in place of *S. marinus*, a different host cell transformed with a nucleic acid construct encoding tetrabrachion protein or a fragment thereof is introduced into the solution. Accordingly, there is provided a method for recovering polycyclic aromatic hydrocarbons and/or PCBs from a solution, wherein the method comprises: (a) introducing a nucleic acid construct encoding tetrabrachion protein or a fragment thereof into a host cell; (b) incubating the host cell under conditions that permit expression of the nucleic acid construct, thereby producing tetrabrachion protein or the fragment thereof; (c) introducing the host cell into the solution or suspension; and (d) providing conditions that permit the tetrabrachion protein or fragment thereof to detect and/or recover the polycyclic aromatic hydrocarbons and/or PCBs from the solution and/or suspension.

As provided above, if desired, the codons of the nucleotide sequence encoding tetrabrachion or a fragment thereof may be optimized for the host cell expressing the nucleic acid construct.

In one embodiment, the nucleic acid construct encoding tetrabrachion protein or a fragment thereof may comprise a nucleotide sequence encoding tetrabrachion or a portion thereof operatively linked to a regulatory region. In an exemplary embodiment, the nucleotide sequence may be the sequence according to SEQ ID NO: 3 (see Fig. 6), encoding the 52 amino acid residue sequence (SEQ ID NO: 2) (Fig. 5) of the RHCC structure of tetrabrachion, which is codon optimized for expression in *Escherichia coli*.

In an alternative embodiment, the nucleotide sequence may be a codon optimized sequence that encodes the full-length tetrabrachion protein according to the amino acid sequence of SEQ ID NO: 1 (see Fig. 4).

The nucleic acid construct of the present disclosure may be expressed in any suitable host cell that is transformed by the nucleotide sequence, or nucleic acid constructs, of the present disclosure. Examples of suitable host cells include, but are not limited to, bacterial cells such as *Escherichia* sp., *Pseudomonas* spp., *Bacillus* sp., yeast, such as *Saccharomyces* sp., *Schizosaccharomyces* sp. and *Candida* sp.

Recovery of polycyclic aromatic hydrocarbons and/or PCBs from a water system and/or a suspension and/or a sludge and/or a sediment is important for reducing damage to both the natural environment as well as to man-made environments.

Surprisingly, the inventors have discovered that the tetrabrachion protein or a fragment thereof from *Staphylothermus marinus* disclosed herein are useful for detection and recovery of recovering polycyclic aromatic hydrocarbons and/or PCBs from natural waterways, from municipal water supply systems, from industrial supply systems, from wastewater storage facilities, and from water purification and recycling facilities, using the methods disclosed herein.

Fig. 8 is a schematic illustration of a RHCC polypeptide chain fragment 10 of tetrabrachion expressed by the nucleotide sequence set forth in SEQ ID NO:3, having two pairs of naphthalene molecules 30 sequestered within the hydrophobic channel 20. Fig. 9 is an electron density presentation from a cross-section of a co-crystal structure of the RHCC polypeptide chain fragment 10 shown in Fig. 8, with two naphthalene molecules 30 within the hydrophobic channel 20.

Stock solutions of 20 µM naphtahlene/fluorene were prepared in a 10 mM bicine buffer (pH=8, I-154mM NaCl). To improve the solubility of fluorene, a pre-stock solution in 30v/v % ethanol was prepared. The solutions were incubated at room temperature for 30 minutes in 10mg/ml stock solutions of RHCC expressed by the nucleotide sequence set forth in SEQ ID NO:3. Steady-state fluorescence spectra were measured on a Fluorolog-3 Horiba Jobin Yvon spetrafluorometer (Edison, NJ).

The reaction was thermostatically controlled at 25C by a Jeio-Tech refrigerating bath circulator. The data shown in Fig. 7 show that increasing concentrations of the present RHCC polypeptide chain fragment of tetrabrachion expressed by the nucleotide sequence set forth in SEQ ID NO:3 sequestered increasing amounts of naphthelene.

Accordingly, the present disclosure can be used in relation to the detection and/or recovery of polycyclic aromatic hydrocarbons from waste water and/or slurries produced during recovery and processing of mineral ores, mine tailings and sludges from tailings ponds, from waste solutions recovered from other types of industrial processes. Examples of polycyclic aromatic hydrocarbons that can be detected and recovered by the methods disclosed herein include naphthalene, napthenic acids, anthracene, tetracene, phenanthrene, acenaphthene, acenaphthylene, fluorine, and their derivations.

For example, the present disclosure can be used in relation to the detection and/or recovery of PCBs from waste water and/or slurries produced during recovery and processing of mineral ores, mine tailings and sludges from tailings ponds, from waste solutions recovered from other types of industrial processes. . Examples of polycyclic aromatic hydrocarbons that can be detected and recovered by the methods disclosed herein include biphenyl, monochlorobiphenyl, dichlorobiphenyl, trichlorobiphenyl, tetrachlorobiphenyl, pentachlorobiphenyl, hexachlorobiphenyl, heptachlorobiphenyl, octachlorobiphenyl, nonachlorobiphenyl, and decachlorobiphenyl.

CLAIMS

What is claimed is:

1. Use of a tetrabrachion protein or a fragment thereof for detecting or recovering a polycyclic aromatic hydrocarbon and/or a PCB from a solution or a suspension.

2. Use according to claim 1, wherein the tetrabrachion protein or fragment thereof comprises an amino acid sequence that shares at least 85% sequence identity with one of SEQ NO ID: 1 or SEQ NO ID: 2.

3. A composition for detecting or recovering a polycyclic aromatic hydrocarbon and/or a PCB from a solution or a suspension, the composition comprising a tetrabrachion protein or a fragment thereof and a carrier therefor.

4. The composition of claim 3, wherein the tetrabrachion protein or fragment thereof comprises an amino acid sequence that shares at least 85% sequence identity with one SEQ NO ID: 1 or SEQ NO ID: 2.

5. Use of the composition of claim 3 for recovering a polycyclic aromatic hydrocarbon and/or a PCB from a solution or suspension.

6. Use of a microbial cell that expresses a polypeptide molecule comprising an amino acid sequence that shares at least 85% sequence identity with one of SEQ ID NO: 1 or SEQ ID NO: 2, for recovering a polycyclic aromatic hydrocarbon and/or a PCB from a solution or suspension.

7. A method for detecting or recovering a polycyclic aromatic hydrocarbon and/or a PCB from a solution or a suspension, the method comprising:

(a) commingling with and culturing therein the solution or the suspension the composition of claim 3, said composition comprising a plurality of right-handed coiled coil polypeptide chain fragments of tetrabrachion;

(b) separating a portion of the plurality of right-handed coiled coil polypeptide chain fragment of tetrabrachion from the solution or suspension; and

(c) assessing the separated portion of the plurality of right-handed coiled coil polypeptide chain fragments of tetrabrachion to detect a polycyclic aromatic hydrocarbon and/or a PCB therein.

8. The method of claim 7, additionally comprising the step of recovering the polycyclic aromatic hydrocarbon and/or the PCB from the separated portion of the plurality of right-handed coiled coil polypeptide chain fragments of tetrabrachion.

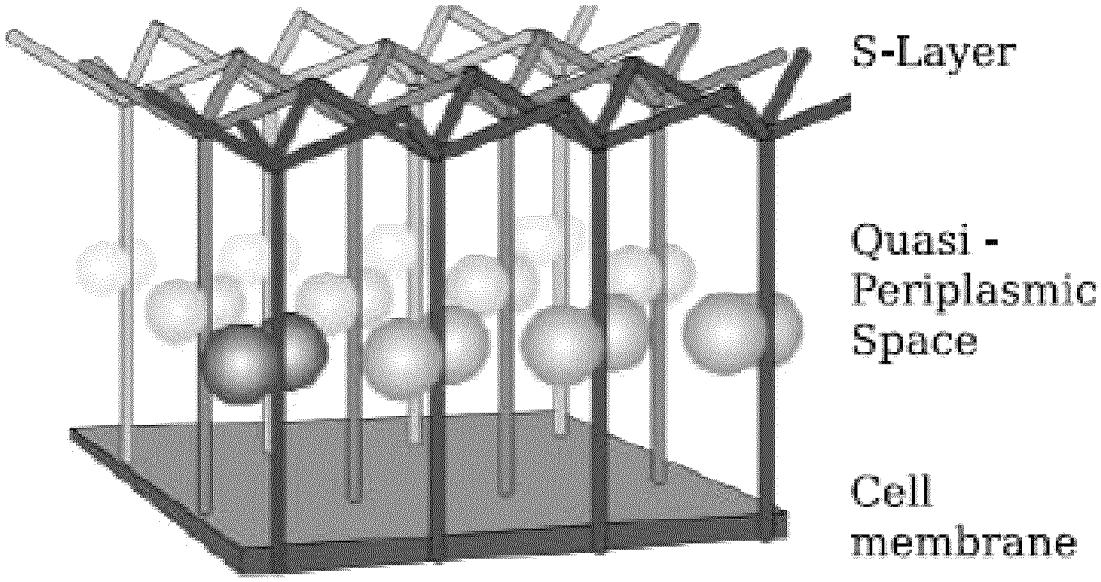

Fig. 1

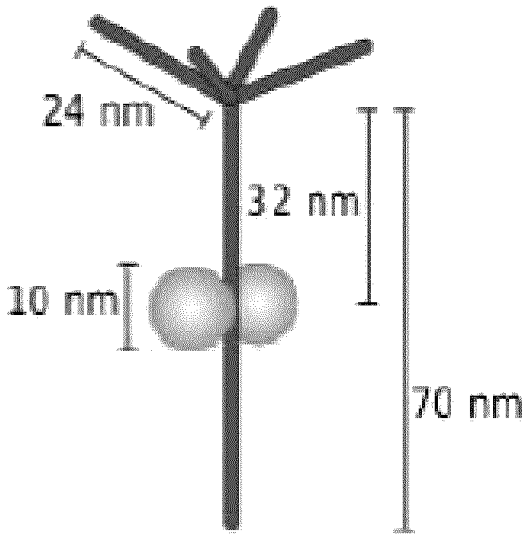

Fig. 2

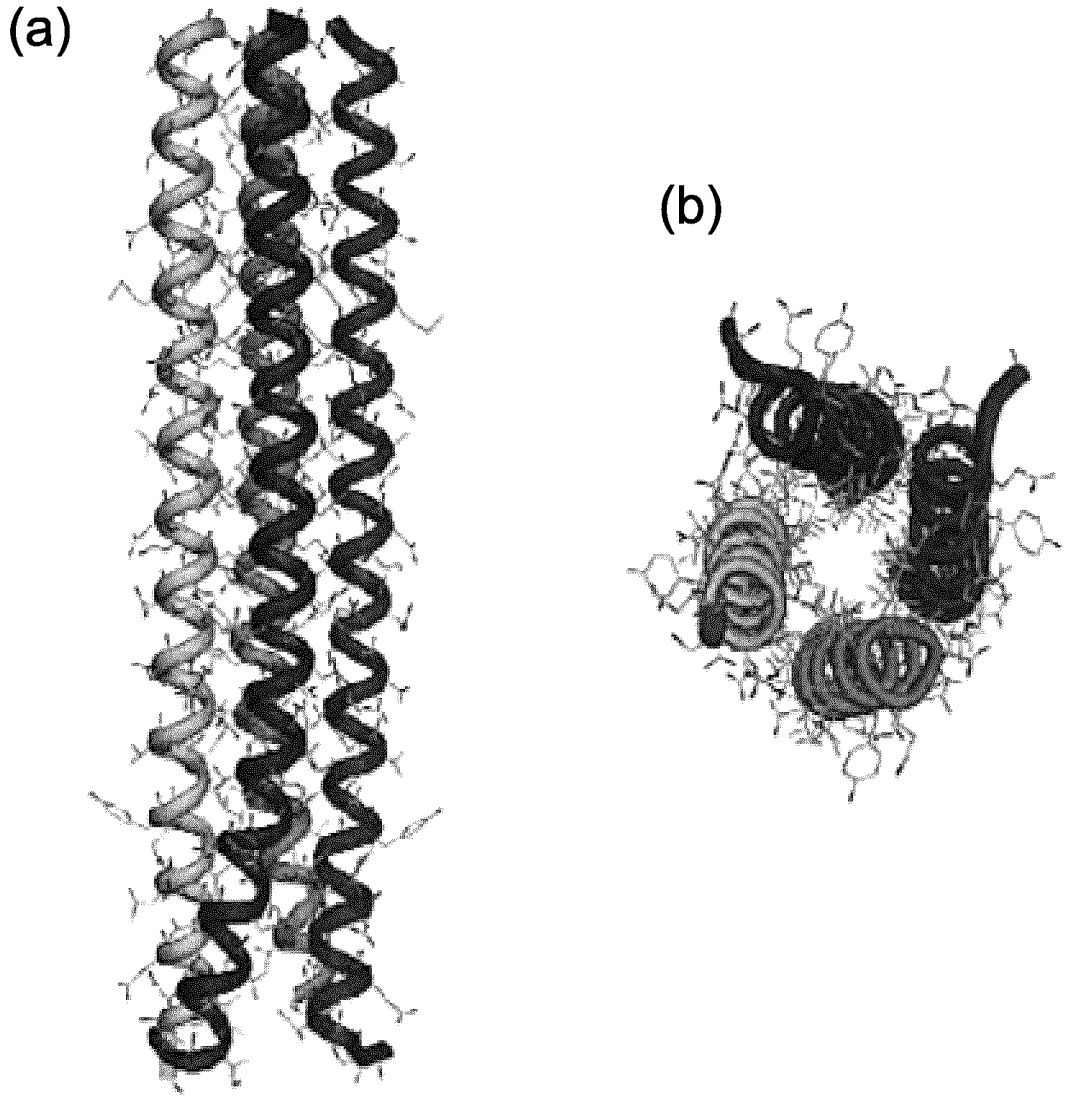

Fig. 3

```
   1 mnrvlaysll aimtlsllii papgiaqrit vgvsvkagty nfynitpttq tvevtdngml
  61 rviinrteat elgttirlaf ildtdkydpn vggyflnvsn igvyapsdpt qspyggvidi
 121 tqnstltdgt qvvgnvtvvn ggnnviilid lsklpdlqnv vyitnvytet sttanltntl
 181 lrvkafdaas wdavisgnqf kilyipslck yvkinvihsp aivgtnvdvi vsfhkyfslv
 241 qsiagvsldi tvdnktqlnm tnyynateny vlatfingnl tvggsevfss ptvtvinast
 301 fkysgqvkdy aptvadtatp wvrtlnkfev efrheivnxt hdlifyihcd sdtvsydtwp
 361 flivnasldi tttevafnst tinpgdivnf tahnvplqyl tatnygvlrf qlinpalvvy
 421 vpvsnmtlsa ntttgiings fvlpdapygg ldyltylvfn dgkfiangyi tvspcietyv
 481 ltntsayaed agssyigrfv pgytsvpgdy ivikgygfal snltgftvsi nntdviilna
 541 tynastgkii ilaklldtng tpipvgagfi rvgqngttni ayapfnvtrn sglekvlfnp
 601 rwfyngtyyi ehdklgdpyl yfpvdyplvn ntfttemwpf nttievigwp tntftlkafn
 661 kefnlsfnll tlsltngynm tnlynltipf lpygnytlle gtllsvnnrt vftvhmginv
 721 dldscgngtl sitvvgaapn teynftfgyq vhdlnygitr yispqwngtw nislvtdiyg
 781 tgstsvplit lyptsyvina twdvitwlrl sgsgtldllf svdvsyngft dnlttpityv
 841 fgpsdttpgs fniyvnttyn vsvvrvavdy lprtnvvisv petvlpgdti tvqifphhne
 901 vwgfieptal fdenqllgwy ltvrlvdpls ntvvervagy yagnlivedv dgdgdnevwf
 961 vvnltaplvl gvdktyrvdv elflavlnps snitgvtavd necyvqldln gtiywnglgs
1021 gimlggdgqi vtvlgvlegk ldtikdgiae inatvndint ylkvnvtdll ktinnsvvmi
1081 kndtatliig qaeikakldd llnltsqvnd tvtmilaccn naskvlnrme gtlnstytgv
1141 lnvksdlstl idtannvvip kfnelydnvt veinasrdli iqkissvnds lttiisagfn
1201 dveamisnln ttllnridel egtllfymta neqrlegiin etaddivyrl tviiddryes
1261 lknlitlrad rlemiindnv stilasignv nltvfnklnd leielgdvna tinagifqiq
1321 snlgnanqli ldtltsskve ilnaissnas aisseihnav nqlstlvlqv ndtltlkitg
1381 eadnilnfls slegsmntgf nnvtstlsav ennilgkitd tsnllsskid ntlstlqdli
1441 tstsndlkns issakndivs slsskvdsst qtlstklddl ksaqesntns innnimlfga
1501 aslillivti glvgyrliar rrvg (SEQ ID NO: 1)
```

Fig. 4

	a	b	c	d	e	f	g	h	i	j	k
1-3									G	S	I
4-11	I	N	E	T	A	D	D	I			
12-18					V	Y	R	L	T	V	I
19-29	I	D	D	R	Y	E	S	L	K	N	L
30-40	I	T	L	R	A	D	R	L	E	M	I
41-51	I	N	D	N	V	S	T	I	L	A	S
52	I										

Fig. 5

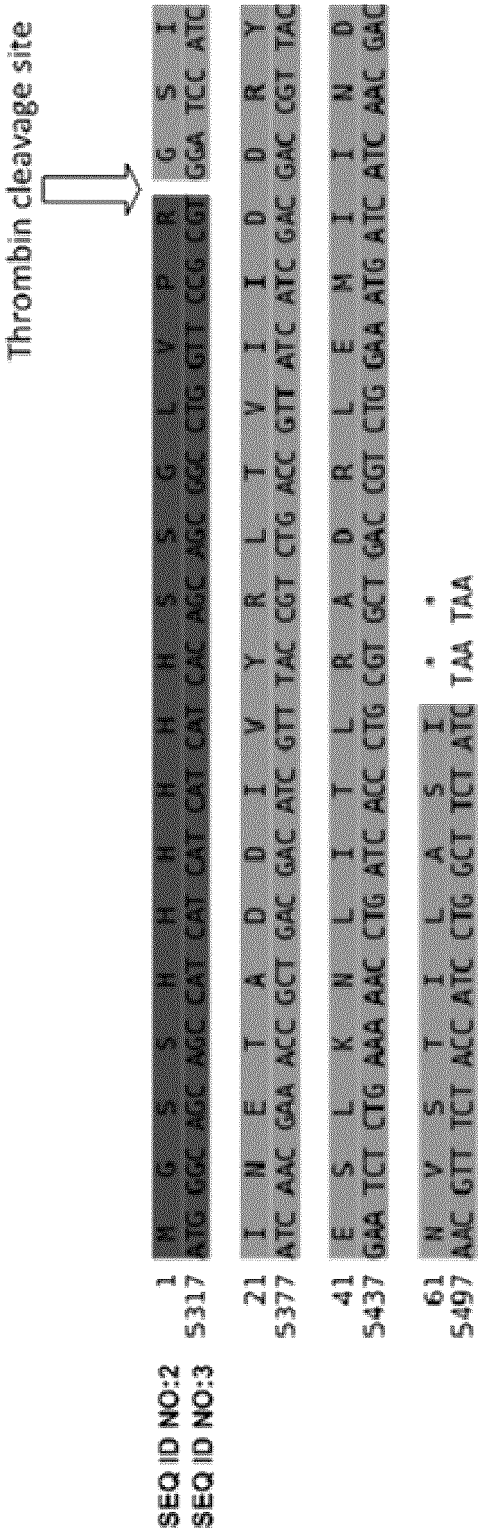
Thrombin cleavage site
M G S S H H H H H H S S G L V P R G S I
ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC AGC GGC CTG GTT CCG CGT GGA TCC ATC
I N E T A D D D I V Y Y R L T V I I D D R Y
ATC AAC GAA ACC GCT GAC GAC ATC GTT TAC CGT CTG ACC GTT ATC ATC GAC GAC CGT TAC
E S L K N L I T L R A D R L E N I I N D
GAA TCT CTG AAA AAC CTG ATC ACC CTG CGT GCT GAC CGT CTG GAA ATC ATC AAC GAC
N V S T I L A S I * *
AAC GTT TCT ACC ATC CTG GCT TCT ATC TAA TAA
SEQ ID NO:2 1
SEQ ID NO:3 5317
21
5377
41
5437
61
5497

Fig. 6

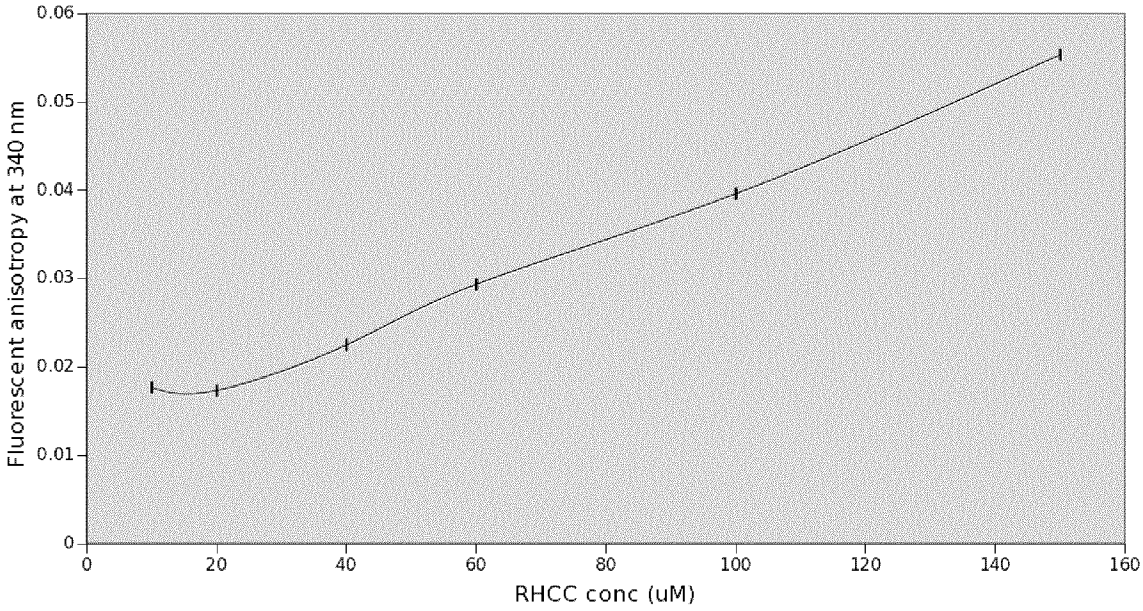
Fluorescent anisotropy at 340 nm
0.06
0.05
0.04
0.03
0.02
0.01
0
0
20
40
60
80
100
120
140
160
RHCC conc (uM)

Fig. 7

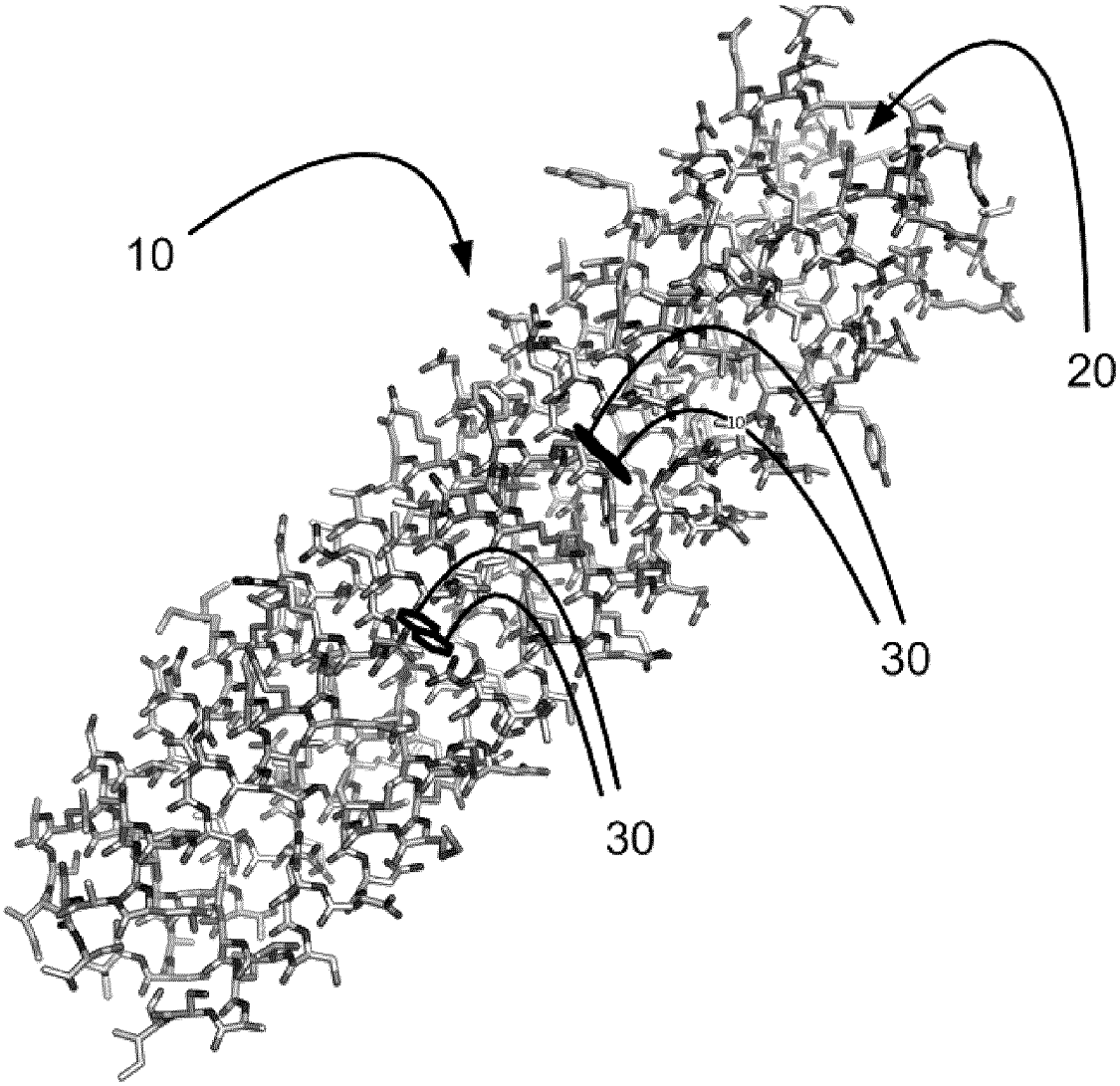

Fig. 8

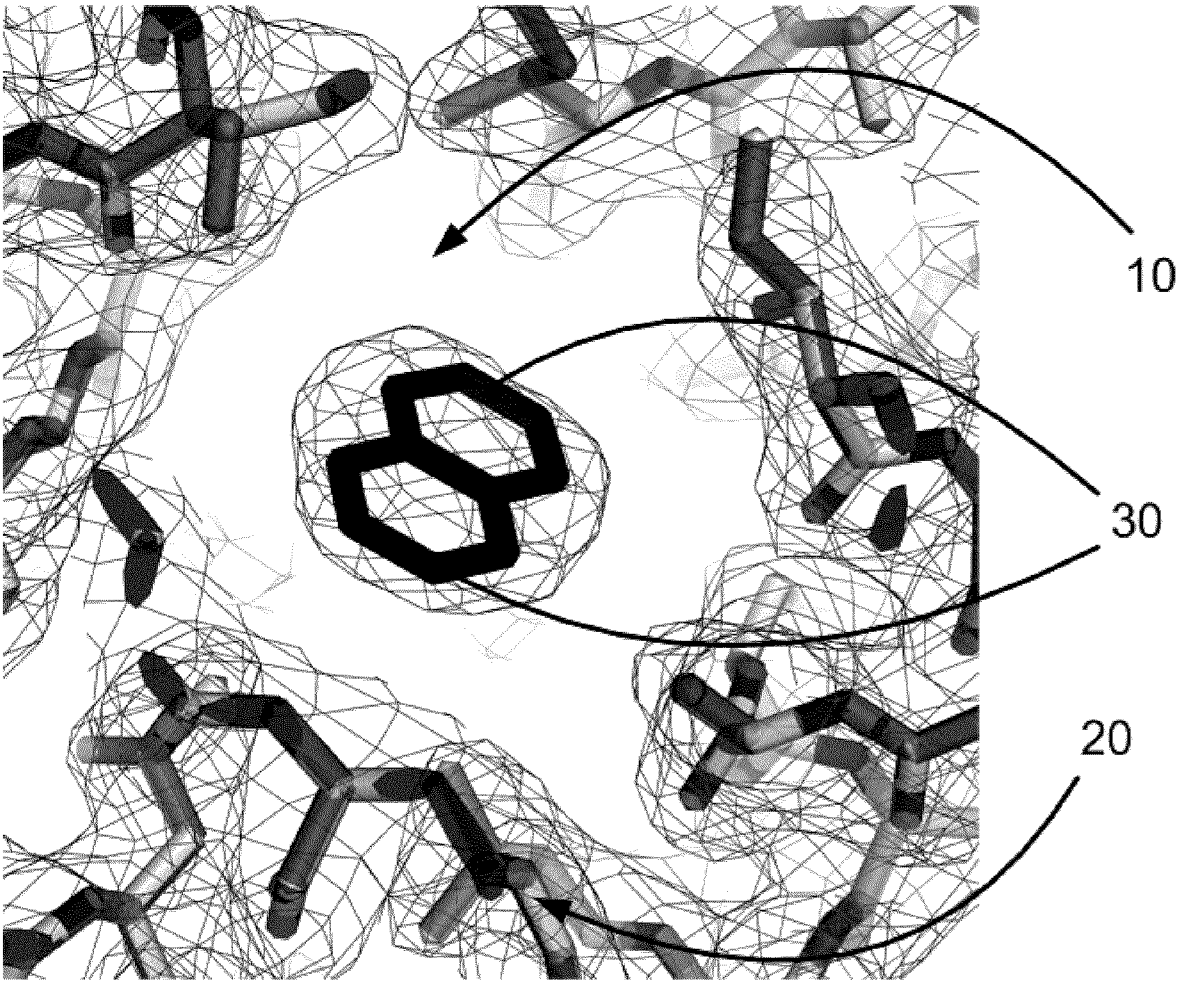

Fig. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2015/050373

A. CLASSIFICATION OF SUBJECT MATTER
IPC: ***G01N 33/18*** (2006.01) , ***B01D 15/08*** (2006.01) , ***C02F 1/28*** (2006.01) , ***C07C 17/38*** (2006.01) , ***C07C 51/42*** (2006.01) , ***C07K 14/195*** (2006.01) (more IPCs on the last page)

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: G01N 33/18 (2006.01) , B01D 15/08 (2006.01) , C02F 1/28 (2006.01) , C07C 17/38 (2006.01) , C07C 51/42 (2006.01) , C07K 14/195 (2006.01)

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

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)
(see first extra sheet)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
T, O	Protein Structure, Function and Malfunction (PSFaM) Meeting, 3rd Annual Meeting, May 7-8, 2015, P23.	
A	WO2013166585 A1 (STETEFELD J, *et al.*) 14 November 2013 (14.11.2013)	
A	WO2006112777 A2 (MUELLER J, *et al.*) 26 October 2006 (26.10.2006)	

☐ 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
"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 04 August 2015 (04-08-2015)	Date of mailing of the international search report 31 August 2015 (31-08-2015)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476	Authorized officer Christian Barrette (819) 994-0477

C12N 1/19 (2006.01) , *C12N 1/21* (2006.01)

(continuation of second sheet, section (b), Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used))

Type	Engine	Dbase or tool	Search string or citing(ed) document
Keyword	Questel -Orbit	Fampat	(tetrabrachion ET protei+)/TI/AB/IW/CLMS/DESC/ODES
Classification	Questel -Orbit	Fampat	(coiled_coil ET (naphtalene OU polycycl+ OU polyarom+ OU biphen+))/TI/AB/IW/CLMS/DESC/ODES/TX ET (G01N-033/18 OU B01D-015/08 OU C02F-001/28 OU C07C-017/38 OU C07C-051/42 OU C07K-014/195 OU C12N-001/19 OU C12N-001/21)/IPC
Inventor	Questel -Orbit	Fampat	((STETEFELD 3M Joerg)/IN/OIN/INH/INV OU (POLONENKO 3M Daniel)/IN/OIN/INH/INV)
Cited by	-	-	(examined application)
Keyword	Google scholar	w/o patents	tetrabrachion
	Google	-	tetrabrachion naphtalene
Author	Google scholar	w/o patents	author:"Joerg STETEFELD" OR author:"Daniel POLONENKO"
Keyword	Intellect	Canadian patents DB	tetrabrachion <in> description

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO2013166585A1	14 novembre 2013 (14-11-2013)	None	
WO2006112777A2	26 octobre 2006 (26-10-2006)	WO2006112777A2	26 October 2006 (26-10-2006)
		WO2006112777A3	07 June 2007 (07-06-2007)
		CN101222940A	16 July 2008 (16-07-2008)
		CN101222940B	08 June 2011 (08-06-2011)
		EP1871422A2	02 January 2008 (02-01-2008)
		EP1871422B1	16 October 2013 (16-10-2013)
		JP2008536916A	11 September 2008 (11-09-2008)
		JP5175714B2	03 April 2013 (03-04-2013)
		US2009214670A1	27 August 2009 (27-08-2009)